

Curso de Introducción a la Ciencia de Datos y Machine Learning

CAPSTONE PROJECT

Mateo Bermann
Gerard Buxó
Olga Llima
Carme Monserrat
Jordi Toset



KAGGLE'S TEAM NICKNAME: GEMJOC

ÍNDICE



1.- CLEANING + DEA





AT A GLANCE



7.000
Bicicletas

La flota se ha actualizado,
3.000 bicicletas ya son
eléctricas.



519
Estaciones

Para la predicción se han
considerado las que están en
activo para todos los ejercicios



137 mil usuarios
y 2 APP's

ANÁLISIS Y EXPLORACIÓN DE DATOS

Origen datos

 Open Data **BCN**

Servicio de datos abiertos del Ayuntamiento de Barcelona

- Ficheros mensuales información estaciones.
- Ficheros mensuales de estado de las estaciones.

Extracción



Carga y análisis



Entorno cloud

Google
colab

jupyter

Notebook

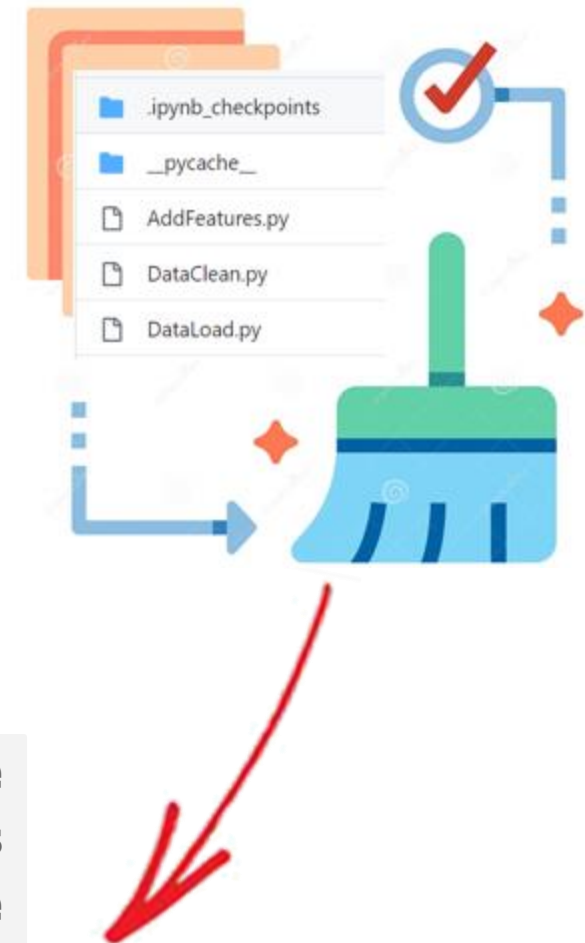


Esquema extracción, carga y análisis de datos

ANÁLISIS Y EXPLORACIÓN DE DATOS

Estudio de las
variables,
limpieza de filas,
etc..

Ejecución de procesos de limpieza de los datos: conversión a valores numéricos, supresión/conversión de NAN, eliminación de variables, creación de variables, etc..



ANÁLISIS Y EXPLORACIÓN DE DATOS

Estructuración de
los datos y
construcción del
modelo



DataReestructure.py

DataTrain.py

Maps.py

MeteoBCN.py

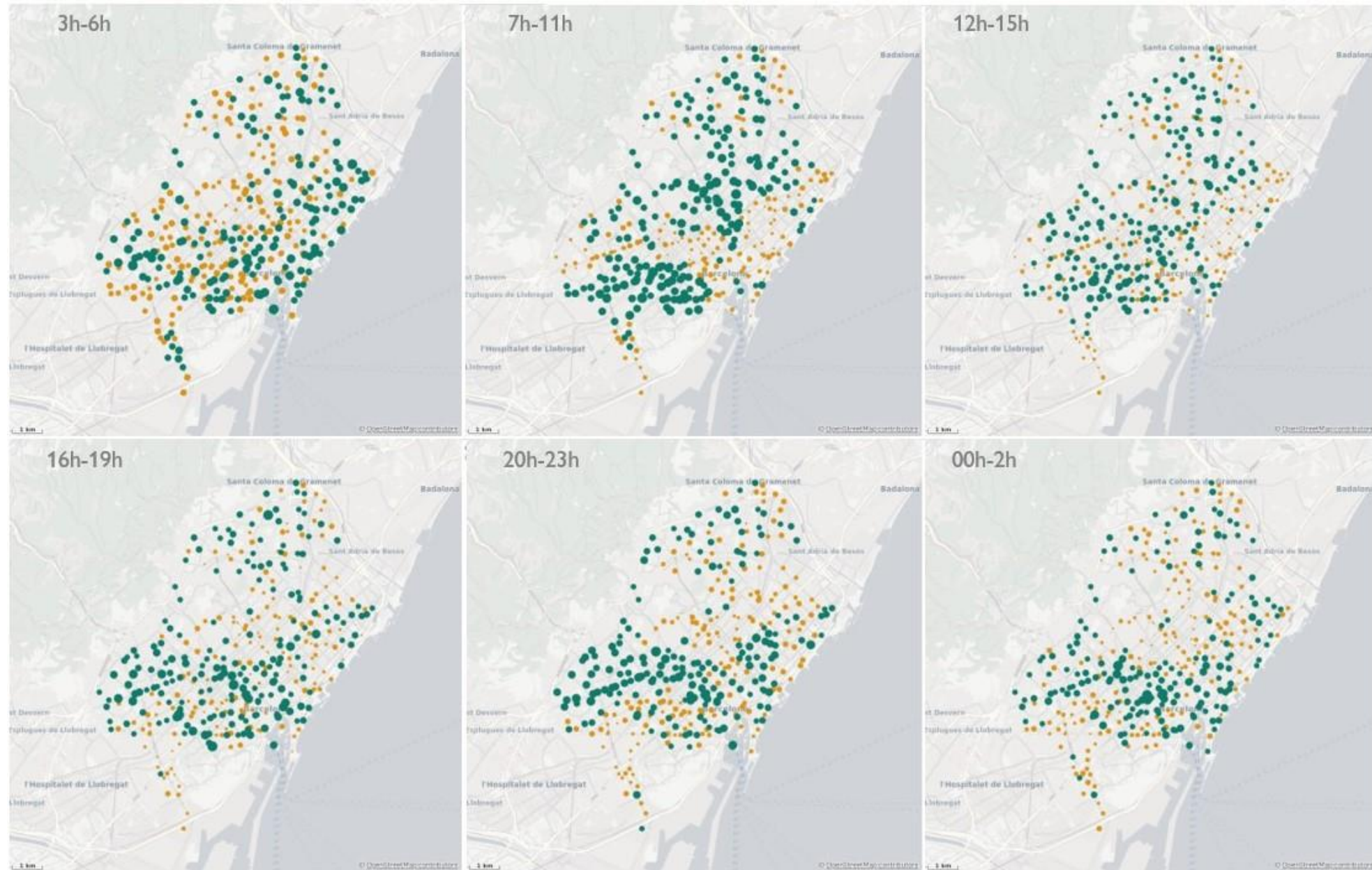
ModelBuild.py

Readme.md

SplitData.py

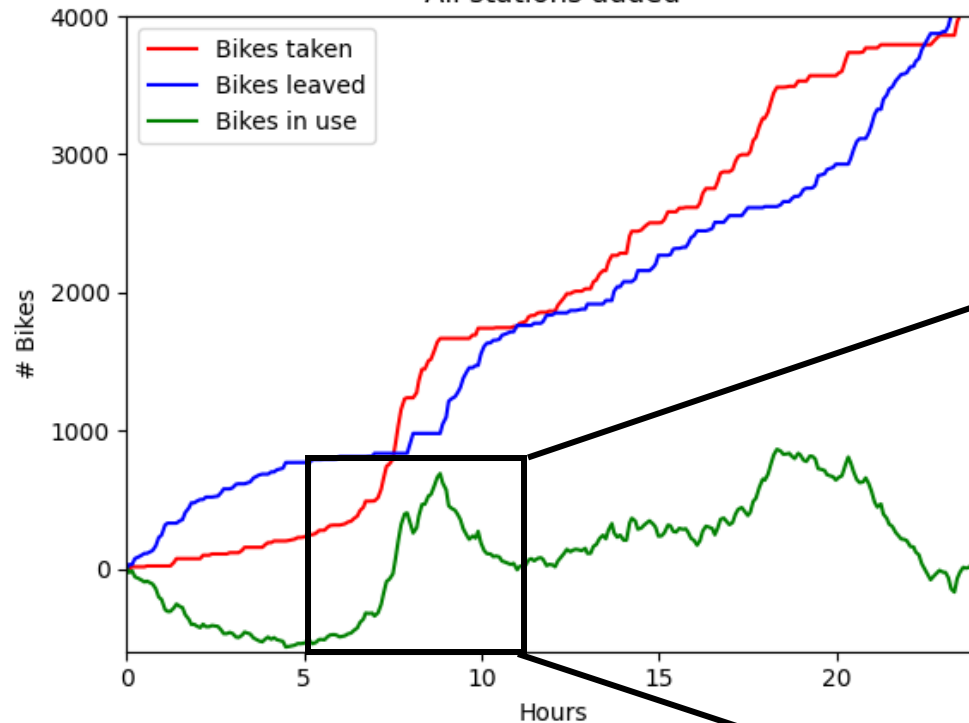


Ajustes de los datos y
establecimiento de los modelos a
entrenar.

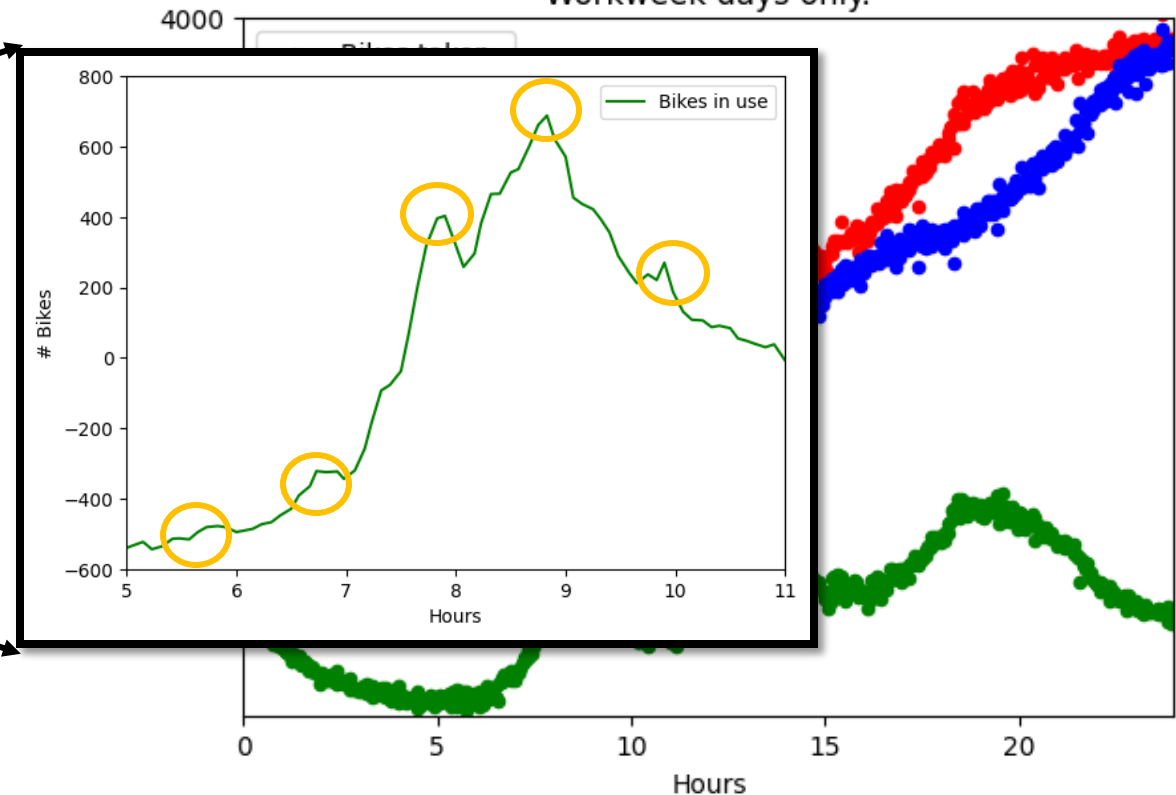


Movimientos acumulativos de un día laborable.

Accumulative bikes moviments in one day (01/06/2022)
All stations added



Average days for all the 06/2022 month.
All stations added.
Workweek days only.



1.239.482

Número de usos mensuales
bicing



16.298.596

Número de usos acumulados
año 2022



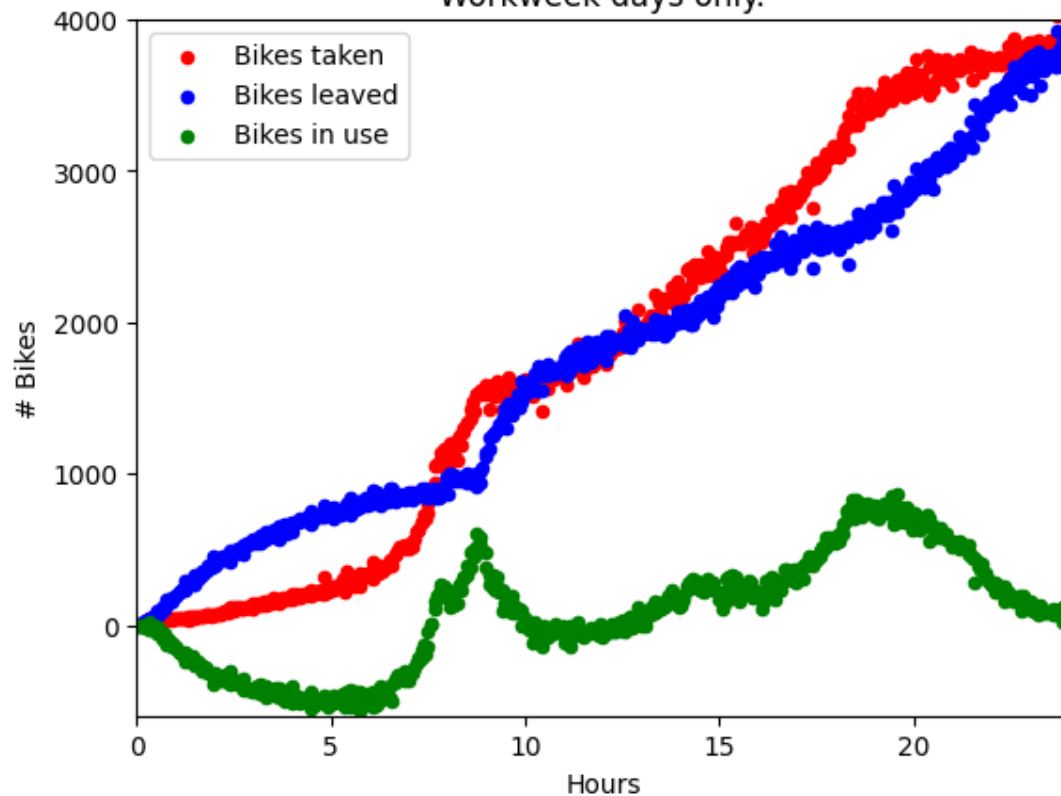
13.5 min

Tiempo medio de viaje con
bicicleta

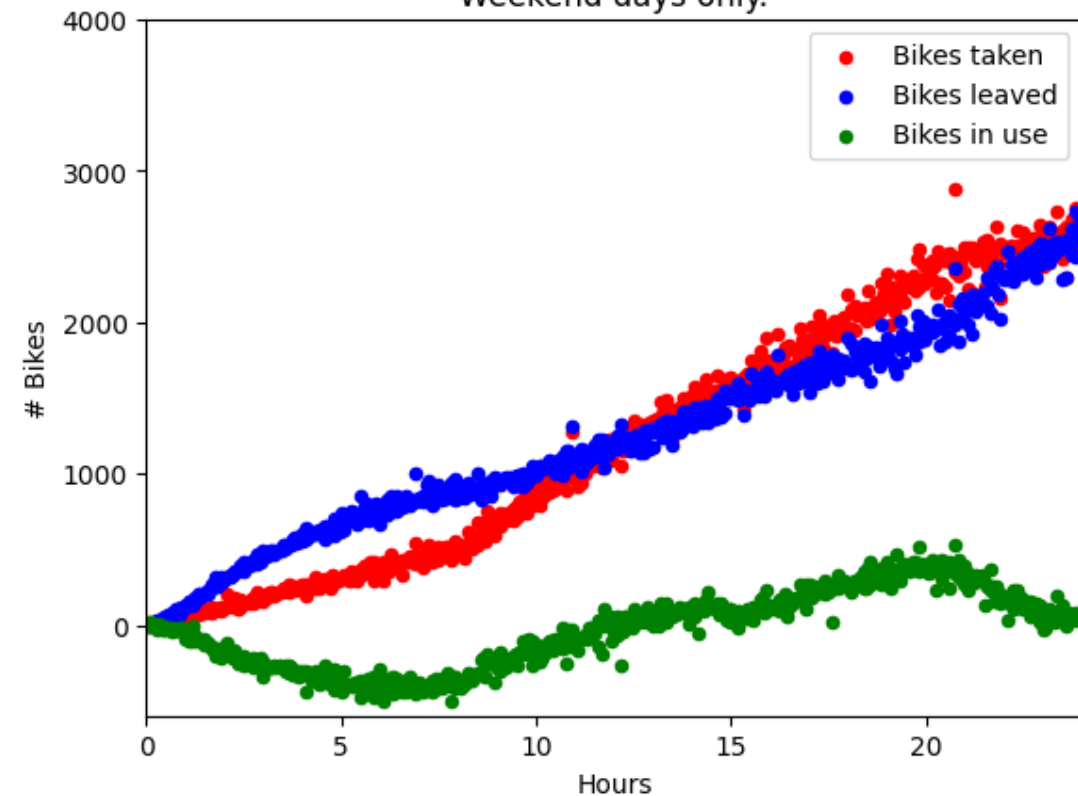
Mostra menys d'un 10% del total

Comparativa de dias laborable vs. festivos para el 06/2022.

Average days for all the 06/2022 month.
All stations added.
Workweek days only.



Average days for all the 06/2022 month.
All stations added.
Weekend days only.





