Progress Report 2

Team PPUC

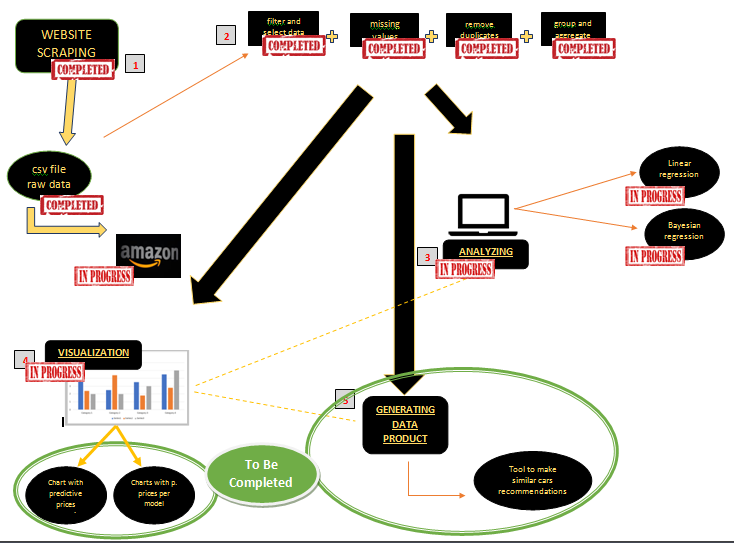
**Team members:**

Brian Colley - Coordinator Sandra Tinta Emmanuel Atanganas David Rodriguez

[bwcolley@icloud.com](mailto:bwcolley@icloud.com) [stinta@gmail.com](mailto:stinta@gmail.com) [emmanuelatanganas@gmail.com](mailto:emmanuelatanganas@gmail.com) [darogi75@gmail.com](mailto:darogi75@gmail.com)

**Overview**: Our team’s project involves a compiled data set of used car advertised prices along with certain variables describing each car. Our hypothesis is that we should be able to make predictions on the asking price of a car, based on its attributes, as well as identify anomalies to show a good buying or selling opportunity.

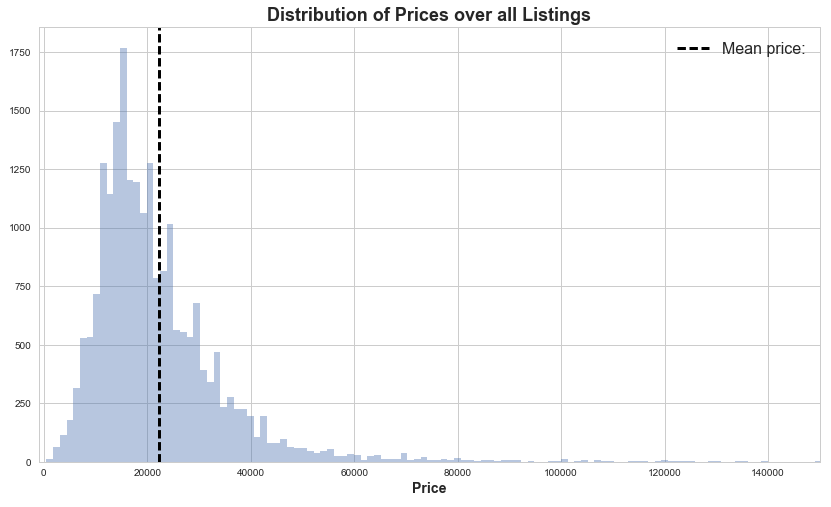
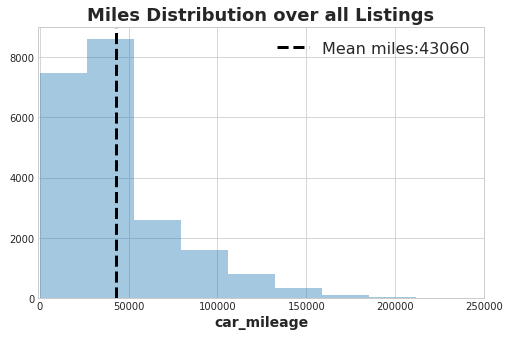
Below is the original architecture diagram as a visual reference to the team’s progress. As is seen, our data has been ingested and cleaned, and is in the process of statistical analysis and visualization.

