Udacity Data Analyst Nanodegree Program

wrangle and analyze data

```
In [78]: df.groupby('stage').describe().iloc[:,0:8]
```

Out[78]:

	Favorite_count							
	count	mean	std	min	25%	50%	75%	max
stage								
doggo	63.0	19356.380952	22208.483825	2593.0	8268.00	12376.0	20502.50	131075.0
floofer	8.0	13701.375000	10516.926166	2262.0	5560.75	11879.0	18623.25	33345.0
pupper	212.0	7424.240566	10793.754867	693.0	2449.75	3393.5	8139.50	106827.0
puppo	23.0	22723.913043	27931.824108	3277.0	7067.00	15359.0	21977.50	132810.0

```
In [79]: df.groupby('stage').describe().iloc[:,8:16]
```

Out[79]:

	Retweet_count							
	count	mean	std	min	25%	50%	75%	max
stage								
doggo	63.0	7125.698413	12868.874516	725.0	2025.50	3327.0	5422.00	79515.0
floofer	8.0	4776.750000	5732.138787	496.0	2167.00	3349.0	4130.75	18497.0
pupper	212.0	2436.594340	3650.522156	103.0	695.75	1256.0	2513.00	32883.0
puppo	23.0	7027.086957	10408.775240	716.0	1721.00	3220.0	7541.50	48265.0

```
In [80]: df.groupby('stage').describe().iloc[:,24:32]
```

Out[80]:

	rating							
	count	mean	std	min	25%	50%	75%	max
stage								
doggo	63.0	11.888889	1.471351	8.0	11.0	12.0	13.0	14.0
floofer	8.0	11.875000	1.125992	10.0	11.0	12.0	13.0	13.0
pupper	212.0	10.726415	2.083713	3.0	10.0	11.0	12.0	27.0
puppo	23.0	12.043478	1.296087	9.0	11.5	12.0	13.0	14.0

```
In [81]: df.prediction.value_counts()
```

Out[81]:
```
No prediction                308
golden_retriever             161
```

insights:

1. the most favorites and retweet came from the iPhone app.
2. the most popular stage is pupper.
3. the most popular breed according to machine learning model is the golden retriever.

visualization: