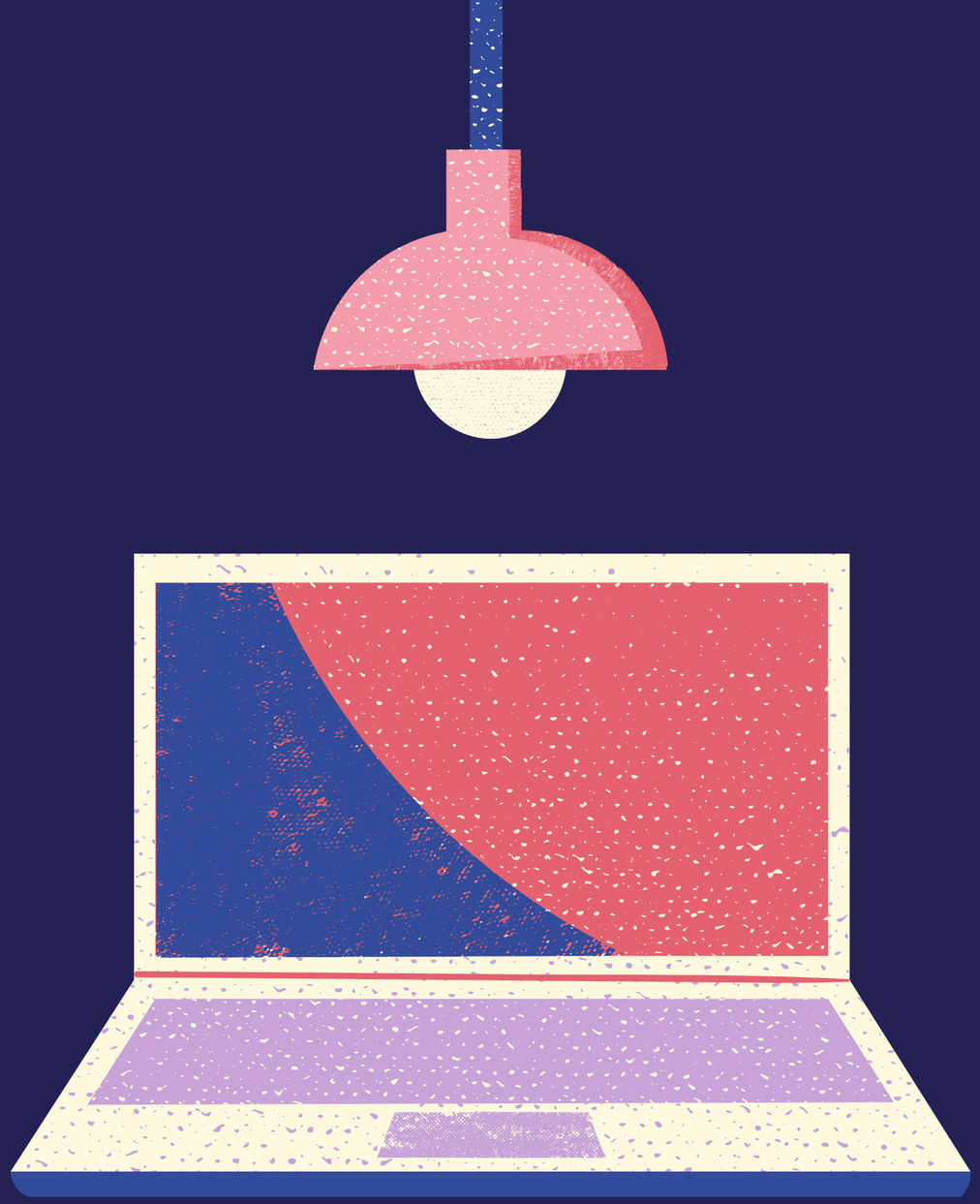


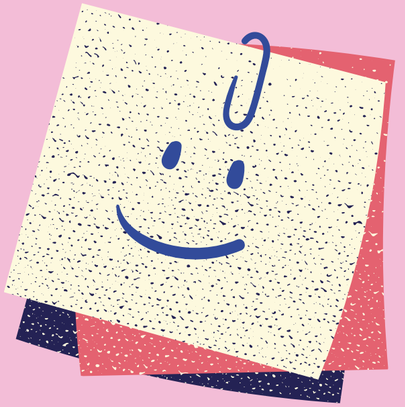
Online Shoppers Purchasing Intention

Taous Bedouhene, Inès Baazia, Elise Barbier

- DIA 1-

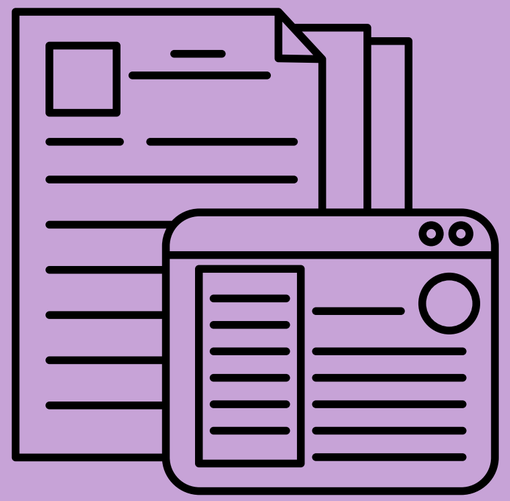
Python for Data-Analysis

Table of Contents



INTRODUCTION

The ins and outs



About 70% of carts are abandoned, which happens when a potential customer chooses products to buy without ever going through the checkout process.

We will produce whether the user buys the product or not

Our Database

- We use a dataset that gives us an estimation of online shoppers purchasing intention.
- We aim to find a model based on the data, which will be used to make predictions on new data to know if the purchase was successful or not.