2.- FEATURES

Más features ?

Meteo

humedad relativa
precipitación
presión atmosférica
temperatura
velocidad viento
dirección viento
radiación solar

Calendario

weekend
holiday
peak hour
season

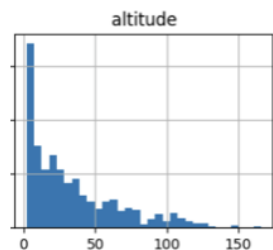
Metadatos

altitude
capacity
post-code
num_pt
dist_to_beach

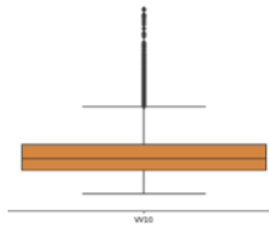
Zona

playas
p. catalunya
uni diagonal
alta
otros

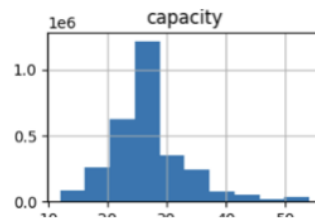
Transformaciones ?



`SimpleImputer(strategy="median")`
`FunctionTransformer(np.log1p)`
`StandardScaler`



`Replace_Outliers_2_NaN()`
`SimpleImputer(strategy="median")`
`StandardScaler`



`SimpleImputer(strategy="median")`
`StandardScaler`

station_id
post_code
year
month
hour

`transform_Num_to_Cat`
`SimpleImputer(strategy="constant", fill_value='Unknown')`
`OrdinalEncoder()`



`SimpleImputer(strategy="constant", fill_value='Unknown')`
`OneHotEncoder(handle_unknown='ignore')`

Estratificación ?

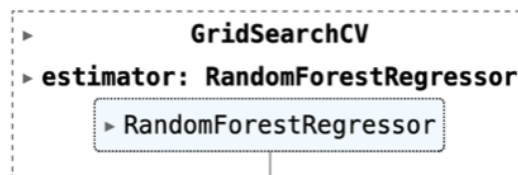
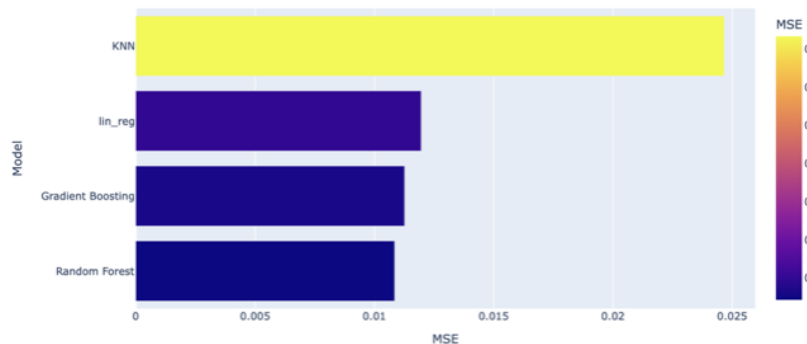


