



AIDA – Anàlisi d'Impacte de la Demanda de l'Aerobús

Informe del projecte



CIDAI-PAI 2025-01

Drets reservats. Aquest treball està disponible sota la llicència Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0).

Segons els termes d'aquesta llicència, podeu copiar, redistribuir i adaptar l'obra amb fins no comercials, sempre que l'obra sigui citada adequadament, tal com s'indica a continuació.

En qualsevol ús d'aquest treball, no s'ha de suggerir que el CIDAI doni suport a cap organització, producte o servei específic. No es permet l'ús del logotip CIDAI. Si adapteu l'obra, heu de llicenciar-la amb la mateixa llicència Creative Commons o equivalent.

Si creeu una traducció d'aquest treball, heu d'afegir la següent exempció de responsabilitat juntament amb la cita suggerida: "Aquesta traducció no la va crear el Centre of Innovation for Data tech and Artificial Intelligence (CIDAI). CIDAI no es fa responsable del contingut ni de l'exactitud d'aquesta traducció. L'edició original en català serà l'edició autèntica i vinculant".

Qualsevol mediació relacionada amb disputes derivades de la llicència es durà a terme d'acord amb les normes de mediació de la World Intellectual Property Organization.

Cita suggerida. CIDAI-PAI 2025-01// Anàlisi d'Impacte de la Demanda de l'Aerobús (AIDA) per Xavier Conill i Ramon Pruneda - AMB Informació, Vanessa Moreno, Francisca Rosell - Eurecat, CIDAI, 2025. Llicència: CC BY-NC-SA 4.0

Abreviacions, acrònims i terminologia específica

A1	Línia d'autobús que connecta Plaça Catalunya i la terminal T1 de l'aeroport
A1IDA	Servei que fa el trajecte Plaça Catalunya-Aeroport T1
A1VUELTA	Servei que fa el trajecte Aeroport T1- Plaça Catalunya
A2	Línia d'autobús que connecta Plaça Catalunya i la terminal T2 de l'aeroport
A2IDA	Servei que fa el trajecte Plaça Catalunya-Aeroport T2
A2VUELTA	Servei que fa el trajecte Aeroport T2- Plaça Catalunya
AMB	Àrea Metropolitana de Barcelona
Boxplot	Tipus de gràfic estadístic que mostra en forma de caixa els quartils i permet representar els valor atípics
data sets	Conjunt de dades estructurades o no estructurades compost per diferent elements que poden ser tractades de forma conjunta
EDA	Anàlisi exploratori de dades (Exploratory Data Analysis)
Mapa de calor	Representació gràfica de les dades on els colors i diferents tonalitats dels colors representen els valors
NaN	Not a Number. S'aplica a un camp d'una estructura de dades quan s'espera un valor numèric i el camp està buit o el valor no és numèric
Outlier	Valor atípic dins del conjunt de dades
PAI	Projectes d'Alt Impacte
SAE	Sistema d'ajuda a l'explotació – eina d'AMB on estan emmagatzemades les dades dels serveis d'autobusos (validacions, hora de pas per parada, entre d'altres dades)
Web scraping	Procediments automatitzats per extraure dades d'una pàgina web

Índex

1. RESUM EXECUTIU

5

2. INTRODUCCIÓ

7

3. TRACTAMENT DE LES DADES

10

4. ANÀLISI DE DADES

27

5. ANÀLISI DE CUES

56

6. MODEL DE PREVISIÓ DE LA DEMANDA

69

7. MODEL D'OPTIMITZACIÓ DE LA FLOTA D'AUTOBUSOS

88

8. CONCLUSIONS

105

9. BIBLIOGRAFIA

108



Resum executiu



El projecte 'Anàlisi d'impacte de la demanda de l'Aerobús', AIDA, s'ha desenvolupat en el marc dels projectes d'alt impacte del CIDAI (*Centre of Innovation in Data technologies and Artificial Intelligence*) amb el finançament de la Generalitat de Catalunya, dintre de les oportunitats relacionades amb la planificació i millora dels sistemes de Transport Públic Metropolità. Els centres/entitats/organitzacions que col·laboren en aquest projecte han sigut AMB Informació i Serveis, l'Ajuntament de Barcelona i Eurecat.

L'Aerobús, servei de titularitat de l'Àrea Metropolitana de Barcelona (AMB), i operat per l'empresa Monbus, dona servei a la línia que connecta el centre de Barcelona i l'aeroport. És un servei 24x7, definit per la freqüència de pas, amb alts nivells de demanda i lligat a una certa estacionalitat. Diferents factors externs poden tenir impacte en la demanda, tal i com poden ser les temporades turístiques, les arribades i sortides d'avions, els grans esdeveniments que tenen lloc a Barcelona i la seva àrea metropolitana, o inclús, la pluja.

L'objectiu del projecte és desenvolupar un model combinat de la predicció de la demanda a curt-mig termini, amb un model d'optimització per fer una estimació de la flota de vehicles necessària per cobrir la demanda, respectant la freqüència de pas establerta. Per validar els models resultants, es prepara una prova de concepte que permeti comprovar la utilitat de la solució tecnològica proposada.

El projecte s'ha abordat en diferents fases. En primer lloc, s'ha fet una preparació de les dades disponibles, recollint-les, processant-les i tractant-les per garantir la fiabilitat dels resultats. A continuació, s'ha procedit a analitzar les dades aplicant diferents eines estadístiques, el que ha permès detectar patrons i correlacions que faciliten la creació del model predictiu i asseguruen uns resultats robustos i fiables. A més a més, els resultats de l'anàlisi permeten detectar punts de millora en els serveis, detectant per exemple possibilitats d'ajustar la freqüència de pas a la demanda, per exemple per necessitats de reforçar o reduir el servei en determinats períodes i hores del dia.

Un cop analitzades les dades, s'ha treballat en el model de predicció, investigant diferents algorismes i alternatives i seleccionant el que millor s'ha comportat segons les dades disponibles. En aquest cas, ha sigut el model generat fent servir un algorisme de tipus 'Gradient Boosting Regressor',

basat en arbres de decisió el que ha donat una millor precisió en el curt-mig termini (entre 2 dies i quatre setmanes).

Finalment, s'ha desenvolupat un model matemàtic de programació lineal que permet optimitzar el nombre de vehicles necessaris per donar servei a la demanda prevista, tot garantint la freqüència establerta. El mateix algorisme també permet identificar quina seria la freqüència mínima de pas per cobrir la demanda prevista. Aquest resultat pot ajudar als gestors a l'hora de definir les necessitats dels serveis.

Tots els participants al projecte van signar un Acord de Confidencialitat per garantir la confidencialitat durant el transcurs del projecte i protegir les dades d'ús privat. AMB informació i serveis va facilitar les dades d'ús de l'aerobús. Eurecat ha definit els models, fet el procés de tractament de les dades i desenvolupat i testejat els algorismes.



Introducció



Aquest projecte està emmarcat en els Projectes d'Alt Impacte (PAI) dintre de les oportunitats amb la planificació i millora dels sistemes de Transport Públic Metropolità. Els centres/entitats/organitzacions que col·laboren en aquest projecte són:

- AMB Informació i Serveis
- Ajuntament de Barcelona
- Eurecat



2.1. Antecedents

El servei Aerobús, dissenyat i licitat per l'Àrea Metropolitana de Barcelona (AMB) i operat per l'empresa Monbus, connecta el centre de Barcelona amb les dues terminals de l'Aeroport Josep Tarradellas Barcelona – El Prat. El servei consta de dues línies d'autobús, A1 i A2, que connecten el centre de la ciutat amb les terminals T1 i T2, respectivament. Ambdues línies tenen actualment tres parades per recollir passatgers en sentit aeroport: Plaça Catalunya, Carrer Sepúlveda i Plaça Espanya. En sentit centre de la ciutat, la línia A1 només admet passatgers a l'origen de línia (T1 Planta 0), mentre que la línia A2 recull passatgers a dues parades (T2B i T2C), tot i que l'origen és el principal punt d'admissió.

Els horaris de les dues línies cobreixen les 24 hores del dia, tots els dies de la setmana i estan definits per la freqüència de pas. Actualment, la línia A1 ofereix una freqüència de pas de 5 minuts durant la major part del dia i d'entre 10 i 20 minuts durant la nit. Per la seva part, la línia A2 té una freqüència de pas de 10 minuts durant la major part del dia i de 20 minuts durant la nit.

Actualment, el dimensionament de la flota que permet operar el servei està ajustada a la freqüència de pas i a la demanda prevista. No obstant, algunes circumstàncies de l'entorn poden impactar sobre la demanda, com poden ser l'estacionalitat del turisme, la freqüència d'arribades i sortides de vols a l'aeroport, la meteorologia o els grans esdeveniments celebrats a Barcelona, ja que aquests últims atrauen un major flux de visitants a la ciutat, generen una major demanda de mitjans de transport alternatius i provoquen més afectacions al trànsit. Les característiques de les línies de l'Aerobús permeten que l'operador del servei pugui respondre de forma àgil a canvis que es puguin produir en l'entorn i que puguin afectar a la qualitat del servei. reveure amb un cert marge de temps aquests canvis pot ajudar a ajustar millor l'oferta a la demanda, i mantenir el nivell de servei esperat.

2.2. Descripció del projecte

L'objectiu del projecte és analitzar la demanda actual del servei d'Aerobús, estudiant el possible impacte d'efectes externs sobre la demanda del servei, per poder ajustar la flota de vehicles a la demanda esperada en el curt i mig termini (no en temps real). El resultat del projecte serà un prototipus de model combinat de predicció + optimització.

El projecte consta de tres fases: un primera fase d'anàlisi de dades, una segona on es definirà i construirà el model de predicció i una tercera per definir i construir el model d'optimització.

Inicialment, es planteja fer servir les següents dades per tal d'esbrinar si tenen o no impacte en la demanda dels serveis de l'Aerobús i, en cas afirmatiu, dimensionar el seu impacte per fer una previsió més ajustada de la demanda:

- Dades històriques d'ús del servei d'Aerobús; proporcionades per l'AMB.
- Dades històriques (2022-2024) i futures dels horaris d'arribades i sortides de vols a l'Aeroport Josep Tarradellas Barcelona – El Prat, amb orígens i destinacions, tant de la terminal T1 com de la T2; proporcionades per AMB Informació a través d'un proveïdor extern.
- Estacionalitat temporal, incloent informació sobre dies festius i períodes vacacionals.
- Calendari de grans esdeveniments a la província de Barcelona dels darrers dos anys (2022-2023) i l'actual (2024); proporcionades per AMB Informació.
- Dades meteorològiques (precipitació) a la ciutat de Barcelona i entorn –històriques i previsions; obtingudes de fonts d'accés públic, com OpenWeather (openweathermap.com) o Open-Meteo (open-meteo.com)

Durant la fase analítica s'aniran refinant les dades necessàries i les disponibles.

Un cop determinats els factors que influeixen en l'ús del servei d'Aerobús i el seu grau d'impacte en la demanda, es construirà un model d'optimització de la flota de vehicles a partir de la previsió de la

demanda. Aquest model ha de permetre ajustar les necessitats del servei en el curt termini.

Durant l'execució del projecte, s'acabarà de definir l'interval de temps a partir del qual caldria realitzar la previsió de la demanda per tal d'anticipar-se a les necessitats del servei a partir de l'optimització de la flota.

Sense necessitat de donar resposta en temps real, els models de previsió + optimització haurien de permetre obtenir solucions en cas de canvis en les condicions inicials (principalment, per canvis en les condicions esperades de meteorologia o celebració d'esdeveniments inesperats).

2.3. Organització del document

El present document recull el treball realitzat en el marc del projecte i detalla els resultats obtinguts. Els continguts s'organitzen de la forma que es descriu a continuació.

En primer lloc, al Capítol 1 es resumeix el treball realitzat i es sintetitzen els principals resultats obtinguts. Al Capítol 2 apareix una llista de les abreviatures i acrònims fets servir, així com termes especials utilitzats en el document. El Capítol 3 presenta la introducció al projecte i les tasques realitzades. En el Capítol 4 s'explica tot el procés fet de preparació de les dades, mentre que al Capítol 5 es detalla l'anàlisi de les dades. En el Capítol 6 es descriu l'anàlisi de les dades de les cues: en ser dades mostrals, s'ha optat per analitzar-les de forma aïllada a la resta. El Capítol 7 explica el procés de construcció del model de previsió de la demanda, i mostra alguns exemples amb les dades disponibles en el moment de preparar l'informe. El Capítol 8 està destinat a descriure el model d'optimització dissenyat per estimar el nombre mínim necessari de vehicles per cobrir la demanda i la freqüència, i presenta els resultats de les proves realitzades amb els exemples de l'estimació de la demanda del Capítol 7. Per últim, al Capítol 9 es presenten les conclusions.

Tractament de les dades



3.1. Introducció

En aquesta fase del projecte ens centrem en l'anàlisi detallada de les dades recollides amb l'objectiu de comprendre els factors que influeixen en la demanda de l'Aerobús.

Per realitzar l'anàlisi de dades, en primer lloc, es realitza una descripció detallada de les dades disponibles, analitzant les variables inicialment proposades i justificant la selecció de les que finalment s'utilitzaran per a l'estudi. Aquesta selecció es basa en criteris de rellevància, qualitat i adequació a les necessitats específiques de l'anàlisi.

Un cop seleccionades les variables, es procedeix a la neteja de les dades, un procés fonamental per garantir la fiabilitat dels resultats. Aquesta etapa implica la identificació i tractament de valors faltants, la correcció d'errors i la transformació de les dades, si escau, per adaptar-les al model analític.

Tots aquests passos permetran realitzar una exploració inicial de les dades que permet identificar patrons, distribucions i relacions entre les variables, i que es descriu en el capítol 5. Aquest pas és crucial per establir les bases de l'anàlisi posterior i assegurar que es compta amb una comprensió sòlida del conjunt de dades abans de procedir amb models més complexos.

3.2. Descripció de les dades

3.2.1. Data sets i origen

Inicialment, es van considerar diversos conjunts de dades amb l'objectiu d'esbrinar si tenien o no impacte en la demanda del servei d'Aerobús i, en cas afirmatiu, dimensionar aquest impacte per realitzar una previsió més ajustada de la demanda. La selecció inicial d'aquestes dades es basa en la hipòtesi de què poden tenir una relació significativa amb la nostra variable objectiu. Concretament, les dades inicialment plantejades per a l'estudi són:

- Dades històriques d'ús del servei d'Aerobús, a partir de les dades del Sistema d'ajuda a l'explotació (SAE) de l'operador i de les dades de què disposa l'AMB.
- Previsió de la demanda de l'Aerobús, proporcionada per l'operador o per l'AMB.
- Dades històriques i futures dels horaris d'arribades i sortides de vols a l'Aeroport Josep Tarradellas Barcelona –El Prat, amb orígens i destinacions, tant de la terminal T1 com de la T2 (dades històriques i futures); proporcionades per AMB a partir de les dades adquirides des de la web AeroDataBox (Timber Bytes LLP, 2024).
- Estacionalitat temporal, incloent informació sobre dies festius i períodes vacacionals.
- Calendari de grans esdeveniments celebrats a la província de Barcelona dels darrers dos anys (2022-2023) i l'actual (2024); proporcionades per AMB Informació.
- Dades meteorològiques (precipitació) a la ciutat de Barcelona i entorn –històriques i previsions, obtingudes de fonts d'accés públic, com ara OpenWeather (Openweather, 2024) o Open-Meteo (Open-Meteo, 2024)

Durant el procés d'obtenció de dades, es va descobrir que alguns conjunts de dades inicialment previstos no estaven disponibles o que hi havia altres dades no considerades inicialment que també podien tenir impacte en la demanda del servei.

A continuació, es detallen els *data sets* finalment utilitzats, incloent el seu origen i les seves característiques principals, així com una anàlisi preliminar per assegurar la seva adequació als objectius del projecte.

Data set 1: Metro

- **Origen del data set:** Aquest data set ha estat creat manualment a partir de la informació proporcionada a la pàgina web oficial de Transports Metropolitans de Barcelona (TMB); concretament, a la secció d'horaris del metro (TMB, 2024)
- **Descripció general:** El data set conté informació sobre els horaris d'operació del metro de Barcelona per dia de la setmana.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és analitzar les dades dels horaris del metro per determinar si tenen una influència significativa en la demanda del servei. En concret, es pretén avaluar si els horaris del metro poden ser considerats com una variable rellevant en el model de predicció de la demanda.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
date	Data	Data	[2022-01-01 al 2024-07-31]	Format YYYY-MM-DD
tipo_de_dia	Categoria de dia per assignar l'horari de metro	Categòrica	[dilluns a diumenge, festiu, vigília de festiu]	
hora_inicio	Hora en què comença el servei	Hora	05:00, 06:00, etc.	
hora_fin	Hora en què finalitza el servei	Hora	23:00:00, 23:59:00, etc.	Format HH:MM

Data set 2: Esdeveniments

- **Origen del data set:** Data set proporcionat per AMB Informació, creat manualment per AMB Informació a partir de les dades dels informes dels anys corresponents "Perspectives de l'activitat turística a la Destinació Barcelona" (Observatori de Turisme a Barcelona, 2024) i aportacions pròpies.
- **Descripció general:** El data set conté informació sobre els esdeveniments rellevants a la província de Barcelona dels anys 2022, 2023 i 2024.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és analitzar les dades dels esdeveniments per determinar si tenen una influència significativa en la demanda del servei. En concret, es pretén avaluar si els esdeveniments poden ser considerats com una variable rellevant en el model de predicció de la demanda.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
esdeveniment	Nom de l'esdeveniment	Text	La Mercè / BAM, Cavalcada dels Reis Mags...	
ubicació	Lloc de l'esdeveniment	Text	Poblenou, Circuit de Catalunya...	
municipi	Municipi de l'esdeveniment	Text	Barcelona, Montmeló	
tipus	Categoria de l'esdeveniment	Categòrica	Art, Concert, Festival...	
data inici	Data inici de l'esdeveniment	Data	17/09/2023...	Format DD-MM-YYY
data fi	Data fi de l'esdeveniment	Data	17/09/2023...	Format DD-MM-YYY
aforament estimat	Aforament estimat per dia	Numèrica	49500, 50000...	

Data set 3: Horari de l'Aerobús

- **Origen del data set:** Aquest data set ha estat creat manualment amb la freqüència de pas de l'Aerobús proporcionada per l'AMB.
- **Descripció general:** La informació del data set s'ha dividit per línia:
 - Línia A1 Anada / Barcelona – Aeroport Terminal 1
 - Línia A1 Tornada / Aeroport Terminal 1 – Barcelona
 - Línia A2 Anada / Barcelona – Aeroport Terminal 2
 - Línia A2 Tornada / Aeroport Terminal 2 – Barcelona
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és analitzar les dades dels esdeveniments per determinar si tenen una influència significativa en la demanda del servei. En concret, es pretén avaluar si els esdeveniments poden ser considerats com una variable rellevant en el model de predicció de la demanda.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
abril-octubre	Inclou l'hora de pas de cada autobús en horari d'estiu	Hora	1:09:00, 1:29:00...	Format HH:MM
novembre-març	Inclou l'hora de pas de cada autobús en horari d'hivern	Hora	5:49:00, 3:49:00...	Format HH:MM

Data set 4: Precipitació

- **Origen del data set:** Aquest data set s'ha obtingut fent *web scraping* de la pàgina web del Meteocat; concretament, de les dades històriques de l'estació del Prat de Llobregat (Meteocat, 2024).
- **Descripció general:** El data set conté la informació de la quantitat de precipitació acumulada (mm) cada 30 minuts.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és analitzar les dades de pluja per determinar si tenen una influència significativa en la demanda del servei. En concret, es pretén avaluar si les condicions de pluja poden ser considerats com una variable rellevant en el model de predicció de la demanda.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
data	Data	Text	2022-01-01T00:00Z, 2022-01-02T00:00Z	Format YYYY-MM-DD+'T00:00Z'
precipitació	Cadena on s'inclou les dades de precipitació acumulada cada 30 minuts.	Array	"0.0, 0.0, 0.0, 0.0,..."	Array

Data set 5: Expedicions

- **Origen del data set:** Aquest data set s'ha obtingut del servidor de dades de Àrea Metropolitana de Barcelona (AMB).
- **Descripció general:** El data set conté la informació de cadascuna de les parades realitzades per l'Aerobús durant el servei.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és extreure la variable "tripcode" i incloure-la al data set principal per identificar la ruta específica del servei. Això permetrà analitzar la relació entre la ruta del servei i la demanda.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
date	Data del servei. Pot no coincidir amb la data d'inici d'expedició, per exemple per les línies nocturnes, en les que la expedició comença a la matinada del dia següent.	Data	2023/06/28...	Format YYYY/MM/DD
sae	Identificador del SAE	Categòrica	[4]	Valor únic
concession	Identificador de la concessió	Categòrica	[AEROBUS]	Valor únic
agencycode	Identificador de l'operador	Numèrica	[49]	Valor únic
linename	Nom comercial de la línia (nom de la línia, per exemple LH1)	Text	['A1' 'A2']	
routeid	Identificador de la ruta. Ha de coincidir amb el valor passat en els serveis de temps real de pas	Categòrica	[2969 2973 2968....]	
tripcode	Identificador del viatge.	Text	['A1VUELTA0246' 'A1VUELTA2234' 'A2VUELTA1348' ... 'A2VUELTA1250']	
tripid	Codi identificador únic de cada viatge	Numèric	[81432 81660 82091 ... 892401850 81099956 102376956]	
theoreticaldistance	Kilòmetres calculats de la expedició	Numèric	[17.07540131 13.82209969 17.63630104]	km
theoreticalspeed	Velocitat mitja teòrica (km/h)	Numèric	[37.94533539 30.7157753 40.69915466]	
theoreticalstarttimestamp	Data i hora planificada en la que el bus ha d'arribar a la parada.	Numèric	[1.64100156e+09 1.64098644e+09 1.64095488e+09]	
realdistance	Kilòmetres reals recorreguts	Numèric	[17.07540131 13.82209969 17.63630104]	
realstarttimestamp	Data i hora real en la que el bus ha arribat a la parada.		[1.64100156e+09 1.64098644e+09 1.64095488e+09]	
realspeed	Velocitat mitja real (km/h)		[37.94533539 30.7157753 40.69915466]	
type		Categòrica	[1]	Valor únic
realspeed	Velocitat mitja real (km/h)		[37.94533539 30.7157753 40.69915466]	
calendar	Calendari associat	Categòrica	T-Baja Nov 2022 Viernes,	

Data set 6: Expedició-parada

- **Origen del data set:** Aquest data set s'ha obtingut del servidor de dades de l'AMB.
- **Descripció general:** El data set conté la informació de cadascuna de les parades realitzades per l'Aerobús durant el servei, incloent el volum de passatgers que pugen i baixen a cada parada.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és el principal i conté la variable a predir: el nombre d'usuaris que pugen a cada parada, que representa la demanda del servei. Inicialment, el data set s'utilitzarà per fer un Anàlisi Exploratori de Dades (EDA), que permetrà comprendre el funcionament del servei, identificar patrons i detectar possibles anomalies. En un segon pas, les dades serviran per construir un model de predicció de la demanda del servei, integrant variables com la pluja, els esdeveniments especials i els horaris del metro, per tal de millorar la gestió i planificació de la xarxa de transport.

Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
referencestop	Indica si es parada de referència	Binària	[True False]	
theoreticaldeparture	Data i hora planificada en la que el bus ha d'arribar a la parada.	Timestamp Unix	[1674197760, 1674198060 ...]	
realdeparture	Data i hora real en la que el bus ha sortit de la parada	Timestamp Unix	[1.67419785e+09 1.67419813e+09...]	
theoreticalarrival	Data i hora planificada en la que el bus ha d'arribar a la parada.	Timestamp Unix	[1674197760 1674198060	km
theoreticaltimeinstop	Temps planificat que es troba en espera el bus en la parada (temps de regulació)	Timestamp Unix	[0.]	
realarrival	Data i hora real en la que el bus ha arribat a la parada.	Timestamp Unix	[1.64100156e+09 1.64098644e+09 1.64095488e+09...]	
usersup	Passatgers reals que han pujat al bus	Numèric	[0. nan 29. 26...]	
usersdown	Passatgers reals que han baixat del bus.	Numèrica	[0. 2. 4. 1. nan 9...]	
complete	Indica el percentatge de l'ocupació. Valors admesos de 0-100. Si no es disposa de la informació, el camp estarà en blanc.	Numèrica	[7.78 5.56 1.11 0. nan...]	
sae	Identificador del SAE	Catègorica	[4]	Valor únic
agencycode	Identificador de l'operador	Numèrica	[49]	Valor únic
concession	Identificador de la concessió	Catègorica	[AEROBUS]	Valor únic

Data set 7: Expedició-parada

- **Origen del data set:** Aquest data set s'ha obtingut del servidor de dades de l'AMB.
- **Descripció general:** El data set conté les característiques dels autobusos que operen el servei.
- **Objectiu d'ús:** L'objectiu d'ús d'aquest data set és extreure la variable "seat" i incloure-la al data set principal per assignar a cada servei la quantitat de seients de l'autobús. Aquesta dada, es el valor màxim d'ocupació de l'autobús i servirà per calcular l'ocupació mitjana dels autobusos a l' Anàlisi Exploratori de Dades.

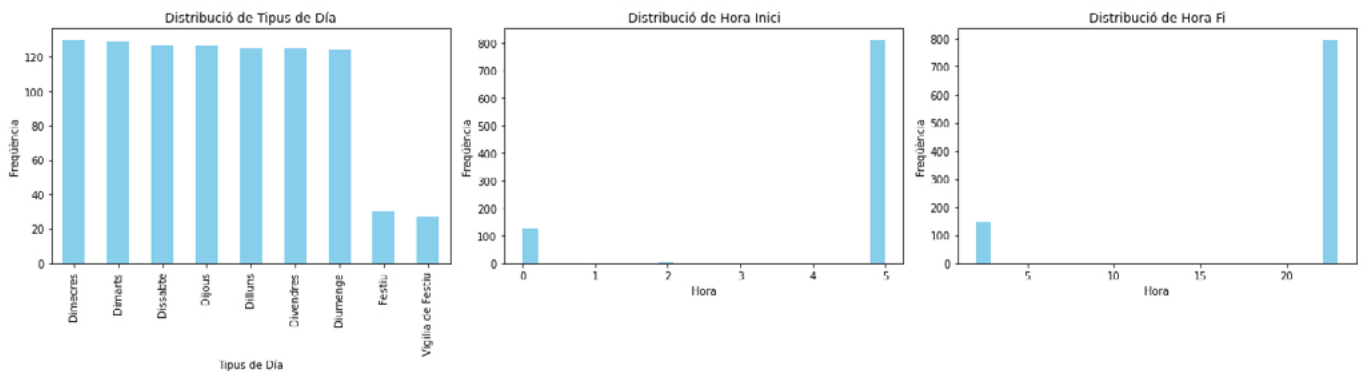
Taula de variables:

Variable	Descripció	Tipus dada	Valors possibles/Rangs	Comentaris
id	Identificador únic bus	Categòrica	[2772 2773 2774 ...]	
code	Codi utilitzat per l'operador. Calca del vehicle (podria coincidir amb l'id de vehicle).	Categòrica	[3340 3350 3336 ...]	
wheelchair	Plaques per cadira de rodes	Numèric	[1.]	Valor únic
plate	Matrícula del vehicle	Alpha-numèric	['4132LRS' '4138LRS' ...]	
seat	Plaques per seure	Numèric	[49. 48.]	
stand	Places per anar de peu	Numèric	[41. 59.]	
distanceout	Kilòmetres acumulats realitzats fora de ruta. Per exemple sortida o tornada a cotxeres	Numèric	[472.21665005 579.01120103 ...]	
distancein	Kilòmetres acumulats realitzats en ruta	Numèric	[472.21665005 579.01120103 ...]	
sae	Identificador del SAE	Categòrica	[4]	Valor únic
agencycode	Identificador de l'operador	Numèrica	[49]	Valor únic
concession	Identificador de la concessió	Categòrica	[AEROBUS]	Valor únic

3.2.2. Anàlisi inicial

A continuació, realitzarem una visualització general de la distribució dels valors en els data sets principals, mitjançant gràfics, amb l'objectiu de detectar possibles errors o anomalies en les dades.

Data set 1: Metro

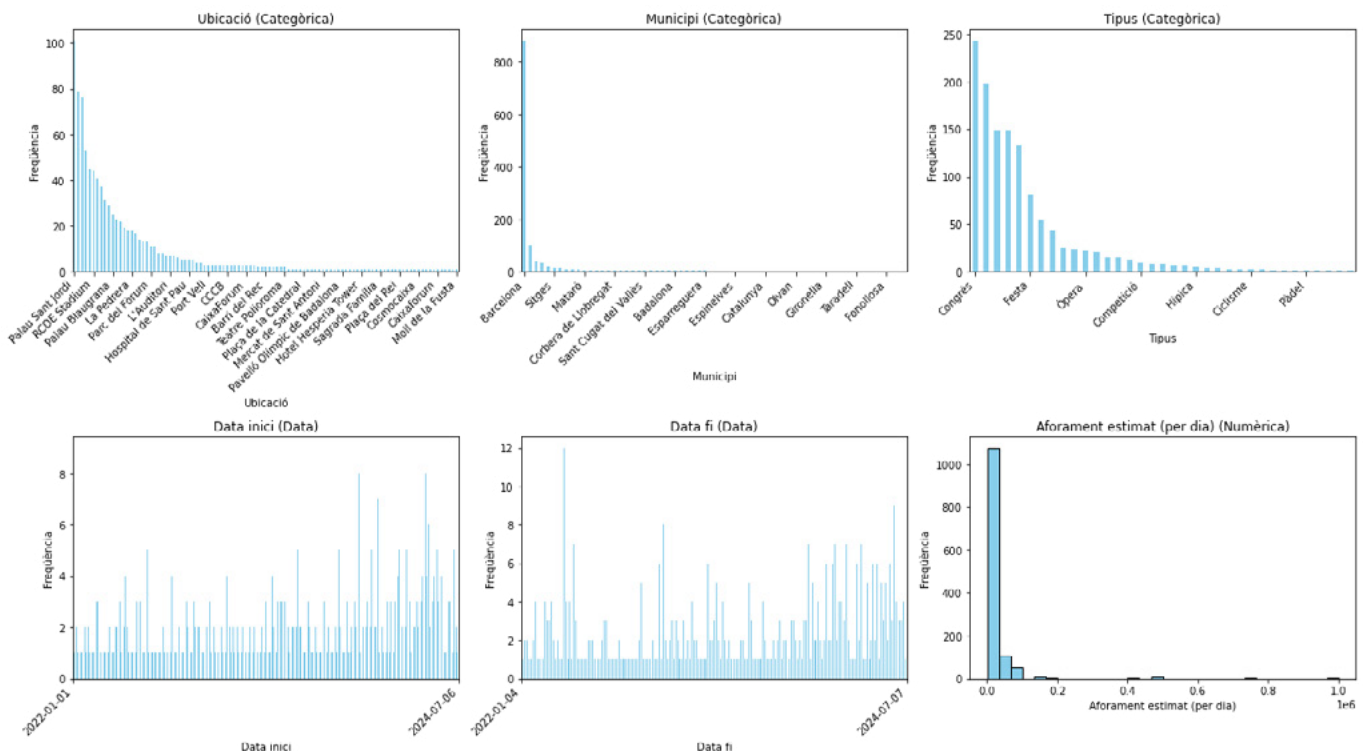


Observem que els dies de la setmana presents al *data set* es distribueixen de dilluns a divendres en quantitat proporcional i després tenim, en menor proporció, festius i vigílies de festiu.

L'hora d'inici és a les 0:00h i a les 5:00h del matí (majoritàriament, aquest segon) i l'hora de fi, a les 23:59h i a les 02:00h.

No es detecten inconsistències als valors del *data set*.

Data set 2: Esdeveniments

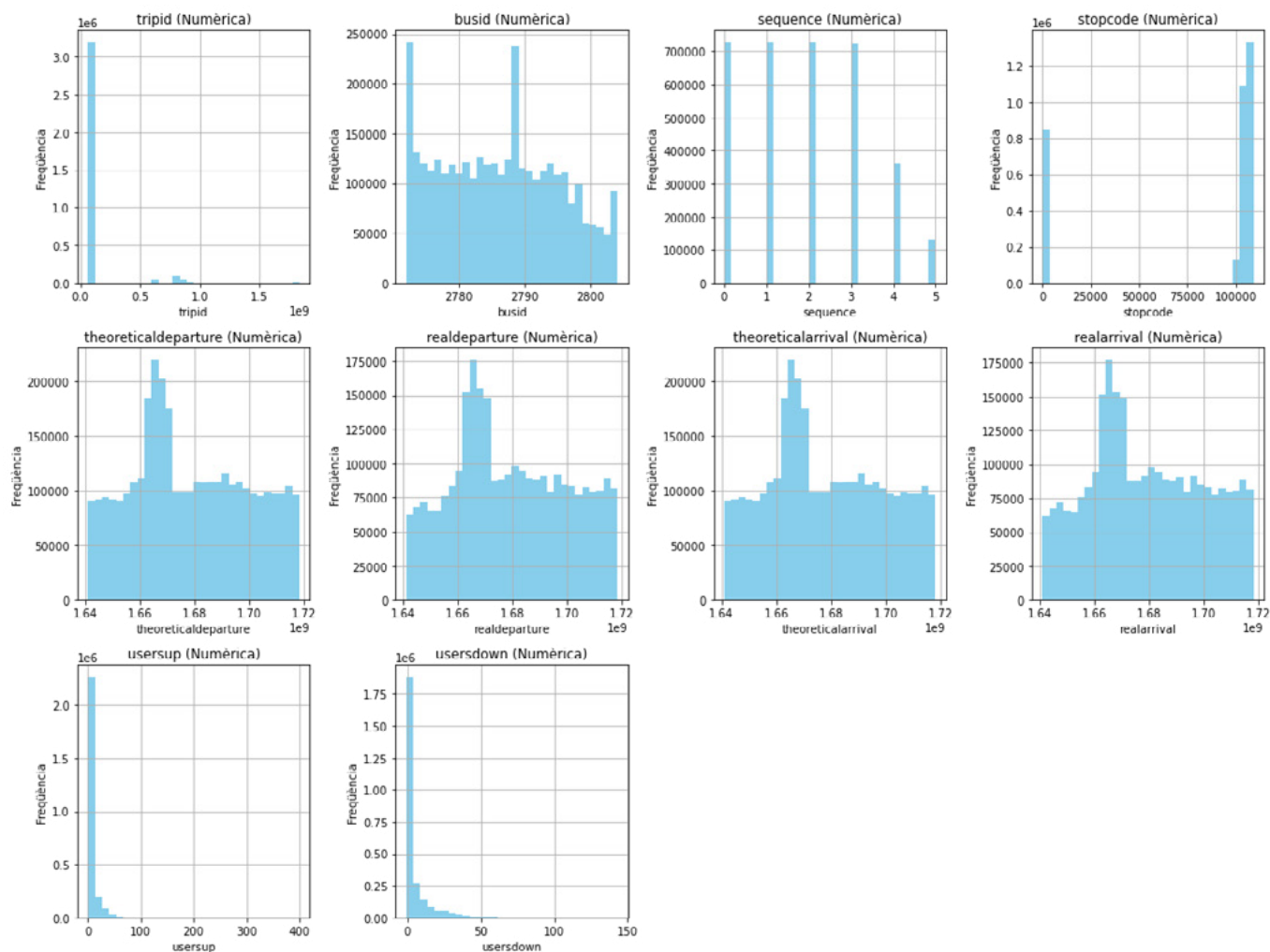


Els esdeveniments es distribueixen del 01-01-2022 al 06-07-2024. Els valors de 1.000 son els més comuns pel camp "aforament estimat".

La distribució de la resta de valors semblen consistents. No es detecten inconsistències als valors del *data set*.

Data set 6: Expedició-Parada

És el *data set* principal, ja que conté les dades de demanda del servei ("usersup"), que és la variable que volem predir. Visualitzem les variables principals:



Observem que totes les variables semblen estar correctament distribuïdes. Al fer el preprocessament verificarem els temps teòrics i reals de sortida i arribada, ja que sembla que hi ha una concentració de dades en algun període concret.

En realitzar el preprocessament de dades de "usersup" i "usersdown" haurem de revisar els valors atípics; sobretot, superiors, ja que observem que els valors es distribueixen entre 0 i 100 per "usersup" i entre 0 i 50 per "usersdown". En canvi els valors d'"usersup" arriben fins a 400 (intuint incongruències en les dades) i els de "usersdown" fins a 150.

3.3. Neteja de dades

En aquesta secció de neteja de dades, ens centrarem a corregir les inconsistències detectades durant l'anàlisi inicial. Aquest procés és essencial per garantir la fiabilitat i precisió dels resultats. Això inclourà el tractament dels valors perduts o erronis, l'eliminació de duplicats, la correcció de formats incoherents i la normalització de les variables categòriques i numèriques. A través d'aquestes accions, transformarem les dades crues en un conjunt estructurat i preparat per a una anàlisi més profunda o per ser utilitzat en models de predicció.

Per dur a terme el preprocessament de les dades, hem dividit el data set principal per ruta, creant així quatre subconjunts de dades: "A1IDA", "A1VUELTA", "A2IDA" i "A2VUELTA". Aquesta divisió l'hem realitzat en dues fases:

- **Data set Expedicions:** Hem extret la part de text del camp "tripcode" (que segueix un format com ara "A1VUELTA1119" o "A2IDA1456") per identificar la direcció i ruta del viatge. A partir d'aquesta informació, hem creat un nou camp combinant "date" i "tripid", que serveix com a clau única per a cada viatge. Això ens permet diferenciar clarament cada expedició segons la seva ruta i direcció.
- **Data set Expedició-Parada:** Hem afegit la part extreta del camp "tripcode" (per exemple, "A1VUELTA", "A2IDA", etc.) a aquest data set, el qual conté informació detallada sobre cada parada. Posteriorment, hem fet una fusió amb el data set d'expedició utilitzant com a clau els camps "date" i "tripid" per garantir la coherència entre la informació de l'expedició i les parades associades.

S'eliminen les dades anteriors a 1 de setembre de 2022 per haver-hi encara restriccions COVID.

Aquest procés ens permet organitzar les dades de manera més clara i eficient, facilitant l'anàlisi per a rutes i direccions específiques.

3.3.1. Gestió de valors buits i zeros

A1

A continuació, es presenta una taula que mostra la quantitat total de valors nuls presents en cada variable del conjunt de dades, juntament amb el percentatge que aquests representen respecte al total de registres.

	A1IDA		A1VUELTA	
Variable	Total NaN	Percentatge NaN	Total NaN	Percentatge NaN
date	0	0.000000	0	0.000000
tripid	0	0.000000	0	0.000000
busid	0	0.000000	0	0.000000
sequence	0	0.000000	0	0.000000
stopcode	0	0.000000	0	0.000000
referencestop	0	0.000000	0	0.000000
theoreticaldeparture	0	0.000000	0	0.000000
realdeparture	109060	17.205421	113551	14.624945
theoreticalarrival	0	0.000000	0	0.000000
theoreticaltimeinstop	0	0.000000	0	0.000000
realarrival	109060	17.205421	113551	14.624945
usersup	154906	24.438134	173824	22.387883
usersdown	154906	24.438134	173824	22.387883

Observant la taula, al data set "A1IDA", els camps d'interès per modelar "realdeparture" i "realarrival" contenen un 17,20% de valors faltants (NaN). Els camps "usersup" i "usersdown" contenen un total de 24,43% de valors faltants.

Al data set "A1VUELTA" els camps d'interès per modelar "realdeparture" i "realarrival" contenen un 14,62% de valors faltants. Els camps "usersup" i "usersdown" contenen un total de 22,38% de valors faltants:

- Dins de "realdeparture" i "realarrival" trobarem viatges que no s'han realitzat, així com viatges que sí s'han fet però en què no s'ha efectuat alguna parada perquè l'autobús anava ple. Aquells viatges que no s'han realitzat s'eliminaran, ja que aquests autobusos no existeixen (no es poden modelar) ¹.
- Un cop eliminats els viatges inexistents i feta l'agregació per franges horàries, els valors de "usersup" que no tinguin valor per franja horària hauran de ser estimats.

A2

A continuació, es presenta una taula que mostra la quantitat total de valors nuls presents en cada variable del conjunt de dades, juntament amb el percentatge que aquests representen respecte al total de registres.

	A1IDA		A1VUELTA	
Variable	Total NaN	Percentatge NaN	Total NaN	Percentatge NaN
date	0	0.000000	0	0.000000
tripid	0	0.000000	0	0.000000
busid	0	0.000000	0	0.000000
sequence	0	0.000000	0	0.000000
stopcode	0	0.000000	0	0.000000
referencestop	0	0.000000	0	0.000000
theoreticaldeparture	0	0.000000	0	0.000000
realdeparture	76014	20.901514	100460	18.745335
theoreticalarrival	0	0.000000	0	0.000000
theoreticaltimeinstop	0	0.000000	0	0.000000
realarrival	76014	20.901514	100460	18.745335
usersup	95341	26.215845	131266	24.493581
usersdown	95341	26.215845	131266	24.493581
complete	95341	26.215845	131266	24.493581

Al data set "A2IDA", els camps d'interès per modelar "realdeparture" i "realarrival" contenen un 20,90% de valors faltants. Els camps "usersup" i "usersdown" contenen un total de 26,21% de valors faltants.

Al data set "A2VUELTA", els camps d'interès per modelar "realdeparture" i "realarrival" contenen un 18,74% de valors faltants. Els camps "usersup" i "usersdown" contenen un total de 24,49% de valors faltants.

- Dins de "realdeparture" i "realarrival" trobarem viatges que no s'han realitzat, així com viatges que sí s'han fet però en què no s'ha efectuat alguna parada perquè l'autobús anava ple. Aquells viatges que no s'han realitzat s'eliminaran, ja que aquests autobusos no existeixen (no es poden modelar) ².

¹ Un cop finalitzades les fases de preparació de les dades i entrenament dels models, s'ha sabut que aquests registres que s'han associat a viatges inexistents en realitat són viatges que sí s'han fet però els autobusos no estaven introduïts al sistema SAE, i per això els viatges no tenen dades. S'ha tingut en compte després en les proves del model d'optimització

² Vegeu el comentari en el peu de pàgina anterior

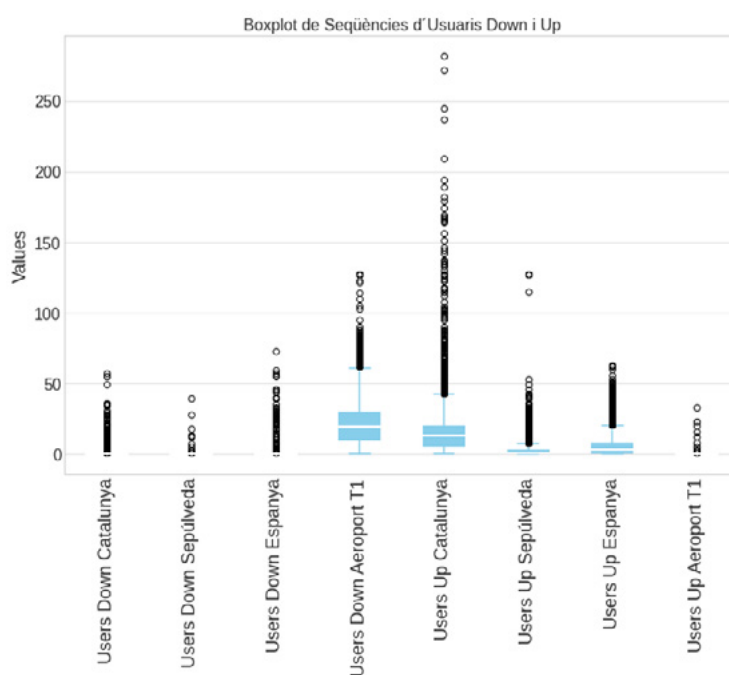
- Un cop eliminats els viatges inexistents i feta l'agregació per franges horàries, els valors de "usersup" que no tinguin valor per franja horària hauran de ser estimats.

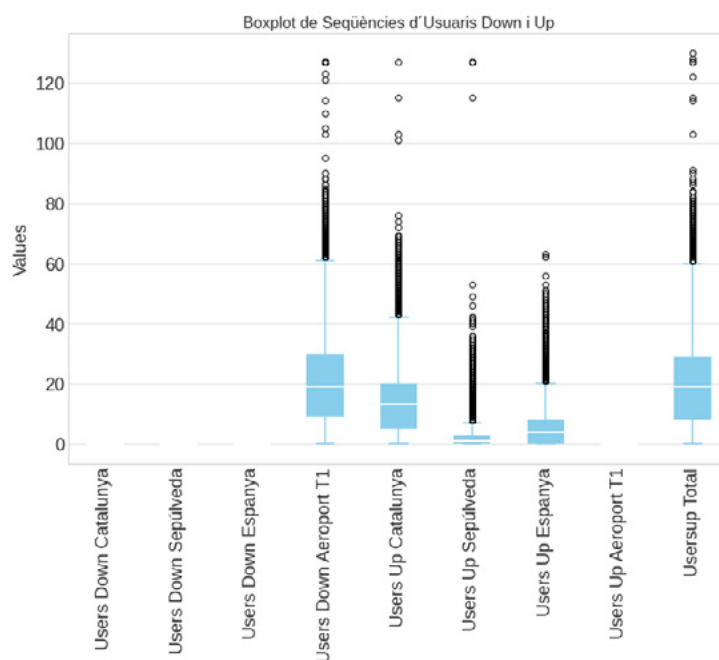
3.3.2. Outliers

En aquesta secció, ens centrarem en la identificació i tractament dels *outliers* o valors atípics dins del conjunt de dades. Els *outliers* són observacions que es troben significativament allunyades de la resta de dades i poden distorsionar els resultats de les anàlisis o dels models predictius. Aquests valors poden ser deguts a errors de mesura, dades introduïdes incorrectament o, en alguns casos, poden ser indicadors d'informació rellevant. En aquesta etapa, analitzarem quins *outliers* són problemàtics i aplicarem tècniques adequades per gestionar-los, ja sigui eliminant-los, ajustant-los o utilitzant mètodes robustos que minimitzin el seu impacte en l'anàlisi.

