Dogs Wrangle & Analyze Data

Essam Almasad

11/2/2019

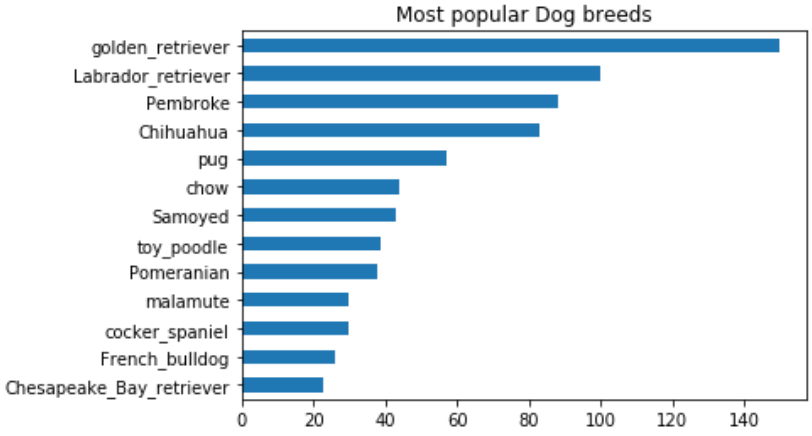
In this project I gathered , assessed and cleaned data. There will be visualization and analysis to help us reach to the best form of completing this project. The data will be provided offline from Udacity, because there is not enough time for me to contact Twitter. The subject of this project was dogs in twitter in matter of dog types, pictures, likes, retweets etc…

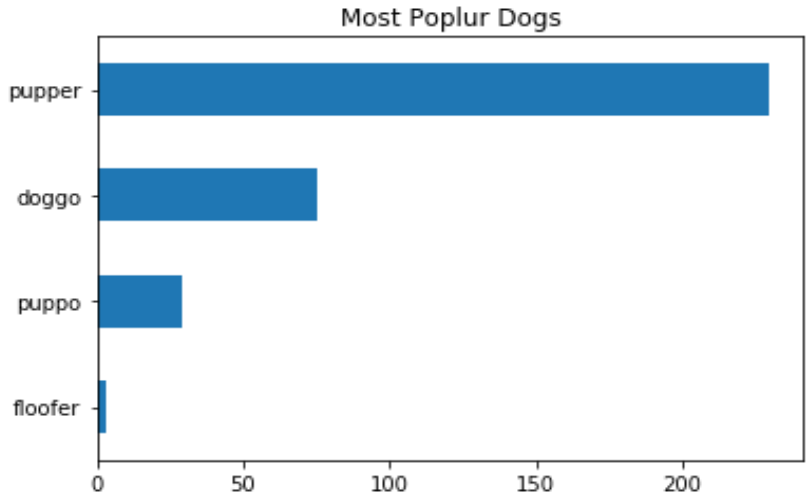
For this analysis, I gathered data from three different sources. Then, I stored these data in new tables to help me manipulate the data in a better way. The first file was csv file provided to me by Udacity. The second file was a tsv file also provided by Udacity. For the third file, which I was supposed to get by contacting Twitter; but due to the limited time I had, I used the tweet json text file also provided by Udacity.

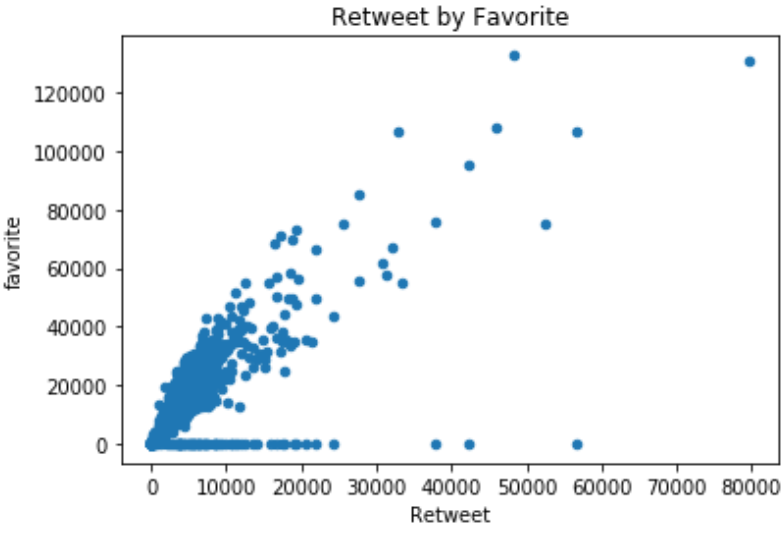
In the assessing part of the project, I provide all the necessary visual and programmatic assessments to help showing the three tables in a clear way. Moreover, by addressing the data quality and tidiness issues, the assessing was complete and we are ready for cleaning the data.

The cleaning part of the data was based on what we highlighted in the previous step. Make the data more tidy and clean. By merging all three tables in one, removing unwanted columns for this analysis, and extracting the once cleaned data to a new separate data csv file.

Finally, we created visualizations to show and compare between columns in our new cleaned data. As shown below, the three visualizations for our analysis. The first visual is about the most popular dogs and how it was determined after the data was cleaned. Second was the retweets vs the favorites which has this compliment relationship, with higher retweets comes higher number of favorites. The third one is about the most popular dog breeds.


Most popular Dog breeds
golden_retriever
Labrador_retriever
Pembroke
Chihuahua
pug
chow
Samoyed
toy_poodle
Pomeranian
malamute
cocker_spaniel
French_bulldog
Chesapeake_Bay_retriever
0
20
40
60
80
100
120
140


Most Poplur Dogs
pupper
doggo
puppo
floofer
0
50
100
150
200


Retweet by Favorite
favorite
Retweet
0
20000
40000
60000
80000
100000
120000
0
10000
20000
30000
40000
50000
60000
70000
80000