----------Fatemeh Kiaie----------

Origination of Data Set

The selected data set is downloaded from UCI website. Data set is about direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. In the data set, 20 features are captured for 45,211 entities.

Data Cleaning

In order to clean data set, the missing data needed to be handled. Fortunately, this data set does not have any missing value.

Knowing the demographic information of the input data

The input columns are basically related to age, type of the job, marital status, education level, financial status of the customer, month that the customer been contacted and the duration of each call. Fig.1 shows the age density of the clients that been contacted, the age of contacted customers is between 18 to 95 years old and the average is 40.9. The most frequent contacted client age is 33 years old. As can be seen in Fig. 2, the clients are in 12 different job categorization and the bank targeted mostly the “blue collar” and “managers” to offer the new product. It is also evident that the majority of contacted customers are well-educated. From Fig. 4 and 5, the majority of targeted customers re not paying loan but they are owning house. It is evident that the bank are contacting the more stable customers, both in financial and family point of view.

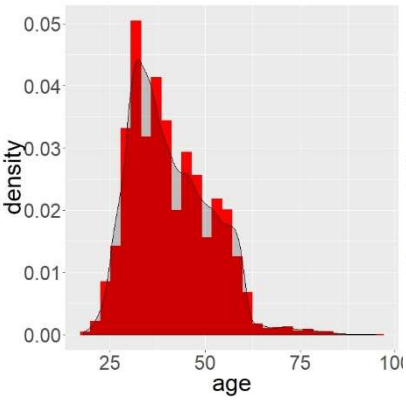

Fig.1. Age density

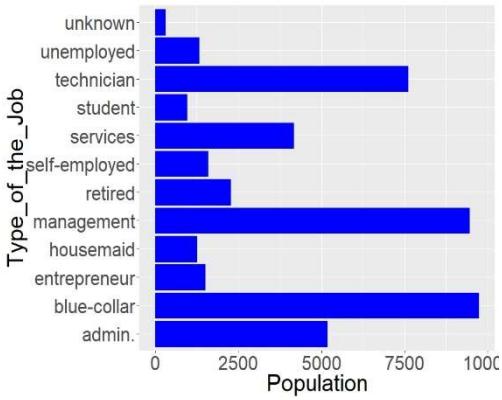

Fig.2. Type of the job

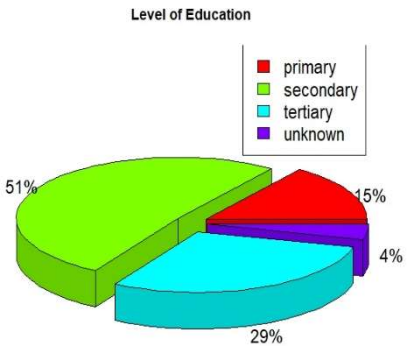

Fig. 3. Level of education

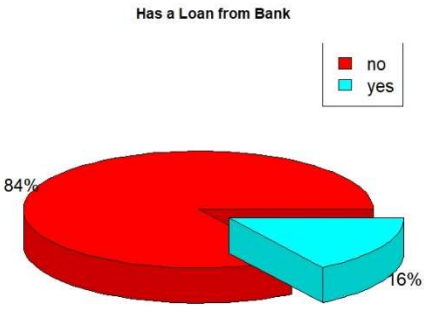

Fig. 4. Percentage that have loan

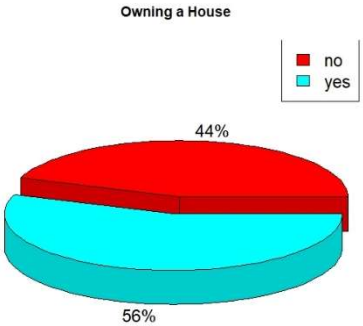

Fig. 5. Owning house loan

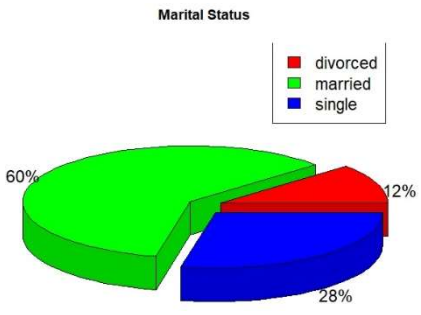

Fig. 6. Marital status

Moreover, Fig. 7 illustrates the month of the year that the contact been made and as can be seen majority of the call been made in May and minimum number of calls happened in December. Duration of calls in Fig.8, showing that the average call lasts 258.2 second and maximum call is 4918 second. Number of calls for the client are shown as the number of campaign in Fig. 9. The maximum number of calls is 63 trial and the average number of calls are 2.76.

