

Estimació d'Audiències



Generalitat de Catalunya
**Corporació Catalana
de Mitjans Audiovisuals**

Estimació d'Audiències per a la Corporació Catalana de Mitjans Audiovisuals



Karla Trejo

Responsable d'Innovació
Fundació i2CAT

01

Introducció

02

Objectius

03

Enfocament

04

Resultats

05

Conclusions

- La recerca de mercats és un aspecte fonamental per a poder prendre les millors decisions estratègiques. En l'àmbit de la comunicació i la publicitat, **conèixer qui és la teva audiència, quins són els seus interessos**, on es troba, les dades més rellevants dels diferents suports i mitjans, i els sectors que integren el mercat, és imprescindible.
- El mesurament d'audiències televisives es realitza a través d'**audímetres** connectats al televisor. Aquest audímetre registra de mode continu les persones veient la televisió (incloent convidats) i el programa que estan visualitzant.
- L'audímetre envia cada nit aquesta informació recol·lectada durant el dia al centre de dades de Kantar. Allà es processa aquesta informació i es posa a la disposició dels clients **a primera hora del matí de l'endemà**.



- La demora en el lliurament de dades d'audiència televisiva fins a l'endemà impacta a la Corporació Catalana de Mitjans Audiovisuals (CCMA, ara 3Cat) en **restringir la seva capacitat per a realitzar anàlisis immediates** i fonamentades en informació recent.
- 3Cat és l'ens públic que gestiona els mitjans de comunicació audiovisual de la Generalitat de Catalunya. Ho fa en compliment de la seva missió de **servei públic** i amb un ferm compromís de qualitat, independència i promoció de la cultura i la llengua catalanes.
- Dins de 3Cat, TV3 destaca com el seu principal canal de televisió de cobertura autonòmica, sent el **mitjà de comunicació de referència** a Catalunya per a la gran majoria dels seus ciutadans.



OBJECTIU: Utilitzar dades històriques d'audiències i dispositius connectats de TDT (televisors) del canal de TV3 per entrenar un model d'Intel·ligència Artificial que demostrï que es poden fer estimacions d'audiències en temps real.

IMPACTE: Poder adaptar els continguts en funció de l'audiència estimada en cada moment per oferir els que siguin més interessants per als espectadors o per 3Cat.



- El *dataset* de MATR conté les dades en temps real de totes les SmartTV connectades, amb el *timestamp* i el nombre de dispositius connectats.
- El *dataset* de Kantar proporciona tres tipus de dades diferents: *aud_mitj_milers*, *aud_mitj_percent* i *quota*. La primera variable representa el total de persones que observen la televisió en milers, la segona representa el percentatge de població de Catalunya que observa TV3 i la quota representa el percentatge de persones que observen TV3 sobre el total de persones que miren la televisió.
- Els *datasets* van de l'1 de juliol de 2023 al 08 de febrer de 2024 (8 mesos).
- La dada principal a predir és l'audiència. No obstant això, durant l'anàlisi exploratòria de les dades, identifiquem unes certes propietats o *característiques rellevants que impacten a la sèrie temporal*.
- En transformar aquestes característiques en *metadades de tipus booleà*, augmenta la flexibilitat i el poder explicatiu del model.



- Predictor booleà: **Estacionalitat**
 - Diferència entre els mesos d'estiu (juny-agost) i els mesos d'hivern (desembre-febrer)
 - “Temporada d'estiu o hivern?”
- Predictor booleà: **Dia de la setmana**
 - Distinció entre setmana laboral o cap de setmana
 - “És un dia laboral o festiu?”
- Predictor booleà: **Hora del dia**
 - L'error es concentra entre les 5-8 hores i les 13-17 hores
 - “Matinada o migdia?”
- Predictor booleà: **Tipus de Programa**
 - Meta-grup amb més alt percentatge
 - Els programes de notícies constitueixen un 44% del total
 - “El programa és o no un noticiari?”



