RAMP project: predicting cyclist traffic in Paris

Master Data science for Business X-HEC, 2021

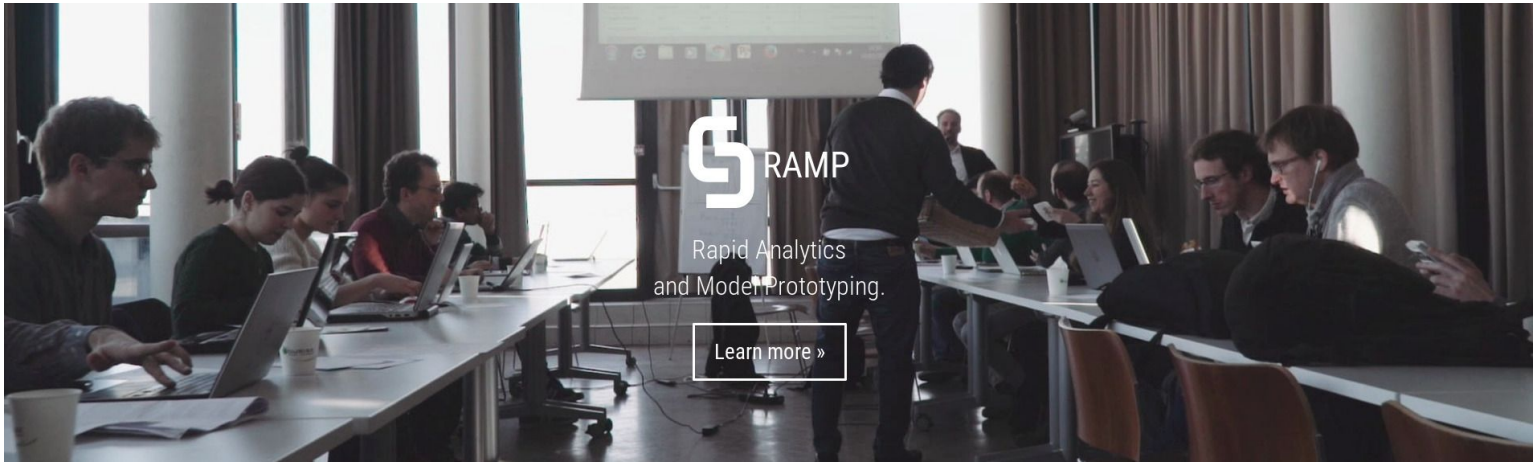
Roman Yurchak,
Mathurin Massias
Hicham Janati

October 2, 2020

RAMP platform

Collaborative data science challenges

- create a model locally using a public data set
- submit on RAMP platform, evaluate on a hidden test set



