CSC343 Project Discussion

Discover about Domain

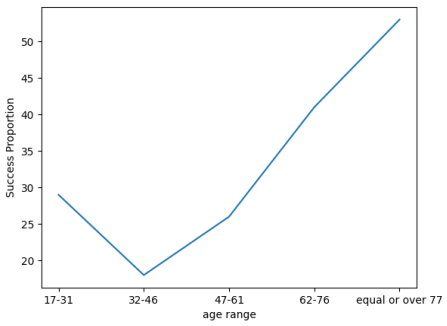
Firstly, our domain is about the direct banking marketing campaign. The dataset we chose contains information about customers of a Portuguese banking institution through one of the direct marketing campaigns, phone calls.

Basically, we are interested in exploring this bank's target customers that would have a higher chance of being promoted successfully through the marketing campaign and customers who have had credit in default. Then, we mainly focused on the outcome of the previous marketing campaign, {'poutcome': 'success', 'failure'}, the existence of credit in default, {'credit': 'yes', 'no'}, and discovered the relationship between other attributes and the proportion of poutcome being 'success' or credit being 'yes'. We will use the average poutcome in this report to represent the proportion of successful previous campaign results.

Investigative Questions' Answer

At first, for numeric variables, we used each of their values as a group to generate the corresponding average poutcome. However, since the numbers of groups for these variables are large and the number of people in each group is kind of small, then we divided each value into several groups and regenerate the result. Similarly, categorical variables would have too many categories when given, and the differences in the number of people between each category are quite large, which may lead to some bias when comparing across groups. Thus, we regrouped them into more generic groups.

Among all 13 attributes, we found out that a person's age, job, education level, the number of contacts this bank performed before current marketing campaign, CPI which means an average change in prices over time that consumers pay for a basket of goods and services(Fernando, 2021), CCI which means how optimistic consumers are regarding the bank(McWhinney, 2021), and whether this person subscribed a term deposit or not all have a relationship with the outcome of the previous marketing campaign for this person, respectively.



For the first investigative question, as the age increases(see figure at right), the average poutcome first goes down and reaches the minimum at around aged 32 to 46 with a value of 0.18, and then goes

up until the end. The average poutcome of CPI has a similar variation with age. When CCI increases, the average poutcome always increases. Also, the highest average poutcome for the groups of each individual variable respectively are the group with an age of 92 or over{average poutcome: 0.66}, do not have a job{0.40}, with high school or lower education level{0.39}, the bank performed multiple contacts before the current marketing campaign{0.47}, with CPI between 94.0 and 95.0{0.53}, with CCI between -33 to -24{0.44}, and subscribed a term deposit{0.59}.

For the second investigative question, we are interested in the type of people that would have the highest average poutcome. Based on the queries in queries_q3.sql, we found out that people with age between 32 and 46, do not have a job, have an education level higher than high school, and that were contacted by the bank multiple times before the current marketing campaign would have the highest average poutcome, which is 0.81. People that have CPI between 94.0 and 95.0, CCI between -51 to -42, and subscribed to a term deposit would have the highest average poutcome, which is 0.77.

In the third question, we are interested in investigating if marital, education and job attributes have a relation with people's default on credit which is failing to pay an expected debt (Chen, 2021).

We first investigate the proportion of existent default credit in each marital group by using Query #1 (in file queries_q3.sql). We found only 'married' people have an existent credit proportion of 0.00016 which is the highest proportion among another group of marital status. Then by Query #2 (in file queries_q3.sql), we found for education attribute, the proportion of existent default credit of the group whose education level is 'professional. course' is 0.00045 which is the highest. For the job attribute, we use Query #3 (in file queries_q3.sql) and found the proportion of existent default credit is 0.00129 for the group labelled as 'unemployed' and is the highest among other jobs.

Even though the proportion of existence of default credit does vary in different marital, education and job groups, it is not a strong evidence for us to conclude that these attributes have a relation with people's default on credit. Firstly, the proportion of existent default credit is all close to 0 and does not have a significant difference in each group. Secondly, in the result of all queries in the file queries_q3.sql, each group's number of people that have default credit varies small compare to the total number of people in this group. In fact, we found there are only 3

observations that have default credit among all 32,591 observations. Therefore, we cannot conclude people's marital, education and job attributes have a relation with their default on credit.

Conclusion and Future Recommendation

Our study mainly focuses on helping banks to campaign more efficiently by choosing the right target population and reducing the loss of customers being default on credit. After investigating our first two questions, we suggest that when CPI is between 94.0 and 95.0, the bank should pay more effort to contact frequently with people that have attributes: age between 32 and 46, no job, education level higher than high school, CCI between -51 to -42, and subscribed to a term deposit. We found this type of people has a higher proportion of accepting the bank's marketing campaign. In future studies, the result can be further explained by models such as logistic regression so that banks can use it to predict the result of the campaign. Regarding default on credit, we do not find strong evidence that the attributes we explore have any relation to it due to the limitation of data.

Reference

Chen, J. (2021, December 1). Default. Investopedia. Retrieved December 3, 2021, from https://www.investopedia.com/terms/d/default2.asp.

Fernando, J. (2021, December 2). Consumer price index (CPI). Investopedia. Retrieved December 3, 2021, from https://www.investopedia.com/terms/c/consumerpriceindex.asp#:~:text=The%20Consumer%20Price%20Index%20measures,widely%20used%20measure%20of%20inflation.

McWhinney, J. (2021, December 2). Understanding the consumer confidence index. Investopedia. Retrieved December 3, 2021, from https://www.investopedia.com/insights/understanding-consumer-confidence-index/.