Report: Data Storytelling on Bank Customer Churn dataset

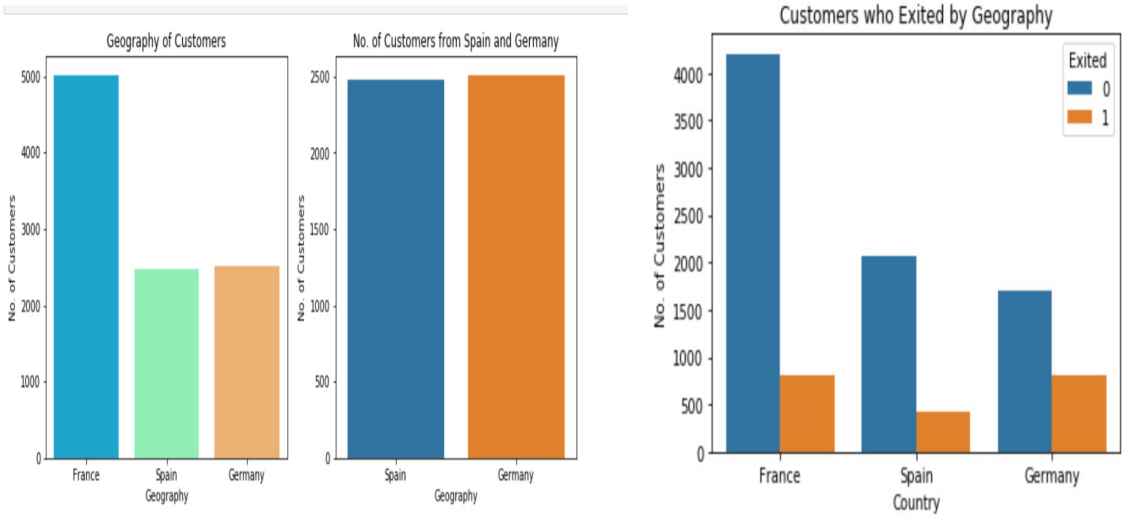
The Dataset is about bank customers churning and can be found on Kaggle:

https://www.kaggle.com/barelydedicated/bank-customer-churn-modeling

Disclaimer: The dataset above is simulated.

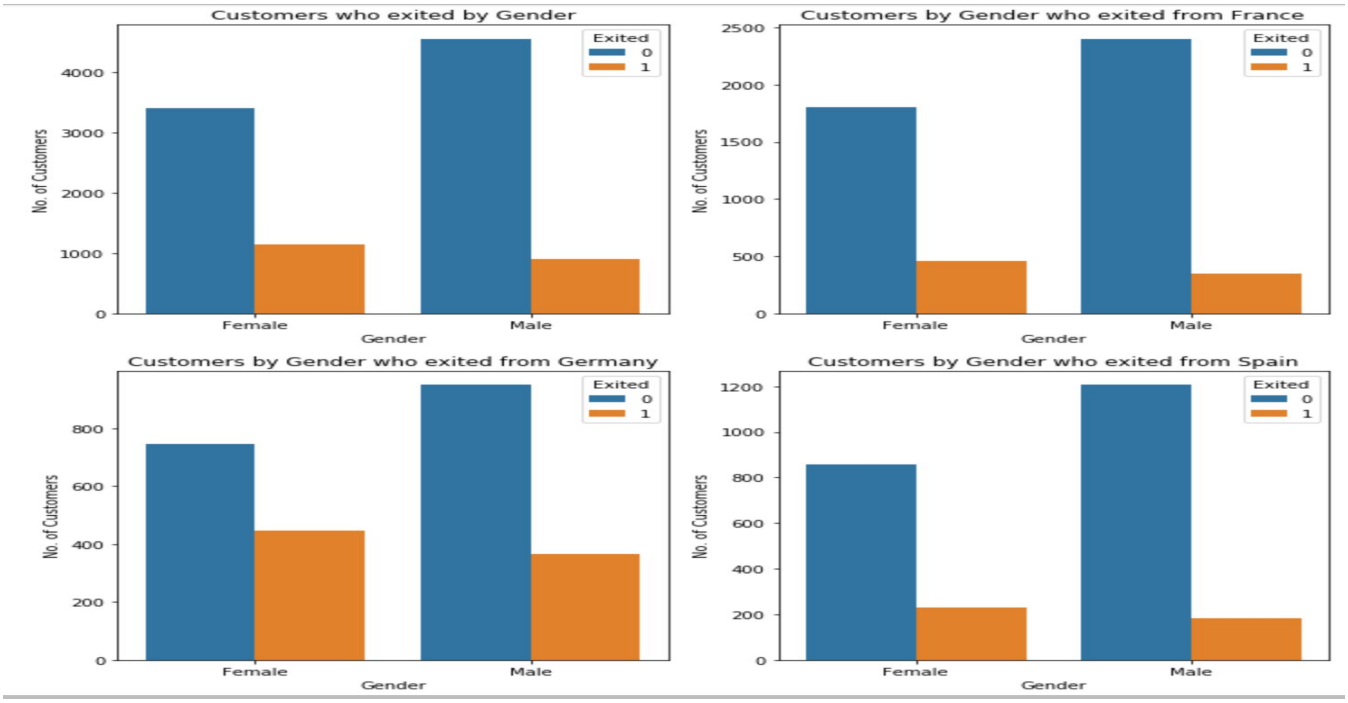
The aim of this exercise was to investigate, by visualizing the data, to uncover interesting information about the data and learn more about the customers in the bank who are exiting or not.

