The Hypothetical question I have considered as part of this course for the project is if the car prices are actually related to the no. of kms a car has driven also called as mileage in U.S.

I took the Kaggle data from a website called cardekho that has the details of used cars. I took 5 variables out of 13 variables that are present in the data set to perform the Analysis. Based on the steps performed, it seems that the car prices are indeed related to the no. of kms driven. They both are negatively correlated. As the no of kms driven increase, the selling price of the car decrease automatically. The correlation value I got is -0.225 which means that even though they are negatively correlated, the relation seems to be weak one based on the data set we analyzed. R2 is ~0.05 which means only 5% of the total data is predictable from the independent variable. More Analysis needs to be done to understand if there are any other variables as well that are affecting the selling price of a car.

I felt that the analysis could have been more extensive if I know more tests that can be performed on the data set. I feel that some data cleaning and transformations would have helped in performing the analysis much better. There are few other variables that are left out in the data set such as engine, max power and torque because they are not integers. Some transformations such as converting all of them into the same units would have helped the analysis much better in my opinion.