**AI and Machine Learning**

**Project Overview**: In this project we were presented with a database of a manufacturing company and were asked to use machine learning and learn how to use the database.

**Problem Statement**: The problem that we were asked by the company was to predict when a package would be received by a customer.

**Data**: The database was provided by KOCH and accessed through the Snowflake program. In the database there are 1.1 million rows of data that include different vendors, dates, and product descriptions.

First, we noticed that the data needed to be cleaned. Many of the rows contained NULL values and had to be removed. [use the NULL.txt]

To further clean the data, we noticed that all of the date columns had dates stored as integers using the format YYYYMMDD (I.e. 20210807), which can't be used to make calculations. We changed the dates by converting them into CHAR and then used the TO\_DATE function to convert it into a date datatype. [use the DateConvert.txt]

Once we did this, we were able to start using the data in calculations. [use th e Aggregate.txt]

You can see in the Aggregate.txt file that the secret to our prediction calculations was based on subtracting the FIRST\_GR\_POSTING\_DATE from the DELIVERY\_DATE. This gets us the difference between the expected delivery date and the actual delivery date.

**Methods**: Using the data that we found we created a pop window within python that the company can use to enter a vendor ID and the expected date of delivery given by the vendor and compare that to the Median of delivery dates that the vendor has had in the past and output an expected delivery date based on that information.

**Results**: Using the median the company can see all vendors and their delivery dates and using that can makes decisions if they want to keep using that vendor based on time that they normally deliver their products. If a vendor has a long delivery time then the company can make a decision whether or not they want to continue to buy from that vendor.

**Visualizations**: Using Python we then created a graph to show the medians for each vendor. Each number on the graph is the median number of days that each vendor took to deliver the product. If the number is negative that means the vendor is usually early in comparison to its predicted delivery date and if the number is positive then the vendor had a tendency to deliver its products after its expected date.

**Code**: We used the code in snowflakeCode.py to generate the graph for visualization along with creating a UI that allows the user to input in the vender ID and the date that the vender estimated the package would arrive and then give a more accurate estimate based on the data provided. The connection.json file allows for access into the Snowflake database. https://github.com/Tuck-Danielle/Querious\_George/tree/main/SnowStuff

Challenges: One of the challenges that we faced was using are time wisely. While working on the QuickStart we took too much time trying to get everyone to do the QuickStart rather then one person doing on one computer and we all learn togeher. Another challenge that we had was that we did not use one anothers strength in the beginning. Had we used each other strenghts sooner we could have completed mutiple parts of the project more efficiantly. The final problem that we ran into was recording information. We were not very good at writing down our steps in the early parts of our process so we had to do it later which makes it harder.

**Future Work**: One thing that would enhance the project would be a clairifcation in the naming of the columns. There was a lot of confusion as to which columns meant what and took some digging to find which ones we needed to use in order predict the date of delivery. We would recomned that the company finds a more standardized method of collecting data from the vendors. Within the data there are a lot of null values and a lot of formating issues. With theese new corrections the company could by more easliy evalueate the data

**Conclusion**: We found that between all of the vendors that the company purchaes from there are more vendors that have a tendency to deliver early however of the vendors that deliver late they tend to deliver well past their expected deliviery date. The company would need to decide if they want to continue working with those vendors or have a converstaion with those vendors and maybe find the reason that those vendors tend to deilver late. We found that if you could