Analysis of a Marketing Campaign: From descriptive analyses to predictive modelling

Jacopo De Angelis

PhD Candidate

Scientific Consultant

Data Analysis & Data Science

Topics

- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

Topics

- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

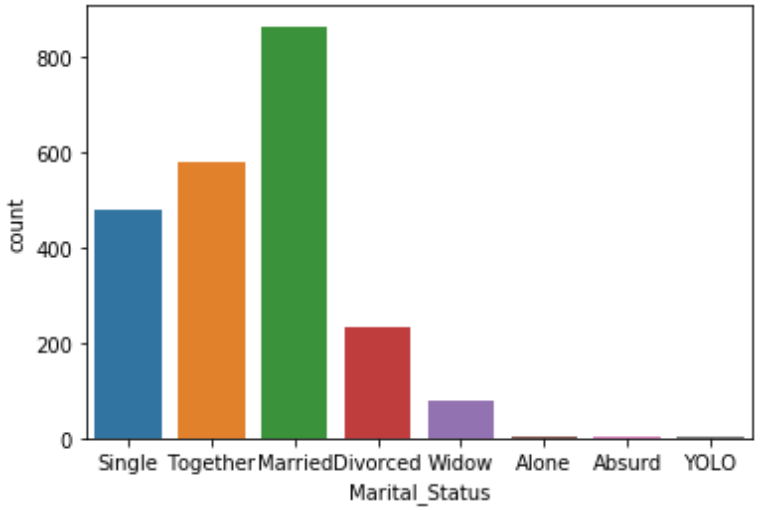
Aims and dataset description

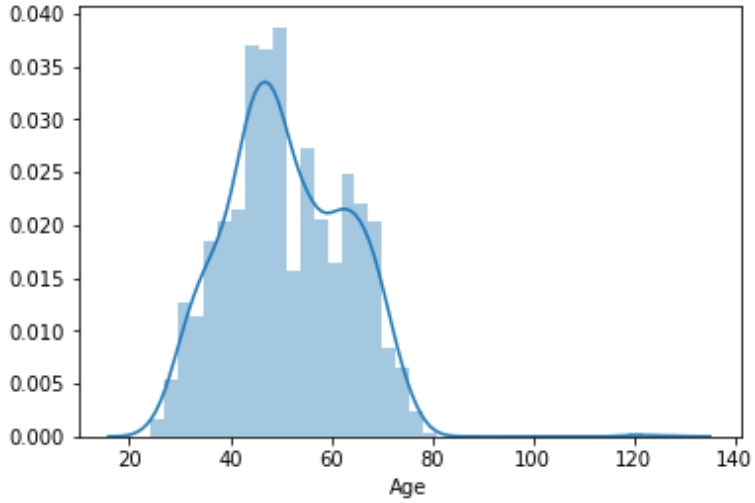
- The primary aim of this project is to develop models predicting whether a customer will respond, or not, to the marketing campaign;
- The secondary aim is to analyze which specific factors mostly predict customer responding or not to the campaign;
- The dataset has been taken from Kraggle (https://www.kaggle.com/rodsaldanha/arketing-campaign). It contains profile and online behavior user's data, 2240 rows x 29 columns.

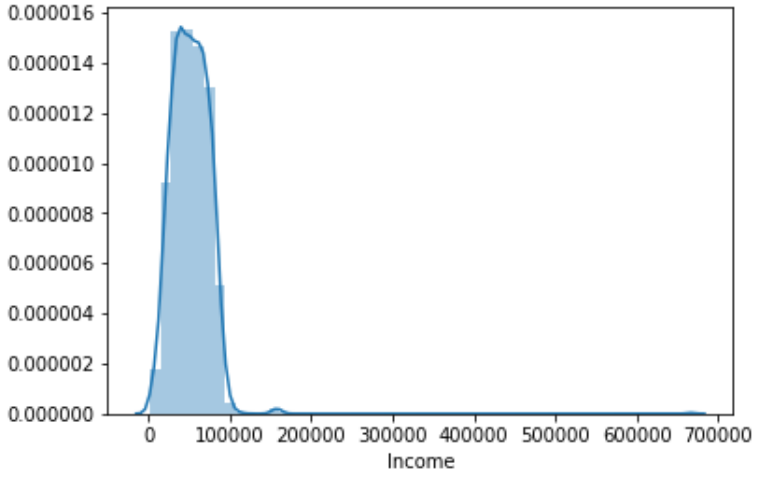
Topics

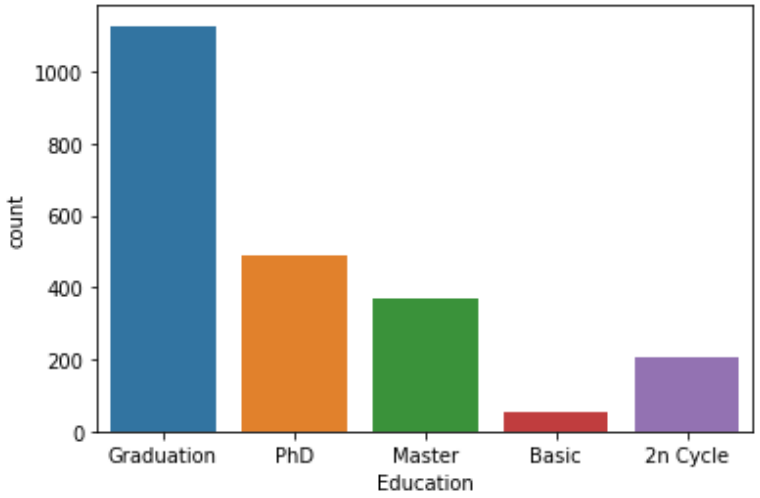
- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