disponibilidad docks

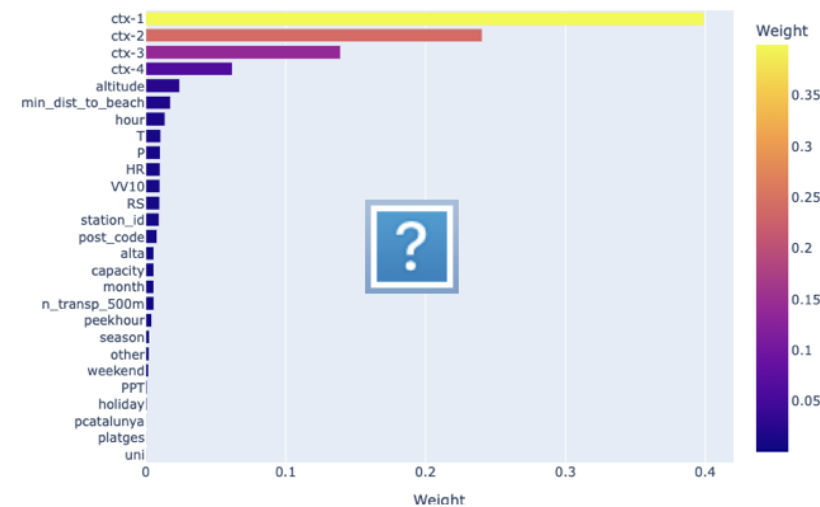
3.- MODEL



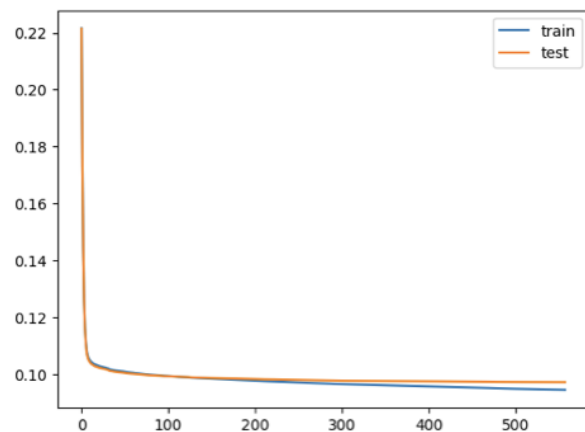
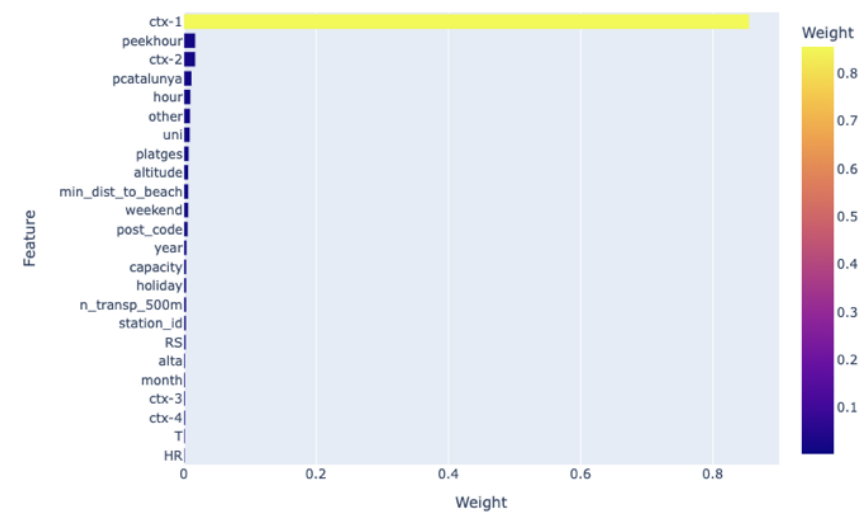
Models Comparison



XGBoost



Feature Importance



MSE validación:
0.009439

Intervalo de confianza
(0.009366, 0.009512)



