Bank Marketing Campaign

Tony Nguyen

4/16/2024

Background & Objective

- ABC Bank wants to sell its term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

- Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing, etc.) can focus only to those customers whose chances of buying the product is more.

- This will save resource and their time (which is directly involved in the cost (resource billing)).

- Develop model with Duration and without duration feature and report the performance of the model.

Business Understanding

- The bank wishes to use an ML-model to shortlist customers who are more likely to buy their product so that their marketing channel can focus only to those customers who have a higher chance of buying their product.

- To create an ML-model, we use a potential customer's provided bank information to isolate significant variables that help determine whether a customer is a good candidate.

Data Understanding

Data Type

• We have clean data with an assortment of numeric, categorical, and binary data. It is worth noting that the output variable, or predictor variable, "y" is binary.

Data Issues

• We see unknown values.

• We see outliers and skewness in several variables. We also see a potential error in one variable

• It is worth noting that despite being outliers, these values are crucial as the data is sensitive which means the outliers could provide useful information.

Data Solution/Approaches

• We have three issues we need to solve for our data: unknown values, outliers, and skewness.

• Unknown values: Will be deleted as they are missing values.

• Outliers: As stated above, outliers in our case were deemed noteworthy. It is best not to remove them.

• Skewness: If we calculate the natural logarithmic value of the data, we can normalize the distributions.

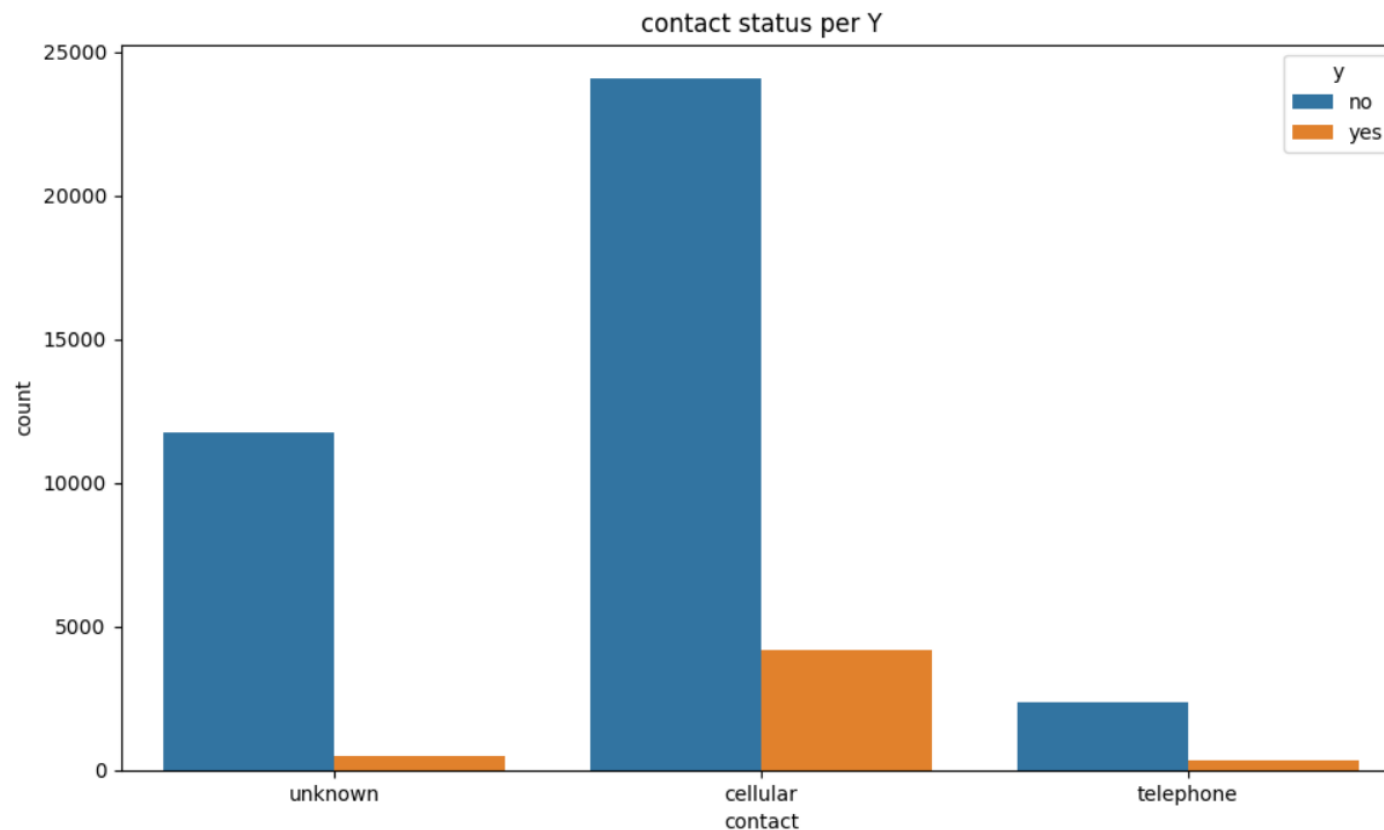|	
Total number of observations	**45211**
Total number of files	1
Total number of features	16
Base format of the file	.csv
Size of the data	4.503 MB