Personal Info

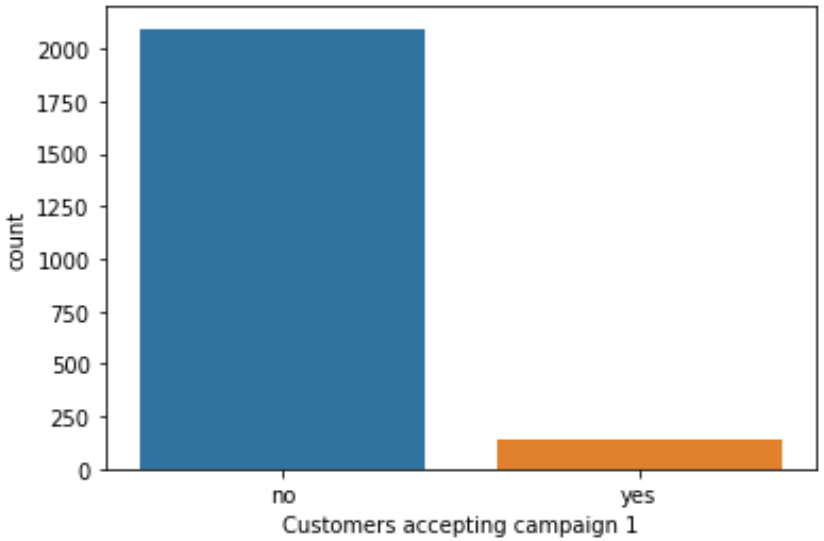


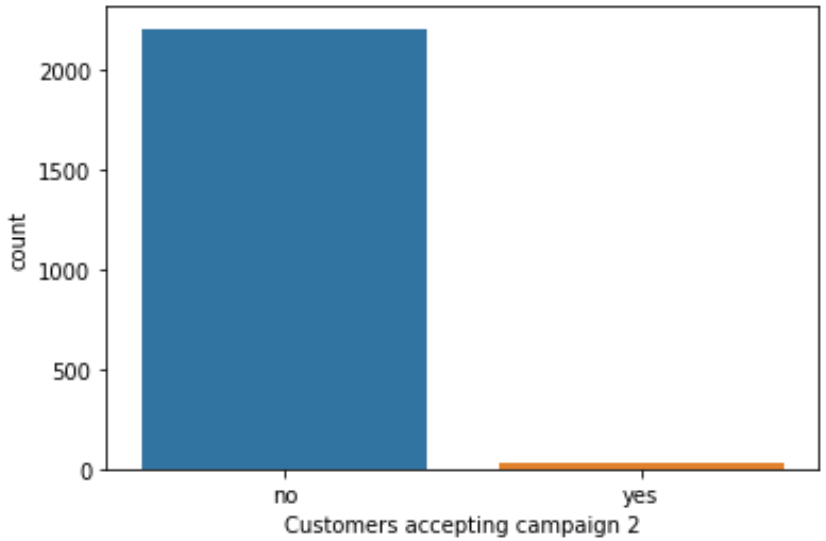


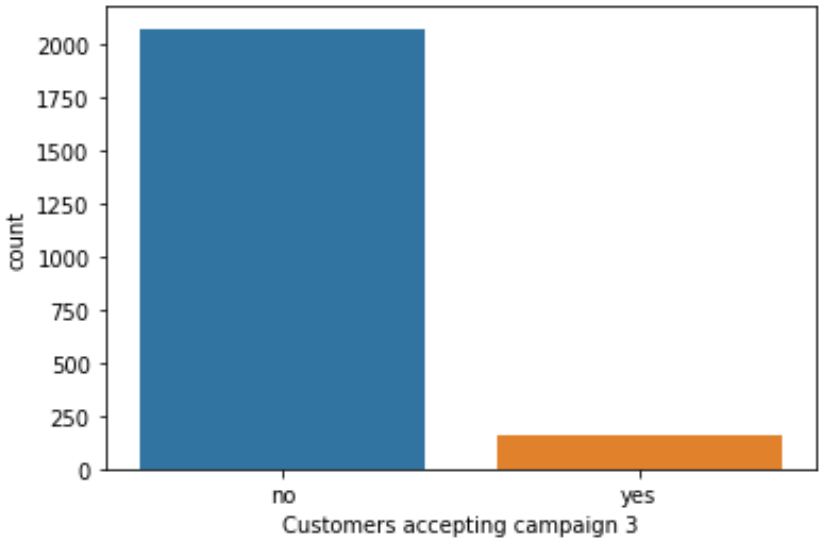


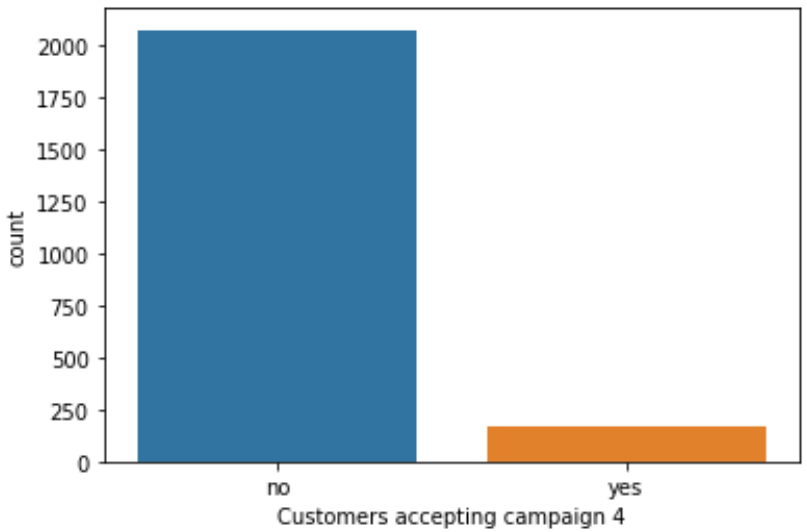


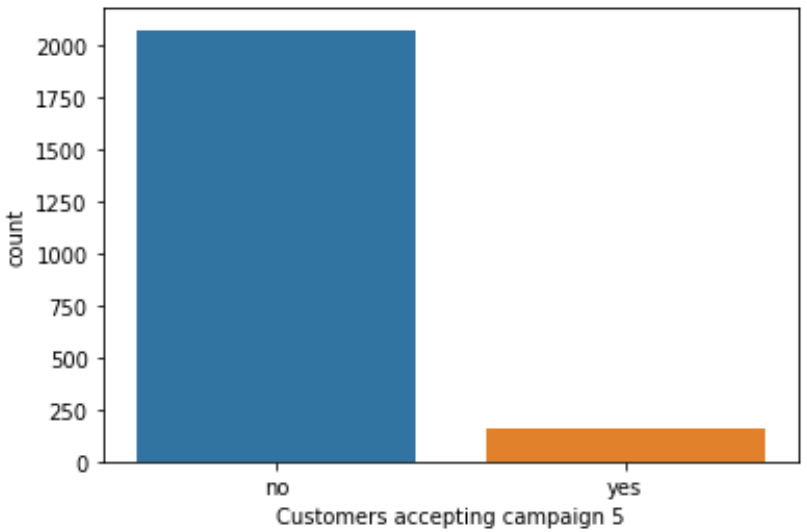
Responses to previous campaigns