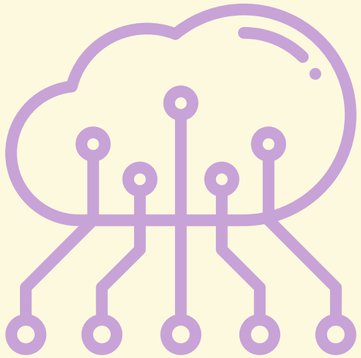
Attributes

- Administrative
- Administrative Duration
- Informational
- Informational Duration
- Product Related
- Product Related Duration
- Bounce Rate
- Exit Rate
- Page Valuer un sous-titre
- Special Day
- Mounth
- Operating Systems
- Browser
- Region
- Traffic Type
- Visitor Type
- Weekend
- Revenu

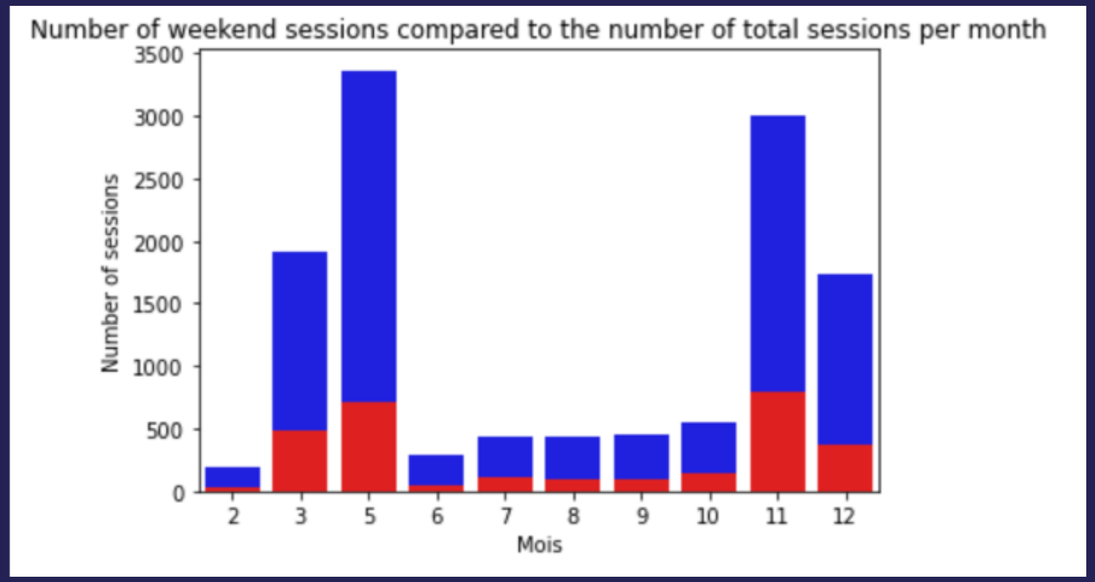
The variable we are trying to predict is the **revenue**

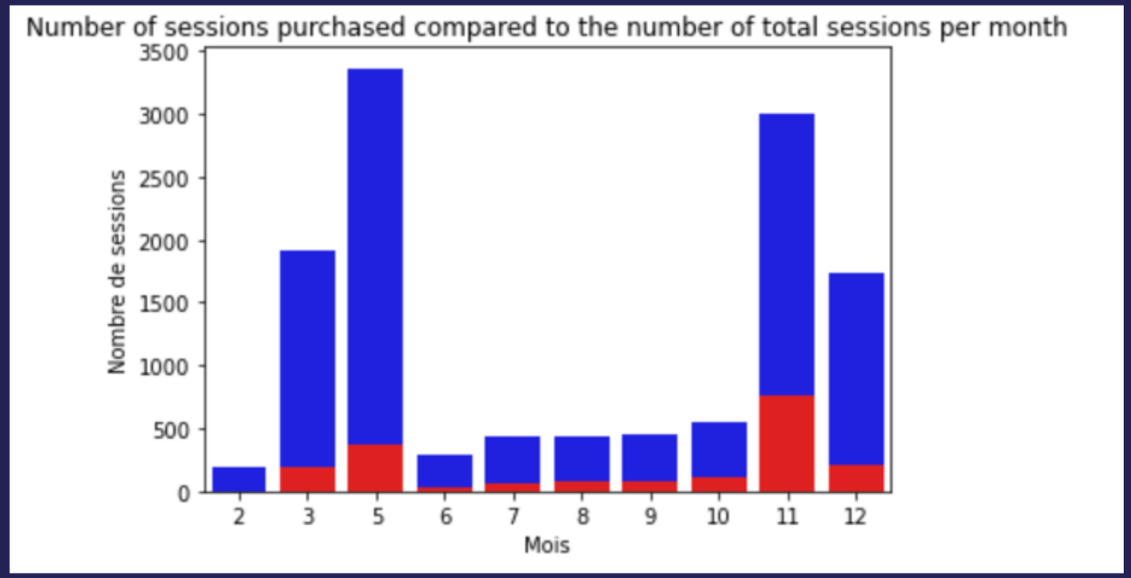
New values

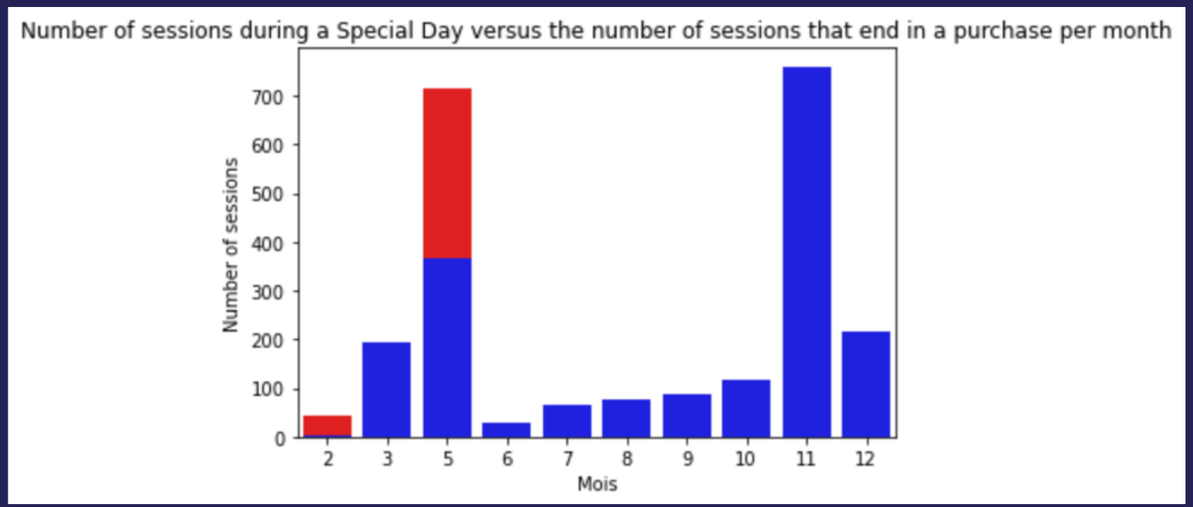
- "Moyenne_Administrative": mean of the time spent on the administrative part
- "Moyenne_Informational": mean of the time spent on the informational part
- "Moyenne_ProductRelated": mean of the time spent on the product related part



