* Hello everyone, my name is George Athitakis and for the next 5 minutes, I’ll be providing some insight on credit card fraud and how we go about detecting it.

SLIDE

* So, why is this an issue? Somewhat of a rhetorical question but let’s put it in to perspective.
* In just the United States, there are about 108 million credit card transactions every day. That comes to roughly 40 billion every year.
* In 2021 alone, 5.8 billion dollar’s worth of fraudulent transactions were reported. That’s up 70% from the 3.4 billion dollars reported in 2020 and it does not look to be slowing down.
* Now, the two main stakeholders affected by this are the Service providers, TD, VISA and so on, as well as the consumers. That being you and I, those who use our credit cards every day to make purchases.

SLIDE

* How did I go about handling this task? What was the plan?
* I created a model to detect fraudulent credit card transactions using a classical machine learning approach, utilizing statistical algorithms to get my results. I went down this route instead of implementing neural networks primarily to limit the computational stress on my machine as I was working with quite a large amount of data.
* I was able to obtain a dataset compiled by IBM that included just over 24 million transactions as well as the account information and the card holder information.
* To supplement the transaction data from IBM, I sourced unemployment records from 1991 to 2020 for all 50 states in the US and averages for all 6 continents, excluding the United States from the North American average.

SLIDE

* After creating my model, how did I know it was working? I looked primarily at three related metrics.
* The first being overall accuracy. Now on the surface, 90% seems promising but when you look further into the data and see that out of the 24 million transactions only 30,000 were fraudulent, you realize that it doesn’t tell the full story as if the model guessed non fraudulent every time, it would have an accuracy of 99.99%.
* So, I looked further. Next, was the metric Recall. This shows that out of all the fraudulent transactions, my model was able to detect 85% of them. Now this is encouraging as we want to catch as many fraudulent charges as possible. Being able to decern 25,500 out of the 30,000 total instances of fraud, though not perfect is a great starting point to work from.
* And lastly, I looked at the precision of my model. In addition to being able to recognize fraudulent transactions, I wanted to know how many times was my model was mis-classifying either fraudulent charges as not fraudulent or vice versa. I found my model to have a level of precision of 74%.

SLIDE

* Next, I wanted to see what features were impacting my model and rank their importance based on if they dictate whether a transaction is fraudulent or not.
* The chart here shows the top 10 features from my model ranked by, you guessed it, their importance to the classification.
* The feature with the most importance is the unemployment rate of the merchant location. This feature is specific to not only the location of the transaction but to the year it was processed as well to give an accurate reading of the unemployment rate at the time of the transaction.
* The next feature on the list is the transaction type. This lets us know whether the transaction was processed online, if it was a swipe transaction or if it was a chip insertion transaction.

SLIDE

* As I dove in to the two aforementioned features, I was able to find some interesting bits of information that help make sense of the rise in credit card fraud.
* We can see here in the following chart that 62% of all fraudulent transactions occurred in areas where unemployment was at 10% or above. From this we can infer that the more people that are out of a job, the likelier they are to resort to fraud in an attempt to cover expenses.
* In addition, 62% of all fraudulent transactions were committed online. This comes as no surprise as majority of our purchases have moved online as well, and it only makes sense that scammers are able to utilize this to their advantage in both gathering sensitive information and processing their scams.

SLIDE

* Lastly, as me and my classmates know oh so well, there is no such thing as perfection in the world of data science as the information is always changing. So how can I go about making improvements?
* I would like to obtain and implement information regarding data breaches to verify whether the card used for said transaction has been compromised. Data breaches are becoming an increasing issue as seen in the chart below.
* Shown here are 5 of the notable data breaches that occurred in the past 10 years amassing an amount of 509 million affected accounts.
* Now after everything I’ve shown, I hope I haven’t scared you from using your credit card ever again but instead reassured you that we data scientists are working diligently to help keep us and our funds protected. Thank you for your time, and once again, my name is George Athitakis.