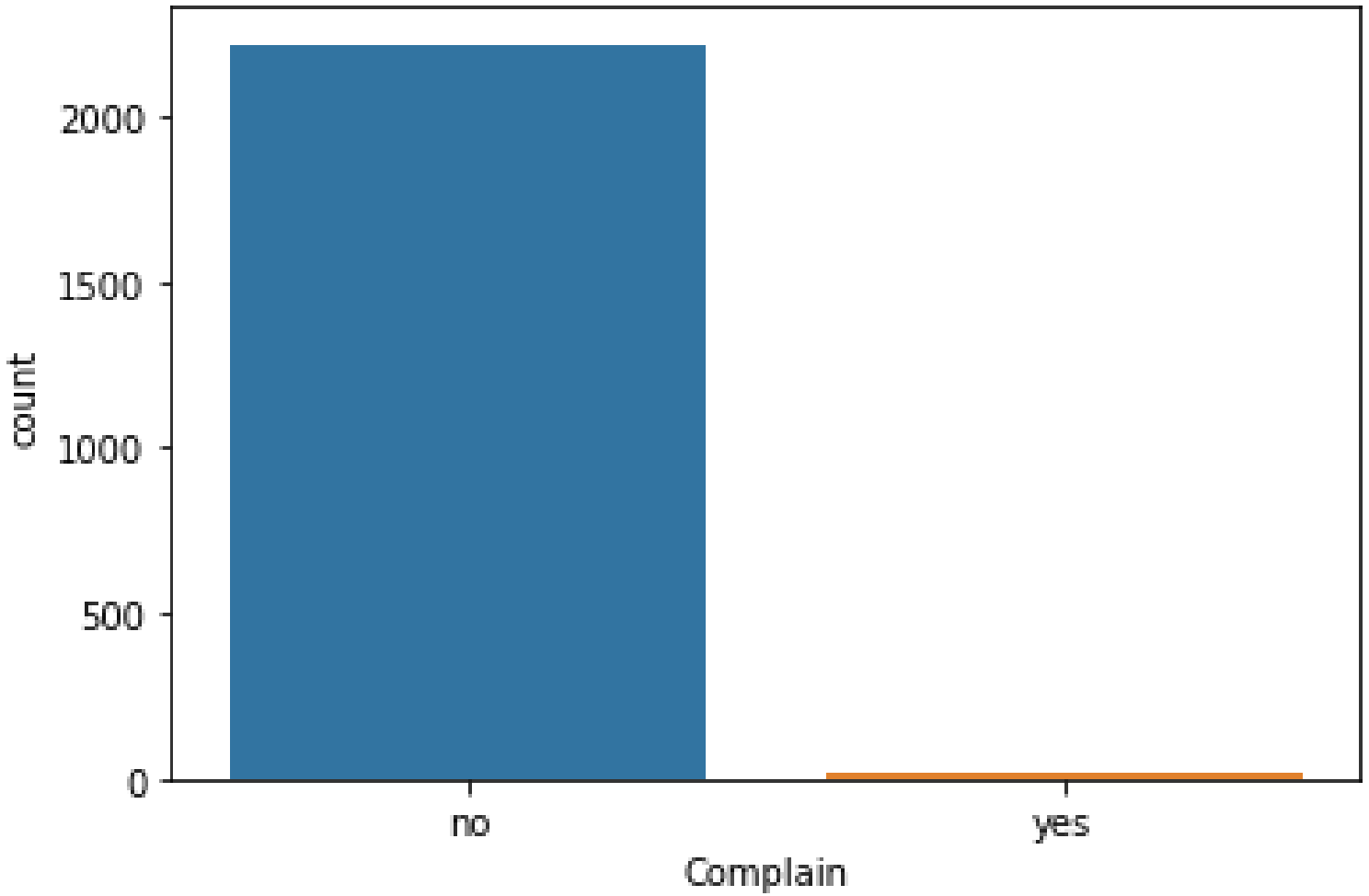






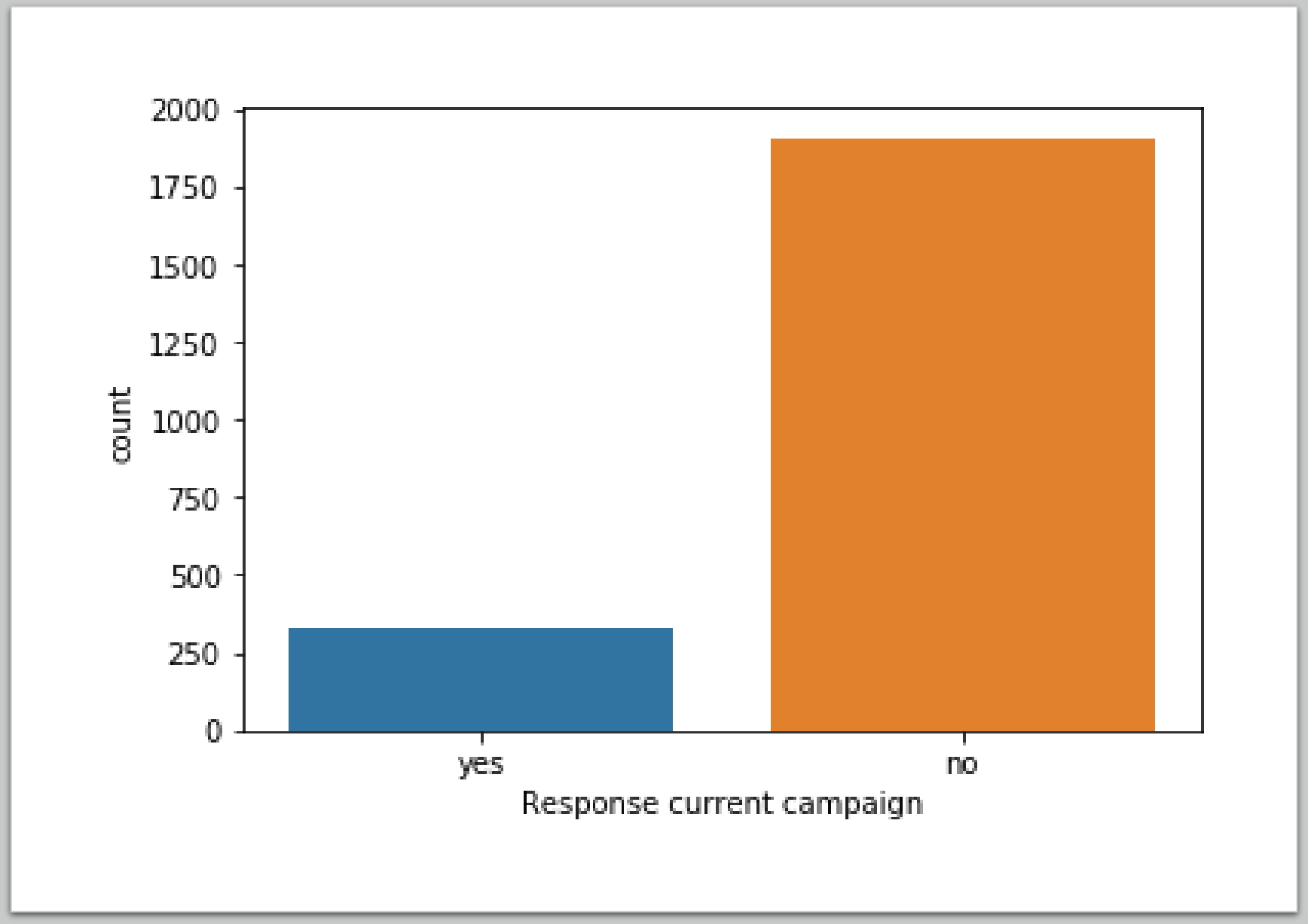




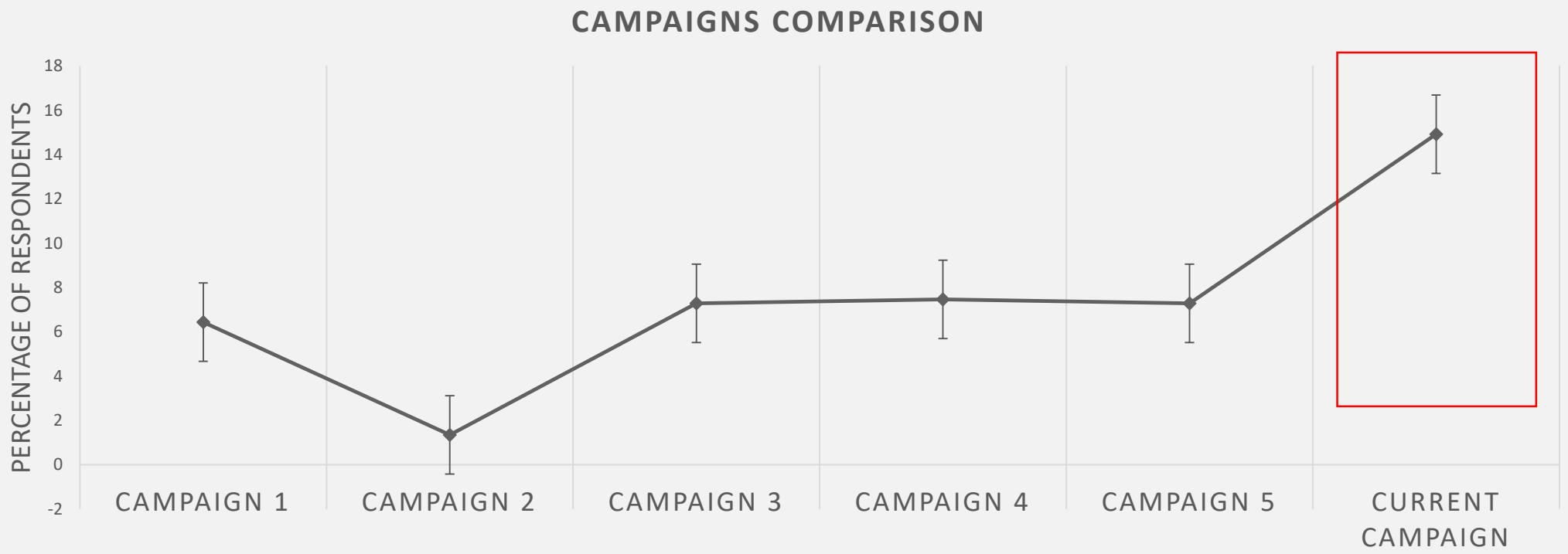


Frequency of complains

Number of responses to the current campaign