A1IDA

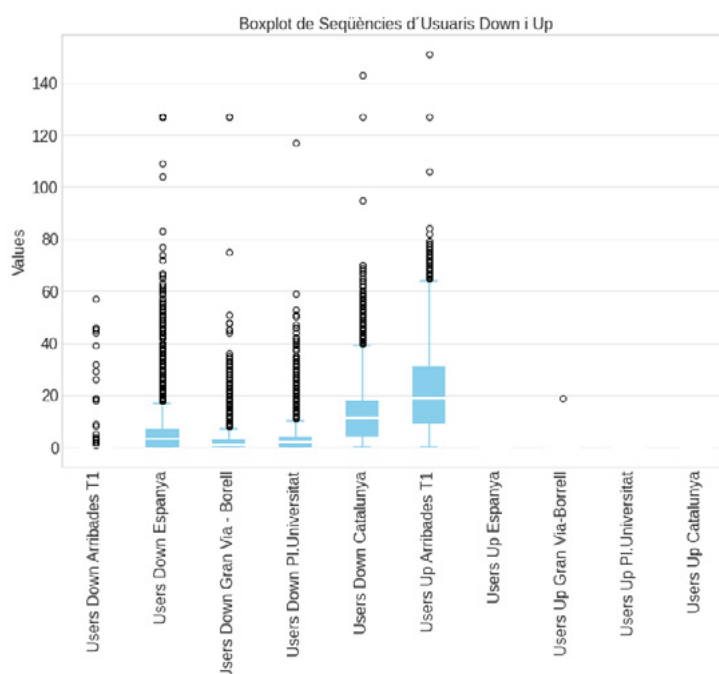
El següent *boxplot* mostra la distribució del nombre de passatgers que pugen ("usersup") i baixen ("usersdown") a cada parada pel trajecte "A1IDA", representada per la variable "sequence". Aquesta variable indica l'ordre de les parades de l'autobús al llarg del recorregut.





A1VUELTA

El següent *boxplot* mostra la distribució del nombre de passatgers que pugen ("usersup") i baixen ("users-down") a cada parada pel trajecte "A2VUELTA", representada per la variable "sequence". Aquesta variable indica l'ordre de les parades de l'autobús al llarg del recorregut.

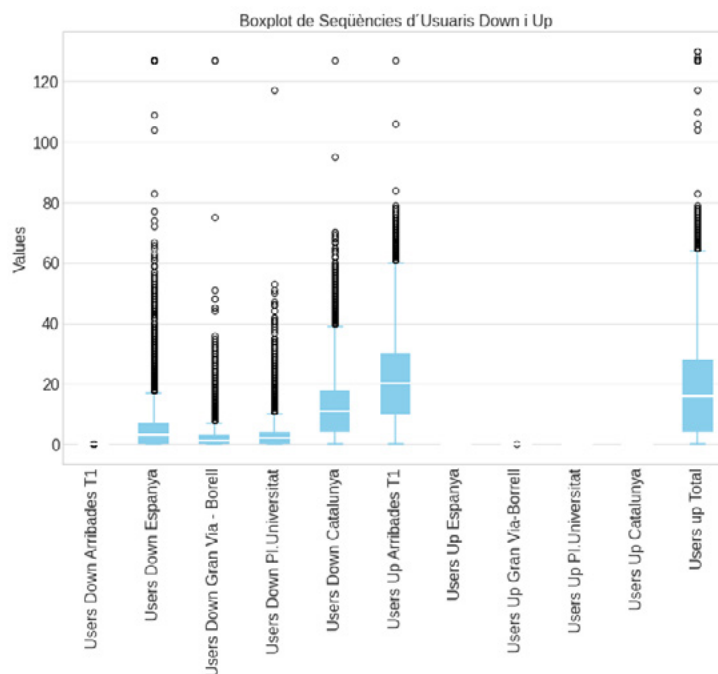


Observem que hi ha dades de "usersdown" a Arribades T1 (baixarien al començament de la ruta) i de "usersup" a "Gran Via Borrell" (pujarien enmig d'una ruta de tornada de l'aeroport). Aquestes dades no tenen caixa i tenen una mitjana de 0 (són dades atípiques), per tant, no les considerarem ja que ens interessa el trajecte:

Vial arribades T1 (Usersup) -> Plaça Espanya (UsersDown) -> Gran Via - Borrell (UsersDown) -> Plaça Universitat (UsersDown) -> Plaça Catalunya (UsersDown).

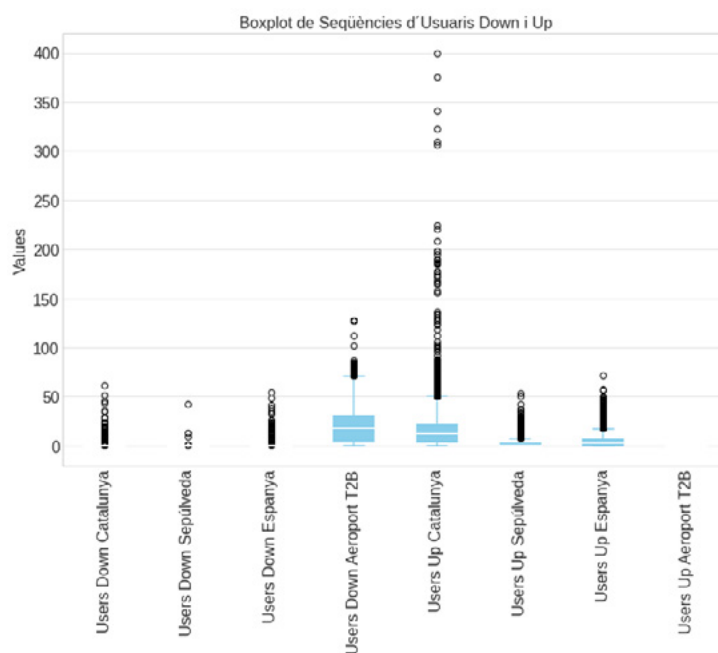
Observem també que tenim valors superiors a 130 de “usersup” a la parada inicial. Corregim aquests valors extrems amb la mediana per franja horària d’aquesta seqüència.

Un cop imputats els valors comparem “usersup” a l’Aeroport amb la suma dels “usersdown” de totes dels parades i validem que estigui correcte si les caixes són similars.



A2IDA

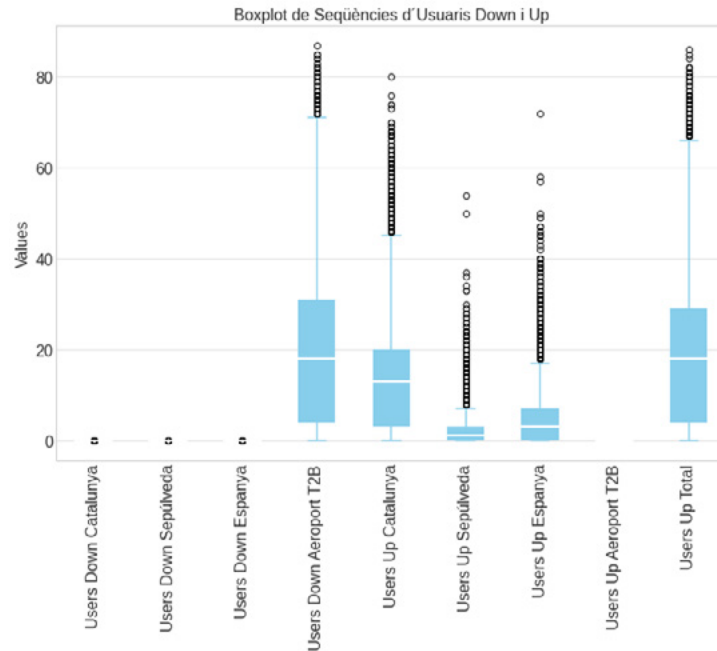
El següent *boxplot* mostra la distribució del nombre de passatgers que pugen (“usersup”) i baixen (“usersdown”) a cada parada pel trajecte “A2IDA”, representada per la variable “sequence”. Aquesta variable indica l’ordre de les parades de l’autobús al llarg del recorregut.



Observem que hi ha dades de “usersdown” a les parades Catalunya, Sepúlveda i Espanya, i de “usersup” a l'Aeroport T2B (pugen a l'aeroport en un autobús amb ruta d'anada) quan el servei és d'anada cap a l'aeroport. Aquestes dades no tenen caixa i tenen una mitjana de 0 (són dades atípiques), per tant, no les considerarem, ja que ens interessa el trajecte:

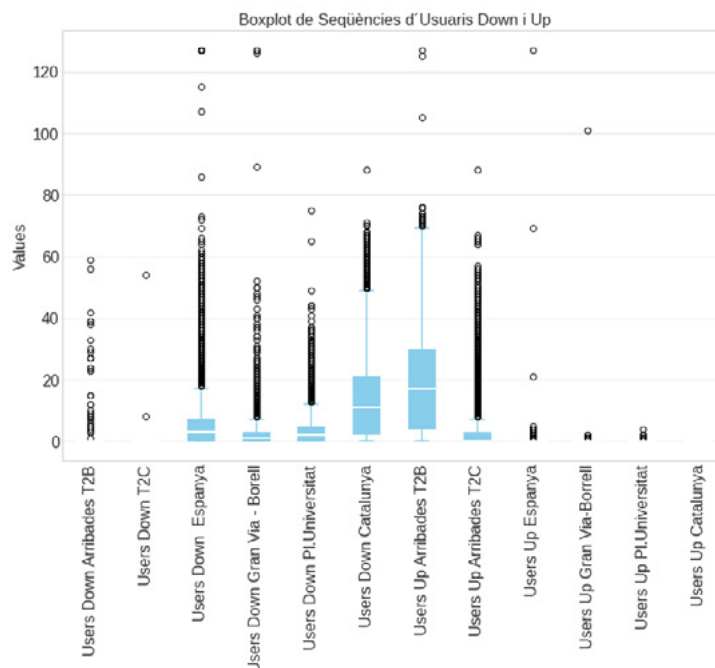
Catalunya (Usersup) -> Sepúlveda (Usersup) -> Espanya (Usersup) -> Aeroport T2B (UsersDown).

Imputem valor 0 a les dades anòmales d'aquestes seqüències i la mediana per franja horària als valors atípics extrems (aquells que s'han calculat amb un temps superior a 0,25 hores entre arribada i sortida de l'autobús a una parada)



A2VUELTA

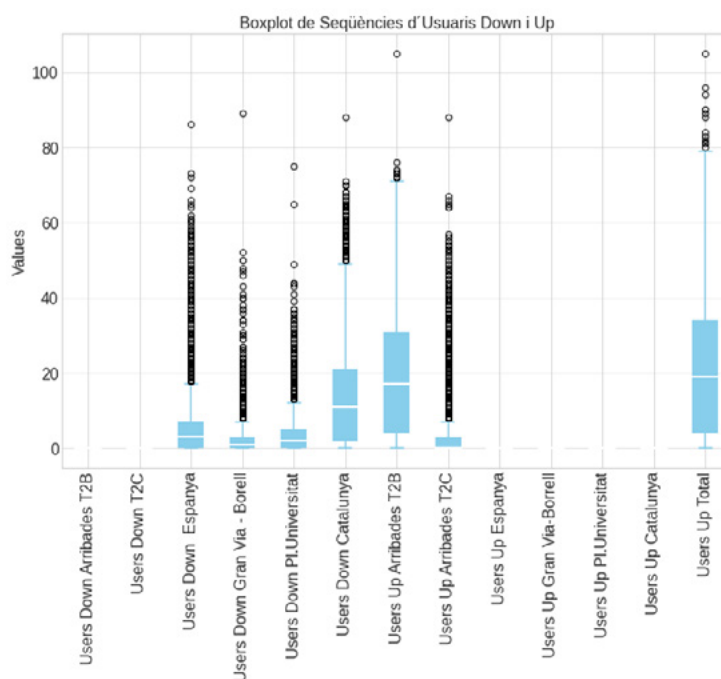
El següent *boxplot* mostra la distribució del nombre de passatgers que pugen (“usersup”) i baixen (“users-down”) a cada parada pel trajecte “A2VUELTA”, representada per la variable “sequence”. Aquesta variable indica l'ordre de les parades de l'autobús al llarg del recorregut.



Observem que hi ha dades de “usersdown” a Arribades T2B (baixarien al començament de la ruta) i a Arribades T2C (baixarien a la següent parada, encara a la mateixa terminal de l'aeroport). El mateix passa amb els “user-sup” a Espanya, Gran Via Borrell i Pl. Universitat. Aquestes dades no tenen caixa i tenen una mitjana de 0 (són dades atípiques), per tant, no les considerarem, ja que ens interessa el trajecte:

Vial aeroport T2B (Usersup) -> Vial aeroport T2C (Usersup) -> Plaça Espanya (UsersDown) -> Gran Via - Borrell (UsersDown) -> Plaça Universitat (UsersDown) -> Plaça Catalunya (UsersDown).

A més de imputar valor 0 a aquests valors, imputem la mediana per franja horària als valors atípics extrems (aquells que s'han calculat amb un temps superior a 0,25 hores entre arribada i sortida de l'autobús a una parada).



3.4. Transformació de dades

3.4.1. Creació de variables

Hem creat diverses variables per preparar el data set per a la predicció de la demanda del servei. Com que la predicció es realitza sobre una sèrie temporal, aquestes variables ens permetran capturar patrons estacionals, tendències i influències temporals específiques en el comportament de la demanda:

- **"is_wknd"**: Indica si el dia correspon a cap de setmana ("True" si és dissabte o diumenge), ja que els patrons de demanda poden canviar en dies no laborables.
- **"is_first_fortnight"**: Indica si el dia del mes està comprès entre l'1 i el 15, per captar possibles diferències de comportament entre la primera i la segona quinzena de cada mes.
- **"day_of_week"**: Indica el número del dia de la setmana, de l'1 (dilluns) al 7 (diumenge), per identificar variacions segons el dia de la setmana.
- **"week_of_year"**: Indica el número de la setmana dins l'any, per detectar tendències setmanals o patrons estacionals.
- **"day_of_month"**: Indica el número del dia dins del mes, que pot influir en la demanda per esdeveniments recurrents o patrons mensuals.
- **"year"**: Indica l'any del registre, per tenir en compte l'evolució de la demanda al llarg dels anys.
- **"rain"**: Indica si ha plogut (sí/no), ja que el clima podria tenir un impacte significatiu en la demanda del servei.
- **"holiday"**: Indica si el dia és festiu ("True" si és festiu). Aquesta variable és crucial, ja que la demanda del servei pot variar significativament en dies no laborables o festius.
- **"bank_holiday"**: Indica si el dia és l'anterior o el següent a un festiu ("True" si és un dia abans o després d'un festiu). Aquesta variable ens permet detectar patrons de demanda que puguin estar influenciats pels dies propers a festius, ja que solen generar comportaments diferents, com ponts o increments de mobilitat.

A més de les variables temporals que hem creat per a la predicció de la demanda del servei, també s'han afegit les següents variables relacionades amb esdeveniments especials. Únicament s'ha contemplat els esdeveniments amb aforament estimat:

- **"inici_esdeveniment"**: Suma dels esdeveniments que s'inicien per dia.
- **"fi_esdeveniment"**: Suma dels esdeveniments que finalitzen per dia.
- **"esdeveniment"**: Suma total dels esdeveniments presents per dia.
- **"aforament"**: Suma de l'aforament total de tots els esdeveniments per dia.

Anàlisi de dades



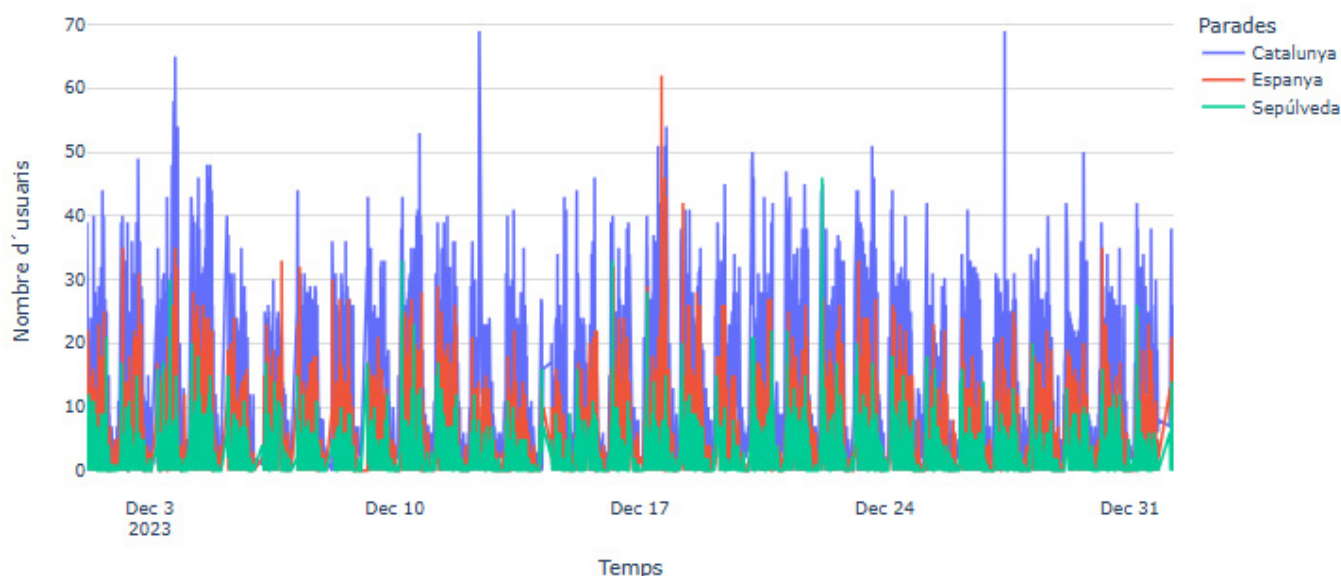
L'anàlisi exploratòria de dades (EDA) en aquest projecte té com a objectiu entendre el funcionament del servei d'autobusos a partir de les dades disponibles. A través de l'EDA, explorarem aspectes clau com el volum de passatgers en diferents parades, la distribució horària dels viatges i la variació de la demanda segons el dia de la setmana i, en festius o esdeveniments especials. Aquesta anàlisi ens ajudarà a identificar patrons en l'ús del servei, detectar possibles ineficiències i entendre com factors externs com el clima o els esdeveniments impacten sobre la mobilitat dels usuaris. A més, mitjançant visualitzacions i estadístiques descriptives, l'EDA ens proporcionarà una visió global que facilitarà la presa de decisions per dissenyar els models de predicció de la demanda del servei.

4.1. Visualització general de la sèrie temporal

En aquesta secció, es realitzarà una visualització general de la sèrie temporal per obtenir una primera visió sobre com es comporta la demanda del servei d'autobusos al llarg del temps. Aquesta anàlisi ens permetrà identificar tendències globals, patrons estacionals i possibles fluctuacions en l'ús del servei. Mitjançant la representació gràfica de la sèrie temporal, podrem observar com variables com el nombre de passatgers varien al llarg dels dies. A continuació, es visualitza el volum de passatgers en totes les línies de l'Aerobús durant el mes de desembre de 2023, a mode d'exemple, per observar si existeixen tendències, cicles o comportaments particulars. Aquest gràfic ens permet detectar també si hi ha hores amb més afluència o una tendència creixent o decreixent en el temps.

A1IDA

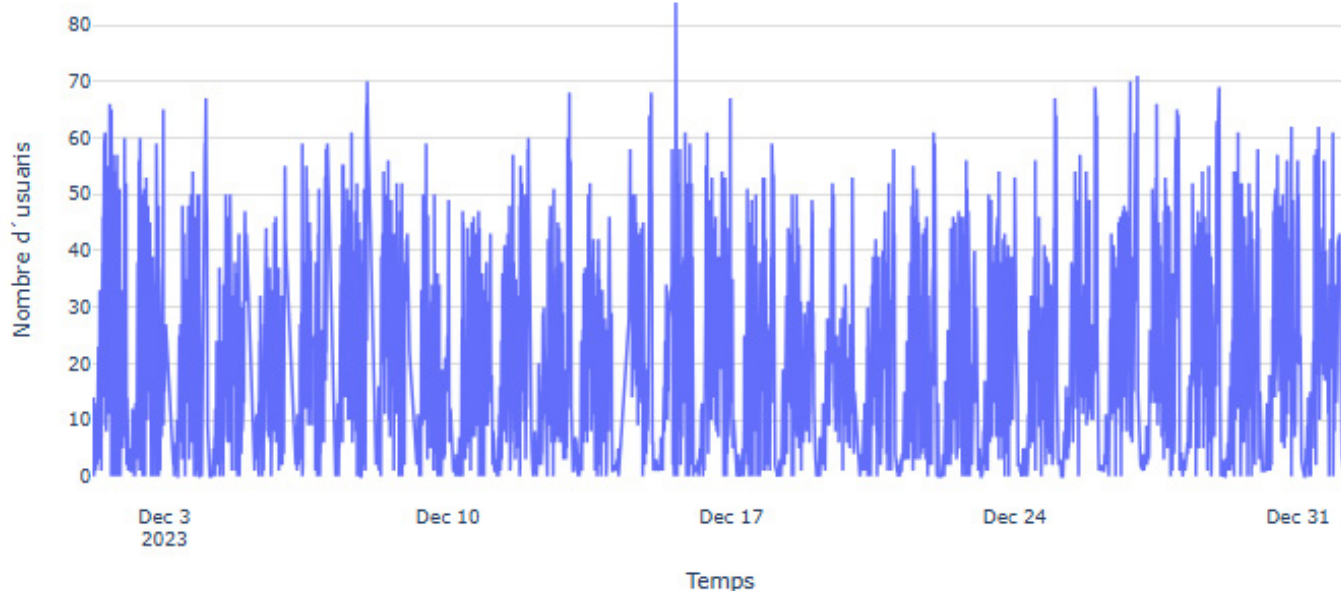
Usuaris pujant per parada (Users Up) - A1IDA



En aquesta sèrie temporal s'observa un patró clar de repetició diària, on els passatgers pugen en major quantitat a la [seqüència 0 \(Catalunya\)](#), seguit per la [seqüència 2 \(Espanya\)](#) i, finalment, la [seqüència 1 \(Sepúlveda\)](#). Tot i que aquest comportament es repeteix cada dia, es percep una variació en la quantitat de passatgers entre dies. Això indica que, tot i mantenir un cicle diari consistent en l'ordre de parades amb més afluència, la demanda del servei d'autobusos varia segons el dia. Aquesta variació pot estar influenciada per factors com el dia de la setmana, esdeveniments específics o altres condicionants externs. El patró general es manté, però la quantitat de passatgers presenta fluctuacions entre dies.

A1VUELTA

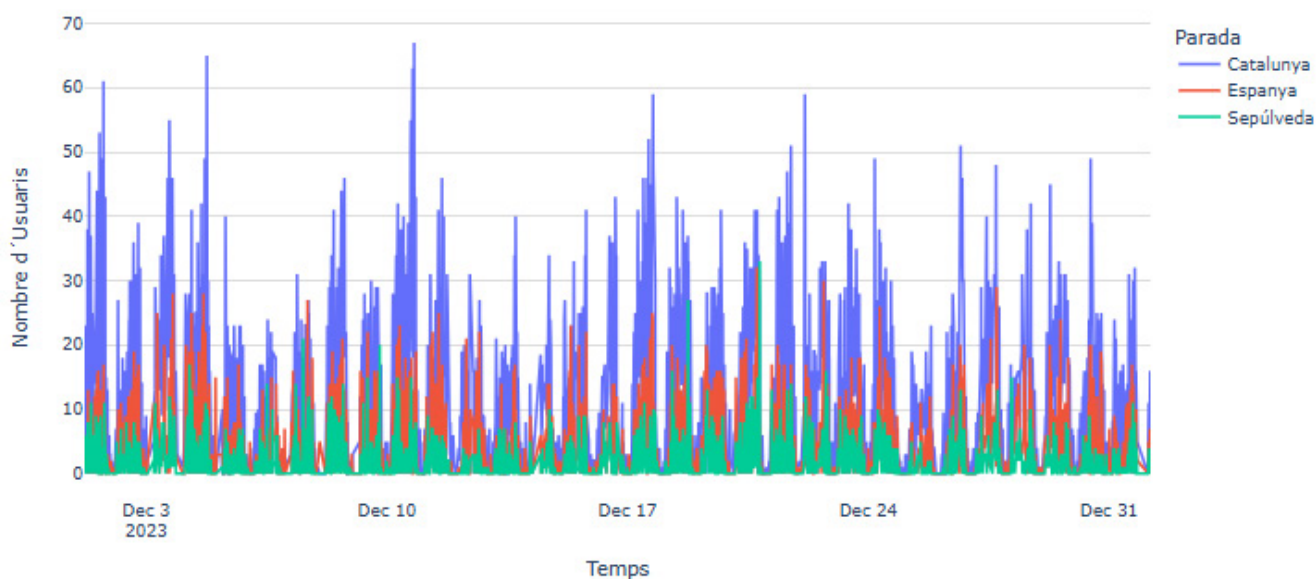
Usuaris pujant a l'aeroport - A1VUELTA



Aquest gràfic mostra l'evolució dels usuaris que pugen a la parada arribades T1, la qual és la dada clau que ens interessa analitzar. Al llarg del temps, podem observar un patró de repetició diària consistent, on el comportament de la demanda de passatgers es repeteix de manera similar cada dia. Aquest patró es caracteritza per pics recurrents en moments específics de la jornada, seguits de descensos, indicant hores de més afluència i hores de menys activitat a la parada d'arribades T1.

A2IDA

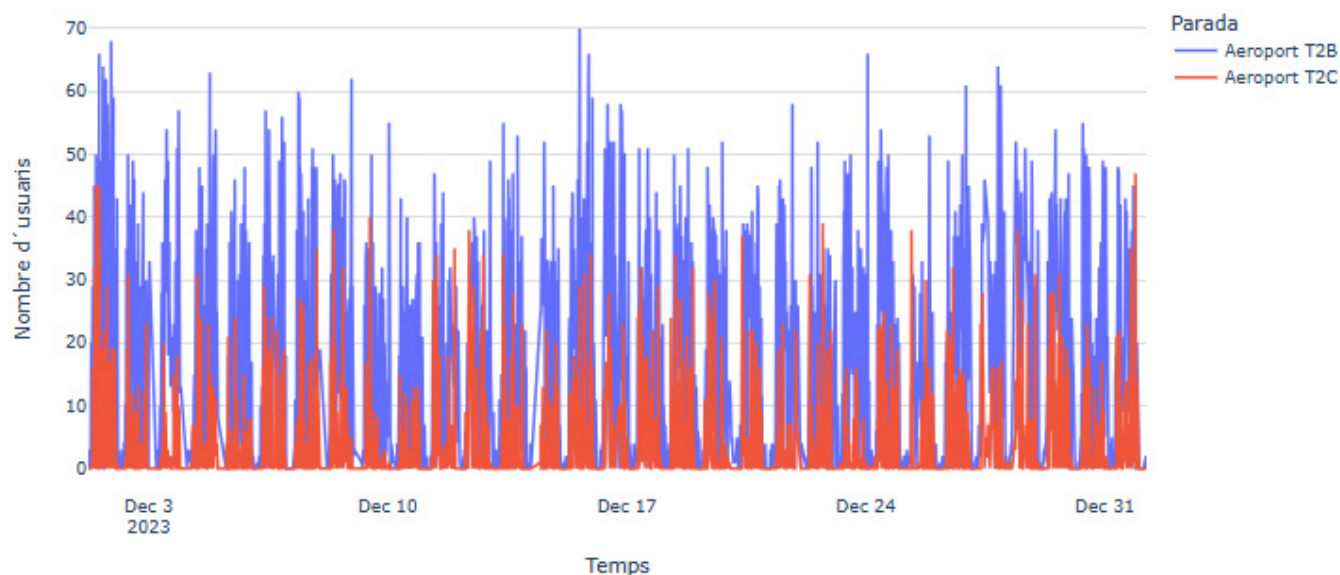
Usuaris pujant per parada (Users Up) - A2IDA



En aquesta sèrie temporal, com a les anteriors, s'observa el mateix patró de repetició diària, on els passatgers pugen en major quantitat a la [seqüència 0 \(Catalunya\)](#), seguit per la [seqüència 2 \(Espanya\)](#) i, finalment, la [seqüència 1 \(Sepúlveda\)](#). Existeix una variació en el número total de passatgers entre dies.

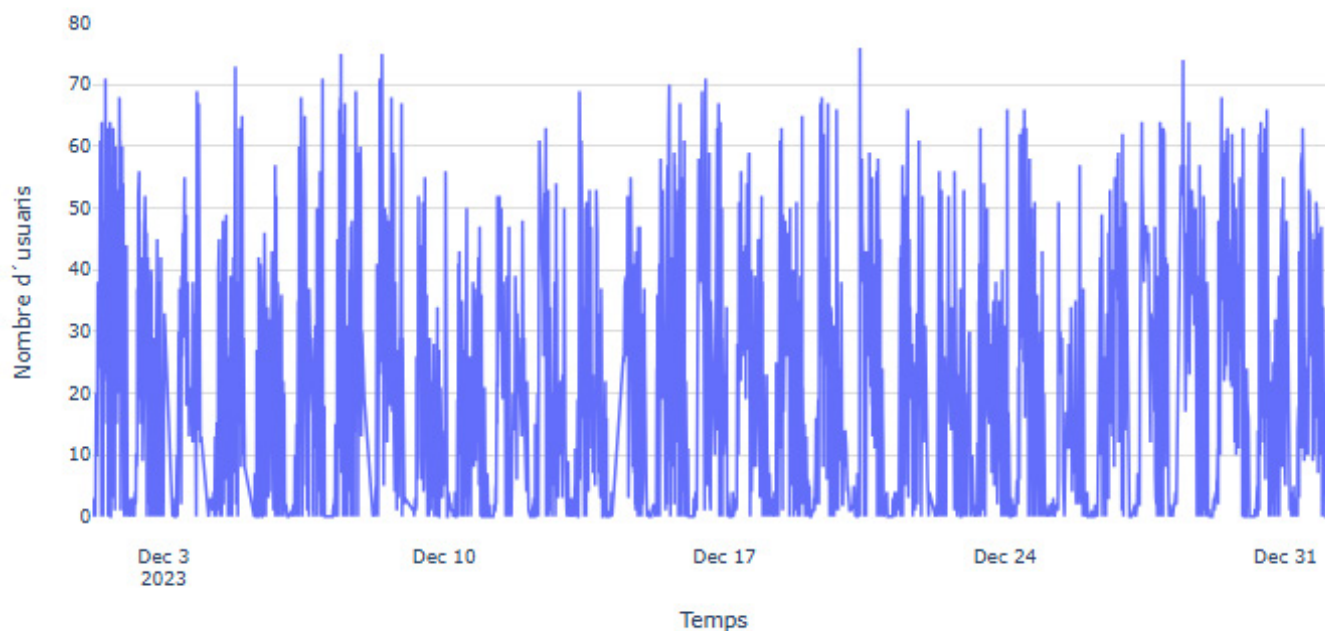
A2VUELTA

Usuaris pujant a l'Aeroport T2 - A2VUELTA



Aquest primer gràfic mostra el comportament dels passatgers que pugen a les parades de T2B i T2C a l'aeroport. En aquest gràfic, es pot observar clarament que el nombre d'usuaris que pugen a T2B és sempre superior en comparació amb T2C, cosa que indica una major afluència de passatgers en aquesta parada. Tot i que la diferència entre les dues parades es manté constant, amb T2B dominant, ambdues mostren un patró diari recurrent, amb pics i descensos que segueixen una pauta regular en funció del dia.

Usuaris totals T2



El segon gràfic presenta la suma total dels passatgers que pugen tant a T2B com a T2C. Al igual que en el primer gràfic, es pot veure clarament el patró diari de repetició, on cada dia segueix una dinàmica de flux de passatgers similar, amb variacions de volum al llarg del dia. Aquesta suma reflecteix una tendència similar a la vista en altres gràfics, amb cicles diaris consistents i oscil·lacions moderades entre dies, la qual cosa confirma un comportament cíclic i previsible en el flux de passatgers al llarg del temps.

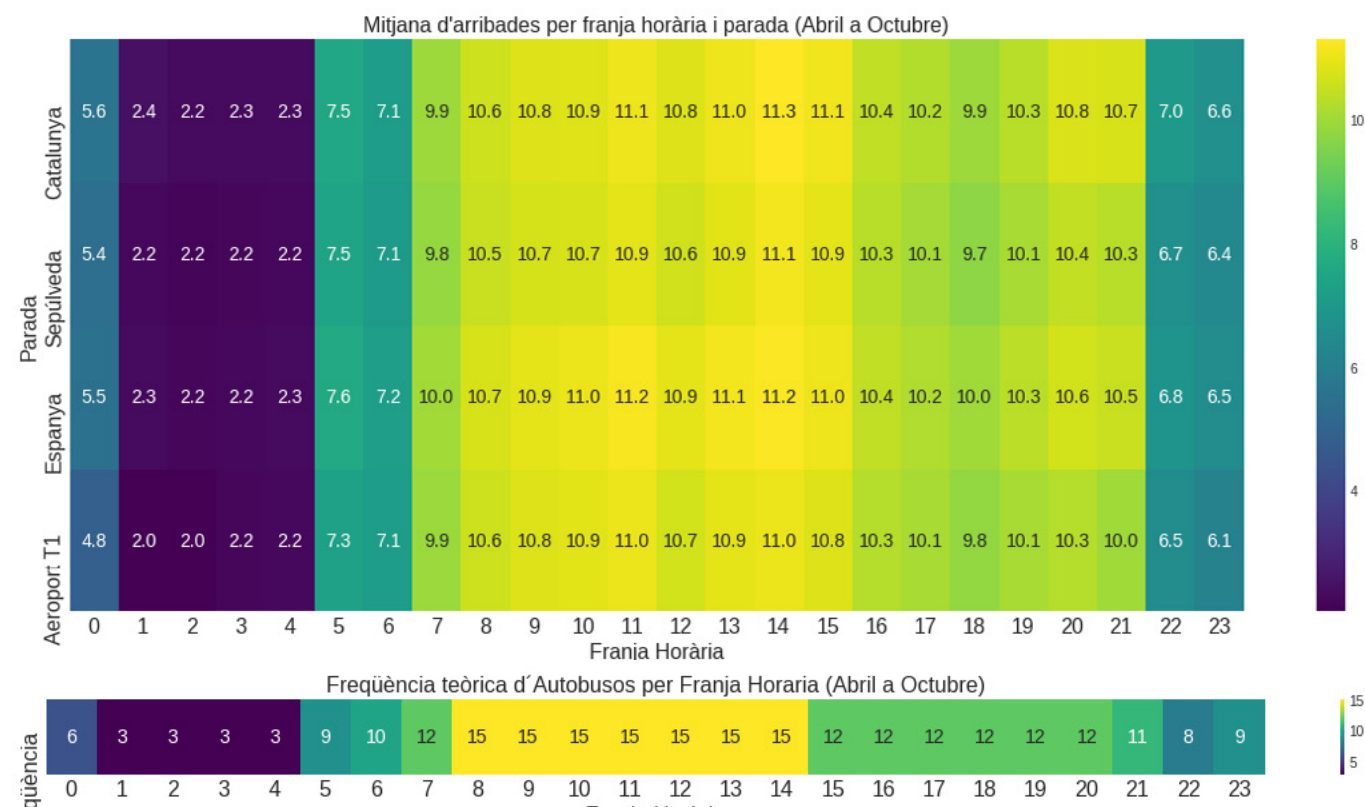
4.2. Quantitat d'autobusos per franja horària

En aquesta secció, presentem una sèrie de mapes de calor que mostren la distribució dels autobusos totals per parada i franja horària, comparant-los amb els viatges previstos. Com ja s'ha comentat anteriorment, s'observa una diferència notable entre les expedicions realitzades i les previstes, degut a l'existència d'aproximadament un 20-26% de viatges faltants. Aquest percentatge inclou tant viatges que no es duen a terme com aquells en què falta alguna parada.

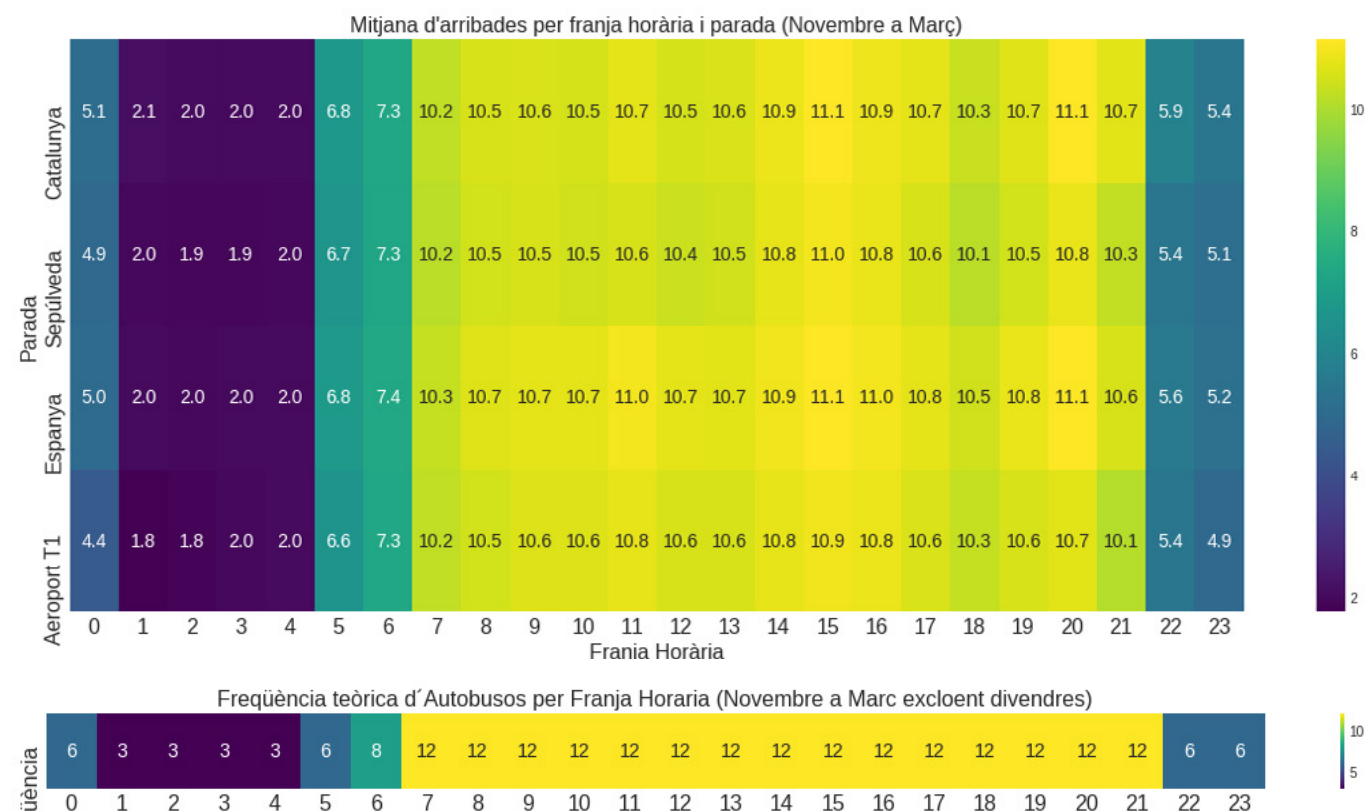
Per tal de proporcionar una visió més detallada, hem dividit les dades en dues temporades: **temporada d'hivern** (d'octubre a març) i **temporada d'estiu** (d'abril a octubre). S'han seleccionat aquests dos períodes perquè la freqüència de l'Aerobús varia entre aquestes dues temporades segons els horaris facilitats. Les gràfiques que es presenten a continuació donen una visualització dels serveis en relació amb la distribució de passatgers i permet identificar si les freqüències dels autobusos es distribueixen d'acord amb el nombre de passatgers. Aquest enfocament ens permet identificar possibles variacions estacionals en la freqüència dels serveis d'autobusos, així com avaluar en quina mesura les discrepàncies entre els viatges realitzats i els previstos afecten el servei al llarg de l'any. Els mapes de calor ofereixen una visualització clara de les diferències en funció de la parada i l'horari, destacant els punts crítics on es concentren aquestes diferències.

A1IDA

Temporada d'estiu



Temporada d'hivern



Els dos mapes de calor presentats il·lustren com varia la freqüència d'autobusos al llarg de les diferents franges horàries, tant a l'hivern com a l'estiu. A través d'aquests mapes de calor, es pot observar clarament l'evolució del nombre d'autobusos al llarg del dia.

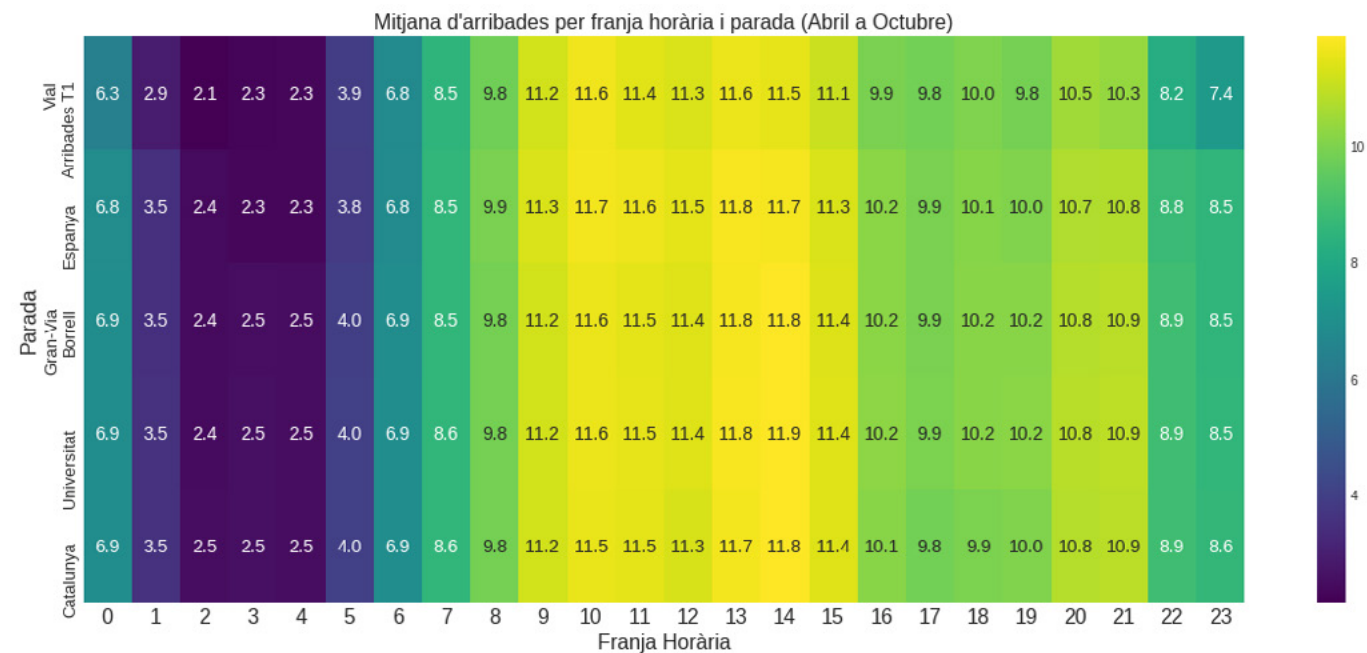
A partir de les 5h del matí, la freqüència d'autobusos comença a augmentar de manera progressiva fins a assolir el seu màxim a partir de les 8h del matí, moment en què es registra la major concentració de serveis. Aquesta elevada activitat es manté durant diverses hores, coincidint amb els horaris de més demanda. Posteriorment, es detecta una reducció gradual en el nombre d'autobusos, que varia segons la temporada.

- A l'estiu, la freqüència d'autobusos comença a disminuir cap a les 15h.
- A l'hivern, aquesta disminució es produeix més tard, a les 22h.

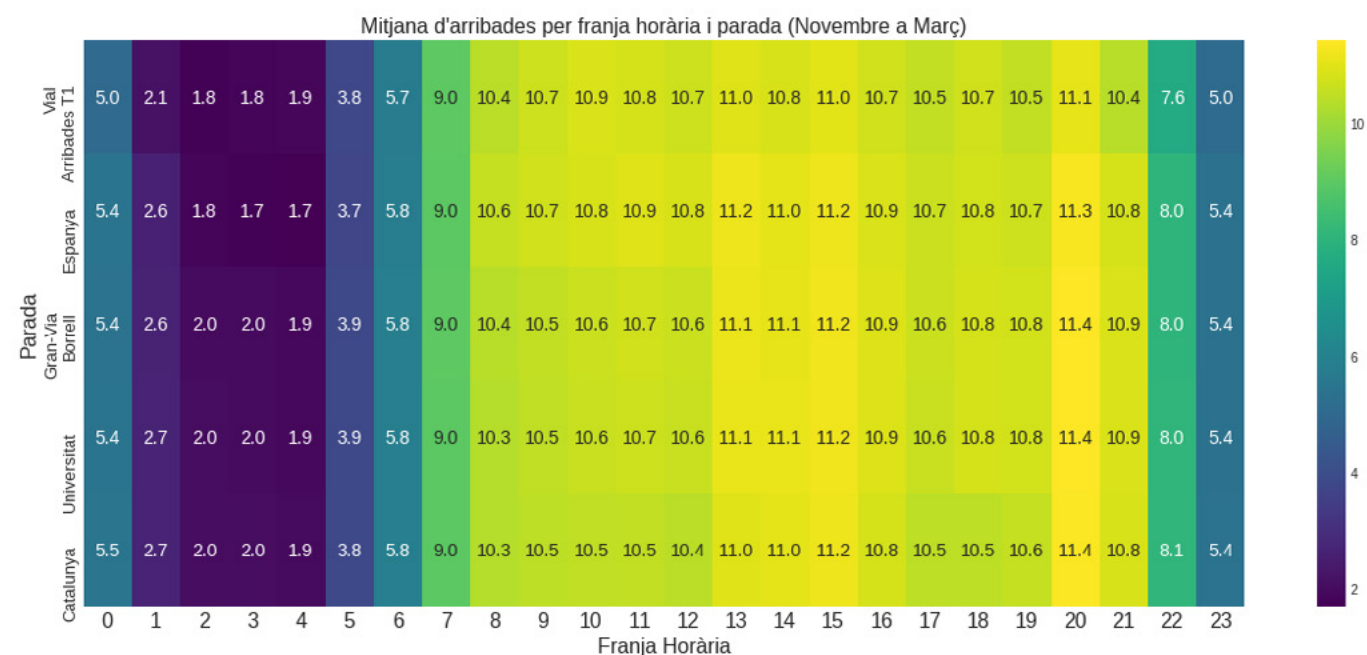
Finalment, a mesura que s'apropa la nit, la freqüència d'autobusos continua disminuint fins a arribar al seu mínim cap a la 1h de la matinada. Aquesta informació ens permet comprendre millor els patrons horaris del servei d'autobusos i les diferències estacionals en la seva operació.