Duration

duration: last contact duration, in seconds (numeric)
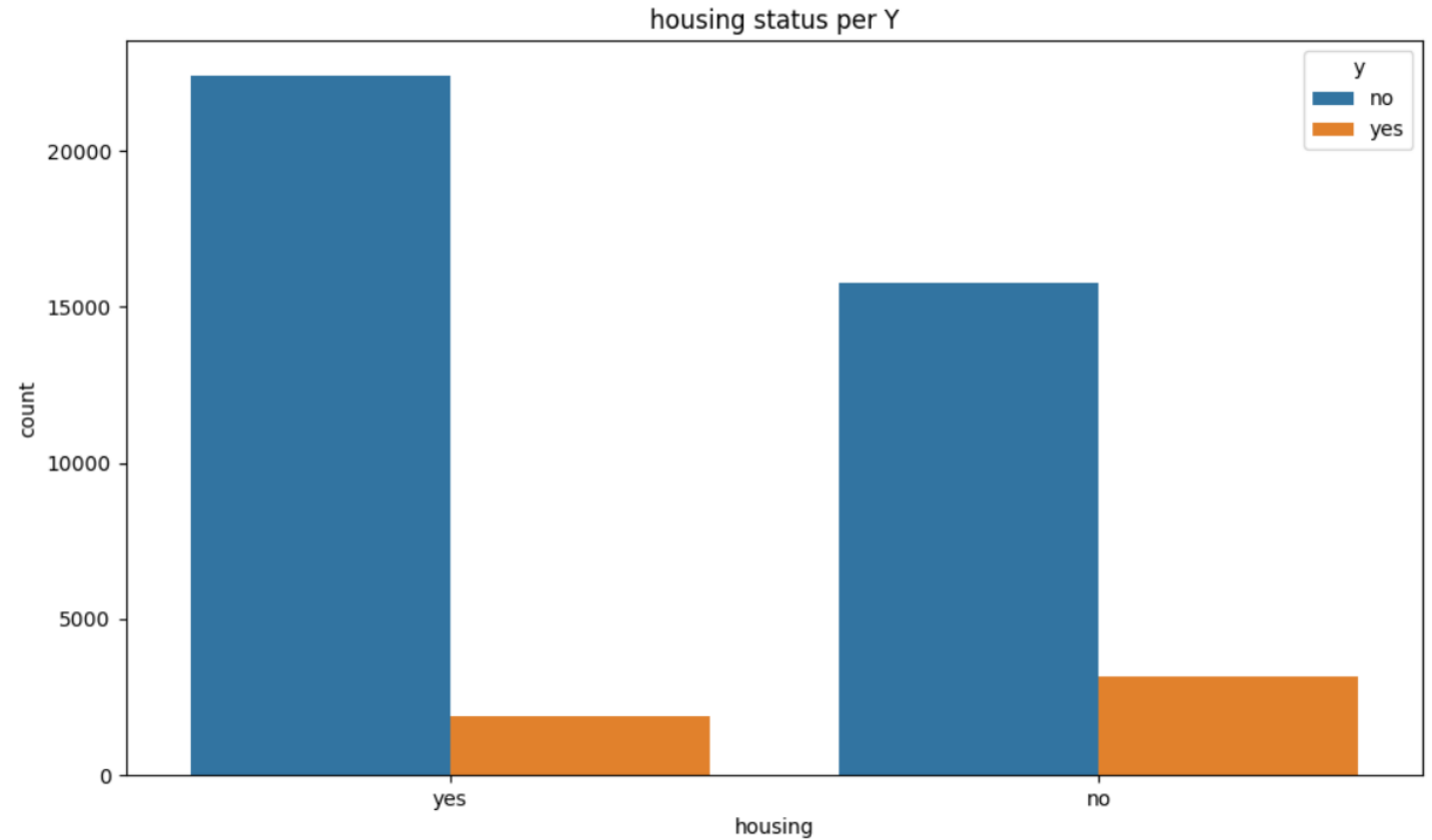

Normalized Duration
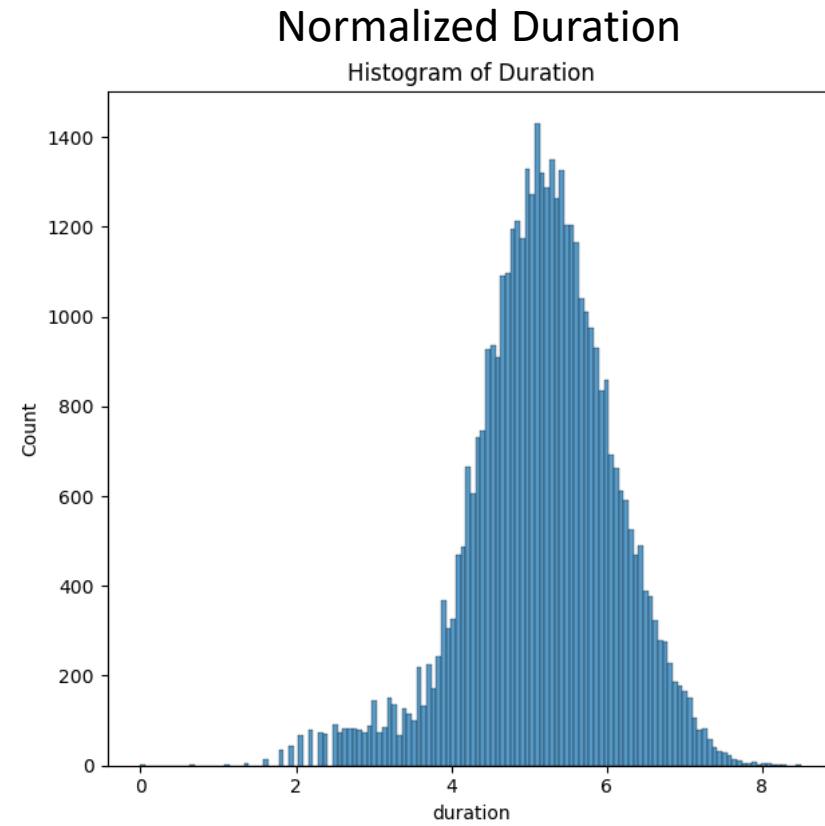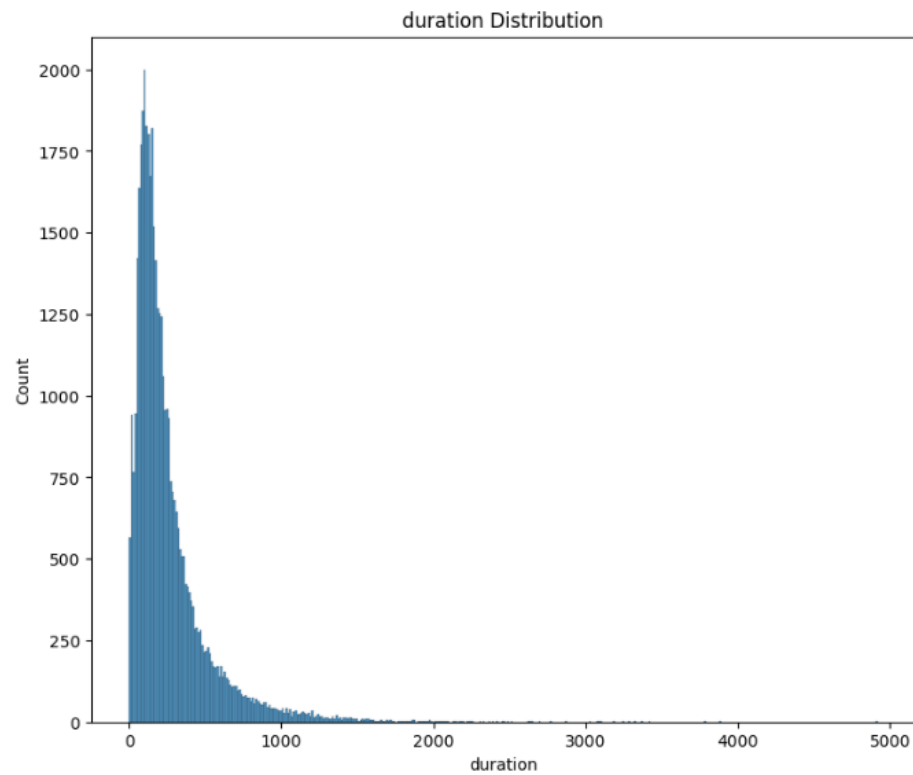
Contact