I wanted to learn more about the bank customers and the dataset so I began with visualizing where the customers were from and found that there were twice as many French customers than from Germany and Spain. I also found Germany had the most customers who exited followed by France and Spain. This can be seen below in the 3 plots.

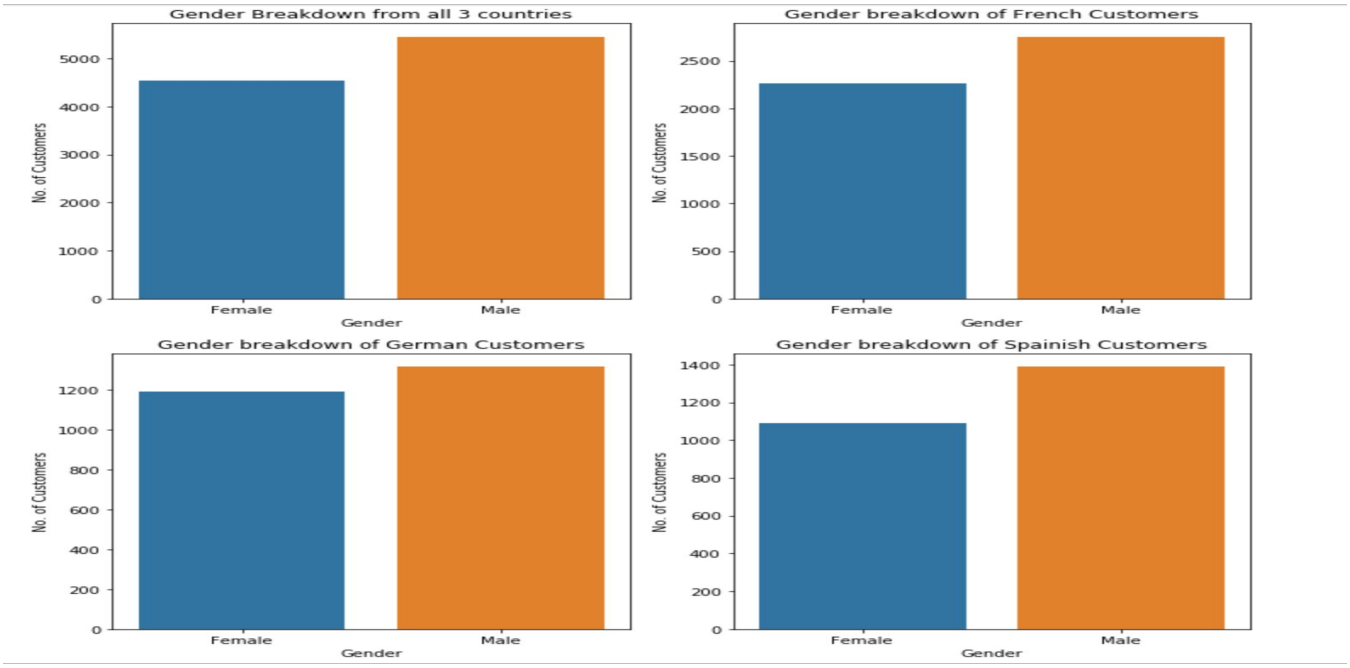


I wanted to find out the gender breakdown of customers who exited in the whole dataset as well as the gender breakdown of customers who exited from each country. Overall, females customers left the bank more than male customers and Germany had

