Luna McBride

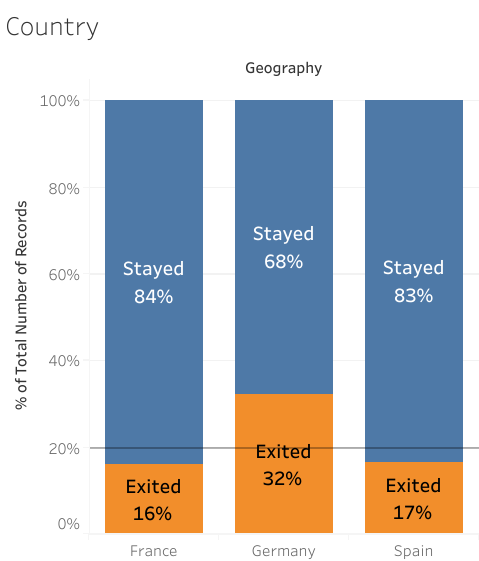
Udemy: Data Science A-Z

**Homework: Tableau Chi-Squared**

**Opening**

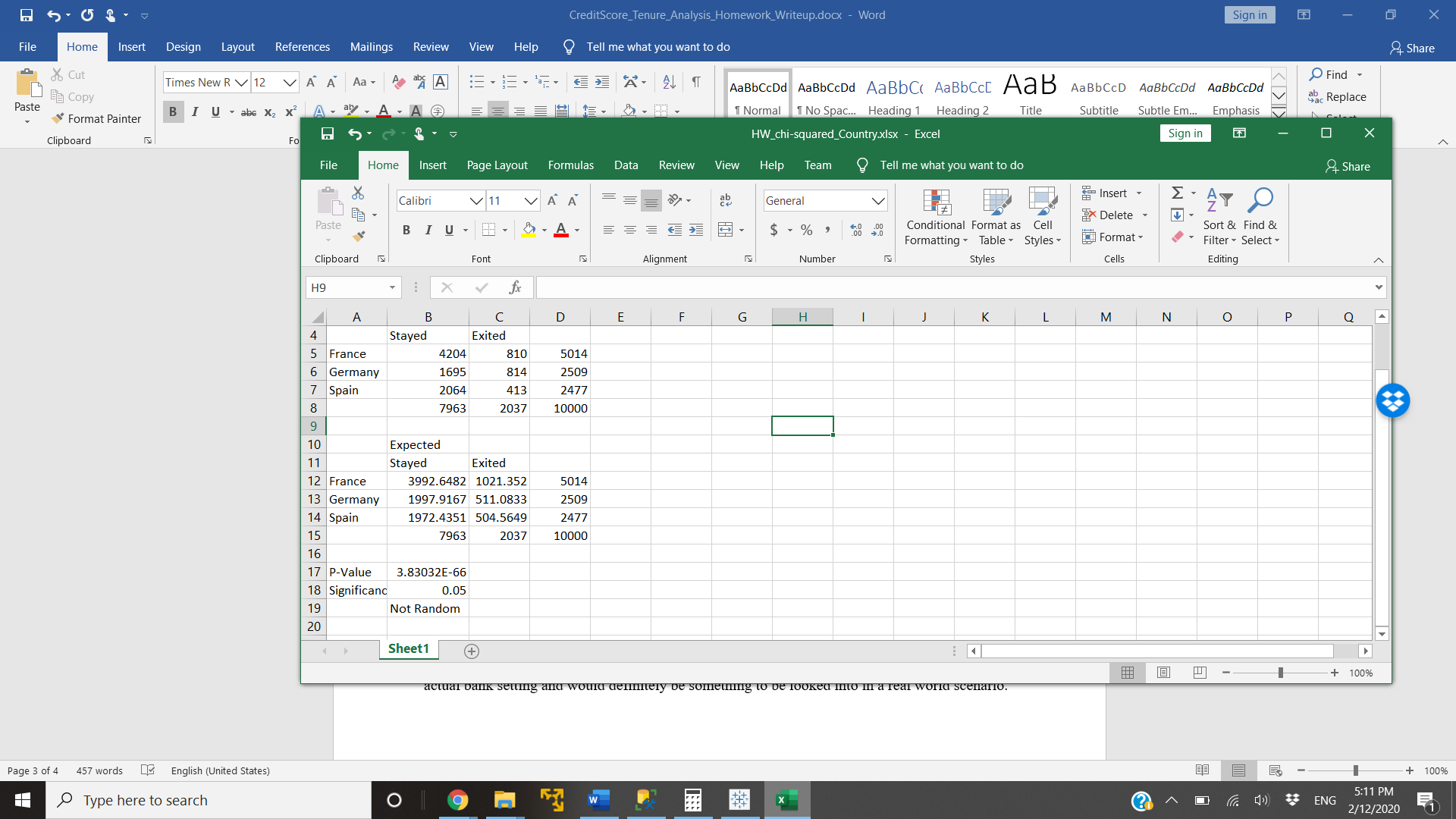
This is a write-up for the second Tableau of the Data Science A-Z class found on Udemy. This one covers a data set used primarily through this section, P12-Churn-Modelling.xlsx. This homework was a quick way to put some statistics behind our data, using Tableau to approximate the end result then using chi-squared testing in Excel to add more credence to our guesses.

**Country**

****

This was the graph used to make assumptions about the significance of the country. Though France and Spain are around average, Germany is quite high in the amount of people leaving (exemplified by the orange bar). This means that the country likely matters in the amount of people exiting, but we have to be a bit more confident, hence the Chi-Squared test.

**Chi-Squared Test in Excel**



The Excel sheet has three main parts for creating our result. The first is in lines 4-9, where we have a table of people who stayed and exited from each country, which were pulled from modifications to the country table in Tableau (switch from percent to counts). They are surrounded by total numbers for each row or column and a total count in the bottom right corner, mostly for calculations and sanity checks. The second is lines 11-15, which are calculations of what should happen at the baseline exit rate of 20%. These were calculated using the previous table’s overall numbers to get what it should be, not rounded to the nearest human. The count for each row or column is also still shown around them as a sanity check, as they should be the same as the original table. They are, so there is no issue.

The final sections are lines 17-19. Line 17 gets a P-Value using the CHITEST function in Excel. It is passed in the important values, which are those underneath the stayed and exited columns for each country on each graph, and gives a value to test to the significance level I have decided on, being 5% (line 18). The final line is an if statement checking if the p is less than our significance. It is, so these exit figures are statistically significant, thus returning “Not Random”.

**Conclusion**

There is more to seeing data than how it looks. Though intuition said that being in Germany was a likely factor to leaving after seeing the data, using statistics to back our intuition is also quite important. This case came to the same conclusion as the initial thoughts, but this might not always be the case. Doing the work to strengthen the claim is just as important as seeing the value, so keeping Chi-Squared and similar tests in your arsenal is very important to guarantee your forethoughts are correct.