UNIVERSITAT DE
BARCELONA



4.- Streamlit

■ c46d-34-03-143-3.mgrok-free.app



Remote Image

Bicing Prediction

ID_STATION

119

1

519

HOUR

4

1

24

DAY

3

1

31

MONTH

3

1

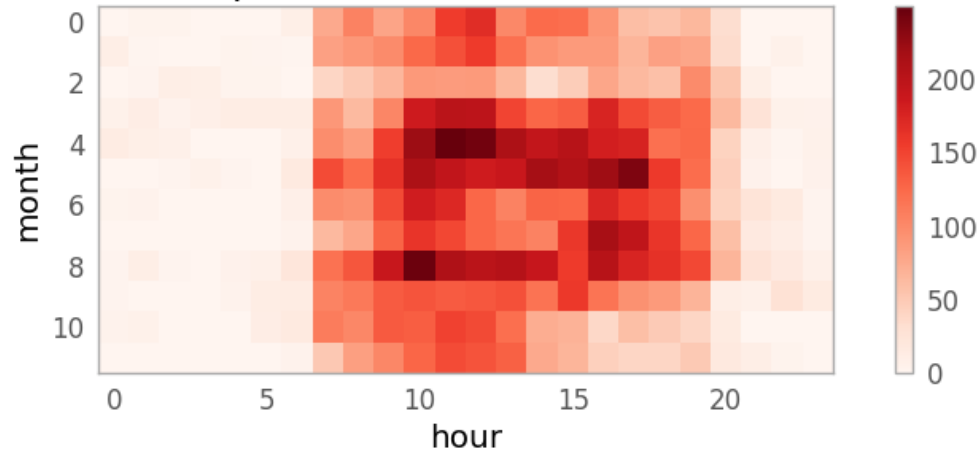
12

THE AVAILABILITY PREDICTION IS: 0.6324859261512756

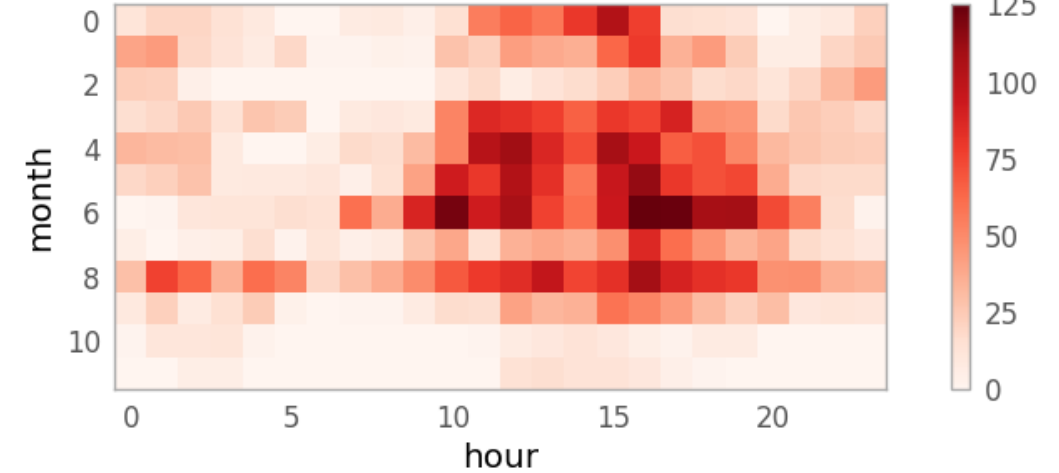


HEATMAPS FALTA DISPONIBILIDAD ESTACIONES EN LA PLAYA

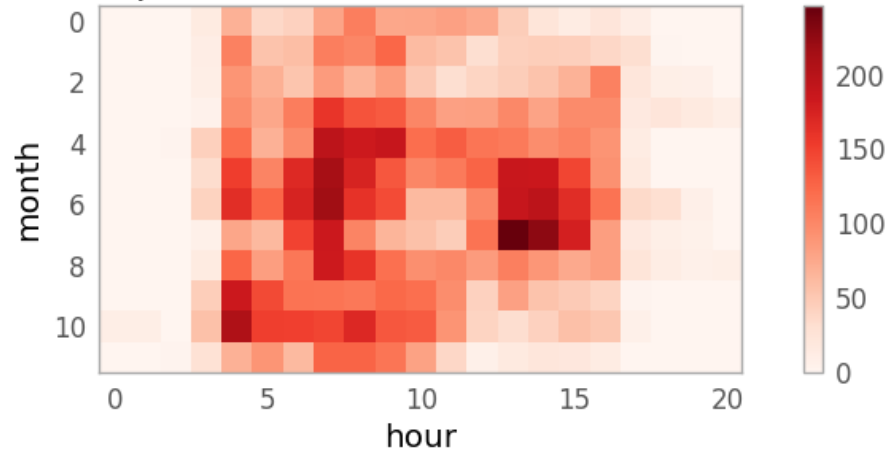
Heatmap of the station: 31 PL. DEL MAR



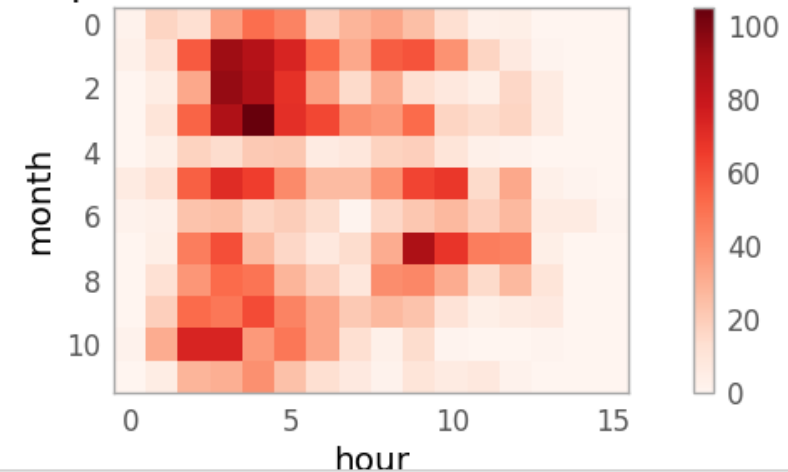
Heatmap of the station: 33 C/PONTEVEDRA-JUDICI



