Wrangle-report

The Rate Dogs Project

Udacity DAND –

Wrangle and Analyse Data

Project Details

Your tasks in this project are as follows:

- Data wrangling, which consists of:
- Gathering data (downloadable file in the Resources tab in the left most panel of your classroom and linked in step 1 below).
- Assessing data
- Cleaning data
- Storing, analyzing, and visualizing your wrangled data
- Reporting on 1) your data wrangling efforts and 2) your data analyses and visualizations

Gathering Data

The data, the most important step is to gather the data. First to gather the data from file than, download the image predictions. Lastly from the udacity's classroom resources I used Panda library to read files and download it. But unfortunately I got rejected from twitter developer to get access to use the library tweeby. Anyway I download the file from udacity resources.

Assessing Data

After collecting the data and gathering it I used multiple function from Pandas library to explore the data to understand it better , to know what I am dealing with. Then I started to clean and wrangle it so I can do an explanatory analysis on the data , I found couple of issues in the files

Quality

- **Twitter archive data:**

1. Wrong names in name table
2. The tweet-id is not str
3. Deleting unnecessary tables as example
4. The table “Timetemp” wasn’t datetype
5. In “Rating_Denomintor” table the rating is not fixed one of rating out of 10 was 180
6. I merged it two tables together Rating_numertor and Rating_Denomintor and alter their type into float
7. Merged it 4 tables as Dog_types instead of being individual tables

- **Image prediction data:**

1. Changing the Tweet-ID types into str

- **Tweet-Json**

2. Retweets table convert into int
3. Favourite table converts into int

Tidyness

So after changing the quality I needed to merge all the 3 datasets to analyse and understand the given data and visualize it.

Visualization:

After wrapping up the wrangling portion which is include gathering, surveying, and cleaning the given data, here we need to reply a few questions by utilizing visualizing. We utilized a few qualities after we had gathered, evaluated, and cleaned our data, such as, retweets rate, top choices rates, puppy categories and puppy names.

I used pandas libraries to figure out from dogs_type which of the dogs are the favourite with the most high rating among all the group. Another visual is to understand the rate of the dogs by dividing the rating_denominator / rating_numerator

Limitation that I found

The most hard thing is was to get twitter developer account approval. Also, I had very hard time understand the different types of dogs and their rating cause I am not a pet person so I had to understand it in order to do a really perfect analysis. It was fun.

Conclusion

The Twitter account of (@dogs_rates) is devoted to laughable reviewing pictures of dogs and their adorable photos

the rating on scale of one to ten . but more given rating in excess of the maximum 16/10. it has acquired over million of followers

if I was thinking of adopting a dog I will definitely adopt a puppo. I hope this analysis makes understand what kind of pet you want to adopt