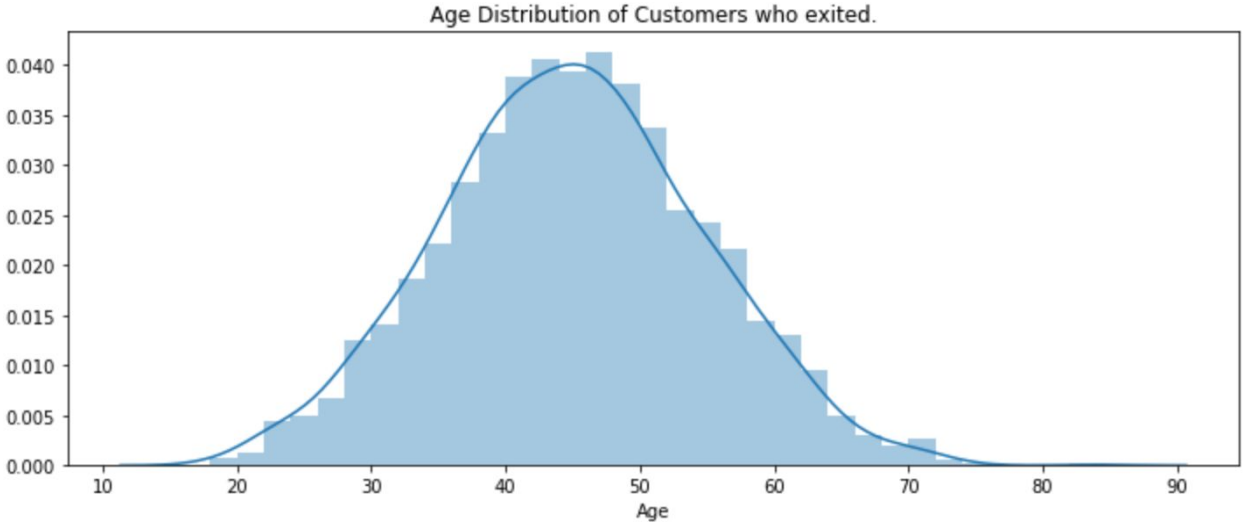
the highest number of customers who left the bank. France had the highest number of female customers who exited the bank and Germany had the highest number of males who exited the bank.

