Untitled1

December 8, 2020

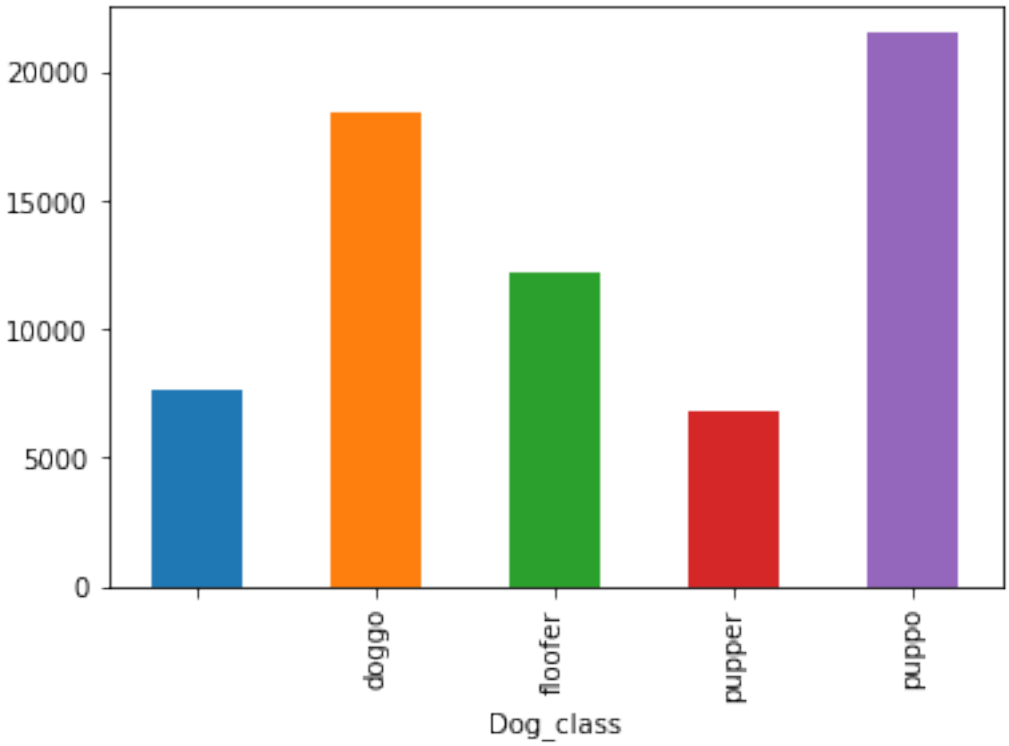
1 Introduction

This report is a resume of the Wrangle and Analyze Data Project, and aims to shows the insights observed in the wrangle_act.ipynb file.

2 Exploratory Data Analysis

3 Research Question 1: Which type of dogs got the highest favorite counts?

```
In [2]: import pandas as pd
        import seaborn as sns
        import os
        import numpy as np
        import matplotlib.pyplot as plt
        %matplotlib inline
        import requests
        import tweepy
        import json
        from tweepy import OAuthHandler
        from timeit import default_timer as timer
        from pandas.api.types import CategoricalDtype

In [6]: twitter_archive_master = pd.read_csv('twitter_archive_master.csv')

In [7]: twitter_archive_master.groupby('Dog_class')['favorite_count'].mean().plot(kind = 'bar')

Out[7]: <matplotlib.axes._subplots.AxesSubplot at 0x7fb8650c7350>
```



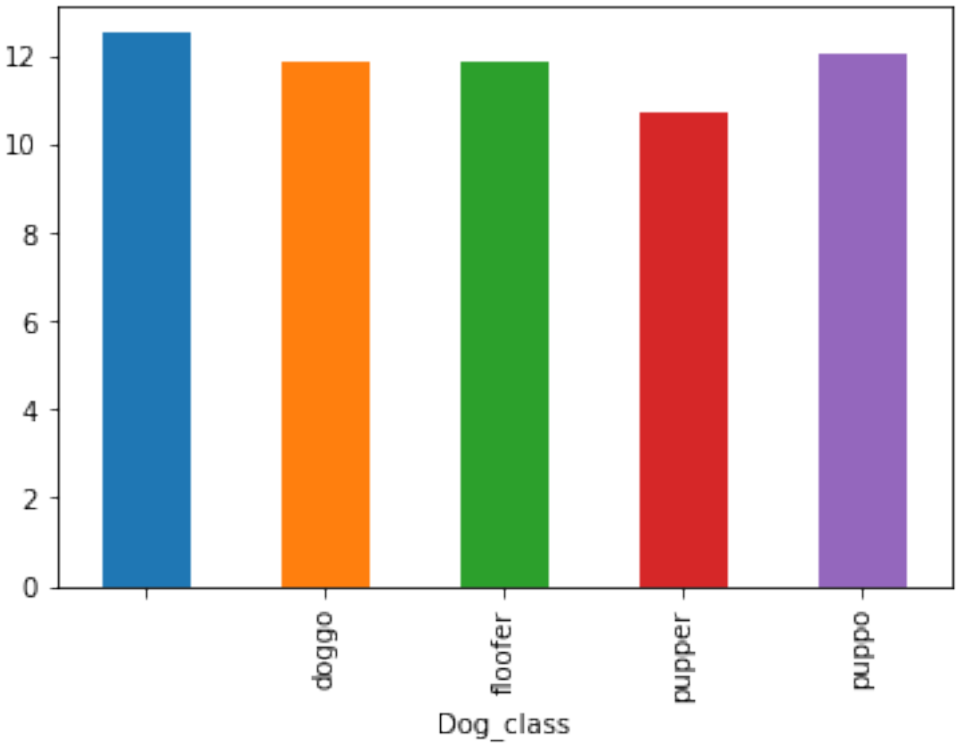
4 the puppo classification has the highest favorite counts