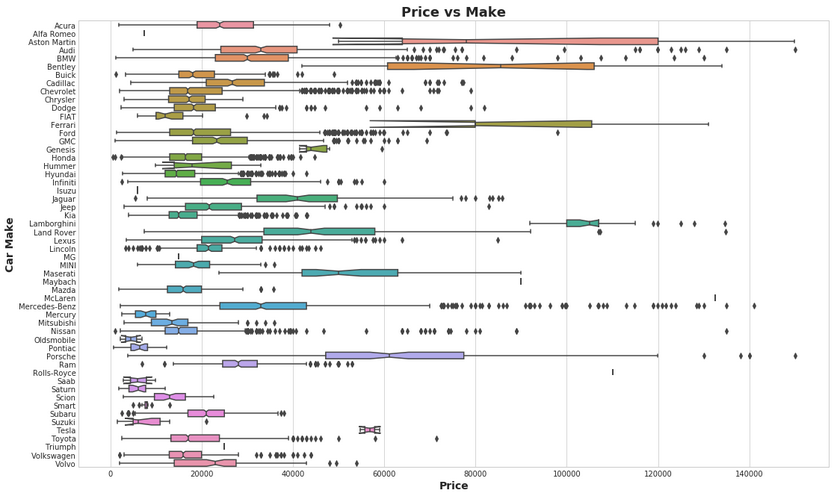


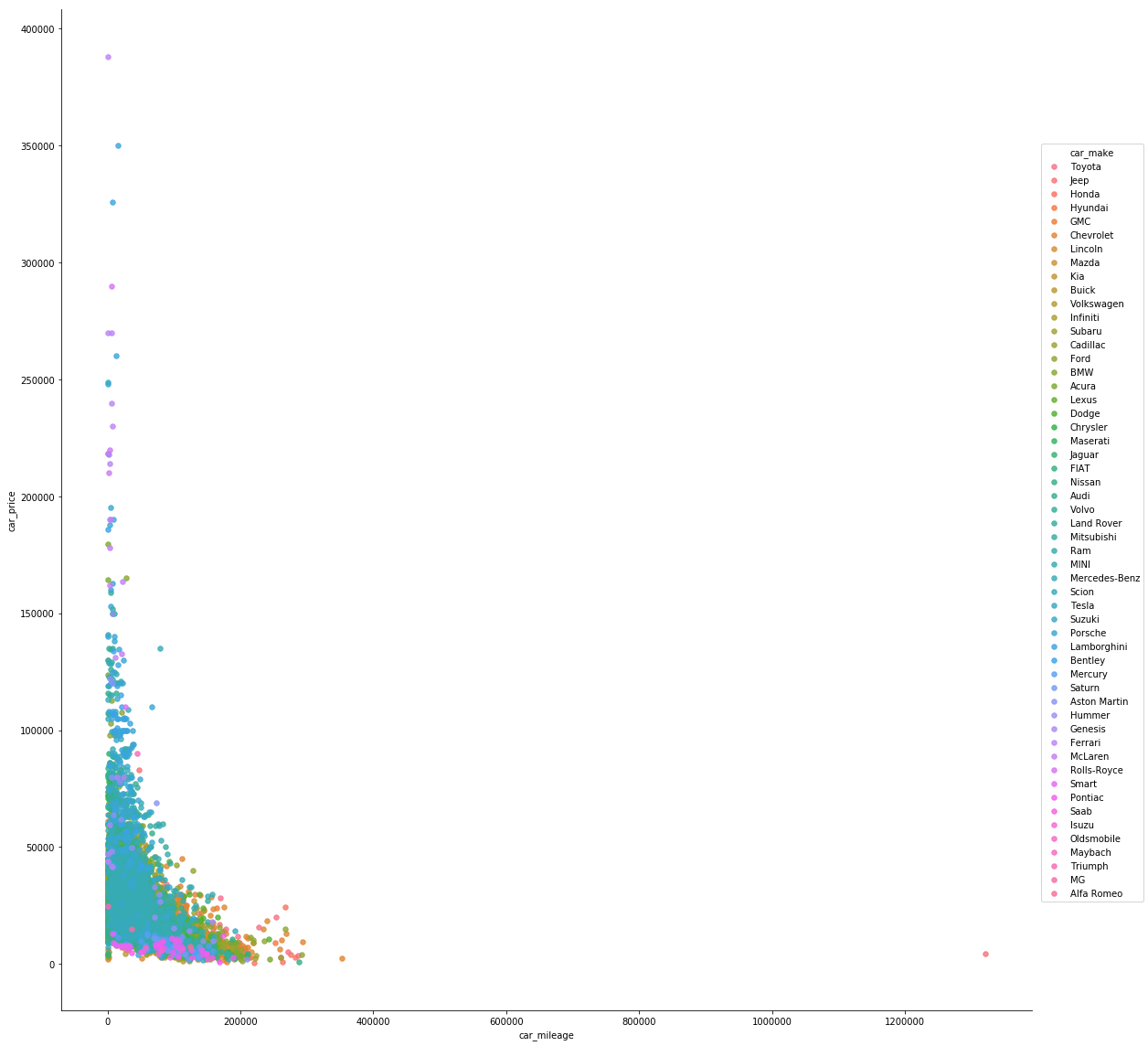
**Completed tasks:**

The majority of the time and effort spent since the last Progress Report has been on cleaning and merging the data set. For this task we have worked in Jupyter Notebook using pandas and numpy. The major issues that needed to be resolved were the removal of NaN values, resolving values that were placed into the incorrect column, removing duplicate rows, and resolving instances where there were negative values under the car\_mileage column. Other cleaning involved correcting technical duplicates of car\_make (ie, INFINITY & Infinity), input needed

* We have completed some initial exploratory visualization analysis using Seaborn:







**In Progress Tasks:**

* Perform analysis of the data to find interesting correlations
* Apply statistical models (linear and Bayesian regressions)
* Properly store the dataset

**To-Be-Completed Tasks**

* Apply Machine Learning models for predictions
* Determine method for classifying similar cars
* Build further visualizations