

Seoul Bike Sharing Demand

Inès PEREZ & Mathilde SALAÜN

DIA 5



Overview

Importation, NaN values,
definition of new variables

Encoding

Best prediction and best
model, what can we say ?



Ploting : matplotlib, seaborn,
plotly express, bokeh

Regression and Classification
models, comparaison

Flask



Exploring our dataset

WHAT WE HAD

- Date : year-month-day
- **Rented Bike count - Count of bikes rented at each hour**
- Hour - Hour of the day
- Temperature-Temperature in Celsius
- Humidity - %
- Windspeed - m/s
- Visibility - 10m
- Dew point temperature - Celsius
- Solar radiation - MJ/m2
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday/No holiday
- Functional Day - NoFunc(Non Functional Hours), Fun(Functional hours)

Décembre 2022

OUR GOAL

- Study the impact of all the variables on the number of rented bikes

Exploring our dataset

2 GENRES OF VARIABLES

- **Temporal ones :**

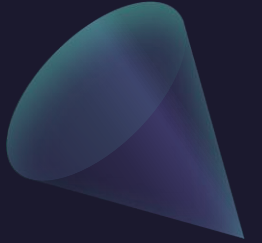
Date
Hour
Holiday
Functional Day

- **Meteorological ones :**

Temperature
Humidity
Windspeed, Visibility
Dew point temperature
Solar radiation
Rainfall
Snowfall
Seasons

