



# **Data Analyst Program**

## **Wrangle and Analyze Project**

**By**

**Aisha M AL-mutairi**

**In gathering step :**

Gathering three pieces of data set using Pandas , requests and tweepy

**The output from this step is :**

Display and reading the three pieces of data set

**In assessing step :**

I was use two types of assessment : Visual assessment and Programmatic assessment to detict and doucument quality and tidiness issues using some tools like (.info ) ,(.sample) and (.sort\_values)

**The output from this step is :**

Extracted eight quality issues :

- issue 1: we don't need to "in\_reply\_to\_status\_id","in\_reply\_to\_user\_id" ,"retweeted\_status\_id" ,"retweeted\_status\_user\_id" and "retweeted\_status\_timestamp" columns
- issue 2: data tybe for timestamp is a string not a date time
- issue 3 : data type for " rating\_denominator"and " rating\_numerator " are an integer not a float
- issue 4 : data type for " tweet\_id " is an integer not a string
- issue 5 :data type for " doggo , floofer , pupper , puppo " are a string
- issue 6 : Missing — Null — NaN Values in name column
- issue 7 :source column needed to be clear more
- issue 8 :some of rating\_numerator is higher than 12
- issue 9 :some of rating\_denominator is higher than 12
- issue 10 :tweet\_id column most hav the same data type for each table

**And two tidiness issues :**

- issue 1 : merge archive, image\_ and tweet\_df
- issue 2 Restructuring image\_ table

**In the cleaning step :**

After taking a copy from data set

I was clean all issues that I document and make a test for each issue

**The output from this step is**

Successfully cleaned all issues

**In the storing step :**

I have saved the gathered , assessed and cleaned to a CSV file named ("new\_table ")

