WERATEDOGS

A Twitter Account Study Result

ABSTRACT

Throughout this report, and by using visuals, we will review the end result and conclusion for the study of "WeRateDogs" Twitter account data.

Maged Baheig

Udacity Student

11/13/2020

Contents

Background 3

Study Context 3

Study Objective 4

Datasets 4

Output Datasets 4

Data Analysis Result 5

1. Research Questions 5

2. Insights and Conclusion 5

Q1: What are the main devices/apps that WeRateDogs' users use? 5

Q2: Is there a relationship between dog rates and retweet count? 6

Q3: Is there a relationship between dog rates and favorite count? 6

Q4: Is there a relationship between favorite count retweet counts? 7

Q5: What time that most of tweets are tweeted at? 7

Q6: Is high confidence prediction meet reality more than low ones? 8

3. Additional EDA Figures 9

Background

Study Context

- **@dog_rates**, also known as **WeRateDogs** is a **Twitter account** that rates people's dogs with a humorous comment about the dog.

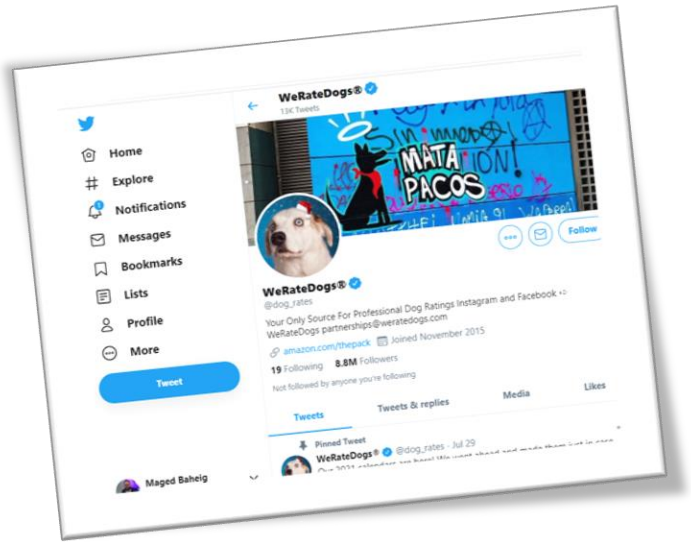


Figure 1: ***WeRateDogs Twitter Account***

- These **ratings** almost always have **a denominator of 10**. The numerators, though? Almost always greater than 10. 11/10, 12/10, 13/10, etc.

Figure 2: ***WeRateDogs Tweet Example***

- Nowadays, WeRateDogs has **over 8 million followers** and has received international **media coverage**.
- **WeRateDogs profile URL:** https://twitter.com/dog_rates?s=20

Back to Contents

Study Objective

- Wrangling and **analyzing WeRateDogs Twitter account's tweet data** (tweet ID, timestamp, text, etc.) for 5000+ of their tweets as they stood on **August 1, 2017**.
- **We will work on:**
    - **Original ratings only** (no retweets and/o no replies).
    - Tweets that **have images only.**
- Our **goal** is to **create interesting and trustworthy analyses and visualizations**. The Twitter archive is great, but it only contains very basic tweet information.

Datasets

*WE started with using **3 datasets** as "**data input**", **processed** them by **making** required assessing and cleaning actions, finally ended with 2 datasets in hand as "**data output**". Our analysis done using output datasets.*

Output Datasets

Output_File #1**: **twitter_archive_master.csv

twitter_archive_master.csv

```
1 twitter_archive_master.sample()
```

	tweet_id	timestamp	source	text	expanded_urls	name	dog_stage	dog_rating	date	time	hour	favorite_count	retweet_c
824	733828123016450049	2016-05-21 01:13:53+00:00	Twitter for iPhone	This is Terry. The harder you hug him the fart...	https://twitter.com/dog_rates/status/733828123...	Terry	NaN	100.0	2016-05-21	01:13:53	1	3489	

```
1 twitter_archive_master.to_csv('twitter_archive_master.csv', index=False)
```

