|  |
| --- |
|  |
| Importing statsmodels as library in python, calculated OLS model for the data.  No stage is used as baseline for dummy variables  By holding other stage as 0 doggo, floofer, pupper, puppo stage has positive effect on rating.  However, puppo and floofer has too high p-value to pass the 0.1 confidence level.  Img\_num and p1\_conf has positive effect on rating.    Retweet\_count has negative effect and favorite\_count has positive effect on rating. However, these two have strong multicollinearity problem. |

|  |
| --- |
|  |
| retweet\_count and favorite\_count variables were dropped to eliminate multicollinearity problems.  Using no stage as a baseline for stage dummy variables, doggo, floofer, pupper and puppo has positive effects on rating. However, pupper stage had too high p-value to pass 0.1 confidence level.  Img\_num and p1\_conf also has positive effect on rating. |

|  |
| --- |
|  |
| By using backward elimination on the model, pupper stage were removed.  Floofer stage had highest positive effect on rating than other stage puppo were second and doggo were last.  As image number increase there were higher rating and also best prediction confidence.  Due to high multicollinearity and p-value, retweet\_count, favorite\_count and pupper variables were removed. |

|  |
| --- |
|  |
| Correlation matrix were plotted on the project.  As discussed in OLS model, there is high multicollinearity problems on retweet\_count and favorite\_count varibles. |

|  |  |
| --- | --- |
|  |  |
|  |  |
| Floofer stage had highest positive effect on rating on our OLS model. However, floofer stage had only 8 rows. It is not accurate to predict with such a small variance.  Puppo stage were second highest rating on the OLS model. However, puppo stage also had low observations 23 rows.  Doggo stage recorded lowest rating on the OLS model. Doggo stage had 74 observation which is still low.  Pupper stage had highest observation 211 row. However, it was dropped out of our OLS model because of high p-value.  In conclusion, it is not accurate to predict rating with stage correlation.  There has been bias variance trade of in stage. | |