A1VUELTA

Temporada d'estiu



Temporada d'hivern



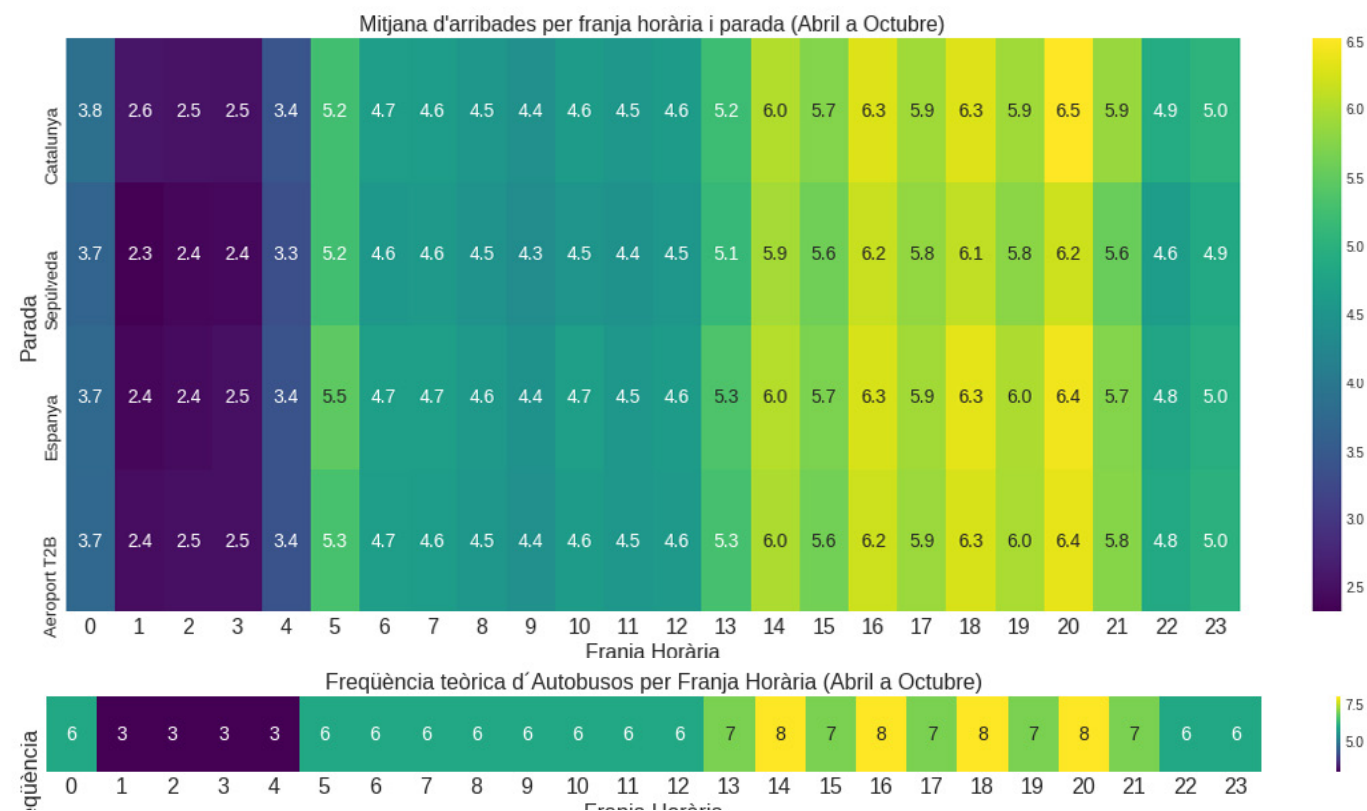
A partir de les 5h del matí, la freqüència d'autobusos comença a augmentar de manera progressiva fins a assolir el seu màxim a partir de les 9h del matí, moment en què es registra la major concentració de serveis. Aquesta elevada activitat es manté durant diverses hores, coincidint amb els horaris de més demanda. Posteriorment, es detecta una reducció gradual en el nombre d'autobusos, que varia segons la temporada.

- A l'estiu, la freqüència d'autobusos comença a disminuir cap a les 15h.
- A l'hivern, aquesta disminució es produeix més tard, a les 22h.

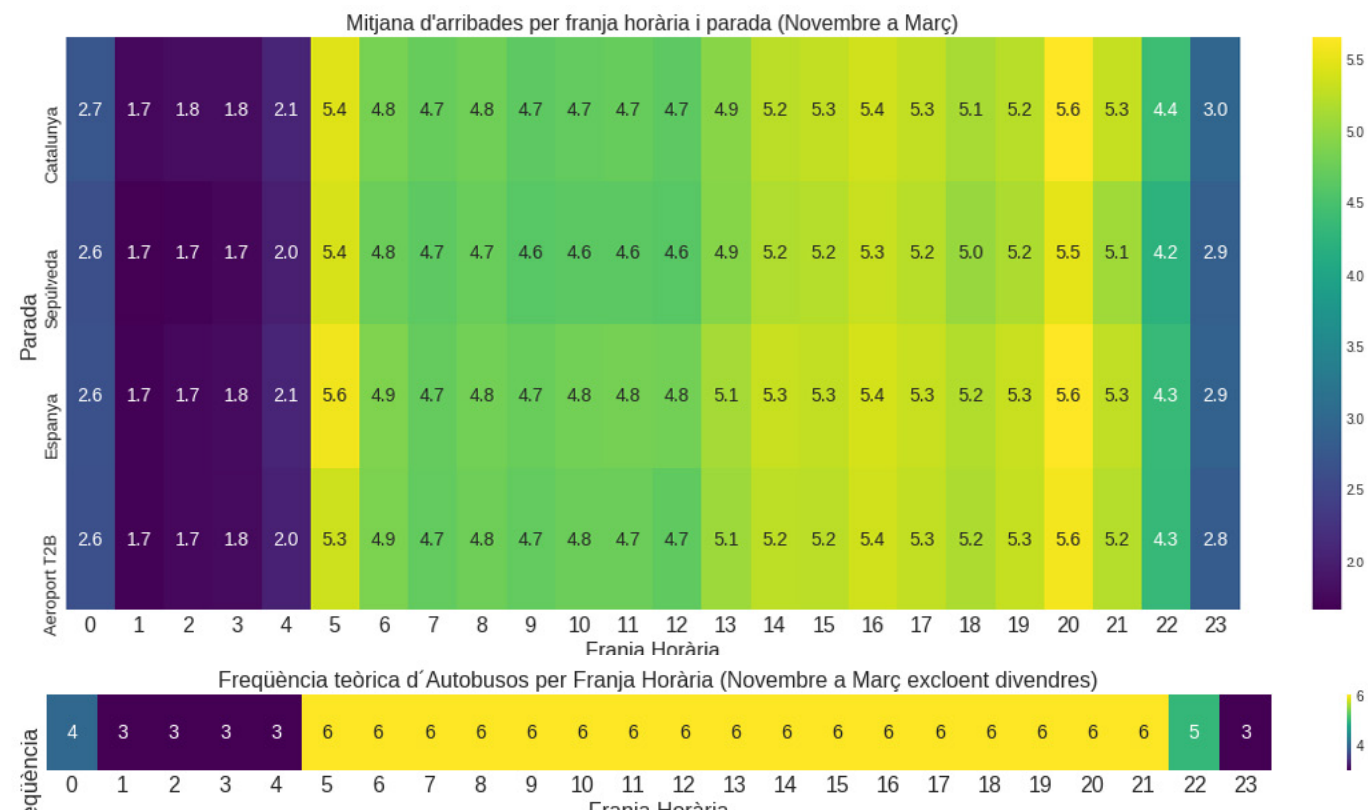
Finalment, a mesura que s'apropa la nit, la freqüència d'autobusos continua disminuint fins a arribar al seu mínim cap a les 2h de la matinada. Aquesta informació ens permet comprendre millor els patrons horaris del servei d'autobusos i les diferències estacionals en la seva operació.

A2IDA

Temporada d'estiu



Temporada d'hivern

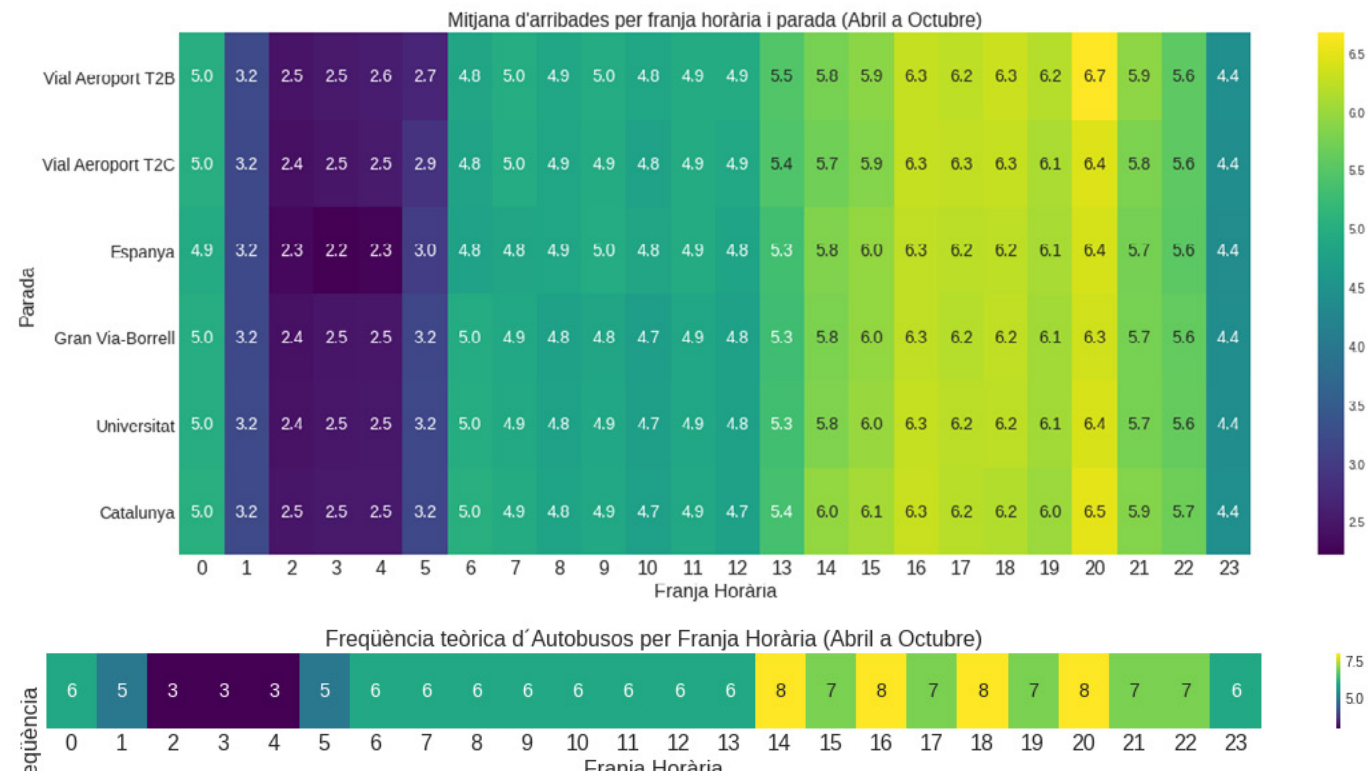


Per als viatges d'anada a la terminal 2 des de plaça Catalunya s'observen dos patrons diferenciats per temporada.

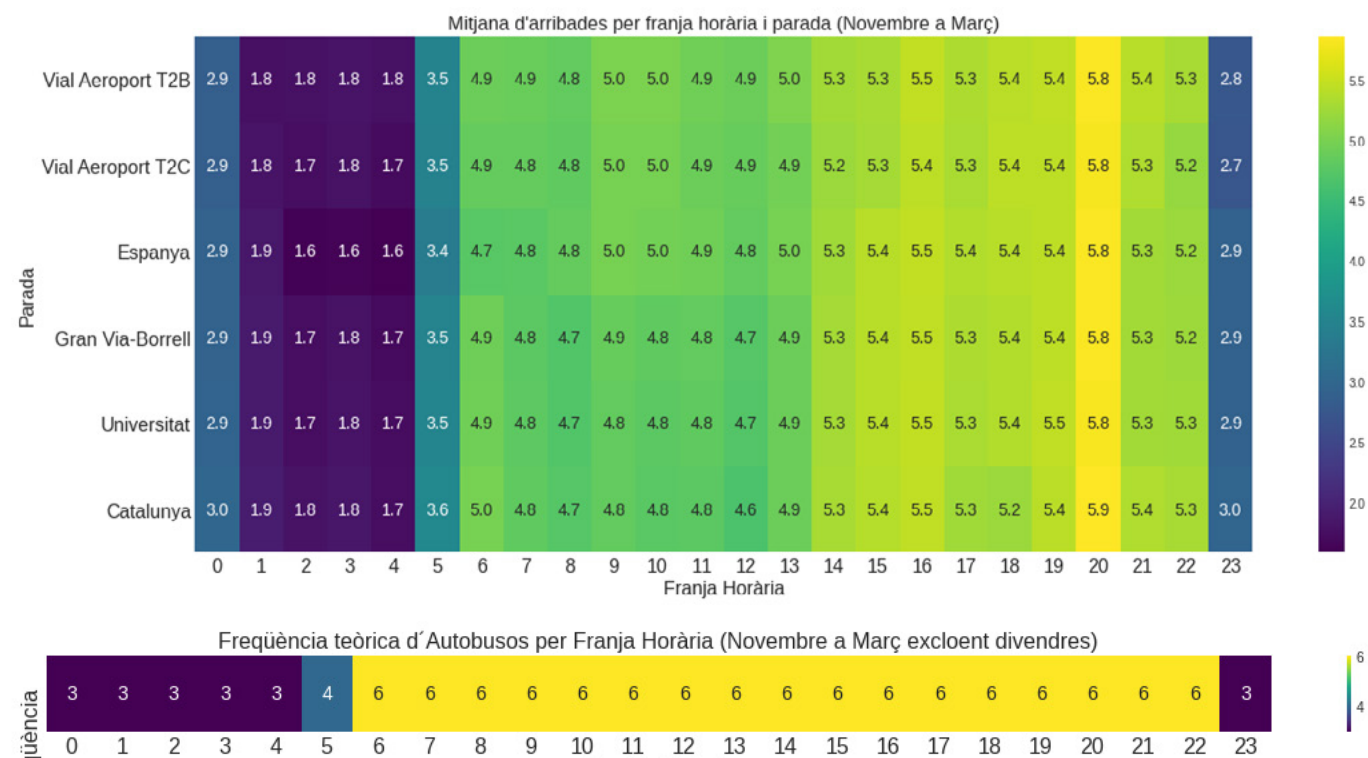
- A l'estiu, a partir de les 5h del matí, hi ha un primer augment de la freqüència d'autobusos, situant-se en aproximadament 4,5 autobusos per hora. Posteriorment, hi ha un segon augment a les 13h, arribant a 6 autobusos per hora. A les 22h, la freqüència disminueix a 5 autobusos per hora.
- A l'hivern, tot i que l'horari preveu una freqüència homogènia de 5h a 22h, s'observen dues franges diferenciades: de 6h a 12h, amb una freqüència de 4,7 autobusos per hora, i de 13h a 21h, amb una freqüència de 5,3 autobusos per hora.

A2VUELTA

Temporada d'estiu



Temporada d'hivern



El patró observat en el data set "A2IDA" es repeteix de manera similar al data set "A2VUELTA" per a les mateixes temporades.

- A l'estiu, es manté l'augment inicial de la freqüència d'autobusos a partir de les 5h del matí, amb valors al voltant dels 5 autobusos per hora, seguit d'un increment a les 13h. A les 22h, la freqüència torna a disminuir entre 3 i 1,5 autobusos per hora.
- A l'hivern, es repeteixen les dues franges diferenciades: de 6h a 12h, amb una mitjana de 4,7 autobusos per hora, i de 13h a 21h, amb 5,3 autobusos per hora, seguint el mateix patró observat a l'anada.

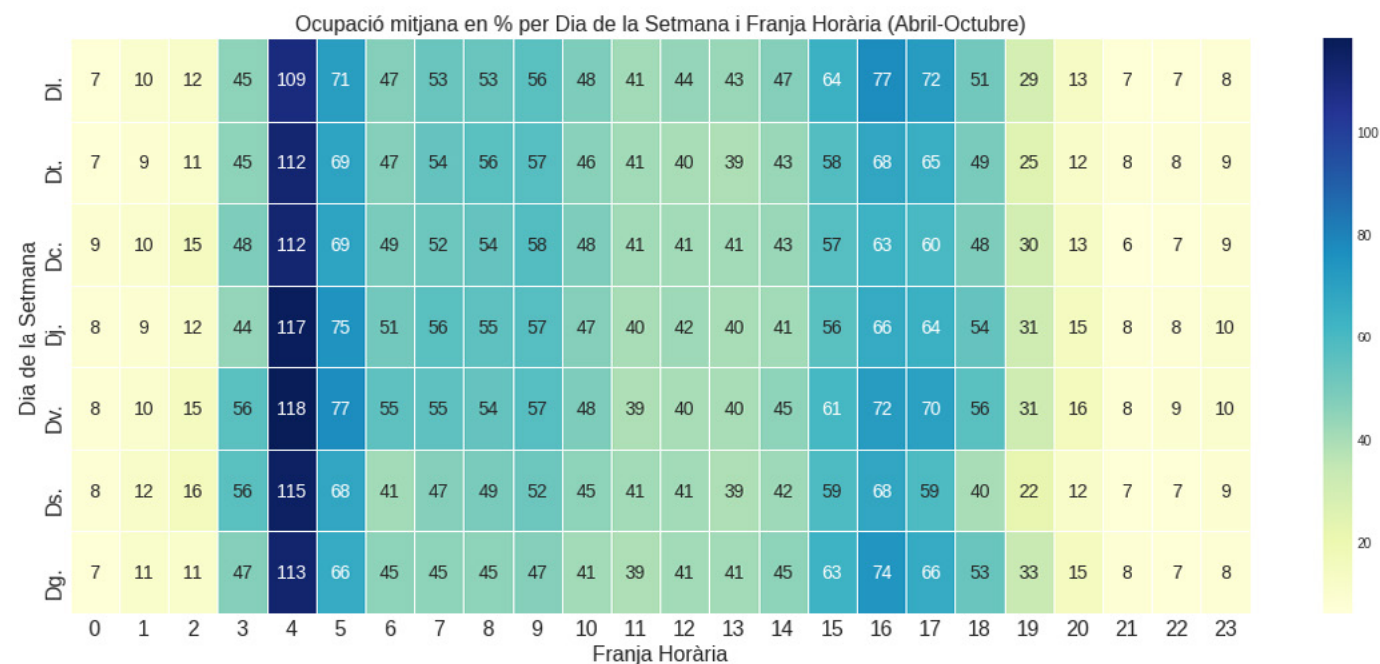
4.3. Percentatge d'ocupació

Els següents mapes de calor mostren el percentatge d'ocupació dels autobusos i el percentatge d'autobusos que supera l'ocupació màxima per dia de la setmana i franja horària, diferenciats per temporada d'hivern i temporada d'estiu, i el percentatge de viatges que excedeixen l'ocupació màxima. Aquests mapes permeten visualitzar de manera clara com varia l'ocupació dels autobusos en funció del dia i de les diferents franges horàries, oferint una visió detallada del comportament de la demanda durant les dues temporades.

Per calcular l'ocupació mitjana i l'excés d'ocupació s'ha fet servir la dada de quantitat de seients per autobús, que es situa entre 48 i 49 seients.

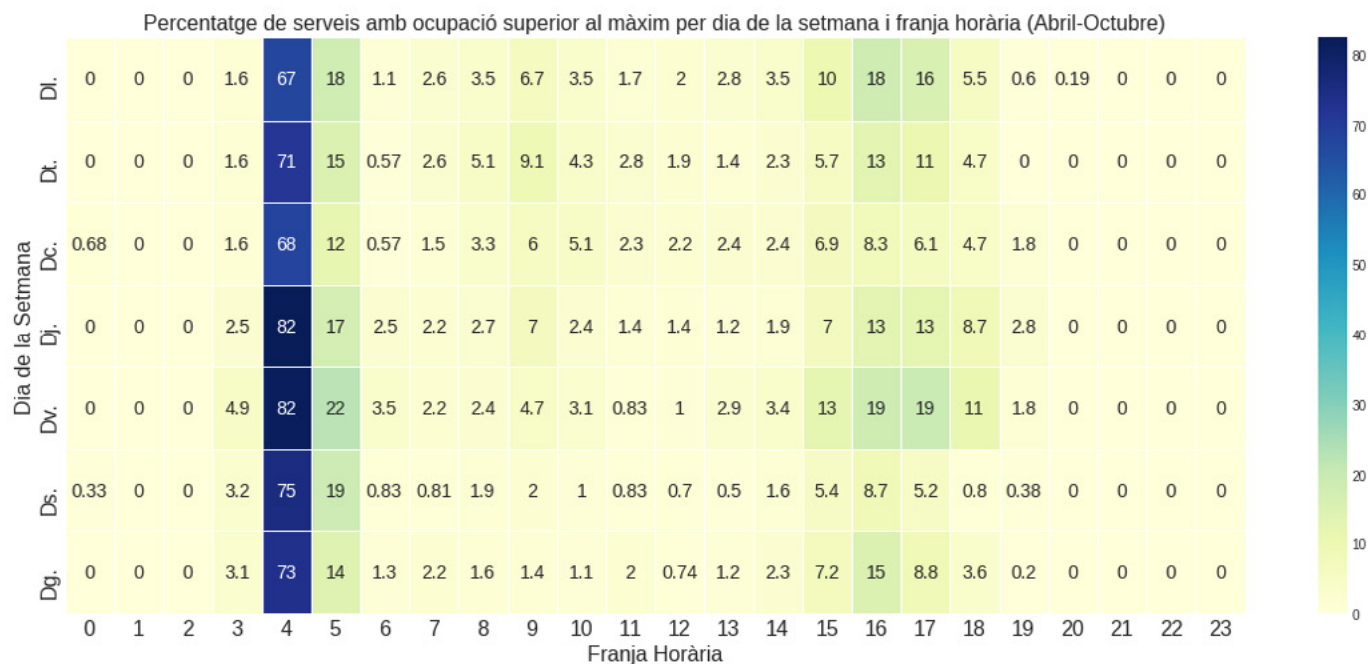
A1IDA

Temporada d'estiu



Aquest mapa de calor mostra l'ocupació mitjana en percentatge per cada dia de la setmana i franja horària, corresponent al període d'abril a octubre. En el gràfic es poden identificar clarament pics d'ocupació més elevats durant certes hores del dia, especialment entre les 3h-5h i les 16h-17h.

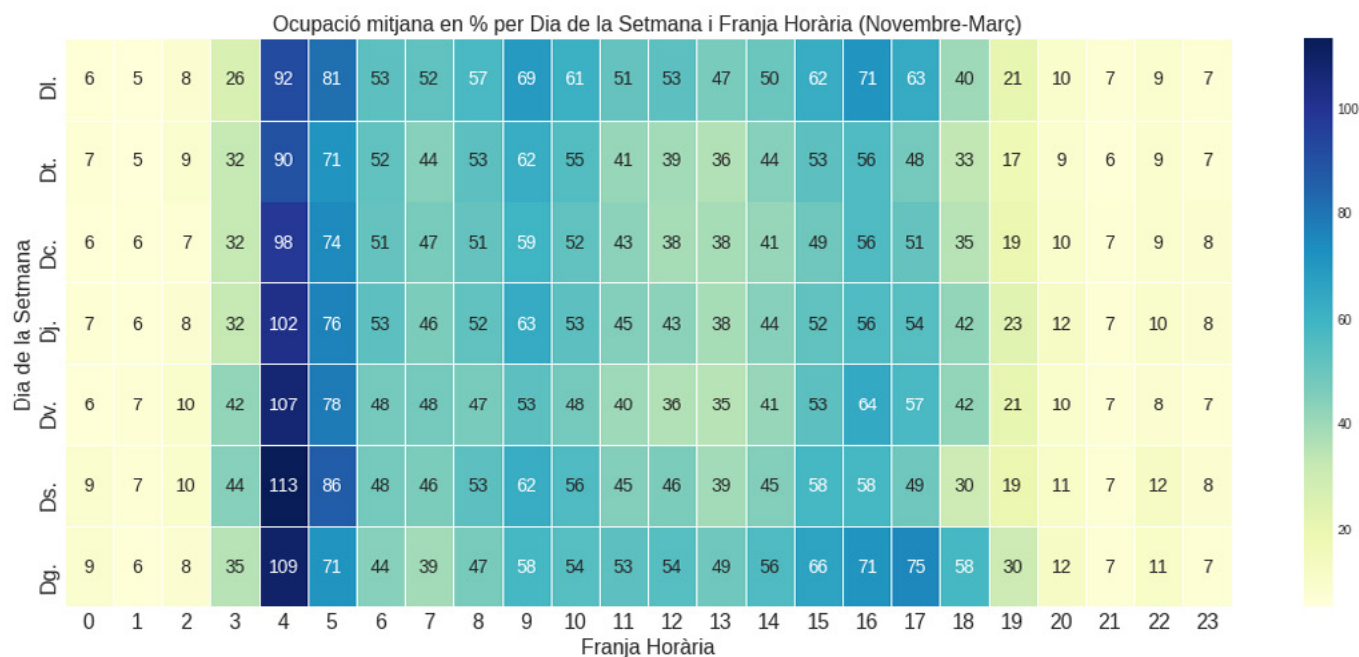
- **Franges horàries clau:** Es pot observar que les franges horàries compreses entre les 3h i les 18h tenen els percentatges d'ocupació més elevats, destacant especialment la franja de les 4h del matí, amb una ocupació mitjana per sobre del 100% tots els dies de la setmana.
- **Ocupació més baixa:** A partir de les 18h, l'ocupació baixa considerablement, arribant a mínims durant la nit, especialment entre les 20h i les 00h, on les ocupacions es mantenen per sota del 10%.



Aquest mapa de calor mostra el percentatge de serveis d'autobusos amb una ocupació superior al màxim permès per dia de la setmana i franja horària, per al període d'abril a octubre. Això indica quantes expedicions estan superant la capacitat màxima d'ocupació en cada franja horària i dia de la setmana.

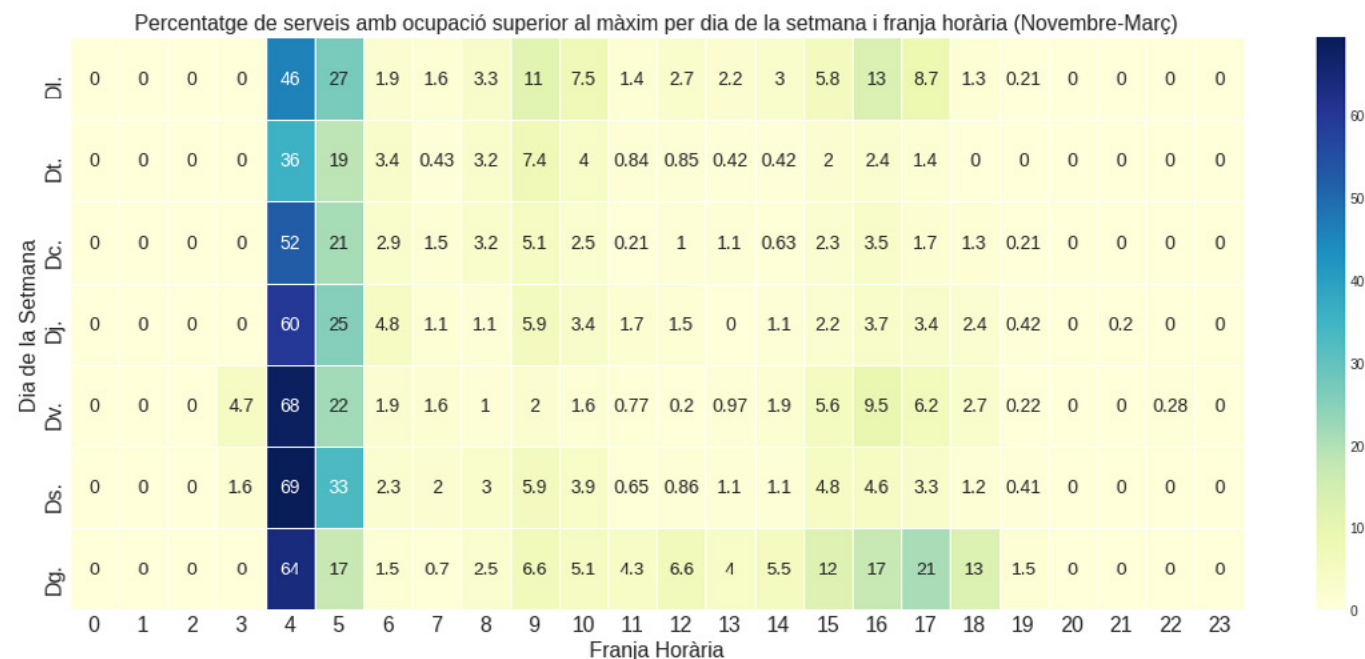
- **Franja horària amb més sobreocupació:** La franja de les 4h del matí és clarament la més crítica, amb valors que arriben fins al 82% de serveis sobreocupats en dies com dijous i divendres. En altres dies, com el dissabte i el diumenge, aquesta franja també registra un percentatge alt de sobreocupació, entre el 73% i el 75%.
- **Reducció progressiva durant el dia:** A partir de les 6h del matí, el percentatge de serveis amb sobreocupació es redueix notablement. Tot i això, continuen existint pics d'ocupació superior en algunes franges, especialment entre les 15h i les 18h, amb valors que arriben fins al 19% de serveis sobreocupats.
- **Menys sobreocupació:** Després de les 17h, el percentatge de serveis amb ocupació superior al màxim disminueix dràsticament, amb gairebé cap servei sobreocupat a partir de les 20h.

Temporada d'hivern



En el **període d'hivern (novembre-març)**, es repeteix el patró d'ocupació elevat durant les primeres hores del matí, especialment entre les 4h i les 5h del matí amb sobre ocupacions freqüents en diversos dies de la setmana. Les franges d'ocupació més altes es distribueixen entre les 3h i 18h, seguides d'una disminució progressiva durant la tarda i la nit.

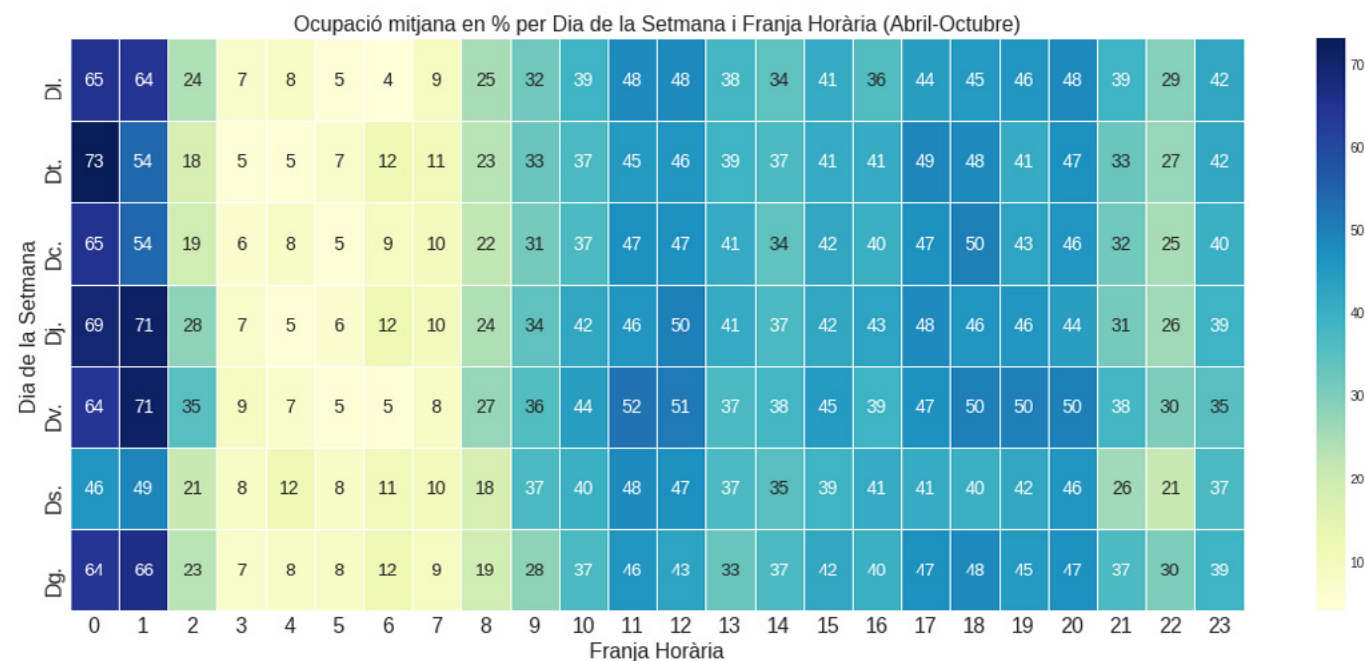
El que sí s'observa és que l'ocupació, tot i repetir el patró de l'estiu, és, generalment, més baixa en tots els dies i franges horàries.



Igualment, en el **període d'hivern (novembre-març)**, es repeteix el patró de sobreocupació elevada durant les franges de 4h a 5h (amb pic a les 4h del matí) i amb baixada de l'ocupació de progressiva durant el dia. Hi ha un lleuger augment de la sobreocupació entre les 9h i les 10h i entre les 15h i les 17h. Clarament, els dies de la setmana amb una sobreocupació més alta són els dilluns, divendres i diumenges.

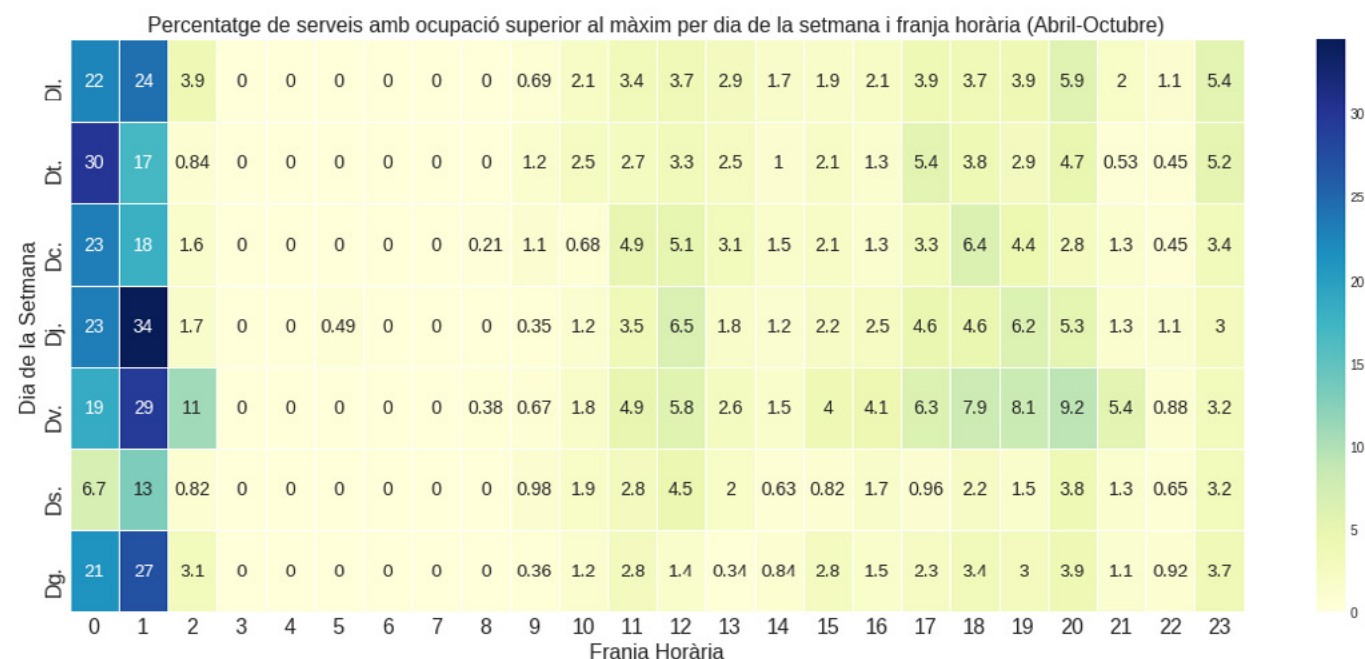
A1VUELTA

Temporada d'estiu



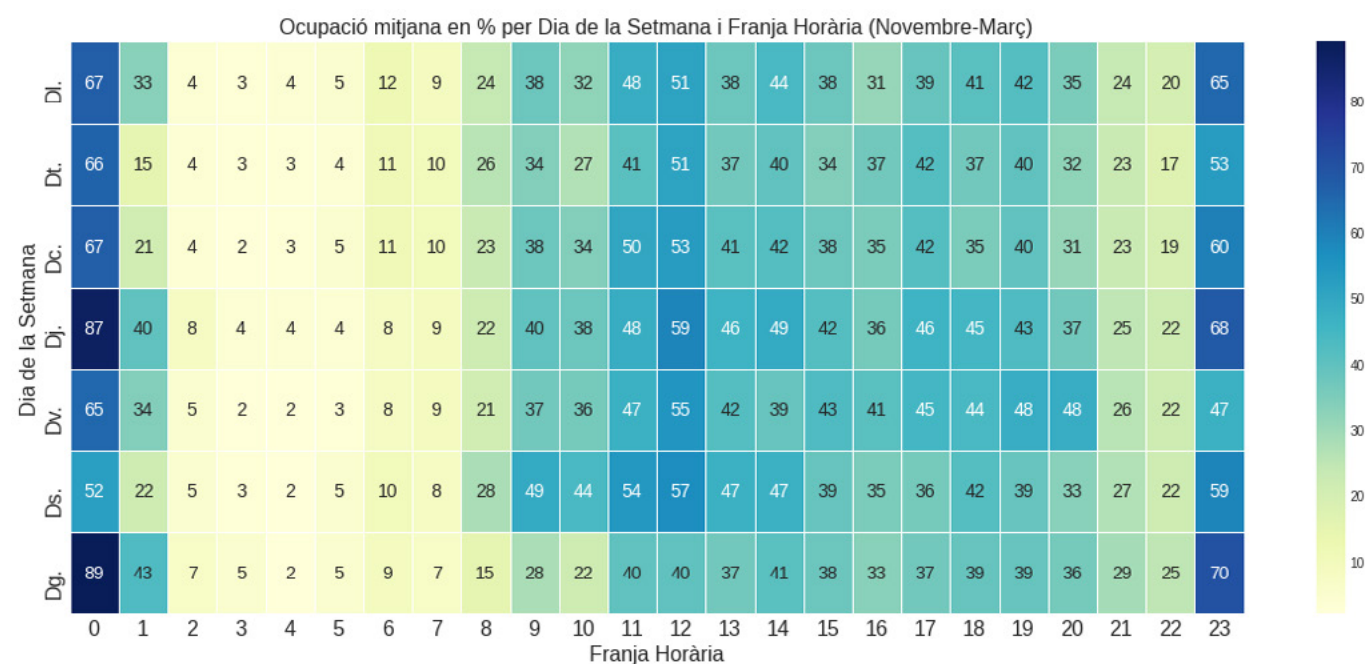
En aquests mapes de calor que mostren l'ocupació pels viatges de tornada des de l'aeroport T1, es poden identificar les franges horàries i els dies de la setmana en què l'ocupació és més alta:

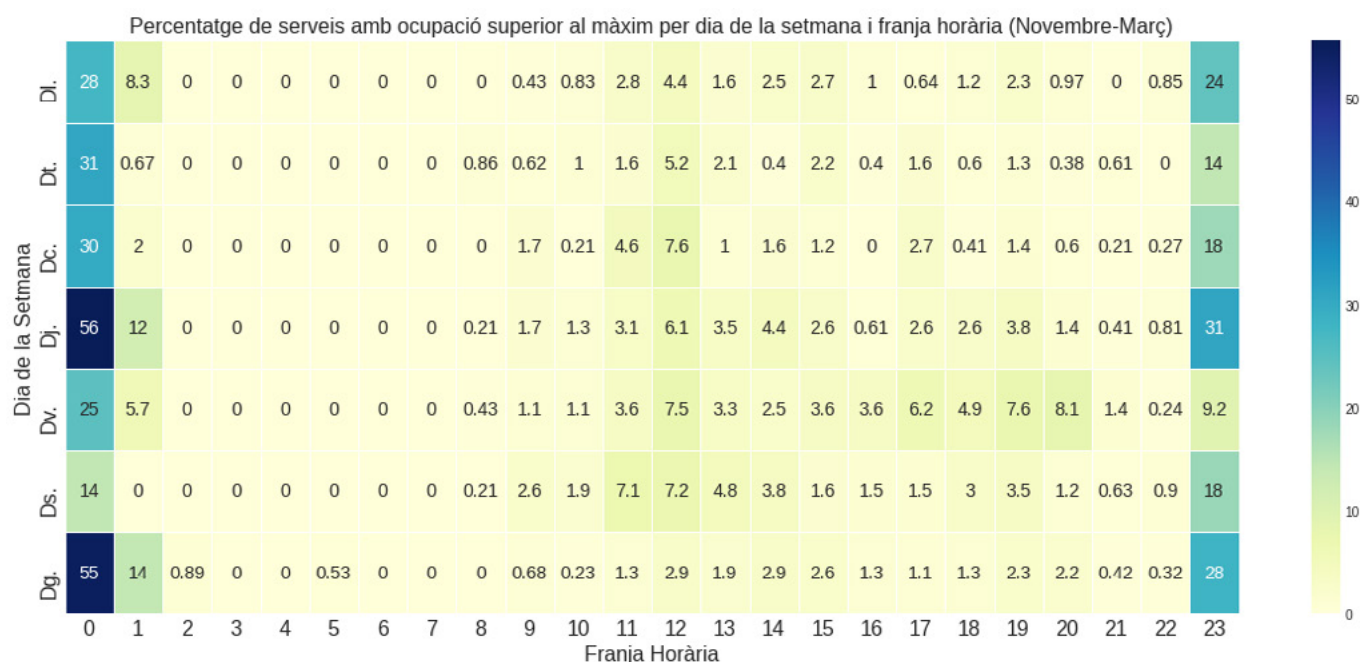
- **Franges horàries clau:** S'observa com l'ocupació va incrementant-se des de les 8h fins a la 1h de la matinada, per disminuir des de les 2h a les 7h (no arriben vols a l'aeroport). El pic d'ocupació es troba a la franja entre les 00h i la 1h de la matinada.
- **Ocupació més baixa:** L'ocupació disminueix significativament a partir de les 2h de la matinada, on els percentatges arriben fins a valors inferiors al 10% en moltes franges horàries.



Coincidint amb les dades d'ocupació, s'observa que les franges horàries on hi ha més sobreocupació d'autobusos són entre les 00h i la 1h de la matinada. A partir d'aquí, el període entre 3h i 7h pràcticament no pateix sobreocupació, a partir del qual s'incrementa fins a situar-se entre el 0-6%. S'observa que els divendres són els dies de la setmana amb les sobre ocupacions més elevades.

Temporada d'hivern

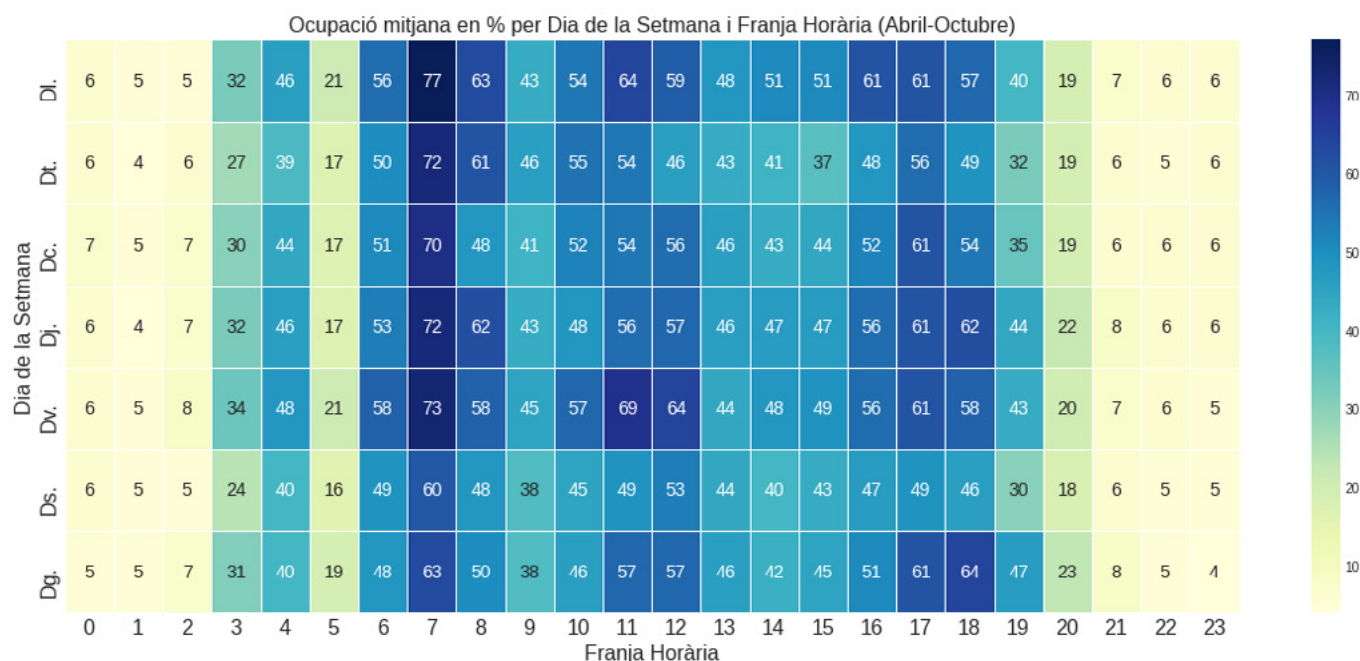




La temporada d'hivern repeteix el mateix patró d'ocupació mitjana i sobreocupació que la temporada d'estiu diferenciant-se en què la franja de les 2h a les 22h presenta valors inferiors d'ocupació mitjana respecte a l'estiu i que el pic d'ocupació s'avança una hora, entre les 23h i les 00h.

A2IDA

Temporada d'estiu

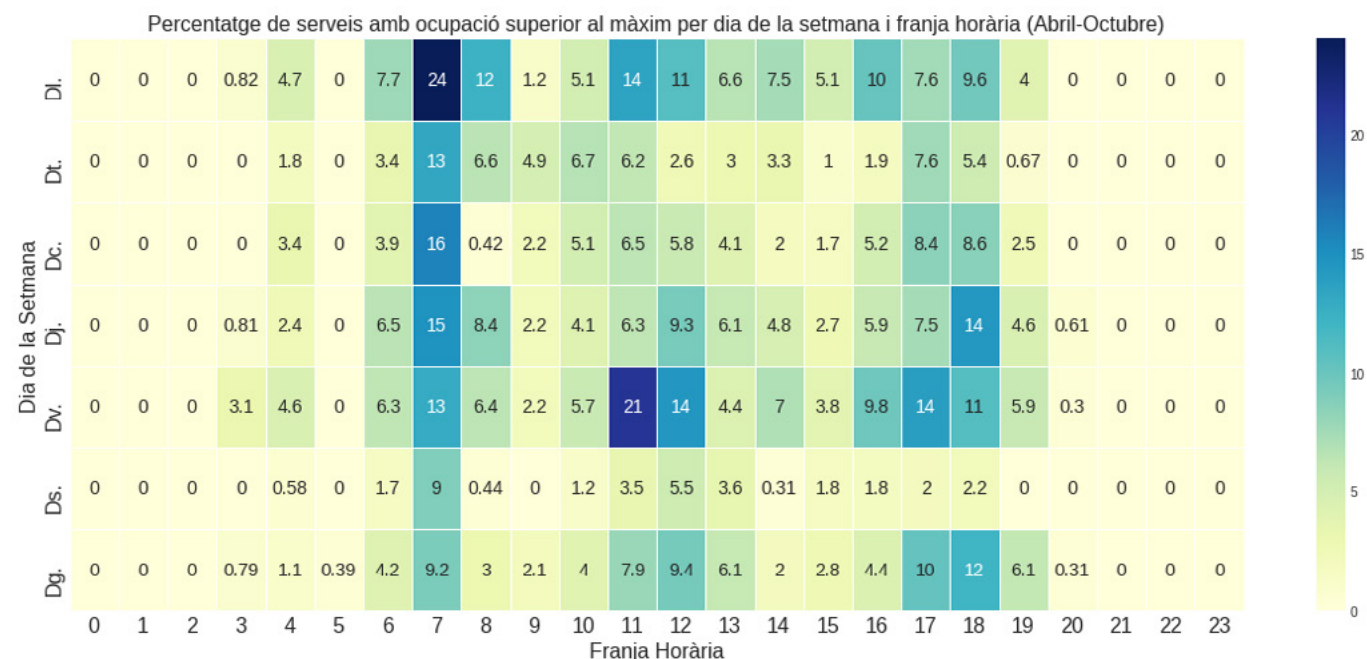


En el cas dels serveis que van a la T2, s'observa un patró similar a la línia T1 amb algunes diferències a destacar:

- La **franja horària amb més ocupació** és la de les 7h del matí, arribant els divendres a una ocupació mitjana dels serveis del 73%.

