act_report

Introduction

This is a data wrangling and analyzing project aiming to wrangle and analyze data from three separated datasets : twitter-archive-enhanced dataSet that contains 2356 row and 17attributes (tweet_id , in_reply_to_status_id , in_reply_to_user_id , timestamp , source , text , retweeted_status_id , retweeted_status_user_id , retweeted_status_timestamp , expanded_urls , rating_numerator , rating_denominator , name , doggo , floofer , pupper and puppo) , image-predictions dataSet thst contain 2074 row and 12 column (tweet_id , jpg_url , img_num , p1 , p1_conf , p1_dog , p2 , p2_conf , p2_dog , p3 , p3_conf and p3_dog) and tweet_json dataSet that is supposed to be by Twitter's API which contains 2354 row and 31 column .

Insights

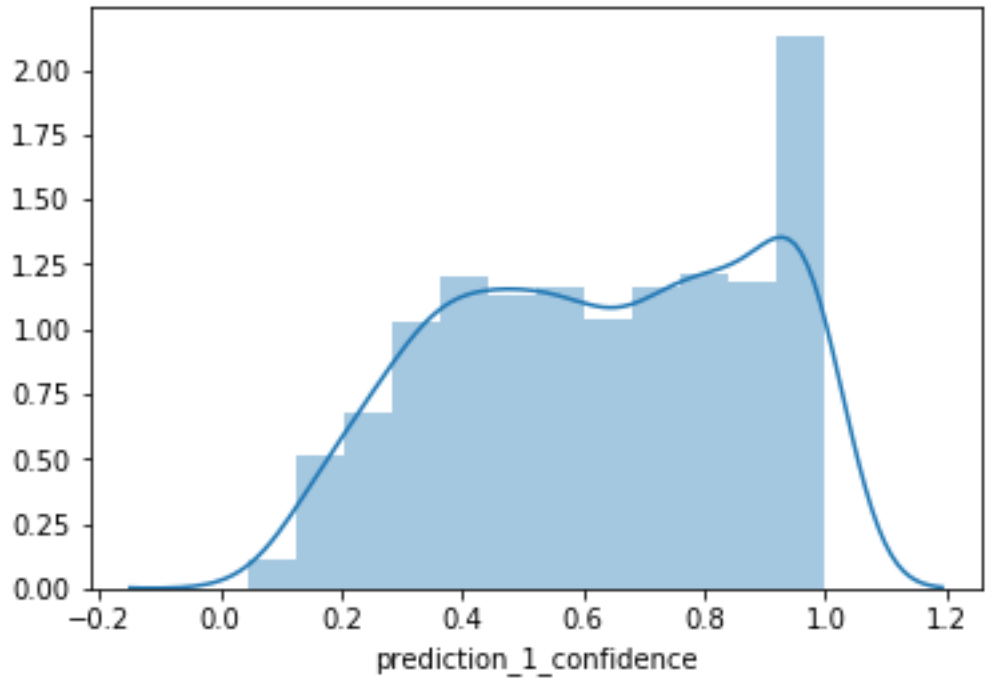
After the gathering, assessing, and cleaning processes, the datasets were analysed. This data analysis aimed to produce three (3) insights and more than one (1) visualization :

1. Calculate the mean and median for the predictions 'prediction_1_confidence', 'prediction_2_confidence' and 'prediction_3_confidence' .
2. Describe the data records of 'prediction_1_confidence', 'prediction_2_confidence' and 'prediction_3_confidence' distributed .
3. Extract the names of the dogs with the highest prediction rates of 'prediction_1_confidence', 'prediction_2_confidence' and 'prediction_3_confidence' with the doge image , retweet_count , favorite_count , text and prediction_1(breed of dog) .

PREDICTION_1

Based on the column prediction_1_confidence it could be clarified that the mean of prediction_1_confidence is 0.627221 and the median is 0.631501 . This The distribution of this data (prediction_1_confidence) is Negative/left skew which means that the data is mostly high values, long left tail , the dog name is not availble (NaN) , 1552 retweet-count , 3748 favorite_count , text " Happy Wednesday here's a bucket of pups. 44/40 would pet all at once "

, breed of dog is Labrador_retriever and with distribution of this prediction_1_confidence data the dog image as shown below .