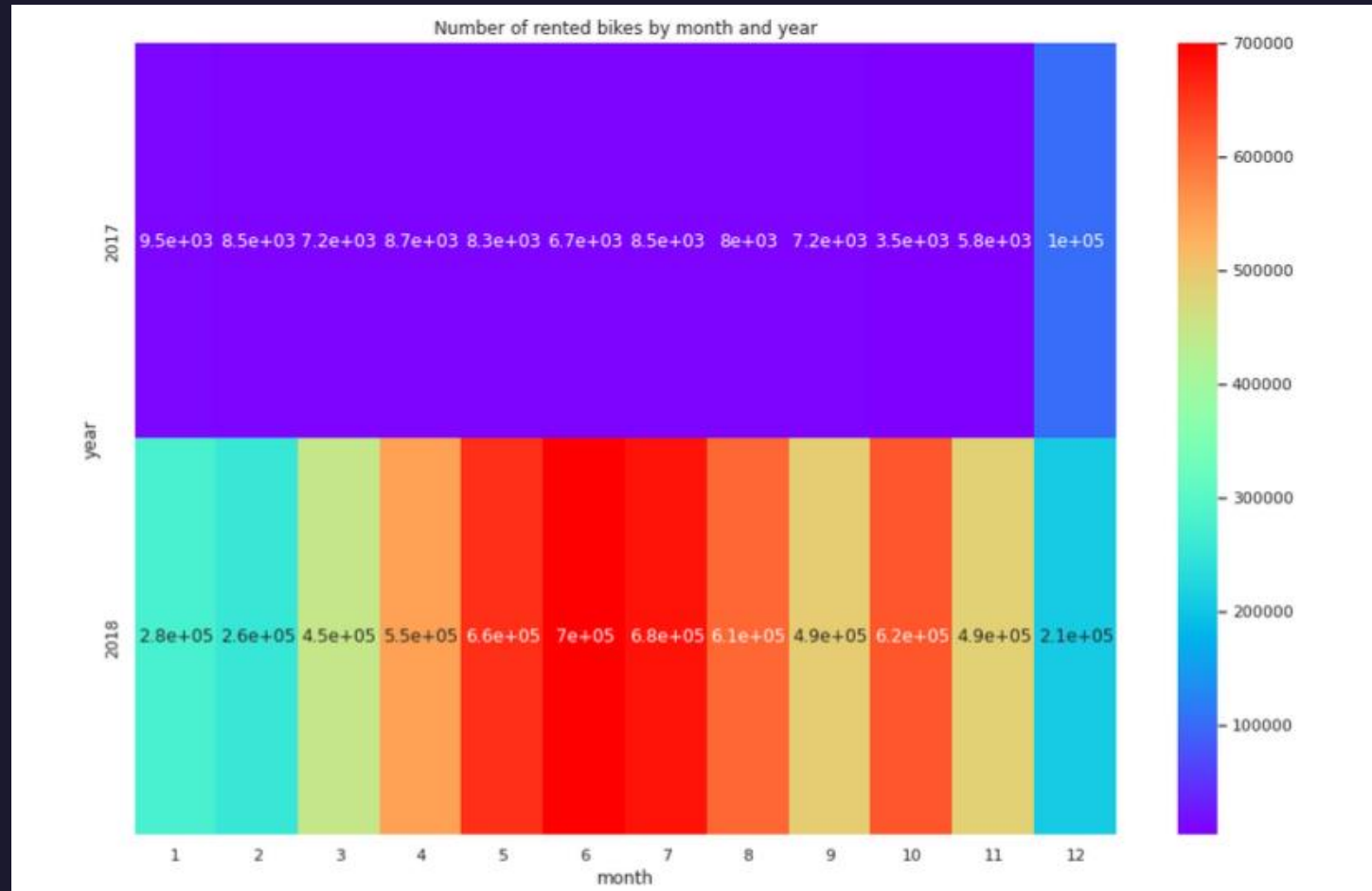
WHAT WE HAVE DONE

- Fixing our target : Rented Bike count
- No NaNs values
- Rename columns
- New variables : day, month, year, Moment_of_day and bike_affluence
- Correction of the season label



Visualisation

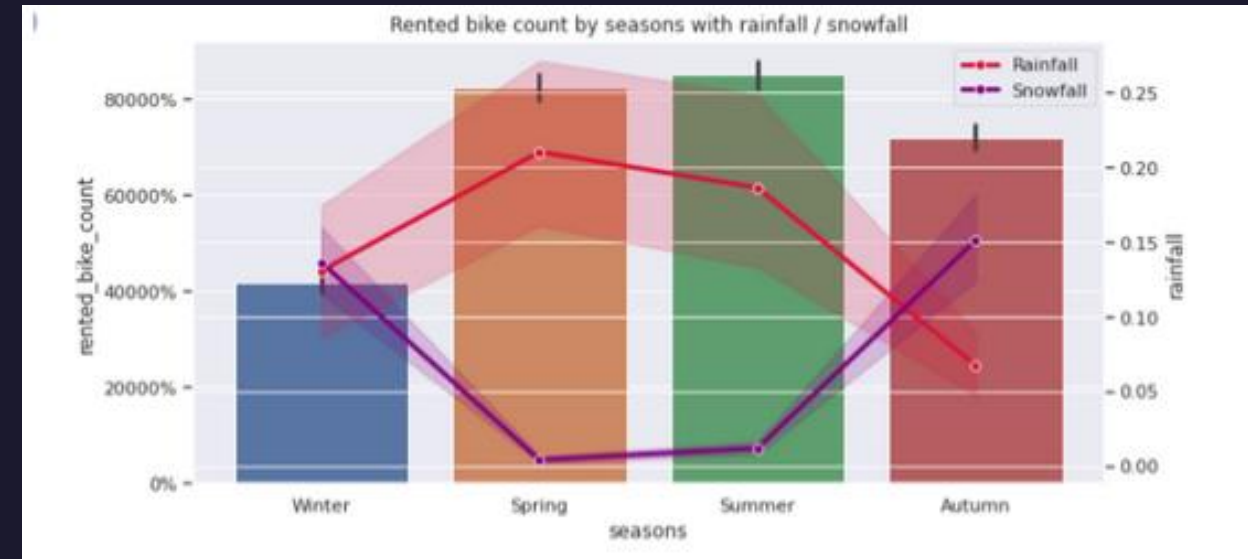
TEMPORAL VARIABLES



Visualisation

METEOROLOGICAL VARIABLES

- Less influence : Humidity, Rainfall
- More Influence : Wind speed, Temperature, Snowfall