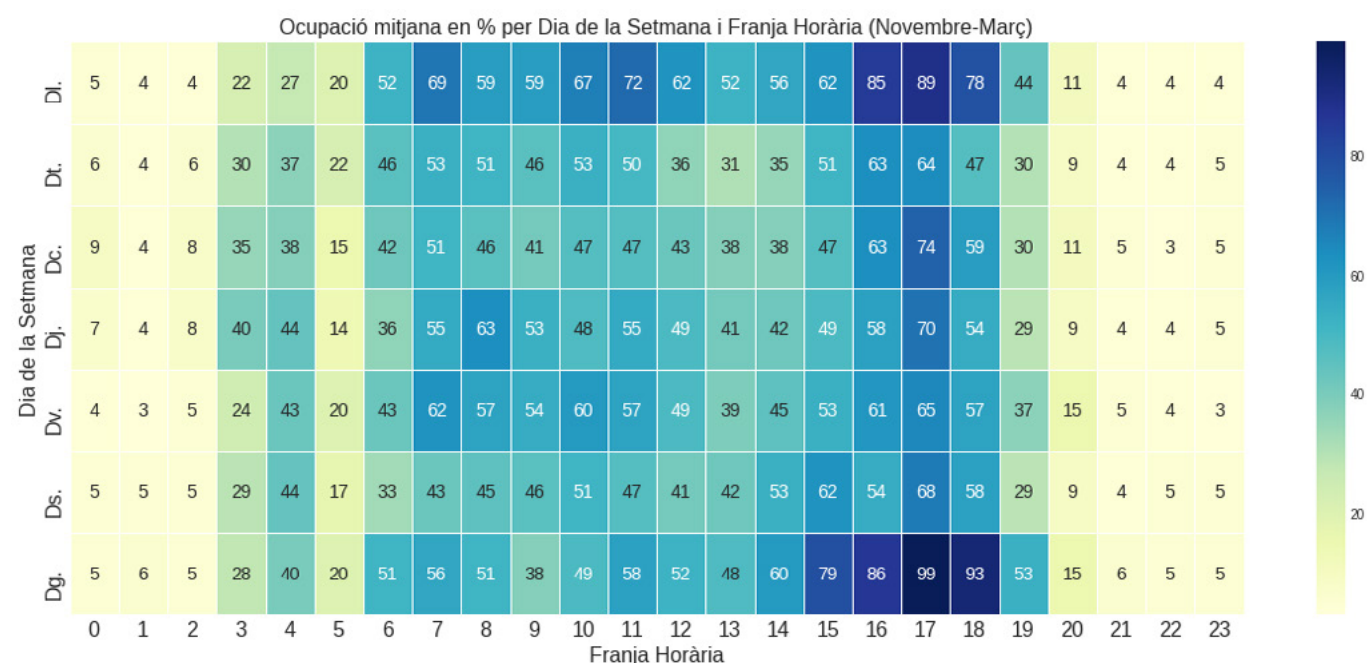
- **L'ocupació augmenta** a partir de les 3h del matí fins a les 20h (una hora més a que als serveis amb direcció a la T1). S'observa una disminució de l'ocupació a la franja de les 5h del matí.
- **L'ocupació es redueix** a partir de les 21h i fins a les 2h, on es manté per sota del 10%.

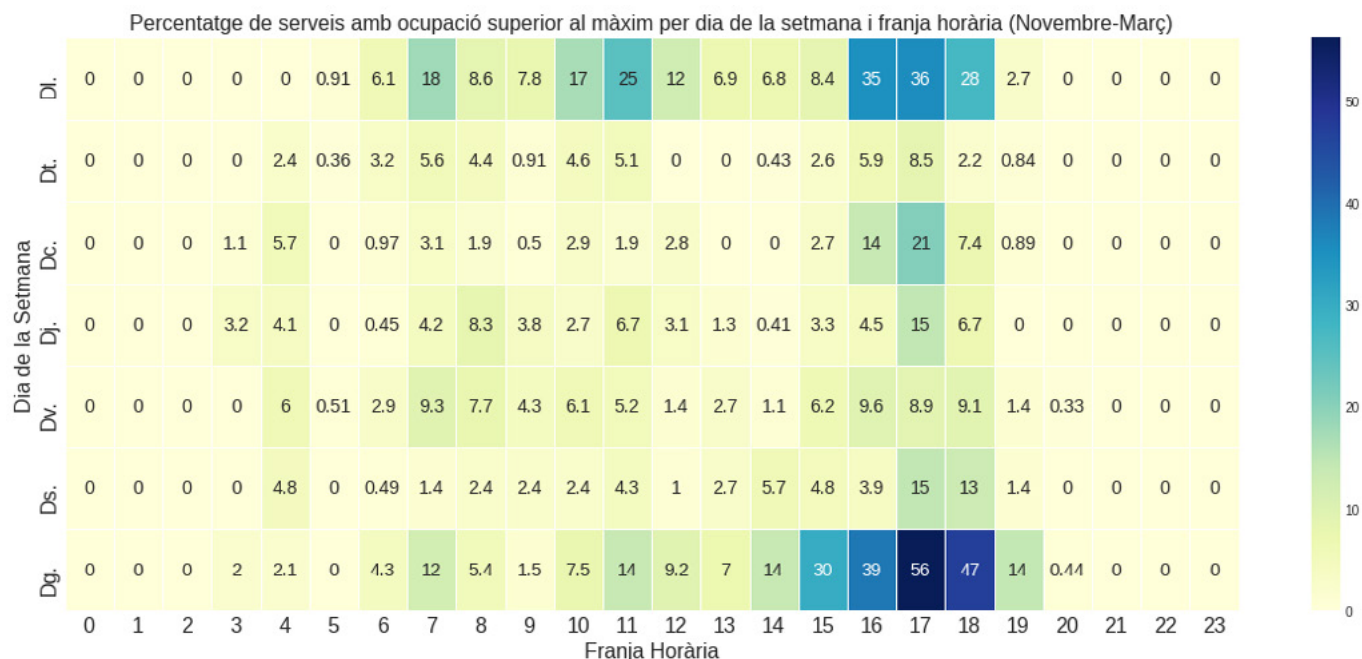
De manera general, les ocupacions mitjanes dels serveis en direcció a la T2 són superiors a les dels serveis en direcció a la T1.



S'observa que la sobreocupació dels serveis es manté pràcticament a 0 entre les 20h i les 2h del matí. A partir d'aquí, i coincidint amb el patró d'augment d'ocupació, la sobreocupació augmenta fins a les 19h de la tarda, arribant al pic màxim a la franja de les 7h (hi ha una franja, la de les 5h, on torna a 0).

Temporada d'hivern



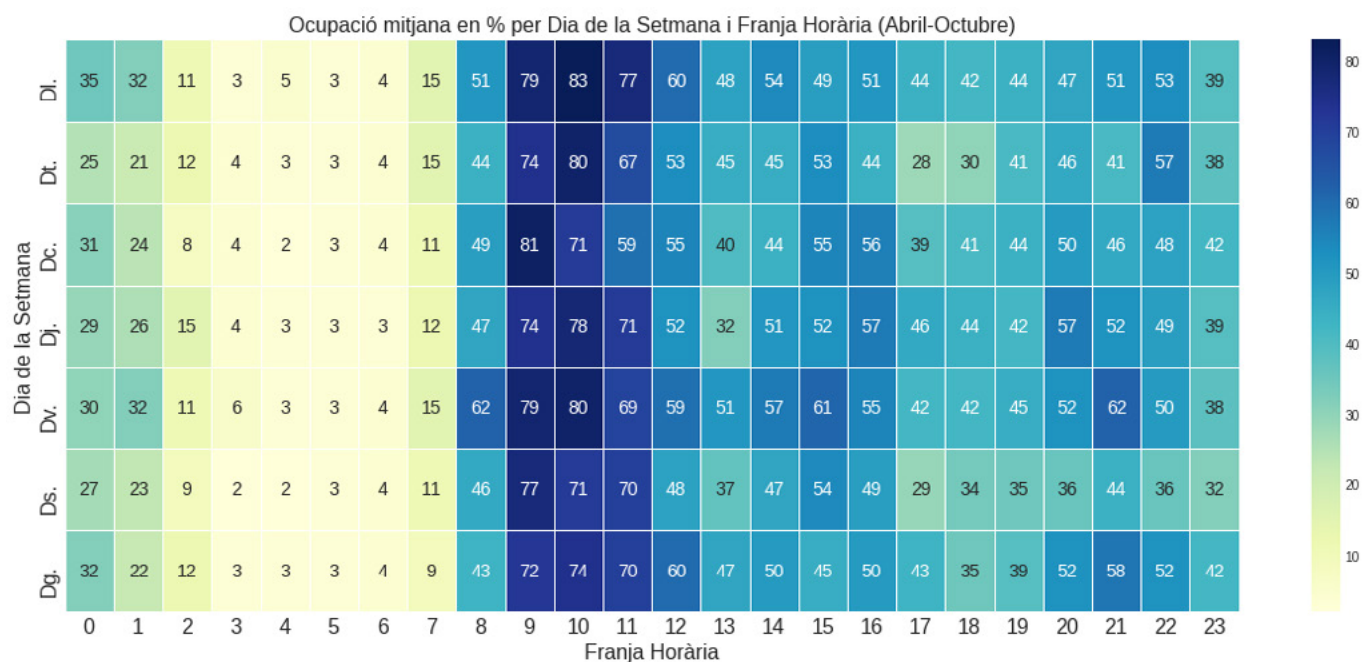


Els patrons observats a l'estiu es repeteixen clarament a l'hivern. Es pot veure com els dimegnes i els dilluns presenten una ocupació mitjana més elevada respecte a la resta de dies. En particular, hi ha un període entre les 15h i les 18h on l'ocupació arriba a valors mitjans del 99%.

Pel que fa a la sobreocupació, aquest patró també es manté, destacant els dimegnes i els dilluns, on entre les 16h i les 18h hi ha línies que estan sobreocupades entre el 35% i el 56% de la seva capacitat.

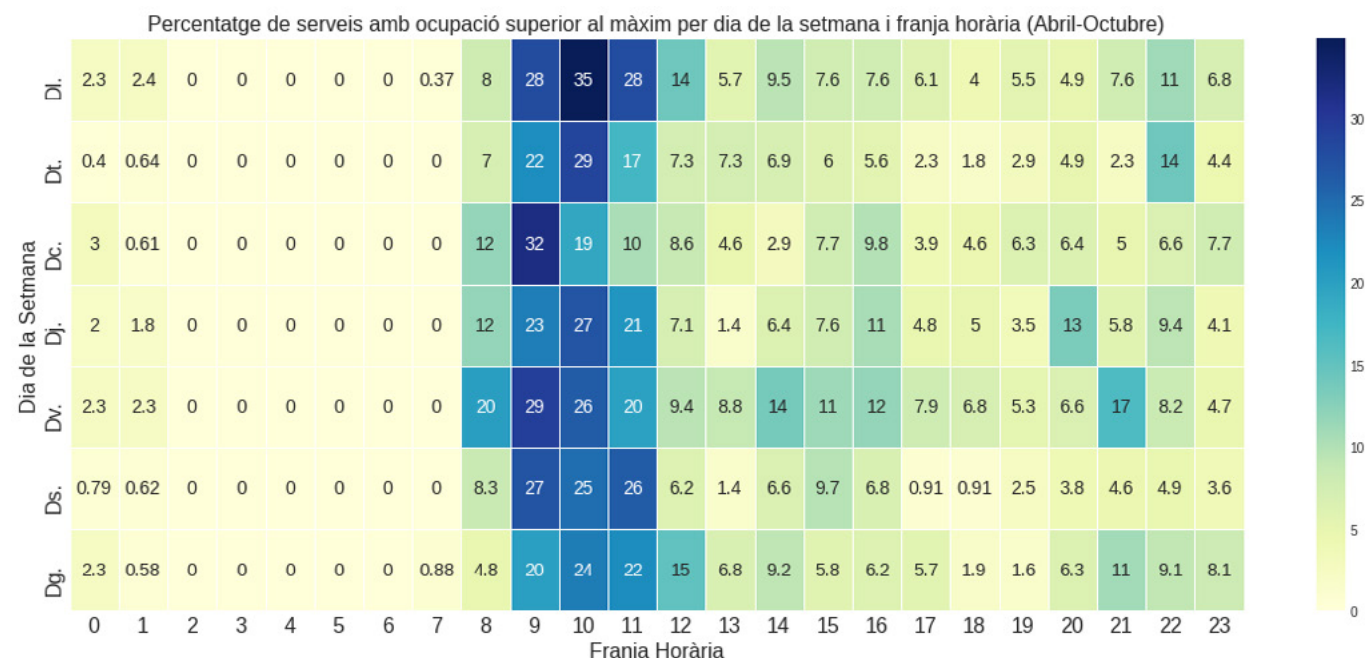
A2VUELTA

Temporada d'estiu



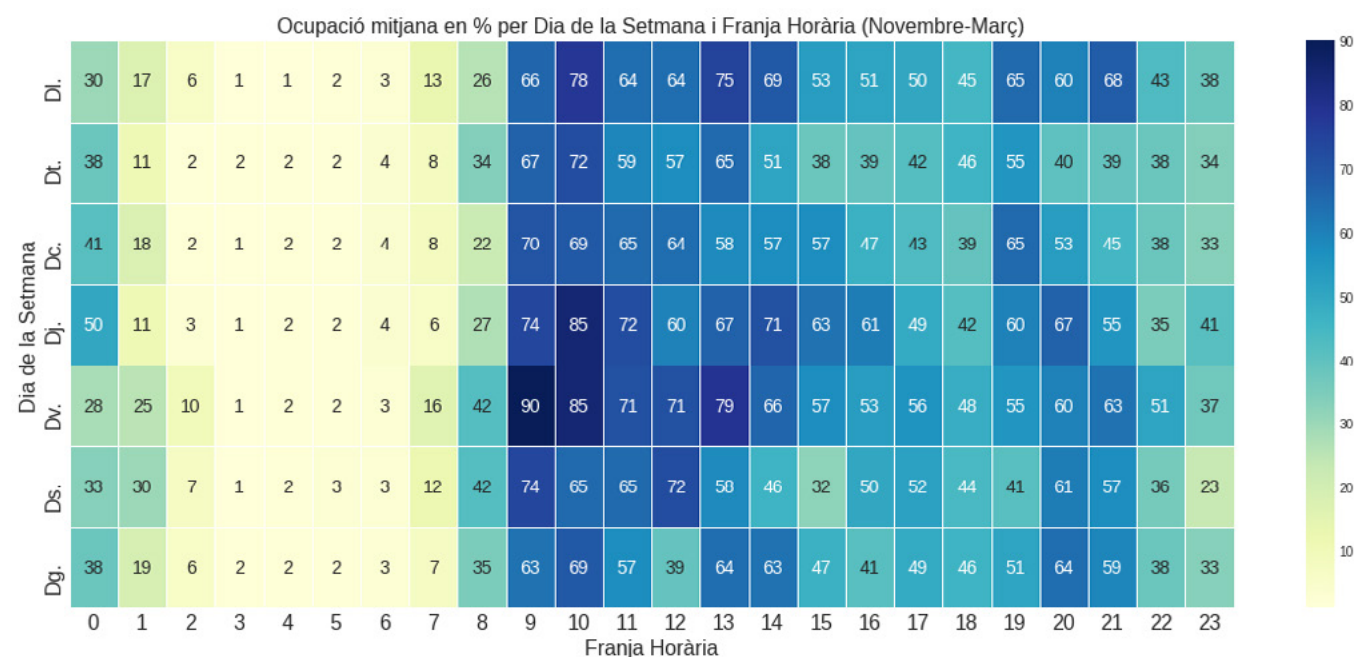
Aquest mapa de calor mostra clarament un augment ràpid de l'ocupació a partir de les 7h del matí, començant amb un 4% i arribant al seu màxim, al voltant del 80%, entre les 9h i les 11h del matí. Després d'aquest pic,

l'ocupació baixa novament cap a les 12h, situant-se al 50%, i es manté estable entre el 30% i el 50% fins a les 23h. A partir d'aquesta hora, l'ocupació disminueix ràpidament, oscil·lant entre el 10% i el 15% fins a les 2h de la matinada, i baixant encara més, fins a un 3%, durant la franja de les 3h a les 6h del matí.

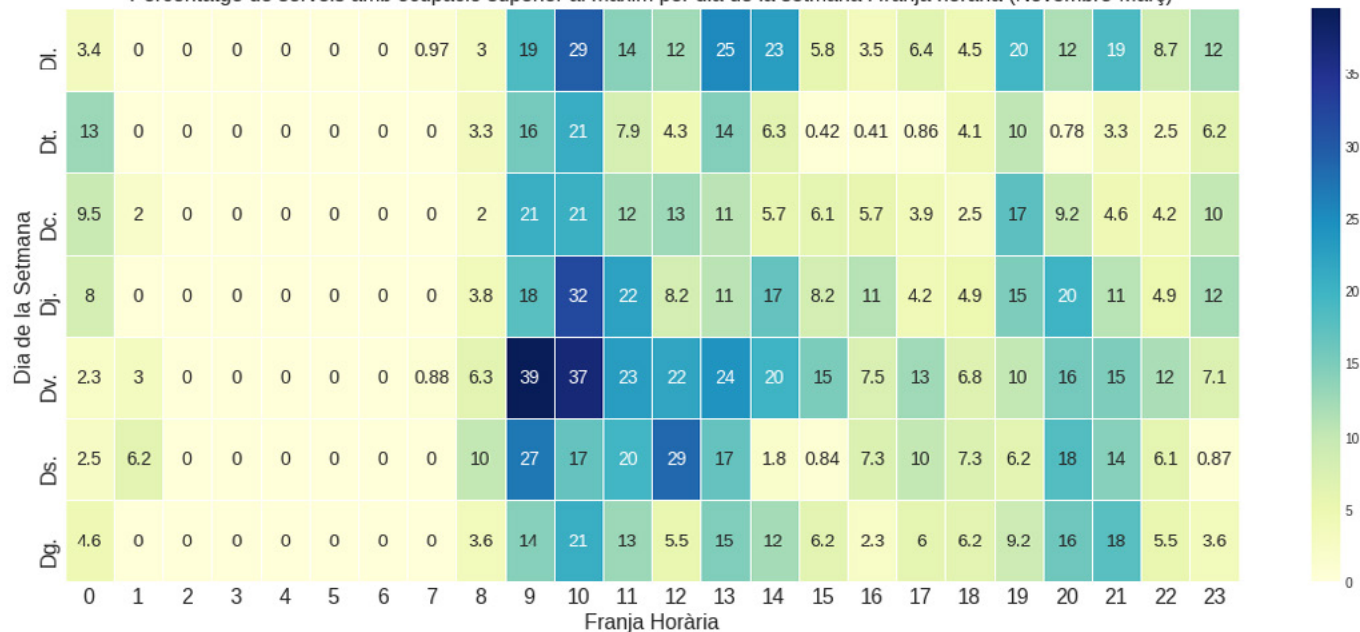


Pel que fa a la sobreocupació, aquesta segueix un patró similar, amb el seu pic màxim entre les 9h i les 11h del matí, on fins al 36% dels serveis superen els 49 passatgers. Després de les 12h, la sobreocupació disminueix gradualment, mantenint-se per sota del 10% fins a les 23h. En aquesta hora, es produeix una segona caiguda significativa, amb valors que oscil·len entre el 4% i el 8%, fins a arribar a un mínim d'entre 0,6% i 3%. A partir de les 2h de la matinada i fins a les 7h del matí, cap servei supera la capacitat de passatgers.

Temporada d'hivern



Percentatge de serveis amb ocupació superior al màxim per dia de la setmana i franja horària (Novembre-Març)



Els mapes de calor de l'hivern comparteixen patró amb els d'estiu.



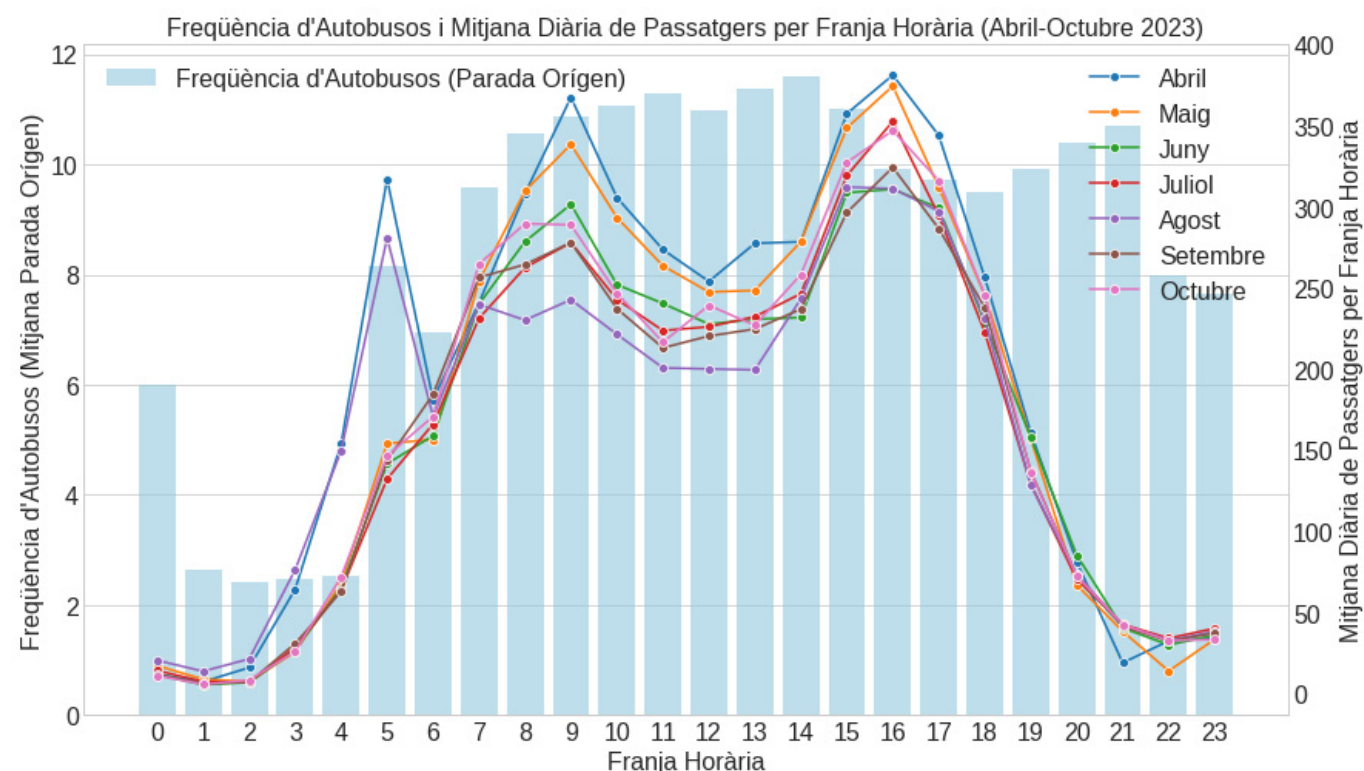
4.4. Distribució horària de passatgers i freqüència mitjana d'autobusos

En aquesta secció, es presenten diversos gràfics que comparen la distribució horària de passatgers (mitjana per franja horària) i la quantitat d'autobusos per franja horària, diferenciats per temporada d'hivern i d'estiu. Aquests gràfics ens permeten visualitzar l'evolució del nombre de passatgers al llarg del dia en comparació amb la freqüència dels autobusos (en nombre de trajectes), amb l'objectiu de contrastar si la variació en la demanda de passatgers està alineada amb els horaris dels serveis d'autobusos o si existeixen discrepàncies entre ambdues variables.

A l'eix principal y, trobem la dada de freqüència mitjana d'autobusos, que correspon amb les barres del gràfic, mentre que l'eix secundari indica la quantitat mitjana de passatgers per franja horària i, amb diferents colors, per mes.

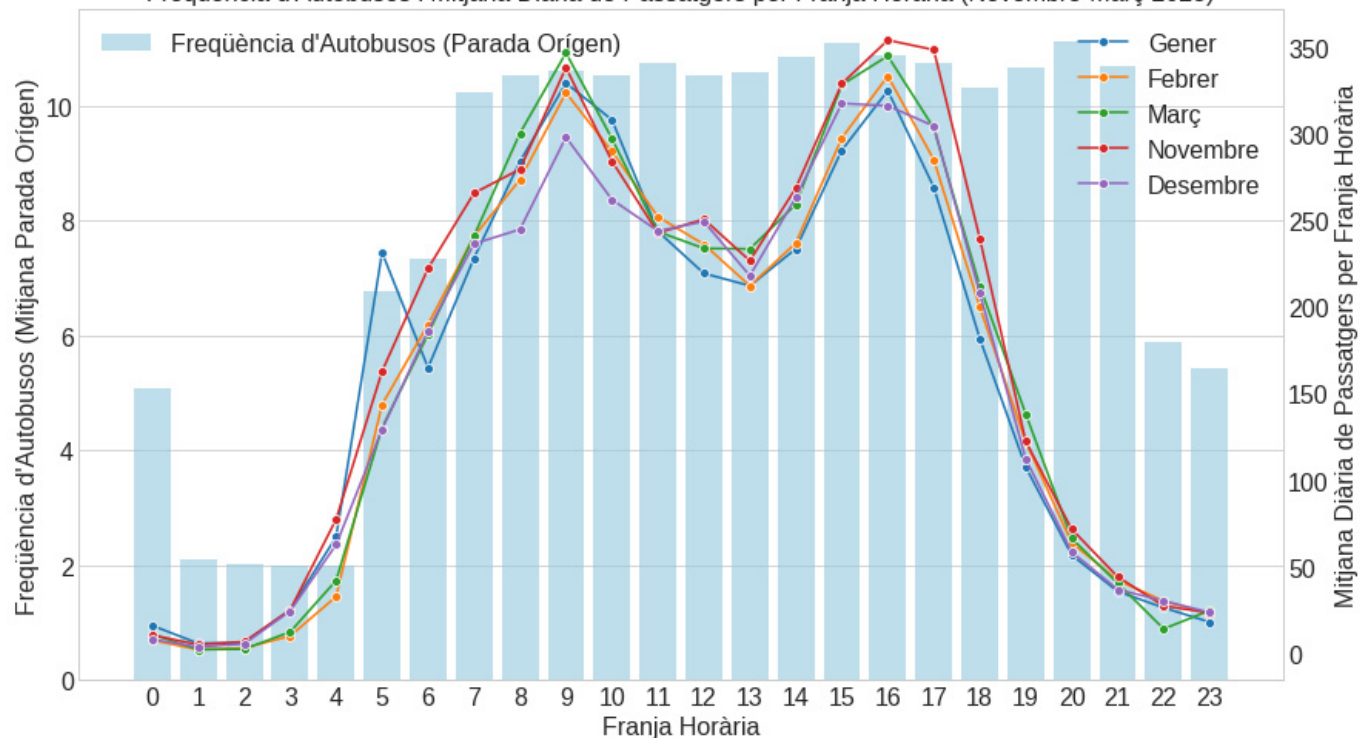
A1IDA

Temporada d'estiu



Temporada d'hivern

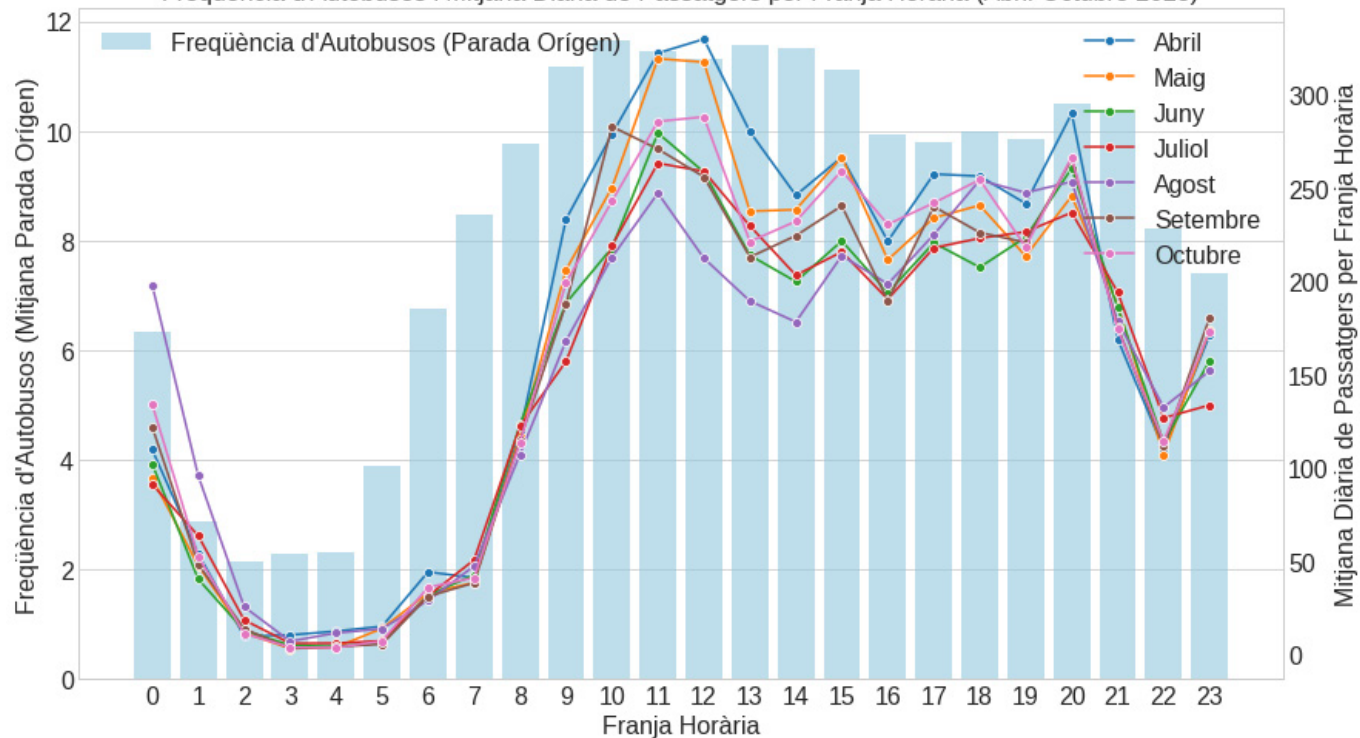
Freqüència d'Autobusos i Mitjana Diària de Passatgers per Franja Horària (Novembre-Març 2023)



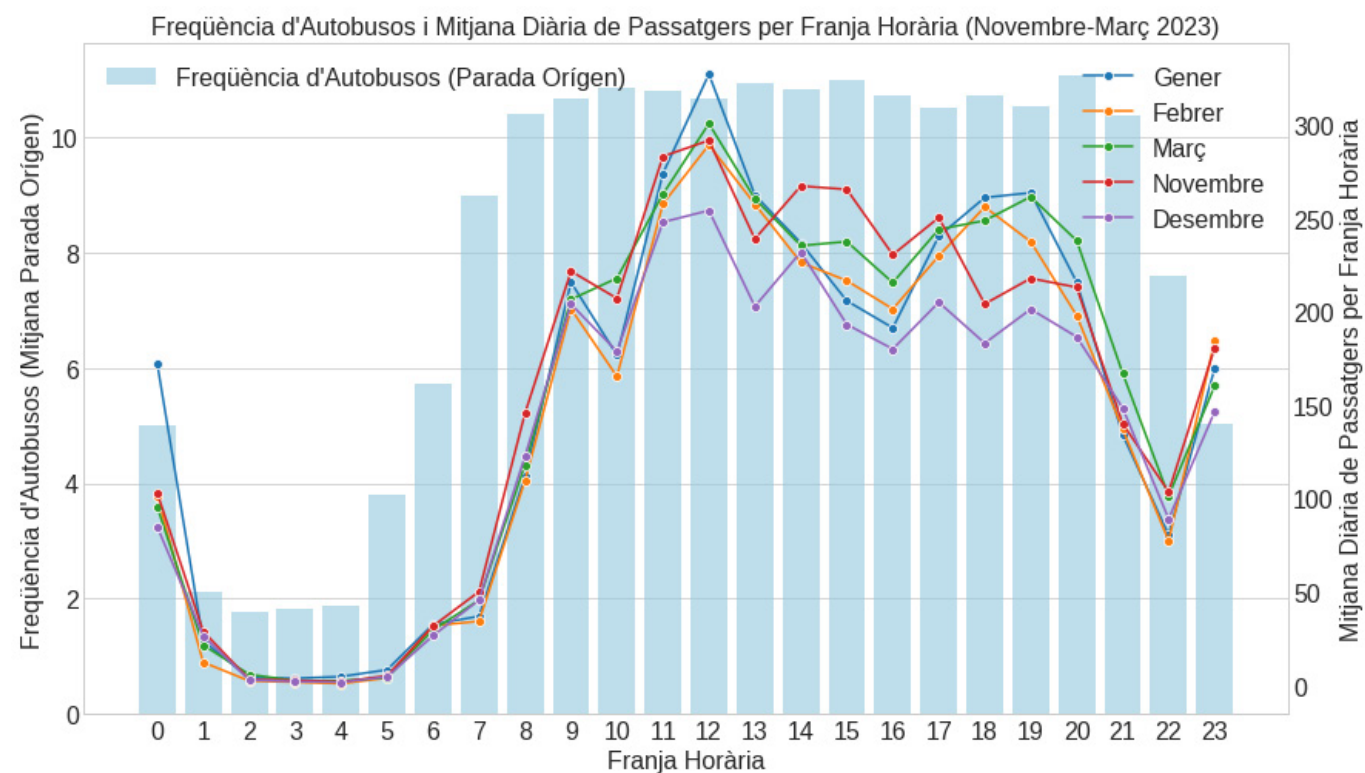
A1VUELTA

Temporada d'estiu

Freqüència d'Autobusos i Mitjana Diària de Passatgers per Franja Horària (Abril-Octubre 2023)

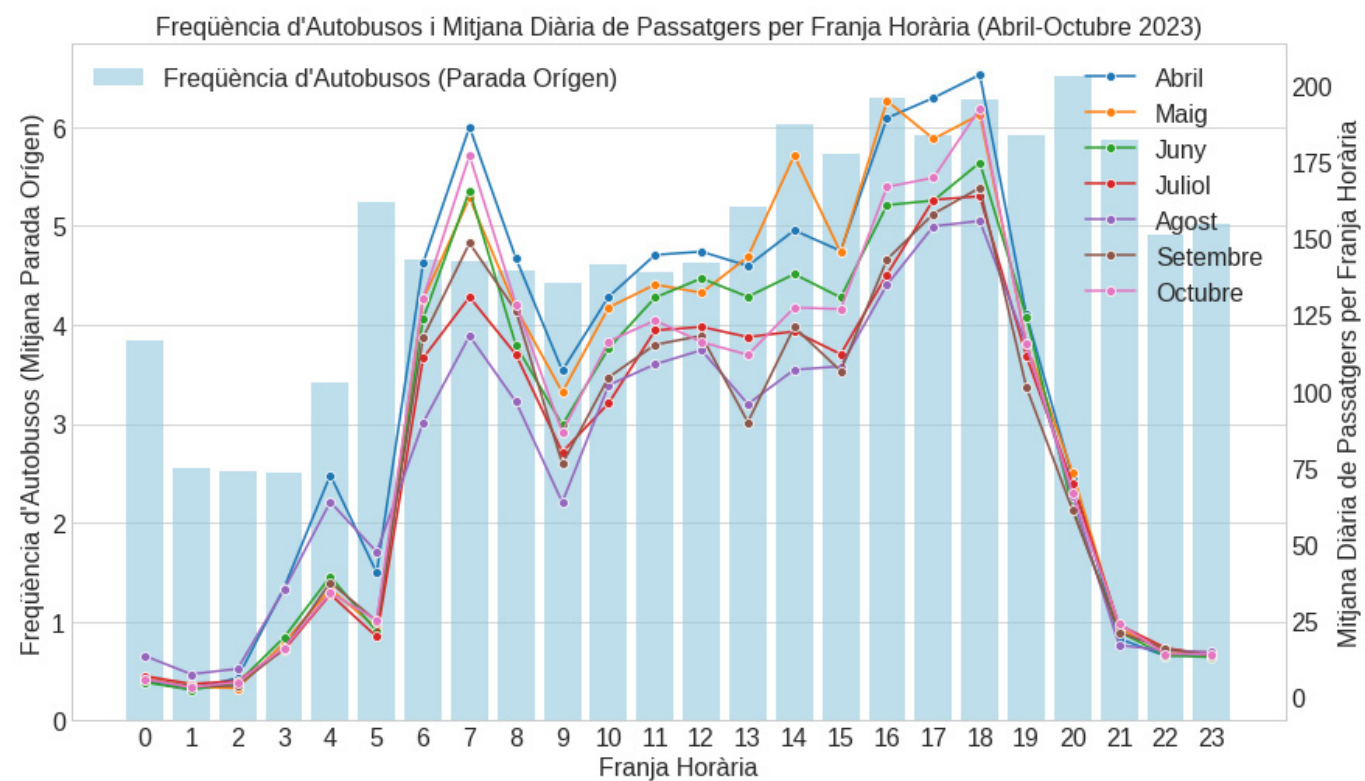


Temporada d'hivern

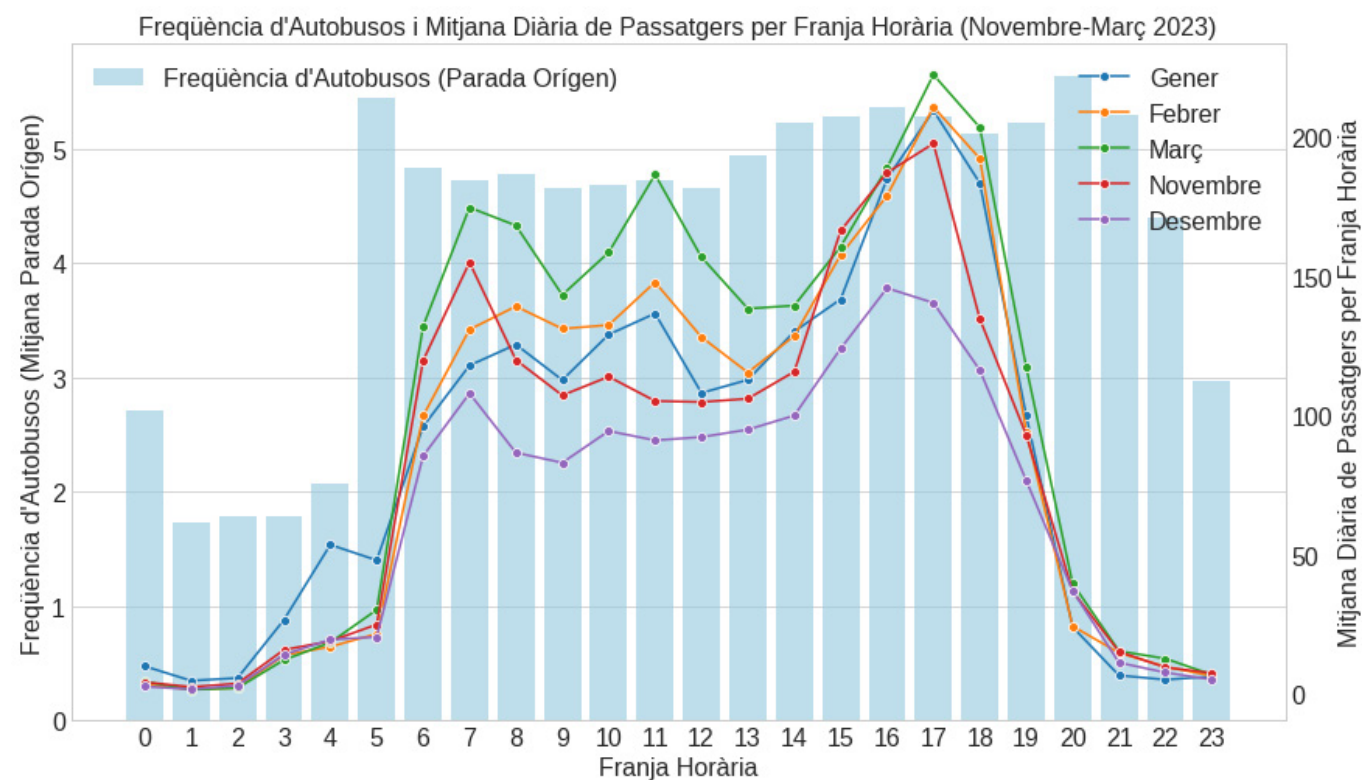


A2IDA

Temporada d'estiu

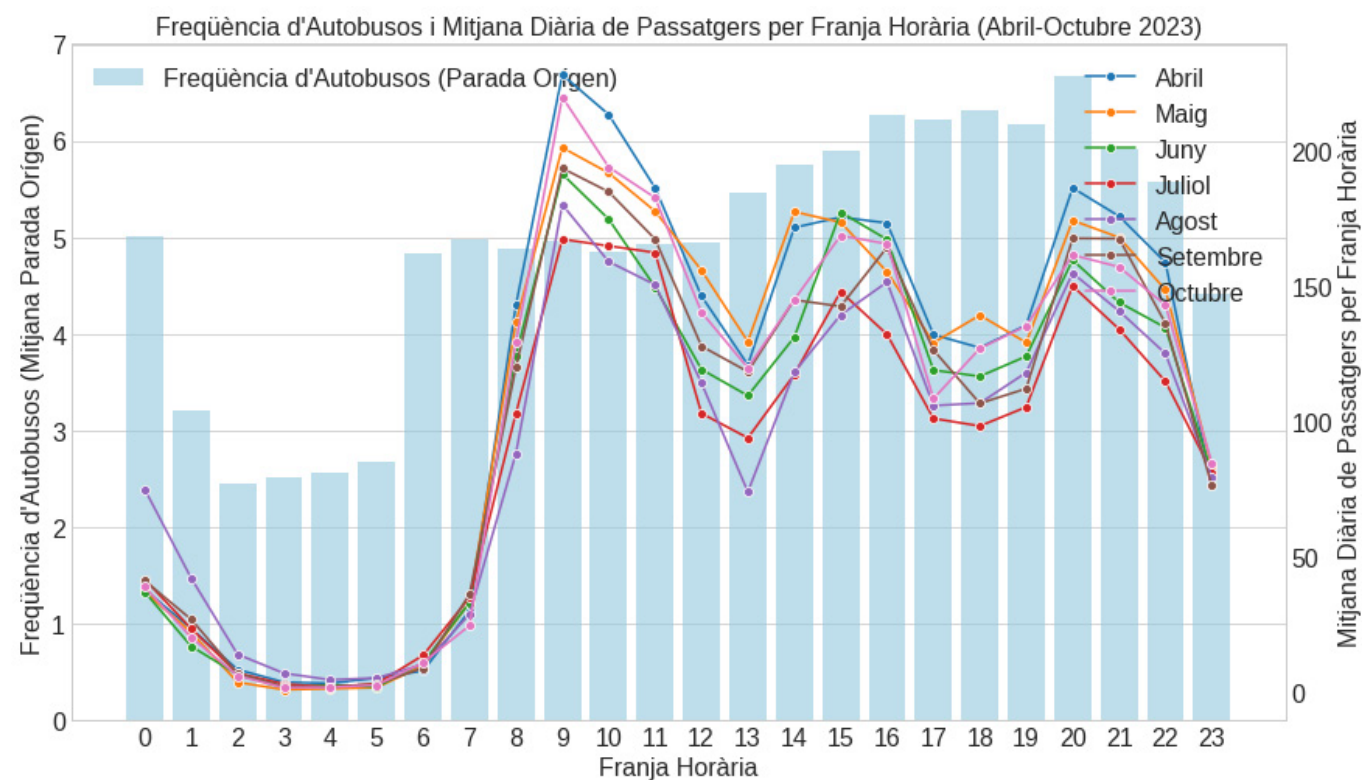


Temporada d'hivern

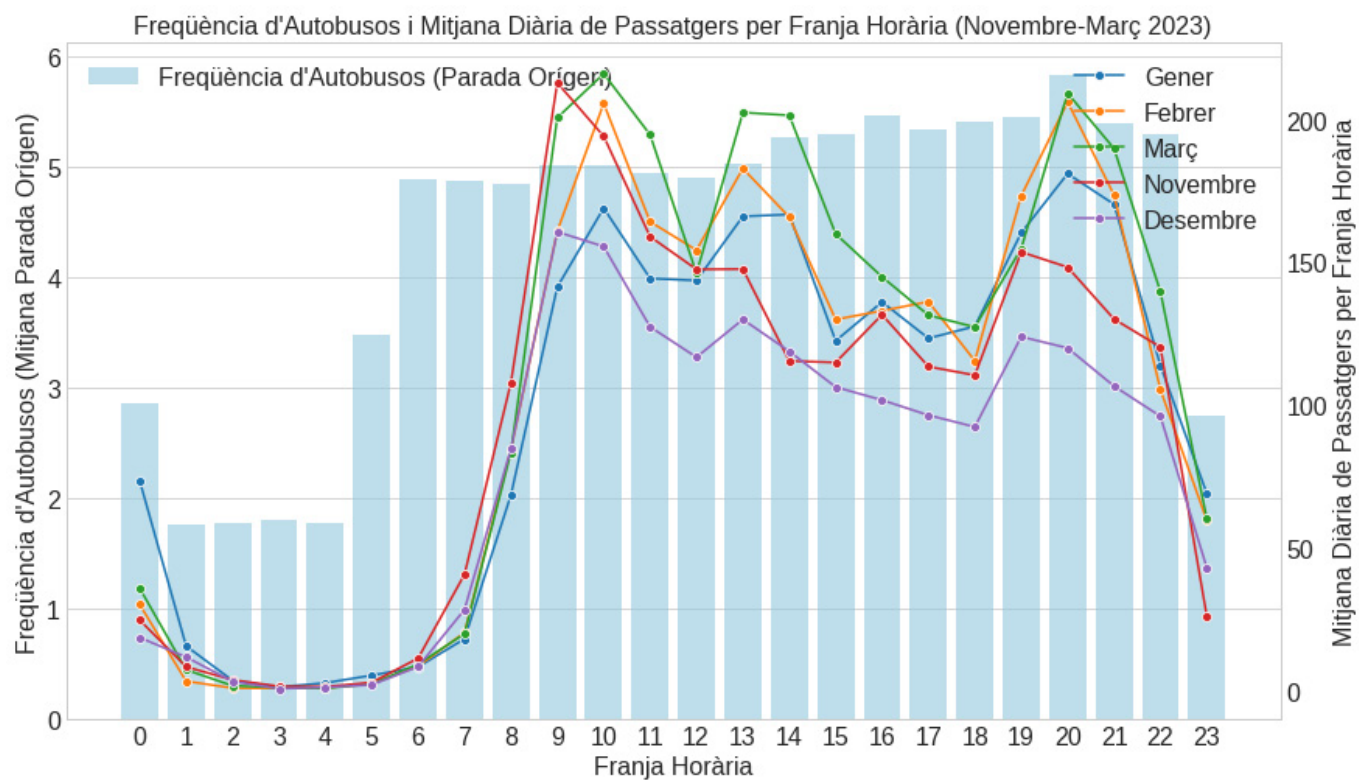


A2VUELTA

Temporada d'estiu



Temporada d'hivern



4.5. Temps mitjà dels viatges per franja horària

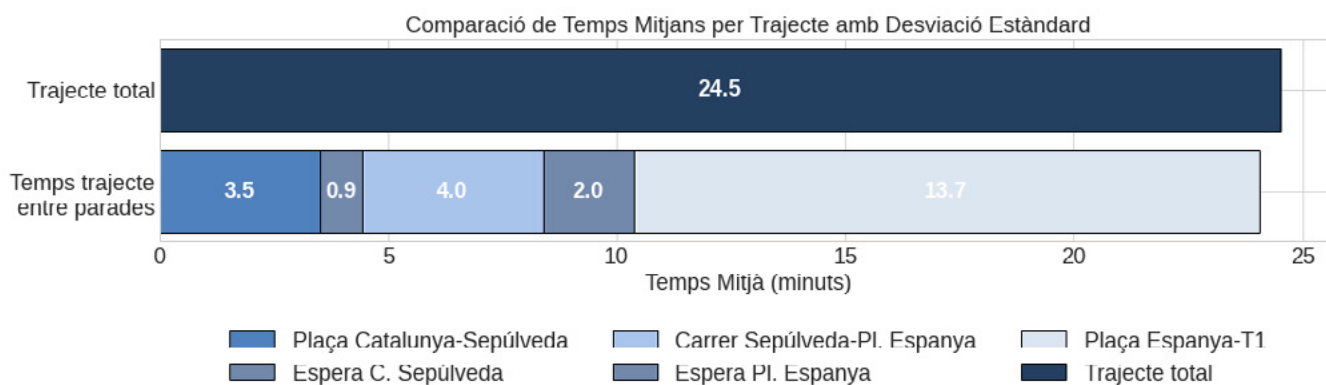
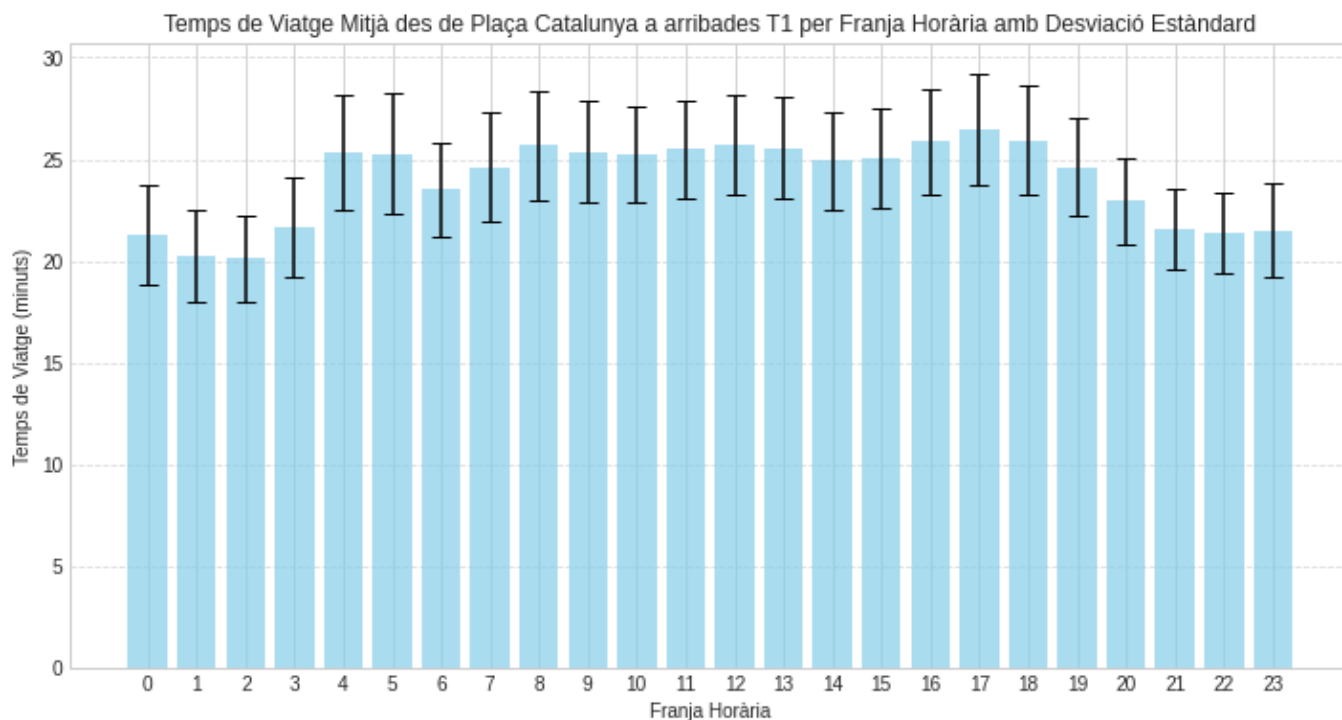