- La **regressió lineal múltiple** (per les seves sigles en anglès, MLR) és una tècnica estadística que utilitza diverses variables explicatives per predir el resultat d'una variable de resposta.
- L'objectiu de la regressió lineal múltiple és modelar la relació lineal entre les variables explicatives (independents) i les variables de resposta (dependents).
- La regressió calcula un polinomi que pot emprar-se per a avaluar l'efecte de cadascuna de les variables.

Valor Kantar
(audiència)

Predicted value

$$y = \alpha + \beta_1(x_1) + \beta_2(x_2) + \dots + \beta_n(x_n)$$

Bias

Weight 1

Feature 1

Weight 2

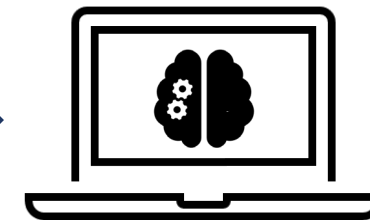
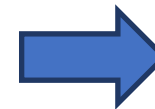
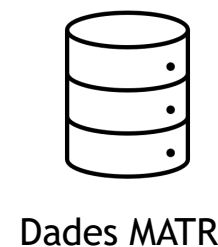
Feature 2

Weight n

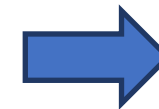
Feature n

Valors de n-
característiques
MATR

Coeficients apresos pel model



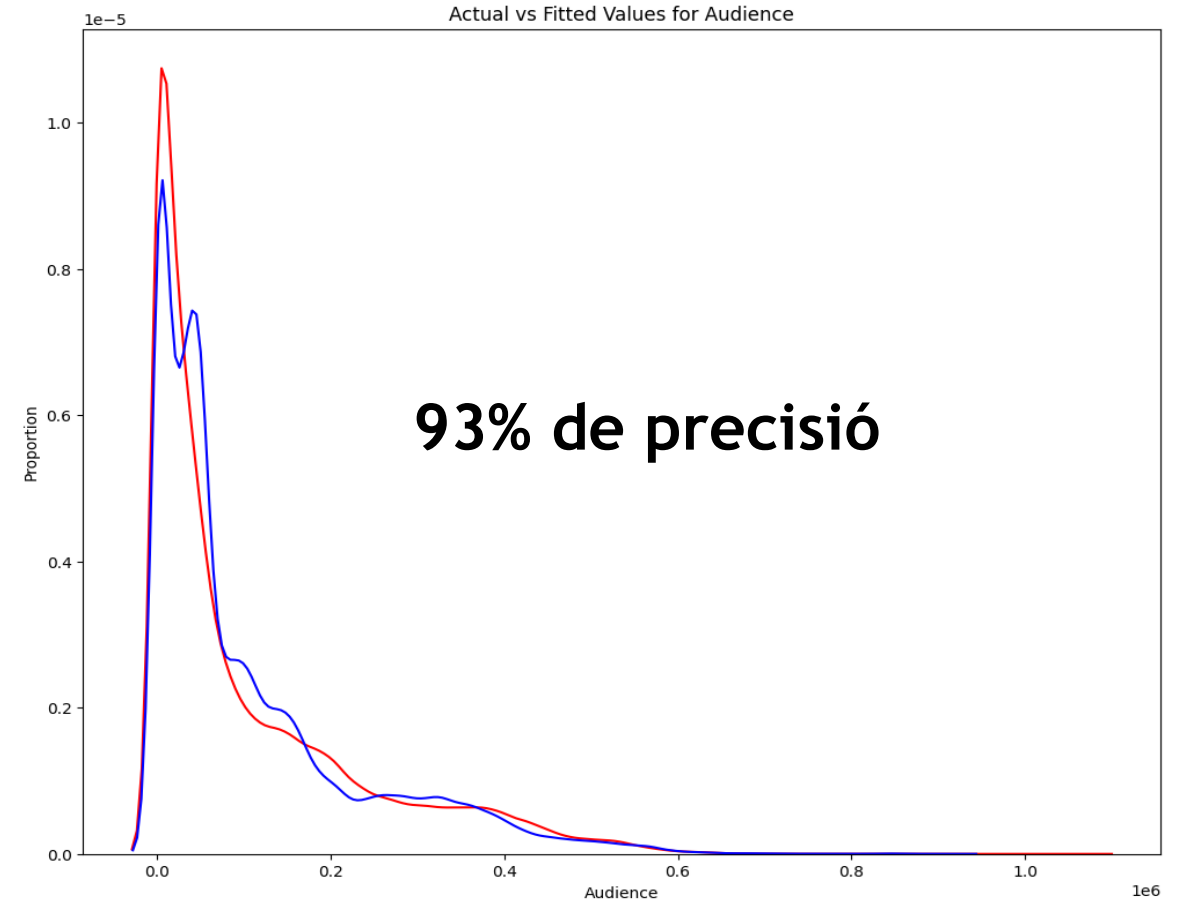
Model MLR



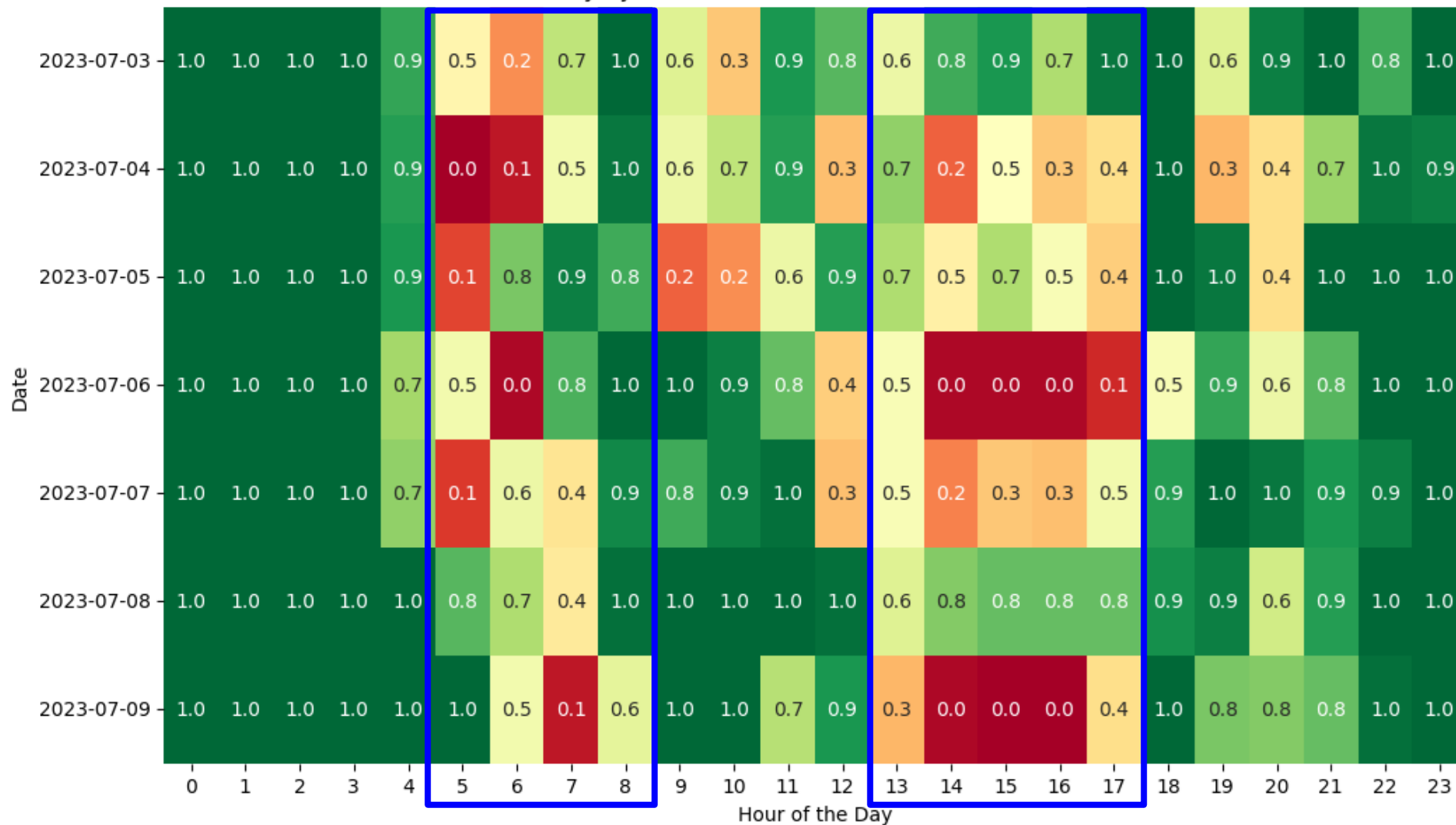
Predicció
d'audiència

- **Mean Absolute Error (MAE)**
 - Calcula la diferència absoluta entre els valors real i predit
 - Robust a valors atípics
- **Root Mean Squared Error (RMSE)**
 - Calcula què tan distants estan les observacions de la línia de regressió
 - Penalitza errors grans
- **Coeficient de determinació (R^2 score)**
 - Mesura la força de la relació entre el model i la variable dependent
 - Com més proper a 1, millor és el model a l'hora d'explicar la variabilitat en les dades