Comparing current campaign outcome with previous campaigns outcomes

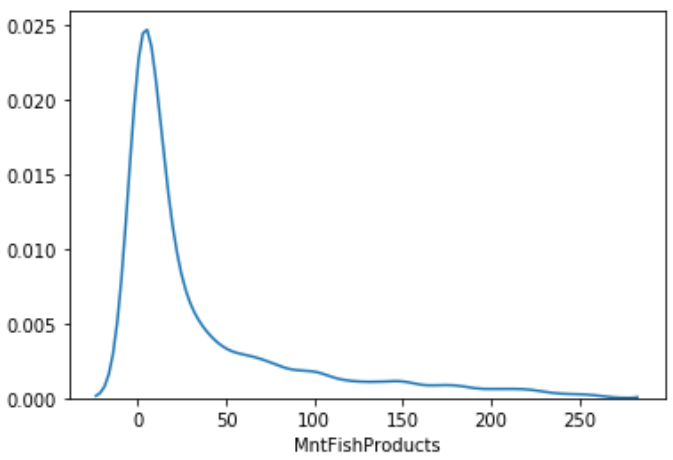


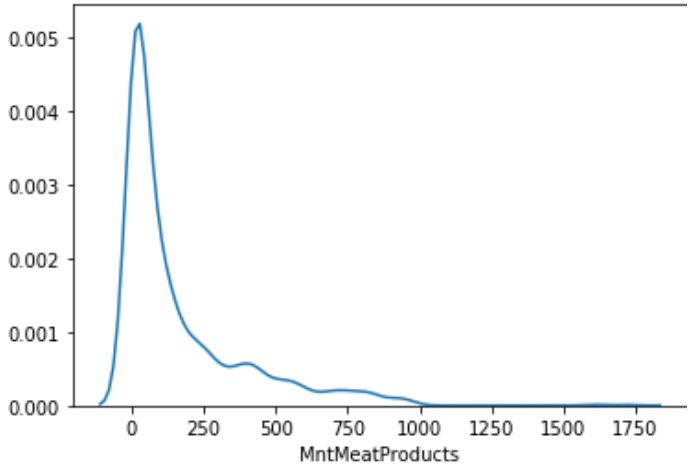
Binomial Test comparing current campaign proportion with the average of previous campaigns' proportion significant (p <.001)

