**Summary of CARS DATA ANALYSIS project**

My project was about data analysis on Cars dataset. The goal of the project was to combine cars data from 3 different sources, combine them using common key and perform analysis on that.

Source 1: Flat file(csv) – Kaggle.com

Source 2: Website data –

https://en.wikipedia.org/wiki/List\_of\_best-selling\_automobiles

Source 3: API data -https://vpic.nhtsa.dot.gov/api/vehicles/GetMakesForVehicleType/car?format=json

Step 1 was to fetch, clean and format the data in the csv file. Out the three data sources, the csv file dataset took multiple steps to clean as it had many inconsistent values. While performing data cleaning, learned about fuzzy logic which was very useful to bring inconsistent values in ‘Make’ column. Car ‘Make’ or ‘Brand’ along with model was the common key. Hence it was important to remove inconsistent values as much as possible before merging.

Step 2 was to get and clean the website data. During this process, learned about reading data from html using pandas. There was an easy method which I came to know during research to read the website data. Since the data is loaded form Wikipedia, there were some hyperlinks on the data that had to be cleaned. Performed few string clean techniques to format the data.

Step 3 was to get and clean the API data. This process was bit challenging and interesting for me as I had to call the API source twice to get the required data. The first API call was to get all the ‘Make’ and the next API call was to pass the fetched ‘Make’ values and then get all corresponding ‘Models’.

Step 4 was to combine the fetched data. I stored the data from steps 1 to 3 in csv file separately. They were combined and loaded to database. Initially the plan was to just combine the data based on just ‘Make’ as key but it did not give me expected results. Hence used ‘Make’ and ‘Model’ as key to merge the data.

Overall, I learned about various data cleaning methods, data visualization techniques and SQLlite database to complete the project.