*Figure 3: **twitter_archive_master.csv***

Output_File #2**: **imagee_prediction.csv

image_prediction.csv

```
1 image_prediction = image_df_clean_melted.copy()
```

```
1 image_prediction.sample()
```

	tweet_id	jpg_url	prediction_number	prediction_result	prediction_confidence	prediction_match_breed	prediction_match_breed
1785	679511351870550016	https://pbs.twimg.com/media/CW4b-GUWYAAa8QO.jpg	p1	Chihuahua	0.761972	True	

```
1 image_prediction.to_csv('image_prediction.csv', index=False)
```

*Figure 4: **imagee_prediction.csv***

Back to Contents

Data Analysis Result

1. Research Questions

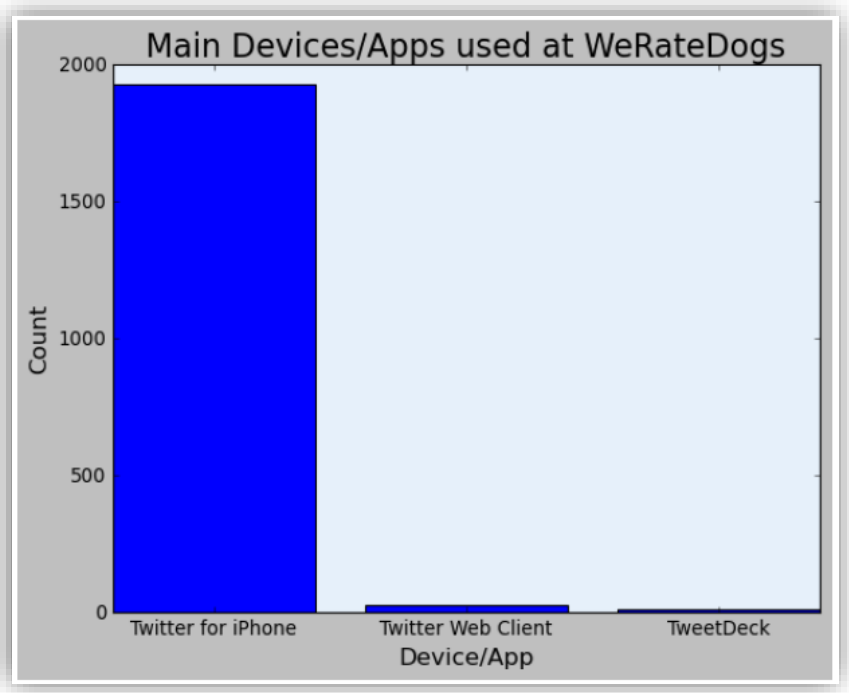
Below are what we tried to find out...

- **Q1** What are the main devices/apps that WeRateDogs' users use?
- **Q2** Is there a relationship between dog rates and retweet count?
- **Q3** Is there a relationship between dog rates and favorite count?
- **Q4** Is there a relationship between favorite count retweet counts?
- **Q5** What time that most of tweets are tweeted at?
- **Q6** Is high confidence prediction meet reality more than low ones?

2. Insights and Conclusion

Q1: What are the main devices/apps that WeRateDogs' users use?

Chart



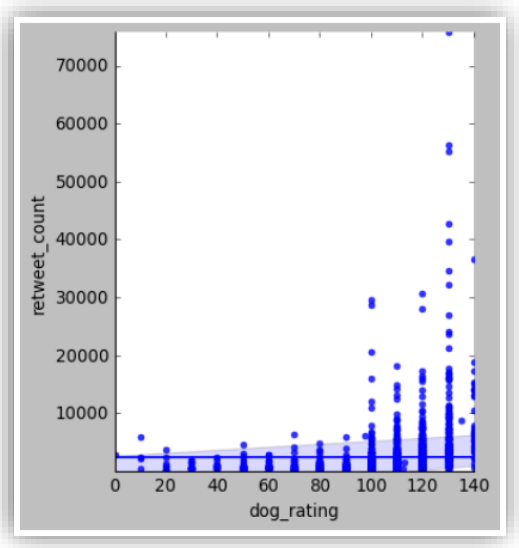
Conclusion

- **Most** of WeRateDogs' users are using **'Twitter for iPhone'**.