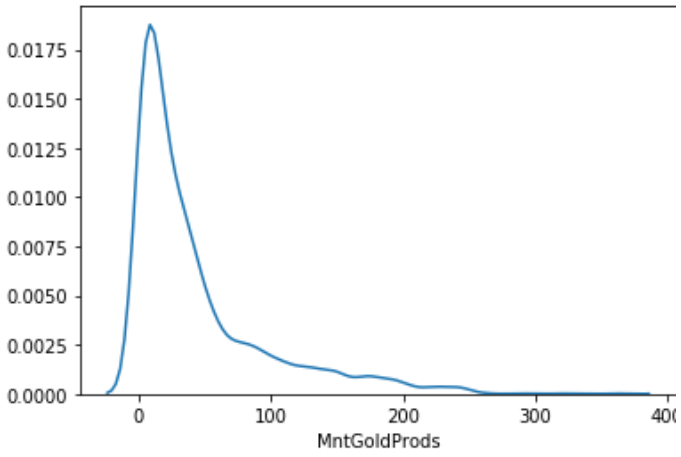
Amount spent on products in the last 2 years

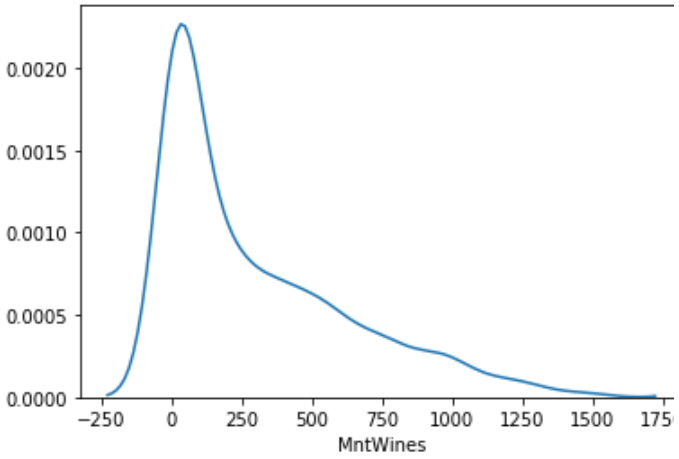
Amount spent on:

- Fish
- Gold
- Meat
- Wines





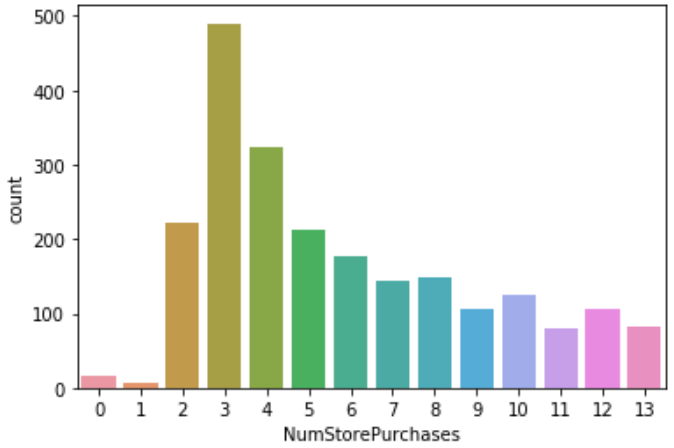


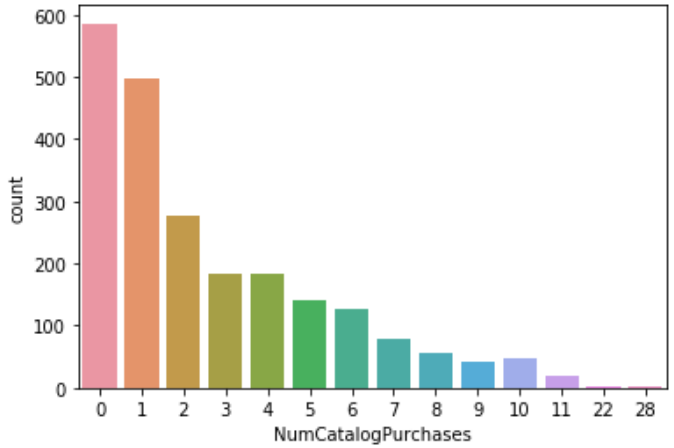


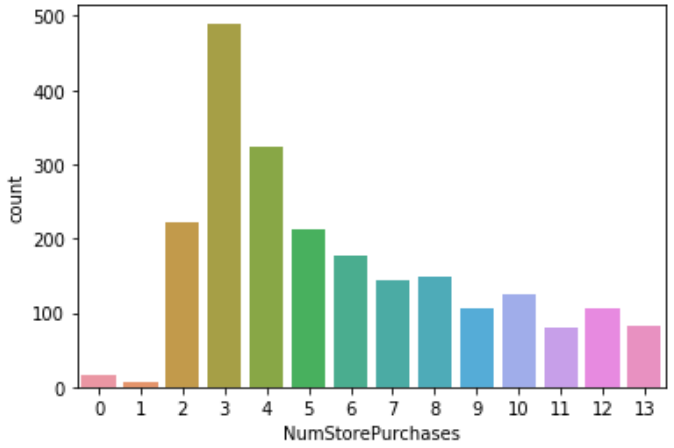
Number of Purchases

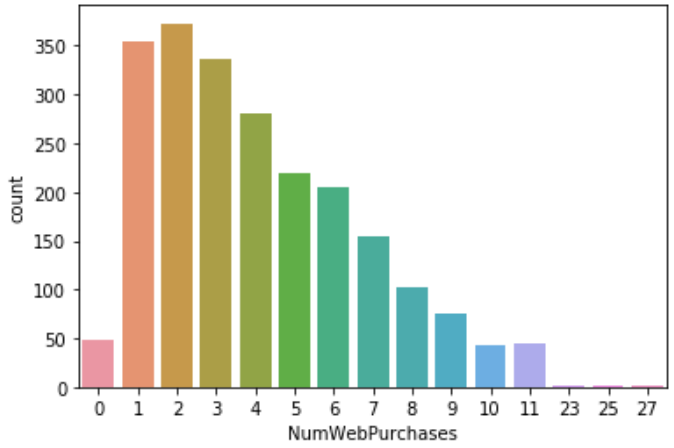
Options:

