Report for Data Analysis and Visualisation of the WeRateDogs Data

Background

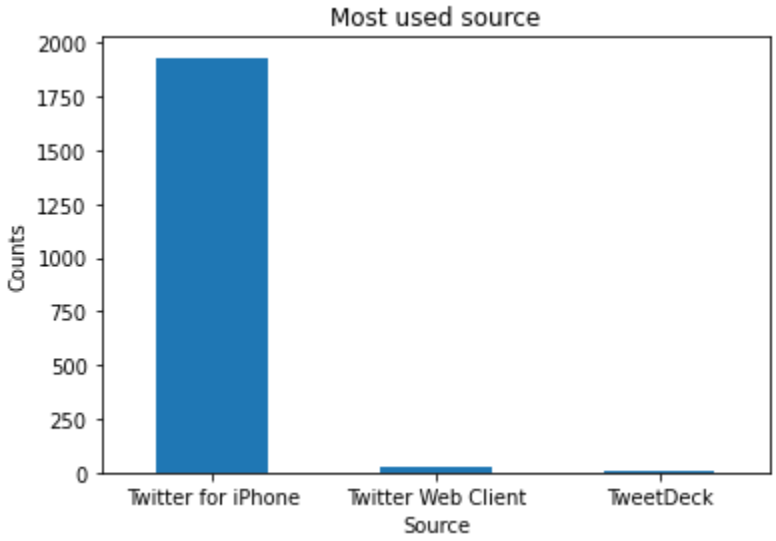
The aim of this project is to put knowledge and skills gained in the Data Wrangling course to practise. Therefore, major concepts and techniques discussed in the course have been used in the **Wrangle and Analyze Project** to practise data wrangling skills.

The datasets used are twitter archive data and image predictions relating to the tweet history of WeRateDogs, also known as Twitter user @dog_rates. WeRateDogs is a Twitter account that rates users' dogs and includes a cute story about the animal.

This report communicates the insights derived from performing analysis and visualisation on the cleaned master dataset. Below are some of the insights derived after performing Exploratory Data Analysis (EDA) and Data Visualisation on the master cleaned dataset:

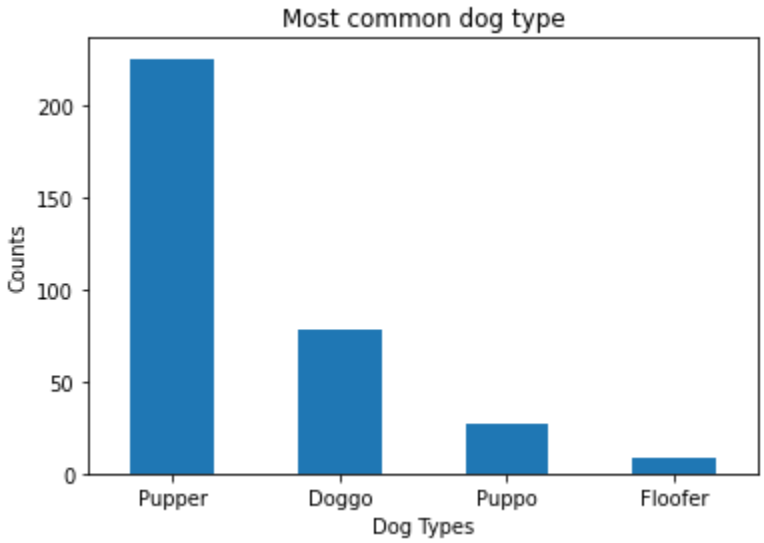
Insights

1. **Which Twitter source is the most used source?**



Narrative: *Twitter for iPhone* is the most used source. From the chart above, it can be deduced that the majority of users that tweeted made use of iPhones.

2. **Which dog stage is the most common?**



Narrative: *Pupper* is the most common dog type (stage). From the chart above, it can be deduced that the majority of dogs rated are in *Pupper* stage.

3. **Does retweet count correlate with the favourite counts of tweets ?**

	retweet_count	favorite_count
retweet_count	1.000000	0.913014
favorite_count	0.913014	1.000000

Narrative: From the analysis above, it can be concluded that there is a strong positive relationship between retweet counts and favourite counts of tweets because the correlation score is 0.9 (close to 1).

Conclusion

I have developed data wrangling skills through the tasks performed in this project