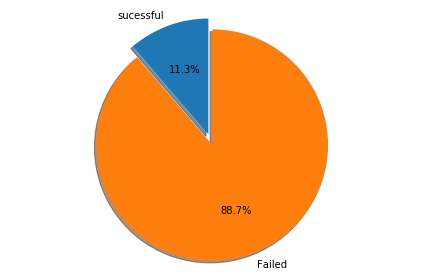
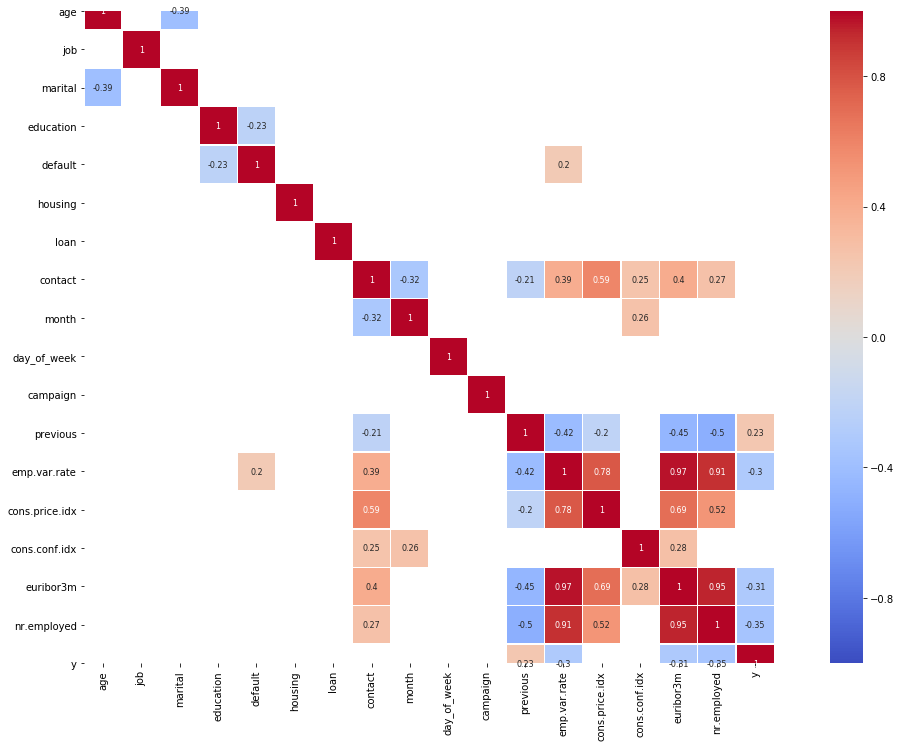
**Exploratory Data Analysis and Data Visualization**

The need for this project arose because of the very low subscription rate to ‘y’ of the term deposit. Only 11.3% of bank customers subscribed, and the EDA will give an initial indication of customers most likely to subscribe. This will be further analyzed using various machine learning techniques.

This document shows some of the techniques used in EDA for this project. For a detailed review of all features, see, the Capstone Project documentation (Jupyter notebook).

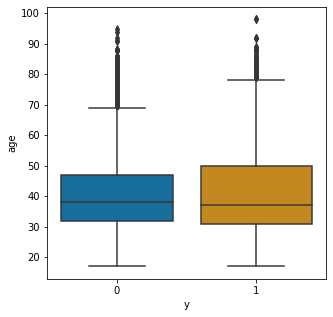
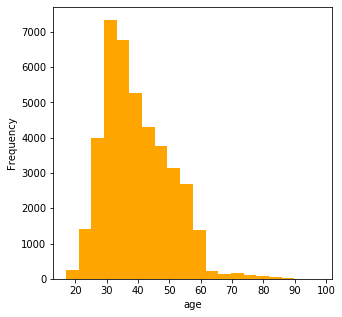
****

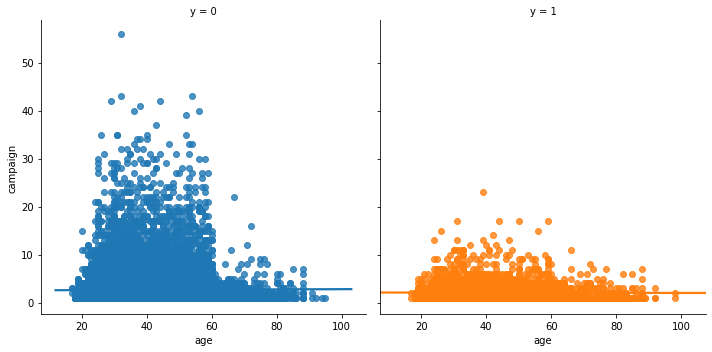
I used the heat map below as a quick way to evaluate correlation between different features and variable Y. The findings from this chart have been explained in detail using more detailed charts. A review of this chart reveals that the features with the greatest correlation to variable ‘y’ are, contact, previous, emp.var.rate, euribor3m and nr.employed.