Back to Contents

Q2: Is there a relationship between dog rates and retweet count?

Chart

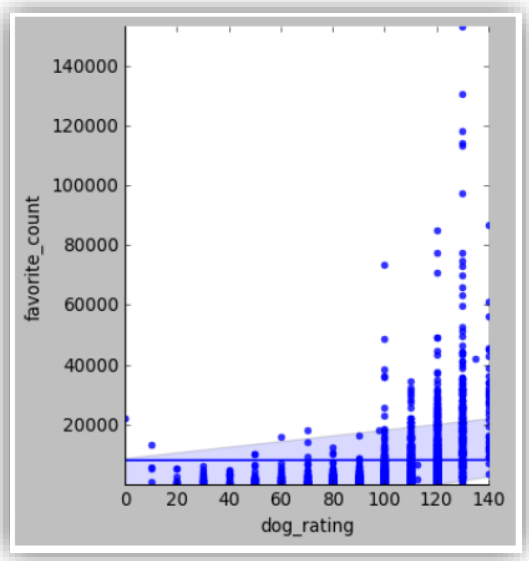


Conclusion

- It is clear that there is a **Positive relationship** between dog_rating & retweet_count.

Q3: Is there a relationship between dog rates and favorite count?

Chart



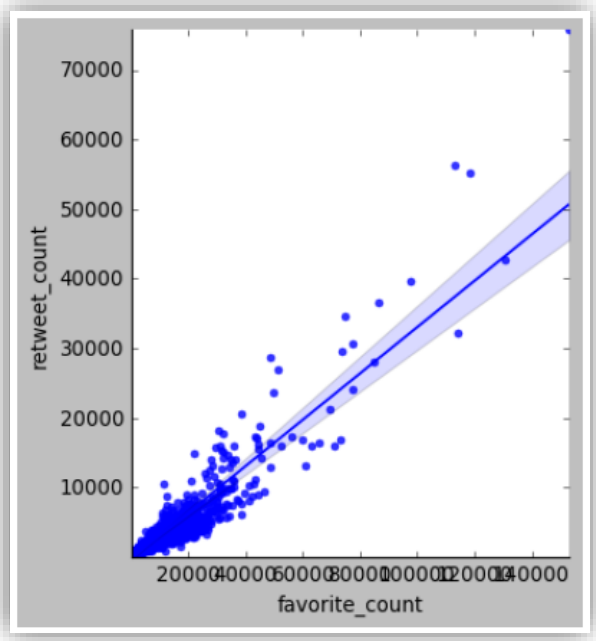
Conclusion

- It is clear that there is a **Positive relationship** between dog_rating & favorite_count.

Back to Contents

Q4: Is there a relationship between favorite count retweet counts?

Chart

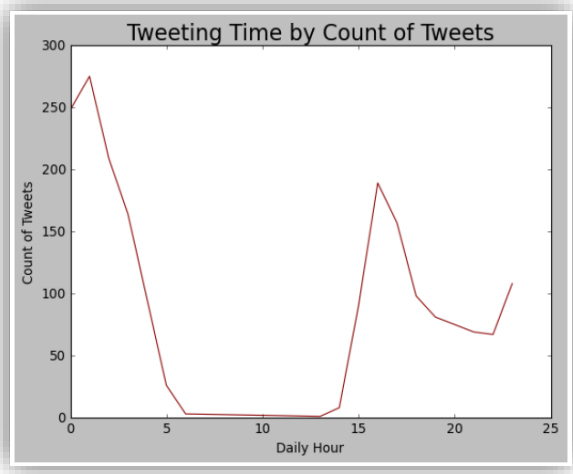


Conclusion

- It is clear that there is a **Strong Positive relationship** between favorite_count & retweet_count.