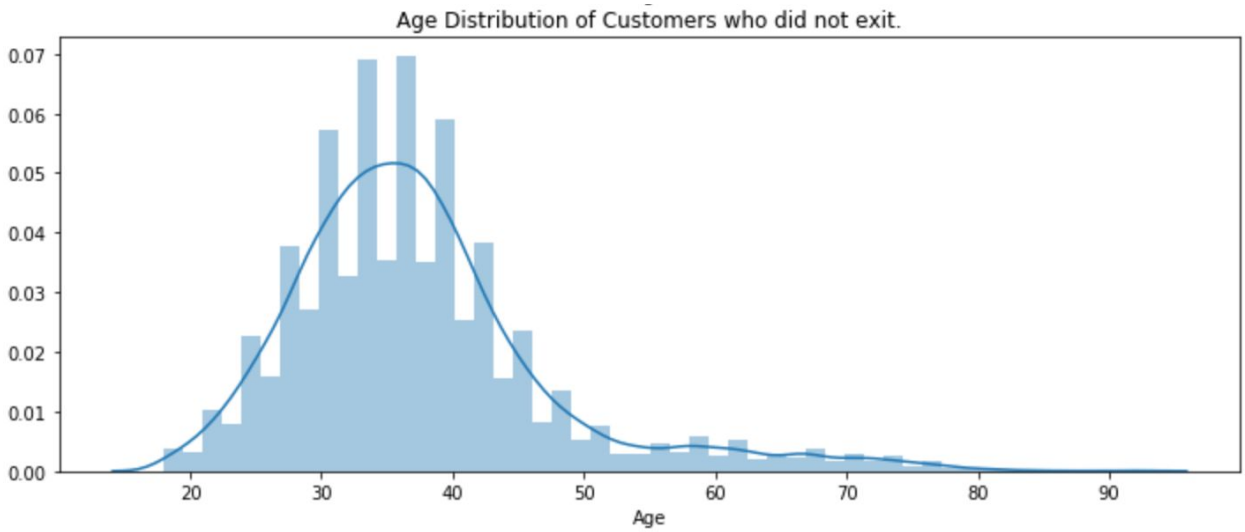


Looking at the gender breakdown of the bank customers, I found the highest number of female and male customers to be from France as seen below.

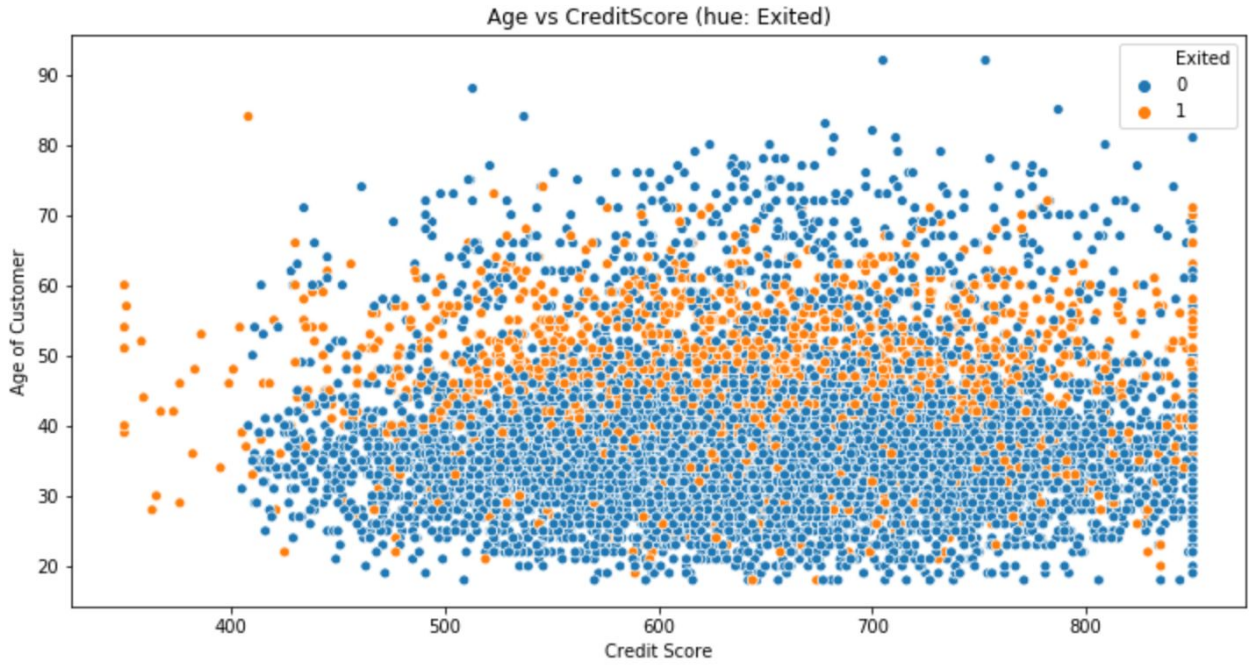


I looked at the age distribution of the customers who exited the bank and it was close to looking like it is normally distributed.

