The Dataset

The data used for the three following data analysis questions in available from the UCI Repository's Bank Marketing Data Set (Dua and Graff 2019). The data set is the bank-full.csv from the Data Folder provided from the URL below. Where the phone-based marketing campaign from a Portuguese banking institution produce a data set of 17 attributes that represents the banks client data. The objective of the phone marketing campaign was to see if clients would be or not subscribed to a bank term deposit.

Available at: http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Format of the Dataset

A summary of each column is represented in the table below as well as the following summary of the format.

The age column represents the age, job, marital and education columns represent the age of the client in terms of years, the type of occupation they have, their marital status and their level of education. Columns such as default, balance, housing, and loan represents if they have any credit in default, the average yearly balance in euros, whether they have the existence of a housing loan or a personal loan. The method of contact communication, the last contact day of the month in numerical value, and categorical last contact month of the year is represented by the columns contact, day and month. With the last contact duration in seconds represented as duration, the campaign column shows the number of times the client was contacted during this campaign and in general. Pdays indicates the number of days that passed by after the client was last contacted from a previous campaign where previous represents how often a client was contacted before this campaign. Poutcome is the outcome of the previous marketing campaign and y represents is a client subscribed a term deposit.

Format of the Dataset

The format of the dataset was the following:

Attribute No.	Dimension	Further Description	Data Type
1	Age	Current Age	Numeric
2	Job	Type Of Job	Categorical: “Admin.", "Unknown"," Unemployed" ,"Management", "Housemaid"," Entrepreneur", "Student", "Blue-Collar", "Self-Employed", "Retired", "Technician", "Services")
3	Marital	Marital Status	Categorical: "Married","Divorced","Single"; Note: "Divorced" Means Divorced Or Widowed)
4	Education	Level Of Education	Categorical: "Unknown","Secondary","Primary","Tertiary")
5	Default	Has Credit In Default?	Binary: "Yes","No
6	Balance	Average Yearly Balance, In Euros	Numeric
7	Housing	Has Housing Loan?	

Attribute No.	Dimension	Further Description	Data Type
9	Contact	Contact Communication Type	Categorical: "Unknown","Telephone","Cellular")
10	Day	Last Contact Day Of The Month	Numeric
11	Month	Last Contact Month Of Year	Categorical: "Jan", "Feb", "Mar", ..., "Nov", "Dec")
12	Duration	Last Contact Duration, In Seconds	Numeric

Attribute No.	Dimension	Further Description	Data Type
13	Campaign	Number Of Contacts Performed During This Campaign And For This Client	Numeric, Includes Last Contact
14	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)
15	Previous	Number Of Contacts Performed Before This Campaign And For This Client	Numeric
16	Poutcome	Outcome Of The Previous Marketing Campaign	(Categorical: "Unknown","Other","Failure","Success")

Attribute No.	Dimension	Further Description	Data Type
17	Y	Has The Client Subscribed A Term Deposit?	(Binary: "Yes","No")

Motivation in Selecting this Dataset

The reason this dataset was chosen was because it had a significant number of attributes that could provide a variety of interesting insights that could improve future marketing campaign for the bank and increase success rates and reduce their costs by allocating their resources more effectively. This provides relevant information that not only the Portuguese bank but every financial institution could take the information to improve their decision making in terms of marketing. The other reason the data set was chosen was because of my personal interest in financial services after previously working within three different types of financial institutions. Where I used to perform phone marketing campaigns for financial services. These are the main motivations in the selection of the Bank Marketing dataset.

Useful information Available in this Dataset

The dataset provides a variety of characteristics on their clients which will be explored in the analysis questions. Age, marital status, and level of education allow an insight to clients' social background which is useful in predict future clients financial prospect. By holding the knowledge of these three things the bank could make predict who is more like to default on their credit or have a personal loan when combined with columns such as housing, loan and credit. The dataset could be used to perform a classification prediction which would be very insightful and useful for a bank when preparing to perform a new marketing campaign or when deciding who to offer specific loans too new customers. Overall, the variety and vast number of attributes can be incredibly insightful for both financial or marketing strategies companies would want to implement.

Data Analysis Task 1

Data Script is found under bankQ1 and the results under bankQ1Results.

In society marketing via phone calls or cold calling have been known to be very ineffective in attracting customers. This data analysis task will aim to support the theory that phone marketing is not very effective. By calculating what percentage of calls result in a successful outcome previously and how many attempts are used on average to create a successful outcome. As the dataset does not have a customer identification number or a primary key to uniquely identify a customer, it is assumed that each instance represents the unique customer.