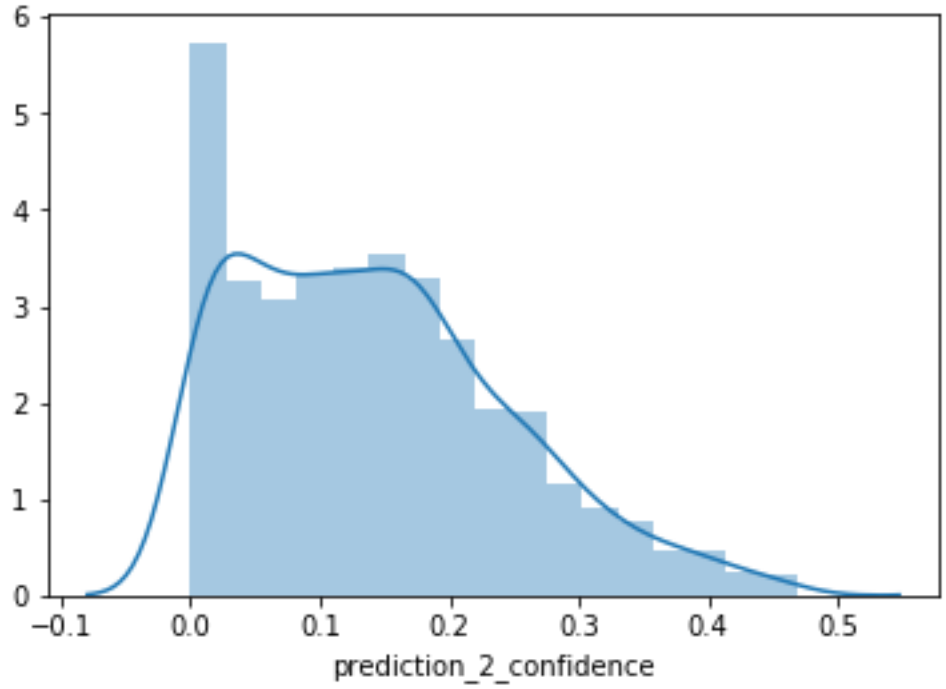



N-2300
@dog_rates

PREDICTION_2

Based on the column prediction_2_confidence it could be clarified that the mean of prediction_2_confidence is 0.143309 and the median is 0.130726. This The distribution of this data (prediction_2_confidence) is Positive/right skew which means that the data is mostly low values, long right tail , the dog name is Pablo , 1483 retweet-count , 3748 favorite_count , text " Say hello to Pablo. He's one gorgeous puppo. A true 12/10. Click the link to see why Pablo requests your assistance "

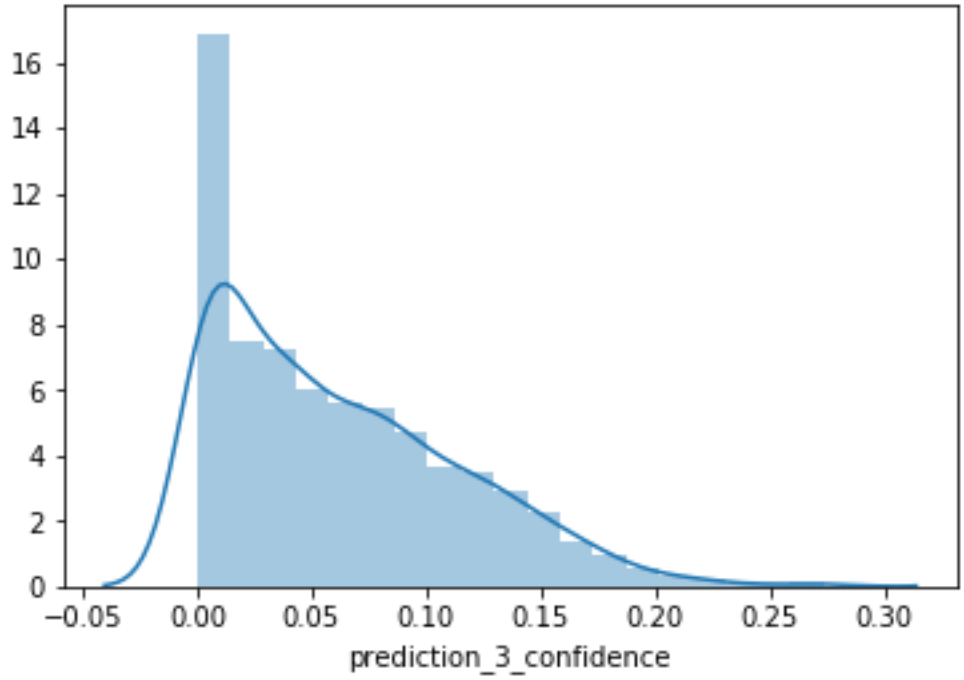
, breed of dog is Siberian_husky and with distribution of this prediction_2_confidence data the dog image as shown below .