**Age:**

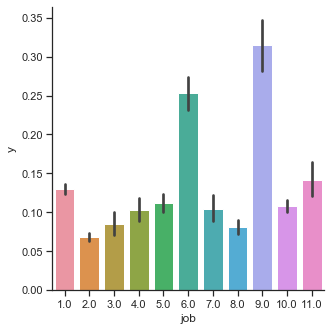
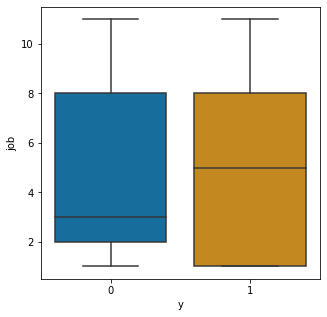
|  |
| --- |
| From the boxplot, the age of customers does not appear to have to have an impact on variable ‘y’. The Average age of subscribers ‘yes’ and ‘no’ is between 35 and 40 years old for both customers that subscribed, and those that did not. The distribution for Age is skewed to the left.  A review of the scatterplot for age and campaign indicates that most customers regardless of subscription and age, were contacted less than 10 times. |

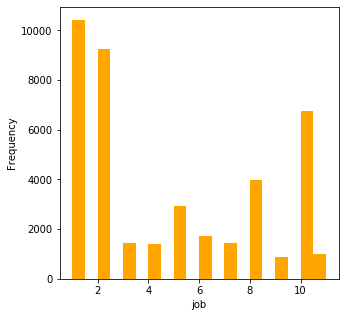
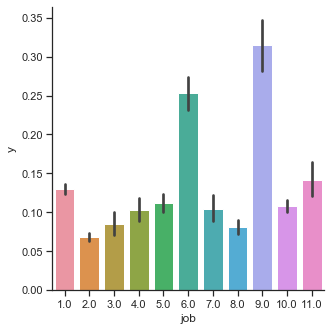


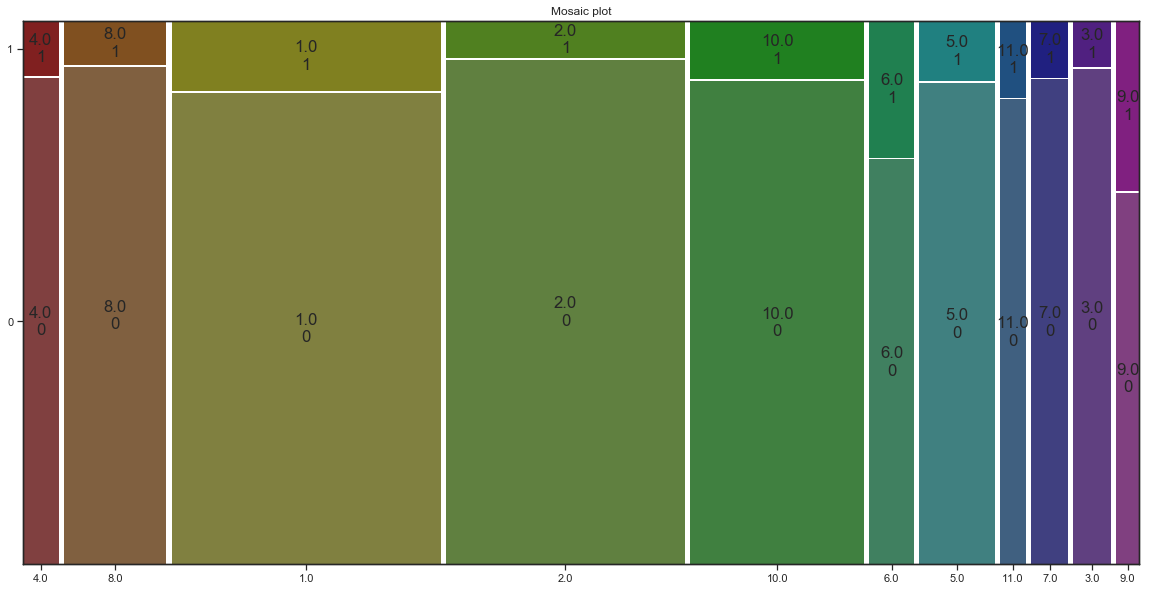


**Job:**

The type of Job that a customer does, appears to have an impact on whether they will subscribe to the term deposit or not(‘y’). Retirees and students in particular are more likely to subscribe, while blue collar (the second largest category, is least likely to subscribe.

