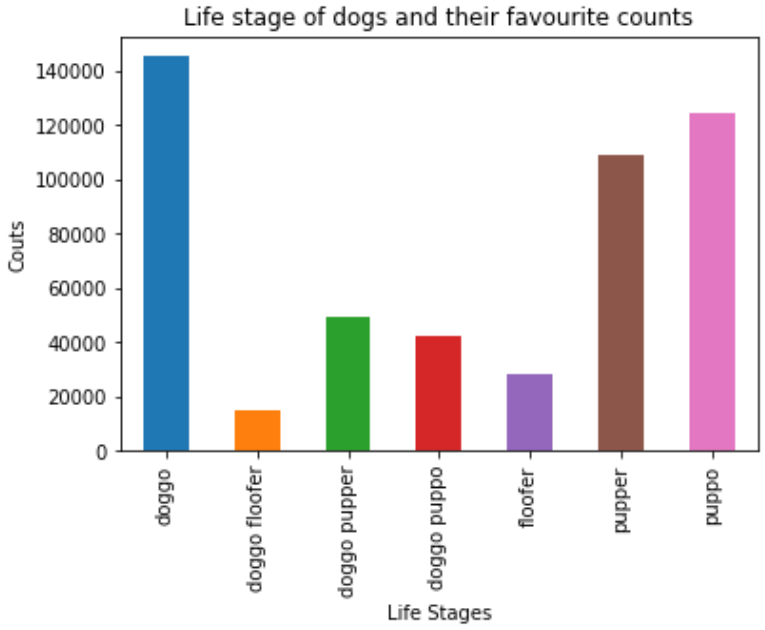
Udacity Data Wrangling with Twitter's @WeRateDogs Account

Data analysis, according to Wikipedia is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis of @dog_rates, a Twitter account that rates people's dogs with a humorous comment about the dog was carried out. It is important to note here that @dog_rates also known as WeRatDogs has over 4 million followers and has received international media coverage.

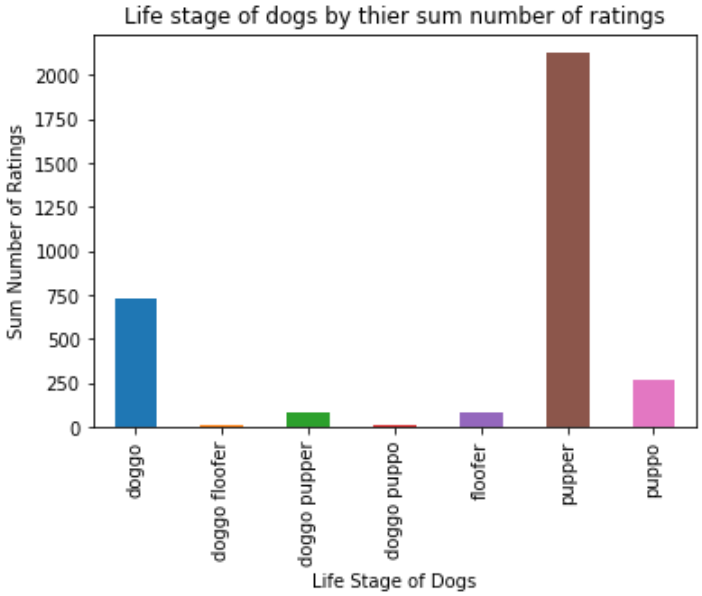
Real-world data rarely come clean. That being said, after gathering, assessing and cleaning all relevant data needed for analysis, the following questions were asked to guide the analysis of the @dog_rates dataset.

- What is the life stage of the dog with the most favourite count?
- What life stage of dog did owners feel best about?
- How many posts included the life stage of the dog?
- Does having the life stage of the dog included in a post increase the possibility of a higher favourite count?
- Is there a correlation between favourite counts and the number of retweets a post gets?

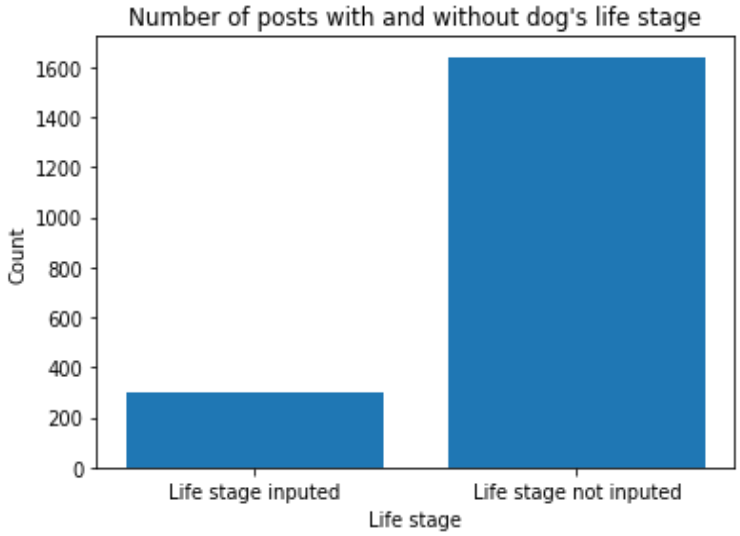
It was found out that the life stage of a dog with the most favourite counts was the doggo development stage with a maximum favourite count of 145,441 with the leas being floofer with a maximum favourite count of 28235.