Heatmap of the station: 32 CN BARCELONETA

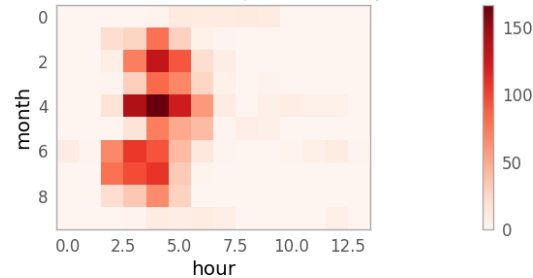


Heatmap of the station: 124 C/NOVA BOCANA

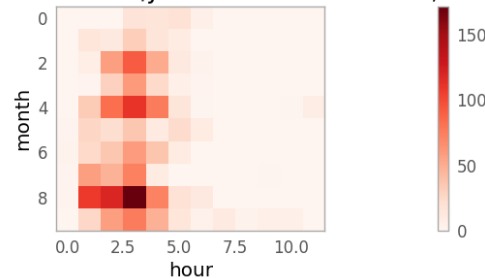


HEATMAPS FALTA DISPONIBILIDAD E STACIONES UB-CAMPUS ECONOMIA

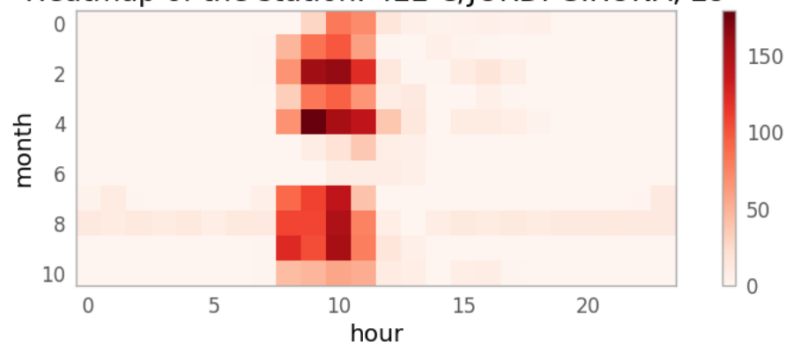
Heatmap of the station: 430 C/ALFAMBRA, 2-DIAGONAL



Heatmap of the station: 433 C/JOHN MAYNARD KEYNES, 2-DIAGONAL

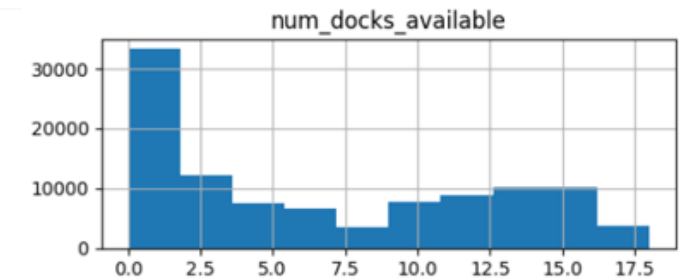


Heatmap of the station: 422 C/JORDI GIRONA, 29

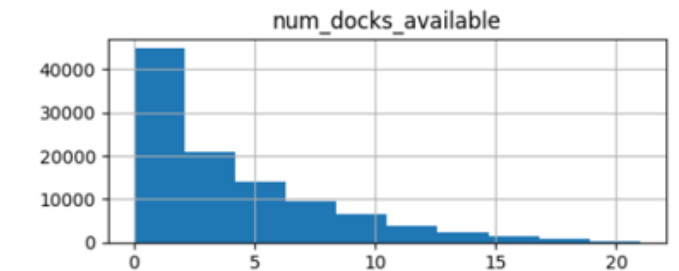


HISTOGRAMA ESTACIONES ATÍPICAS

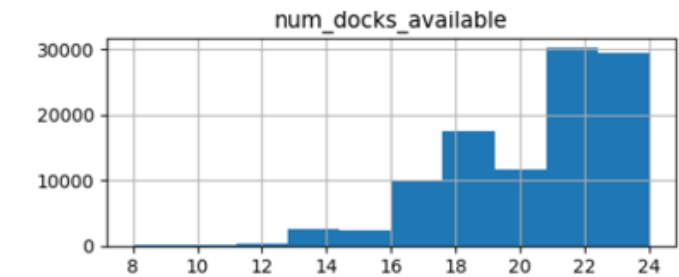
ESTACIÓN 31
Poca disponibilidad



ESTACIÓN 378
Poca disponibilidad



ESTACIÓN 482
Alta disponibilidad



CONCLUSIÓN FINAL:

Las estaciones de algunos distritos como Ciutat Vella y Sant Martí, son las que tienen un elevado índice de falta de disponibilidad, debido a que son céntricas o bien están en zonas próximas al mar, zonas de ocio, oficinas o bien centros comerciales.

Algunos distritos como Ciutat Vella y Gràcia tienen calles estrechas que imposibilitan la instalación de estaciones con muchos docks.

Para elevar el uso del Bicing es necesaria la ampliación del volumen de carriles bici y conexión con otros puntos de transporte, área metropolitana, para hacer hubs de intercambio, fomentando la distribución.

La orografía de Barcelona hace que el uso de la Bicicleta sea dispar, aunque el uso de la bicicleta eléctrica ayuda a repartir las Bicicletas hacía los barrios del norte de la ciudad.

La climatología en Barcelona hace que el uso de la bicicleta sea agradable en casi todos los meses del año, menos en pleno invierno o verano.





**iii MUCHAS GRACIAS
POR VUESTRA
ATENCIÓN!!!!**

