

EDA Report

Self-Driving Car Training-Dataset

Introduction:

This data set is taken from Kaggle self-driving car for training a car. This dataset contains the images for right, left and center. The images are taken from a camera installed on the windshield of a car.

Attributes of Data Set:

This data set contains following attributes

1. Right side images
2. Left Side Images
3. Center Images
4. Steering angle
5. Throttle
6. Brake
7. Speed

This dataset contains following data

1. 8036 images of left side
2. 8036 images of Right side
3. 8036 images of center

Total 8036 Rows are there in this dataset. Based on these images Steering angle, Throttle, Brake and speed are analyzed.

Techniques to be applied for EDA on this Dataset:

On this dataset we will use Univariate Analysis. In this we can analyze the following attributes

1. Find the central tendency (mean, median, mode)
2. We can find missing values and outliers
3. Total sum of each attribute.

By this observation we can understand and deal with other side images.

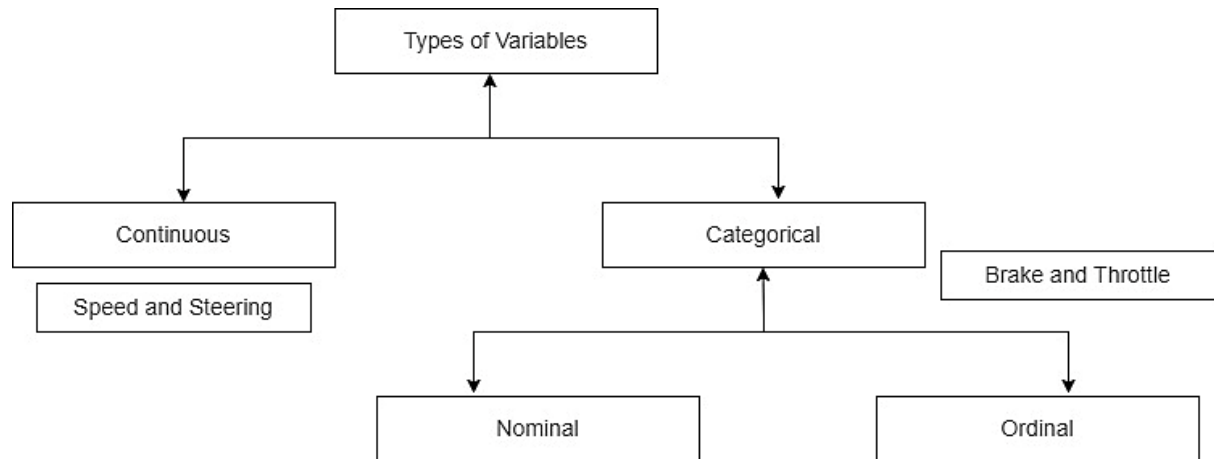
On this dataset we can only apply univariate analysis but not bivariate because it includes images that can be analyzed using OpenCV. Practical we can perform EDA on images by using OpenCV and Pandas Library in python.

In this dataset we have **continuous** data variable like **speed and steering** so here we will use two types of method to perform univariate analysis.

1. Tabular Method: used to describe central tendency and missing values

2. Graphical Method: we can use Histogram for understanding of distribution of data and Boxplot is used for outlier detection.

For **categorical variables** like **brake and throttle** we will use same methods as we have used for continuous variable, but we will only use Bar plot only.



Group Members:

Sagar Kumar

Gul Munir Hassan