



# Python Project

**Facebook Comment Volume Dataset**

# the ins and outs of the problem

---



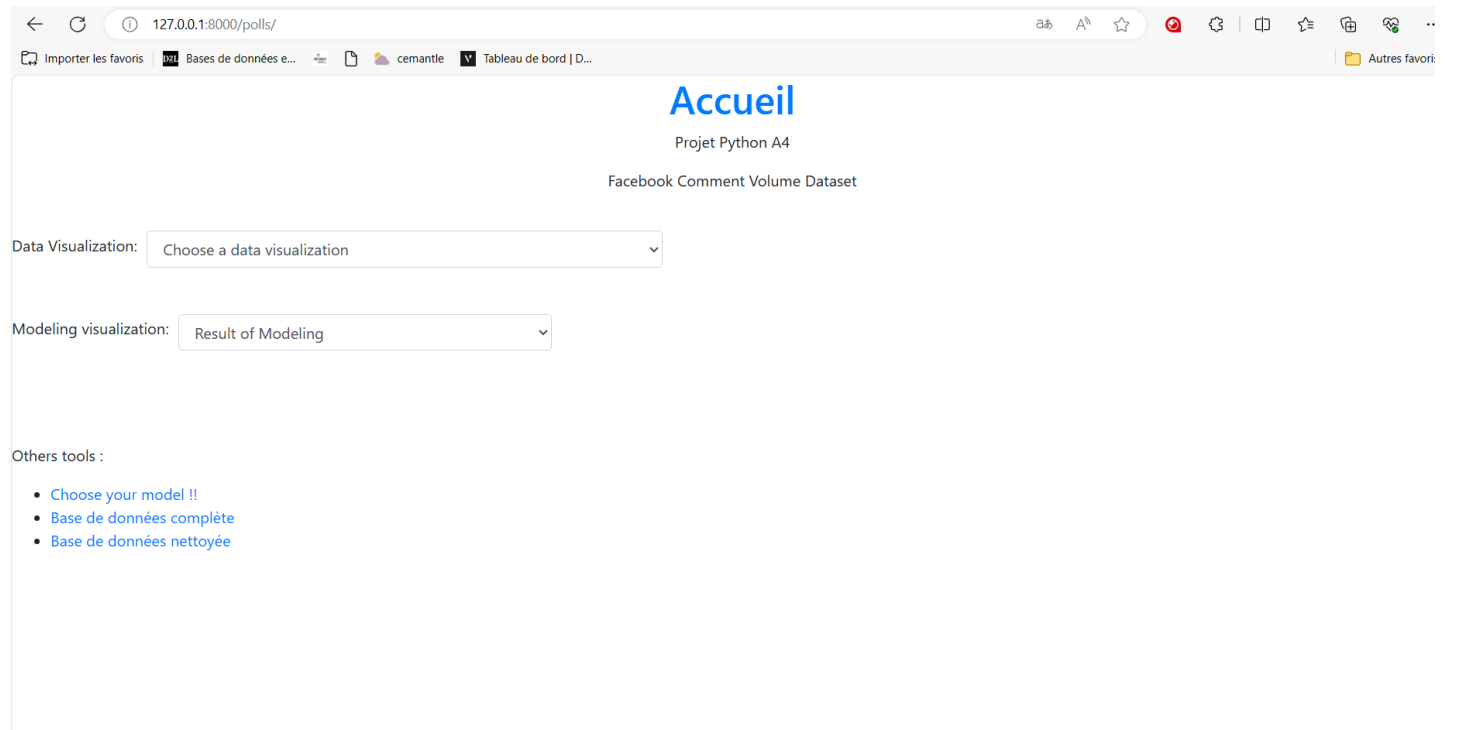
- Our dataset is consist of 53 features.
- This dataset is relevant for understanding engagement patterns on Facebook posts. It offers insights into what factors might influence the comment volume a post receives. Predicting the best features to increase comment count can help social media managers or analysts optimize their content strategy and posting schedules to maximize engagement.
- How can we predict the best way to increase the number of comments that a Facebook post will receive in the next few hours?



# Our thoughts on the asked question

- This dataset encompasses a range of features, allowing for diverse analysis and modeling approaches. Feature engineering plays a crucial role here to derive more insightful predictors and improve the predictive capacity of the model.
- The target variable, the number of comments in the next H hours, serves as the focal point for prediction. Understanding the interplay between various features and their influence on comment volumes can be pivotal in developing a robust predictive model.
- The dataset's timestamped nature allows for time-series analysis, enabling the exploration of temporal patterns and trends in comment volumes.

# Django site



# Data visualization :

Data Visualization:

Choose a data visualization



Choose a data visualization

Top 10 page category published

Correlation with target (week)

Correlation with target (post length)

Correlation with target (share post)

Median like and visitors linked to the target

Linked between CC2 and CC5 and linked between CC4 and the target

Modeling visualization

# Model

Modeling visualization:

Result of Modeling



Result of Modeling

Analyze of different model

Top 10 features importance (ExtraTreesClassifier)

Top 10 features importance (SelectKBest)

PCA with 2 features

feature importance linear regression

Others tools :

- [Choose your model !!](#)

# Model personalized

## Accueil

Choose the model between 'ETC' (ExtraTreesClassifier) 'SKB' (SelectKBest) or 'LR' (Linear regression) :

Choose the number of features wanted :

Save and start

Your model can be long te be started please wait a minute

---

## How the problem fits in the context of the study

- The problem of predicting the number of comments a Facebook post will receive fits within the context of the study as it aims to model user patterns and study the effectiveness of machine learning predictive modeling approaches on the leading social networking service, Facebook. The study recognizes the massive amount of data being uploaded to social networking services and the need to understand the dynamic behavior of users towards these services.





# Conclusion



We can conclude that the features importances linked to the target can vary depending of the model.



The Random Forest model appears to be the most performant for predicting our target variable

---

Random Forest - RMSE: 0.015299100122301353

Random Forest -  $R^2$ : 0.6272726239776909