- Catalogue
- Discount
- Online
- Store

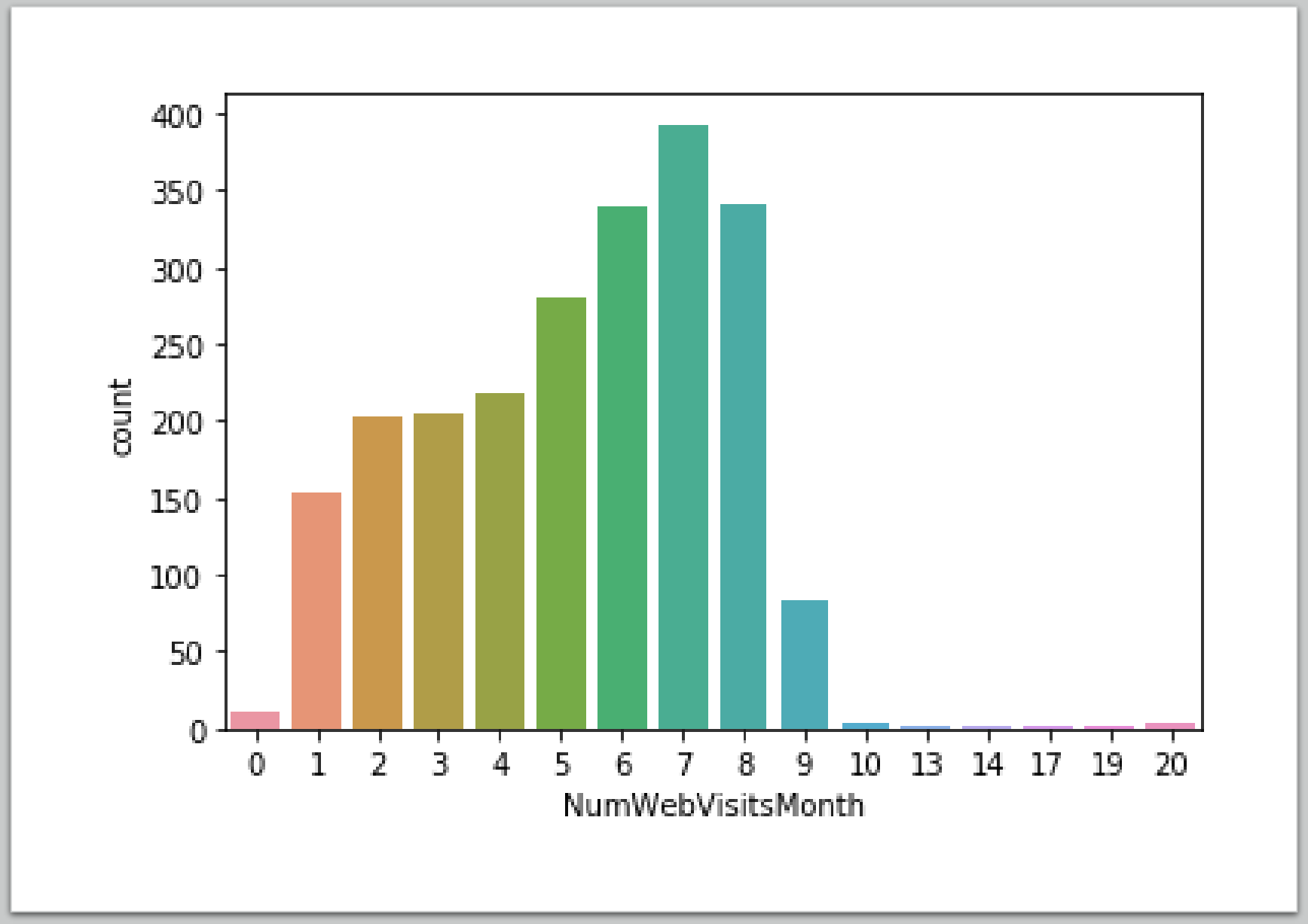








Number of web visits per month

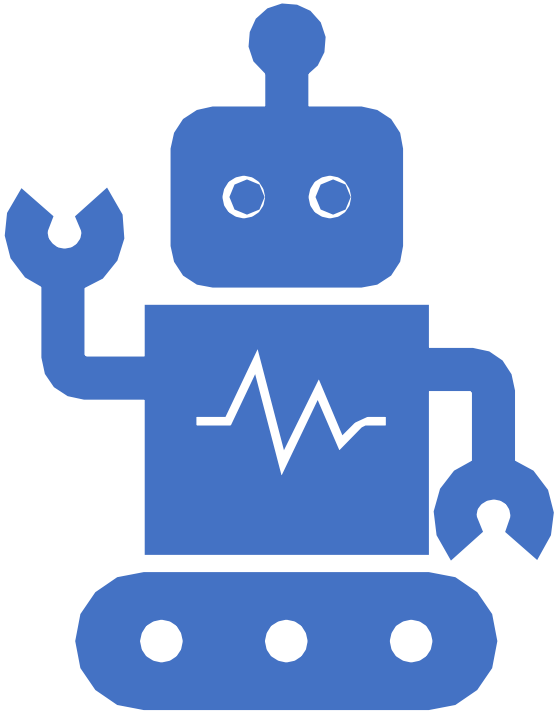


Topics

- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

Steps

- Checking and filling missing values (i.e., imputer strategy);
- Feature engineering (i.e., computing age, transforming education and marital status);
- Feature scaling (i.e., standardization);
- Input selection (i.e., selecting most relevant features through recursive feature elimination methodology);
- Training Test Split (test size = 0.3);
- SMOTE procedure to balance the dataset;
- Machine Learning algorithms