The analysis task aimed to measure if the phone marketing campaign had a previous outcome of success by filtering the data to retrieve only data that had a previously successful outcome. Once this was done it was ordered to produce the top 5 results of what the highest number of times the client was contacted that results in success. The top values were 11, 9, 8, 8, 8. This process was repeated to produce the top number of contact a client received that was unsuccessful were 15, 14, 13, 12, 12 to gain a full understanding the number of calls influenced the success of the outcome. The average number of successful calls was compared with successful calls having an average of 1.8 and failed calls having an average of 1.98 contacts. Finally, a percentage was taken of out of all clients that had an outcome of either success or failed a percentage stated only 23.56519 percent of calls made through the duration of the campaign was successful.

To determine how successful the phone marketing campaign was it was important to distinguish what a successful outcome means. According to the evidence within the description of the data set, the attribute outcome can appear as “success”, “failure”, “other”, or “unknown.” This proposes an initial issue with the adoption of the results as a representation of the success. As the data is not in a binary form of failed or successful calls and suggest that both failed and successful calls could be hidden or unrecognized within the data of “unknown” or “other”. This also suggest the company may not have a clear criterion of what constitutes a successful call or not. If using the assumption that a success previous outcome signifies an effective call and that a failed call signifies an ineffective call. Then when evaluating the results of the analysis it seems that only around 24 percent of calls made through the duration of the phone marketing campaign were effective. This results in over three quarters of the calls made through the marketing campaign were ineffective. This is a very high values of inefficiency and could suggest that the if the overall outcome of the campaign was more likely to be unsuccessful, then the company would be best to find a different approach to marketing. Any marketing strategy that aimed to produce a response greater than 24% could be much more effective for cost reduction and better allocation of resource. However, this percentage does not provide the full scope of the business strategy. There is no evidence from the dataset or the analysis that indicated the costs of running the marketing campaign. If the campaign was significantly cheaper to implement than other forms of marketing such as television marketing, billboard marketing or radio advents. Then perhaps a success rate of 24% is quite effective relative to the opportunity costs.

The task also analysed if there were several calls that determine whether the call is successful or not. For example, if there was a maximum number an employee would call before the call was deemed unsuccessful or if the other issues also arise within number of contacts with the customer that was the top number of contact attempts during the campaign that resulted in a success. When looking at the highest number of calls from the failed data it shows the highest number of contacts for a failed call reached 15 throughout the duration of the

campaign. This compared to the successful calls was 11. These do not provide great insight or a significant difference in a successful call to a failed one. You could argue no more that 11 calls are likely to result in an effective outcome. However, the number of calls, but the exact nature of the calls is not clear from the data attributes of the dataset. When calculating the average number of calls in a campaign that result in a failed. Both averages round to two. This could be a significant indicator that a clear outcome could be reached after two calls.

For future decision making the bank could set specific criteria that states a call has been successful or failed. The description of the data set describes the outcomes of a previous campaign therefore an outcome of a finished campaign should be known. In general, a call could fail or be deemed as unknown or unsuccessful if the client never answered or hung up. But these outcomes do not provide any indication of why that could be. Therefore, for future marketing campaigns the bank should identify common outcomes or must ensure those carrying out the calls have descriptions of the outcome. This would allow the bank to extract much more information on what makes calls successful or fail to improve future marketing campaign.

Overall, in terms of the objectives of the data analysis task, from the evidence available from the dataset phone marketing calls are ineffective at bringing a successful outcome. Although several factors and future work would need to be implemented to come to a concrete conclusion.

Data Analysis Task 2

Data Script is found under bankQ2 and the results under bankQ2Results.

Financial institutions tend to have seasonality within their uptake of financial products especially with low rates in January and higher rates in March (Investopedia 2021). The objective of this analysis is to discover if seasonality influences the success rate of the marketing campaign and identify when the most effective time of year for a marketing campaign.

To identify if seasonality does exist within the marketing campaign data, each data is split into quarters assuming January, February, and March, are Quarter 1 (Q1), April, May, June is Quarter 2 (Q2), July, August, September is Quarter 3 (Q3) and October, November, December are Quarter 4 (Q4). These quarters are measured for how many calls are successful per quarter. The results found Quarter 1 has a total of 267 successful calls, Quarter 2 has a total of 483 successful calls, Quarter 3 has a total of 438 successful calls and Quarter 4 has 323 successful calls. These outcomes can be compared to which month of the year is

the most successful. With May, August and February containing totals of 236, 207 and 143 respectively. The lowest success rates of the previous phone marketing campaign indicate that January, March, and December have the lowest values of 64, 60 and 51.

