Wrangling and Analytical Report

A good data wrangler knows how to integrate information from multiple data sources, solve common transformation problems, and resolve data cleansing and quality issues.

Introduction

Data Analyst Nanodegree Program from Udacity taught me various data wrangling methods that are being used in the today's data technology world and given an opportunity to showcase my skills and encourage me to use these learned techniques in different projects under this nanodegree program which are part of real world data.
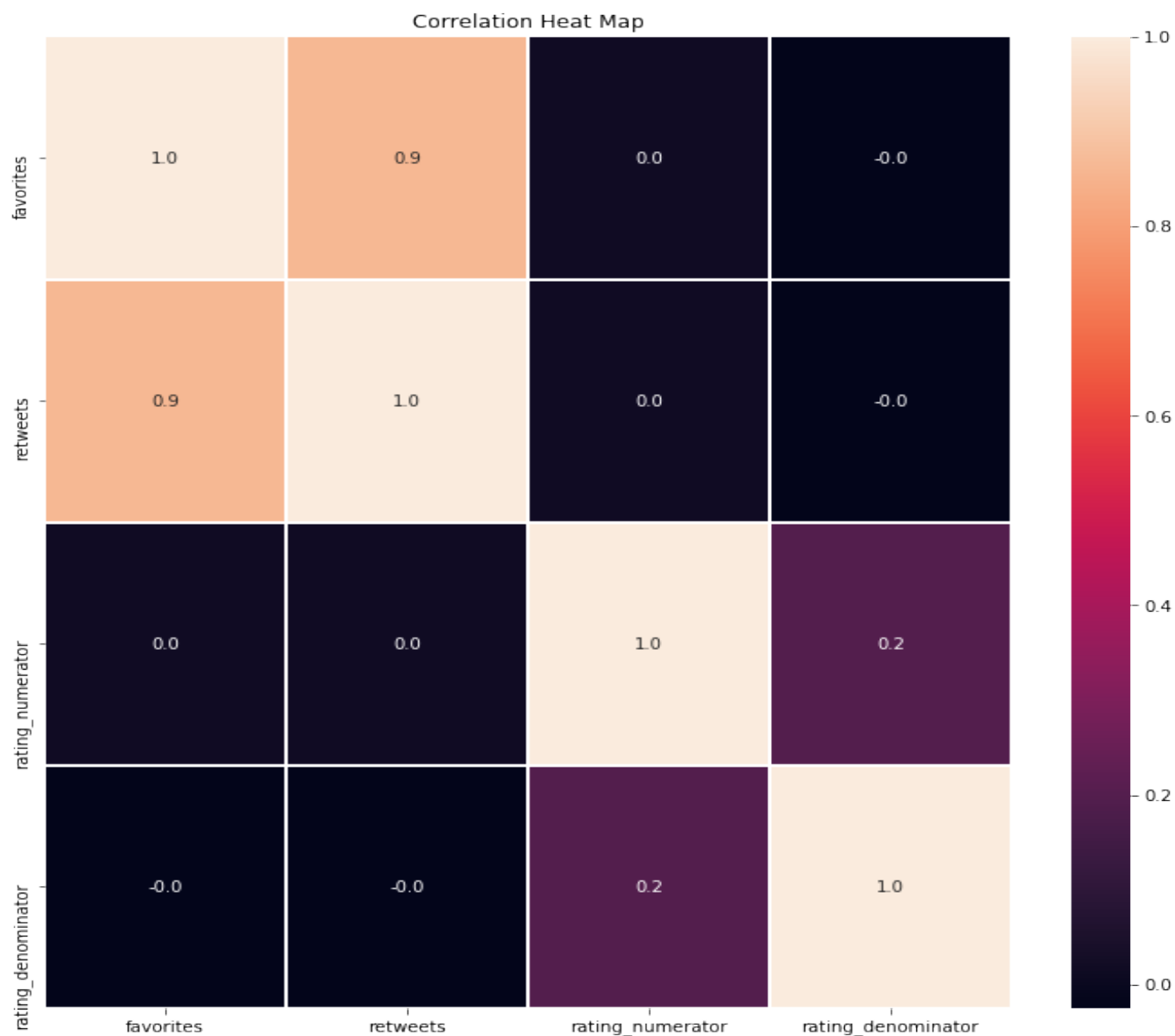
This project aimed to analyse twitter WeRateDogs community where we need to draw meaning insights from the data gathered and try to answer some questions such as most favourite dog name in top 10 or most popular dog breed in all the tweets archived.