Selected model features/inputs

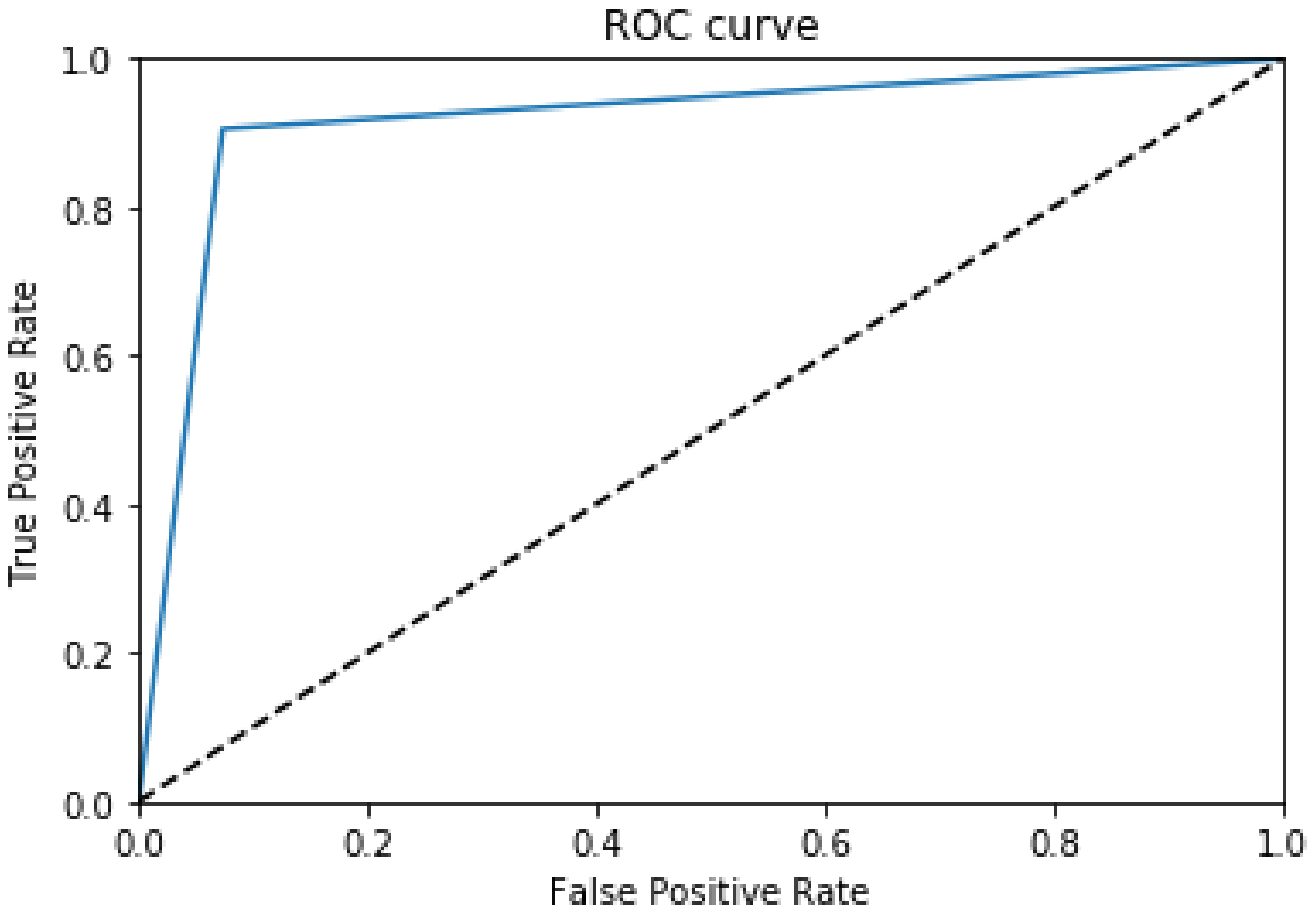
- Accepted Cmp 1 (i.e., respondents to campaign 1)
- Accepted Cmp 2
- Accepted Cmp 3
- Accepted Cmp 4
- Accepted Cmp 5
- Education
- Marital Status
- MntMeatProducts (i.e., Amount spent for meat products)
- NumStorePurchases (i.e., number of purchases at the store)
- NumWebVisitsMonth (i.e., number of visits on the website)
- Recency (i.e., number of days since the last purchase)
- Teens at home

ML alghorithms performances

Model	Precision	Recall	F1-score	Accuracy
SVM*	0.89	0.93	0.90	0.90
Decision Tree	0.88	0.91	0.89	0.90
Random Forest*	0.92	0.91	0.91	0.92

