*Statistical/Hypothetical question*:-

**The higher the price of the car the higher the deprecation.**

*Outcome of EDA*:

**Up on close inspection, Mercedes- Benz make which is higher in price in fact shown lower depreciation compared to other high priced luxury cars like BMW , Jaguar and Porsche.**

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

***Felt the size of the data is not large enough to decipher the permutations***

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

***Insurance rating would have helped to see the safety of the car & would play a role on deprecations aspect.***

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

***The assumptions of higher priced luxury cars would depreciate lot higher than regular cars was found to be incorrect.***

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

***Logo CDF was not as flat line as I thought it would be. Have to read it over and over to understand if the approach was correct and has to go back and forth to cross verify the understanding.***

*Submit a link to your repository to the assignment link during the final week of class.*

[Data Set – Kaggle Auto Mobile Data set](https://www.kaggle.com/toramky/automobile-dataset/version/2) - https://www.kaggle.com/toramky/automobile-dataset/version/2

Context

This dataset consist of data From 1985 Ward's Automotive Yearbook. Here are the sources

Sources:

1) 1985 Model Import Car and Truck Specifications, 1985 Ward's Automotive Yearbook.  
2) Personal Auto Manuals, Insurance Services Office, 160 Water Street, New York, NY 10038  
3) Insurance Collision Report, Insurance Institute for Highway Safety, Watergate 600, Washington, DC 20037

Content

This data set consists of three types of entities: (a) the specification of an auto in terms of various characteristics, (b) its assigned insurance risk rating, (c) its normalized losses in use as compared to other cars. The second rating corresponds to the degree to which the auto is more risky than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is more risky (or less), this symbol is adjusted by moving it up (or down) the scale. Actuarians call this process "symboling". A value of +3 indicates that the auto is risky, -3 that it is probably pretty safe.

The third factor is the relative average loss payment per insured vehicle year. This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/speciality, etc…), and represents the average loss per car per year.

Note: Several of the attributes in the database could be used as a "class" attribute.

This data set consists of three types of entities:

(a) Technical specs of an auto in terms of various characteristics

(b) Automobile assigned insurance risk rating

(c) Normalized/Depreciation losses in use as compared to other cars

* Symbolling rating: It is a degree to which autos are marked based on the risk and price index. It could go up and down. It’s also referred by actuaries/insurance auditors as "symbolling"
* Normalized losses: Depreciation losses in use as compared to other cars
* Make: Manufacturer name
* fuel-type: Fuel type uses by auto. Gas or Diesel
* Aspiration: Naturally aspirated engine or turbo charges/super charged
* Num-of-doors: Number of doors to the automobile
* Body-style: Sedan/Coupe/SUV/Hatchback/Hardtop...etc
* Drive-wheels: Front wheel drive, rare wheel drive, all wheel drive
* Engine-location: Location of engine
* Wheelbase: Length of the car from front wheel to rare wheel
* Length: Length of the car from bumper to bumper
* Width: Width of the car side to side
* Height: Hight of the car from ground to roof
* Curb-weight: Empty vehicle weight with fuel
* Engine-type: Engine type, dual overhead, single overhead cams, etc
* Num-of-cylinders: Number of cylinders in the engine 4,6,8 or 12
* Engine-size: Volume of air and fuel that's pushed through the engine by its cylinders fuel-system
* Bore: Diameters of each cylinder
* Stroke: stroke is the length that it travels when moving from bottom position to the top position.
* Compression-ratio:  Ratio of the volume of the cylinder and the combustion chamber when the piston is at the bottom, and the volume of the combustion chamber when the piston is at the top.
* Horsepower: The power an engine produces
* Peak-rpm: Rotation per min of the engine at its peak
* City-mpg: Millage per gallon in city driving conditions
* Highway-mpg: Millage per gallon in highway driving conditions
* Price: Price of the vehicle