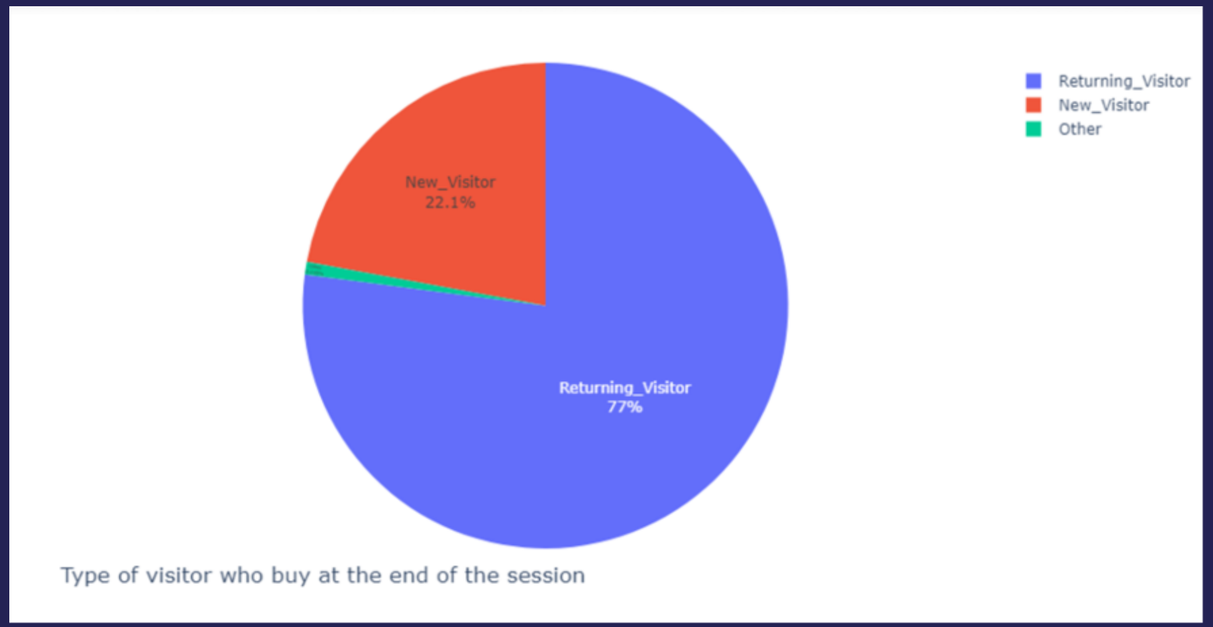
Data-visualization

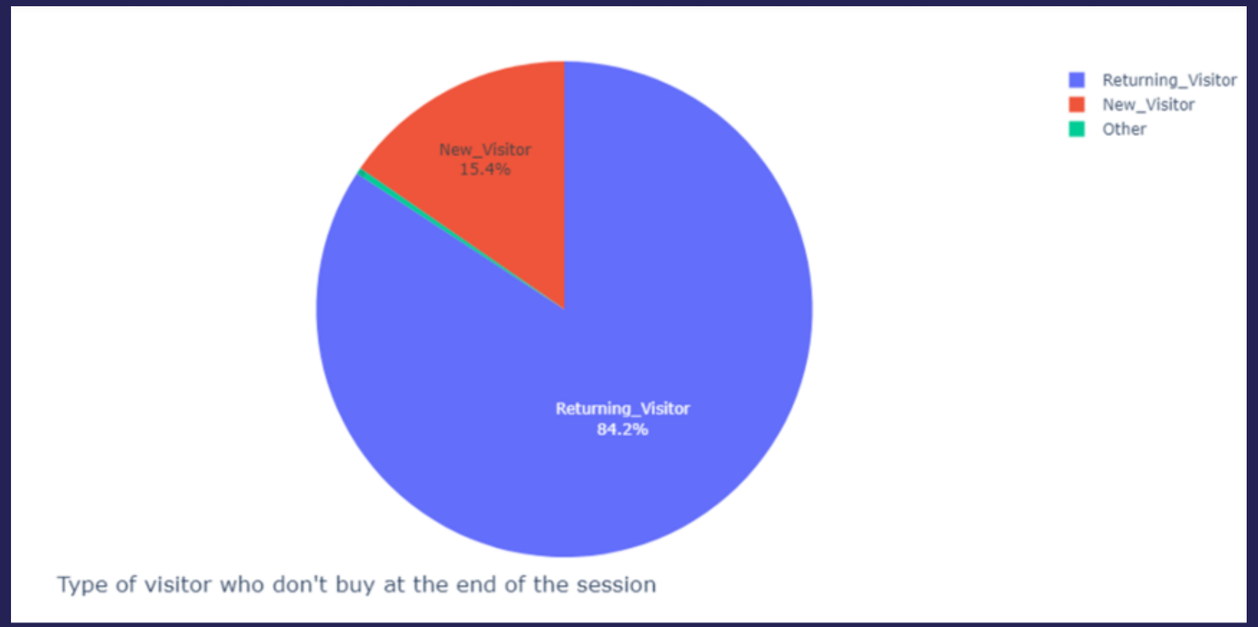




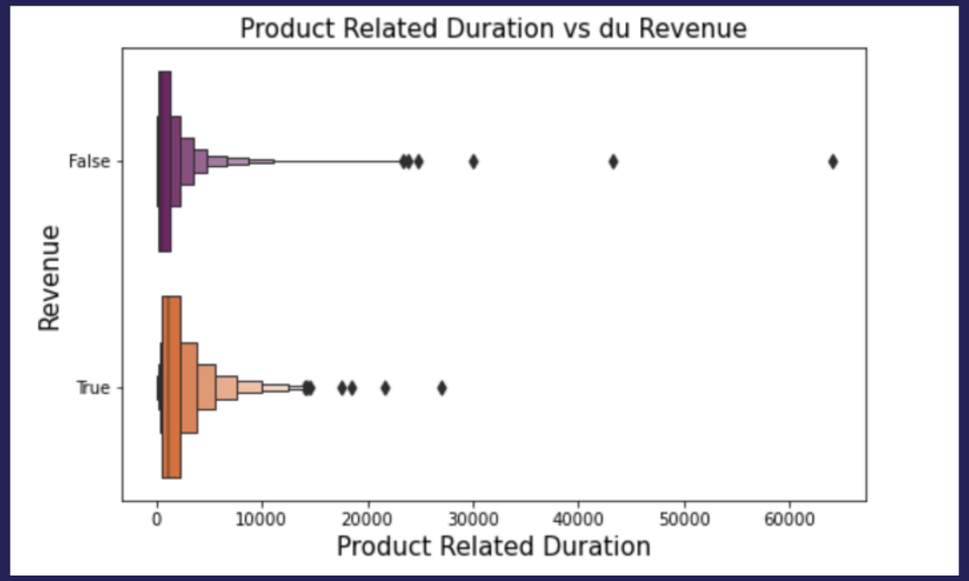


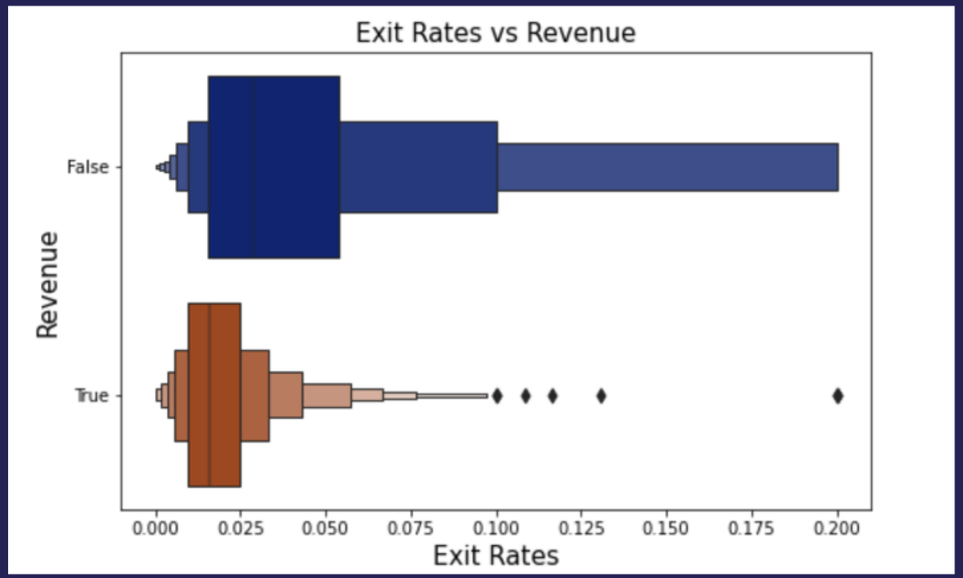
Data-visualization