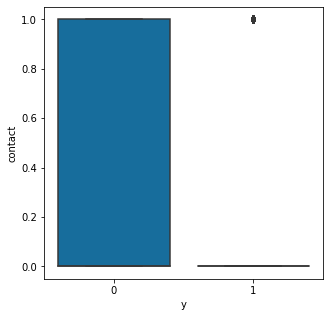
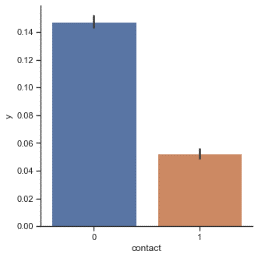
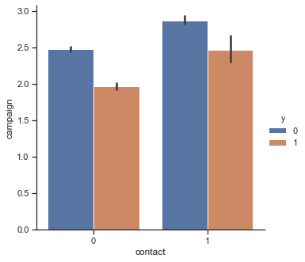






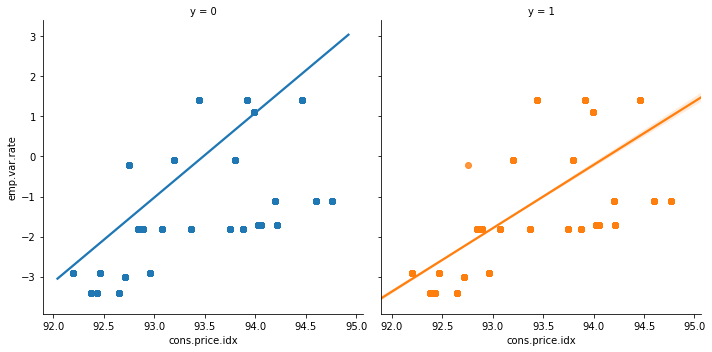
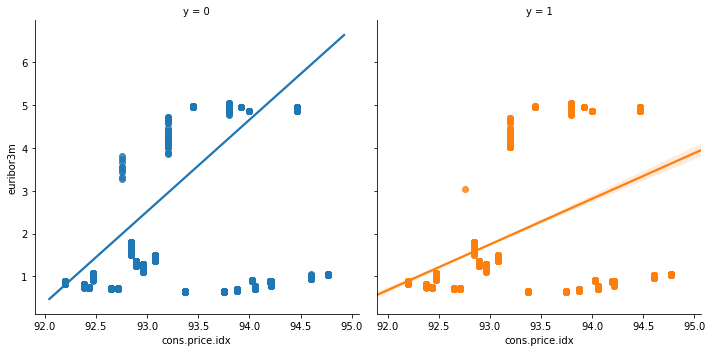
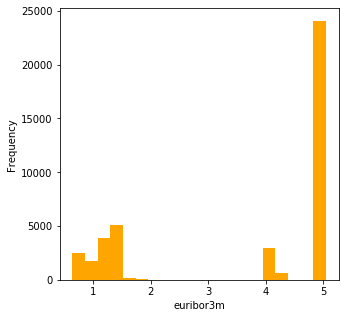
**Contact**