Q5: What time that most of tweets are tweeted at?

Chart



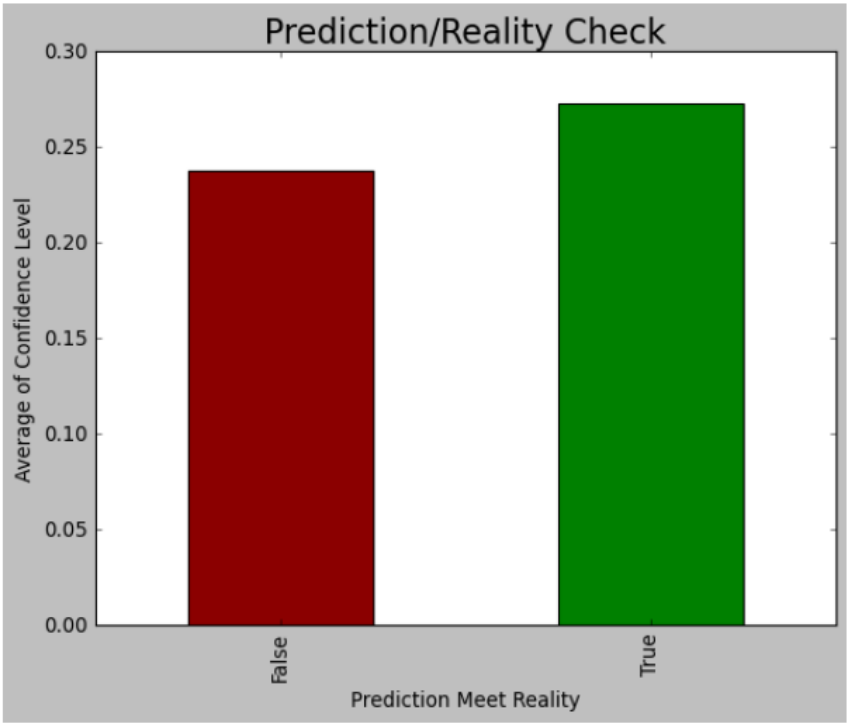
Conclusion

- The **majority of tweets** were tweeted during hours **[12:00 AM to 03:00 AM]** and **[03:00 PM to 06:00 PM]**

Back to Contents

Q6: Is high confidence prediction meet reality more than low ones?