In [ ]: ---

5 Research Question 1: Which type of dogs got the low rating_numerator?ű

In [8]:
```
twitter_archive_master.groupby(['Dog_class'])['rating_numerator'].mean().plot(kind = "ba
```

Out[8]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fb8650e8d50>
```



6 the pupper classification has lowst rating

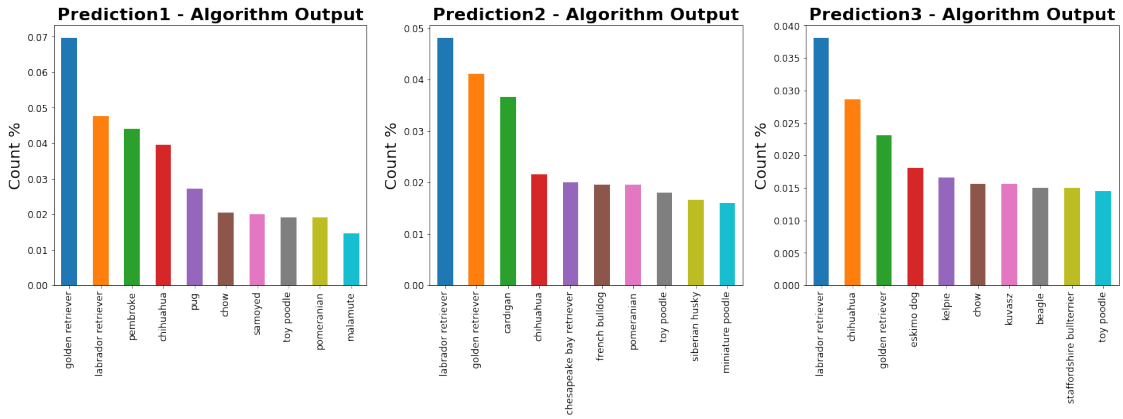
7 ----------

In [ ]: *# which dog type predictions has highst value*

In [12]:
```
# plt.figure(1)
plt.subplot(131)

twitter_archive_master['Dog_Type1'].value_counts(normalize=True).head(10).plot.bar(figs
plt.title('Prediction1 - Algorithm Output', fontweight="bold", fontsize = 22.0)
plt.ylabel('Count %', fontsize = 20.0)

plt.subplot(132)
twitter_archive_master['Dog_Type2'].value_counts(normalize=True).head(10).plot.bar(figs
plt.title('Prediction2 - Algorithm Output', fontweight="bold", fontsize = 22.0)
plt.ylabel('Count %', fontsize = 20.0)

plt.subplot(133)
twitter_archive_master['Dog_Type3'].value_counts(normalize=True).head(10).plot.bar(figs
```