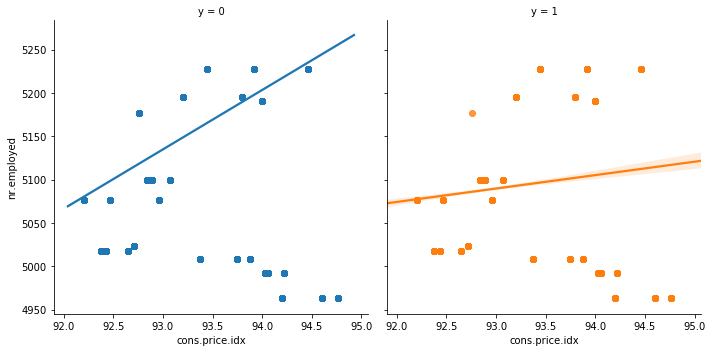
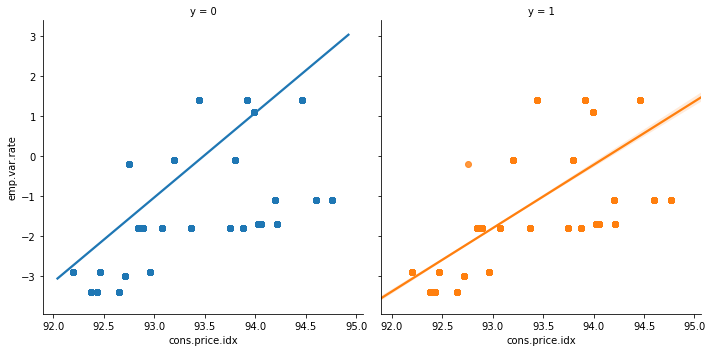
Of the customers contacted, the customers contacted by telephone appeared to be most likely to subscribe.

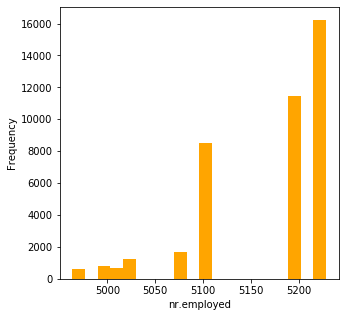


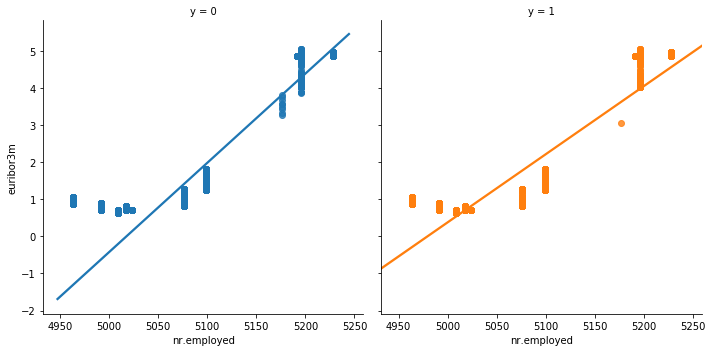
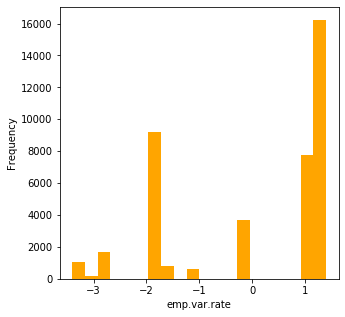
**cons.price.idx', emp.var.rate', euribor3m, nr.employed**

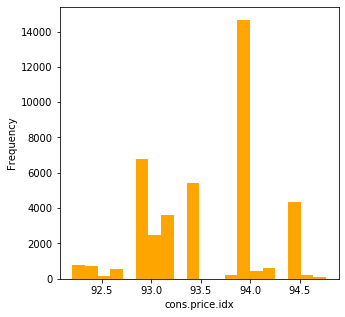
The features above have some level of correlation, and were all multimodal, while all the other features (as shown in the Jupyter notebook of this exercise), did not appear to have any meaningful correlation











APPENDIX

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Level** | **Education** | **Month** | **Day** | **Job** | **Marital** | **Housing** | **Loan** | **Default** | **Contact** | **y** |
| 0 | Illiterate | mar | mon |  | Divorced | No | No | No | Cellular | No |
| 1 | basic.4y | apr | tue | admin | Single | Yes | Yes | yes | Telephone | Yes |
| 2 | basic.6y | may | wed | blue-collar | Married |  |  |  |  |  |
| 3 | basic.9y | jun | thu | entrepreneur |  |  |  |  |  |  |
| 4 | high.school | jul | fri | housemaid |  |  |  |  |  |  |
| 5 | Professional.course | aug |  | management |  |  |  |  |  |  |
| 6 | university.degree | sep |  | retired |  |  |  |  |  |  |
| 7 |  | oct |  | self-employed |  |  |  |  |  |  |
| 8 |  | nov |  | services |  |  |  |  |  |  |
| 9 |  | dec |  | student |  |  |  |  |  |  |
| 10 |  |  |  | technician |  |  |  |  |  |  |
| 11 |  |  |  | unemployed |  |  |  |  |  |  |