**Capstone Project – Supply Chain Analytics**

**Data set**

* Cycle data.
* Location data.
* Delay data.

**Tools Used**

* Jupyter Notebook
* MY SQL workbench 8.0
* Power BI.

**Data Cleaning and EDA**

* The Given data set is loaded into Python notebook .
* Basic checks like Shape ,size ,Data types is checked .
* After checking doing the basic checks analyzed data set for missing value and found out there were many columns are with more than 50% of missing value ,hence checked the reason for it to find is it a really a missing info or the info is genuinely missed and found out that most of the missing information is genuine .So ,didn’t dropped any columns even though the missing percentages are higher .Instead of Dropping the columns replaced the missing value with some of the appropriate value so that the data set wont get affected for further analysis.
* Data types are checked and changed it to the Appropriate data types.
* Soft capping of Outlier Treatment is Done for the columns which is used in the further analysis .
* After the outlier treatment loaded the 3 data sets into MYSQL by establishing connection between python and MYSQL.
* After loading the data into MYSQL created the master tables as required.
* Stored Procedure is created for all the key metrics and derived some new tables and columns like running time ,quality ,good count which is needed for the visualization.
* After analyzing the data using stored procedures ,connected the MYSQL work bench with Power BI to Analyze the data using the help of visualization to derive more insights.