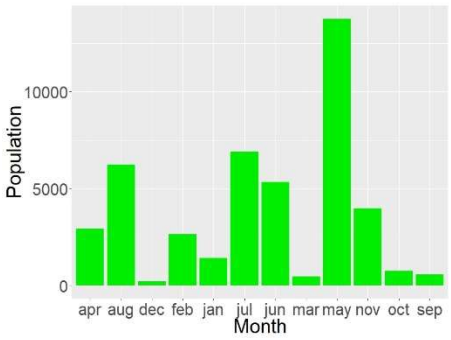

Fig. 7 Month that call been made

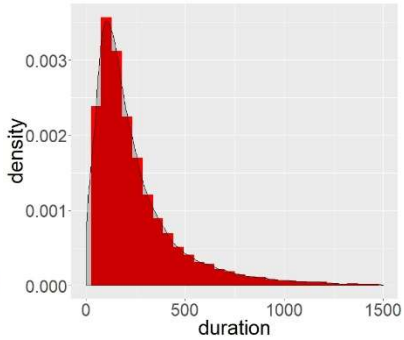

Fig. 8. Duration of call density

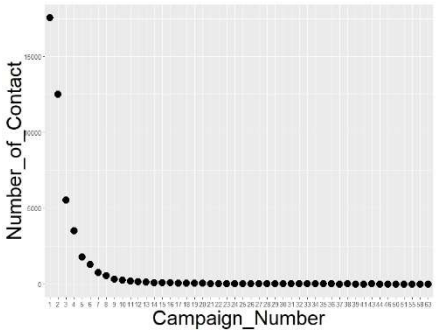

Fig. 9. Number of calls for each customer

Knowing the relations among the input data

The relationship among some input feature been illustrate for further data exploration. Duration of calls for different cclient categorizations are shown in Fig. 10 to Fig. 12. As can be seen the call duration is longer for unemployed and retired group. However, the call duration is not affected much by education and marital status

