

# UDACITY PROJECT- WRANGLE AND ANALYZE DATA

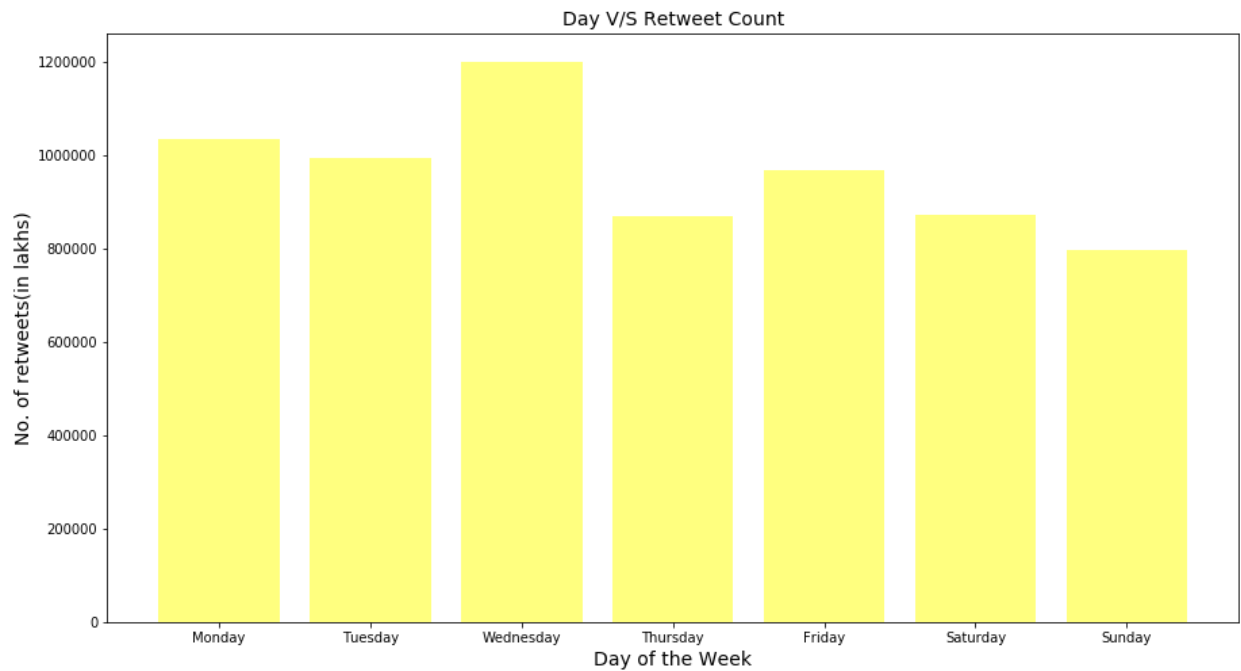
## DATA ANALYST NANODGREE

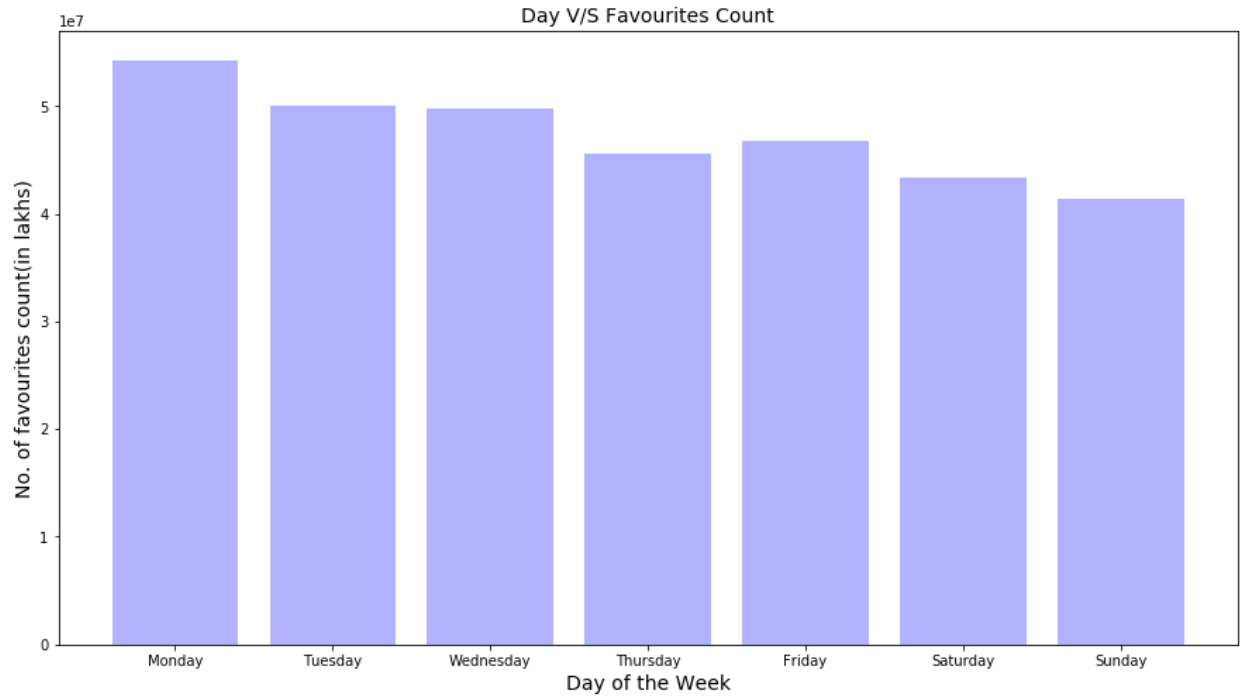
### ACT REPORT BY NEHA PATIL

#### INTRODUCTION

In this project we have to wrangle (gather, assess and clean) dataset. The dataset is a twitter archive of twitter user @dog\_rates also