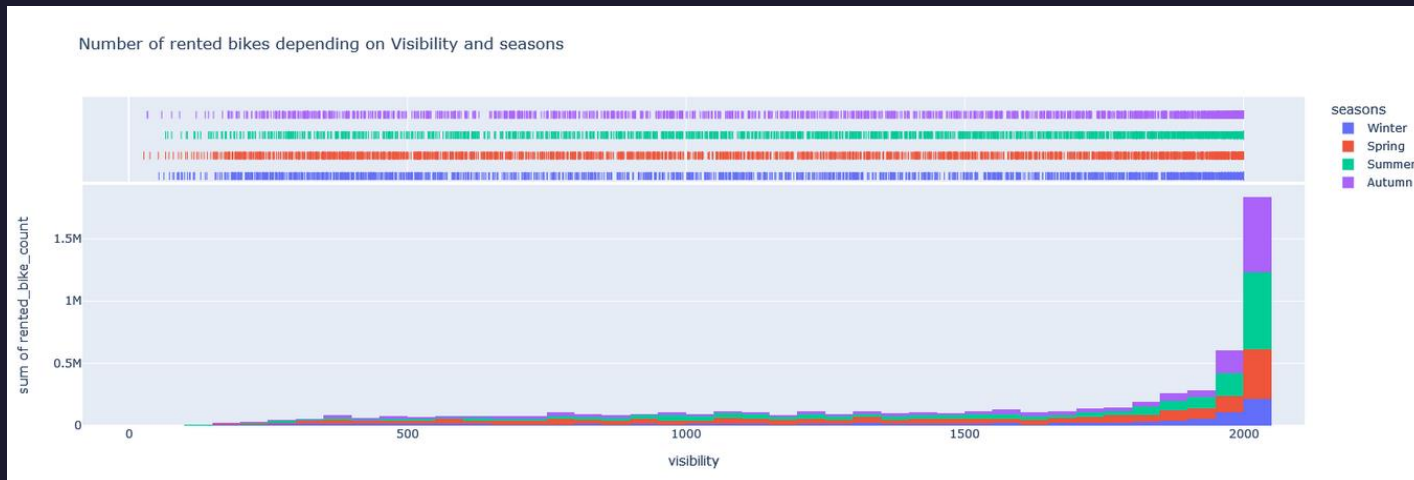
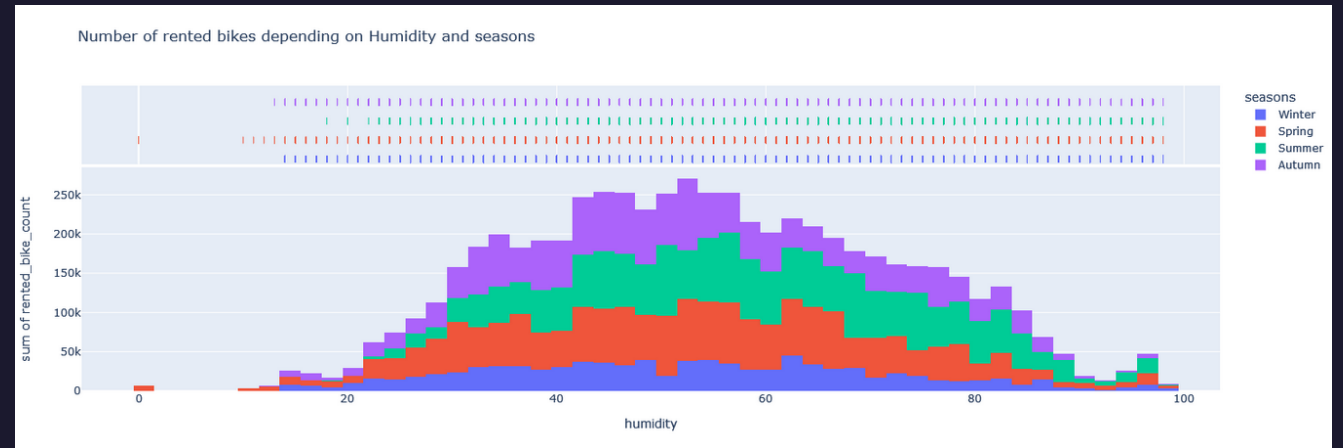
Sunburst of the number of rented bike throughout time and seasons