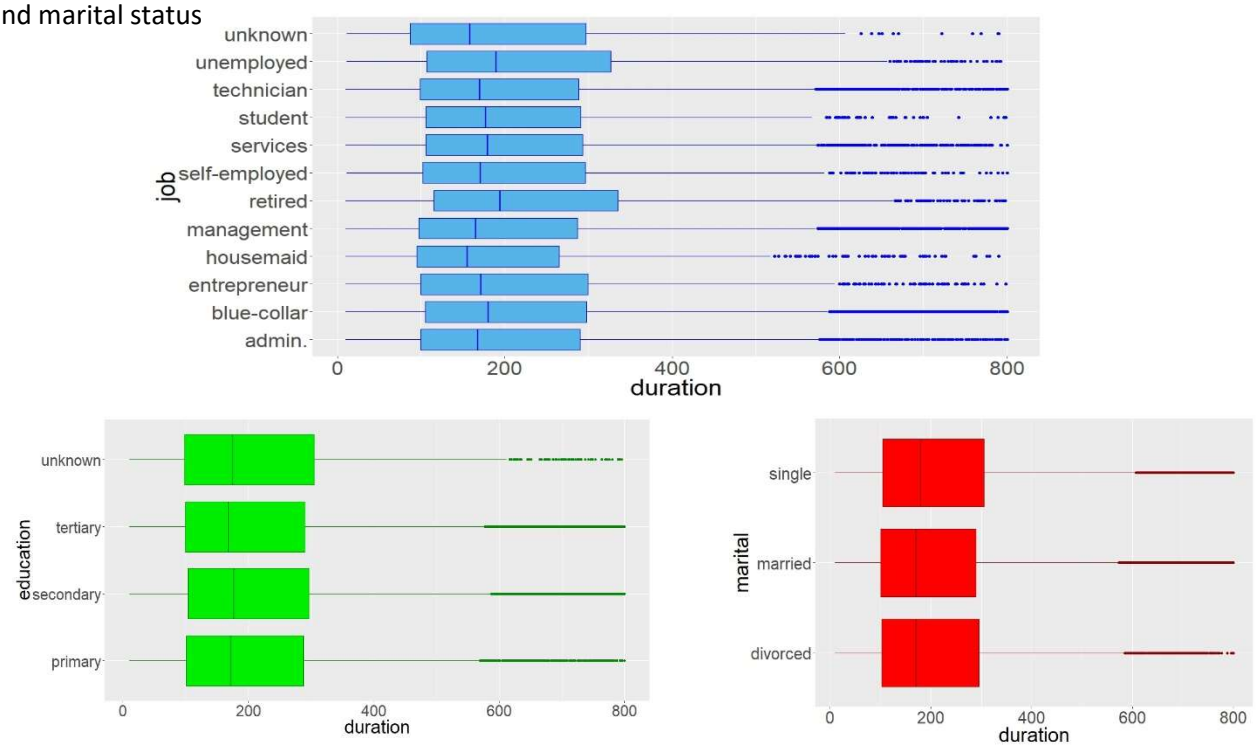

Fig. 10-12 Duration of calls in different jobs, education, and marital status, respectively

Fig. 13 and Fig. 14 illustrating the population of the education level of the customers that have loan and owning the house, respectively. Comparing these two figures shows that higher education group are more popular to owning the house than getting loan.

Fig. 13-14 population of the education level over the house and owning loan

Similarly the population of the loan and house for different jobs are shown in Fig. 15 and 16. Fig. 15 shows that entrepreneurs more tend to have loan more than the other group and 23.9% of them currently have the loan payment and "student" with 1.2% have the minimum population for the loan. Similarly as can be seen in Fig. 16, majority of clients that are working as "blue collar", "manager", or "technician" are owning the house and "blue-collar" with the population of 72.4% has the maximum probability of owning house and in "unknown" group only 9.03% are owning house.

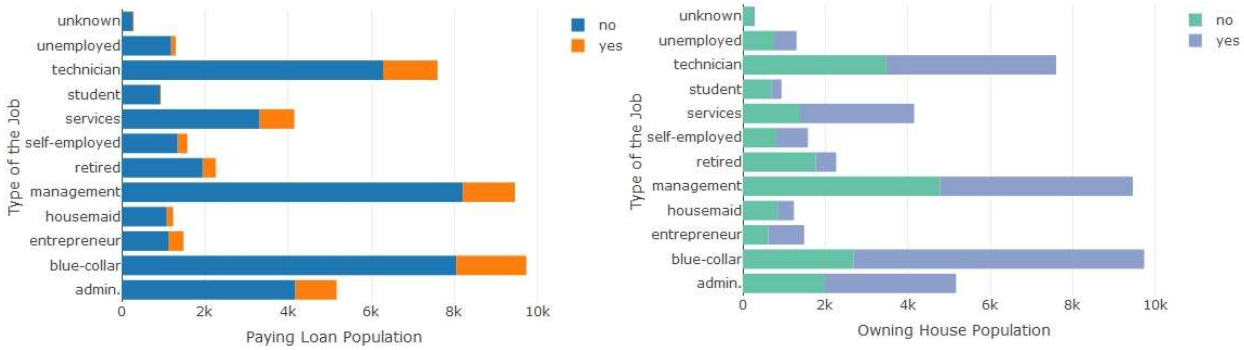

Fig. 15-16 population of the jobs for owning house

Knowing the relationship between input and output data

Output of the dataset is the subscription acceptance. As can be seen from the relationship shown in Fig. 17 and Fig. 18, there is not any clear relationship between the age and accepting the offer; however, the clients with a very low or very high balance are most likely not accepting the offer