**The output from this step is**

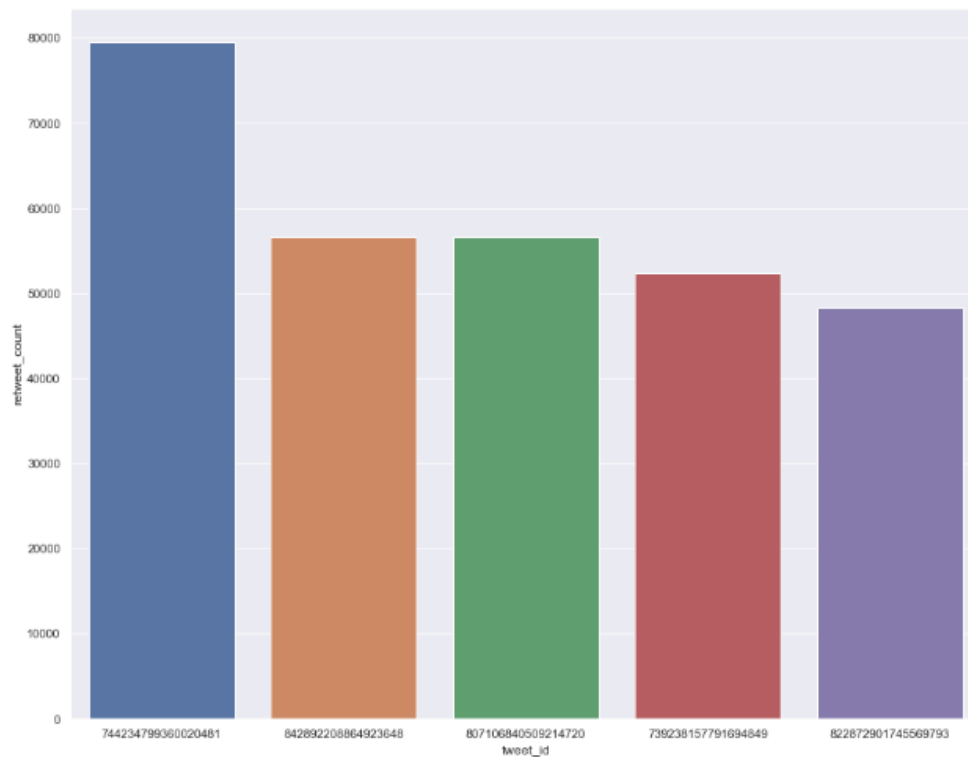
Successfully saved the file

**In Analyzing and Visualizing Data step:**

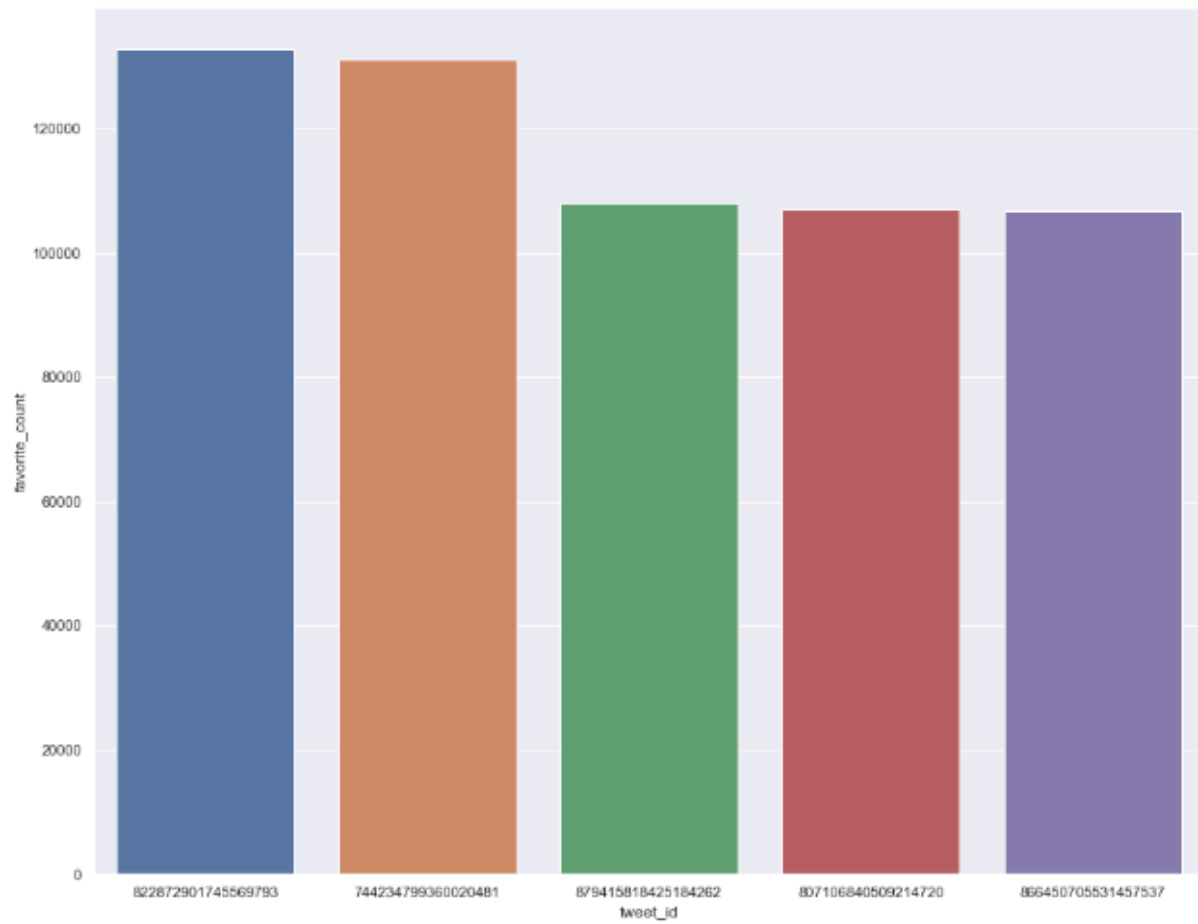
produce at least three (4) insights and make a visualization

**The output from this step is**

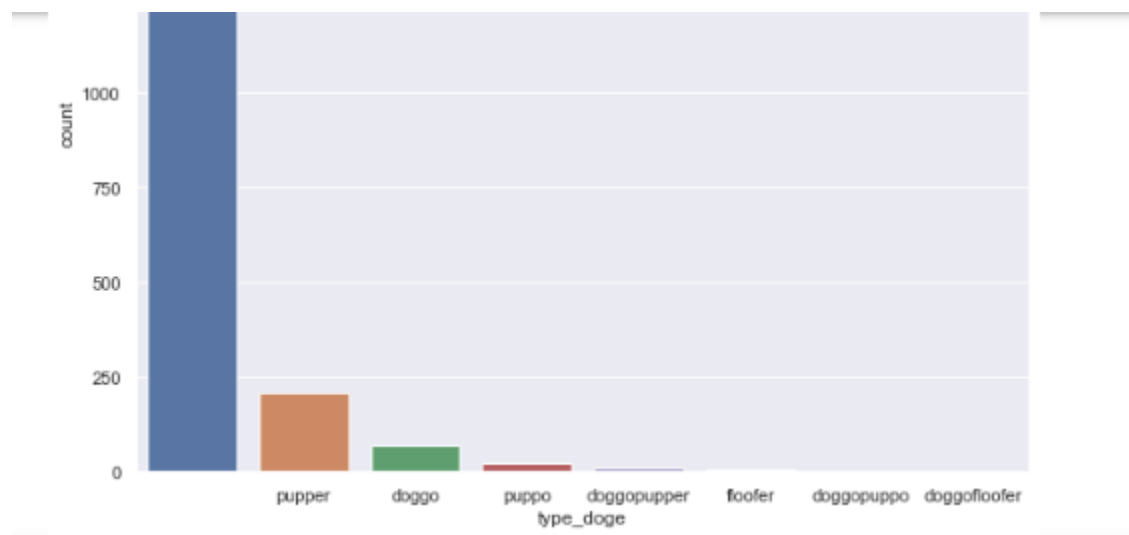
**What is the top 5 tweet's has retweets**



What is the top 5 tweet's has likes



What is The most popular types of dogs from high to low



The relationship between Retweets & Favorites

