**Bank Marketing Classification**

**Numerical Analysis**

High Level Stats

Table

Description automatically generated

Figure

*Figure 1* displays a large standard deviation across both age and balance, suggesting our data is very spread out and not likely to be centered around the mean. This is also backed up by the range between our min and max values, thinking ahead to when we pre-process our data, we may decide using a MinMax scaler isn’t appropriate.

Exploring the Range and Distribution

|  |  |  |
| --- | --- | --- |
|  | Skewness | Kurtois |
| Age | 0.68 | 0.32 |
| Balance | 8.36 | 140.75 |

Chart, histogram

Description automatically generated

Figure

*Figure 2* shows our age variable to be slightly positively skewed by 0.68 with a kurtois of 0.32. This suggest our data isn’t perfectly normally distributed however, I do think the skewness is a huge cause for concern.

Balance on the other hand, has a skewness of 8.36 and a very high kurtois of 140.75, which further backs up our extremely spread-out data which is to be expected as we have collected data from clients with varying job roles and varying age and places in their life, we should expect everyone to varying balances in their bank account.

Chart, box and whisker chart

Description automatically generated

Figure

Understanding the correlation

Chart, scatter chart

Description automatically generated

Figure

There is very little correlation between age and balance, i.e., various ages have varying balance amounts in their accounts. You could argue slightly that ages over 60 exceed 10k far less that ages under and including 60.

You could hypothesize this and say that over 60's are less likely to subscribe as they have less money lying idle and may need easy access to their funds as opposed to committing the money to a long-term bank deposit.

A picture containing graphical user interface

Description automatically generated

Figure

We can further visualise this in *Figure 5*, with a correlation coefficient of 0.098.

**Categorical Analysis**

Target variable split

Chart, bar chart

Description automatically generated

Figure

At a high level we can see straight away that we have an unbalanced dataset, with a ratio of 1 subscriber to every 8 non-subscribers.

We can run the data through our model and see how this affects our results, but we may consider sampling. Due to not having millions of rows of data, we could consider random over-sampling however, we must be mindful that this could cause overfitting on the training data and as a result our accuracy could come up short when predicting on our test data.

Job analysis

Chart, bar chart

Description automatically generated

Figure

Chart, icon, bar chart

Description automatically generated

Figure

*Figure 7* clearly shows that our majority client base work in blue-collar, management and technician roles and at a first glance, you would assume that these job types are more likely to subscribe as they do have the largest count of subcribers compared to other roles.

However, when we look at the actual percentages, we can see that, retired clients and students have a much higher conversion rate.

This is interesting because we assumed in our numerical analysis that as clients over 60 had a balance of less than 10k, we assumed that they would be less inclined to commit their money to a long-term deposit but, this may not be the case.

This information also gives us an insight into the clients we should be targeting more, perhaps blue-collar or entrepreneur clients need further incentive to subscribe, perhaps like discount on spending or cashback everytime they save a certain amount of money.

Marital status analysis

Chart, bar chart

Description automatically generated

Figure

Chart, bar chart

Description automatically generated

Figure

Most of our client base are married which makes sense as most of our client base are in jobs rather than retired or students. The conversion rate of married clients is the lowest albeit not by a lot. This does make sense; perhaps married clients have familys and find it harder to find more disposible income to commit to a long term deposit.