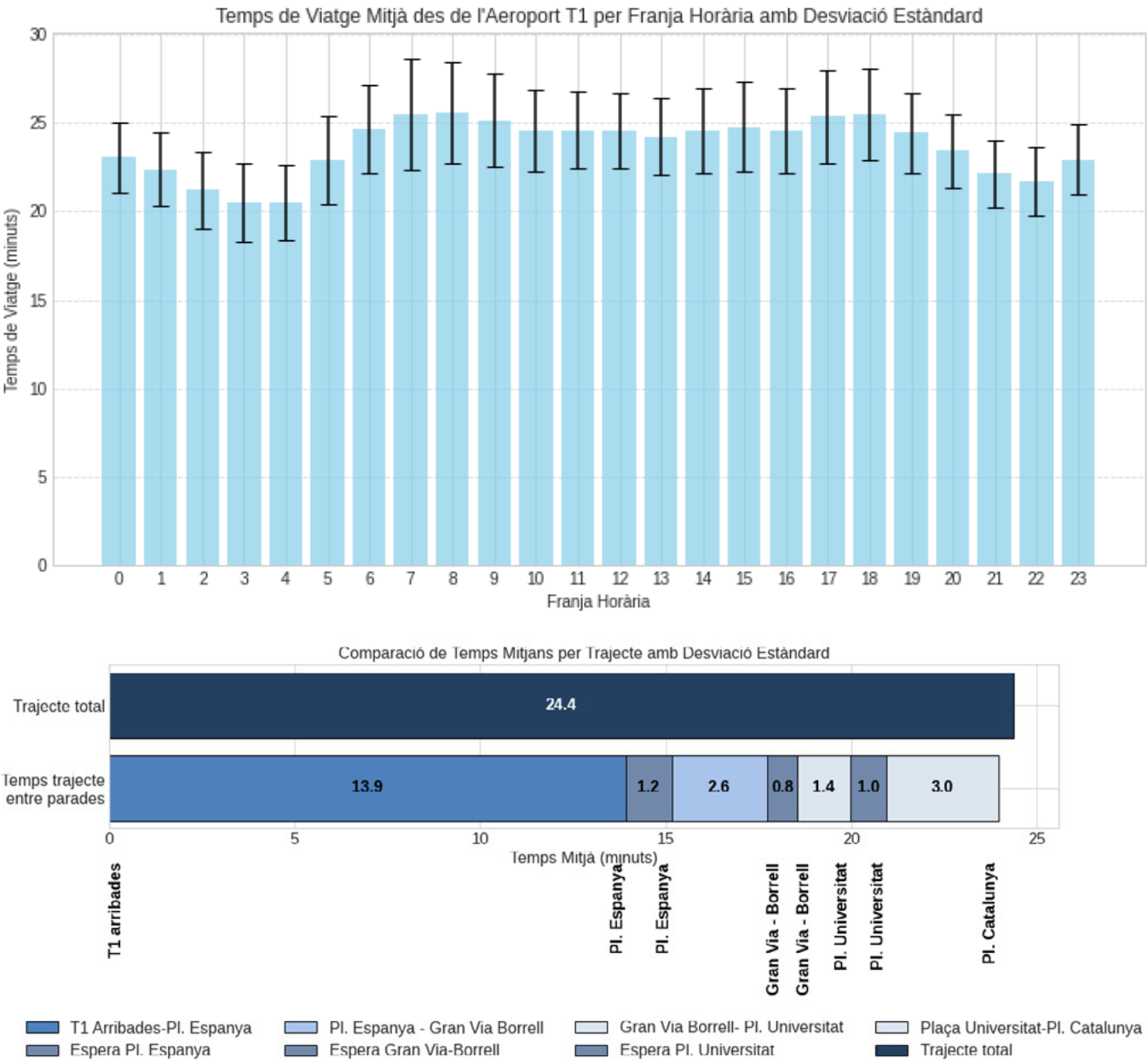
Finalment, en aquesta secció s'inclouen dues taules amb els temps mitjans dels trajectes d'anada i tornada tant a la T1 com a la T2, acompanyats de la desviació estàndard. Aquests temps s'han desglossat per franja horària amb l'objectiu d'identificar si hi ha moments del dia en què el trànsit afecta la durada del viatge, provocant que aquest sigui més llarg.

A més, es presenta un gràfic per cada trajecte, que mostra les dades dels temps mitjans i la seva desviació estàndard. També s'inclou un cronograma del trajecte per parades, on es detalla la duració del trajecte entre cada parada i els temps mitjans d'espera necessaris per a l'embarcament dels passatgers durant el trajecte (no s'inclou el temps de pujada en origen ni el temps de baixada en destinació). Aquesta informació permet una anàlisi més precisa dels temps de viatge i ajuda a identificar punts crítics d'espera o de variabilitat en la duració del recorregut.

A1

	A1IDA		A1VUELTA	
Franja horària	Mitjana (min)	Std (min)	Mitjana (min)	Std (min)
0	21.307327	2.474441	23.036143	2.006204
1	20.290041	2.256542	22.382076	2.114170
2	20.126281	2.083309	21.192984	2.206000
3	21.661756	2.438782	20.470402	2.193064
4	25.333809	2.850192	20.481602	2.121036
5	25.282007	2.967620	22.900283	2.501857
6	23.512213	2.304573	24.640655	2.489520
7	24.634558	2.655564	25.484117	3.156510
8	25.672026	2.672919	25.544961	2.841467
9	25.368764	2.476887	25.114414	2.621469
10	25.261654	2.347720	24.585207	2.317893
11	25.482870	2.365195	24.577803	2.164805
12	25.716441	2.468203	24.537740	2.133886
13	25.560656	2.482460	24.216690	2.144261
14	24.935248	2.373843	24.567649	2.401935
15	25.082817	2.434471	24.749491	2.528460
16	25.881628	2.579805	24.554994	2.369700
17	26.505014	2.725586	25.363808	2.632563
18	25.940608	2.676408	25.481354	2.563668
19	24.627874	2.390658	24.417337	2.295027
20	22.951131	2.131675	23.418496	2.099285
21	21.601326	1.962452	22.122020	1.883845
22	21.373908	1.952028	21.729666	1.929002
23	21.512239	2.283310	22.938441	2.024777

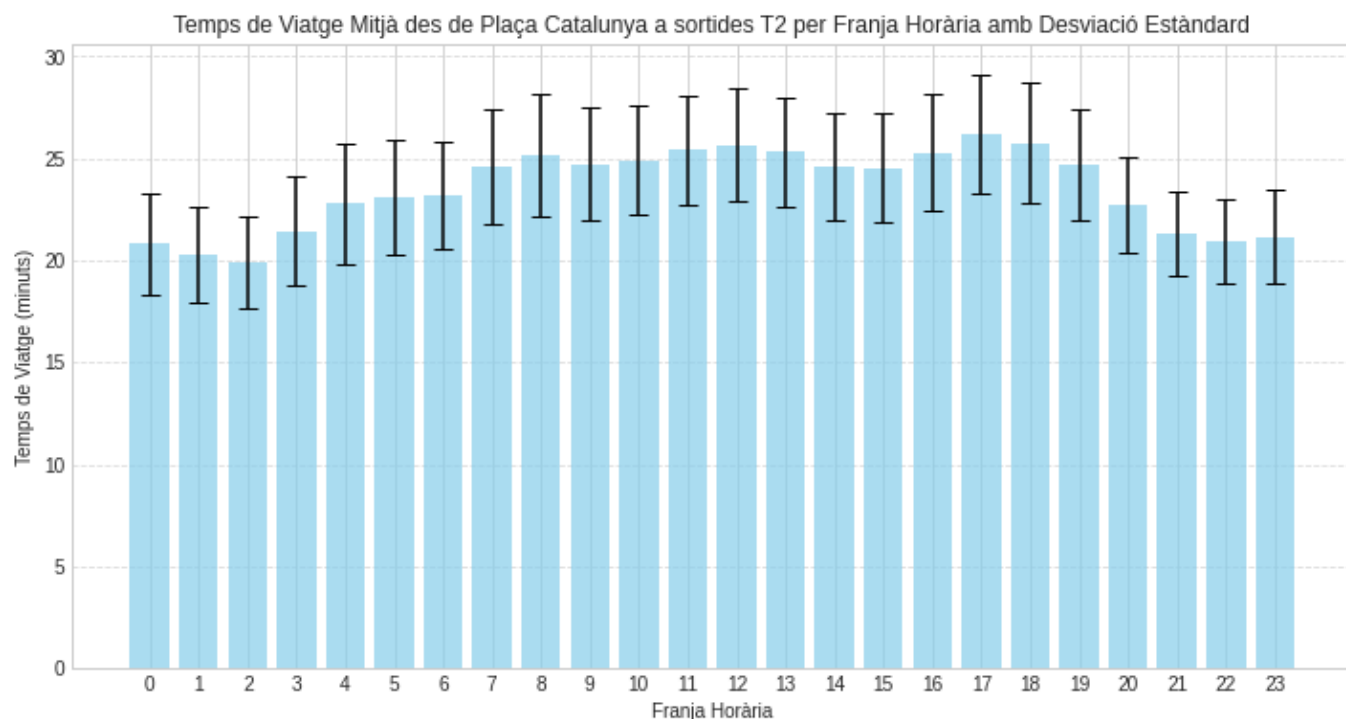


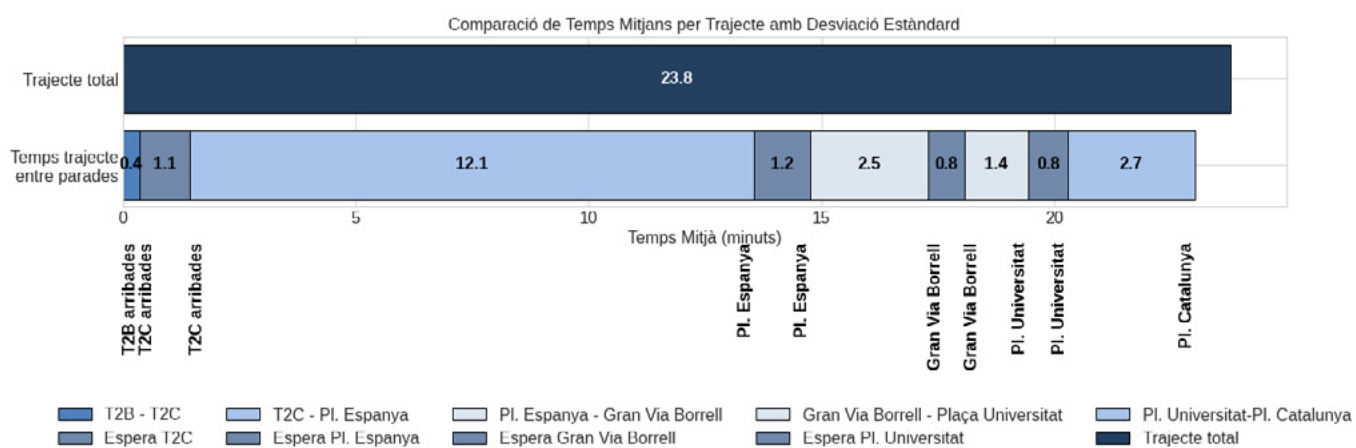
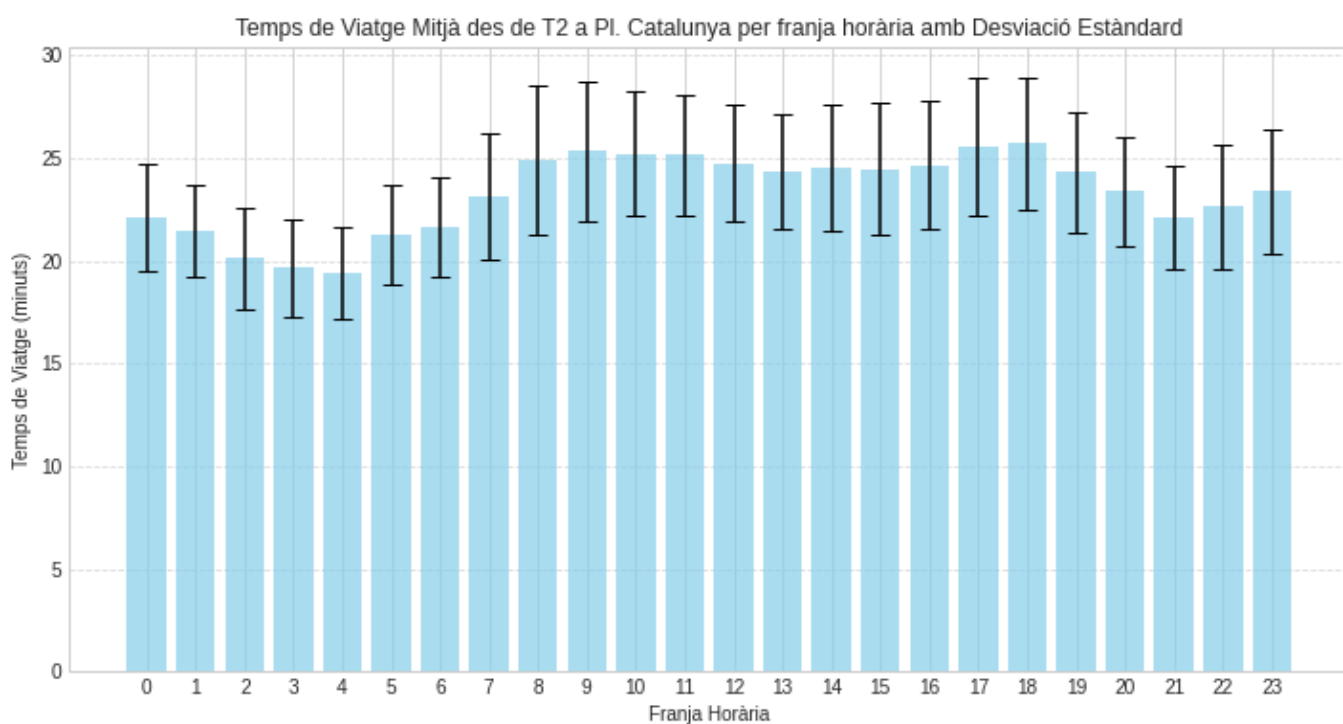
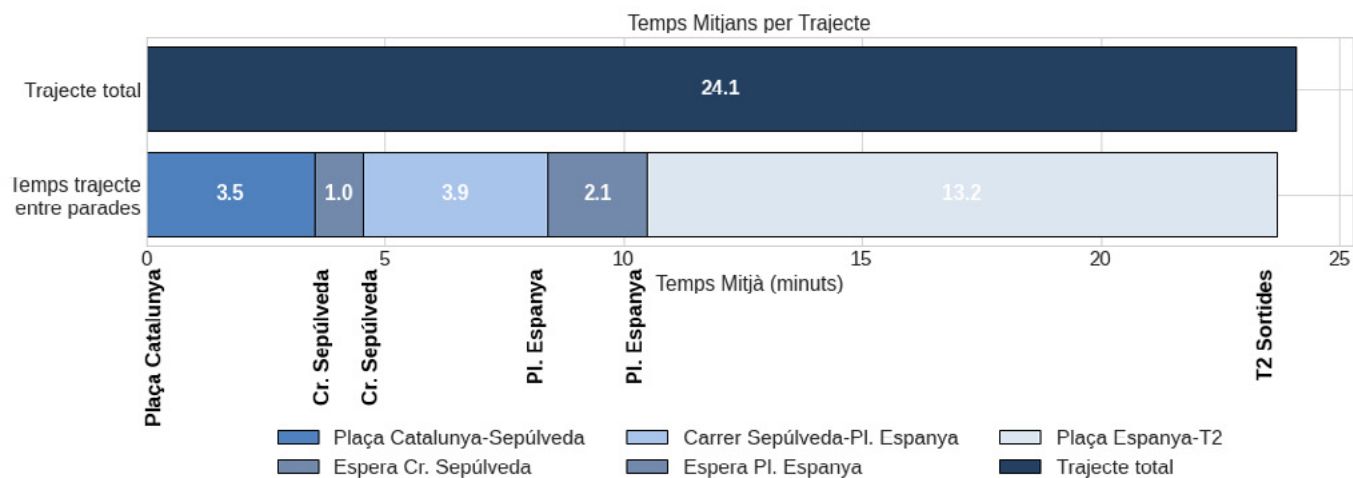


A2

	A2IDA		A2VUELTA	
Franja horària	Mitjana (min)	Std (min)	Mitjana (min)	Std (min)
0	20.817537	2.490205	22.133464	2.598358
1	20.266805	2.342633	21.405020	2.231855
2	19.935129	2.245635	20.115025	2.480103
3	21.443390	2.661207	19.642352	2.367673
4	22.803617	2.966856	19.354179	2.234201
5	23.101704	2.774867	21.281011	2.441118

6	23.235828	2.618478	21.631447	2.457831
7	24.645081	2.820139	23.113921	3.065423
8	25.178328	2.988398	24.855303	3.629228
9	24.732871	2.757948	25.345587	3.392833
10	24.908224	2.682281	25.215197	3.010660
11	25.410868	2.697590	25.129279	2.969945
12	25.669604	2.762558	24.731595	2.821890
13	25.316417	2.701947	24.316650	2.793707
14	24.609187	2.633309	24.530916	3.092290
15	24.515619	2.675538	24.449610	3.198288
16	25.301225	2.844281	24.649026	3.090907
17	26.202706	2.936439	25.542236	3.326669
18	25.777896	2.912988	25.697590	3.227728
19	24.725282	2.716973	24.286200	2.904431
20	22.737789	2.383605	23.369951	2.666597
21	21.331157	2.043491	22.110619	2.538543
22	20.939466	2.040197	22.623552	3.052118
23	21.178292	2.295314	23.385408	3.025111





Anàlisi de cues



5.1. Descripció de les dades

5.1.1. Data sets i origen

En aquest apartat s'analitzen els fitxers que contenen dades sobre els mostrejos de cues i els passatgers que accedeixen a l'Aerobús en els diferents punts clau de les seves rutes:

- **A1IDA:** Plaça Catalunya (direcció Aeroport).
- **A1VUELTA:** Aeroport (direcció Plaça Catalunya).
- **A2IDA:** Plaça Catalunya (direcció Aeroport).
- **A2VUELTA:** Aeroport (direcció Plaça Catalunya).

Els fitxers comprenen dades recopilades des de l'any 2021 fins al 2023. No obstant això, l'anàlisi se centra exclusivament en el període que s'inicia al setembre de 2022. Aquesta decisió es basa en l'observació que les dades anteriors a aquesta data presenten un impacte significatiu dels efectes de la pandèmia de la COVID-19, com es reflecteix en el nombre de passatgers i patrons d'ocupació. Per tant, per assegurar que l'anàlisi reflecteixi un escenari més estable i representatiu de la normalitat en l'ús del servei, s'ha descartat l'estudi del període previ a setembre de 2022.

Per fer l'estudi, els data sets s'agrupen en un únic conjunt anomenat Cues.

Data set 1: Cues

Taula de variables:

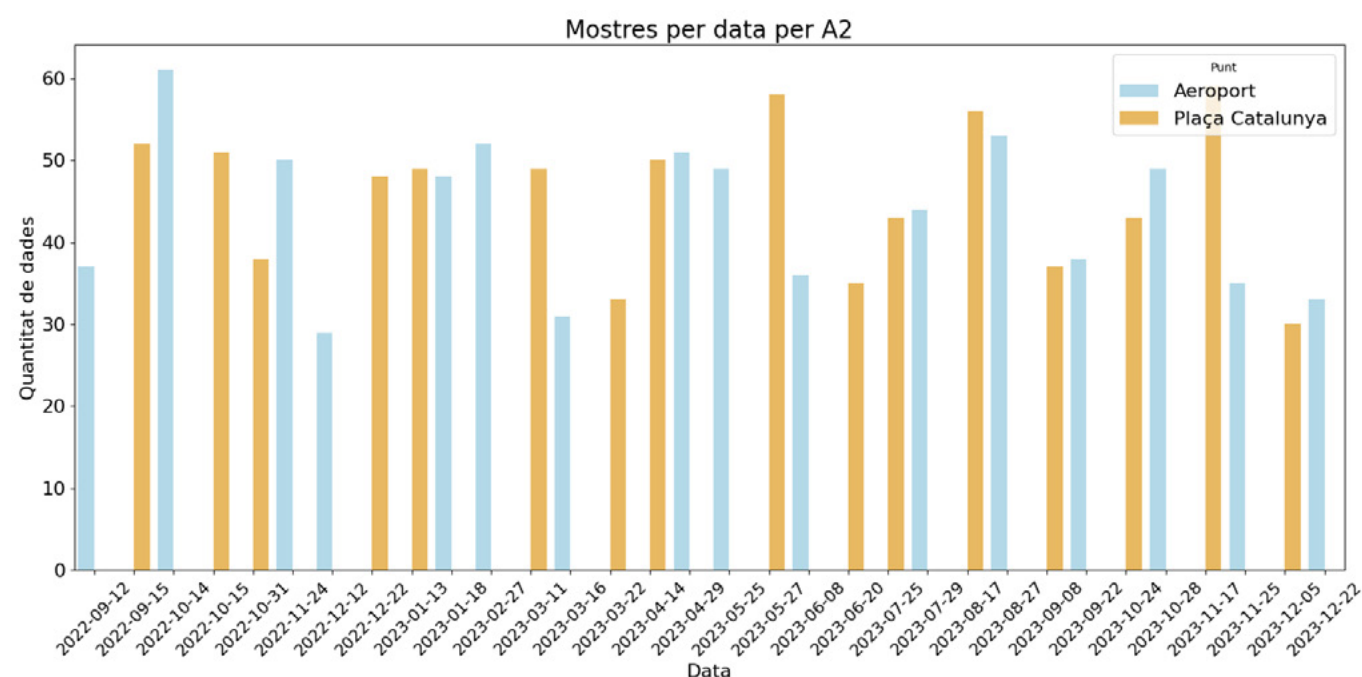
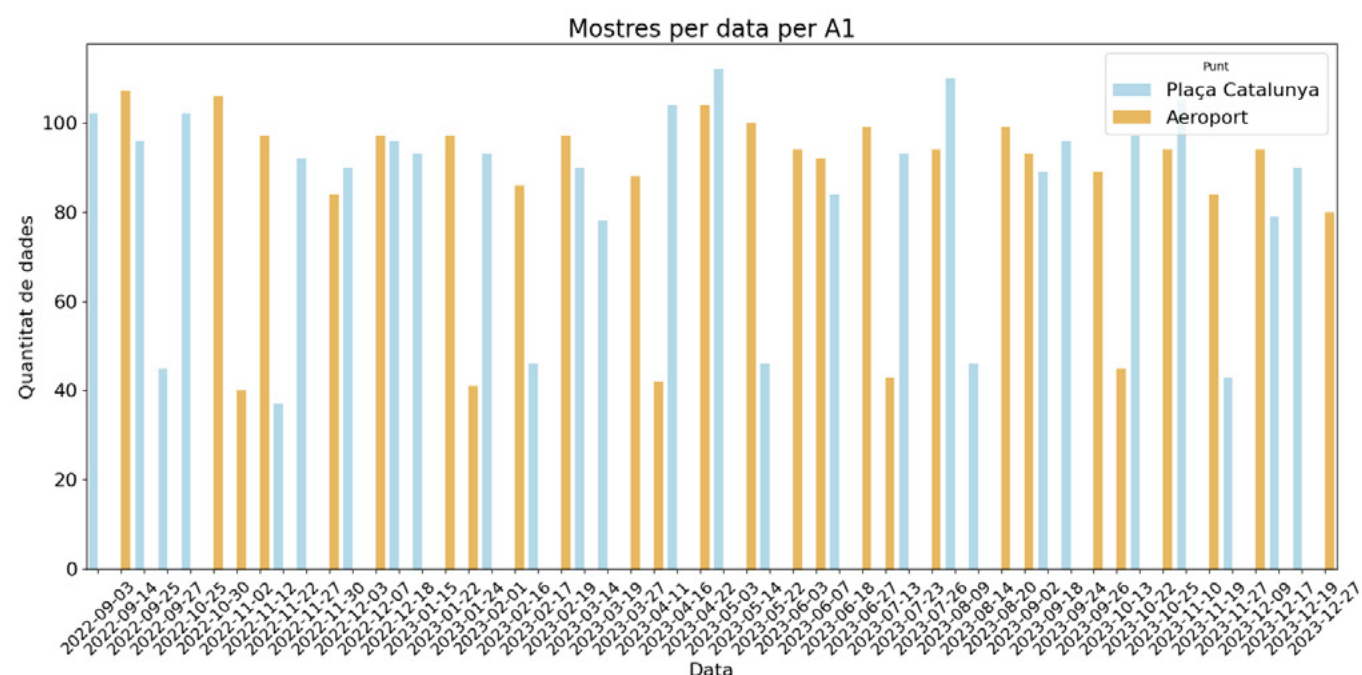
Variable	Descripció	Tipus dada	Valors possibles/ Rangs	Comentaris
trimestre	Trimestre de l'any	Categòrica	[2022-01-01 al 2024-07-31]	Format YYYY-MM-DD
Punt	Parada	Categòrica	[Aeroport i Pl.Catalunya]	
Línia	Línia	Categòrica	[A1, A2]	
Dia	Dia de mostreig	Data	[2022-09-27,..., 2023-09-22]	Format YYYY-MM-DD
Cua d'accés al vehicle (quan obre portes)	Quantitat de gent esperant per pujar a l'autobús	Numèrica	0, 1, 2,...	
Cua d'accés a màquines d'autovenda (quan obre portes)	Quantitat de gent esperant per comprar bitllet al obrir portes	Numèrica	0, 1, 2,...	
Nº total de persones fent cua	Suma de cua accés al vehicle més gent comprant bitllet	Numèrica	0, 1, 2,...	
Hora-rec	Franja horària	Categòrica	de 23h a 24h...	Format "de 23h a 24h"
Hora sortida	Hora de sortida real de l'autobús	Hora	00:06:00, 23:59:00...	Format HH:MM:SS
Freqüència sortida	Freqüència de sortida teòrica	Hora	00:14:00, 00:24:00...	Format HH:MM:SS
Ocup. Vehicle sortida (usuaris que pugen al bus)	Quantitat de gent que puja a l'autobús	Numèrica	0, 1, 2,...	
Cua d'usuaris esperant quan marxa l'autobús	Gent que queda a la parada al marxar l'autobús	Numèrica	0, 1, 2,...	

5.1.2. Anàlisi inicial

A continuació, realitzarem una visualització general de la distribució dels valors al data set, mitjançant gràfics, amb l'objectiu de detectar possibles errors o anomalies en les dades.

El següents gràfics mostren la distribució de les dades. A l'eix X es mostren els dies de mostrejos separats per parada i per línia.

En tots dos gràfics s'observa, de manera general, una distribució equilibrada de les mostres (Aeroport i Plaça Catalunya). La quantitat de mostres varia entre 60 i 30 per dia.



La següent taula indica la quantitat de mostres per punt:

Línia	Punt	Quantitat
A1	Aeroport	2286
A1	Plaça Catalunya	2254
A2	Aeroport	696
A2	Plaça Catalunya	696
Total mostres		5967

No es visualitza inconsistències en les dades.

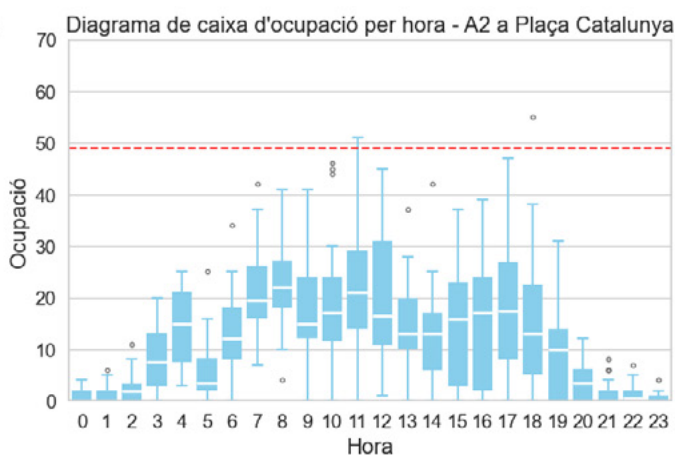
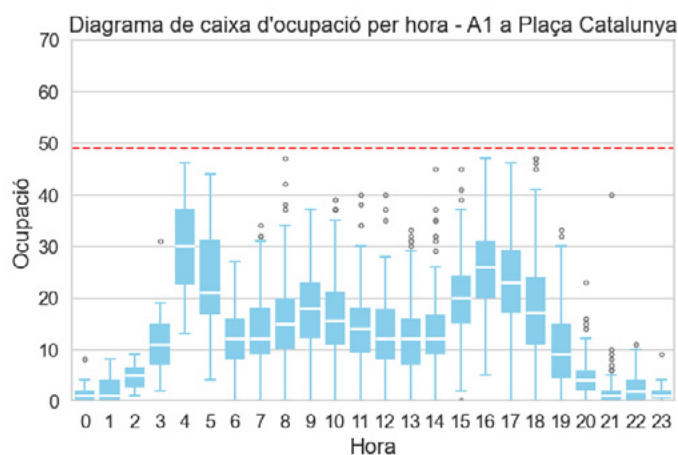
5.2. Distribució de l'ocupació als punts de control per franja horària

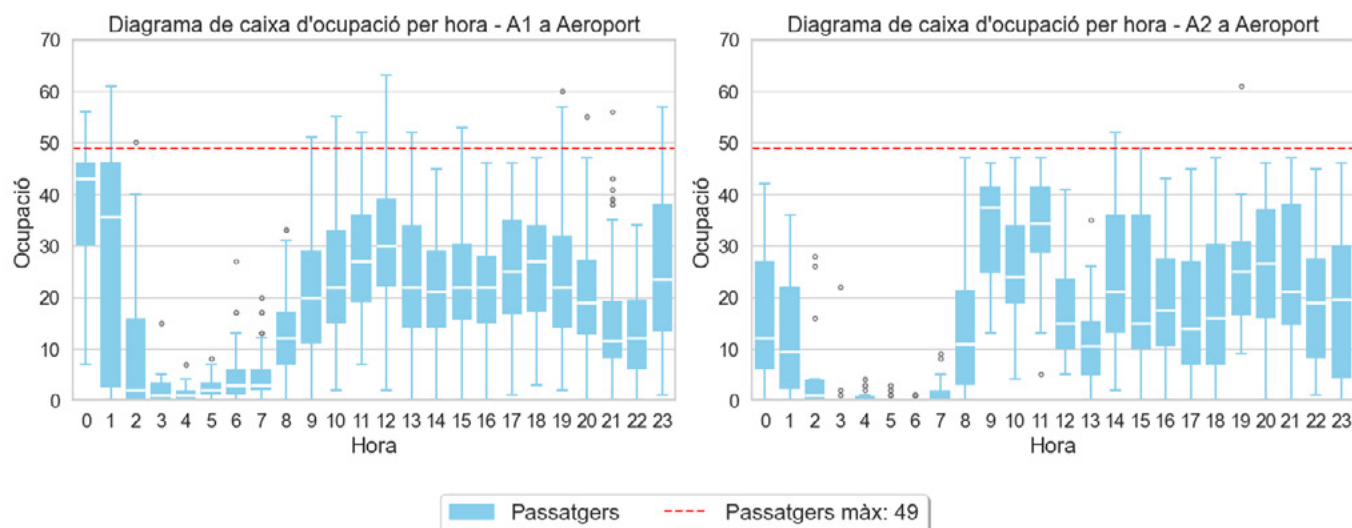
En aquest apartat es presenta l'anàlisi de distribució de l'ocupació per hora i per punt (Plaça Catalunya i Aeroport) per a cadascuna de les línies d'autobús incloses en l'estudi. L'objectiu principal és oferir una visió clara de la variabilitat de l'ocupació al llarg del dia i entre diferents línies, identificant les franges horàries de màxima i mínima demanda.

Cal destacar que, en tots els casos, s'analiza l'ocupació en la primera parada del trajecte, motiu pel qual la dada registrada reflecteix l'ocupació inicial de l'autobús en cada línia.

Els gràfics tipus diagrama de caixes presentats en aquest apartat mostren de forma visual i clara l'ocupació dels autobusos per franja horària en els punts de control de Plaça Catalunya i Aeroport per a cadascuna de les línies analitzades.

- **Eix X (horitzontal):** Representa les franges horàries en què s'ha realitzat l'observació de l'ocupació (passatgers), permetent identificar com varia la demanda al llarg del dia.
- **Caixa i bigotis:** Representa el rang interquartílic (IQR), que inclou el 50% central de les dades (rang os es concentren la major part de les observacions). La línia blanca representa la mediana i els bigotis s'estenen des de el valor més gran al més petit que no es considera atípic. Els punts representen les observacions atípiques (fora de l'habitual).
- **Línia vermella:** Representa la capacitat màxima de l'autobús. Aquesta línia serveix com a referència per visualitzar el grau d'aprofitament de la capacitat del vehicle en cada franja horària.





A partir de l'anàlisi dels diagrames de caixes d'ocupació per franja horària a Plaça Catalunya i Aeroport per a les línies A1 i A2, es poden ressaltar les següents observacions clau:

Línia A1 - Plaça Catalunya:

- **Punt destacat a la matinada:** S'observa un pic significatiu d'ocupació a les 4 h i les 5 h de la matinada, coincidint amb les dades automàtiques prèvies. En aquesta franja, la mediana d'ocupació és la més alta del dia, possiblement degut a la baixa freqüència d'autobusos en aquesta franja horària.
- **Reducció de l'ocupació de 6 a 14 h:** Entre les 6 h i les 14 h, la mediana d'ocupació es redueix de forma progressiva, coincidint amb un augment de la freqüència dels autobusos, fet que distribueix millor els passatgers.
- **Increment de l'ocupació de 15 a 18 h:** A partir de les 15 h i fins a les 18 h, l'ocupació torna a augmentar, arribant a valors de fins a 30 passatgers segons la mediana.
- **Descens progressiu a partir de les 18 h:** Després de les 18 h, l'ocupació comença a disminuir de forma constant fins arribar a valors inferiors a 10 passatgers a partir de les 22 h, mantenint aquesta baixa ocupació fins a les 2 h de la matinada.

Línia A1 - Aeroport:

- **Sobreocupació en diverses franges horàries:** L'Aeroport és el punt on més franges horàries registren ocupacions superiors a la capacitat màxima de l'autobús (límit marcat en 49 passatgers). En alguns casos, s'observen màxims de més de 60 passatgers.
- **Punt màxim d'ocupació a mitjanit:** A les 0 h i la 1 h de la matinada, la mediana d'ocupació es situa entre 35 i 45 passatgers, mentre que els valors màxims superen els 60 passatgers.
- **Baixa ocupació de 2 a 7 h:** Entre les 2 h i les 7 h de la matinada, l'ocupació es redueix notablement, amb una mediana d'ocupació inferior a 10 passatgers. Aquest comportament és consistent amb la reducció general de l'activitat aeroportuària durant aquestes hores.
- **Augment progressiu de 8 a 12 h:** A partir de les 8 h, l'ocupació comença a pujar progressivament fins a 30 passatgers, per disminuir i estabilitzar-se sobre les 12 h, en valors propers als 20 passatgers.
- **Estabilitat de 12 a 20 h:** De 12 a 20 h, l'ocupació es manté estable entre 20 i 30 passatgers per hora.
- **Descens entre les 21 h i les 22 h:** A les 21 h i les 22 h, l'ocupació baixa significativament a una mediana d'uns 10 passatgers per hora.

- **Nova pujada de 23 a 1 h:** Entre les 23 h i la 1 h, es detecta una pujada de l'ocupació, amb medianes que tornen a situar-se entre 30 i 45 passatgers, evidenciant un patró d'augment de la demanda nocturna o disminució de freqüència d'autobusos.

Línia A2 - Plaça Catalunya:

- **Patró similar a l'A1:** L'ocupació de la línia A2 a Plaça Catalunya segueix un patró similar al de la línia A1, tot i que amb petites diferències en alguns moments clau.
- **Increment d'ocupació de les 3 a les 8 h:** L'ocupació augmenta progressivament des de les 3 h de la matinada fins les 8 h. Hi ha un descens a la franja de les 5 h que pot coincidir amb l'augment de la freqüència dels autobusos.
- **Estabilitat de les 8 a les 18 h:** De les 8 a les 18 h l'ocupació mediana es manté entre 20 i 10 passatgers. La variabilitat de l'ocupació augmenta a partir de les 11 del matí (augment de la mida de la caixa).
- **Descens a partir de les 17 h:** A partir de les 17 h, l'ocupació comença a disminuir progressivament fins arribar a valors inferiors a 10 passatgers per hora a partir de les 20 h. Aquesta tendència es manté fins a les 2 h de la matinada.

Línia A2 – Aeroport:

- **Superació de la capacitat a les 14 h:** L'ocupació només supera la capacitat màxima de l'autobús en la franja de les 14 h.
- **Variabilitat entre les 7 h i les 13 h:** De les 7 a les 13 h, es detecta una variabilitat significativa en la mediana d'ocupació, la qual cosa pot indicar la presència de pics d'ocupació en certes hores d'aquest interval.
- **Estabilització de les 14 h fins a la 1 h:** Des de les 14 h fins a la 1 h de la matinada, l'ocupació es manté estable entre 20 i 10 passatgers per hora. Aquest comportament és molt més regular que en altres períodes del dia.
- **Ocupació pràcticament nul·la entre les 2 h i les 7 h:** Entre les 2 h i les 7 h de la matinada, l'ocupació de la línia A2 Aeroport es manté pràcticament a 0 passatgers.
- **Inici d'increment a partir de les 8 h:** A partir de les 8 h, es registra un augment dràstic de l'ocupació.

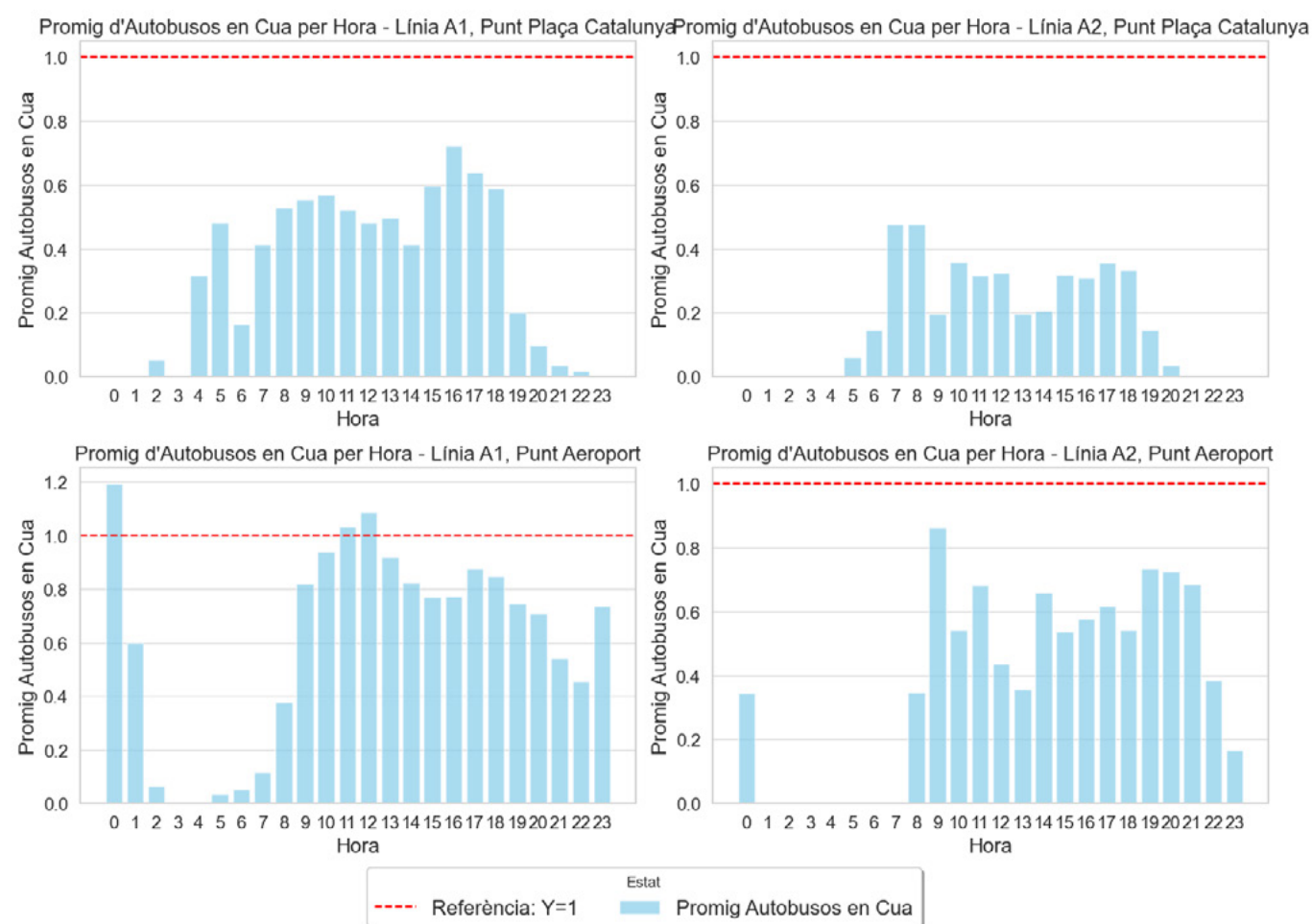
Els resultats de l'anàlisi mostren que les ocupacions obtingudes a partir del mostreig coincideixen amb les ocupacions recopilades automàticament pel sistema. Aquesta coincidència corrobora la qualitat i la fiabilitat de les dades generades pel sistema automàtic, garantint així la precisió dels indicadors d'ocupació presentats i la validesa de les conclusions extretes en els apartats anteriors.