MAE	21,910.40
RMSE	33,507.00
R^2 Score	93.30%

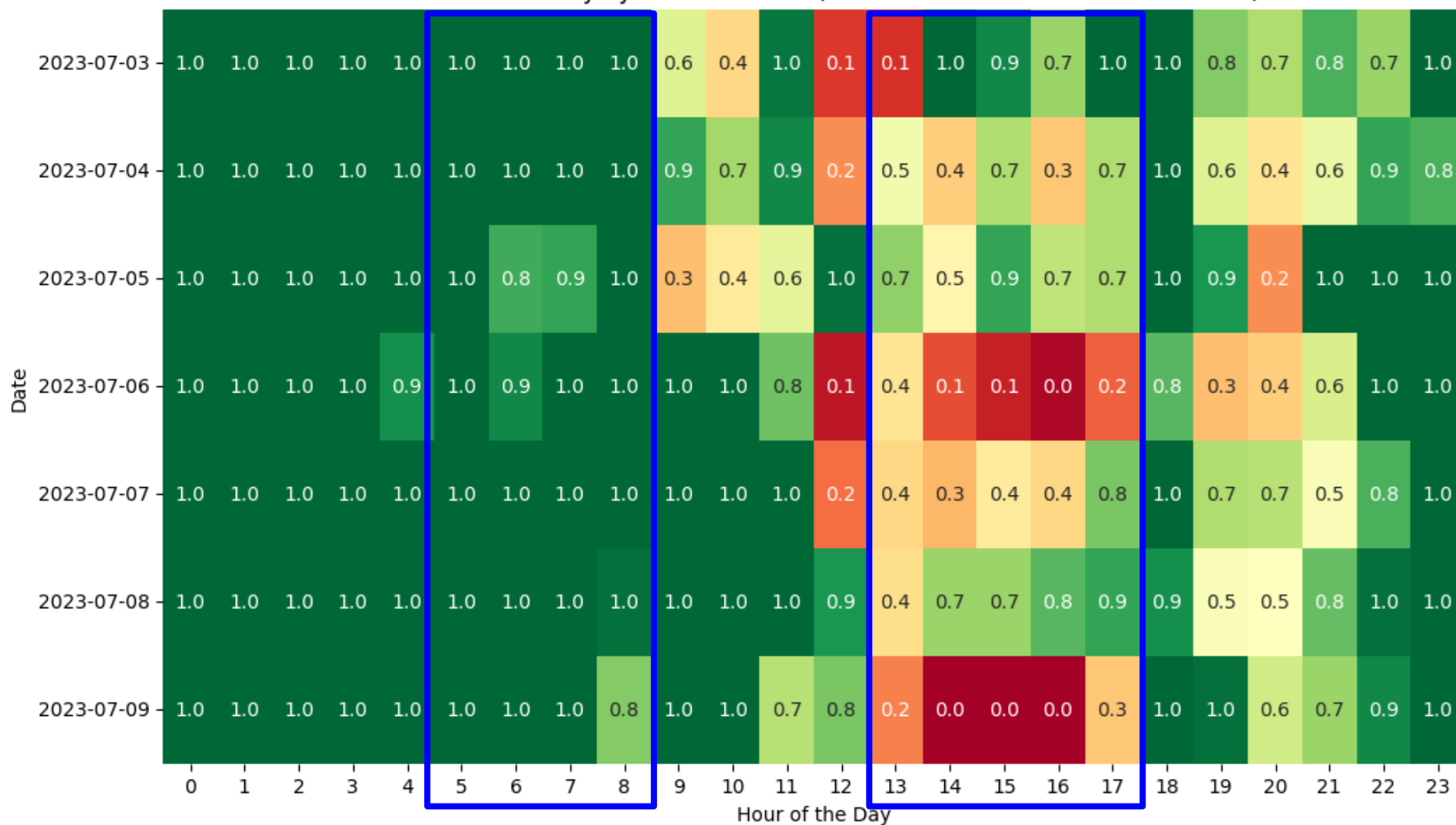


Prediction Accuracy by Date and Hour (Threshold Condition Met - One Week)

