**REPORT – BIKECOUNT PROJECT**

# INTRODUCTION

Paris is a city where cycling is … It is an option which is relatively much more environmentally friendly.

In this context, this project was about ….

The main objectives for our work were (coherent project with relevant hypotheses/ clear storyline, a bit of creativity, great score(what is a great score?)) …

In the following sections, the methodology followed to reach those objectives was explained. Afterward, analysis and interpretation of the results obtained were conducted.

# METHODOLOGY

## Strategy

### Data discovery and quick construction of a machine learning program

The main goals were to get familiar (domain knowledge by data source documentation, articles on bicycles count devices, etc) with the features for which data were recorded to form the datasets, do a basic Explanatory Data Analysis (EDA), and to build a Machine Learning (ML) model as a sort of baseline for the next steps of the work. Indeed, the results obtained will be used to deal better with the model parameters, the …, … This base model and related EDA were strongly inspired by the notebook uploaded by the professo

### Missing values, outliers analysis, categorical variables

### Theories/propositions to deduct pool of features.

The two main goals here were to:

* 1. Improve the baseline model built in the previous part by adding relevant additional features
  2. Try new combinations of features based on a specific assumption (e.g: bikes are used only in sunny period, so the features and observations of … are discarded)

To do so, the pool of features was determined following a key question: “Why do people use bikes in big cities?” To answer this question, we went through a few research articles, social media content (Bikes-related threads, hashtag, and videos from X, TikTok, Instagram, Reddit, and Youtube). We could have also surveyed people using bikes outside if were able to. Then, from knowing why people use bikes, we would make suggestions on when (time, season, special events, strikes, etc) Parisians use bikes.

### Models to choose: How complex will it be?

We thought of having a good-performing model that can be interpreted, and if possible a black box model which the best performance we could get.

From our understanding of the different classes we had so far, and from what we know about what the community says, XG Boost was used as a base model. In addition, a modified …. was implemented.

### What to do if extra time

More niche features like visibility, dew point temperature, bridge day, day of strike,

### Relevant resources for guidance

Academic research papers, Stackoverflow, Medium, online classes from udemy-coursera.

## The three (can be only 2 if you want) propositions

## Proposition 1

## ~~Proposition 2~~

## ~~Proposition 3~~

## Workflow

* Git & Github (notebooks sharing)
* Google drive (docs and slide)

# RESULTS

## Proposition 1

## Proposition 2

## Proposition 3

Use this webite to insert screenshots of our cod https://carbon.now.sh/

# ANALYSIS AND INTERPRETATION

# CONCLUSION (and future work?)

FACTORS INFLUENCING THE ACCURACY OF THE COUNT (trotinettes being counted for example, how old is the counting device,

# REFERENCES