Visualisation

METEOROLOGICAL VARIABLES

- Less influence : Humidity, Rainfall
- More Influence : Wind speed, Temperature, Snowfall



Pre-processing

CORRELATION BETWEEN VARIABLES

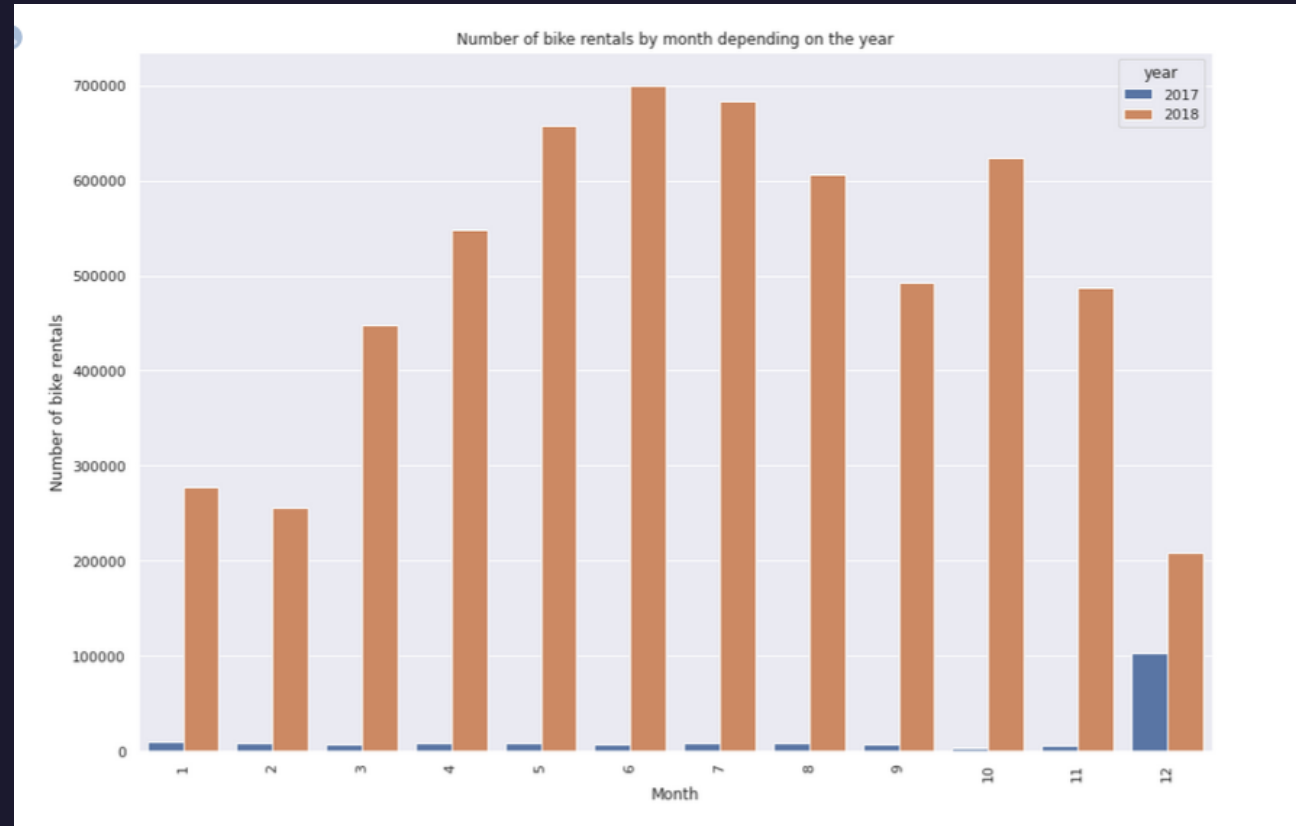
- Temperature : 0.53
- Hour : 0.45
- Dew point temperature : 0.36
- -> but extremely correlated with temperature (0.91)
- Seasons extremely correlated to month (0.91)