known as WeRateDogs. This twitter account rates people's dogs with ratings almost always greater than 10. After gathering the data, assessing it and then cleaning it we have to analyze and record our findings and insight in this document.

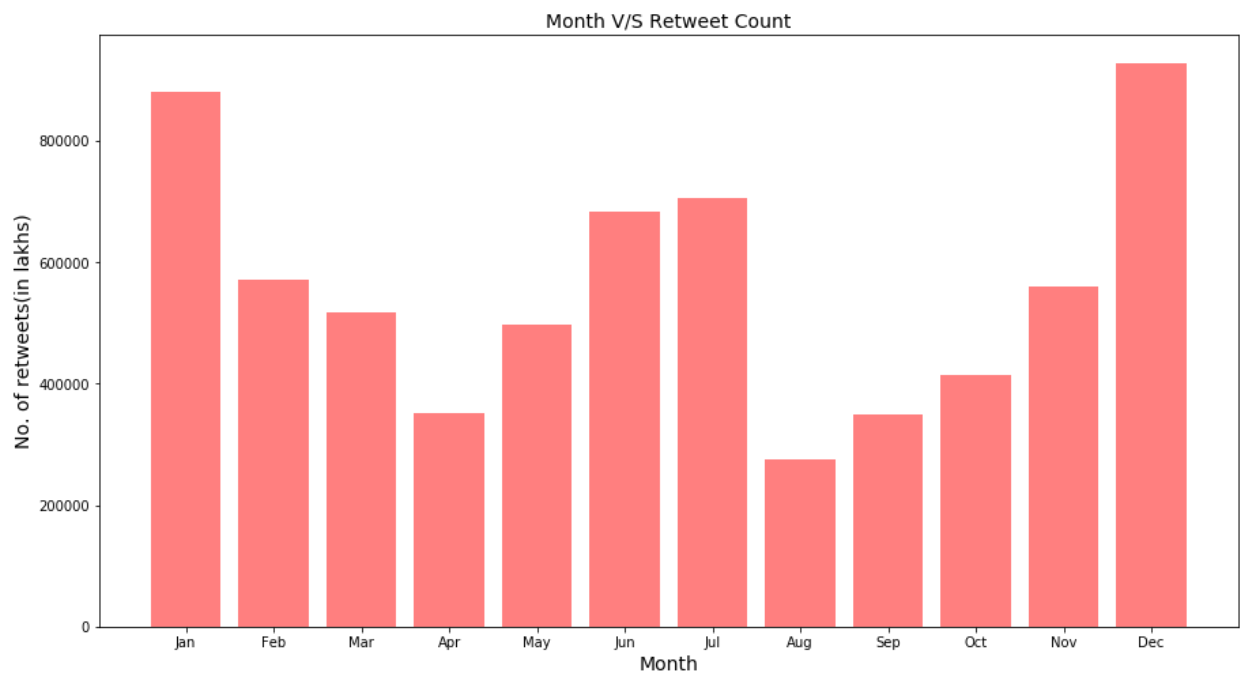
#### DAY OF THE WEEK V/S COUNT (FAVOURITES AND RETWEET)