Chart

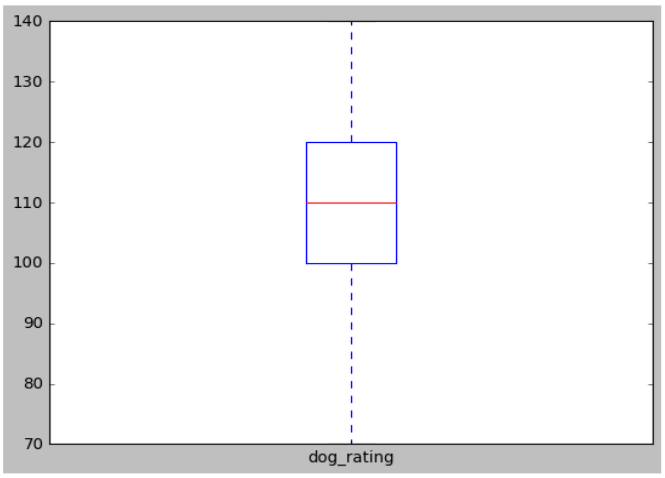


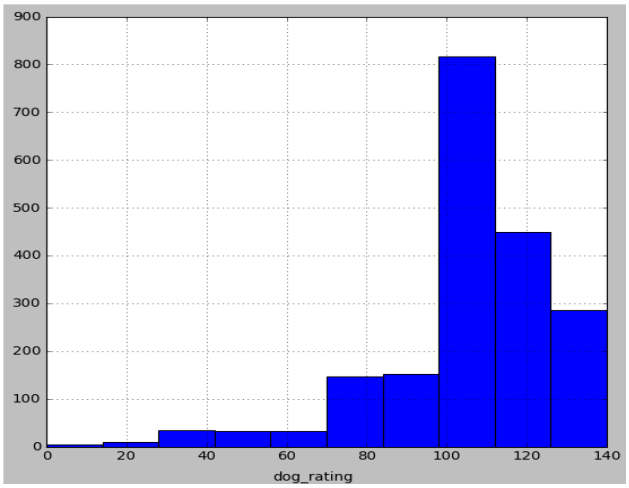
Conclusion

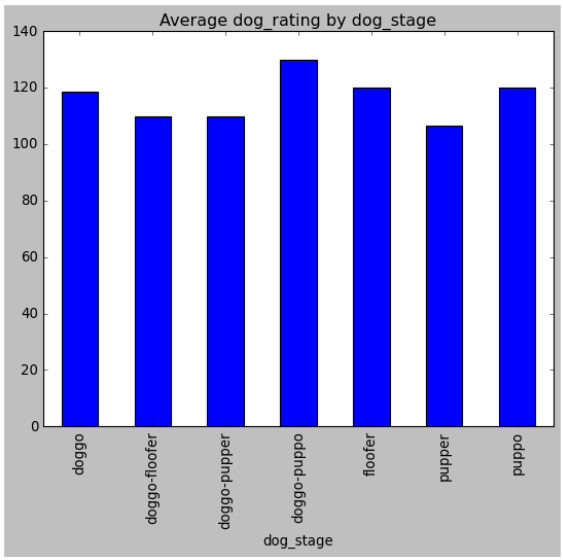
- It seems that **High prediction confidence level** usually has good **True** reality than **Low confidence**.
- That proves the efficiency of the image prediction model.

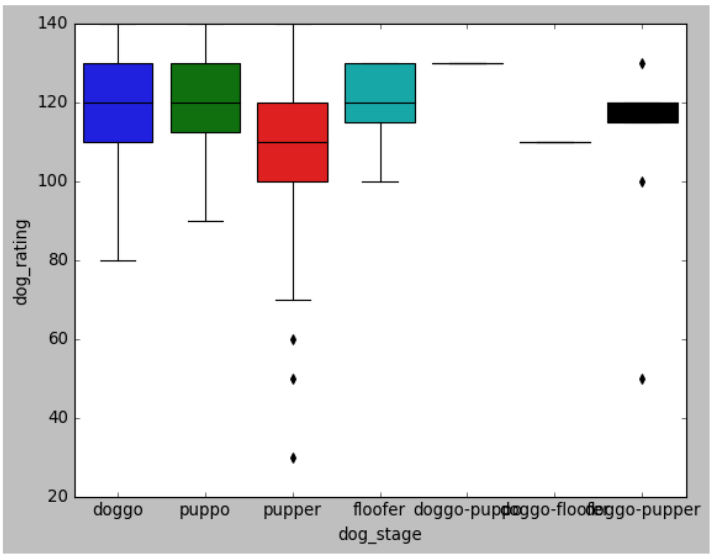
Back to Contents

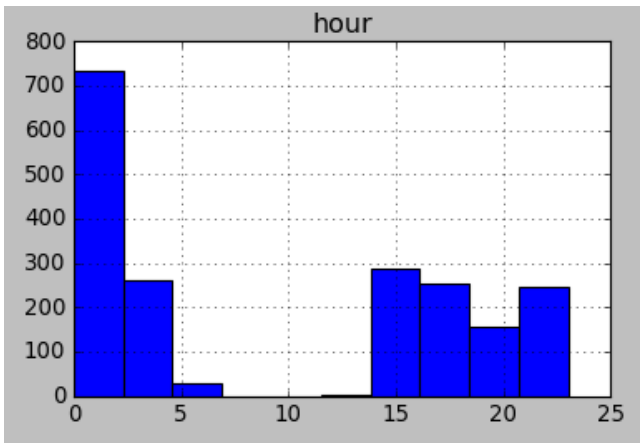
3. Additional EDA Figures











Back to Contents

--- End of Report ---