DAND: Data Wrangle and Analyze

Wrangle Report

Introduction

Data Analysis nanodegree illustrates the concept of Data analysis and how to move over all processes to get clean and understandable data. For over years, data has become more valuable to solve problems in real life and male accurate decisions. Data analysis programs show the best to take the data from early stage and agater to clean and get some available information. The process starts with data wrangling which is when you gather the data from several resources and assess then clean. This report will show how we performed the data analysis over Udacity project 5 or what we call it (we rate the dog project).

Data Analysis

After cleaning the Data I have become able to implement the analysis and investigate the facts. I have implemented three insights that might be helpful for me. The first insight about top five breeds got highest count of favorite and retweets which they are (labrador_retriever, eskimo_dog, chihuahua, lakeland_terrier and english_springer) with different numbers of favorites and retweets but they are the highest. For the second insight I decided to count the value of the new column that I created based on the ratio between two rating columns. Then I created a new category rating based on the results of that division (ratio of dividing two rating columns). I notice that most of it was poor (see figure 1) which I can observe how low users rate some of the pics. The third insight for which platform is the most popular or has been used the most is tweets or rait and that for Iphone. At the end I decide to visualize top five breeds in prediction model (see figure 2).

```
# vaalue counts of my new raiting category
df.users_rating.value_counts()
```

```
0     1214
1      761
2       49
3       19
5       10
10       4
Name: users_rating, dtype: int64
```

```
df.rating_category.value_counts()
```

```
poor         1214
good          761
nice           68
excellent      10
exotic          4
Name: rating_category, dtype: int64
```

Figure 1: show category rating based on new user rating.

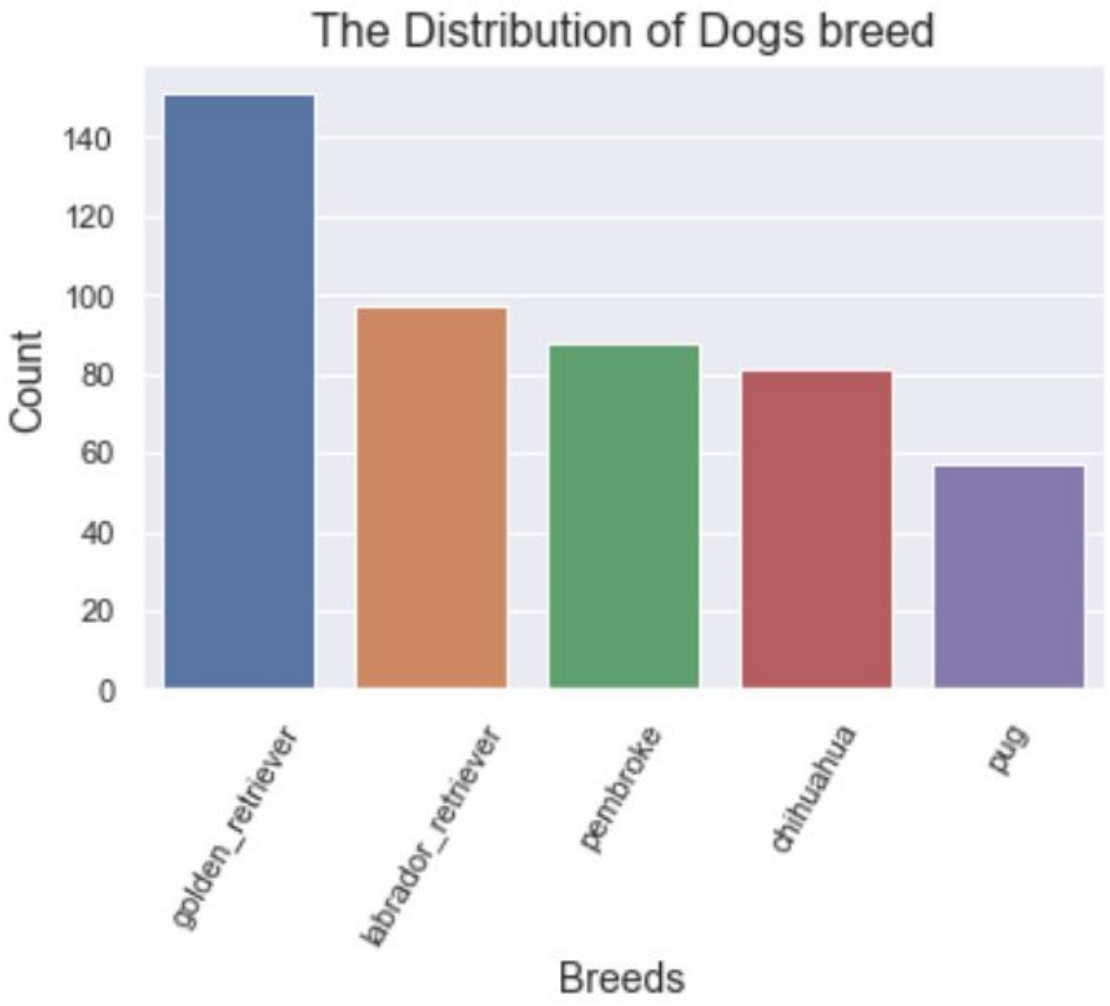

Figure 2: this visualization show top five breeds in prediction model