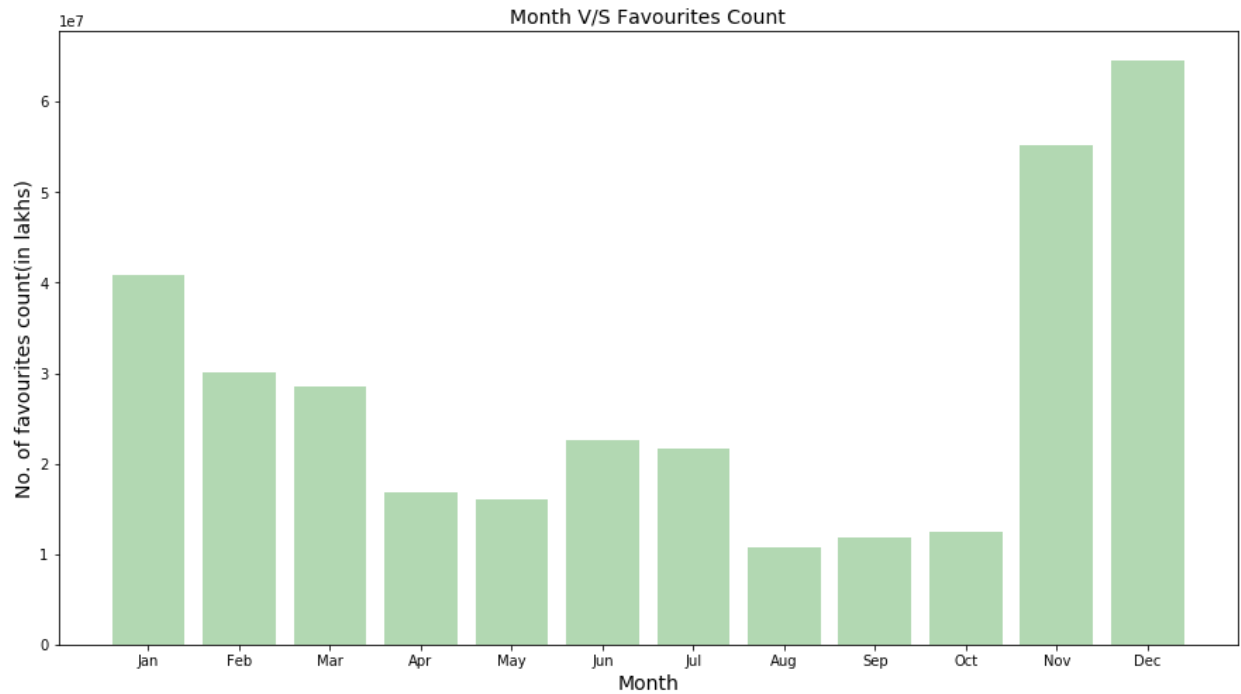




As we can see from the above two plots that the number of retweets differs from the favourites count for each day. Maximum number of retweets happen on Wednesday whereas maximum number of favourites happen on a Monday. Since we are dealing with numbers in millions, the difference in count from day to day is quite significant.

## MONTH V/S COUNT (FAVOURITES AND RETWEET)





Above bar graphs are almost similar in their plot expect the difference in count between retweets and favourites. As we can see from the two graphs that the activity on twitter or in this case for the account @WeRateDogs is more during the month of January, December, June and July compared to months of March, April, August and September. This may be because they are the holiday months and people have more time to dedicate to social media. It is highest in January and December because most parts of the world experience winter season during that time and people tend to stay indoors during time compared to June and July.

## RETWEETS COUNT V/S FAVOURITES COUNT

From the graph below we can see that retweet count and favourites count are positively co-related. This means that more the favourites count increases the more the chances of the tweet getting retweeted increases.

