

PREDICTING CYCLIST TRAFFIC IN PARIS

Solal Zana & Adam Berdah

SOMMAIRE

- I. Project Presentation**
- II. EDA & Preprocessing**
- III. Model Selection**
- IV. Evaluation of the final model**
- V. Future Research**
- VI. Conclusion**

I. PROJECT PRESENTATION

- **GOAL:** Predict bike count at a given hour across 56 counters in Paris.
- **METRIC:** Root mean square error (RMSE)
- **TIMEFRAME:**
 - Training data: September 1st, 2020 – September 9th, 2021
 - Public Test data: September 10th, 2021 – October 18th, 2021
- **DATA:**
 - training dataset partly furnished
 - Testing set: 50% public, 50% private.

II. EDA

ADDING EXTERNAL DATA

Cleaning:

- Missing values
 - Deleted columns containing > 10% missing values
 - Replaced the rest with the median
- Deleted columns with a single value
- Deleted some highly correlated columns

Added features:

- French school holidays (zone C)
- Bank holidays
- Quarantines
- Seasons
- Distance from city center

ISSUE: merging both datasets gave a huge error (RMSE \approx 2.15)

II. EDA

ENCODING

Standard Scaler	One-Hot	Cyclical encoding
All numerical columns	Counter name Years Weekday	<ul style="list-style-type: none">• Hour• Day• Weekday• Month

IV. MODEL EXPERIMENTATION & SELECTION

- Models tested:
 - Ridge Regression
 - Random Forest
 - XGBoost
 - Catboost
 - LightGBM
- Evaluated model performance on XGBoost using cross-validation and RMSE as metrics.
- Grid Search on the most probant model

V. EVALUATION OF FINAL MODEL

OUR FINAL BEST RESULT:

Submission and Description

Private Score ⓘ

Public Score ⓘ

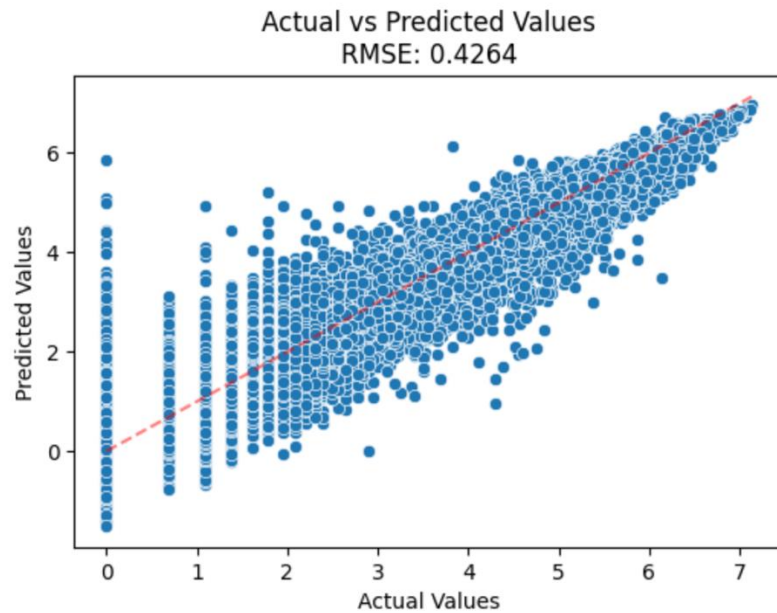


submission_xgboost.csv

Complete (after deadline) · Solal Zana · now

0.7247

0.7366



VI. FUTURE RESEARCH

- Analysis of construction work in Paris from additional datasets ?
- Test other types of encoders ?
 - Gap encoder or Target encoder
- Integrating additional data like geographical traffic density or inhabitants density
- Exploring neural networks

The background image is a faded, light-colored photograph of a Parisian street scene. In the background, the Arc de Triomphe is visible, partially obscured by the text. In the foreground, three cyclists are riding from left to right. The cyclist on the left is wearing a red long-sleeved shirt, dark pants, and a white face mask. The middle cyclist is wearing a light grey hoodie, dark pants, and a blue helmet. The cyclist on the right is wearing a blue jacket, a yellow high-visibility vest, a blue helmet, and has an orange bag in their front basket. The street is lined with trees and lampposts, and a few cars are visible in the distance.

THANK YOU FOR YOUR ATTENTION