**Hypothesis: Reducing number of trips reduced the subsequent COVID-19 Case Count – Regression**

**Steps taken for this analysis:**

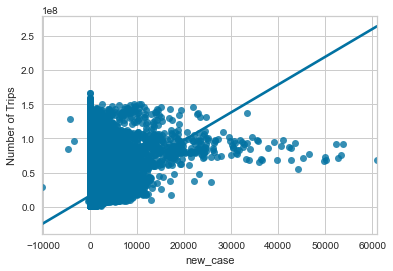
**1) after\_covid data is used for this analysis**

**2) Heatmap is created to show that highly corelated data should not be paired together. Result saved in output\_data/Corr\_heatmap.png**

**Chart, histogram

Description automatically generated**

**3) Regplot to show that the data is Linear.The straight line shows that the data is linear. Result saved in output\_data/Plot\_Linear\_regression.png**



**4) Linear Regression is run on "New COVID Cases" and "Number of Trips". Train data is 80% and Validation (Test) data is 20%**

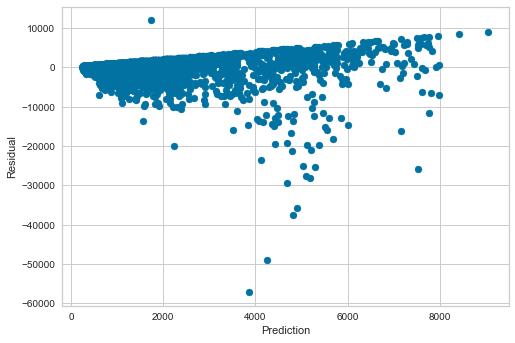
**The train data R square is 0.217 and validation data R square is 0.211. The coefficient result is 0.0005 which shows that the data is not corelated. Less value of R square means that the model is not very accurate**

**5) Residual plot to prove the Linear Regression is created. The result is saved as is saved as output\_data/Residual\_Plot.png**

**The predictor value(X-axis) is "Number of Trips" and " residual value(Y-axis) is "New Cases".**

**Residual = Observed – Predicted**

**The negative value on Y-axis means predictor is too low**

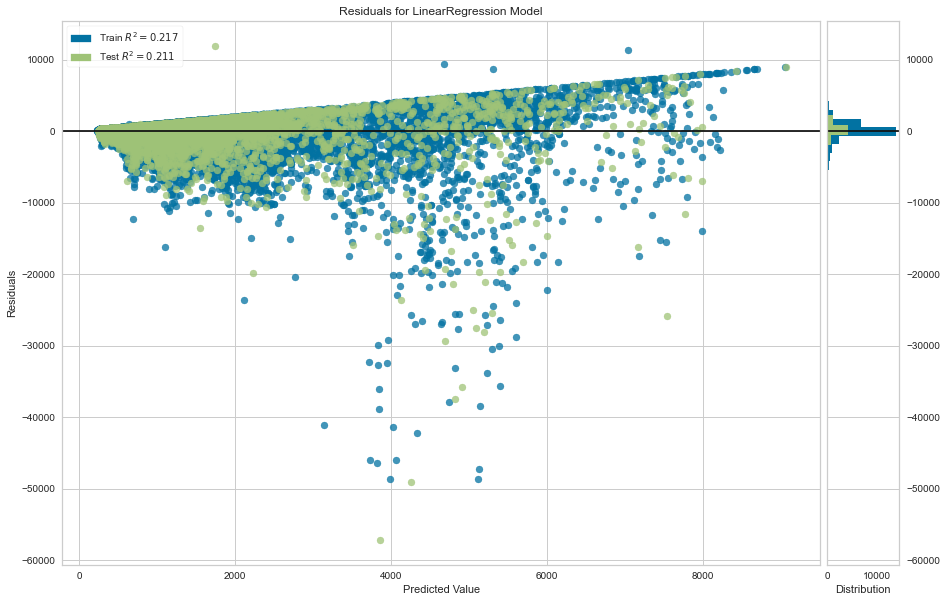


**6) Fancy residual plot using Visualizer is created to prove the Linear Regression. The result is saved as is saved as output\_data/visualizer\_residual\_plot.png**

**The predictor value(X-axis) is "Number of Trips" and " residual value(Y-axis) is "New Cases".**

**Residual = Observed – Predicted**

**The negative value on Y-axis means predictor is too low**



**7) QQ-Plot and Histogram is created to show that the results are left-skewed. The results is saved as output\_data/QQ-Plot.png**

**As the data is left skewed (negative skew), it proves that the data is not normally distributed**