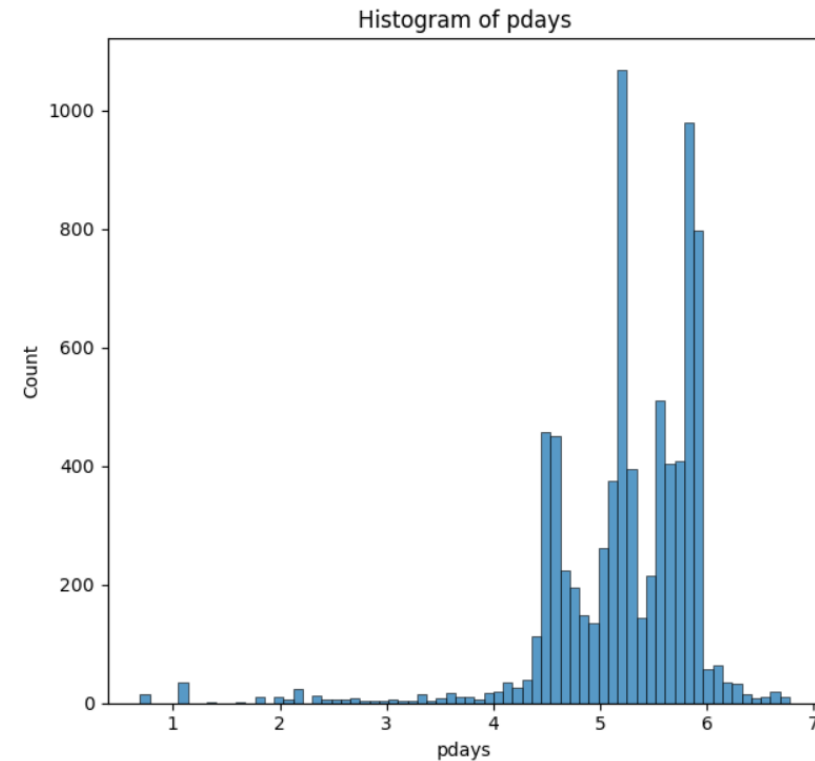contact: contact communication type (categorical: "unknown", "telephone", "cellular")
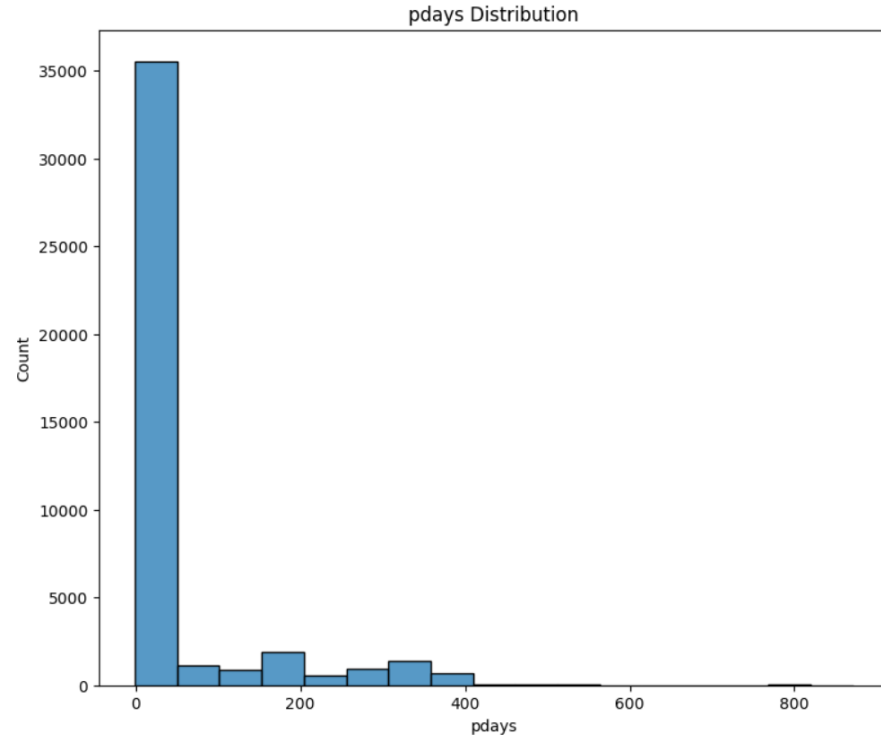
Housing

housing: has housing loan?
(binary: "yes", "no")

pdays

- pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric, -1 means client was not previously contacted)

Normalized pdays

Proposed Solutions

To increase customers that make a deposit:

Duration: Increase follow up contact (Limitation shown in pdays)

Contact: Use cellular to contact customers.

Housing: Focus on customers without a housing loan.

pdays: Do not contact customers in an excessive amount.

Technical Page

Model Building

- Logistic Regression: 72.03981837233671%

- Decision Tree Classifier: 95.96140412155081%

- Random Forest Classifier: 96.62504366049598%

- XGBoost: 92.44673419490046%

- AdaBoost Classifier: 83.7190010478519%

Conclusion:

Decision Tree Classifier and Random Forest Classifier yielded the best accuracy while the other models, such as logistic regression and AdaBoost Classifier, performed poorly in comparison.

To conclude, Random Forest Classifier had the highest accuracy of the various models.

Thank You