Summary

My project was based on the Paycheck Protection Program conducted by the United States government in the year 2020. The main purpose of this program was to support small businesses and companies amidst the Covid 19 situation and prevent unemployment from rising. The government had put in $800billion into this program. I was to analyze two data sets, one being the existing loans given to businesses and the other are the removed loans from the program. Some of the reasons for these removed data was either the applicant withdrew it, Requiring more information from the lender or Fraud.

I began my research by going through the data sets I was given as well as going through the internet in search of any data that would be useful for me. I mainly used to excel to support my predictions as I was not well versed in other data analysis software's. I started by merging some of the columns which I thought showed any sort of relationship. By comparing columns and plotting graphs I could get a basic idea on how to answer the questions at hand. As I progressed through the project gathering more data, I could get an idea on the bigger picture.

Overall, the whole project was a bit challenging and with my low experience in data science, I found myself in many tough situations. But at the end of the day this helped me understand how the PPP loans worked and how we can identify any loans which seems fraudulent. In conclusion it was ta great experience for me and I loved it.