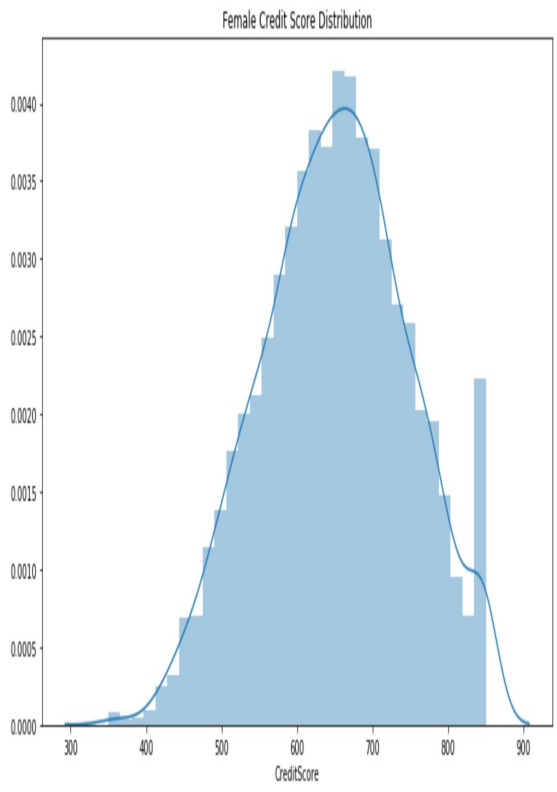


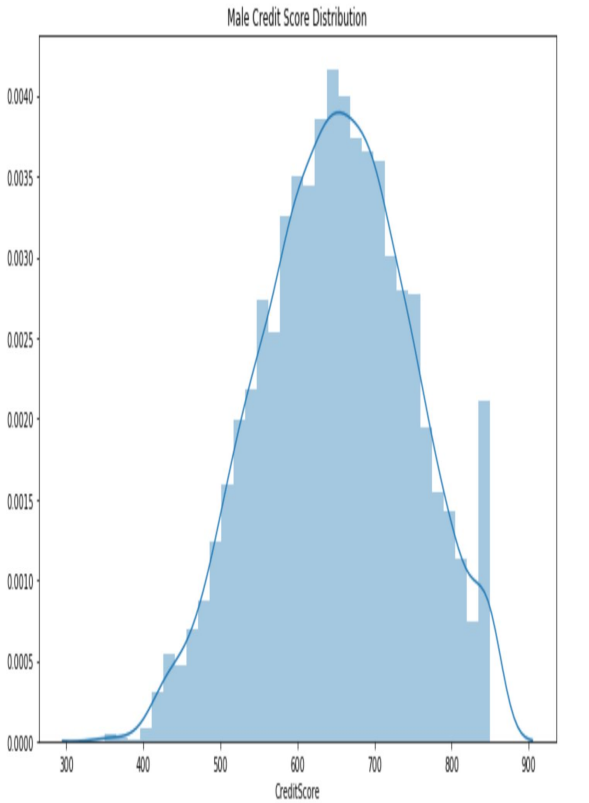


A scatterplot between credit score and age revealed that the majority of the customers who exited were between 40 years and 60 years of age. It would be interesting to see the reason why this particular group of people aged between 40 and 60 years old are exiting the bank? Are they getting better offers from other banks?