PREDICTION_3

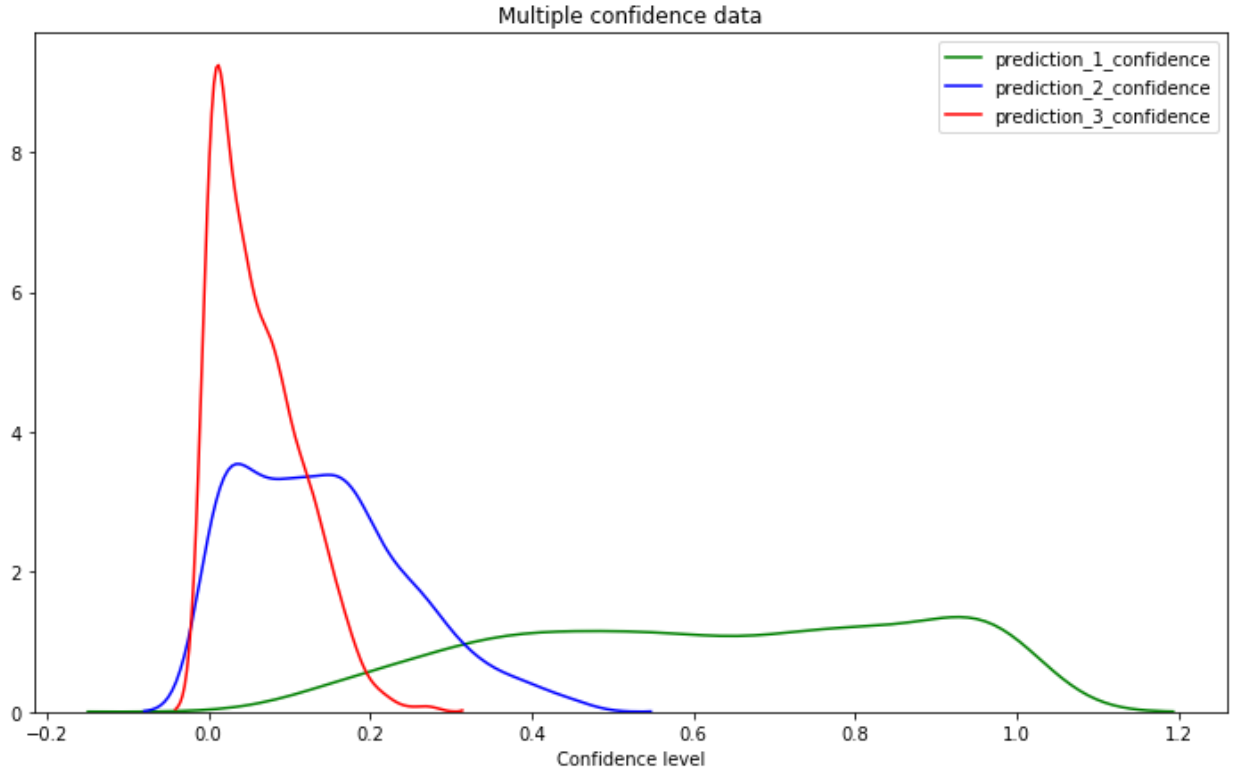
Based on the column prediction_3_confidence it could be clarified that the mean of prediction_3_confidence is 0.061627 and the median is 0.0486655. This The distribution of this data (prediction_3_confidence) is too Positive/right skew which means that the data is mostly low values, long right tail , the dog name is Bluebert, 247 retweet-count , 2574 favorite_count , text " This is Bluebert. He just saw that both #FinalFur match ups are split 50/50. Amazed af. 11/10 "

, breed of dog is Eskimo_dog and with distribution of this prediction_3_confidence data the dog image as shown below .