**Grid Search applied for hyperparameter optimization*

ROC Curve Random Forest

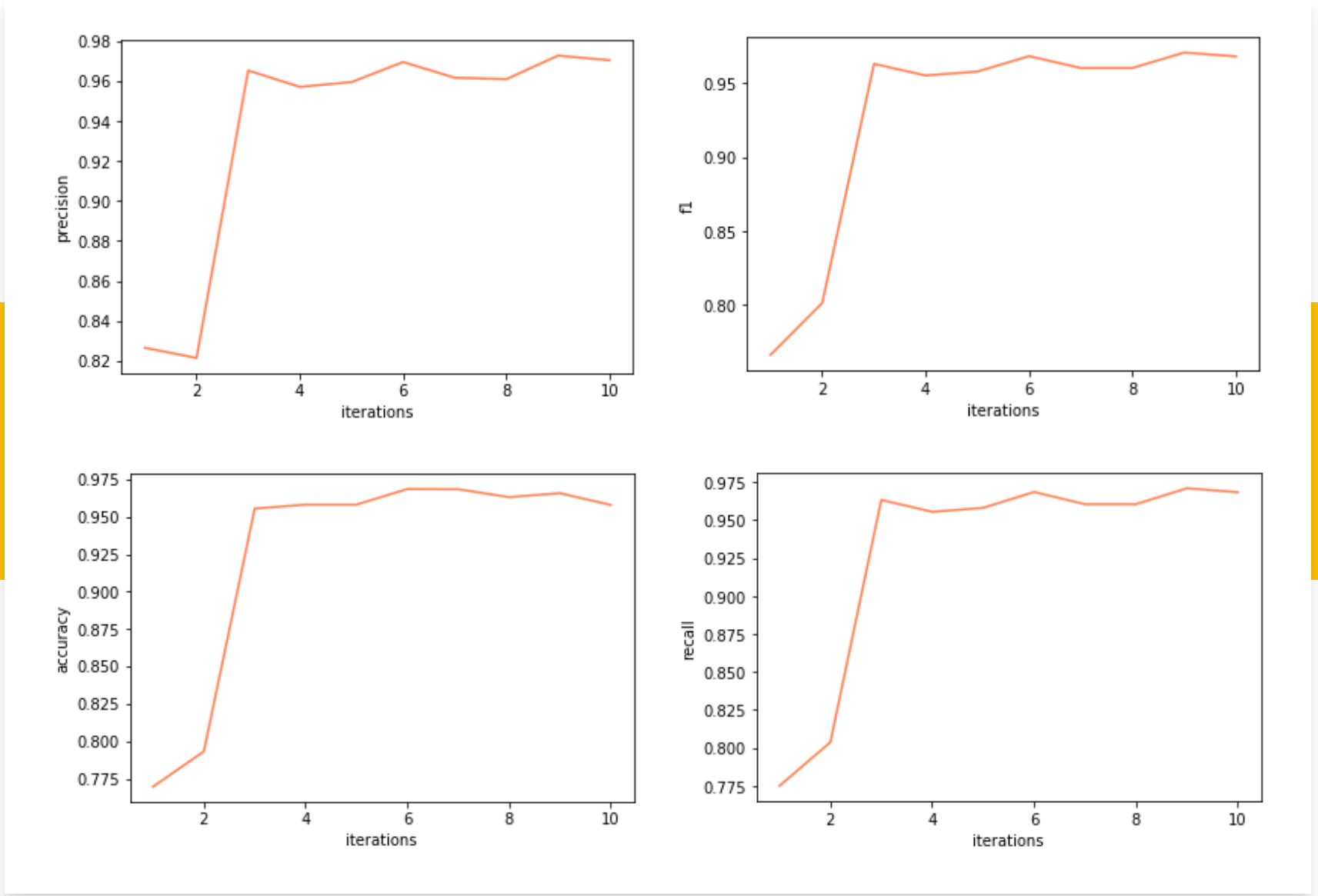


Area Under the Curve (AUC) = 0.91

Topics

- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

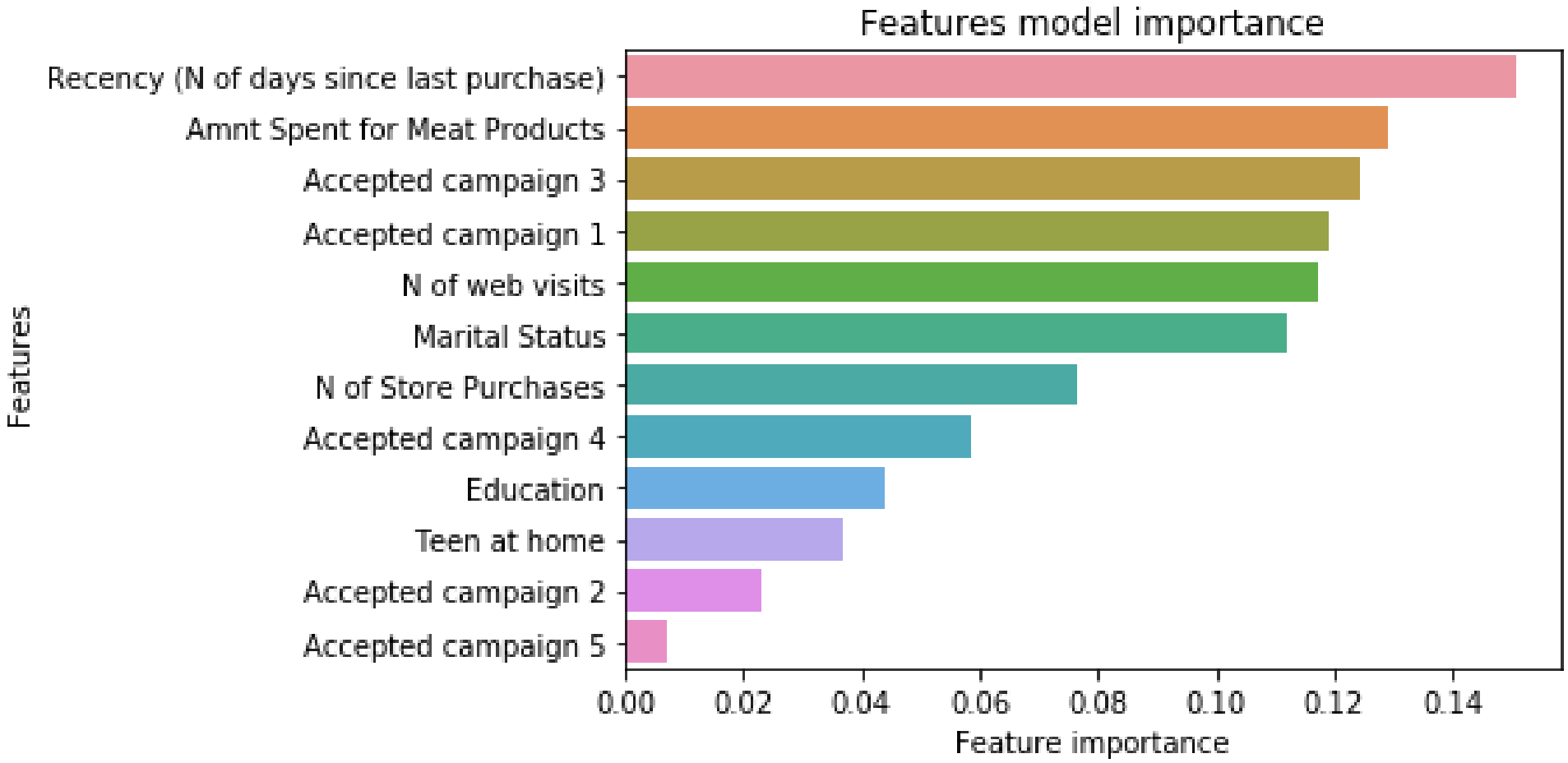
10 – fold Cross-validation



Topics

- Aims of the project and dataset description
- Data exploration and visualization
- Predictive modelling
- Model validation
- Most relevant predictors of campaing success

Features importance



Conclusions and future steps

- The models reached good predictive performance metrics (about 90 % accuracy);
- Random Forest was the best performing alghorithm;
- Random Forest can be tested for model deployment predicting consumer behavior;
- The number of days since last purchase, the amount spent for meat products, and behaviour in previous campaigns are the most substantial predictors of consumer response to the current marketing online campaign.