# act_report_PDF

July 3, 2020

0.0.1 Introduction

In this report, we will analyse the dataset "Master_tidy_data.csv" which contains tweets from WerateDogs Twitter account, and were published in 2017. This analysis aims at answering the following three questions: - What breed of dog is most popular - How do dogs from the most popular breed look like? - When there is no dog identified on the photo, what is there instead?
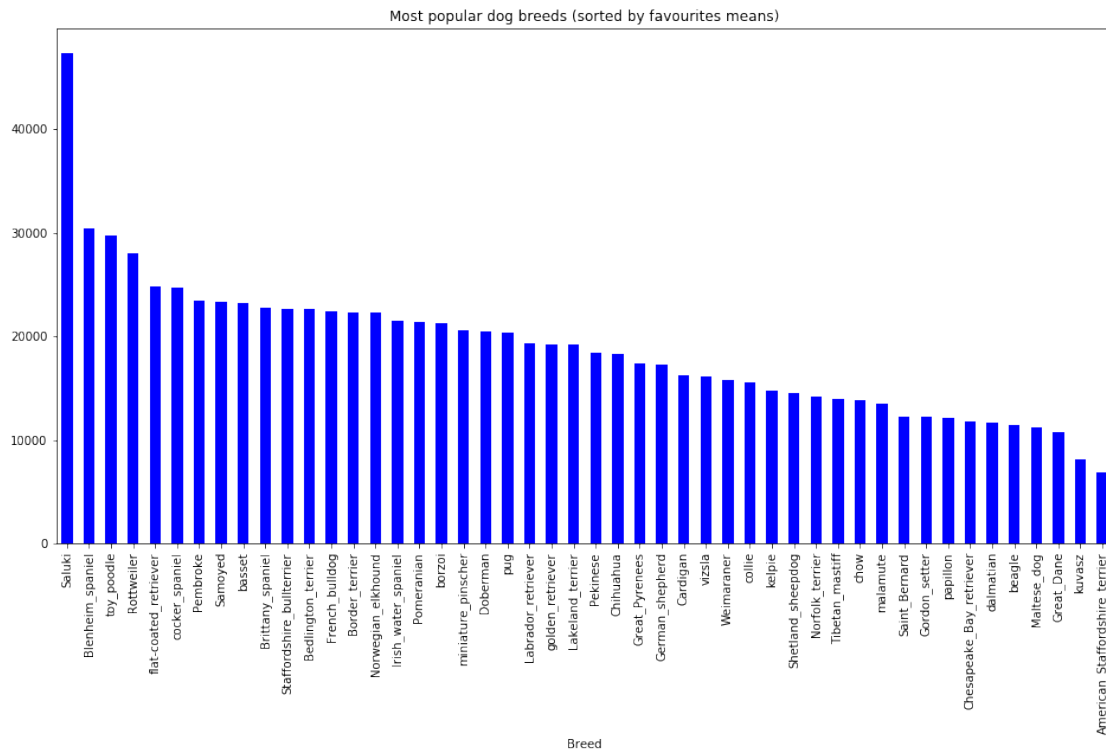
- What breed of dog is most popular?