RAMP instance used for this course: https://ramp.studio

Project problem statement

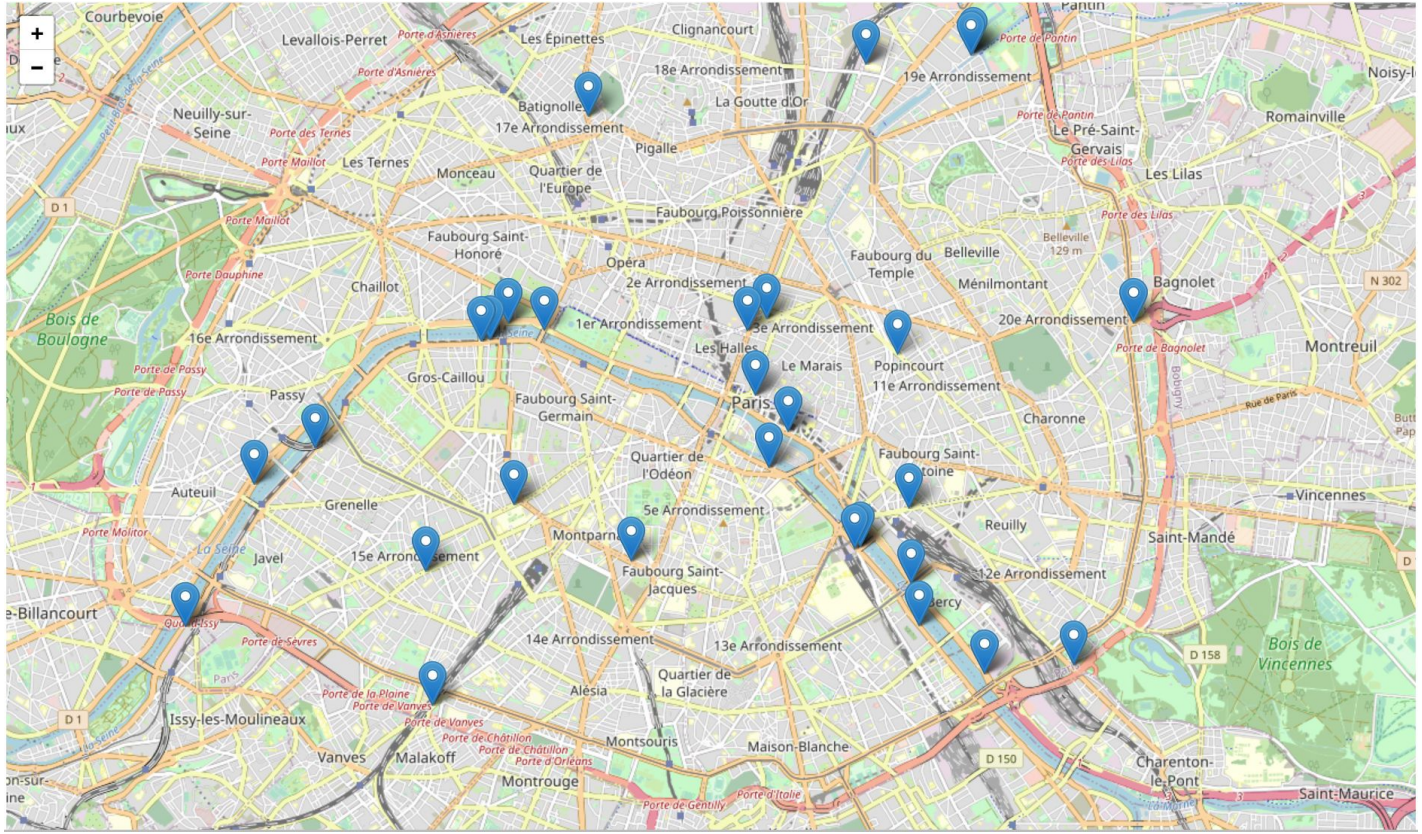
Bike traffic counters installed by Paris city,

- 30 counting sites
- multiple directions
- hourly measurements

Questions

1. Can we predict mean daily traffic for the next few months?
2. Can we predict spikes during rush hours?

Cyclist counter stations in Paris



Predicting the number of cyclists

Available historical data,

- counter name
- counter site name
- date
- counter installation date
- latitude and longitude

\+ any relevant publicly available data you can find.

Starting kit: https://github.com/rth/bike_counters

Workflow

1. Download the starting kit https://github.com/rth/bike_counters and install the environment

2. In groups of 2 create a private Github repo; add the starting kit files

3. Develop and evaluate the machine learning model locally (for instance, in a jupyter notebook)

4. Put your solution in a RAMP format, upload to https://ramp.studio, see results

5. Iterate with 3.

Starting with the RAMP platform

1. Create an account on https://ramp.studio
    - wait for account approval
2. Join the event

    https://ramp.studio/events/bike_counters_mdsb2021

Approach

Approach

- Explore different ML models
- Data preprocessing, feature extraction
- External features / datasets
- Tune hyper parameters

Bonus (advanced)

- model interpretability