**Identification of significant variables to drive the price of used cars on eBay**

**Nitin Mahajan, DSC530 – Bellevue University**

**Data Set**

Used Cars Database from Kaggle

<https://www.kaggle.com/piumiu/used-cars-database-50000-data-points>

**Question**

Identification of significant variables to drive the price of used cars on eBay

**Outcome of your EDA**

There is significant correlation exist between certain variables, however we have to be cautious to make any conclusion. We are able to see strong relationship between car price and other factors like power of the engine, number of years old and number of days listed online.

**What do you feel was missed during the analysis?**

If we had another data set with some other country, that would be nice to compare how the variable effect the car sale in one country vs other.

**Were there any variables you felt could have helped in the analysis?**

If we could see actual selling price of the cars it would add our ability to validate and test our model. Also, if we could find association between the loss of price and the years old and model etc variables.

**Were there any assumptions made you felt were incorrect?**

I could not find the units of the variables in metadata or in the data sheet. I assumed all the units e.g. for odometer, I assumed it is in kilometers.

**What challenges did you face, what did you not fully understand?**

I understood and was able to apply Python libraries (like pandas, matplotlib, seaborn etc) to explore the data and make the visualization. For calculating thee PMF, CDF etc I used the thinkstat 2module. I believe, I did a fairly good job to understand and explore the data I chose.