The project proposal that I will be presenting is to construct a model that can help predict credit card fraud. The job that I will be starting after graduation is a consultant on a Data Analytics team, working with digital forensics and litigation cases. A lot of the projects I will be working on involve cybersecurity and protecting the integrity of our clients, which sometimes involves detecting fraud.

The dataset that I have found for this project has 23 columns and 1,296,675 rows of data. The data has already been cleaned and will only require some minor adjustments to be made to the format of the column names after exporting it into Python. The data also includes timestamps, which have already been cleaned, but will need a check over to ensure that everything works properly when analyzing. The data is very easily imported from Kaggle. Being that it is a larger dataset, there are many different avenues on which this project can be approached. The data is also very easily transported into Jupyter notebook, to be analyzed and visualized in Python.

Regarding the exploratory analysis, most of my initial research would be centered around the binary variable detecting if there is fraud or not. The column states 1 if that transaction was fraudulent, and 0 if not. Many of the other variables include things such as credit card number, transaction number, name of merchant/customer, timestamps, the category of purchase, location of purchase, and city. I would look to see if I could create a model around the dependent variables I listed above that would be able to accurately predict and detect credit card fraud. There are lots of implications for the stakeholders of this project. First off, consumers would be able to have clarity on when their credit cards get hacked. Businesses and producers would be able to protect themselves better against theft and fraud. Finally, police and law enforcement will have a more efficient way to find criminals. Some of the ethical concerns involve infringing on

the privacy of civilians in way of monitoring their credit card purchases. Additionally, there are many legal implications of this model and it would have to be deemed legal to be able to prosecute criminals.

Other Ideas:

- Will use observational data
- Potential ideas:
- Housing affordability
    - Lots that we can do with housing
    - Can do predictive modeling
    - Can also do cost of living in certain areas
- How are local economies affected by large universities.
- Most pandemic resistant industries
    - Jobs that are recession resistant
    - Industries that did not waver during covid
- Finding new ways to deter credit card fraud
    - Really interesting because of the job I'm going to do
    - Building models to detect CC fraud
    - Not sure how easy that would be to implement
    - Two datasets, both from Kaggle.
- Consumer spending during economic downturns
    - Shift from necessity items to luxury items
    - Do companies shift their producing priorities
    - Which companies succeed and which don't