5.3. Distribució de cues en punts de mostreig

5.3.1. Distribució de cues al servei

En aquest apartat es presenten una sèrie de gràfics que mostren la distribució de la cua existent en el moment en què l'autobús parteix, desglossada per franja horària. Cada franja horària permet observar la variabilitat en la longitud de la cua, destacant una línia de referència que representa la capacitat màxima de l'autobús. Quan la cua supera aquesta línia, significa que l'última persona present a la cua haurà d'esperar més d'un vehicle per poder accedir al servei.

A més, es presenten gràfics que mostren el percentatge de situacions de cua. Aquestes situacions es classifiquen en tres categories: (1) quan l'autobús parteix sense cua de passatgers, (2) quan hi ha cua i l'autobús surt ple (amb la capacitat màxima assolida), i (3) quan hi ha cua però l'autobús parteix sense arribar a la seva màxima capacitat.



Línia A1 - Plaça Catalunya:

Les cues més elevades es registren entre les 15 h i les 18 h, moment en què la cua es situa entorn al 60% de la capacitat d'un vehicle o el supera. No es detecta en cap moment una cua superior a 1 vehicle, la qual cosa indica que no hi ha passatgers acumulats que no puguin accedir al següent servei.

El patró de les cues segueix un comportament similar al de l'ocupació del vehicle. La cua puja bruscament a partir de les 4 h de la matinada i disminueix a les 6 h, coincidint amb l'augment de la freqüència dels autobusos. Des de les 6 h fins a les 15 h, la cua es manté estable en valors d'entre 0,4 i 0,6 autobusos en cua, cosa que indica una afluència constant de passatgers durant aquesta franja horària. A partir de les 18 h, la cua disminueix progressivament fins a valors inferiors a 0,2 vehicles i, finalment, es redueix a 0 a partir de les 23 h, quan la demanda ja és mínima.

Línia A1 - Aeroport:

El pic de cua es troba a les 0 h de la matinada, coincidint amb la franja horària amb l'ocupació màxima del vehicle. Aquest punt de control presenta les cues més grans de tots els punts de mostreig, destacant-se com un dels moments més crítics del servei.

Entre les 9 h i les 18 h, la cua se situa de forma consistent vora els 0,8 vehicles i per sobre en algunes franges. Les franges de 10 a 13 h mostren les cues més altes, amb valors iguals al 90% de la capacitat d'un vehicle. Durant les 11 h i les 12 h, la cua fins i tot supera 1 vehicle, indicant que hi ha passatgers que hauran d'esperar dos serveis per poder accedir al vehicle.

Després de les 18 h, la cua disminueix progressivament fins a les 22 h, moment en què la cua és pròxima a 0,4 vehicles. Tot i això, hi ha un nou augment a les 23 h i a les 0 h, coincidint amb l'arribada de nous passatgers d'últims vols nocturns. A partir de la 1 h, la cua torna a disminuir fins a arribar a 0 a les 3 h i les 4 h de la matinada.

Línia A2 - Plaça Catalunya:

El pic de cua es registra a les 7 h i 8 h del matí, situant-se en 0,5 autobusos, la més alta per a aquesta línia i punt de control.

Després d'aquest pic, la cua es manté estable entre 0,2 i 0,4 autobusos durant tot el període comprès entre les 9 h i les 18 h, la qual cosa mostra una situació controlada en termes d'afluència de passatgers. A partir de les 19 h, la cua comença a reduir-se fins a valors inferiors a 0,2 autobusos i finalment es redueix completament a 0 a les 21 h, coincidint amb la reducció de la demanda de la tarda-nit.

Línia A2 - Aeroport:

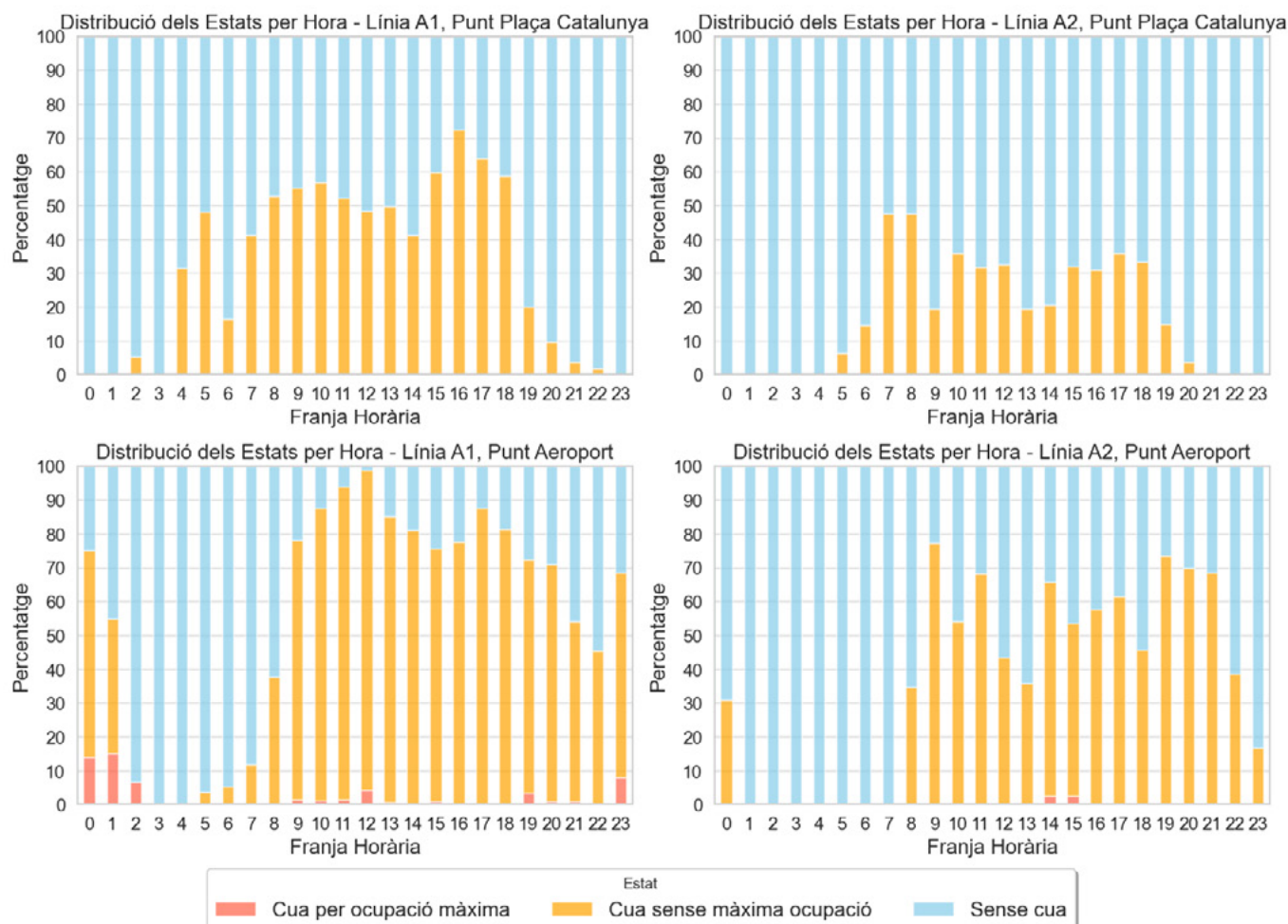
En el punt de control de l'Aeroport, la cua es manté a 0 entre la 1 h i les 7 h de la matinada.

A partir de les 8 h, la cua comença a augmentar, situant-se en valors d'entre 0,3 i 0,4 autobusos en cua. La situació es dispara a les 9 h, on la cua arriba a superar el 0,8 autobusos.

De les 10 h a les 22 h, la cua es manté en valors entre 0,4 i 0,7 autobusos en cua, amb oscil·lacions abruptes entre aquests valors.

A les 23 h la cua es situa en el valor més baix, per sota de 0,2 vehicles, i finalment a la 1 h es situa a 0.

A continuació, es presenten gràfics que mostren el percentatge de situacions de cua:



Línia A1 - Plaça Catalunya:

En aquest punt de control, les cues al partir l'autobús es generen principalment perquè el vehicle no parteix amb la seva ocupació màxima. Això es pot observar en els diagrames de caixes, on en cap franja horària els valors d'ocupació superen els 49 passatgers, ni tan sols els valors atípics. Aquest comportament pot deure's a dues raons principals. En primer lloc, és possible que el conductor segueixi una freqüència estipulada i, per tant, parteixi a una hora determinada encara que l'autobús no estigui completament ple.

En segon lloc, en ser la primera parada de la línia, el conductor podria deixar espai disponible per als passatgers de les següents parades. La major part de les cues observades en aquest punt es corresponen amb la situació en què l'autobús parteix sense ocupar completament la seva capacitat.

Línia A1 - Aeroport:

A l'Aeroport, es detecten diverses franges horàries on l'autobús parteix amb capacitat màxima (49 passatgers) i encara queden passatgers a la parada. Aquestes franges crítiques es concentren en els següents intervals horaris: de 23 a 2 h, de 9 a 12 h, a les 15 h i de 19 a 21 h. En aquestes franges, els bigotis dels diagrames de caixes d'ocupació superen els 49 passatgers, evidenciant que la capacitat màxima de l'autobús ha estat superada. No obstant això, el percentatge de mostres on es produeix aquesta situació és inferior al 15% en aquestes franges, la qual cosa indica que no és una situació constant però sí recurrent.

D'altra banda, es detecta que de 9 a 20 h, el 70% dels mostres realitzats presenten cues, tot i que l'autobús no parteix ple.

Línia A2 - Plaça Catalunya:

A Plaça Catalunya, per a la Línia A2, s'observa que, en la majoria de franges horàries, l'autobús parteix sense ocupar completament la seva capacitat, fins i tot quan queden passatgers en cua. Entre les 5 h i les 20 h, el percentatge de situacions en què queda cua però l'autobús no parteix ple se situa entre el 5% i el 35%, havent-hi dos pics a les 7 h i les 8 h, que arriben fins al 50% de les situacions mostrejades.

Aquest comportament operatiu és molt similar al detectat per a la Línia A1 a Plaça Catalunya, la qual cosa pot deure's al fet que l'autobús parteix segons una freqüència estipulada i no en funció de l'ocupació. Això significa que l'autobús parteix encara que no estigui ple per respectar l'horari de servei. També pot ser que, en ser la primera parada de la línia, es deixi espai per als passatgers de les següents parades, garantint així la capacitat per a altres punts de la ruta.

Línia A2 - Aeroport:

A l'Aeroport, per a la Línia A2, es detecten algunes franges horàries on l'autobús parteix amb la capacitat màxima de 49 passatgers i encara queden passatgers en cua a la parada.

Les franges crítiques d'aquesta situació es concentren a les 14 h i les 15 h, tot i que el percentatge de mostrejos en què es produeix aquesta situació és d'aproximadament el 3%, la qual cosa indica que no és una situació molt freqüent.

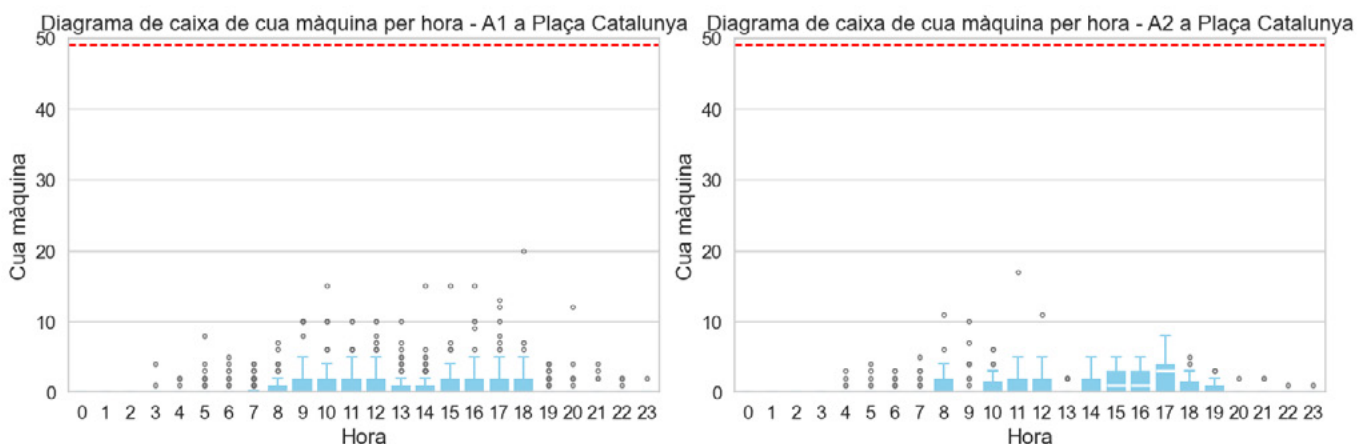
A diferència d'això, les franges on es detecten cues sense que l'autobús parteixi ple són més extenses. Aquestes situacions es presenten entre les 8 h i les 0 h, amb percentatges que varien entre el 20% i el 70% de les mostres realitzades. Un comportament interessant és que, tot i que el diagrama de caixes mostra que les medianes d'ocupació més altes es troben entre les 9 h i les 11 h, no es detecten cues associades a la superació de la capacitat màxima de l'autobús en aquestes franges.

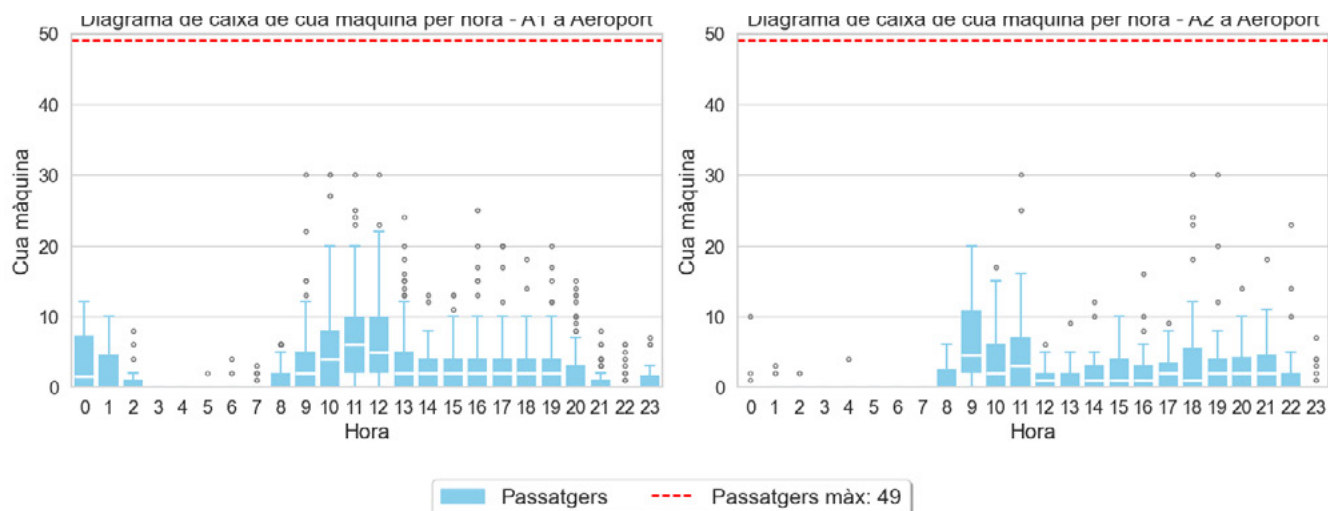
5.3.2. Distribució de cues a les màquines de venda de bitllets

En aquest apartat s'analitzen les cues generades a les màquines de venda de bitllets durant les diferents franges horàries del dia.

Per a l'anàlisi, es presenten dos tipus de gràfics que aporten una visió complementària de la situació. En primer lloc, es mostren els diagrames de caixes per franja horària, que permeten visualitzar la distribució de la quantitat de persones en cua a les màquines en cada franja.

En segon lloc, es presenten els gràfics de barres apilades, que mostren el percentatge de mostrejos amb o sense cua per a cada franja horària.





Línia A1 - Plaça Catalunya:

Els valors de cues en aquest punt segueixen el patró general del servei, amb cues inferiors a 10 passatgers en les franges compreses entre les 7 h i les 18 h, reflectint una demanda controlada i estable durant aquest període. A partir de les 19 h i fins a les 7 h, les cues es consideren valors atípics.

Línia A1 - Aeroport:

Els valors de cues en aquest punt són més elevats que a Plaça Catalunya. Les franges amb la mediana de cues més alta es troben entre les 11 h i les 12 h, amb medianes sobre 5 passatgers, mentre que els bigotis poden arribar fins als 20 passatgers, indicant moments puntuals amb més demanda.

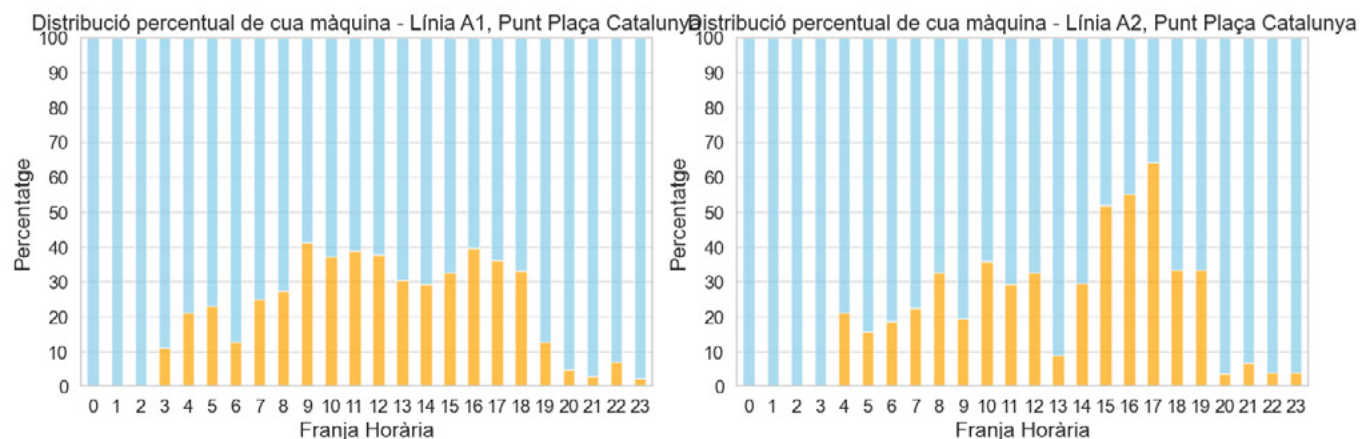
En la resta de franges fins a les 19 h, els valors es mantenen generalment per sota de 5 passatgers, amb variabilitat controlada. Entre les 3 h i les 7 h, no es detecten cues significatives, excepte alguns valors atípics.

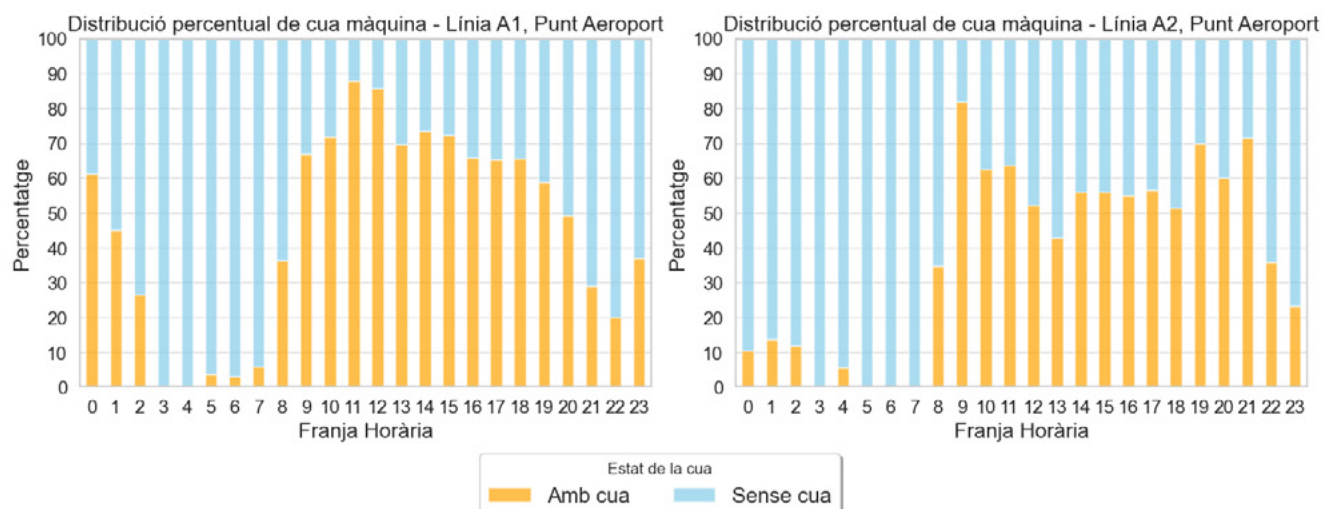
Línia A2 - Plaça Catalunya:

En aquest punt, els valors de les cues segueixen el patró general del servei, amb cues inferiors a 10 passatgers entre les 8 h i les 19 h. Entre les 20 h i les 7 h, les cues es consideren valors atípics.

Línia A2 - Aeroport:

El diagrama de caixa mostra una absència pràcticament total de cues entre les 23 h i les 7 h, amb només alguns valors atípics puntuals. A partir de les 8 h, es detecta un increment progressiu de les cues, amb medianes per sota dels 5 passatgers, però bigotis que arriben fins als 10 passatgers.





Del gràfic de barres apilats es poden extreure les següents conclusions:

Línia A1 - Plaça Catalunya:

Les cues a les màquines de venda de bitllets es concentren entre les 3 h i les 23 h, amb un comportament progressiu. A l'inici del període (3 h a 6 h), el percentatge de mostrejos amb cua fa una corba suau, començant al 10% i arribant fins al 20%. A partir de les 7 h, el percentatge augmenta significativament, establint-se entre el 30% i el 40% dels mostrejos realitzats en la franja de 7 a 18 h.

A partir de les 18 h, el percentatge de cues comença a disminuir de forma progressiva, situant-se per sota del 10% a partir de les 21 h, fins a arribar al 0% a les 0 h. Aquest patró reflecteix una distribució de cues consistent amb les hores de màxima activitat a Plaça Catalunya.

Línia A1 - Aeroport:

Aquest punt presenta els percentatges més alts de cues en els mostrejos realitzats a les màquines de venda de bitllets. A les 0 h, el 60% dels mostrejos presenten cues, i aquest percentatge descendeix progressivament fins a situar-se al 0% entre les 3 h i les 4 h.

Després, entre les 5 h i les 7 h, el percentatge es manté per sota del 10%, però augmenta de forma abrupta fins a arribar al 90% a les 10 h i les 11 h, moment de màxima saturació.

A partir d'aquest pic, el percentatge torna a disminuir de manera progressiva, fins el 20% i el 30% en les franges de 22 h i 23 h. Aquest patró evidencia una forta concentració de cues durant les hores d'alta activitat aeroportuària.

Línia A2 - Plaça Catalunya:

En aquesta línia, les cues comencen a les 4 h de la matinada, creixent gairebé de manera lineal fins a les 17 h, passant del 20% al 60% dels mostrejos. Cal destacar una baixada significativa a les 13 h, on el percentatge de mostrejos amb cues descendeix fins al 10%, abans de tornar a pujar.

A partir de les 17 h, el percentatge de cues comença a baixar de manera progressiva, situant-se per sota del 10% a les 23 h. En les hores nocturnes, de 0 a 3 h, el percentatge de mostrejos amb cua és inexistent (0%), reflectint una baixa activitat durant aquestes hores.

Línia A2 - Aeroport:

El percentatge de mostrejos amb cua a aquest punt es manté en un 10% aproximadament entre les 0 h i les 4 h, reflectint una activitat reduïda durant la matinada. Entre les 5 h i les 7 h, el percentatge descendeix completament al 0%, abans de pujar de manera abrupta fins al 80% a les 9 h.

Després d'aquest pic, el percentatge disminueix lleugerament, estabilitzant-se al voltant del 60% entre les 10 h i les 21 h, un comportament molt més estable en comparació amb altres punts. Finalment, el percentatge de cues comença a baixar a partir de les 21 h, situant-se al 10% a les 0 h.



Model de previsió de la demanda



6.1. Selecció de variables

Per a la selecció de variables predictives, s'han realitzat contrastos d'hipòtesis per avaluar la rellevància de diverses variables inicialment plantejades com a possibles predictors de la demanda. Les variables considerades són:

- **Esdeveniments:** S'han creat quatre variables a partir dels esdeveniments per estudiar la relació d'aquests amb la demanda del servei de l'Aerobús:
 - **Quantitat d'esdeveniments (esdeveniment):** Suma total d'esdeveniments presents el dia objectiu de la predicció.
 - **Esdeveniments que finalitzen per dia (fi_esdeveniment):** Suma dels esdeveniments que finalitzen el dia objectiu de la predicció.
 - **Esdeveniments que s'inicien per dia (inici_esdeveniment):** Suma dels esdeveniments que s'inicien el dia objectiu de la predicció.
 - **Aforament:** suma de l'aforament de tots els esdeveniments presents durant el dia objectiu de la predicció.
- **Pluja:** Les condicions meteorològiques adverses, com la pluja, poden influir en la decisió de la ciutadania a l'hora d'escollir un mitjà de transport. Per incorporar la informació sobre la pluja com a variable predictora en el model, s'han creat dues versions d'aquesta variable per capturar la seva influència de forma més efectiva:
 - **Variable binària (Pluja: Sí/No):** S'ha creat una variable binària que pren el valor "Sí" si s'ha registrat alguna precipitació i "No" si no s'ha registrat cap pluja.
 - **Variable categòrica d'intensitat de la pluja:** Per tal de capturar la intensitat de la pluja i no només la seva presència, s'ha discretitzat la variable de precipitació acumulada en diferents categories. Les franges s'han definit de la següent manera:
 - **Nul·la:** Precipitació acumulada de 0 mm.
 - **Dèbil:** Precipitació acumulada entre 0 i 2 mm.
 - **Moderada:** Precipitació acumulada entre 2 i 15 mm.
 - **Forta:** Precipitació acumulada entre 15 i 30 mm.
 - **Molt forta:** Precipitació acumulada entre 30 i 60 mm.
 - **Torrencial:** Precipitació acumulada superior a 60 mm.
- **Quantitat de vols:** Quantitat de vols d'arribada i sortida per terminal i per franja horària.
- **Festius:** S'ha creat una variable binària "Sí" i "No" que indica, segons el calendari de la ciutat de Barcelona, si és dia festiu.

Els resultats d'aquests contrastos han permès identificar quines d'aquestes variables tenen una influència significativa sobre la demanda. No obstant això, aquesta anàlisi no s'ha pogut aplicar en el cas del metro, ja que aquest mitjà de transport segueix sempre el mateix horari i freqüència de pas, independentment de factors externs com la pluja o els esdeveniments. Per aquest motiu, la seva demanda no es veu tan condicionada per les variables analitzades.

Cal destacar que la selecció de variables es realitzarà únicament en el model de regressió lineal, ja que aquest tipus de model és sensible a la multicolinealitat i la presència de variables irrelevantes pot afectar la precisió i interpretabilitat dels resultats. En canvi, en models més avançats com els arbres de decisió o les xarxes neuronals, no és necessari aplicar una selecció prèvia de variables, ja que aquests models són capaços de ponderar automàticament la importància de cada variable durant el procés d'entrenament

6.1.1. Esdeveniments

Per avaluar la relació entre les variables, s'ha utilitzat el coeficient de correlació de *Spearman* i el p-valor associat al no presentar normalitat les variables a analitzar.

S'ha realitzat l'anàlisi per als 4 *data sets* diferents i per a cada un s'ha calculat la correlació i el p-valor de cada variable. Els valors de p-value inferiors a 0.05 indiquen que la correlació entre les variables és estadísticament significativa.

Taula de variables:

Variable	A1IDA (p, p-value)	A1VUELTA (p, p-value)	A2IDA (p, p-value)	A2VUELTA (p, p-value)
inici_esdeveniment	-0.0038, 0.9235	-0.0074, 0.8512	0.0105, 0.7897	0.0391, 0.3201
fi_esdeveniment	-0.0067, 0.8641	-0.0028, 0.9429	-0.0759, 0.0533	-0.0511, 0.1938
esdeveniment	0.1988, 3.35e-07	0.2933, 2.52e-14	0.1991, 3.23e-07	0.1387, 0.0004
aforament	0.0943, 0.0164	0.2270, 5.08e-09	0.1531, 9.13e-05	0.1115, 0.0045

Identificació de correlacions significatives

A partir dels resultats, es poden destacar les correlacions significatives (p-value < 0.05) per a cada *data set* i cada variable:

A1IDA:

- esdeveniment ($p = 0.1988$, $p = 3.35e-07$)
- aforament ($p = 0.0943$, $p = 0.0164$)

A1VUELTA:

- esdeveniment ($p = 0.2933$, $p = 2.52e-14$)
- aforament ($p = 0.2270$, $p = 5.08e-09$)

A2IDA:

- esdeveniment ($p = 0.1991$, $p = 3.23e-07$)
- aforament ($p = 0.1531$, $p = 9.13e-05$)

A2VUELTA:

- esdeveniment ($p = 0.1387$, $p = 0.0004$)
- aforament ($p = 0.1115$, $p = 0.0045$)

A la Terminal 1, s'observa que existeix la correlació més gran entre la quantitat de passatgers que agafen l'aerobús cap a la ciutat de Barcelona i la quantitat d'esdeveniments i l'aforament total. Concretament, la correlació amb la variable esdeveniment és de $p = 0.2933$ ($p\text{-value} = 2.52e-14$) i amb la variable aforament és de $p = 0.2270$ ($p\text{-value} = 5.08e-09$).

6.1.2. Pluja

Per avaluar si hi ha diferències significatives en la quantitat d'usuaris que pugen a l'autobús entre els dies de pluja i no pluja, hem realitzat una prova t de Student per a cada franja horària.

Franja horaria	A1IDA (t, p-value)	A1VUELTA (t, p-value)	A2IDA (t, p-value)	A2VUELTA (t, p-value)
0	1.027, 0.314	0.732, 0.470	0.704, 0.488	-0.439, 0.664
1	-1.369, 0.181	1.099, 0.282	-0.125, 0.901	-1.156, 0.257
2	-0.122, 0.904	1.252, 0.223	0.718, 0.480	0.301, 0.766
3	-1.718, 0.098	1.093, 0.286	-0.793, 0.436	0.928, 0.364
4	-1.636, 0.115	0.136, 0.893	0.106, 0.916	0.743, 0.467
5	1.179, 0.247	0.571, 0.573	0.244, 0.809	0.083, 0.934
6	-0.664, 0.512	-1.431, 0.165	1.472, 0.153	-0.352, 0.728
7	0.457, 0.651	-0.844, 0.407	1.108, 0.278	-0.873, 0.391
8	-0.939, 0.358	0.435, 0.668	0.206, 0.839	0.765, 0.451
9	-0.579, 0.568	-1.127, 0.272	-0.807, 0.429	-1.155, 0.264
10	-0.180, 0.859	-0.098, 0.922	0.747, 0.463	-2.653, 0.014
11	-0.432, 0.672	-3.563, 0.002	1.338, 0.201	-1.155, 0.266
12	0.609, 0.553	0.061, 0.952	0.807, 0.435	-1.218, 0.242
13	-0.648, 0.526	-0.711, 0.487	1.359, 0.193	1.581, 0.136
14	0.564, 0.583	-0.985, 0.338	2.088, 0.058	0.736, 0.475
15	0.447, 0.660	1.688, 0.106	2.215, 0.038	-0.273, 0.788
16	-0.349, 0.730	1.158, 0.260	-0.754, 0.459	-0.594, 0.559
17	-3.247, 0.003	-1.829, 0.080	1.143, 0.264	1.258, 0.220
18	-0.067, 0.947	0.185, 0.855	-0.441, 0.664	0.407, 0.687
19	-1.468, 0.153	-1.375, 0.180	0.473, 0.640	-0.963, 0.344
20	-0.820, 0.419	-0.560, 0.580	0.343, 0.735	0.063, 0.950
21	-0.494, 0.626	-0.585, 0.563	0.789, 0.439	-0.281, 0.781
22	1.308, 0.203	1.523, 0.140	-1.551, 0.134	0.160, 0.874
23	0.599, 0.556	-0.461, 0.649	1.115, 0.276	0.050, 0.961

Variables amb Diferència Significativa (p-valor < 0.05)

Després de la prova, només es troben unes poques franges horàries on hi ha una diferència estadísticament significativa entre els dies de pluja i els dies de no pluja. Aquestes franges són:

- **A1 IDA:** Franja 17 (p = 0.003)
- **A1 VUELTA:** Franja 11 (p = 0.002)
- **A2 IDA:** Franja 15 (p = 0.038)
- **A2 VUELTA:** Franja 10 (p = 0.014)

Tot i que s'han identificat franges horàries on hi ha una diferència significativa entre els usuaris en dies de pluja i dies de no pluja, es decideix no incloure aquesta variable en el model de predicció per diverses raons. En primer lloc, l'impacte predictiu és limitat, ja que només unes poques franges horàries mostren una diferència significativa. Aquest fet implica que, en la majoria de les franges, la presència de pluja no afecta substancialment la quantitat d'usuaris que accedeixen al servei.

En segon lloc, la inclusió d'aquesta variable comportaria un augment en la complexitat del model, ja que requeriria disposar de prediccions meteorològiques en temps real per poder predir la quantitat d'usuaris amb precisió.

Finalment, la incorporació de la variable de pluja limitaria la finestra de predicció, ja que només es podria predir la quantitat d'usuaris en aquells dies per als quals es disposés d'una predicció meteorològica prèvia.

6.1.3. Vols

Per tal de determinar la relació entre la variable a predir i els vols, s'ha realitzat una correlació de *Spearman* considerant desplaçaments dels vols respecte a la variable de passatgers amb una anticipació d'entre 1 i 6 hores. L'objectiu d'aquesta anàlisi és conèixer amb quina antelació els vols presenten una correlació més forta amb la quantitat de passatgers que accedeixen al servei.

Lag	A1IDA (correlació)	A1VUELTA (correlació)	A2IDA (correlació)	A2VUELTA (correlació)
Lag 1	0.579547	0.385293	0.434301	0.465703
Lag 2	0.714628	0.248240	0.570853	0.344887
Lag 3	0.615811	0.122457	0.581361	0.227535
Lag 4	0.489244	-0.000725	0.411990	0.107077
Lag 5	0.318366	-0.116009	0.227601	-0.026315
Lag 6	0.153239	-0.238385	0.056324	-0.127183
Sin Lag	0.440121	0.600132	0.377808	0.638123

S'han identificat les següents lags (desplaçaments) que mostren la correlació més forta per a cada conjunt de dades:

- **A1 IDA:** La màxima correlació es troba en el *Lag 2* amb un valor de **0.7146**, la qual cosa indica que el volum de passatgers que accedeixen al servei d'anada a l'aeroport T1 està fortament influenciat pels vols que surten 2 hores més tard.
- **A1 VUELTA:** En aquest cas, la correlació més alta es dona a la posició de *Sin Lag* (sense desplaçament), amb un valor de **0.6001**, la qual cosa indica que el volum de passatgers està més relacionat amb els vols que arriben a la mateixa hora.
- **A2 IDA:** La màxima correlació es produeix en el *Lag 3* amb un valor de **0.5814**, cosa que suggereix que els vols que surten 3 hores més tard tenen la major influència en el volum de passatgers.
- **A2 VUELTA:** En aquest conjunt, la màxima correlació es troba en *Sin Lag* (sense desplaçament), amb un valor de **0.6381**, indicant que el volum de passatgers està més associat als vols de la mateixa hora.

6.1.4. Festius

Per tal d'avaluar si hi ha diferències significatives en el comportament dels passatgers entre dies festius i no festius, es realitza la prova de t de Student. Aquesta prova ens permet determinar si la distribució de la quantitat de passatgers en dies festius és diferent de la distribució en dies no festius.

	A1IDA (t, p-valor)	A1VUELTA (t, p-valor)	A2IDA (t, p-valor)	A2VUELTA (t, p-valor)
Resultat	8467.5, 0.607	10714.0, 0.078	7224.5, 0.076	-1.148, 0.260

Els valors de p-valor superiors a 0.05 en tots els data sets indiquen que no hi ha diferències significatives entre els passatgers que accedeixen al servei entre dies festius i dies no festius. Per tant, aquesta variable no aportaria informació rellevant per a la predicció d'usuaris.

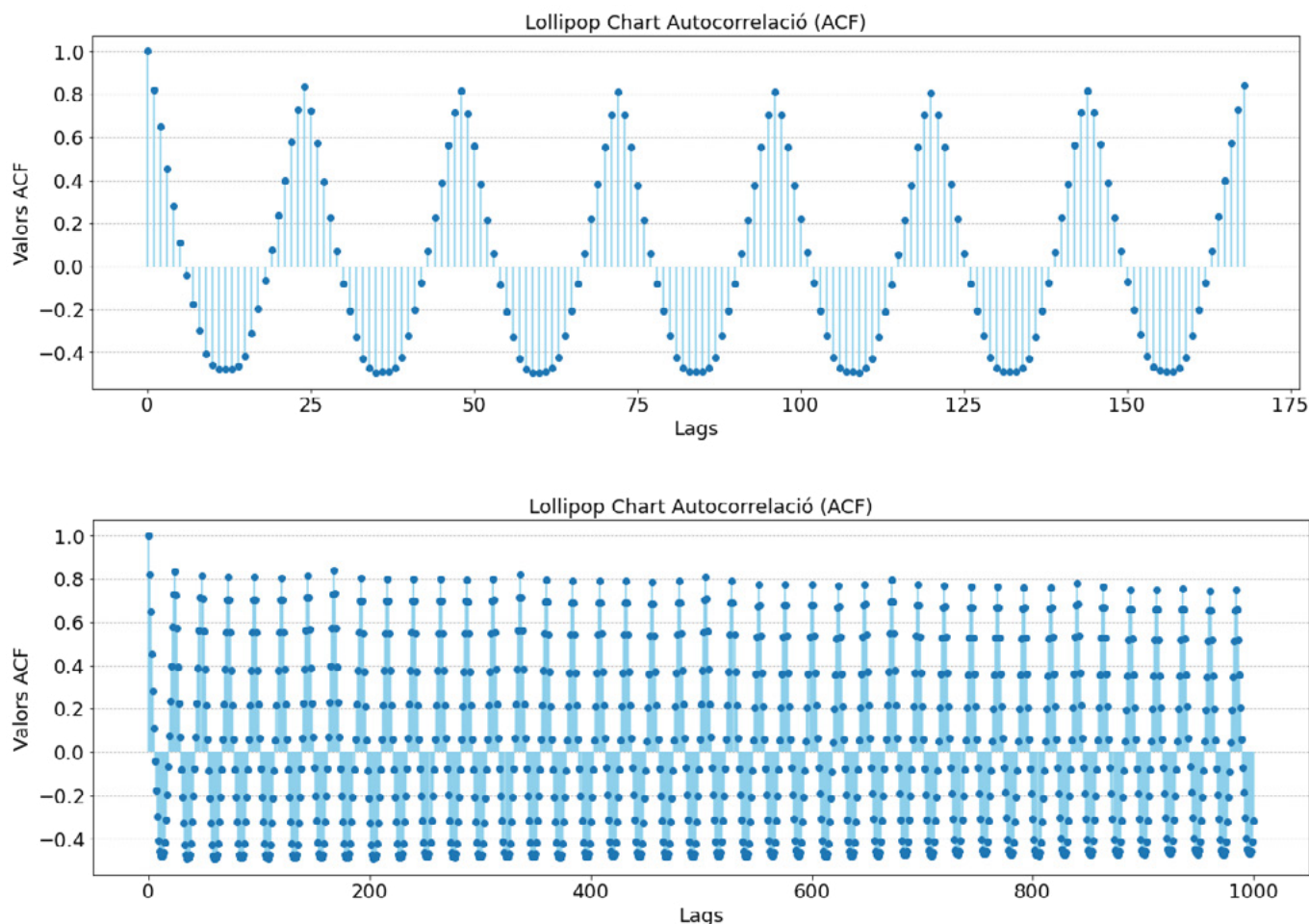
6.2. Autocorrelació

En l'àmbit de la predicció de la demanda, un aspecte fonamental és entendre com els valors de la demanda del passat afecten la demanda futura. Per aquest motiu, s'analitza la autocorrelació de les dades de demanda. L'autocorrelació ens permet avaluar la relació existent entre els valors de la demanda en diferents moments del temps, identificant patrons de repetició o tendències temporals.

Per exemplificar aquesta anàlisi, es presenten dos gràfics d'autocorrelació (ACF) corresponents a dues sèries de demanda generades amb diferents finestres temporals:

1. **Demanda per 1000 períodes (41 dies)**, on cada període representa 1 hora.
2. **Demanda per 168 períodes (1 setmana)**, on cada període també representa 1 hora.

L'objectiu d'aquesta representació és observar com l'autocorrelació es comporta en diferents escales de temps i identificar la presència de patrons estacionals diaris i setmanals.



En el primer gràfic es pot observar clarament autocorrelació diària, ja que el patró es repeteix cada 24 hores aproximadament. Això indica que la demanda en un moment donat està correlacionada amb la demanda del dia anterior a la mateixa hora.

En el segon gràfic s'observa una autocorrelació setmanal, ja que la correlació augmenta cada 168 hores (corresponent a 7 dies). Aquest comportament reflecteix la repetició de patrons setmanals.

Després de calcular l'autocorrelació de la demanda per a cada conjunt de dades, es defineixen els desplaçaments més rellevants que s'inclouran com a variables d'entrada en el model predictiu. La selecció d'aquests desplaçaments es basa en la magnitud de la correlació entre els valors de demanda passats i la variable a predir. Per cada conjunt de dades, s'ha establert un llindar de correlació que determina quins desplaçaments s'inclouen en el model.

Horitzó temporal	A1IDA	A1VUELTA	A2IDA	A2VUELTA
2 dies	Correlació > 0.7	Correlació > 0.6	Correlació > 0.6	Correlació > 0.6
1 setmana	Correlació > 0.7	Correlació > 0.6	Correlació > 0.6	Correlació > 0.6
2 setmanes	Correlació > 0.7	Correlació > 0.6	Correlació > 0.6	Correlació > 0.6
1 mes	Correlació > 0.5	Correlació > 0.5	Correlació > 0.5	Correlació > 0.5

6.3. Selecció de models

Per a realitzar la predicció, s'han seleccionat diferents tipus de models de *machine learning*, incloent models lineals, arbres de decisió i xarxes neuronals, concretament:

Models Lineals

Els models lineals es caracteritzen per la seva simplicitat i interpretabilitat. Assumeixen una relació lineal entre les variables predictives i la variable objectiu.

- **Regressió lineal:** Model bàsic per capturar relacions lineals.
- **Lasso:** Inclou regularització L1, útil per seleccionar les variables més rellevants eliminant les menys importants.
- **Ridge:** Inclou regularització L2 per evitar problemes de sobre ajustament en conjunts de dades amb moltes variables.
- **ElasticNet:** Combina les tècniques L1 i L2 per aprofitar els avantatges de totes dues.

Arbres de Decisió

Els arbres de decisió són models no lineals que seleccionen de manera automàtica les variables més rellevants. Són capaços de capturar relacions més complexes entre les variables i s'han utilitzat en diferents variants:

- **Decision Tree Regressor:** Un arbre de decisió simple que divideix iterativament les dades en funció de les variables més informatives.
- **Random Forest Regressor:** Una combinació d'arbres de decisió independents, que millora la precisió i la robustesa del model reduint el risc de sobre ajustament.
- **Gradient Boosting Regressor:** Una tècnica basada en arbres de decisió que optimitza la predicció iterativament, millorant l'error de cada etapa.
- **SVM (Support Vector Machines):** Tot i ser menys habituals, les màquines de vector suport s'han utilitzat per identificar relacions no lineals entre les variables.

Xarxes Neuronals

Les xarxes neuronals s'utilitzen per modelar relacions no lineals i capturar patrons complexos a llarg termini en les dades. Aquestes tècniques són menys interpretables que altres models, però són extremadament potents en la gestió de grans volums de dades:

- **RNN (Recurrent Neural Networks):** Xarxes especialment dissenyades per processar

dades seqüencials, com ara sèries temporals. Són capaces de retenir informació de prediccions anteriors.

- **LSTM (Long Short-Term Memory)**: Una variant de les RNN amb capacitats avançades per recordar informació a llarg termini, ideal per prediccions que depenen de contextos temporals extensos.
- **GRU (Gated Recurrent Unit)**: Similars a les LSTM, però amb una estructura més senzilla i eficiència computacional superior.

S'ha entrenat un total de 20 models, ja que els períodes de predicció influeixen en la disponibilitat de la informació de la demanda passada. En concret, els models s'han configurat per predir la demanda a partir de diferents finestres temporals:

- **2 dies**: Els primers valors disponibles de demanda passada corresponen a les últimes 72 hores (lag 72):
- **1 setmana**: Els primers valors disponibles de demanda passada corresponen a les últimes 168 hores (lag 168)
- **2 setmanes**: Els primers valors disponibles de demanda passada corresponen a les últimes 334 hores (lag 334)
- **1 mes**: Els primers valors disponibles de demanda passada corresponen a les últimes 720 hores (lag 720)

Aquest disseny es justifica perquè, a mesura que s'allarga el període de predicció, la informació de la demanda passada es redueix progressivament, ja que no està disponible per ser incorporada com a variable predictiva en les finestres temporals més grans.

6.4. Resultats

Després d'entrenar els models de predicció per a cada conjunt de dades (**A1IDA**, **A1VUELTA**, **A2IDA** i **A2VUELTA**), es presenten les mètriques d'avaluació més rellevants.

Aquestes mètriques permeten avaluar la precisió i la capacitat predictiva dels models.

Les mètriques utilitzades són:

- **RMSE (Root Mean Squared Error)**: Mesura la desviació mitjana quadràtica entre els valors reals i els valors predits.
- **MAE (Mean Absolute Error)**: Representa l'error absolut mitjà entre els valors predits i els reals.
- **MSE (Mean Squared Error)**: És la mitjana de l'error quadràtic, que penalitza més fortament els errors grans.
- **R² (Coeficient de Determinació)**: Indica la proporció de la variabilitat de la variable objectiu que el model pot explicar, on un valor proper a 1 indica una bona capacitat predictiva.