In [34]:
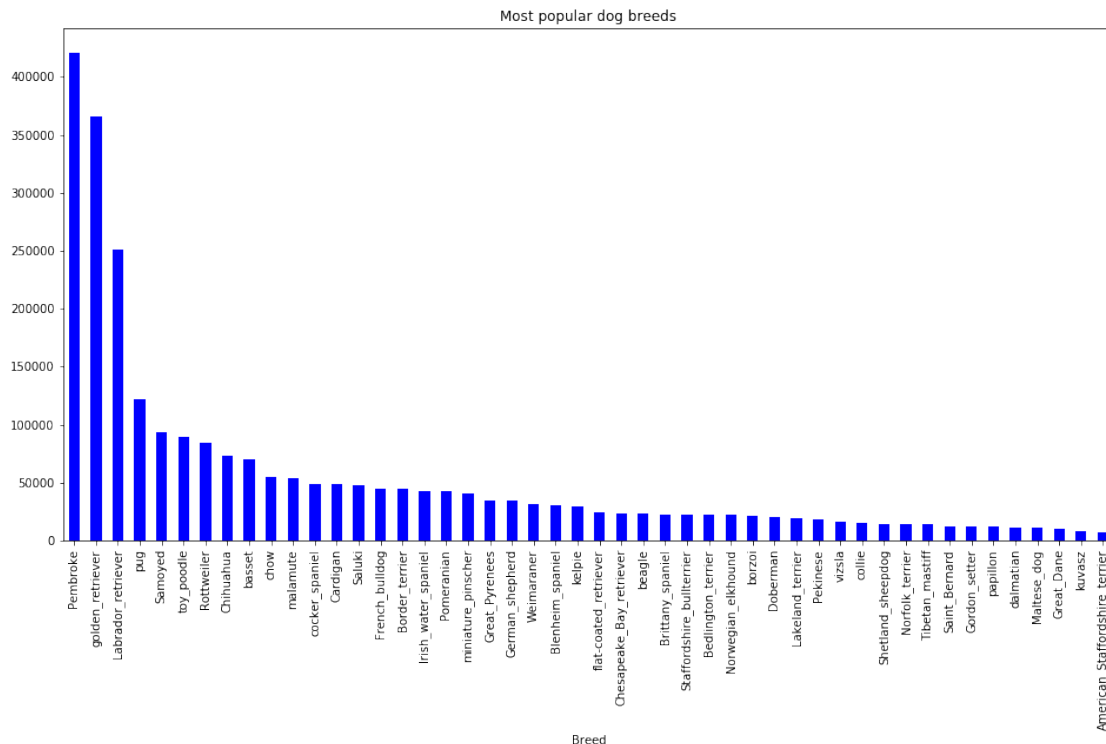
In terms of volumes of total favorites for all tweets, retrievers and Pembrokes are the most popular dogs. The neural network differentiated the different types of retrievers however we can safely say here that those dogs should compete in the same category.

In [36]:

Most popular dog breeds (sorted by favourites means)

When we sort values by tweet number of favorite means, the results are different. The most popular breeds of dogs are Rottweiler and Bleinheim Spaniel. High volumes of favourites that retrievers and Pembrokes received can be due to a higher volume of dogs in tweets. Pembrokes and Retrievers are more common breed of dogs owned by people.

- How do dogs from the most popular breed look like?

We know that the most popular breed of dog is Pembroke. To import pictures from the dataset and display them in the notebook, we will use a query to select only the most favorited pembrokes of the dataset. This will enable us to access their image urls.

In [21]:

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

We do the same for retrievers

```
In [20]:
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

Looking at the most favourited pictures, we can say that most popular dogs belong to the "pupper" category.

- When there is no dog identified on the photo, what is there instead?

With the groupby function, we check if there is any recurrence in the identified objects of the photos. From this analysis, we see that seatbelts are particularly recurrent in photos.

```
In [71]:
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

```
<IPython.core.display.Image object>
```

Dogs are present in those pictures but the neural network recognised a car environment and chose "seat belt" rather than "Labrador" for the photo. The neural network model could be improved by feeding those results back in the system and include those images reidentified as "dogs".

0.0.2 Conclusion

After several iterations of the wrangling process, we managed to build a master dataset 'Master_tidy_data' that we could investigate and draw insights from. One of the main findings of the analysis is that some breeds of dogs are more popular than other, with higher volumes of "favourites" than others. One limitation to this study is the volume of dog for each breed. To have a fair representation of popularity amongst dog breeds, we should create a sample with an even volume of dogs within each breed.

Another interesting finding from this analysis is that we can help improving the image predictions neural network by feeding back in our findings in the AI model.