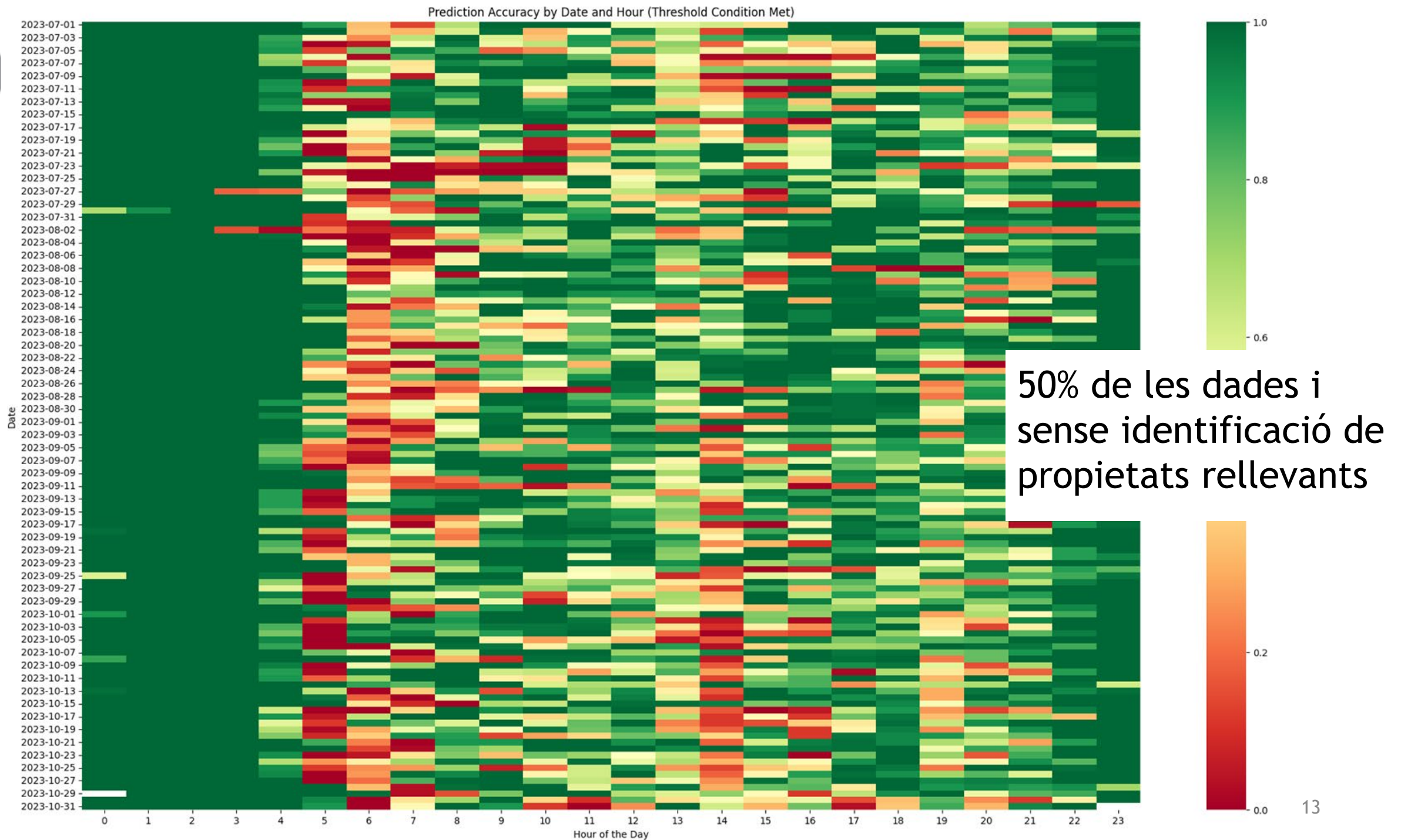


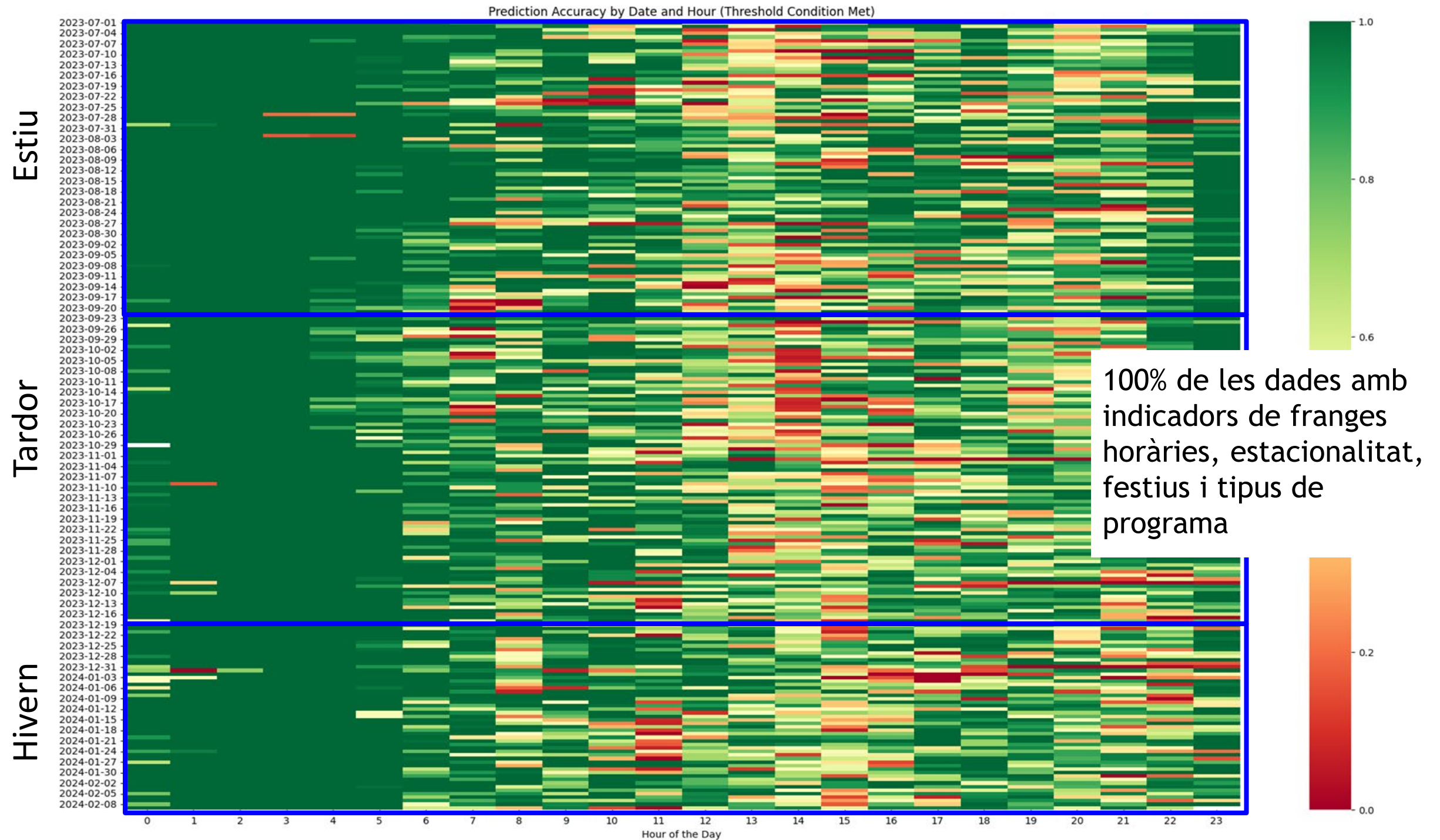
Abans d'afegir
propietats

Prediction Accuracy by Date and Hour (Threshold Condition Met - One Week)

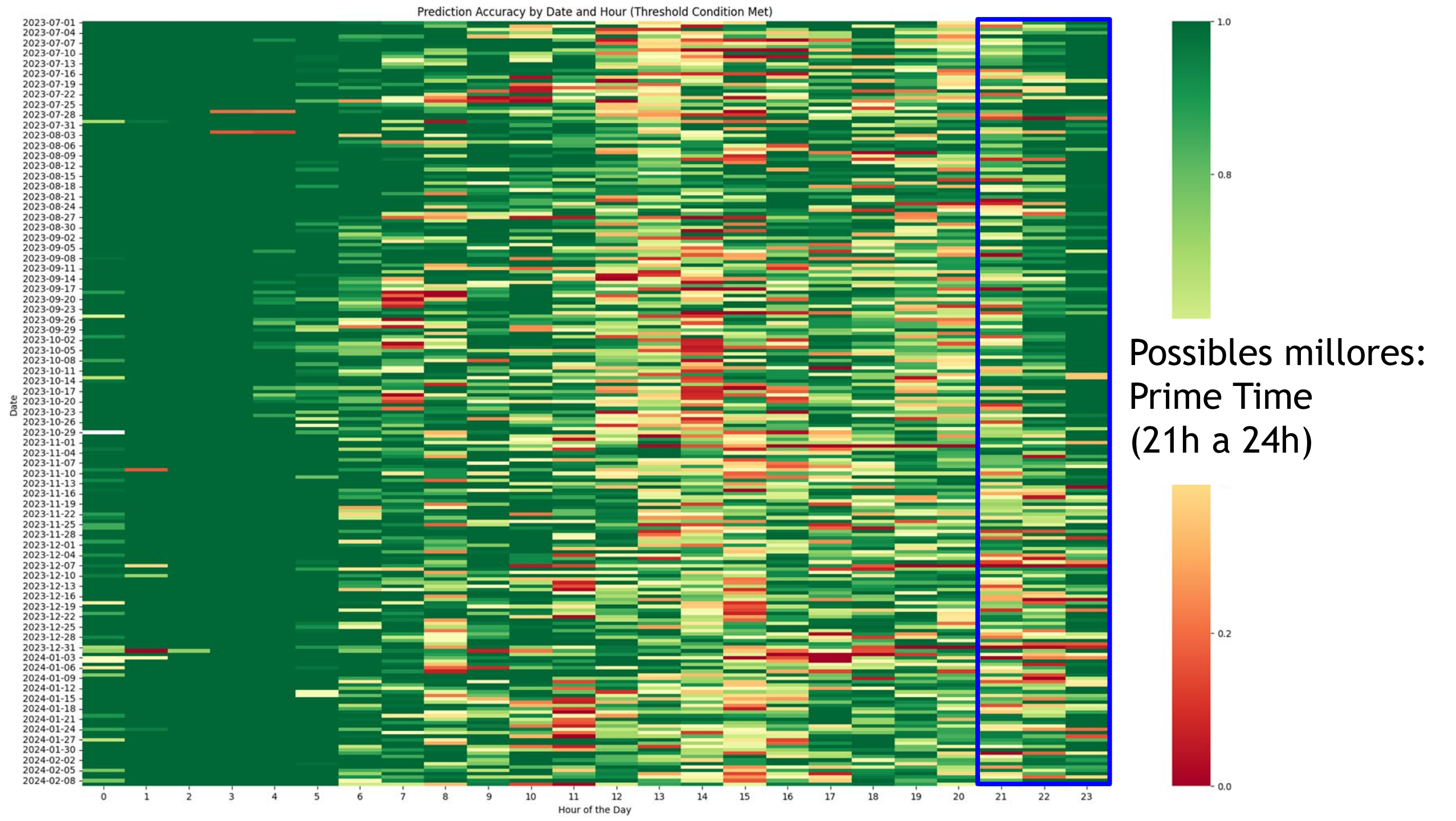


Després d'afegir propietats





- Ampliar el *dataset* d'entrenament incorporant un **conjunt més ampli de dades per a abastar un marc temporal anual complet**, la qual cosa permetria al model aprendre patrons més complets i robustos. Actualment, el *dataset* cobreix únicament 8 mesos i no inclou els mesos de primavera ni els festius associats a aquestes dates.
- Afinar l'anàlisi de la **franja horària del "prime time"**, la qual abasta generalment les hores entre les 21:00h i les 24:00h, considerant la influència de diversos factors interrelacionats. Això implica ajustar el model per a captar amb major exactitud les variacions i comportaments específics durant aquest període crucial de visualització televisiva.



Conclusions

- El model presenta un **R² superior al 90%**, amb millores significatives en zones amb més variabilitat i error inicial.
- L'anàlisi exploratòria ha revelat que **l'hora del dia i el dia de la setmana** són factors clau en les audiències.
- Els mapes de calor entre valors reals i prediccions han aportat **claredat i intuïció** sobre el rendiment del model.
- Els resultats es generen en un **format compatible** amb l'actual, evitant conversions addicionals i facilitant l'ús.
- La integració via Docker+API amb els sistemes de 3CAT ha permès una **visualització directa i eficient** de les estimacions.
- El model permet als especialistes enfocar les seves anàlisis i explotar les dades d'acord amb els seus interessos específics, la qual cosa **millora significativament la utilitat pràctica** de la informació.
- La versatilitat per a l'usuari final s'ha ampliat considerablement, brindant eines més adaptables i útils per a la **presa de decisions**.

Moltes gràcies!

Karla Trejo

Responsable d'Innovació

karla.trejo@i2cat.net