6.4.1. Models lineals

Es presenten els resultats dels models sense millorar els hiperparàmetres amb l'objectiu d'identificar quin model s'adapta millor a les dades. Aquesta fase inicial permet establir una línia de base comparativa per avaluar el rendiment de models lineals.

Data set: A1IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	36.84	3050.52	55.23	0.81	36.46	2874.88	53.62	0.82
	Lasso	36.81	3135.75	55.99	0.81	35.78	2867.00	53.54	0.82
	Ridge	36.84	3050.52	55.23	0.81	36.45	2874.75	53.62	0.82
	ElasticNet	38.29	3255.41	57.06	0.80	37.52	2990.53	54.69	0.81
1 setmana	Linear Regression	37.30	3127.98	55.93	0.81	37.47	3018.27	54.94	0.81
	Lasso	37.17	3183.26	56.42	0.81	36.90	3037.78	55.12	0.81
	Ridge	37.30	3127.98	55.93	0.81	37.47	3018.24	54.94	0.81
	ElasticNet	39.41	3347.21	57.86	0.80	39.06	3186.22	56.45	0.80
2 setmanes	Linear Regression	39.67	3409.63	58.39	0.79	40.66	3396.80	58.28	0.78
	Lasso	39.66	3504.76	59.20	0.79	39.96	3428.85	58.56	0.78
	Ridge	39.67	3409.63	58.39	0.79	40.65	3396.69	58.28	0.78
	ElasticNet	42.91	3783.52	61.51	0.77	43.07	3677.46	60.64	0.76
4 setmanes	Linear Regression	40.28	3438.50	58.64	0.79	40.21	3297.33	57.42	0.79
	Lasso	39.99	3504.95	59.20	0.79	38.62	3247.34	56.99	0.79
	Ridge	40.28	3438.50	58.64	0.79	40.20	3297.07	57.42	0.79
	ElasticNet	40.77	3557.66	59.65	0.78	39.12	3298.57	57.43	0.79

Data set: A1VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	38.92	2997.04	54.75	0.75	37.43	2721.96	52.17	0.73
	Lasso	38.97	3039.55	55.13	0.75	36.69	2658.28	51.56	0.74
	Ridge	38.92	2997.04	54.75	0.75	37.43	2721.74	52.17	0.73
	ElasticNet	39.56	3090.34	55.59	0.74	37.07	2680.59	51.77	0.74
1 setmana	Linear Regression	39.15	3016.39	54.92	0.75	37.74	2756.65	52.50	0.73
	Lasso	39.23	3060.63	55.32	0.75	36.90	2685.84	51.83	0.74
	Ridge	39.15	3016.39	54.92	0.75	37.73	2756.41	52.50	0.73
	ElasticNet	39.83	3113.40	55.80	0.74	37.23	2702.68	51.99	0.73
2 setmanes	Linear Regression	39.94	3106.77	55.74	0.74	38.93	2887.78	53.74	0.72
	Lasso	40.06	3151.46	56.14	0.74	37.47	2753.64	52.48	0.73
	Ridge	39.94	3106.77	55.74	0.74	38.92	2887.45	53.73	0.72
	ElasticNet	40.62	3203.72	56.60	0.74	37.78	2762.05	52.56	0.73

4 setmanes	Linear Regression	42.32	3436.14	58.62	0.72	41.41	3204.36	56.61	0.69
	Lasso	42.42	3489.22	59.07	0.71	39.92	3070.08	55.41	0.70
	Ridge	42.32	3436.14	58.62	0.72	41.40	3204.12	56.60	0.69
	ElasticNet	43.14	3536.91	59.47	0.71	40.21	3077.34	55.47	0.70

Data set: A2IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	24.39	1223.73	34.98	0.76	27.71	1475.62	38.41	0.70
	Lasso	24.22	1265.20	35.57	0.76	26.81	1463.80	38.26	0.70
	Ridge	24.39	1223.73	34.98	0.76	27.71	1475.59	38.41	0.70
	ElasticNet	24.53	1296.22	36.00	0.75	26.96	1463.22	38.25	0.70
1 setmana	Linear Regression	24.52	1238.58	35.19	0.76	27.71	1470.09	38.34	0.70
	Lasso	24.35	1276.31	35.73	0.75	26.89	1465.22	38.28	0.70
	Ridge	24.52	1238.58	35.19	0.76	27.71	1470.07	38.34	0.70
	ElasticNet	24.68	1307.72	36.16	0.75	27.05	1461.79	38.23	0.70
2 setmanes	Linear Regression	25.28	1306.32	36.14	0.75	28.09	1493.17	38.64	0.69
	Lasso	25.06	1346.75	36.70	0.74	27.08	1478.94	38.46	0.70
	Ridge	25.28	1306.32	36.14	0.75	28.09	1493.14	38.64	0.69
	ElasticNet	25.36	1374.08	37.07	0.73	27.38	1486.85	38.56	0.69
4 setmanes	Linear Regression	27.10	1437.91	37.92	0.72	29.03	1554.55	39.43	0.68
	Lasso	26.82	1486.04	38.55	0.71	27.97	1536.62	39.20	0.68
	Ridge	27.10	1437.91	37.92	0.72	29.03	1554.52	39.43	0.68
	ElasticNet	27.12	1530.97	39.13	0.70	28.26	1557.21	39.46	0.68

Data set: A2VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	29.61	1736.84	41.68	0.73	32.59	2070.51	45.50	0.68
	Lasso	29.59	1766.75	42.03	0.72	32.43	2074.54	45.55	0.68
	Ridge	29.61	1736.84	41.68	0.73	32.59	2070.47	45.50	0.68
	ElasticNet	30.34	1815.82	42.61	0.71	32.87	2079.47	45.60	0.68
1 setmana	Linear Regression	29.95	1765.91	42.02	0.72	32.76	2073.31	45.53	0.68
	Lasso	29.92	1793.08	42.34	0.72	32.74	2081.06	45.62	0.68
	Ridge	29.95	1765.91	42.02	0.72	32.76	2073.27	45.53	0.68
	ElasticNet	30.81	1851.33	43.03	0.71	33.32	2085.79	45.67	0.68
2 setmanes	Linear Regression	31.39	1906.67	43.67	0.70	33.54	2158.20	46.46	0.66
	Lasso	31.31	1932.67	43.96	0.70	33.45	2159.83	46.47	0.66
	Ridge	31.39	1906.67	43.67	0.70	33.54	2158.16	46.46	0.66
	ElasticNet	32.46	2011.36	44.85	0.68	34.41	2180.58	46.70	0.66

4 setmanes	Linear Regression	32.82	2037.63	45.14	0.68	34.02	2198.80	46.89	0.66
	Lasso	32.82	2087.58	45.69	0.67	34.49	2250.87	47.44	0.65
	Ridge	32.82	2037.63	45.14	0.68	34.01	2198.81	46.89	0.66
	ElasticNet	33.40	2127.12	46.12	0.66	35.27	2288.80	47.84	0.65

Les conclusions generals pels models lineals son:

- A1IDA és el *data set* amb els millors resultats globals, probablement degut a un patró de dades més regular i previsible.
- La finestra de 2 dies és la que presenta els resultats més precisos en tots els *data sets*, mentre que les finestres més llargues (4 setmanes) tenen un rendiment inferior.
- En general, la regressió lineal ha estat consistentment el millor model per a predir la demanda, superant lleugerament altres models com *Lasso*, *Ridge* i *ElasticNet*.

Es presenten els resultats de la Regressió Lineal després d'excloure les variables significatives identificades en els apartats anteriors. Concretament, s'han eliminat les variables *metro*, *pluja*, *inici_esdeveniment* i *fi_esdeveniment* amb l'objectiu d'avaluar l'impacte d'aquestes variables en la capacitat predictiva del model.

Data set: A1IDA (sense variables significatives)

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	36.79	3064.00	55.35	0.81	36.26	2883.90	53.70	0.82
1 setmana	Linear Regression	37.16	3143.66	56.07	0.81	37.22	3036.58	55.11	0.81
2 setmanes	Linear Regression	39.55	3457.69	58.80	0.79	40.32	3434.39	58.60	0.78
4 setmanes	Linear Regression	40.19	3475.60	58.95	0.79	39.89	3304.39	57.48	0.79

Data set: A1VUELTA (sense variables significatives)

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	38.91	3012.82	54.89	0.75	36.92	2678.85	51.76	0.74
1 setmana	Linear Regression	39.14	3035.86	55.10	0.75	37.14	2709.66	52.05	0.73
2 setmanes	Linear Regression	39.93	3134.85	55.99	0.74	37.98	2817.98	53.08	0.72
4 setmanes	Linear Regression	42.37	3470.49	58.91	0.71	40.79	3173.48	56.33	0.69

Data set: A2VUELTA (sense variables significatives)

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	29.53	1749.48	41.83	0.73	32.35	2065.39	45.45	0.68
1 setmana	Linear Regression	29.86	1781.92	42.21	0.72	32.49	2062.99	45.42	0.68
2 setmanes	Linear Regression	31.32	1932.42	43.96	0.70	33.07	2139.39	46.25	0.67
4 setmanes	Linear Regression	32.88	2073.02	45.53	0.67	33.82	2186.78	46.76	0.66

Data set: A2VUELTA (sense variables significatives)

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Linear Regression	29.53	1749.48	41.83	0.73	32.35	2065.39	45.45	0.68
1 setmana	Linear Regression	29.86	1781.92	42.21	0.72	32.49	2062.99	45.42	0.68
2 setmanes	Linear Regression	31.32	1932.42	43.96	0.70	33.07	2139.39	46.25	0.67
4 setmanes	Linear Regression	32.88	2073.02	45.53	0.67	33.82	2186.78	46.76	0.66

Els resultats obtinguts mostren que no hi ha diferències significatives en el rendiment dels models amb i sense les variables significatives. Per aquesta raó, l'exclusió d'aquestes variables simplifica el model sense comprometre la precisió.

6.4.2. Arbres de decisió

A continuació es presenten els resultats del llançament inicial dels models basats en arbres de decisió sense millorar els seus paràmetres d'hiperoptimització. Aquesta fase té com a objectiu identificar quins models ofereixen el millor rendiment abans d'ajustar-ne els hiperparàmetres.

Data set: A1IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Decision Tree	0.000	0.000	0.000	1.000	54.596	6589.090	81.173	0.579
	Random Forest	13.043	398.339	19.958	0.976	35.916	2864.116	53.517	0.817
	Gradient Boosting	33.028	2435.988	49.356	0.852	36.082	2879.990	53.666	0.816
	SVM	39.423	3421.594	58.494	0.792	42.574	3592.729	59.939	0.770
1 setmana	Decision Tree	0.000	0.000	0.000	1.000	50.871	5726.485	75.674	0.634
	Random Forest	12.848	387.977	19.697	0.977	36.084	2902.639	53.876	0.814
	Gradient Boosting	33.262	2463.898	49.638	0.851	36.881	2963.443	54.438	0.810
	SVM	41.130	3635.434	60.295	0.781	48.035	4354.146	65.986	0.721
2 setmanes	Decision Tree	0.000	0.000	0.000	1.000	55.338	6755.404	82.191	0.568
	Random Forest	13.130	404.288	20.107	0.975	37.125	3046.220	55.193	0.805
	Gradient Boosting	33.969	2548.868	50.486	0.845	37.775	3118.577	55.844	0.801
	SVM	42.942	3900.036	62.450	0.763	48.062	4447.801	66.692	0.716
4 setmanes	Decision Tree	0.000	0.000	0.000	1.000	52.912	6113.073	78.186	0.609
	Random Forest	13.036	388.864	19.720	0.976	36.274	2957.185	54.380	0.811
	Gradient Boosting	34.698	2627.395	51.258	0.840	37.923	3099.318	55.672	0.802
	SVM	42.206	3769.905	61.400	0.770	48.195	4457.706	66.766	0.715

Data set: A1VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Decision Tree	0.000	0.000	0.000	1.000	53.799	5685.420	75.402	0.441
	Random Forest	14.288	417.223	20.426	0.966	37.025	2709.933	52.057	0.734
	Gradient Boosting	35.670	2506.715	50.067	0.793	37.218	2714.048	52.097	0.733
	SVM	40.366	3261.366	57.108	0.731	39.590	2891.878	53.776	0.716
1 setmana	Decision Tree	0.000	0.000	0.000	1.000	56.093	6076.697	77.953	0.403
	Random Forest	14.359	419.535	20.483	0.965	36.812	2710.583	52.063	0.734
	Gradient Boosting	35.876	2528.055	50.280	0.791	37.272	2720.560	52.159	0.733
	SVM	40.595	3281.679	57.286	0.729	40.026	2945.221	54.270	0.711
2 setmanes	Decision Tree	0.000	0.000	0.000	1.000	55.033	5943.455	77.094	0.416
	Random Forest	14.456	425.095	20.618	0.965	37.134	2721.185	52.165	0.733
	Gradient Boosting	36.424	2594.452	50.936	0.786	37.937	2780.515	52.731	0.727
	SVM	41.215	3366.562	58.022	0.722	41.184	3097.213	55.653	0.696
4 setmanes	Decision Tree	0.000	0.000	0.000	1.000	58.437	6647.439	81.532	0.348
	Random Forest	14.352	418.263	20.451	0.965	39.125	2987.852	54.661	0.707
	Gradient Boosting	37.678	2763.303	52.567	0.772	39.805	3002.315	54.793	0.706
	SVM	43.432	3681.542	60.676	0.696	46.683	3876.162	62.259	0.620

Data set: A2IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Decision Tree	0.000	0.000	0.000	1.000	37.495	2947.167	54.288	0.392
	Random Forest	8.723	167.236	12.932	0.968	26.916	1507.381	38.825	0.689
	Gradient Boosting	21.765	992.545	31.505	0.808	27.135	1510.935	38.871	0.688
	SVM	24.456	1329.389	36.461	0.743	27.963	1537.917	39.216	0.683
1 setmana	Decision Tree	0.000	0.000	0.000	1.000	37.386	2878.995	53.656	0.406
	Random Forest	8.718	167.767	12.952	0.968	26.773	1480.040	38.471	0.695
	Gradient Boosting	21.786	997.504	31.583	0.807	26.707	1478.101	38.446	0.695
	SVM	24.615	1343.802	36.658	0.740	28.176	1550.058	39.371	0.680
2 setmanes	Decision Tree	0.000	0.000	0.000	1.000	38.742	3126.474	55.915	0.355
	Random Forest	8.790	171.334	13.089	0.967	27.136	1520.582	38.995	0.686
	Gradient Boosting	22.284	1045.712	32.337	0.798	27.148	1494.461	38.658	0.692
	SVM	25.098	1393.674	37.332	0.730	28.984	1628.843	40.359	0.664
4 setmanes	Decision Tree	0.000	0.000	0.000	1.000	40.455	3368.886	58.042	0.303
	Random Forest	8.608	162.017	12.729	0.969	27.749	1624.148	40.301	0.664
	Gradient Boosting	22.897	1089.003	33.000	0.789	27.433	1532.604	39.148	0.683
	SVM	26.178	1487.946	38.574	0.712	30.403	1764.693	42.008	0.635

Data set: A2VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Decision Tree	0.000	0.000	0.000	1.000	45.661	4188.592	64.719	0.349
	Random Forest	10.964	250.930	15.841	0.961	32.278	2098.259	45.807	0.674
	Gradient Boosting	27.360	1493.496	38.646	0.765	32.469	2117.732	46.019	0.671
	SVM	30.294	1870.572	43.250	0.706	34.576	2255.775	47.495	0.650
1 setmana	Decision Tree	0.000	0.000	0.000	1.000	45.832	4211.683	64.897	0.346
	Random Forest	10.966	249.685	15.801	0.961	32.361	2098.627	45.811	0.674
	Gradient Boosting	27.385	1502.927	38.768	0.764	32.512	2116.465	46.005	0.671
	SVM	30.560	1898.752	43.575	0.702	34.976	2294.862	47.905	0.644
2 setmanes	Decision Tree	0.000	0.000	0.000	1.000	45.393	4244.954	65.153	0.341
	Random Forest	11.131	257.471	16.046	0.960	32.576	2116.825	46.009	0.671
	Gradient Boosting	28.292	1588.198	39.852	0.750	32.640	2131.996	46.174	0.669
	SVM	31.732	2032.503	45.083	0.681	36.287	2454.163	49.540	0.619
4 setmanes	Decision Tree	0.000	0.000	0.000	1.000	46.486	4388.533	66.246	0.321
	Random Forest	11.009	250.243	15.819	0.961	33.391	2211.406	47.026	0.658
	Gradient Boosting	29.136	1676.272	40.942	0.736	33.045	2147.107	46.337	0.668
	SVM	32.341	2102.430	45.852	0.668	37.301	2549.424	50.492	0.606

Les conclusions generals dels models basats en arbres de decisió son:

- Els models *Random Forest Regressor* i *Gradient Boosting Regressor* presenten, en la majoria dels casos, millors mètriques de predicció respecte a altres models com *Decision Tree* o SVM. Les diferències en les mètriques de prova, especialment en MAE, MSE i R², fan que es considerin opcions adequades per a la predicció de la demanda.
- Tot i que *Decision Tree Regressor* presenta un error d'entrenament nul (perfecte ajust), el seu rendiment en la prova és clarament inferior, ja que pateix de sobre ajustament. Per contra, models més robustos com *Random Forest Regressor* i *Gradient Boosting Regressor* presenten un equilibri entre entrenament i prova.
- A mesura que augmenta la finestra de predicció (de 2 dies a 4 setmanes), s'observa un augment de l'error (MAE, MSE i RMSE), fet que evidencia la dificultat de predir a llarg termini.
- Els resultats indiquen que *Random Forest Regressor* i *Gradient Boosting Regressor* són les millors opcions, ja que ofereixen precisió, robustesa i generalització. En la selecció final, es podria prioritzar el model més senzill o el que tingui menor complexitat computacional, sempre que l'impacte en la precisió sigui mínim.

A continuació es presenten els resultats de *Random Forest Regressor* i *Gradient Boosting Regressor* amb optimització d'hiperparàmetres:

Data set: A1IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Random Forest	33.463	2614.429	51.131	0.841	35.531	2835.419	53.249	0.819
	Gradient Boosting	31.967	2326.977	48.239	0.859	35.479	2815.821	53.064	0.820
1 setmana	Random Forest	33.618	2633.074	51.313	0.841	35.530	2848.517	53.372	0.818
	Gradient Boosting	33.108	2480.562	49.805	0.850	35.696	2821.552	53.118	0.820
2 setmanes	Random Forest	34.285	2731.752	52.266	0.834	36.621	2997.498	54.749	0.808
	Gradient Boosting	34.470	2691.439	51.879	0.836	37.141	3021.901	54.972	0.807
4 setmanes	Random Forest	34.935	2781.628	52.741	0.830	36.297	2949.556	54.310	0.811
	Gradient Boosting	33.828	2554.111	50.538	0.844	37.001	2965.528	54.457	0.810

Data set: A1VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Random Forest	37.133	2776.866	52.696	0.771	36.904	2708.450	52.043	0.734
	Gradient Boosting	32.993	2176.507	46.653	0.820	36.439	2620.472	51.191	0.743
1 setmana	Random Forest	37.262	2786.771	52.790	0.770	37.007	2722.241	52.175	0.733
	Gradient Boosting	34.910	2433.320	49.329	0.799	36.897	2672.416	51.695	0.737
2 setmanes	Random Forest	36.937	2721.400	52.167	0.775	37.285	2739.721	52.342	0.731
	Gradient Boosting	35.967	2541.112	50.409	0.790	37.449	2716.508	52.120	0.733
4 setmanes	Random Forest	38.559	2944.567	54.264	0.757	39.160	2991.675	54.696	0.707
	Gradient Boosting	35.802	2533.866	50.338	0.790	39.100	2950.612	54.320	0.711

Data set: A2IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Random Forest	22.231	1094.575	33.084	0.788	27.206	1539.113	39.232	0.683
	Gradient Boosting	20.442	889.340	29.822	0.828	26.491	1450.972	38.092	0.701
1 setmana	Random Forest	22.304	1101.969	33.196	0.787	27.273	1545.003	39.307	0.681
	Gradient Boosting	19.739	826.950	28.757	0.840	26.471	1441.164	37.963	0.703
2 setmanes	Random Forest	22.880	1158.434	34.036	0.776	27.129	1516.543	38.943	0.687
	Gradient Boosting	21.285	963.068	31.033	0.814	26.637	1453.239	38.121	0.700
4 setmanes	Random Forest	23.663	1209.700	34.781	0.766	28.011	1636.915	40.459	0.662
	Gradient Boosting	24.286	1236.957	35.170	0.760	27.448	1531.444	39.134	0.683

Data set: A2VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	Random Forest	27.568	1565.552	39.567	0.754	32.586	2144.328	46.307	0.667
	Gradient Boosting	26.509	1418.334	37.661	0.777	31.897	2047.062	45.244	0.682
1 setmana	Random Forest	27.828	1587.827	39.848	0.751	32.618	2133.609	46.191	0.669
	Gradient Boosting	26.722	1436.884	37.906	0.774	31.934	2047.228	45.246	0.682
2 setmanes	Random Forest	28.572	1673.173	40.904	0.737	32.673	2150.374	46.372	0.666
	Gradient Boosting	27.419	1505.915	38.806	0.763	32.090	2065.514	45.448	0.679
4 setmanes	Random Forest	29.230	1738.029	41.690	0.726	33.617	2250.338	47.438	0.652
	Gradient Boosting	28.614	1637.335	40.464	0.742	32.637	2123.134	46.077	0.672

A partir de les taules comparatives dels resultats optimitzats per als models *Random Forest Regressor* i *Gradient Boosting Regressor* en cadascun dels data sets i finestres temporals, es poden extreure les següents conclusions generals:

- *Gradient Boosting Regressor* sol obtenir millors resultats en termes d'errors més baixos (MAE, MSE, RMSE) i coeficients R² més elevats, especialment en finestres curtes com 2 dies.
- *Random Forest Regressor* manté un rendiment robust i competitiu, amb una diferència mínima respecte a *Gradient Boosting Regressor*, especialment en finestres més llargues (e.g., 4 setmanes).
- A1IDA i A1VUELTA presenten patrons més consistents i predictibles, amb errors més baixos i coeficients R² més alts.
- En les finestres curtes de predicció (2 dies i 1 setmana) els errors són més baixos i els coeficients R² més alts.
- *Gradient Boosting Regressor* és el millor model tot i que *Random Forest Regressor* podria ser una alternativa si es busqués més eficiència computacional.

6.4.3. Xarxes Neuronals

Per últim, es presenten els resultats del llançament inicial dels models basats xarxes neuronals sense millorar els seus paràmetres d'hiperoptimització.

Data set: A1IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	LSTM	36.180	2578.004	50.774	0.846	39.922	3126.403	55.914	0.791
	GRU	36.839	2858.367	53.464	0.829	39.421	3278.664	57.260	0.781
	RNN	39.382	3041.944	55.154	0.818	39.361	3061.801	55.334	0.795
1 setmana	LSTM	32.925	2226.155	47.182	0.868	41.656	3633.176	60.276	0.758
	GRU	35.475	2632.801	51.311	0.844	38.287	3140.214	56.038	0.791
	RNN	42.438	3452.301	58.756	0.795	40.435	3075.300	55.455	0.795
2 setmanes	LSTM	40.204	3204.099	56.605	0.809	42.881	3509.622	59.242	0.765
	GRU	35.383	2608.167	51.070	0.844	42.422	3678.686	60.652	0.754
	RNN	42.273	3270.151	57.185	0.805	42.264	3265.675	57.146	0.782
4 setmanes	LSTM	35.889	2707.354	52.032	0.838	41.394	3480.031	58.992	0.767
	GRU	39.763	3299.030	57.437	0.802	41.101	3443.888	58.685	0.769
	RNN	38.712	2868.568	53.559	0.828	40.660	3271.319	57.195	0.781

Data set: A1VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	LSTM	36.589	2507.876	50.079	0.798	41.628	3080.800	55.505	0.679
	GRU	40.736	3039.800	55.134	0.756	41.509	3172.866	56.328	0.670
	RNN	41.296	3114.126	55.804	0.750	40.008	2914.948	53.990	0.696
1 setmana	LSTM	38.094	2721.692	52.170	0.781	43.455	3374.249	58.088	0.649
	GRU	39.018	2843.434	53.324	0.771	39.892	3018.659	54.942	0.686
	RNN	42.975	3033.921	55.081	0.756	44.766	3223.537	56.776	0.664
2 setmanes	LSTM	37.949	2680.653	51.775	0.785	43.043	3161.212	56.225	0.671
	GRU	40.189	3019.422	54.949	0.757	40.744	3066.712	55.378	0.681
	RNN	39.722	2914.294	53.984	0.766	38.352	2837.035	53.264	0.704
4 setmanes	LSTM	48.269	4111.817	64.123	0.669	55.374	5066.567	71.180	0.471
	GRU	40.160	3070.158	55.409	0.753	39.772	3032.785	55.071	0.683
	RNN	42.222	3263.573	57.128	0.738	39.520	2961.610	54.421	0.691

Data set: A2IDA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	LSTM	23.959	1128.794	33.598	0.787	27.765	1557.744	39.468	0.653
	GRU	24.200	1141.748	33.790	0.784	29.616	1727.177	41.559	0.615
	RNN	25.099	1152.303	33.946	0.782	29.890	1678.493	40.969	0.626
1 setmana	LSTM	22.169	978.409	31.280	0.815	28.470	1672.991	40.902	0.627
	GRU	22.099	987.369	31.422	0.813	27.835	1600.615	40.008	0.644
	RNN	22.210	973.857	31.207	0.816	28.002	1602.615	40.033	0.643
2 setmanes	LSTM	24.112	1097.215	33.124	0.793	29.709	1661.401	40.760	0.630
	GRU	24.567	1217.094	34.887	0.770	30.201	1845.568	42.960	0.589
	RNN	26.611	1256.316	35.445	0.763	31.750	1816.629	42.622	0.595
4 setmanes	LSTM	24.821	1232.439	35.106	0.767	28.708	1679.041	40.976	0.631
	GRU	23.614	1125.079	33.542	0.787	28.857	1686.315	41.065	0.629
	RNN	25.136	1212.165	34.816	0.770	29.958	1758.401	41.933	0.613

Data set: A2VUELTA

Finestra	Model	Train MAE	Train MSE	Train RMSE	Train R ²	Test MAE	Test MSE	Test RMSE	Test R ²
2 dies	LSTM	27.167	1435.295	37.885	0.777	35.979	2495.460	49.955	0.612
	GRU	29.274	1668.991	40.853	0.740	33.965	2219.891	47.116	0.654
	RNN	28.923	1500.890	38.741	0.766	36.007	2332.075	48.292	0.637
1 setmana	LSTM	30.621	1779.864	42.188	0.723	35.422	2321.162	48.178	0.639
	GRU	26.434	1327.743	36.438	0.793	35.587	2353.236	48.510	0.634
	RNN	27.905	1404.340	37.475	0.781	34.983	2297.115	47.928	0.642
2 setmanes	LSTM	34.272	1965.874	44.338	0.694	37.346	2357.667	48.556	0.633
	GRU	26.968	1376.427	37.100	0.786	34.639	2382.756	48.813	0.629
	RNN	30.750	1708.116	41.329	0.734	35.978	2353.266	48.510	0.634
4 setmanes	LSTM	27.732	1464.041	38.263	0.771	35.271	2390.606	48.894	0.631
	GRU	26.227	1310.419	36.200	0.795	37.225	2448.625	49.484	0.623
	RNN	27.530	1437.837	37.919	0.775	36.243	2412.900	49.121	0.628

No s'han optimitzat els hiperparàmetres dels models de xarxes neuronals, ja que els resultats obtinguts mostren rendiments més baixos en comparació amb els models de *Gradient Boosting Regressor* i *Random Forest Regressor*. A més, les xarxes neuronals requereixen un temps d'entrenament molt elevat.

Després d'analitzar els resultats de tots els models provats (*Regressió Lineal*, *Random Forest Regressor*, *Gradient Boosting Regressor* i Xarxes Neuronals), es pot concloure que *Gradient Boosting Regressor* és el millor model en termes de rendiment global.

Model d'optimització de la flota d'autobusos



A partir dels resultats del model de previsió de la demanda, el model d'optimització permetrà calcular una estimació dels vehicles necessaris per cobrir els serveis definits per la demanda i la freqüència establerta.

7.1. Definició del model

Les dades de partida són les següents:

- Període d'estimació (24 hores-48 hores-1 setmana)
- Franges horàries es què es reparteix la demanda
- Línies del servei
- Freqüència per cada línia i cada franja horària
- Demanda per cada línia i cada franja horària
- Temps mig de trajecte per línia i franja horària
- *Temps mig de desplaçament entre dipòsit d'autobusos i origen/destinació*
- Capacitat mitjana per vehicle

Els conjunts del model:

Conjunt	Símbol	Descripció
H	h	Trams horaris, poden ser per franges horàries combinades amb el dia de la setmana (hora punta laborable, hora punta dissabte, hora punta diumenge, etc.)
Γ	ϱ	Línies
S	s	Serveis
I	i	Registres de demanda
J	j	Registres de freqüència.
I(j)	i	Registres de demanda associats al registre de freqüència j
I(h)	i	Registres de demanda associats al tram horari h

Els paràmetres del model:

Símbol	Descripció
T^h	Durada del tram horari h
d_i	Demanda de passatgers per registre de demanda i
s_i	Servei al que correspon el registre de demanda i
t_i	Temps mig de trajecte per registre de demanda i. Inclou l'estimació del temps de pujada i baixada dels passatgers, excepte a les parades inicials/finals (*)
f_j	Freqüència (en número de trajectes) per registre de freqüència j

c_i	Temps estimat recollida de passatgers a la pl. Catalunya / descàrrega a Aeroport /altres
d_i	Temps estimat recollida de passatgers a l'aeroport / descàrrega a Pl. Catalunya / altres
$\hat{t}_i$	Temps total estimat de trajecte, incloent-hi temps addicional de recollida/ descàrrega a origen/destí
K	Capacitat mitjana dels vehicles
A	Factor de pes per ponderar el nombre de vehicles
B	Factor de pes per ponderar l'excés de passatgers

(*) No inclou el temps de desplaçament entre la parada inicial/final i la base

Les variables del model:

Variable	Domini	Dimensió	Descripció
$\rho\lambda^h$	Z^+	$ H $	Número de vehicles necessaris
x_i	Z^+	$ I $	Número d'expedicions per registre de demanda i (freqüència real)
y_i	Z^+	$ I $	Número de passatgers que queden fora, per registre de demanda i

Les condicions del model són les següents:

1) Cal complir la freqüència acordada

$$\sum_{i \in I(j)} x_i \geq f_j \quad \forall j \in J$$

2) Ha d'haver-hi suficients vehicles per cobrir tots els serveis

$$\sum_{i \in I(h)} \hat{t}_i x_i \leq T^h \lambda^h \quad \forall h \in H$$

3) Càlcul de l'excés de demanda sobre la capacitat

$$y_i \geq d_i - Kx_i \quad \forall i \in I$$

4) Domini de les variables

$$x_i, y_i, \lambda^h \in Z^+$$

L'objectiu del model és minimitzar el número de vehicles necessaris per cobrir la demanda. S'afegeix com a segon objectiu minimitzar el nombre de passatgers que queden fora. S'afegeixen dos factors de pes, A y B, per calibrar correctament els dos components de la funció objectiu. S'afegeix com a tercer component la minimització del número de trajectes:

$$\min_{x,y,\lambda} A \sum_{h \in H} \lambda^h + B \sum_{i \in I} y_i + \sum_{i \in I} x_i$$

subjecte a les condicions 1), 2), 3) i 4)

Aquest tercer component és necessari per garantir que la solució obtinguda conté el mínim nombre de trajectes necessaris. En efecte, per les restriccions que s'estableixen, sense aquest tercer component, dues solucions amb diferent nombre de trajectes poden donar lloc al mateix valor de la funció objectiu. Afegint aquest tercer component, ens assegurem d'obtenir la solució més ajustada a la freqüència i a la demanda.

Es pot treure la condició 1), llavors el model decidirà també quina és la freqüència òptima. En aquest cas s'ha d'afegir una condició que mantingui la coherència en els serveis d'anada i tornada, del tipus:

$$\sum_{i \in I(s,j)} x_i - \sum_{i \in I(\hat{s},j)} x_i = 0 \quad \forall s, \hat{s} \in S, \text{ on } \hat{s} \text{ és la tornada del servei } s, \forall j \in J$$

7.2. Exemple aplicat

7.2.1. Valors dels conjunts i paràmetres utilitzats:

Símbol	Descripció	Valor
H	Trams horaris	0:00-5:00 5:00-7:00 7:00-22:00 22:00-0:00
Γ	Línies	A1, A2
S	Serveis	A1IDA, A1VUELTA, A2IDA, A2VUELTA
I	Registres de demanda	Model predicció, per dia, hora i servei
J	Registres de freqüència	Per línia, temporada (alta, baixa) i hora. Calculat a partir dels intervals de freqüència publicats a la web
$I(j)$	Registres de demanda associats al registre de freqüència j	Correspondència entre l'interval hora del registre de demanda i l'interval hora del registre de freqüència, tenint en compte la temporada i la línia
$I(h)$	Registres de demanda associats al tram horari h	Correspondència entre l'interval hora del registre de demanda i l'interval horari del tram horari

T^h	Durada del tram horari h	Calculada a partir dels trams horaris
d_i	Demanda de passatgers per registre de demanda i	Model predicció, per dia, hora i servei
t_i	Temps mig de trajecte per registre de demanda i. Inclou l'estimació del temps de pujada i baixada dels passatgers, excepte a les parades inicials/finals (*)	Model predicció, per dia, hora i servei
c	Temps estimat recollida de passatgers a la pl. Catalunya / descàrrega a Aeroport /altres	A1:9,58 minuts (9'35'') A2: 9,42 minuts (9'25'')
d	Temps estimat recollida de passatgers a l'aeroport / descàrrega a Pl. Catalunya / altres	A1:4,8 minuts (4'48'') A2: 12,55 minuts (12'33'')
$\hat{t}_i$	Temps total estimat de trajecte, incloent-hi temps addicional de recollida/ descàrrega a origen/destí	Si l'origen del trajecte del registre de demanda i és Pl. Catalunya: $\hat{t}_i = t_i + c$ Si l'origen del trajecte del registre de demanda i és l'Aeroport: $\hat{t}_i = t_i + d$
K	Capacitat mitjana dels vehicles	50 persones
A	Factor de pes per ponderar núm. vehicles	10
B	Factor de pes per ponderar l'excés de passatgers	1

7.2.2. Fitxer de demanda prevista

La demanda prevista és el resultat del model de previsió de la demanda, descrit al Capítol 6 amb *Gradient Boosting Regressor*. Les dades generades corresponen al període de l'11-02-2024 fins l'11-06-2024. Del fitxer original s'exclouen les dades del primer i últim dia (10-02-2024 i 12-06-2024) perquè no tenen dades per les 24 hores.

Hi ha un registre de demanda per servei, dia i hora, on es recull la demanda mitjana. No es tenen en compte les dades de les cues: tal i com s'ha explicat al Capítol 6, els passatgers que fan cua agafaran el següent trajecte en la gran majoria de casos, i una petita quantitat l'agafarà dos trajectes més tard del desitjat. Per tant, els passatgers de les cues ja estan contemplats en la demanda calculada.

Es genera una segona versió del fitxer de demanda, en la que s'incrementa la demanda de forma proporcional aplicant els següents percentatges:

Servei	Percentatge
A1IDA	17,20%
A1VUELTA	14,62%
A2IDA	20,90%
A2VUELTA	18,74%

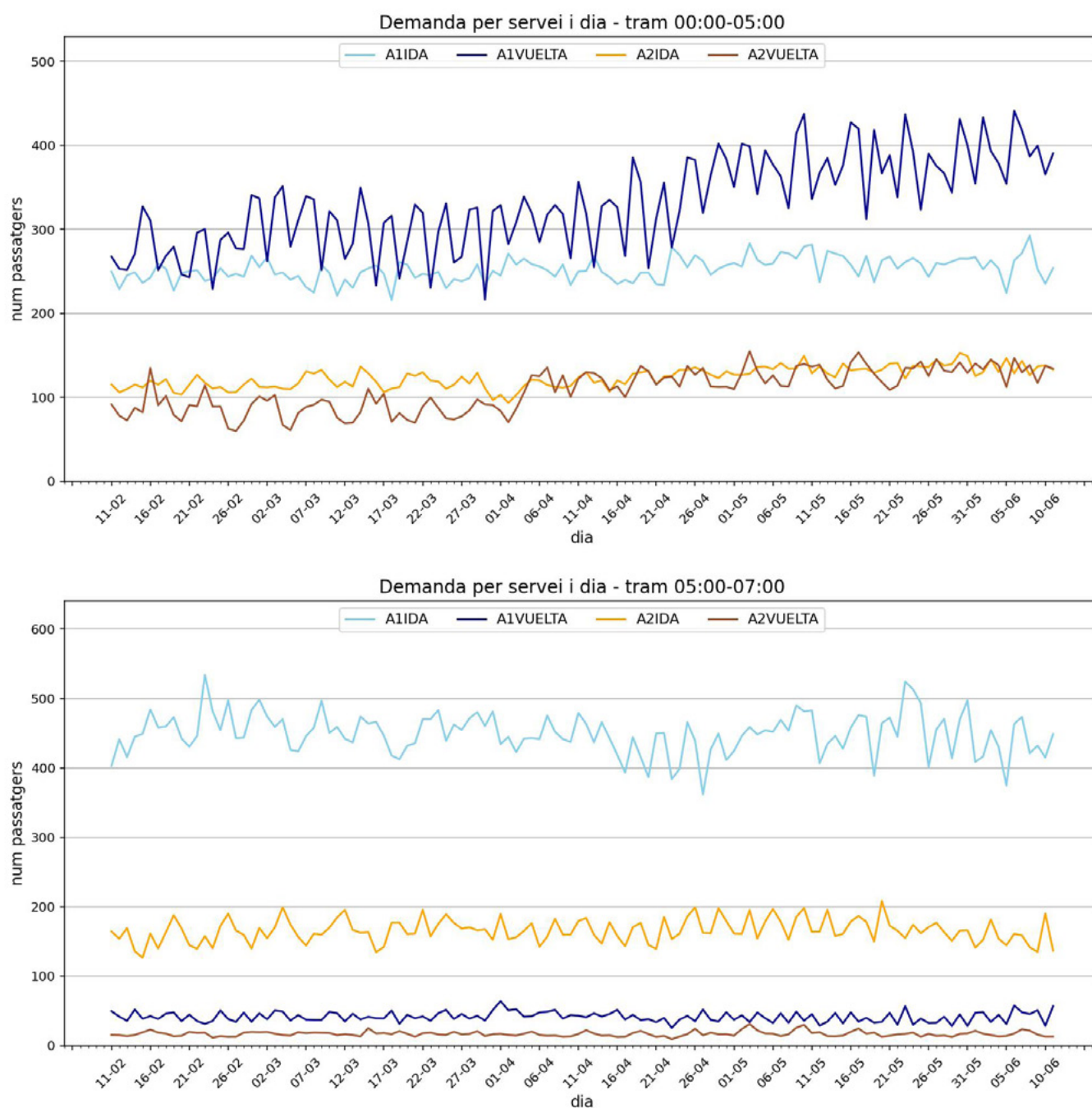
El que es fa és afegir la part de validacions del fitxer Ticketing que s'han ignorat en el procés de neteja per no estar associades a cap servei. Són els percentatges de registres del fitxer Expedicions que s'han descartat perquè els camps 'departure/arrival' contenien valors nuls per a totes les parades, i es va fer l'assumpció que els viatges no existien, quan en realitat són trajectes que s'han dut a terme per autobusos no registrats en el sistema SAE, tal i com es va confirmar quan ja s'havia finalitzat el model de previsió de la demanda.

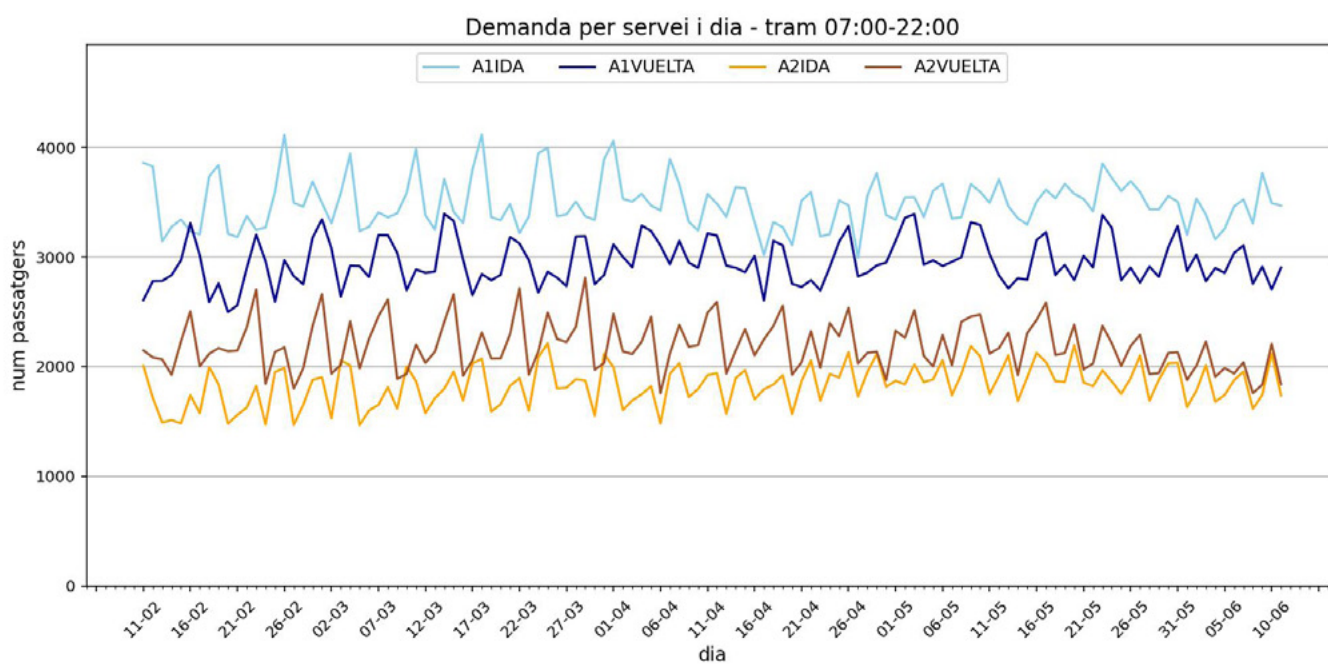
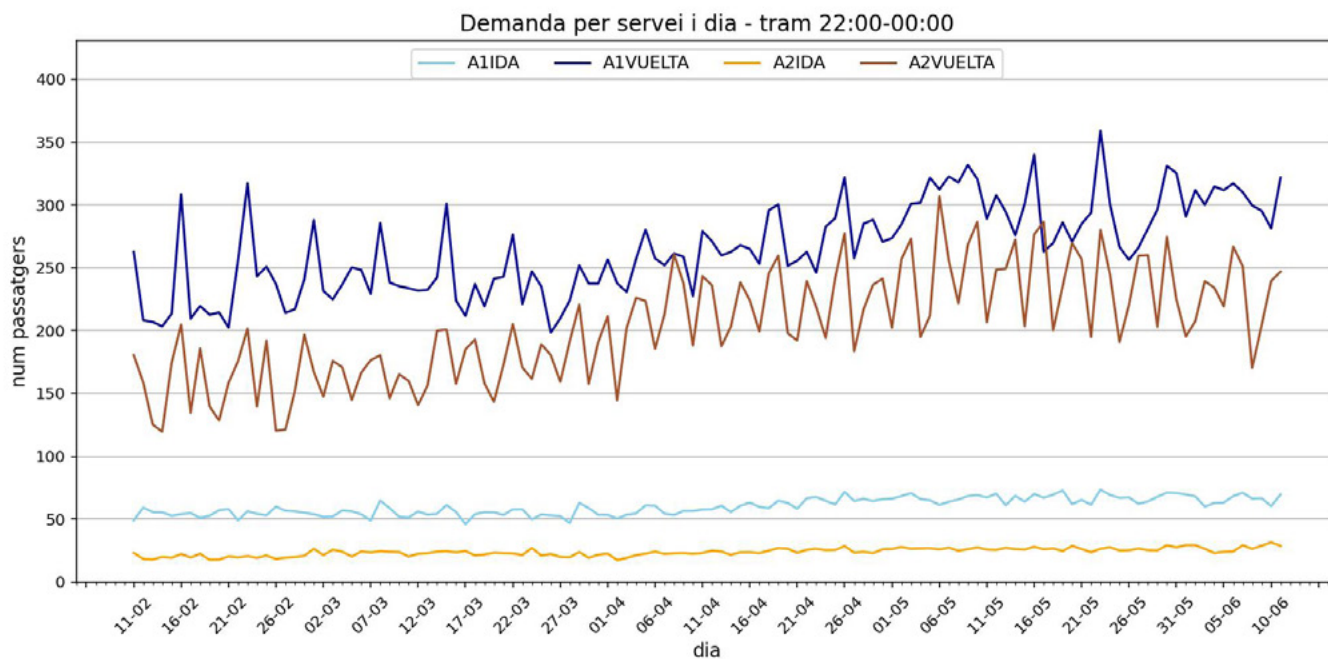
Les següents gràfiques mostren la demanda acumulada per tram horari, per servei i dia, primer pel cas amb la demanda tal i com surt del model de previsió de la demanda ('demanda original'), i després, pel cas amb la demanda incrementada segons s'ha indicat ('demanda incrementada').