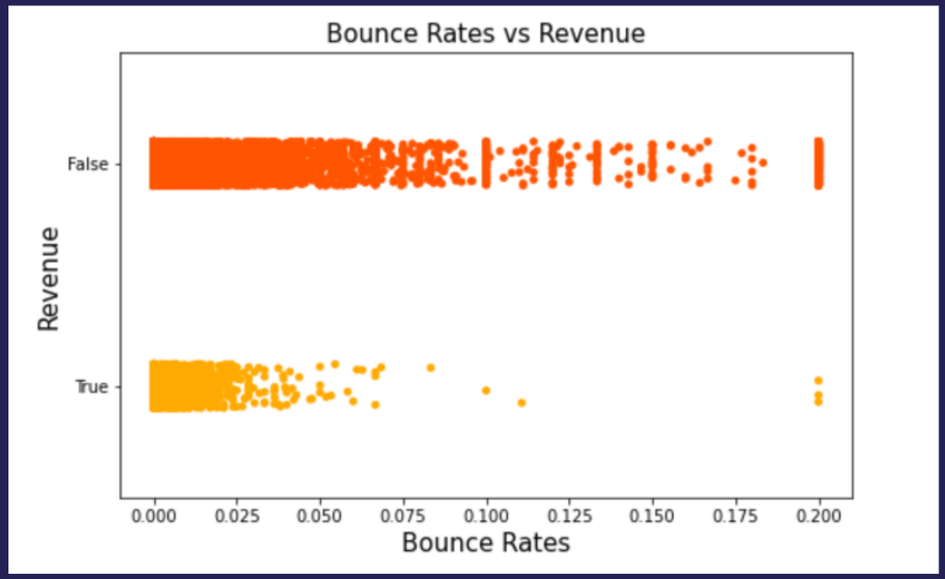


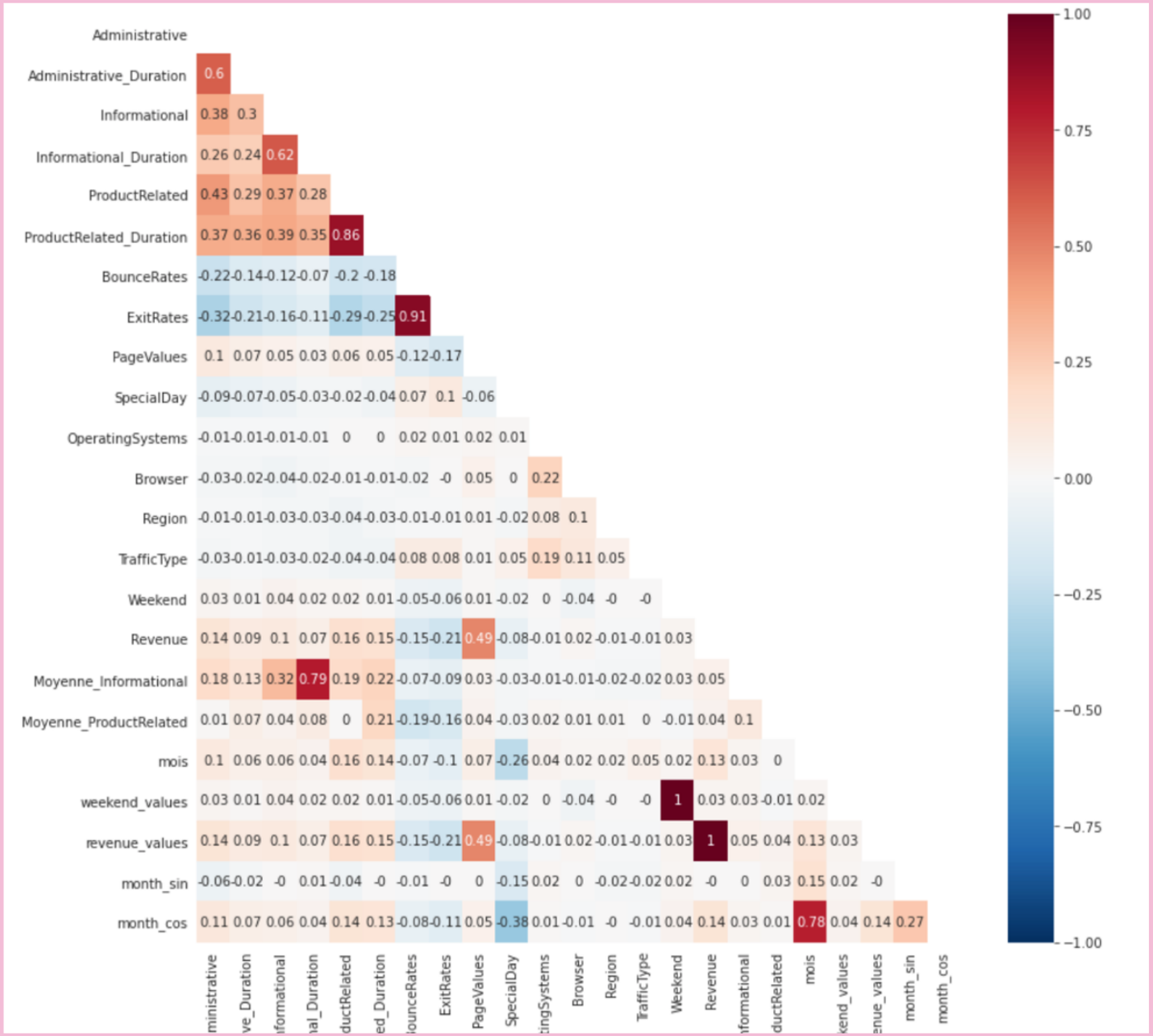


Data-visualization







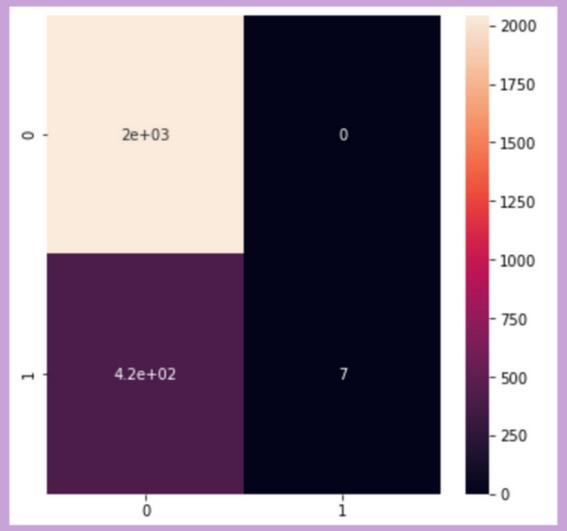


Matrix of correlations

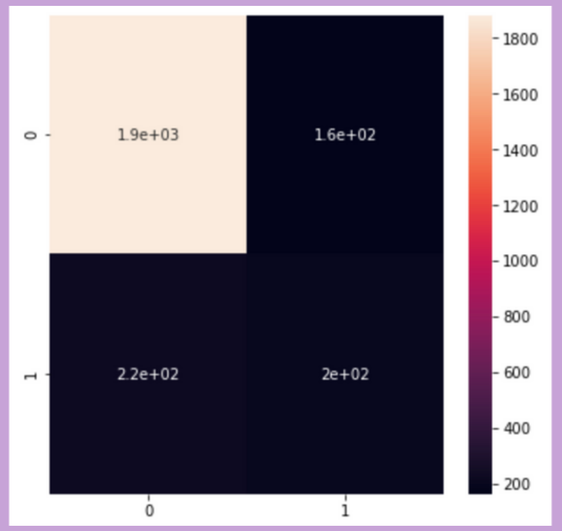
It shows that our target (Revenue) has high correlation with :

- Administrative
- Administrative_Duration
- Informational
- Informational_Duration
- ProductRelated
- ProductRelated_Duration
- PageValues