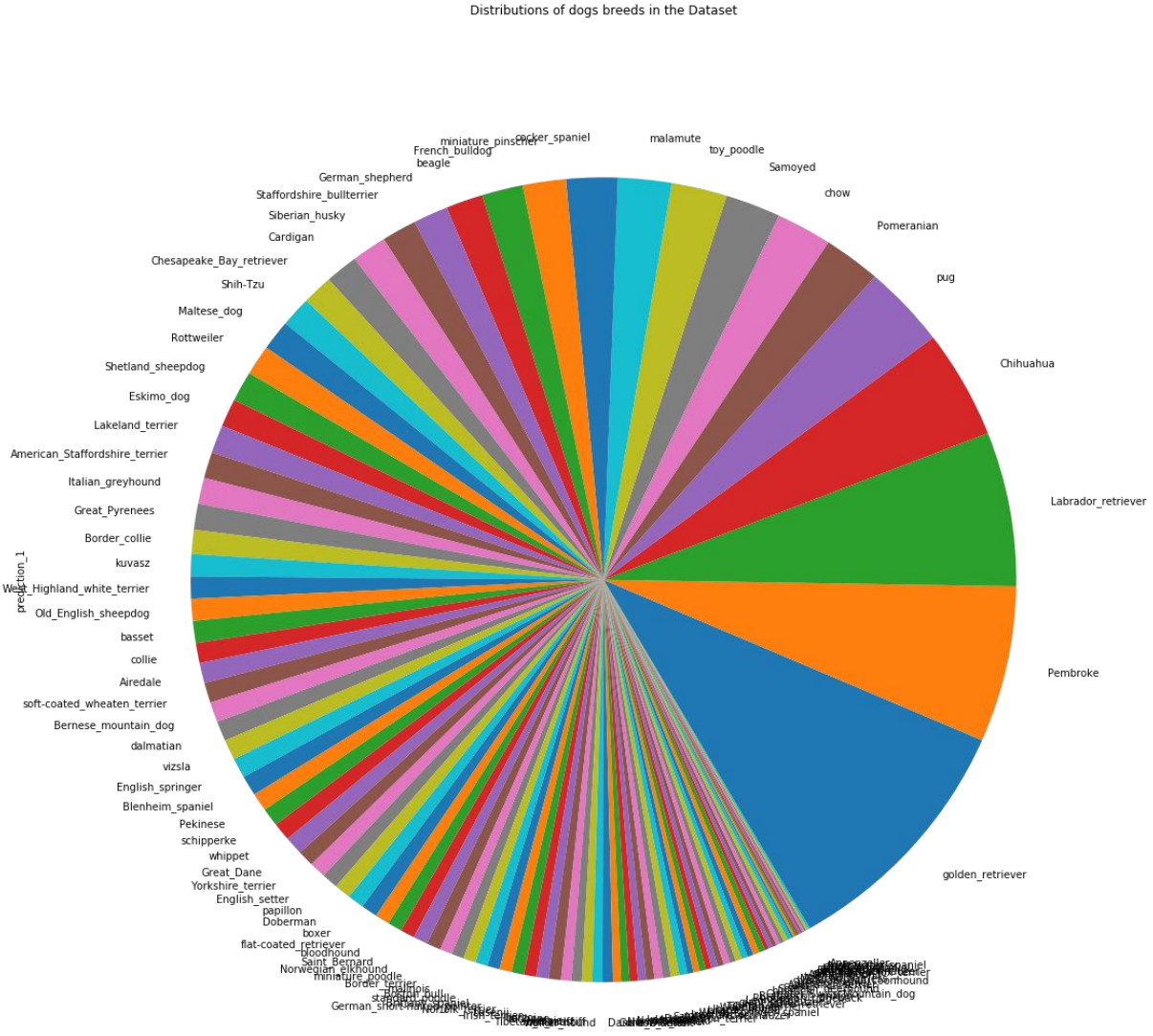



@dog_rates

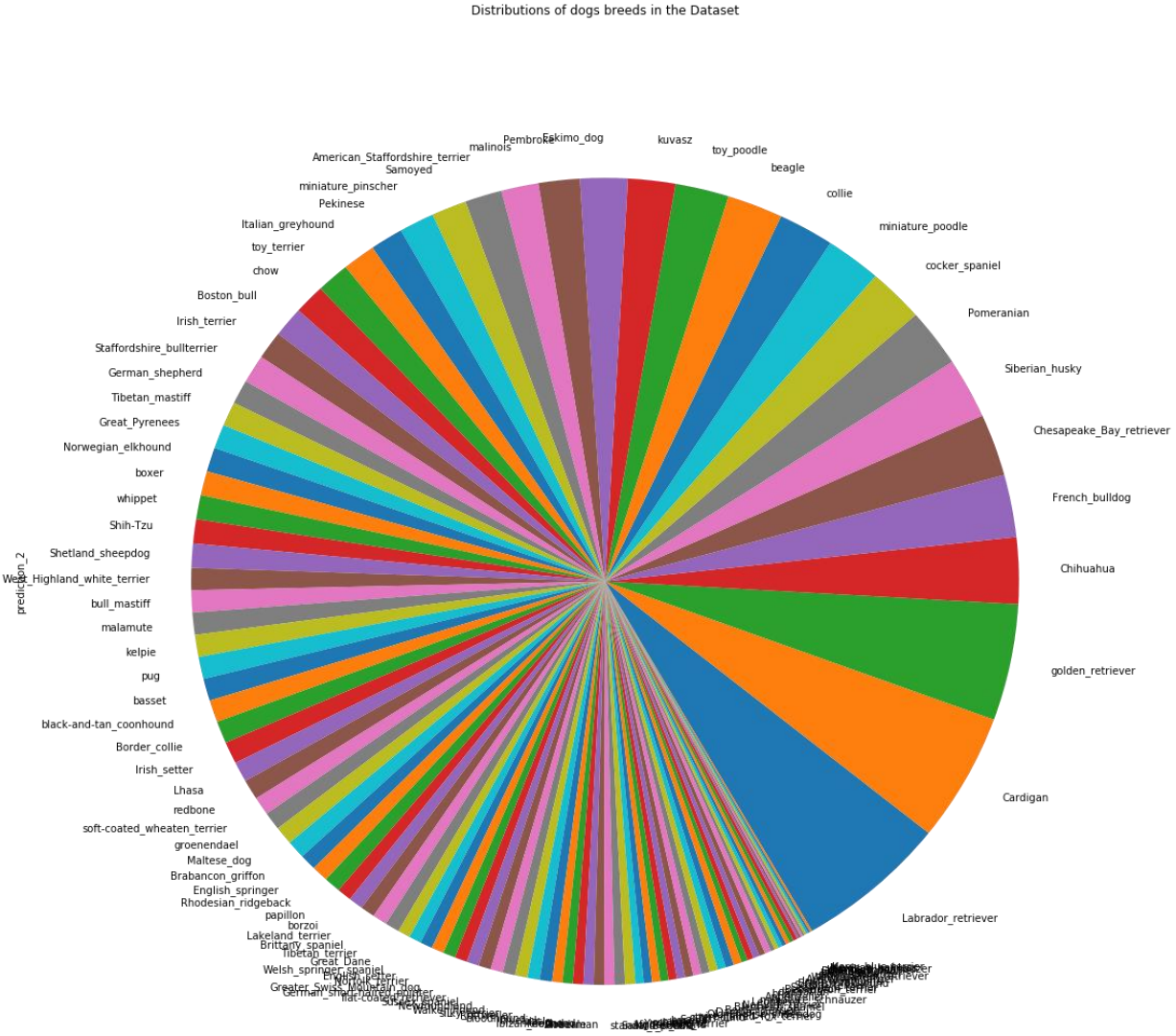
VISUALIZING THE THREE CONFIDENCES OF DATASET



VISUALIZING THE DOGS BREEDS 1 OF DATASET



VISUALIZING THE DOGS BREEDS 2 OF DATASET



VISUALIZING THE DOGS BREEDS 3 OF DATASET