Slideshow

*Slides 1 + 2*: Introduce the group

*Slide 3*: Introduction to what data we are looking for

*Slide 4*: During the presentation we will be looking at the following: the questions we were asking of the dataset and why? Where we found the dataset and how we used it to answer our questions. The clean up and data analysis process and finally the conclusions that we’ve drawn from our data and the implications they may have.

*Slide 5:*

*Slide 6*: Firstly, we wanted to look at if we can compare age and gender to the attrition status of customers, to identify any possible trends. To do this we first selected which columns we would need from the data set and created a separate data frame to allow for easier analysis.

Initially we made pie charts to allow for easier visualization of the percentage of genders that make up the dataset, and of the percentages of attrition flag, which shows us whether a customer is still with the bank or has left. [Show pie charts]

Following this, the .loc method was used to create two data frames that are locked onto Female gender and Male gender, which were then used to get value counts of the attrition flag. This allowed us to create a stacked column chart, showing the numbers of male and female customers that are still with and have left the bank. [Show “customer status by gender” plot]. This allows for easier comparison between the genders.

To look at the customer ages, we first had to create bins to put them in. This allowed for grouping of the ages into a new column, which made the analysis easier. A pie chart was then created to visualize how the age groups make up the data. [Show pie chart]

Using the new data frame with the age group column, the .loc method was used again, this time to create two data frames that are locked onto “Attrited Customer” and “Existing Customer”. These are then used to get the value counts of the age groups, which then had to be sorted using sort\_index, as value counts automatically puts it into descending order.

This was then used to create a stacked column chart showing the number of customers within each age group that are existing customers and have left the bank. [Show “Customer status by Age” plot]

We then created a data frame just to show the mode, mean, median, standard deviation and variance of the age to compare.

*[Should we go into findings on question 1 here or should we do that all at the end?]*