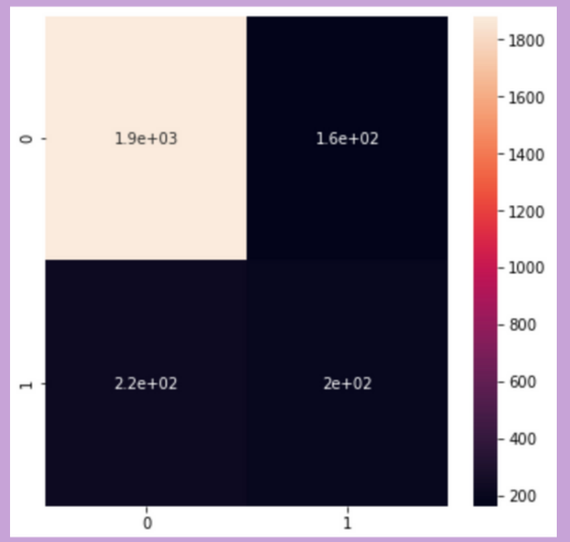
Our model



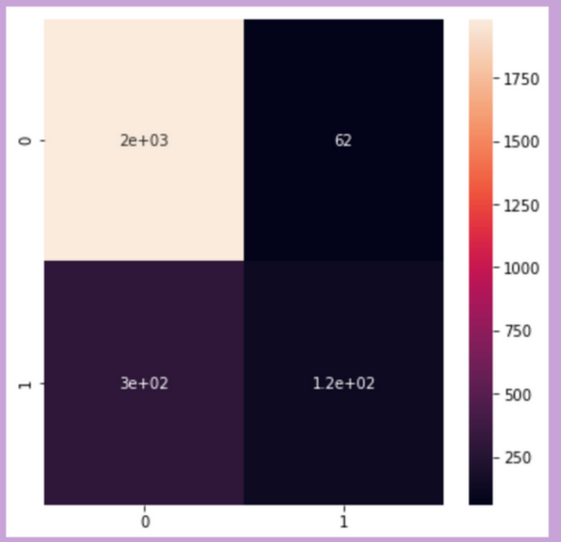
Support Vector Machine



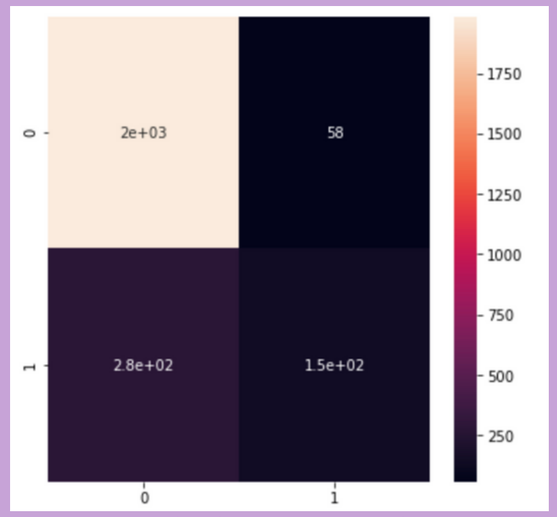
Naive Bayes



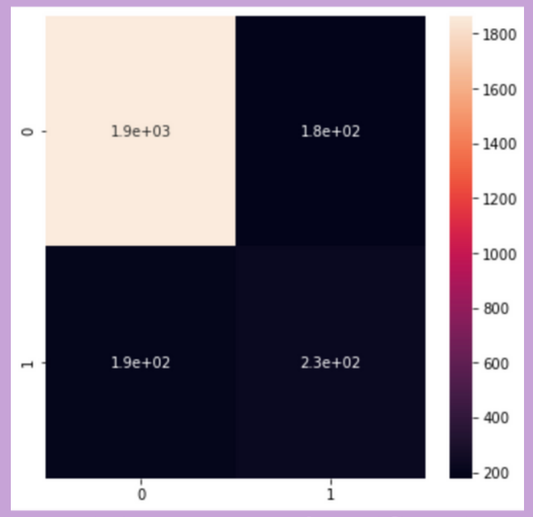
Random Forest Classification



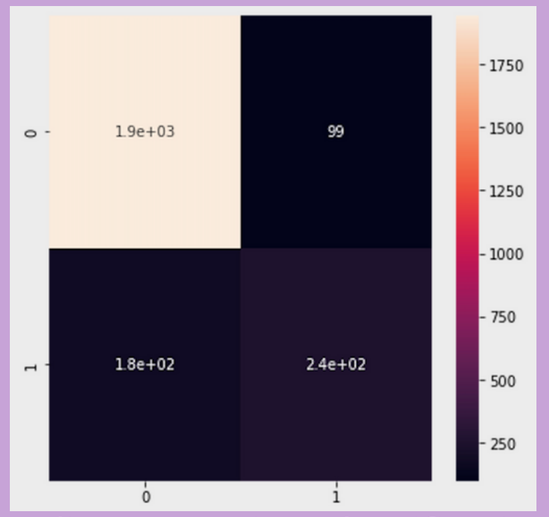
K Nearest Neighbour



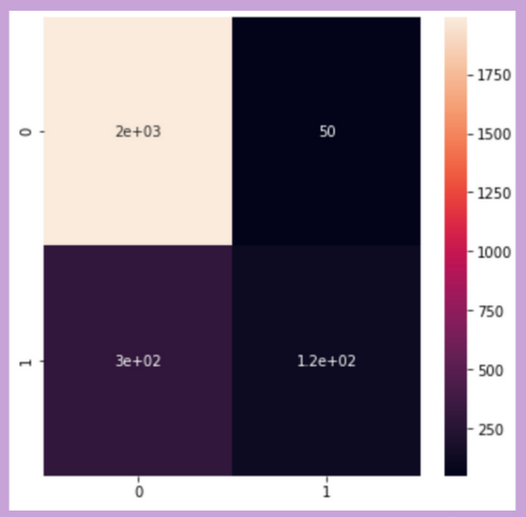
Logistic Regression



Decision Tree Classification



Gradient Boosting Classification

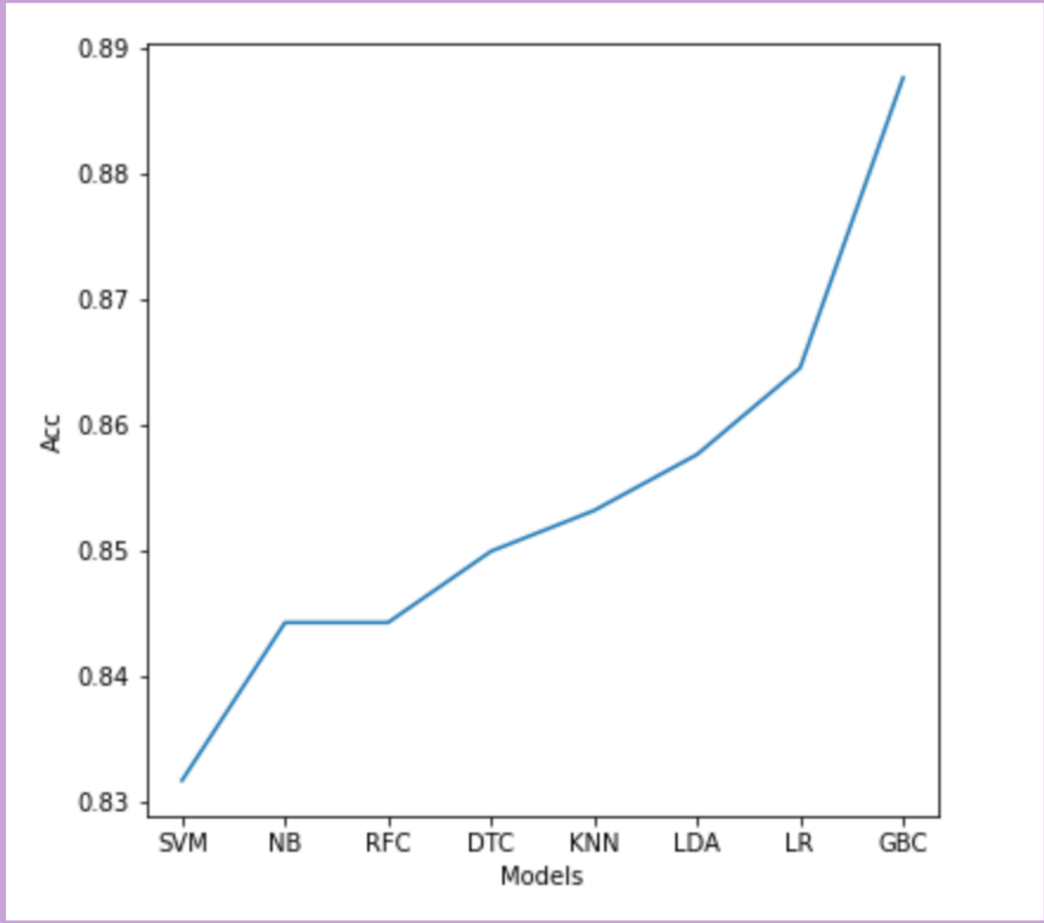


Linear Discriminant Analysis

Our model

Accurancy obtained according to each model studied:

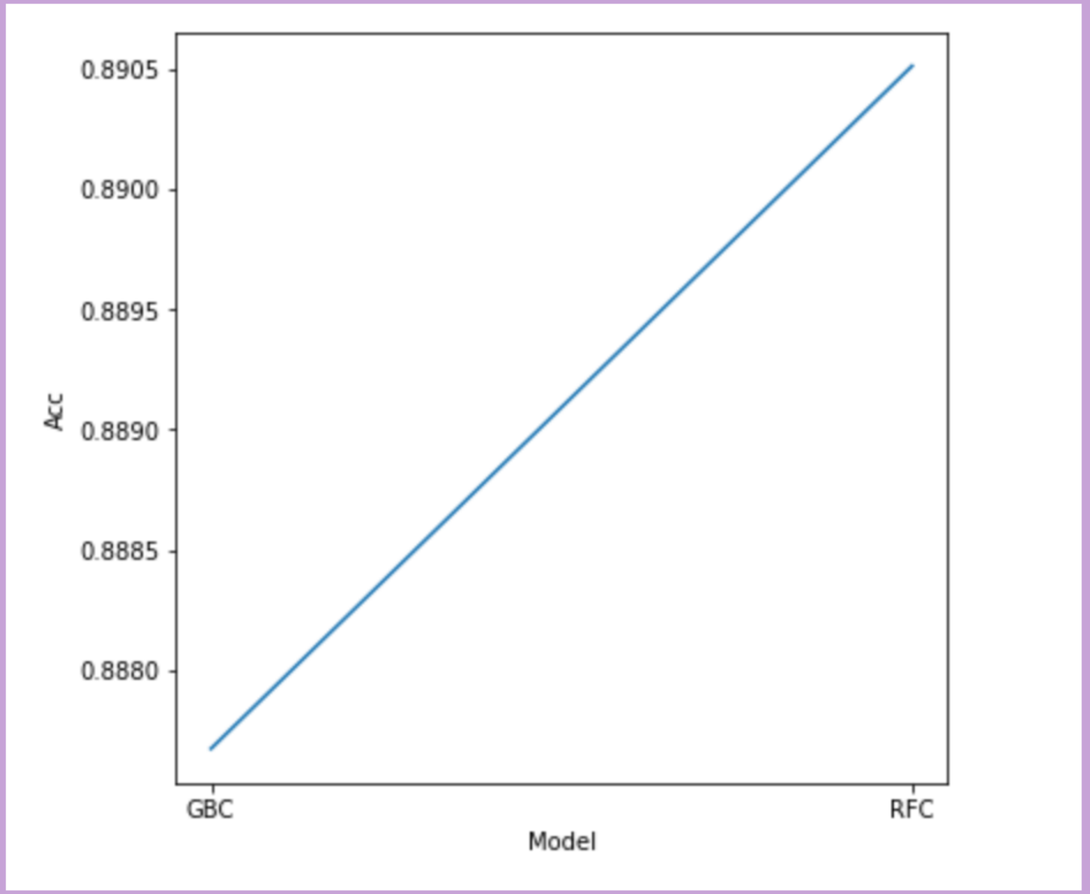
- Support Vector Machine: 0.832
- Naive Bayes: 0.844
- Random Forest Classification: 0.844
- K Nearest Neighbour: 0.853
- Logistic Regression: 0.865
- Decision Tree Classification: 0.851
- Gradient Boosting Classification: 0.888
- Linear Discriminant Analysis: 0.858



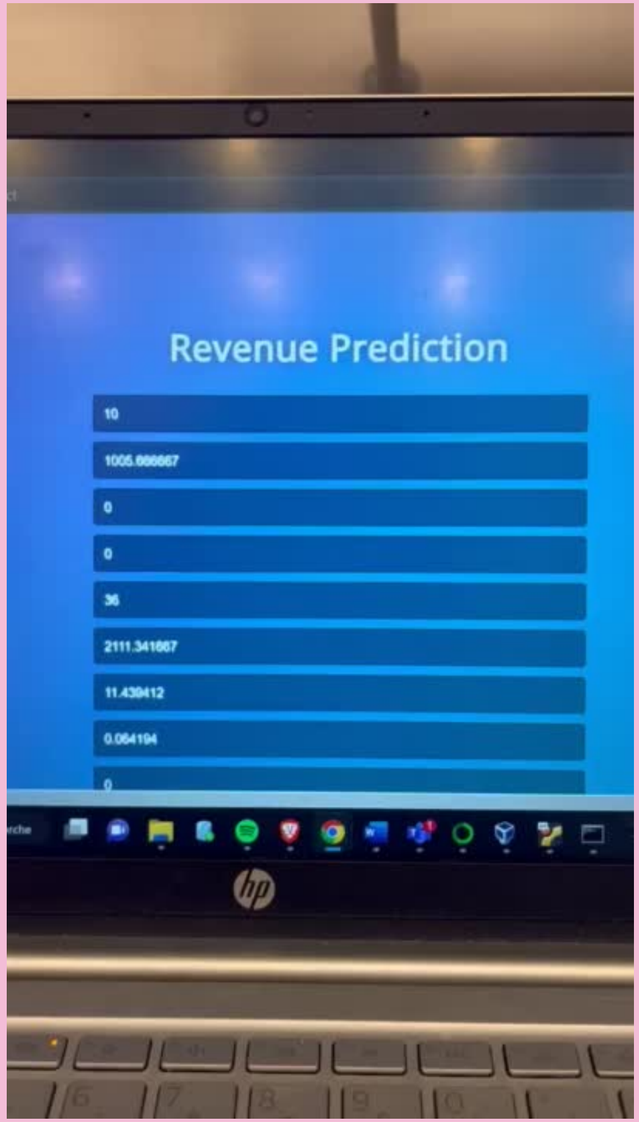
Our best model

Best accurancies:

- Random Forest: 0.891
- Gradient Boosting: 0.888



Graphic interface



THANK YOU