Pre-processing

DROP

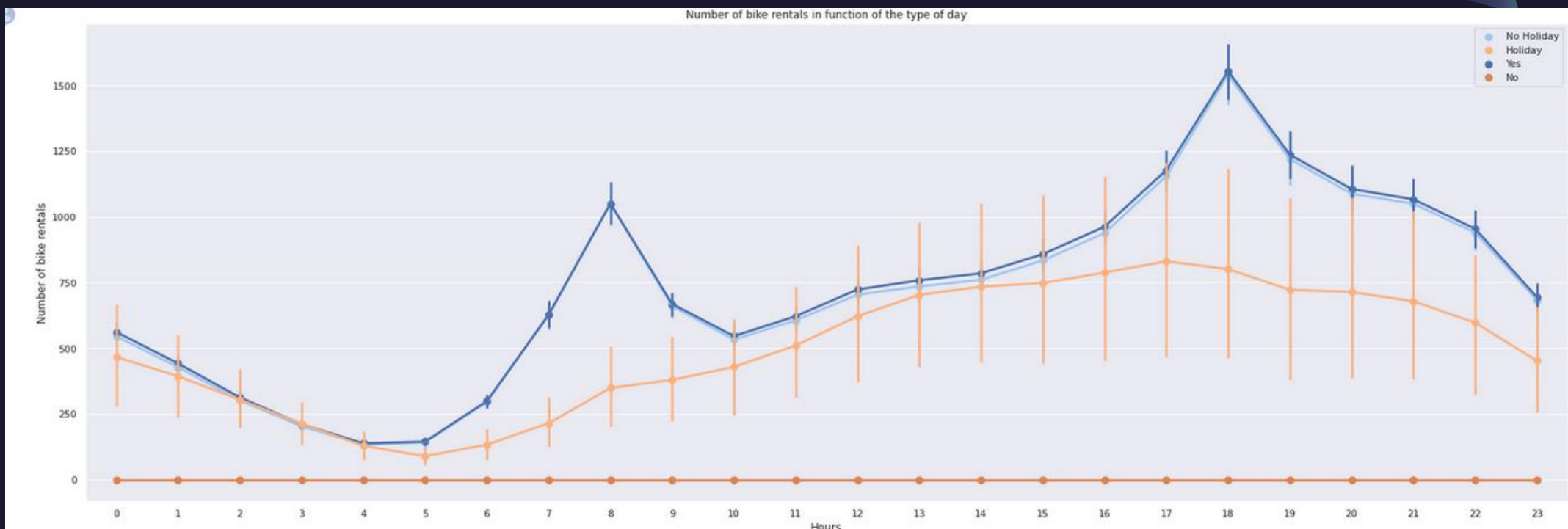
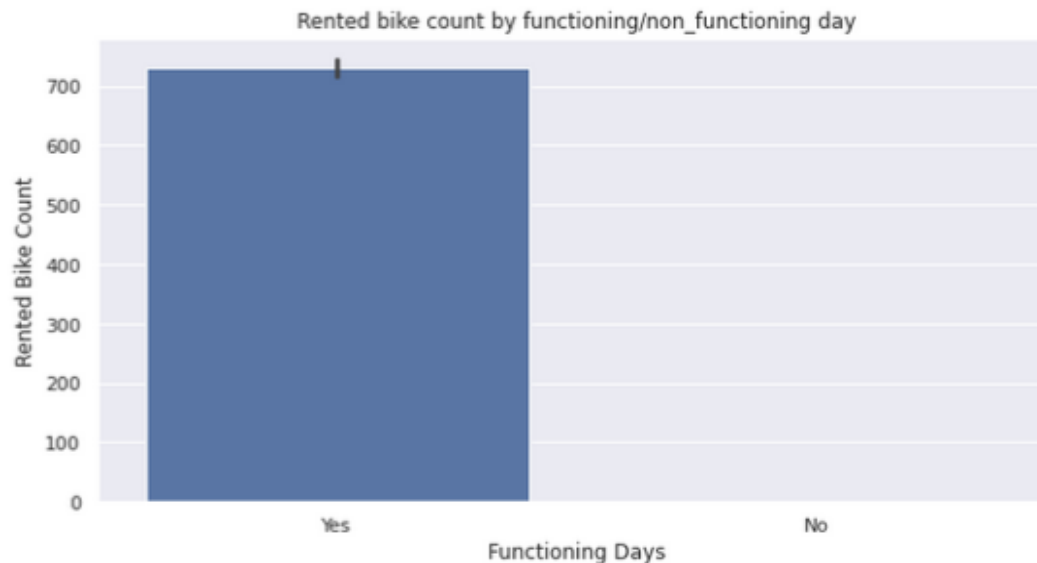
- Moment_of_day
- Functioning_day
- Year
- Seasons
- Dew point temperature
- Date



