* All datasets were merged into one dataset saved as **“twitter\_archive\_master.csv”** after performing data wrangling.
* The merged dataset was grouped by **“year”** to find the most common dog species and the most common dog stage for each year (e.g. 2015, 2016, 2017) by calculating the **“Mode”** value for columns **“species\_prediction”** and **“stage”** respectively.
* In result, the following statistics are shown:

1. **Most Common Dog Species per Year:**

|  |  |
| --- | --- |
| 2015 | Chihuahua |
| 2016 | Golden Retriever |
| 2017 | Golden Retriever |

1. **Most Common Dog Stage per Year:**

|  |  |
| --- | --- |
| 2015 | Pupper |
| 2016 | Pupper |
| 2017 | Doggo |

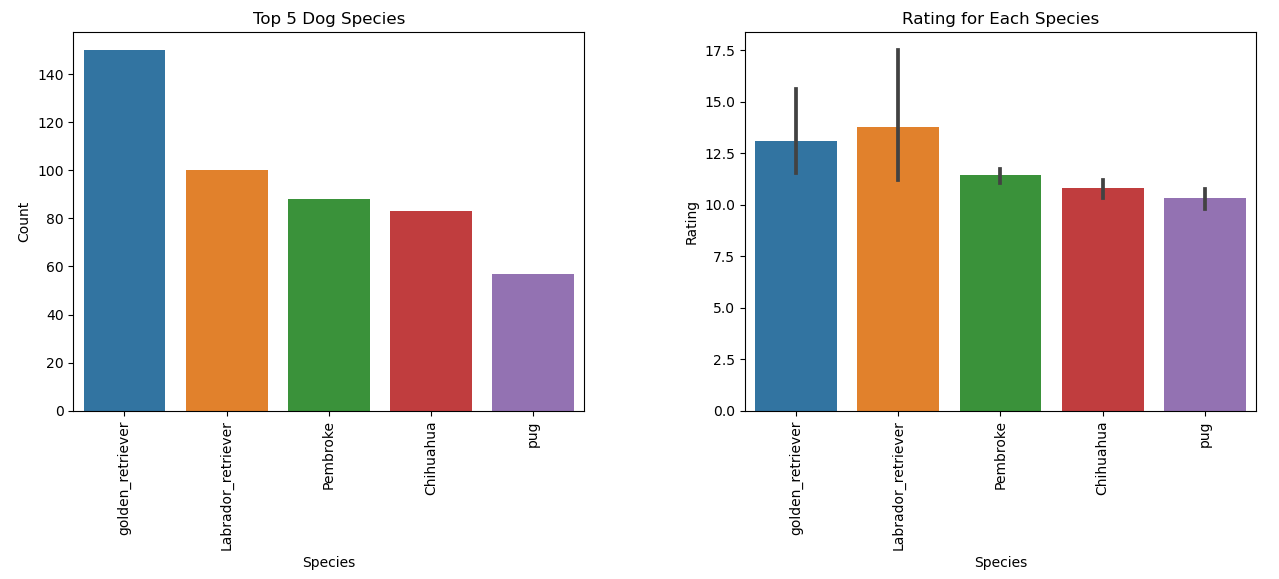
* Some statistics were plotted and found the following:

1. **Most common Dog Species in All Time:**

Using **Seaborn** library and **countplot** function, where **species\_prediction** ison X-axis and **Count** is on Y-axis. It is obvious that **Golden Retriever** is the winner with value **above 140**.

1. **Rating for Each Species:**

Using **Seaborn** library and **barplot** function, where **species\_prediction** is on X-axis and **rating** is on Y-axis. It is shown that **Labrador Retriever** got the **highest rating score** **14**.

****

1. **Number of Retweets for Each Year:**

Using **Seaborn** library and **barplot** function, where **year** is on X-axis and **retweet\_count** is on Y-axis. It is obvious that number of retweets **increases** year after year and in 2017, the retweets count became above 6000 which means the interest in dogs rating is getting more popular.

1. **Favorite Count for Each Stage:**

Using **Seaborn** library and **barplot** function, where **stage** is on X-axis and **favorite\_count** is on Y-axis. It is shown that **Puppo** is most favorite dog stage among the rest and **Pupper** is the least favorite.