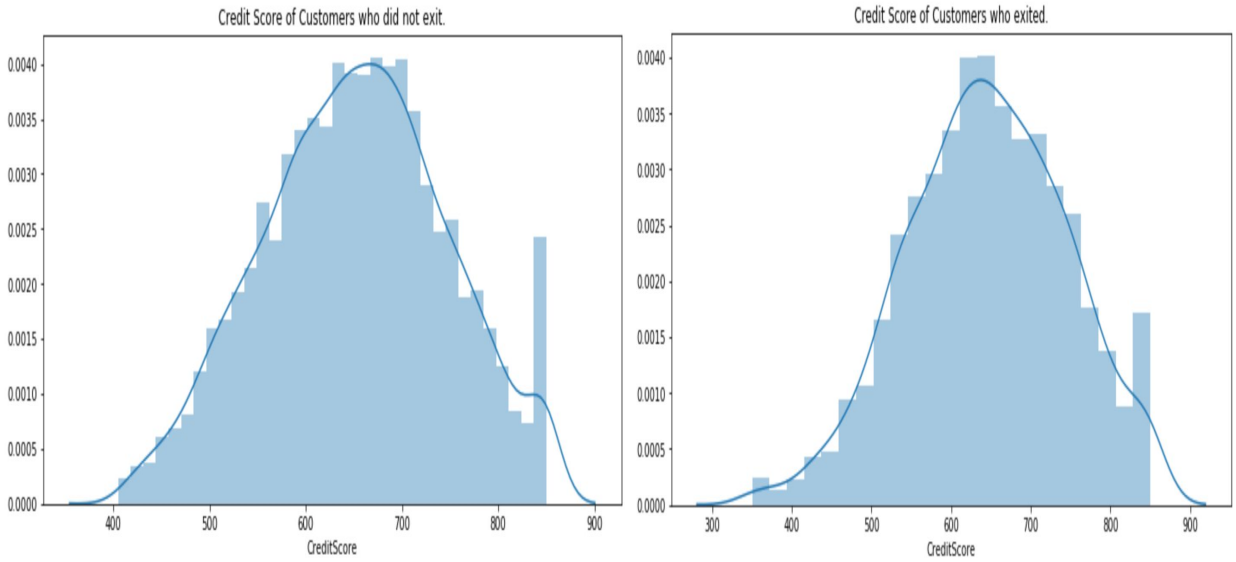


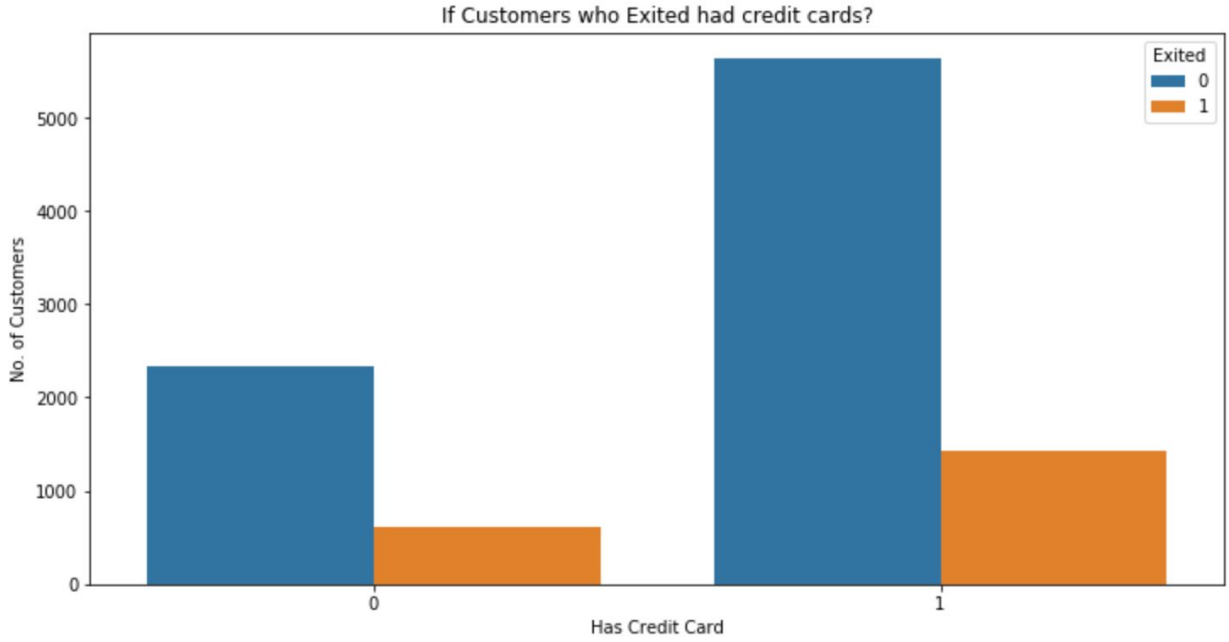
I wanted to look at male and female customers credit score distributions and found no significant difference in the distribution of both distributions.





It is interesting to find that customers who exited had a roughly lower credit score and customers who had a credit card exited more than customers who didn't.





This leads me to ask myself if the customers could get better offers based on their low credit score from other banks that is causing them to exit? Are other banks offering lower APR on their credit cards to lure customers?

After exploring the data visually and extracting information, we are now left with more questions that we can explore further based on the data.