Pre-processing

DROP

- Moment_of_day
- Functioning_day
- Year
- Seasons
- Dew point temperature
- Date



Pre-processing

ENCODER

- Seasons

BINARIZER

- Holiday

NORMALIZE

- All predictors for classification

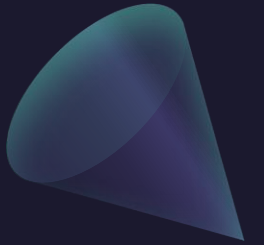
Modeling

REGRESSION MODELS

- KNN Regressor
- Hist Gradient Boosting Regressor
- Gradient Boosting Regressor
- Bagging Regressor
- Extra Trees Regressor
- Random Forest Function
- LGBM Regressor

CLASSIFICATION MODELS

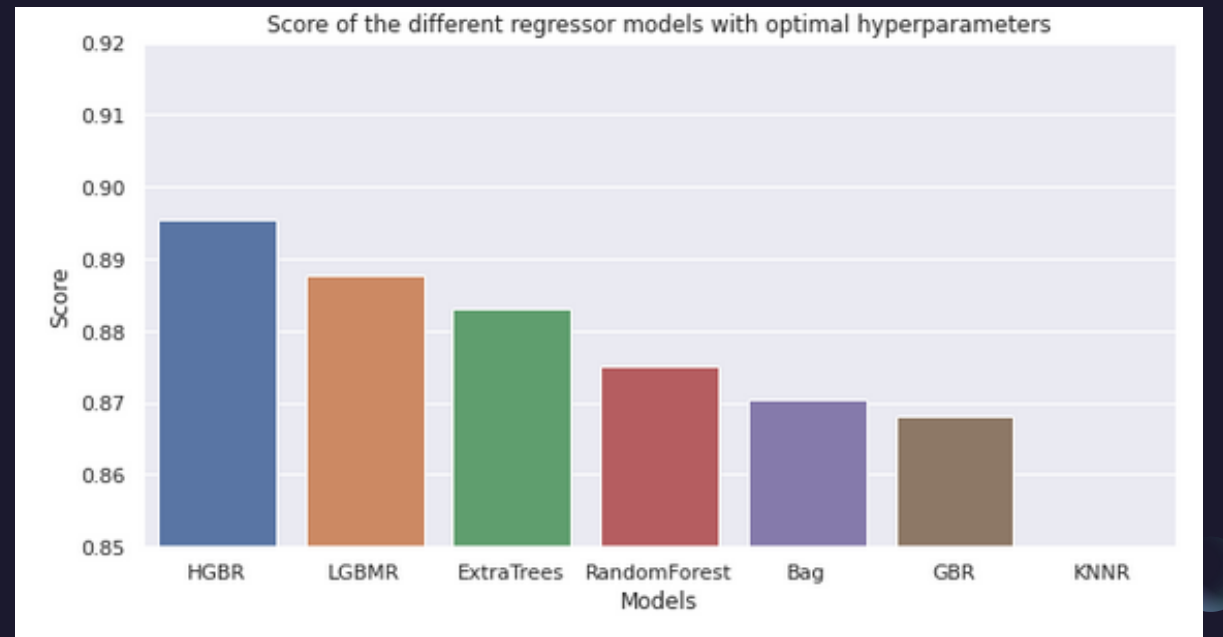
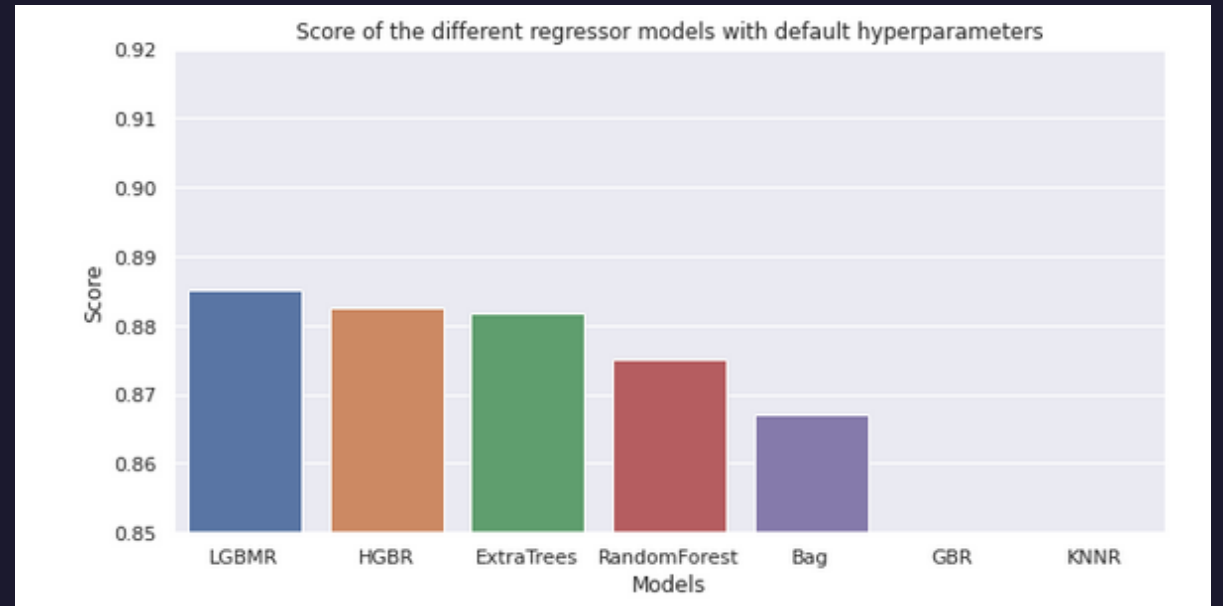
- LGBM Classifier
- Random Forest
- Extra Trees
- KNN Classifier



Modeling

FINDING THE BEST HYPERPARAMETERS

- GridSearch()
- Comparison between defaults and optimal hyperparameters



Modeling : Comparison - Accuracy

	Regressor	Classifier
Multiple Linear Regression	50.78%	-
KNN	57.52%	58%
Gradient Boosting	86.59%	-
Bagging	86.76%	-
Random Forest	87.38%	78.25%
Extra Trees	88.25%	78.35%
LGBM	88.60%	78.82%
Hist Gradient Boosting	89.77%	-

Conclusion

- Hour and Temperature seem to influence the demand for bike a lot
- Our best model is :
Hist Gradient Boosting Regressor
(HGBR) ~89.77%
- Our second best model is :
Light Gradient Boosting Machine Regressor
(LGBMR) ~88.60%



API - Flask

LGBMR model

[GitHub](#)Seoul Bike Sharing DemandBy Inès PEREZ & Mathilde SALAÜN

Rented Bikes Prediction

Fill the inputs to have a prediction :)

Predict Value

Number of rented bikes: 117