```
plt.title('Prediction3 - Algorithm Output', fontweight="bold", fontsize = 22.0)
plt.ylabel('Count %', fontsize = 20.0)
```

Out[12]: Text(0,0.5,u'Count %')

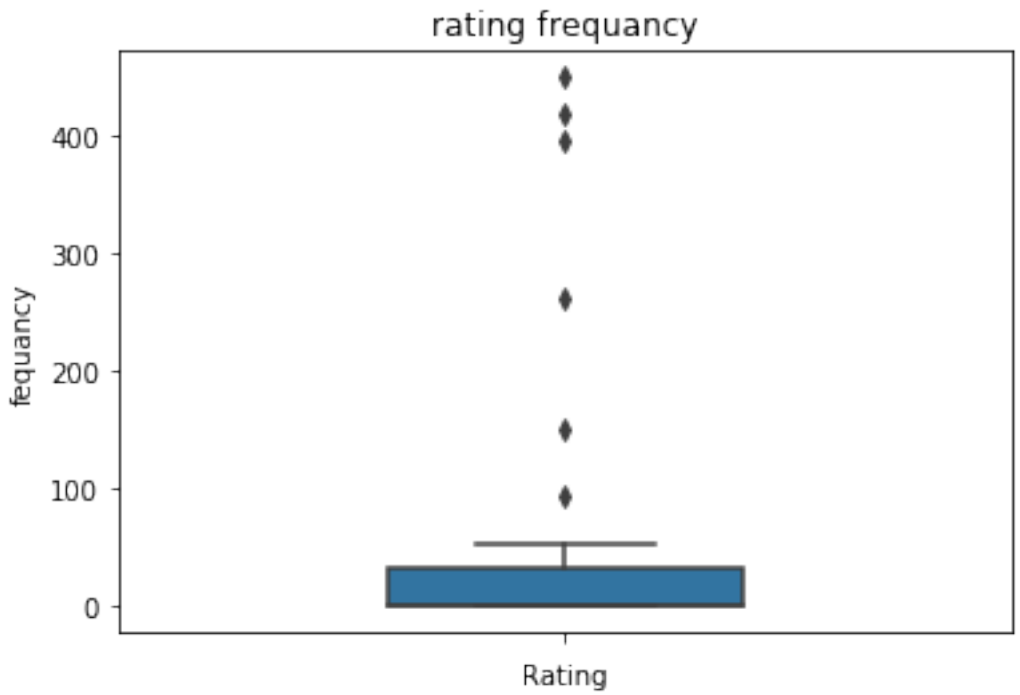


8 the golden retriver has the highst value in the three predictions

In [ ]: ----------

9 lets look for the rating distribution

In [10]:
```
data = twitter_archive_master.rating_numerator.value_counts()
ax = sns.boxplot(data, orient='v', width = .4)
ax.set(xlabel = 'Rating' ,ylabel = 'fequancy',title = 'rating frequancy')
plt.show()
```

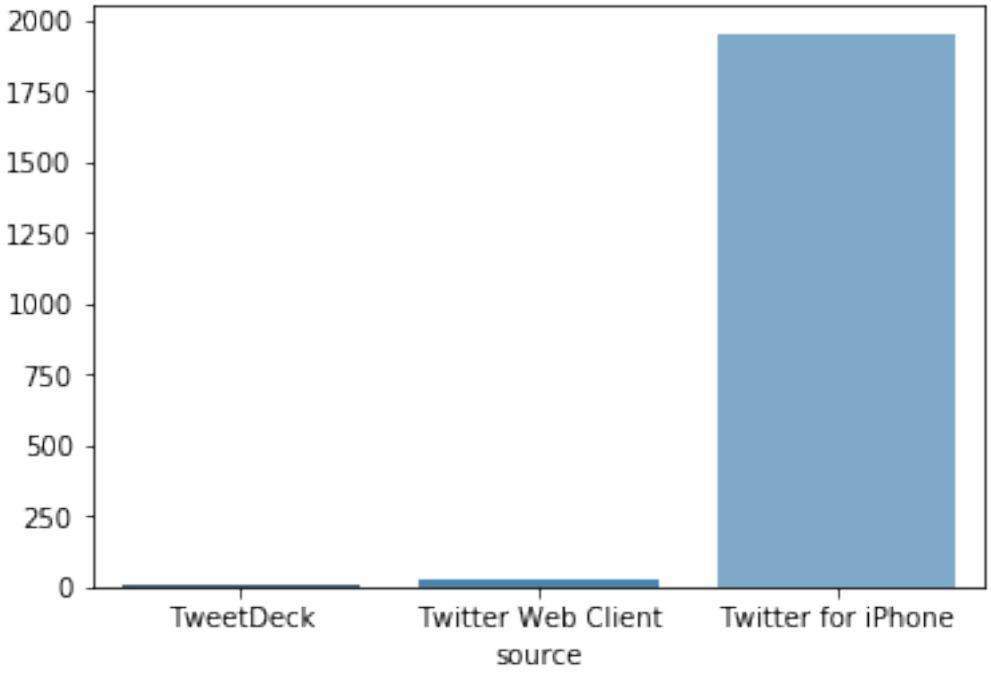


10 there are alot of outliers and the most rating from 0 to 20

11 ----------

In [ ]: *#comparing tweets different source*

```
In [11]: data = twitter_archive_master.groupby('source').count()['tweet_id']
         x = data.index
         y = data.values
         g=sns.barplot(x,y,palette="Blues_d")
         ax.set(xlabel = 'count' ,ylabel = 'tweet source',title = 'tweet source counts')
         plt.show()
```



12 the most source is twitter for iphone

In [ ]: