Bike Availability Prediction

Predict number of bikes using Bicing real data

(Capstone Project)

What is expected?

- To explore data in a ‘real world’ setting
- To identify relevant insights and patterns in the data that can inform business decisions
- To fully analyze data from different sources
- To collaborate with your teammates
- To develop a competitive model
- To present your work in an organized way showing the results you achieved

Important Dates

Submission Deadline

26th June 2024

11.59 p.m.

Short Public Presentation

27th June 2024

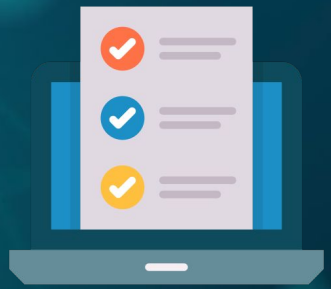
6.00 p.m.

Supervisors

Mariona Carós

Pere Gilabert

Evaluation



Submission delivery

Final Presentation

Organization

Groups of 2-4 people

(3-4 preferably)

Same Problem

Follow up

Sessions

Calendar

18th April - Capstone Project presentation

2nd May - Second face-to-face session to work on the project

& Group formation deadline.

13 th June - Third face-to-face session to work on the project.

26th June - Project delivery

27th June - Short Public Presentation

Proposal: **Bike Availability Prediction**

Two tasks

Prediction

- Predict the percentage of free docks given the historical data of each station.

Study Cases

- Explore new places where stations are needed.
- Explore how different events affect availability.
- ... Your proposal!

Proposal: **Bike Availability Prediction**

The Data

Training / Validation Data: 2020-2023

Test Data (Public & Private): 2024

Bicing stations information: HERE

Use this file to get, for each station:

- capacity (total number of docks)
- coordinates and other information

kaggle

Kaggle competition to evaluate the performance of your models
LINK

To easily download the dataset you can use:

apt-get install p7zip-full p7zip-rar (linux)
+
Run THIS script

Proposal: **Bike Availability Prediction**

Prediction Task

To predict the percentage of free docks for each of the proposed stations given historical data.

index	station_id	month	day	hour	ctx-4	ctx-3	ctx-2	ctx-1	percentage_docks_available
18484	309	3	8	4	0.659091	0.681818	0.666667	0.636364	
50913	114	3	7	21	0.262500	0.041667	0.112500	0.137500	
16655	189	3	14	10	0.232143	0.330357	0.598214	0.711310	
69398	382	3	17	5	0.185185	0.132716	0.129630	0.314815	
11125	284	3	1	3	0.700000	0.719136	0.518519	0.518519	

Proposal: **Bike Availability Prediction**

Where to start?

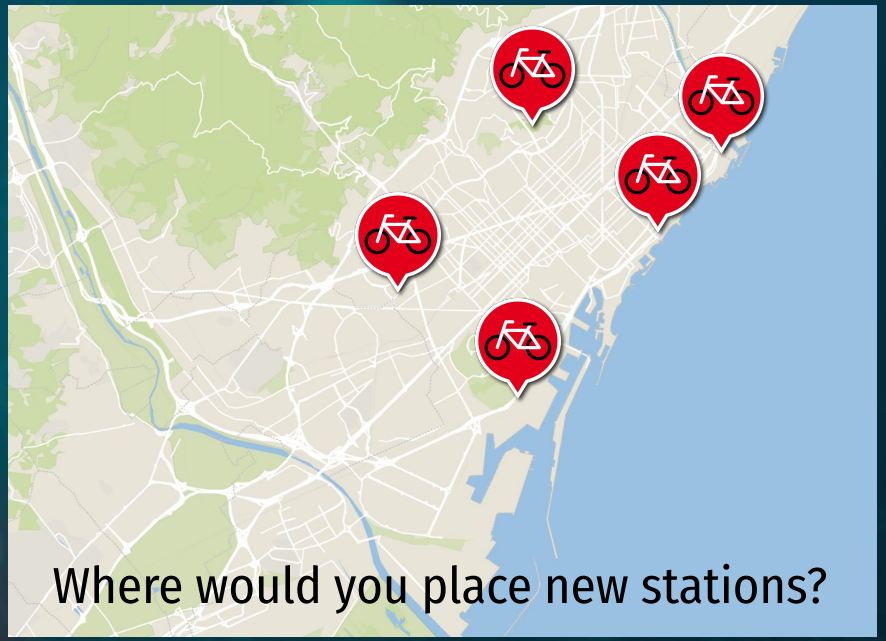
- Start by understanding the data. Take care of possible errors!
- Create a smaller dataset with three partitions: train / val / test.
- Create a regressor model that receives the data and predicts the availability.
- Check the performance... and iterate!
- Include data from other sources: weather, calendar, ...
- Create a Streamlit app to visualize the predictions.

NICE TO HAVE!

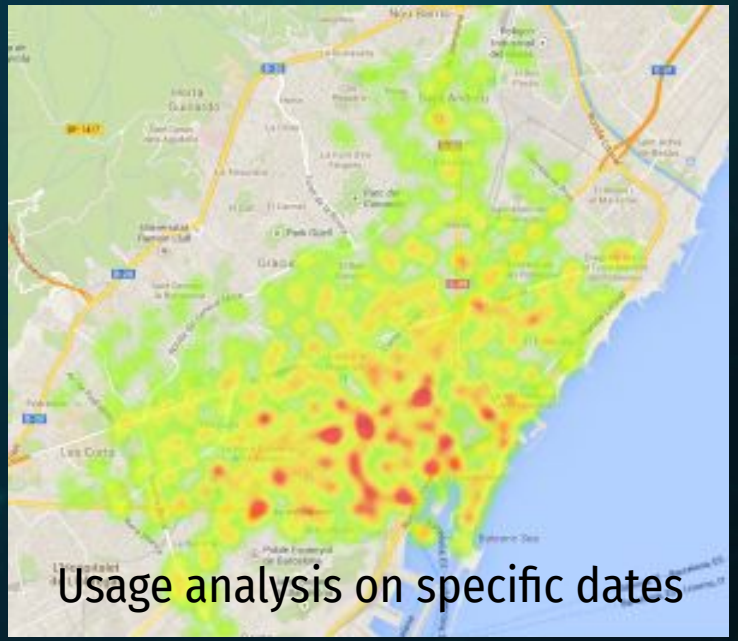
... Sky's the limit!

Proposal: **Bike Availability Prediction**

Case Study Task (Examples)



Where would you place new stations?



Usage analysis on specific dates

Submission Instructions

- Campus Virtual delivery
  - Groups
  - Link to Github Pages
  - User/s used to submit in Kaggle
- Github Pages / Blog / Markdown showing your results and visualizations
- 1 Kaggle submission (at least)

Any Questions?