Analyzing and Visualizing Data

Introduction

This report aims to provide insights and visual display based on the analysis made in the project of the data formed after all the three part of data wrangling

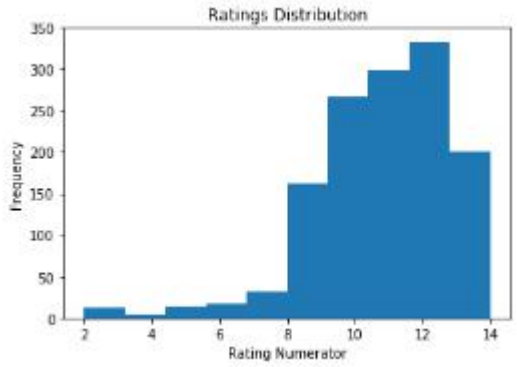
Analysis and Visualizing Data

- The Dataset has 1343 rows and 30 columns
- The dataset info suggest there are some null columns
- There are some missing value in some columns of some rows
- The describetion of the dataset tells us that mean rating_numerator is 12.12 and rating_denominator is 10.0 and mean retweet is 2717.324 and mean likes 9350.049
- The total max number of retweet is 60674 and likes 123762
- The p1 predicts golden retriever as most of the time

- The p2 and p3 predict Labrador Retriever most of the time

- The highest rating is 75/10 and lowest is 2/10 but there is an outliner is 1776/10

- We also made a graph of frequency vs rating numerator which says that the rating 12 is awarded by most people



- The rating_numerator 1776 is retweeted 2693 and 5482 likes
- The rating_numerator 2 is gives to 2 dogs is retweeted 2770 and 896 and 5482 and 1699 likes
- The rating_numerator 75 is retweeted 6810 and 19789 likes
- The scatter plot drawn between likes and retweets to find what type of relation is between them and there is positive correlation between them which says that as likes increases retweet increases and vice versa