WeRateDogs is a Twitter account that rates people's dogs with a humorous comment about the dog. These ratings almost always have a denominator of 10. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc. Why? Because "they're good dogs Brent." WeRateDogs has over 8 million followers and has received international media coverage.
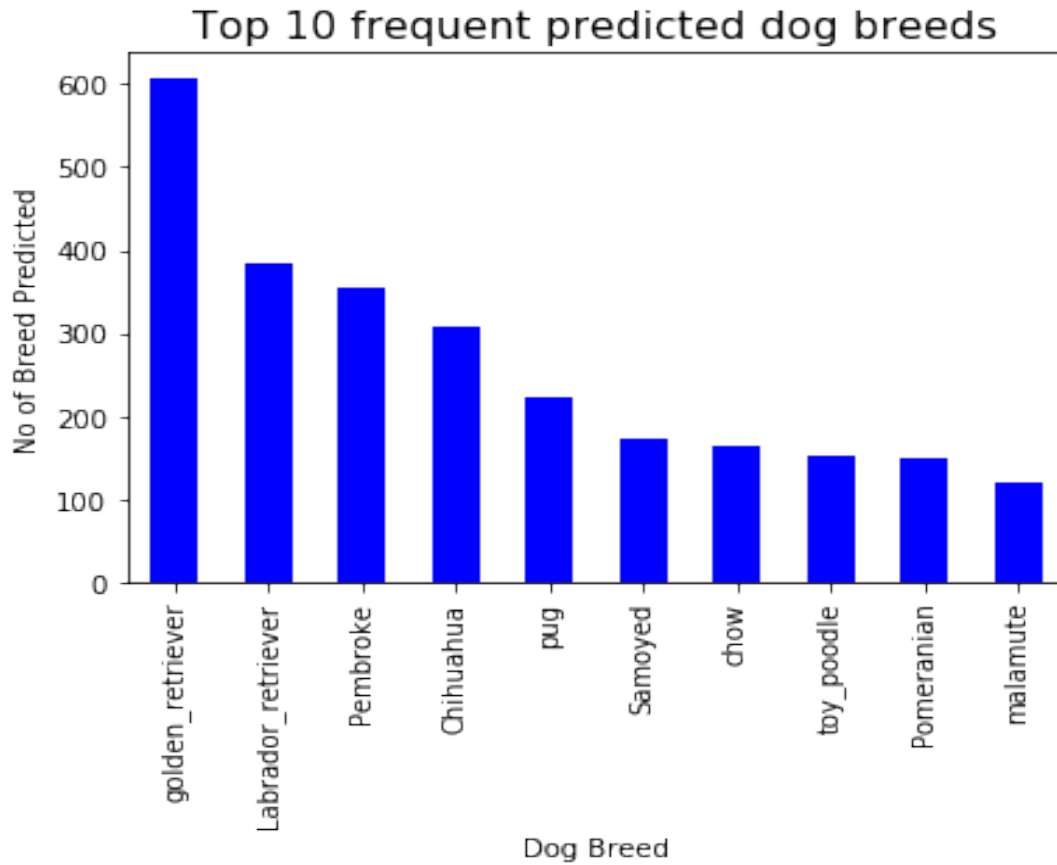
Analysis and insights

1.Using heatmap tried to check correlation between the factors such as retweets and favorites

Analysis

Strong correlation between favorites and retweets, this is as expected.
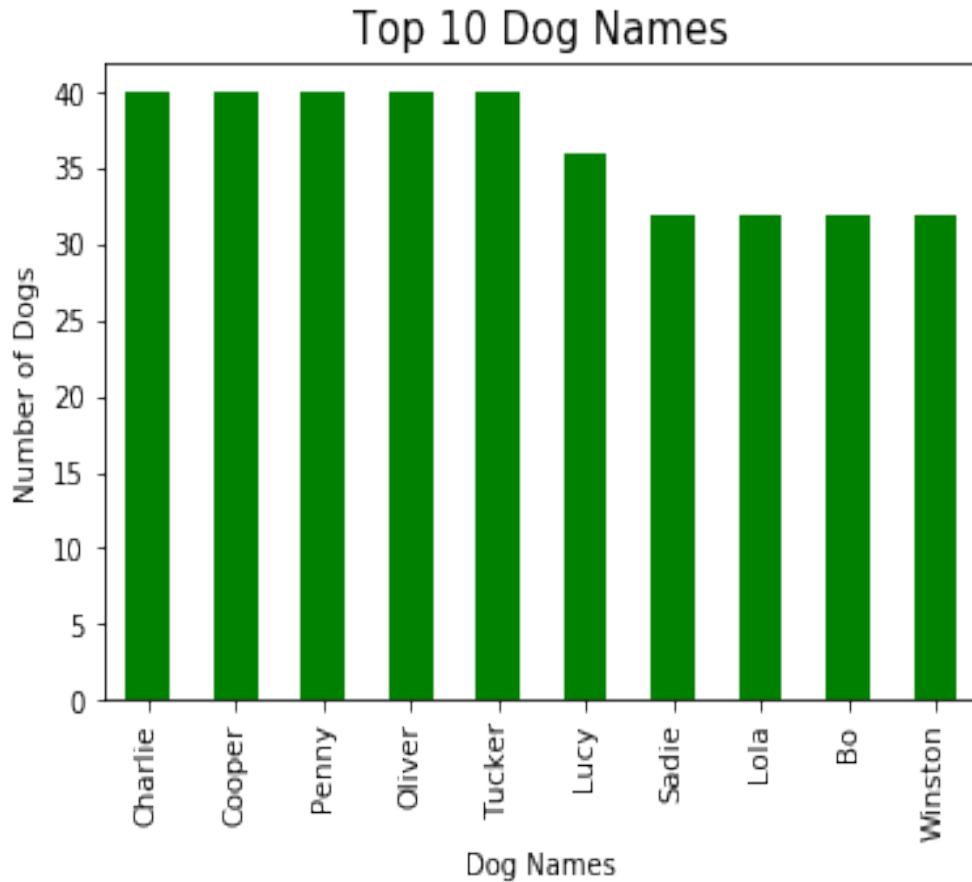
2. Frequent top list for predicted dog breeds

Top 10 frequent predicted dog breeds

Analysis

Golden_retriver seems to be the most frequent predicted list in top 10 with over 600 times predicted correctly

3. Top 10 dog names

Analysis

Penny rocking the favourite top 10 dog names list and almost all top 5 names i.e. Penny, Oliver,Charlie, Tucker and Cooper are rulling top 10 dog names list