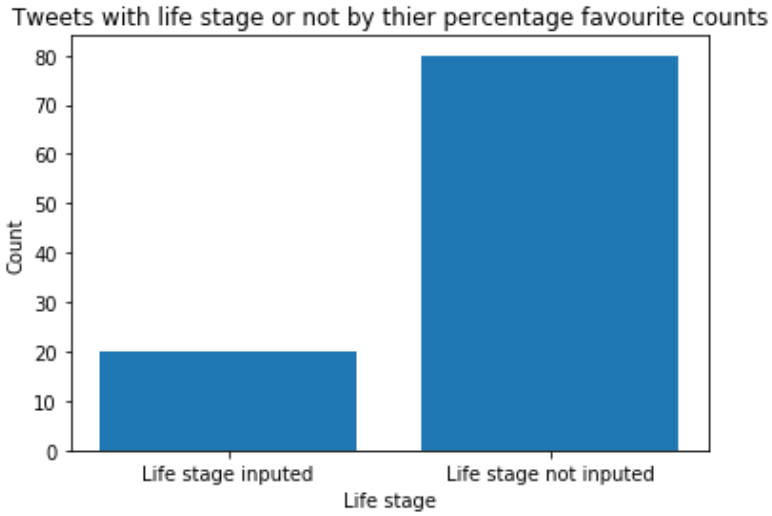
Another insight worthy of mention is that the owners of dogs who posted about their dogs and included the life stages of the dogs felt best about the dogs in the pupper life stage, getting a sum total of 2,127 vote rates by their owners. This might give an insight to business owners who sell or keep dogs as to which developmental stage of dogs to focus more on or market to certain digital spaces.



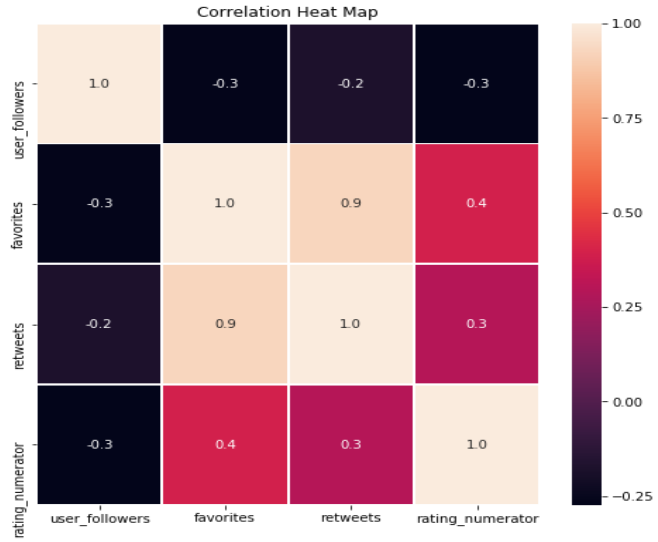
It is also important to note that there were more posts without the life stages or the developmental stages of their dogs included. This might have affected the statistics because a lot of dogs in different developmental stages could not be accounted for in the above insights.

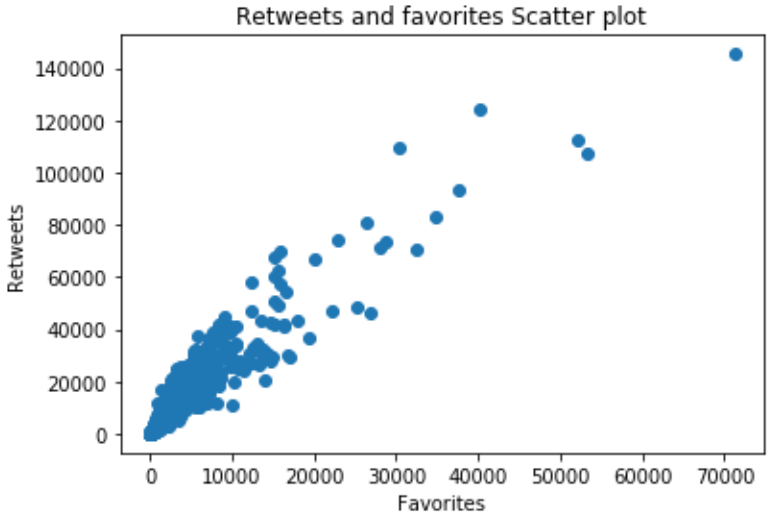


Also, having the life stage of the dog included in the post does not necessarily connote a possibility of a high favourite count. This was found out when posts without an indication of the dog's developmental stage had 80% of the total favourite counts with posts that included the life stages of their dogs having just 20% of the total favourite counts.



It was also found that there was a high positive correlation between the favourite counts and the number of times a post was retweeted with a value of 0.9. Hence, any dog owner who is interested in having more favourite counts, should solicit for retweets.





Finally, the name of the dog with the most favourite count was not given. This shows that having the name of a dog in a post does not increase the number of favourite counts it gets. Having more favourite counts is likely to be a function of the picture posted of the dog. Below is the picture of the doggo life stage dog with the mostt favourite count of 145,364.

Conclusion

Data analysis has the power to shape the way people think, act and make decisions in their everyday lives. This analysis has helped in drawing valuable insights useful for personal and business decision processes, which is the core purpose of data analysis. I hope you enjoyed going through the analysis report and hope it spurred your interest in wanting to check out the Twitter account @WeRateDogs and hopefully own a dog. Thank you for taking the time to read through this report.