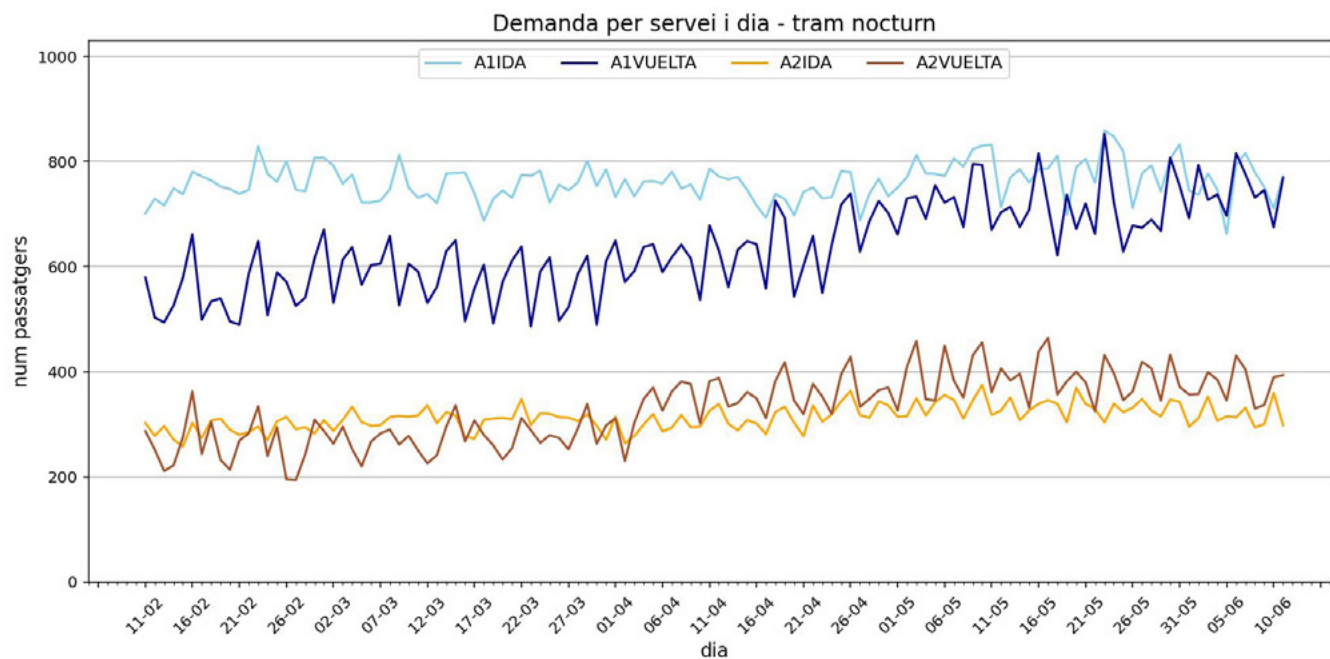
L'eix horitzontal correspon als dies, i l'eix vertical mostra el número de passatgers. Cada color correspon a un servei. Es pot apreciar la variabilitat de demanda segons el tram horari, tot i que s'ha de tenir en compte que cadascun dels trams horaris 0:00-5:00, 5:00-7:00 i 22:00-0:00 acumulen la demanda de 5,2 i 2 hores per cada dia respectivament, mentre que el tram horari 7:00-22:00 s'acumula la demanda de 15 hores per cada dia. Per una millor comparació, s'afegeix una cinquena gràfica (corresponent al tram 'nocturn') que mostra la demanda acumulada per dia dels trams horaris 0:00-5:00, 5:00-7:00 i 22:00-0:00 (9 hores per dia). També cal remarcar que hi ha representades les dues temporades, la temporada baixa des de l'11-02 fins el 31-03, i la temporada alta, des de l'1-04 fins l'11-06. En el tram diürn, de 7:00-22:00 no es pot apreciar diferència entre les temporades, mentre que en els tram nocturns (0:00-5:00, 5:00-7:00 i 22:00-0:00) si hi ha un cert creixement a la demanda a la temporada alta respecte de la demanda de la temporada baixa, més remarcant en els serveis de tornada, que fan la ruta Aeroport-Plaça Catalunya.

En haver aplicat proporcionalitat segons els percentatges dels registres descartats, les tendències de les dades del fitxer amb la demanda incrementada són similars a les de les dades del fitxer de la demanda original, només canvien els valors

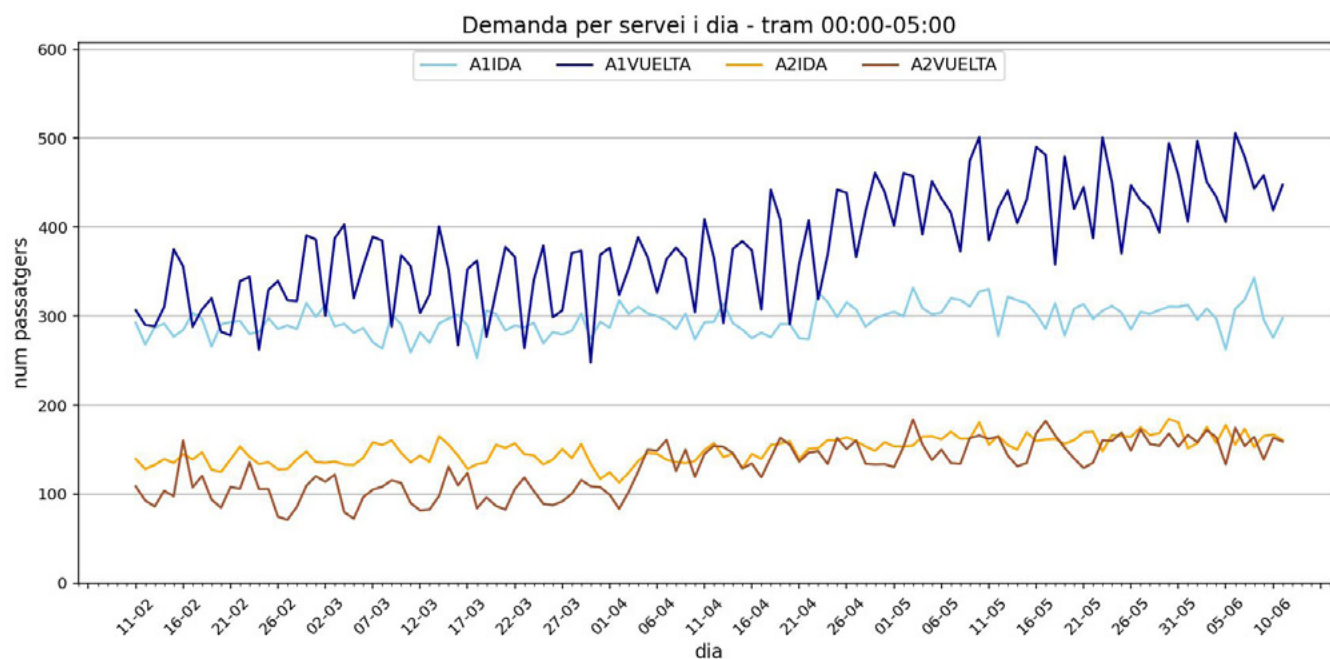
A. Demanda original

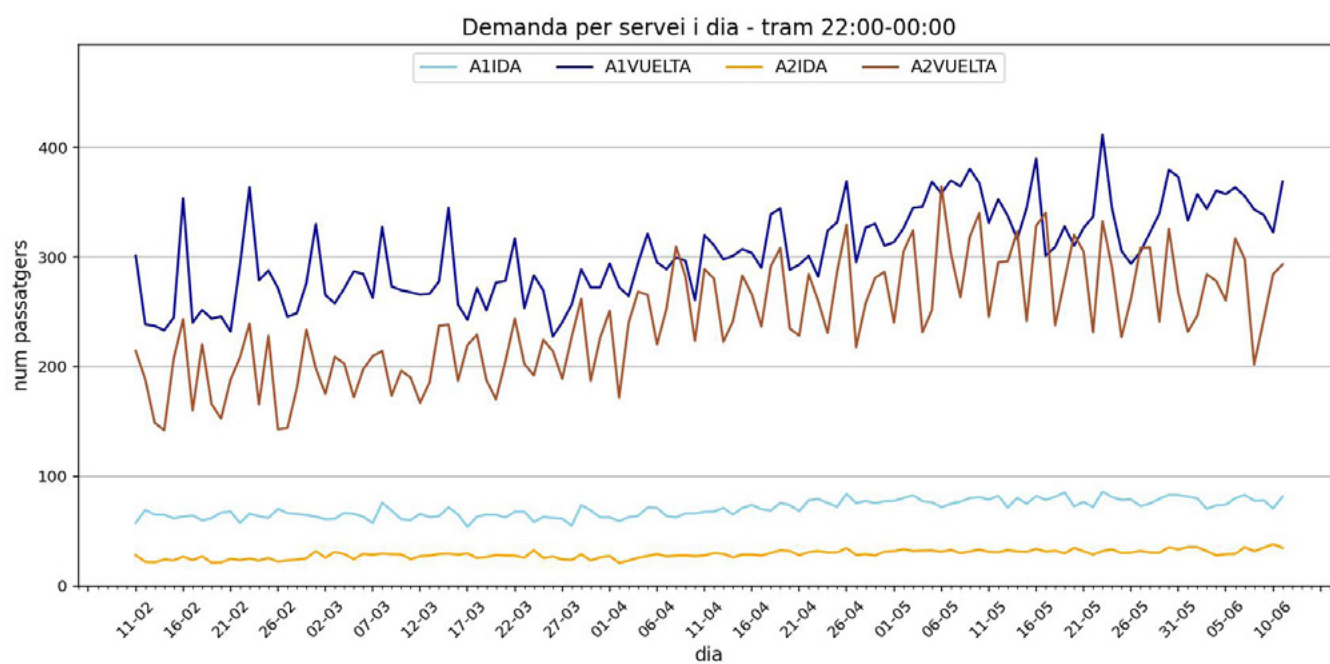
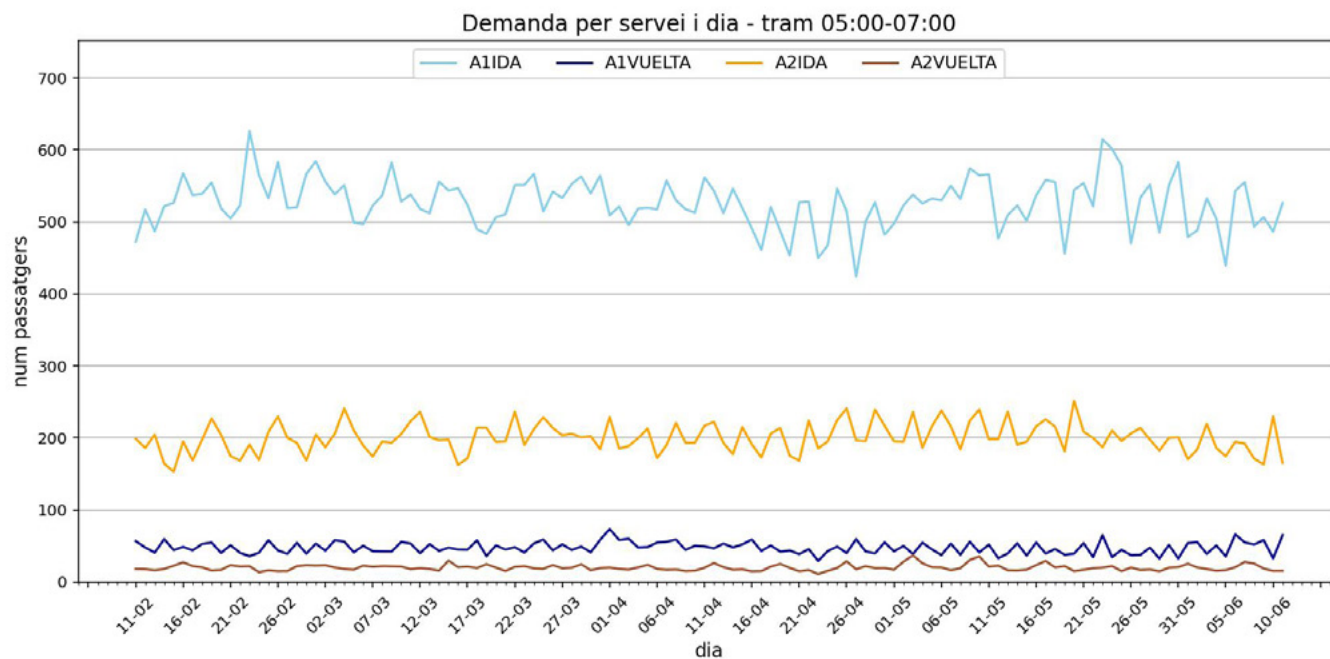


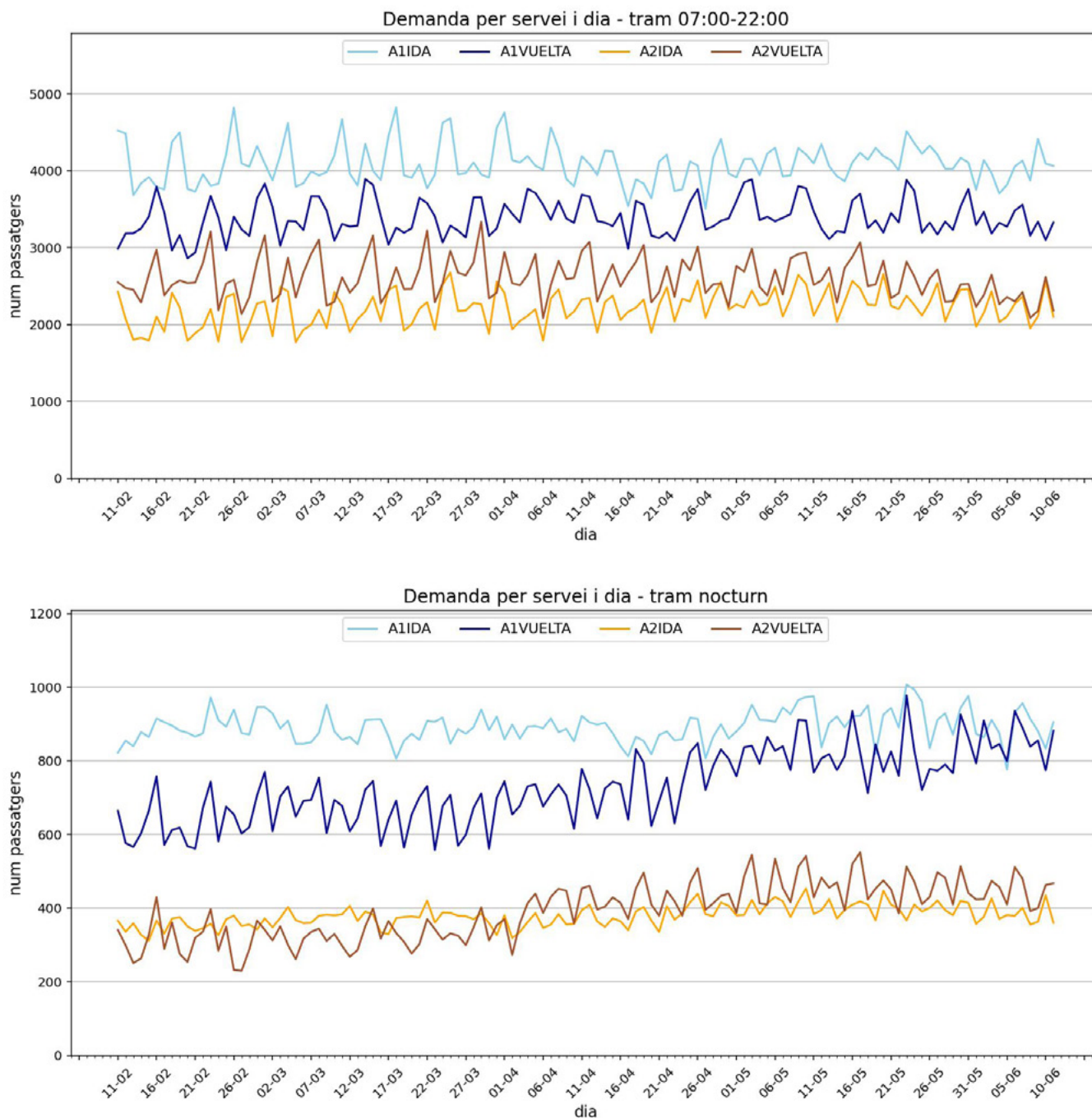




B. Demanda incrementada







7.2.3. Temps mig de trajecte

Cada registre de demanda té assignat un temps de trajecte estimat, sense incloure el temps d'espera a l'inici de parada ni el temps de descàrrega de passatgers a final de parada, segons els resultats que es descriuen a la Secció 5.5. Els temps a l'inici de parada i de descàrrega s'afegeixen als càlculs a partir dels paràmetres definits (subsecció 8.3.1)

7.3. Resultats

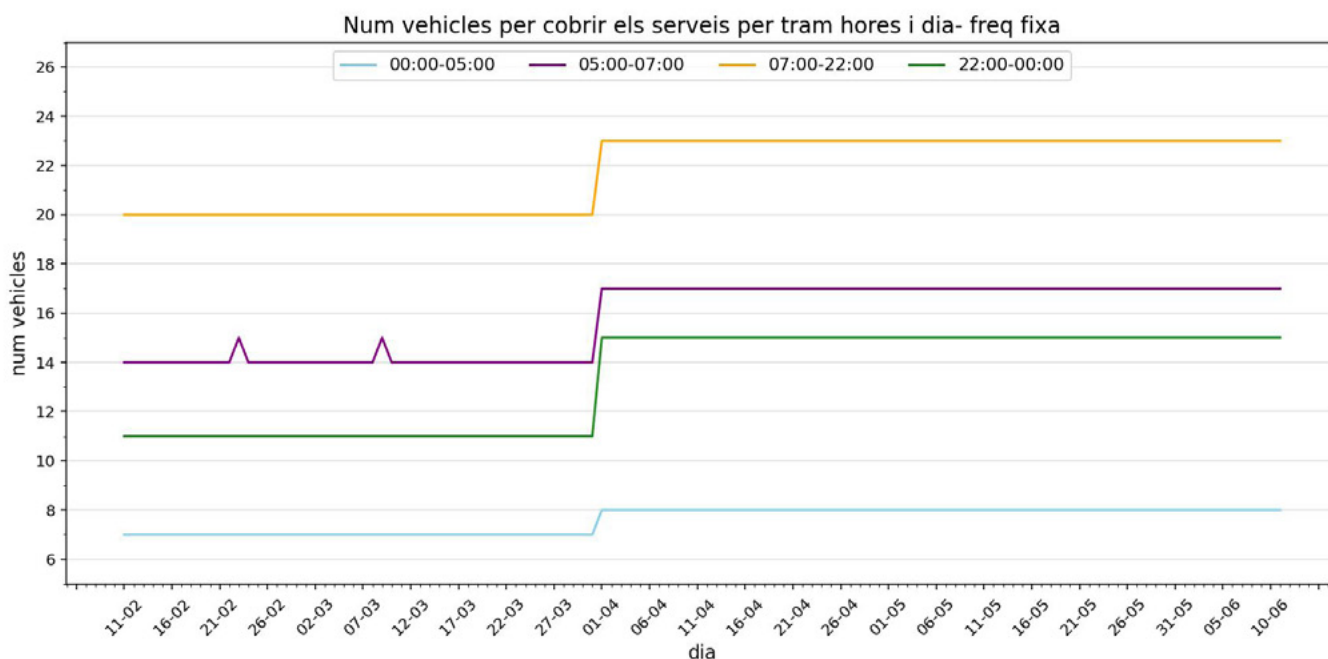
S'han executat dues simulacions, depenent de si es fa servir el fitxer de demanda original o el de la demanda incrementada.

7.3.1. Nombre de vehicles necessaris

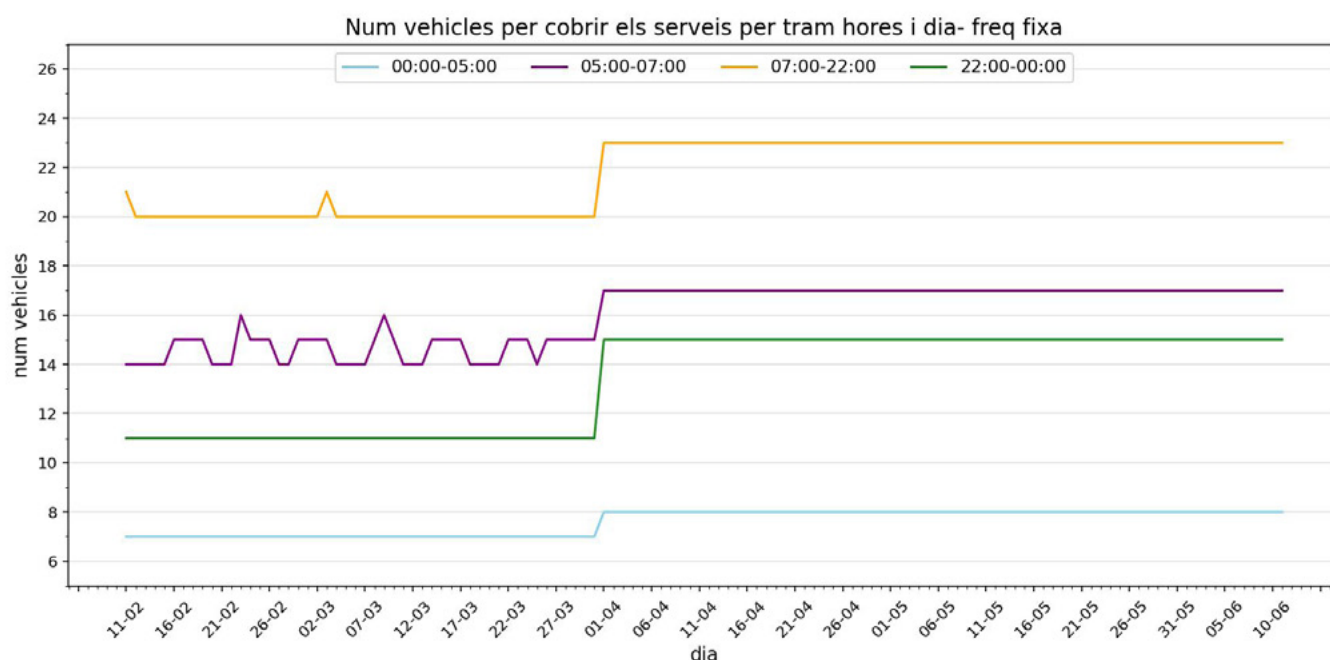
Les següents dues gràfiques mostren el nombre mínim de vehicles necessaris per cobrir la demanda, la primera gràfica pel cas de la demanda original, i la segona amb la demanda incrementada. En general, el total de vehicles és estable al llarg del temps en tres dels quatre trams horaris. Això és un primer indicador de que la freqüència actual permet cobrir la demanda.

En el cas del tram horari 5:00-7:00, amb la demanda original s'intueix que hi ha algun moment en que cal reforçar la flota, en els pics que apareixen al voltant del 21-02 i del 07-03. Això queda més clar en el cas de la demanda incrementada: la necessitat d'incrementar la flota apareix al llarg de tots els dies de la temporada baixa.

A. Demanda original



B. Demanda incrementada



7.3.2. Freqüència de pas

El resultat anterior porta a analitzar què passa amb la freqüència de pas. Es mostren a continuació alguns exemples que permeten comparar la freqüència fixada amb la mínima necessària per cobrir la demanda. Es genera un gràfic per cada hora i depenent si es fa servir la demanda original o la demanda incrementada. Aquí només es mostren els que són més significatius. L'eix horitzontal correspon als dies pels quals s'ha calculat la previsió de demanda, i l'eix vertical representa el nombre de trajectes, es a dir, la freqüència mesurada en nombre d'autobusos. Tots mostren el mateix tipus d'informació: per les dues línies, A1 i A2, en línia contínua es representa la freqüència real i en línia discontinua, la teòrica. Quan només apareix la línia contínua, significa que tant la freqüència teòrica com la real coincideixen.

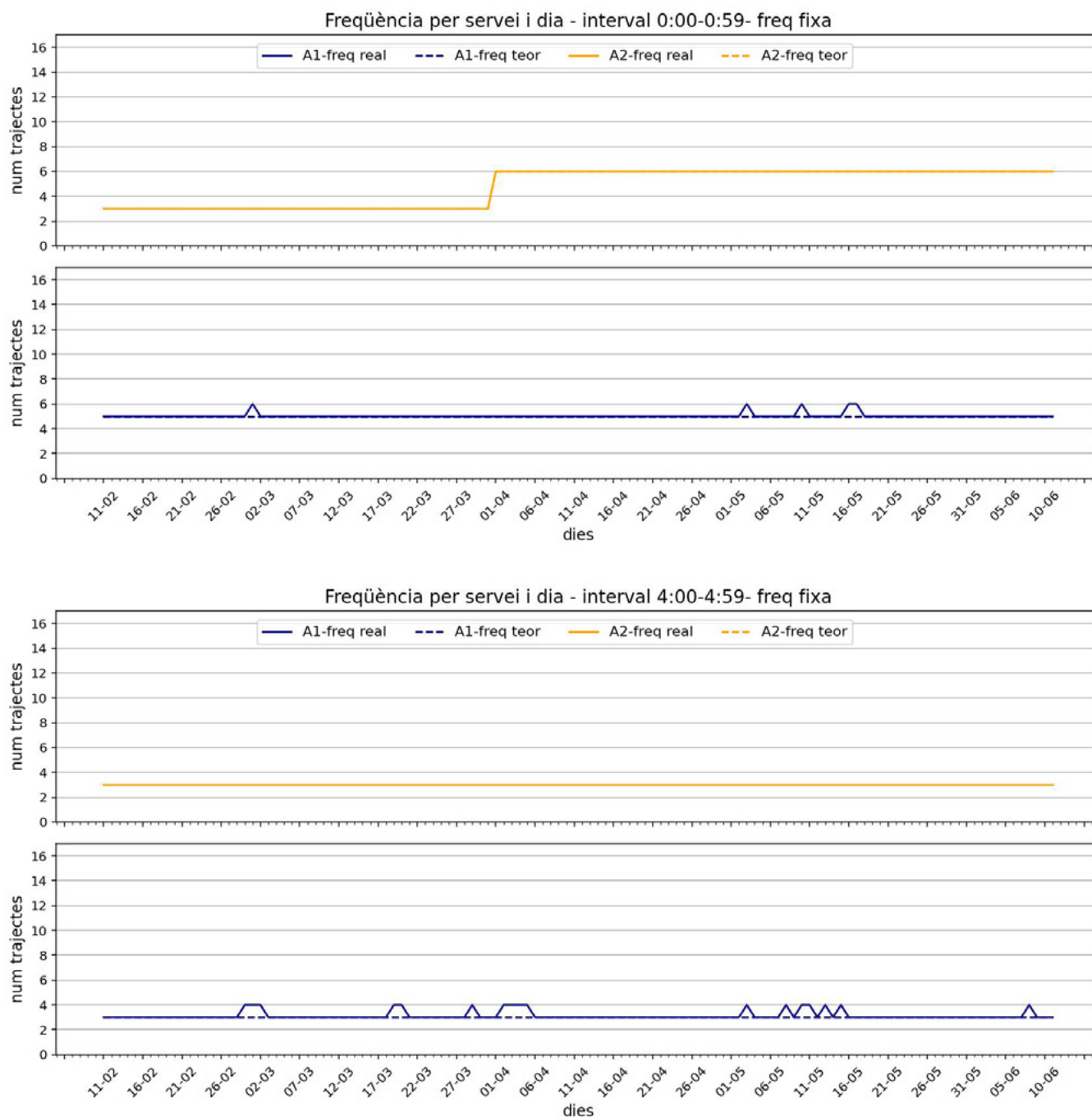
Cal fer notar que en el període analitzat hi ha dates que corresponen a la temporada alta i dates que corresponen a la temporada baixa. La diferència de freqüència de pas en les dues temporades explica el salt que hi ha als gràfics.

El següents quatre gràfics corresponen als intervals d'hores 0:00-1:00, 4:00-5:00, 5:00-6:00 i 6:00-7:00. La resta de gràfics que no s'incorporen a l'informe mostren un resultat similar al gràfic de 6:00-7:00: les freqüències teòriques i reals per totes dues línies coincideixen per tots els dies.

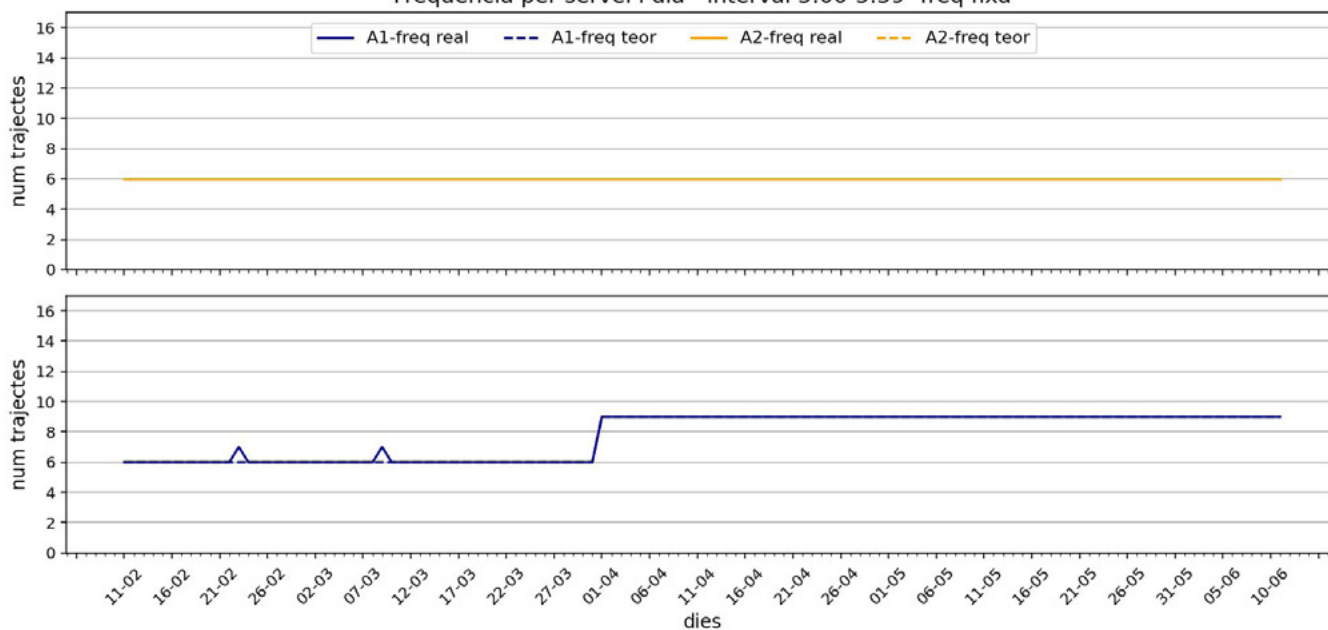
En el cas de la demanda original, en els altres tres exemples (0:00-1:00, 4:00-5:00 i 5:00-6:00), apareixen alguns pics en la línia A1. Això significa que, pels dies corresponents als pics, la freqüència teòrica de la línia A1 no és suficient per cobrir la demanda i cal afegir algun trajecte extra per servir tota la demanda prevista. On això es dona de forma més fre-

qüent és en el tram horari entre les 4:00 i les 5:00, la qual cosa confirma el que ja s'havia detectat en l'anàlisi de les dades del Capítol 5.

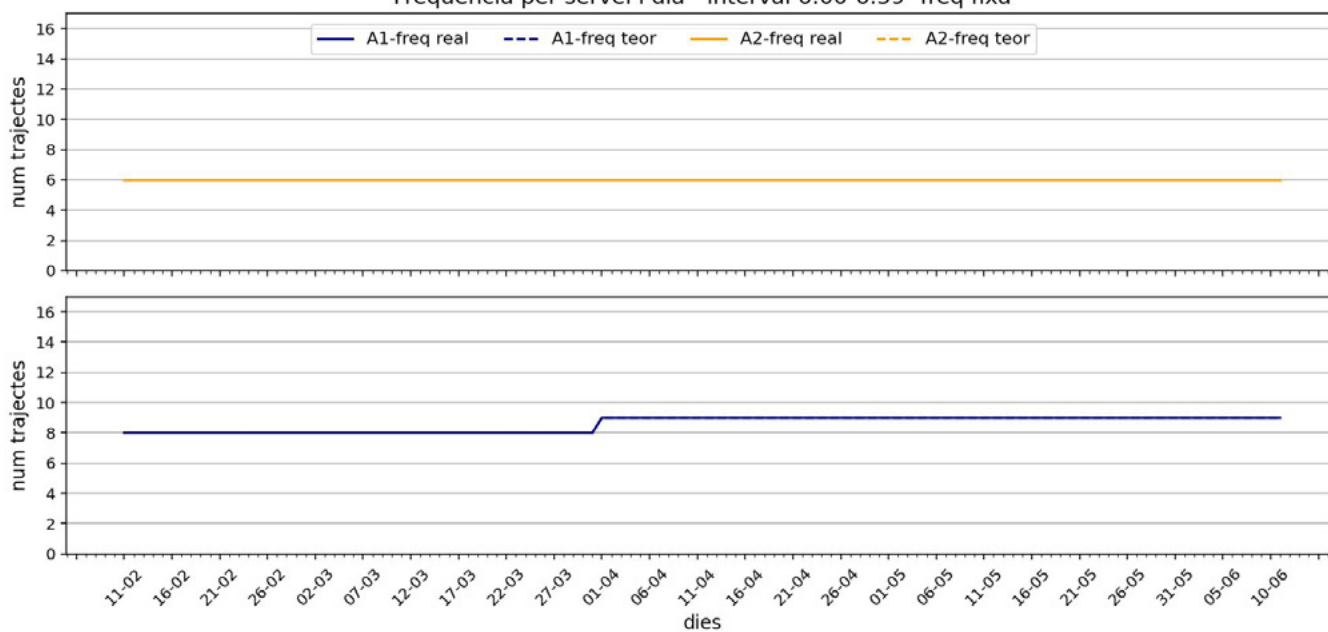
A. Demanda original

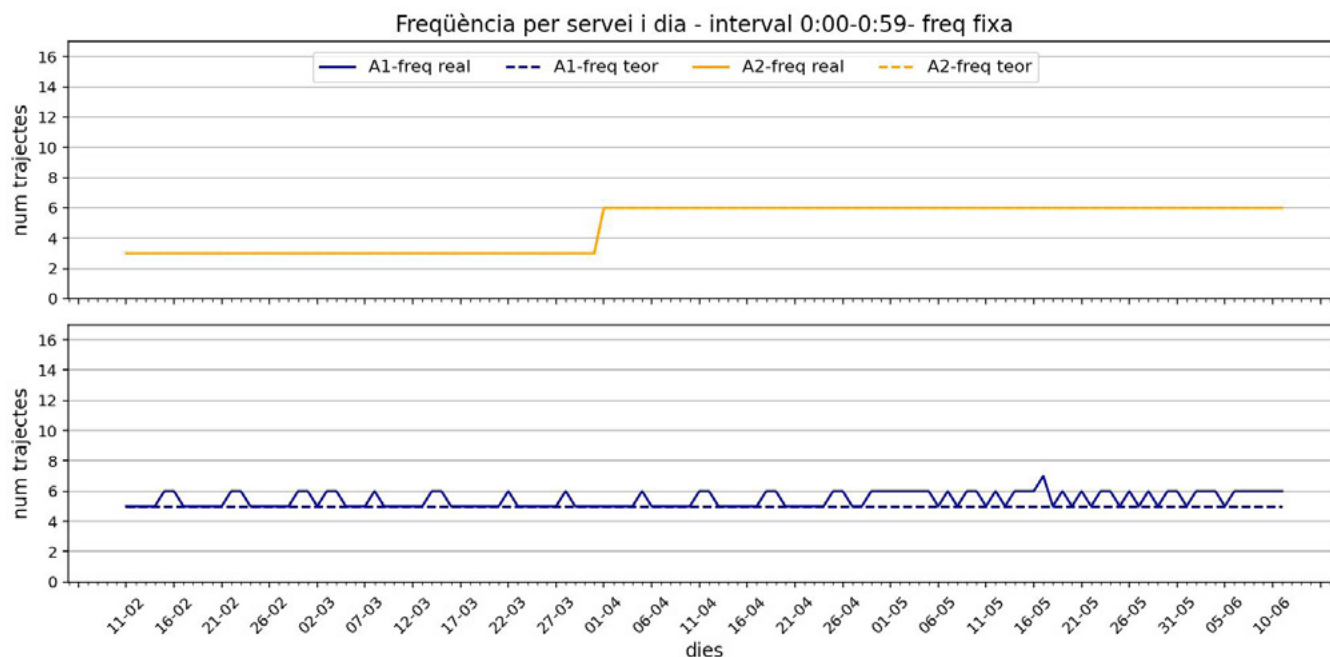


Freqüència per servei i dia - interval 5:00-5:59- freq fixa



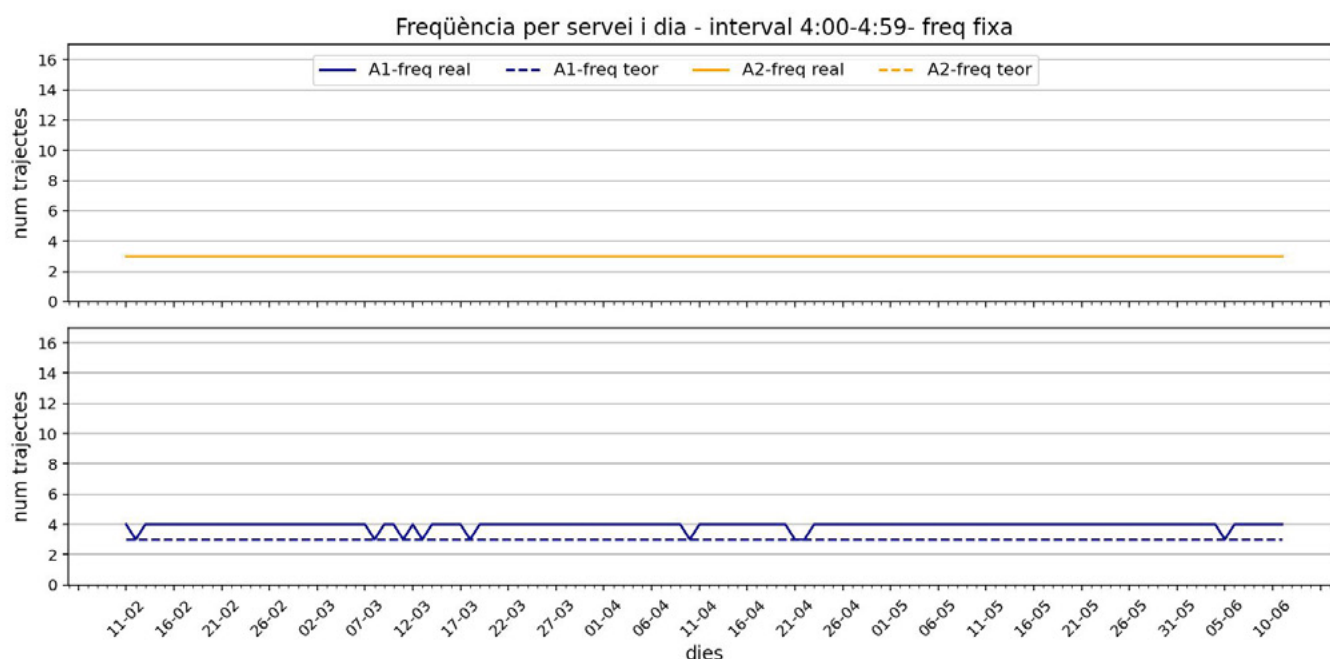
Freqüència per servei i dia - interval 6:00-6:59- freq fixa

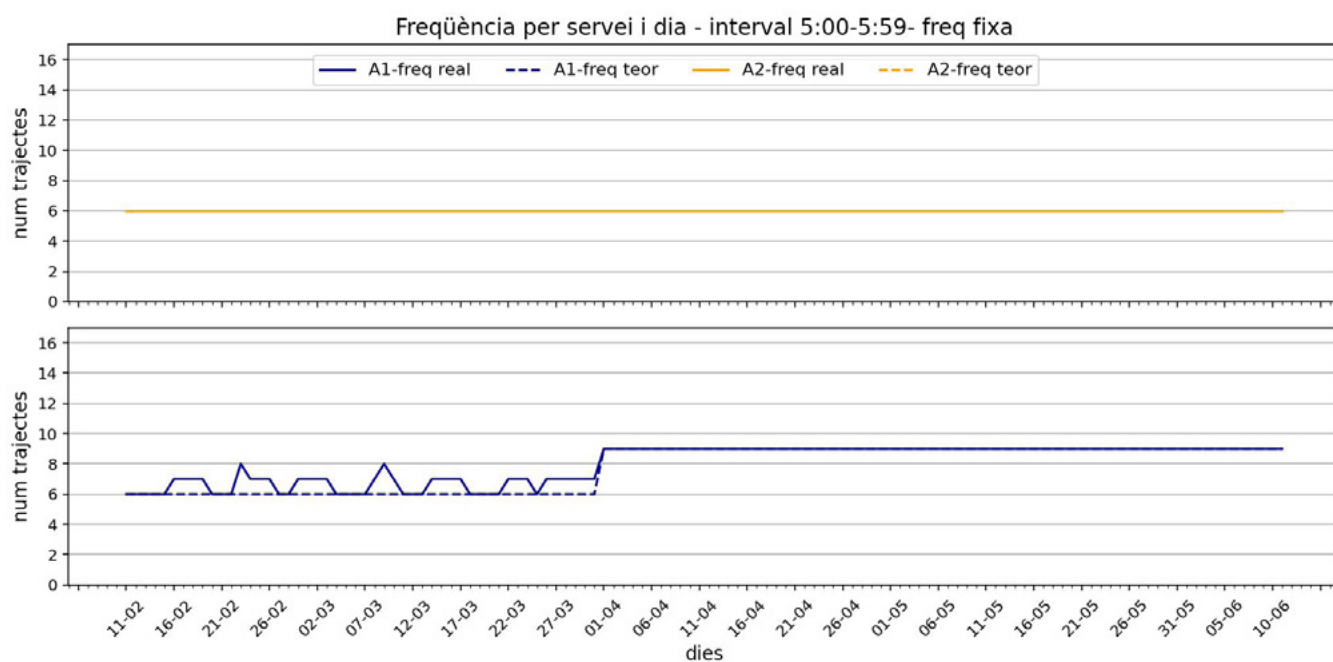




B. Demanda incrementada

En el cas de la demanda incrementada, la necessitat d'incrementar la freqüència de l'A1 en els horaris 0:00-1:00 i 4:00-5:00 es mostra de forma més clara, doncs pràcticament tots els dies la freqüència real està per sobre de la freqüència teòrica. També els resultats indiquen que caldria reforçar la freqüència de pas per la línia A1 entre les 5:00-6:00, al menys durant la temporada baixa. Respecte l'A2, amb la demanda incrementada es manté la freqüència teòrica al mateix nivell que la real, la qual cosa indica que la freqüència actual permet servir sense problemes la demanda incrementada.





Conclusions



L'objectiu de l'estudi que es presenta en aquest informe ha sigut desenvolupar una prova de concepte per validar la utilitat d'un model combinat de predicció de la demanda i optimització de la flota d'autobusos per ajustar el servei a mig-curt termini. En aquest estudi s'ha fet en primer lloc un treball d'anàlisi de dades històriques d'ús dels serveis de transport de l'Aerobús, confrontant aquestes dades amb d'altres candidates a ajudar a explicar la demanda i així poder fer una previsió de demanda futura.

En la primera fase d'anàlisi de les dades històriques, el treball fet ha permès verificar i determinar uns patrons de comportament que per si sols ja faciliten als gestors una bona fotografia del servei, de l'ús que s'està fent i de les necessitats actuals. S'ha contrastat aquest ús amb l'oferta del servei, mitjançant les dades de freqüència de pas, i el nombre real de trajectes per cada hora, reforçant-ho amb dades mostrals de les cues a les parades inicials dels trajectes.

La segona part de l'estudi ha consistit en desenvolupar el model de previsió de la demanda. De les variables inicials proposades per complementar les dades històriques per fer la predicció, les que han mostrat ser rellevants pels casos analitzats han estat la quantitat d'esdeveniments i el seu aforament, i els vols previstos, a més a més del dia de la setmana i el mes de la data a preveure. Les variables descartades són la pluja, el fet que el metro estigui operatiu o no, i que un dia sigui festiu. A més, s'ha dut a terme un anàlisi d'autocorrelació de les dades per identificar la influència de la demanda passada sobre la demanda futura. Aquest anàlisi ha revelat l'existència de patrons diaris molt repetitius, la qual cosa ha portat a la incorporació de les dades de demanda històrica com a variable clau predictora.

S'han provat diferents algorismes, en diferents horitzons de predicció, incloent-hi finestres de 2 dies, 1 setmana, 2 setmanes i 4 setmanes, per finalment trobar el que millor es comporta amb les dades analitzades. Entre d'altres, s'han testejat models de tipus lineal, arbre de decisions i xarxes neuronals, essent l'algoritme de tipus arbre de decisió Gradient Boosting Regresor el que millors resultats ha donat, amb una R^2 entre 0,67 i 0,82, depenent dels casos.

Per últim, la darrera part de l'estudi s'ha dedicat a desenvolupar un model d'optimització que permeti fer una estimació del mínim nombre de vehicles que són necessaris per cobrir la demanda, tenint en compte la freqüència de pas establerta. A partir de l'exemple executat, els resultats del model han mostrat que la freqüència de pas definida dona servei a la demanda en gairebé tots els dies i totes les hores, però que hi ha algunes franges horàries en les que caldria reforçar l'A1. Per complementar els resultats obtinguts, s'ha desenvolupat una segona versió del model en la que la freqüència de pas no és una condició estricta, si no que el propi model suggereix quina podria ser la freqüència si només s'ajusta a la demanda.

La neteja de dades ha portat a descartar trajectes d'autobús i les seves validacions per no estar correctament identificat l'autobús en el sistema SAE: durant el desenvolupament dels models es va fer la assumpció que eren trajectes no realitzats. És un volum significatiu, entre un 14,6% pel servei de l'A1 amb origen l'aeroport, i un 20,9% pel servei de l'A2 amb origen plaça Catalunya. Ignorar aquest registres no afecta a la metodologia que s'ha seguit i que és la part més important de la feina feta però sí pot tenir impacte en els resultats de la previsió

de la demanda. Per això, per fer les proves del model d'optimització s'han fet noves simulacions incrementant la demanda de manera proporcional als percentatges de dades faltant detectades.

Com a resum final, el treball realitzat mostra que amb unes dades de qualitat l'anàlisi de la informació històrica és una eina molt potent per ajudar als gestors dels serveis en les seves tasques de decisió. A més a més, aquestes dades històriques, complementades amb altres dades de qualitat amb les que tinguin una bona correlació i un nivell de significança correcte, permeten generar models de previsió de la demanda fiables. El suport addicional obtingut a partir del model d'optimització de la flota, que a més a més dona informació de la quantitat de serveis indispensables per assegurar la demanda prevista, confirma que la solució tecnològica proposada i validada en aquest estudi pot ser de molta ajuda als gestors dels serveis d'autobusos.



Bibliografia



Meteocat. (06 de 2024). *meteocat.cat - Servei Meteorològic de Catalunya*. Obtingut de <https://www.meteo.cat/observacions/xema/dades?codi=XL&dia=2024-12-11T15:00Z>

Observatori de Turisme a Barcelona. (08 de 2024). *Perspectives de l'activitat turística a la Destinació Barcelona*. Obtingut de <https://www.observatoriturisme.barcelona/ca/arxius>

Open-Meteo. (06 de 2024). *Open-Meteo*. Obtingut de <https://open-meteo.com/>

Openweather. (06 de 2024). *Openweather*. Obtingut de <https://openweathermap.org/>

Timber Bytes LLP. (09 de 2024). *AeroDataBox*. Obtingut de <https://aerodatabox.com/>

TMB. (06 de 2024). *Horaris*. Obtingut de <https://www.tmb.cat/ca/transport-barcelona/horari-metro-bus>

Promotors



Membres