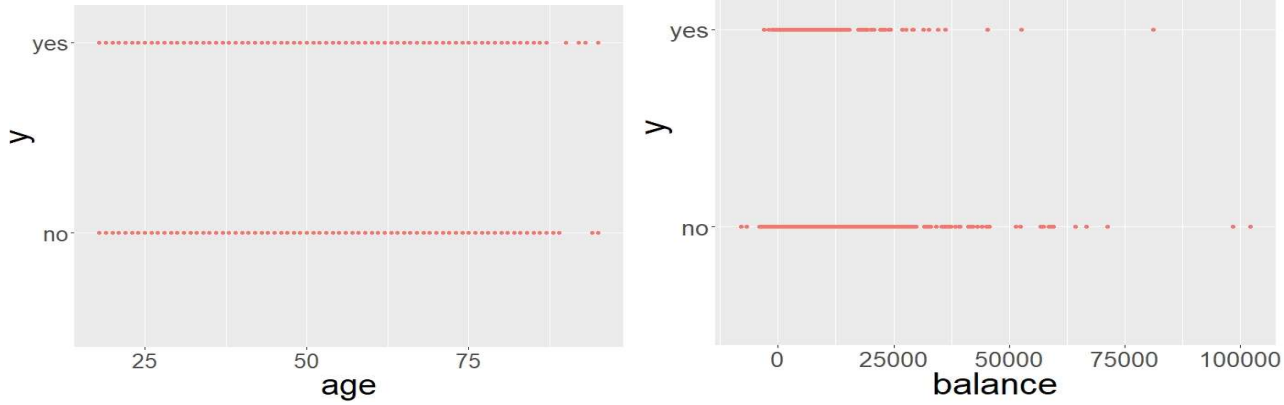

Fig. 17-18 relationship between output and age and output and balance

Acceptance population in different jobs and education are shown in Fig. 19-20. Although from Fig. 2, the bank targeted “managers” and “blue collars” to contact more frequently, “student” group have accepted the offer more than the other type of job with the population of 28.7% and “blue-collar” accepted the offer 7% and which is the minimum acceptance rate among other types of the job. The education level of clients that accepted offer shows that there is not a big difference and the clients with primary education has the minimum acceptance rate of 8.6% while the maximum acceptance rate belongs to the tertiary education with 15% acceptance rate.

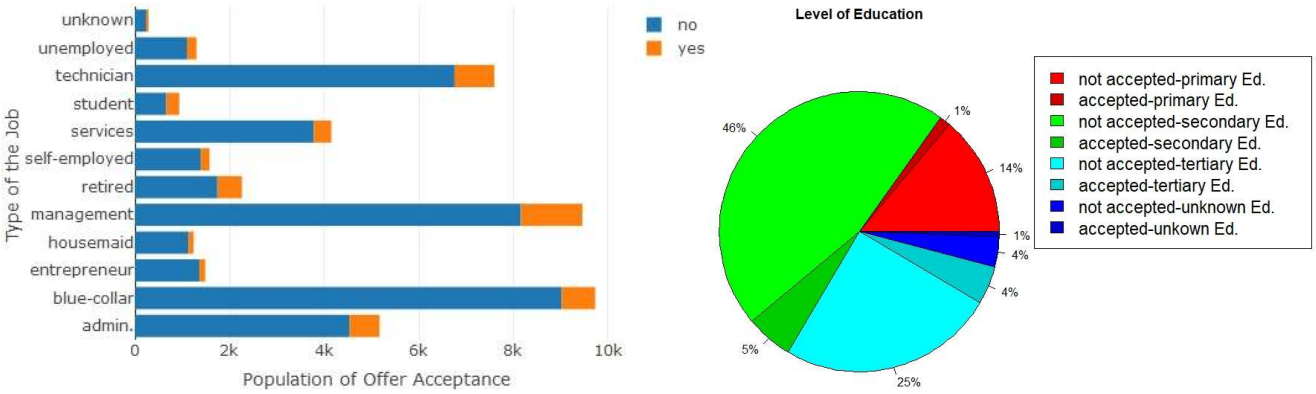

Fig. 19-20 relationship between output and job type and output and education level

The relationship between call duration and acceptance rate in Fig.21 showing that the acceptance rate for the very short call duration is very low. In Fig. 22, the density for the call duration that the client accepted the offer is shown, as can be seen the average call duration for acceptance is 537.29 second.

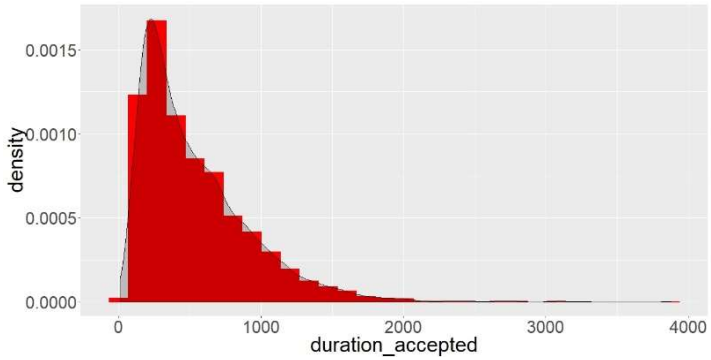

Fig. 21: Density of call duration for the clients that accepted offer

Finally, as can be seen in Fig. 16 only 12% of the offer call been accepted by clients.

Fig. 22: Rate of accepted offer

Conclusion

In order to get the better results, higher education students with the average amount of balance are a good target and more likely accept the offer with the average call duration of 537.29 second.