The results from the total successful calls per quarter do indicate that the quarter two and quarter three in the year represent the highest number of successful calls within the year. The first quarter of the year indicated a significantly lower number that the rest of the year. This suggests that the insight provided by Investopedia could be accurate in terms of noticeable seasonal changes effective the banking sector. With January and December having just over one hundred successful calls over the two months with over double that appearing in May. The Christmas period is around these months were traditionally known for high spending habits, with January especially implementing purse tightening behaviour. This would indicate that the marketing campaign would be effective if targeting the most common months of May and August. This would mean limited resources would be more cost effective than focusing on the most successful times of year in the six months between April and October.

Interesting March appears very low on the number of successful calls, which goes against the original analysis of Investopedia suggesting high rates beginning then. This could be the result of how many campaigns were running throughout the year. If the bank ran multiple phone marketing campaigns through the months of May, August, February, and November this could account for the higher success rate. Suggesting that seasonality is not a factor on the highest successful months. Also, in terms of employees throughout the Christmas period could have less staff working throughout the bank in a marketing campaign.

Overall, there does seem to be a significant difference in the first and final quarters of the year in terms of successful outcome. Which would suggest that the bank aim their targeting campaign throughout the second quarter of the year. The analysis still fails to provide a full picture of when is best to contact clients. The data gather was also taken years ago which if time is a significant factor, it would be beneficial to investigate how patterns of seasonality have affected over the years. Therefore, from the existing data available seasonality does seem to influence the success rate to a certain extent.

Data Analysis 3
Data Script is found under bankQ3 and the results under bankQ3Results.

Marketing strategies tend to have a target market. This data analysis aims to compare the age demographics targeted by the phone marketing campaign and identify the most popular socioeconomic factors that could be used for future marketing.

The minimum age within the dataset is 18 and the maximum age is 95. This is used to create four age categories to compare the socioeconomic characteristics of education level, existence of credit default, the appearance of a housing loan and the holding of a personal loan. Where default, housing and loan are binary, and education is categorical as unknown, secondary, primary, and tertiary. The age categories are between 18 to 34, 35 to 54, 55 to 74, and 75 to 95. Each age category is ordered to include education, default, housing, and loan data. This is then grouped to count each combination of socioeconomic characteristics found for each age group.

The most common socioeconomic characteristics in successful outcomes for the age range of 18- to 34-year-old with 174 counts of those with tertiary education, with no credit defaults, no housing loan, and no personal loans. This is followed by 129 counts of similar credit, housing, and personal status. Theses counts together create most data with every over 167 success counts in the rest of the characteristics. This suggests there is a correlation with the success of calls with those who have no indicators of financial struggle. If it is assumed that you have no credit defaults, no housing loans and no personal loan represent a higher credit score or higher financial stability. The bank could interpret the results as within the age range of 18 to 34 to focus on clients with these characteristics to achieve higher rates of successful calls. They could also adopt these findings into other forms of marketing such as tv or social media, targeting a demographic without existing loans in the future.

For the age range of 35 to 54 there is a higher count of success within financial stability of no credit defaults, no housing loans, and no personal loan. This is very similar numbers in terms of the top two values within a few counts difference in comparison to the 18to 34 demographic. However, there is a substantial difference in the counts of those with equal status in education, credit default and personal loans however the counts of 130 and 77 for successful calls within the clients do have substantially more successful calls with housing loans that 18- to 34-year-olds. This would make sense in terms of a general view of age and banking. It is common for increased financial stability and accumulation of assets as individuals age. This could reflect the common adult aged 35-54 having a home with ta housing loan of some form so when the phone marketing speaks to them most adults have existing housing loans in this age range. This assumption would account for the different in the two age ranges.

In comparison to both the age ranges between 55 to 74 and 75 to 95 all age ranges have the same pattern of higher success outcomes when the client has no default credit and not personal loan. However, education becomes a more important indicator of how successful the outcomes would be compared to the other factors. The top four outcomes of the successful count all have no housing loan compared to all the two younger demographics. This could be the result of after a significant period most individuals contacted had paid off any housing loan or never needed one to begin with.

In general, these finding suggest that the company should aim to target their marketing to users who show evidence of financial stability as that appears from the data collected to be the highest demographic for success. In terms of age category by focusing on those between the ages of 35 to 54. There seems to be a highly likely hood of successes. However overall, in terms of education, default credit status and no personal loan, these seem to be the target audience that would respond the best to future marketing campaigns.