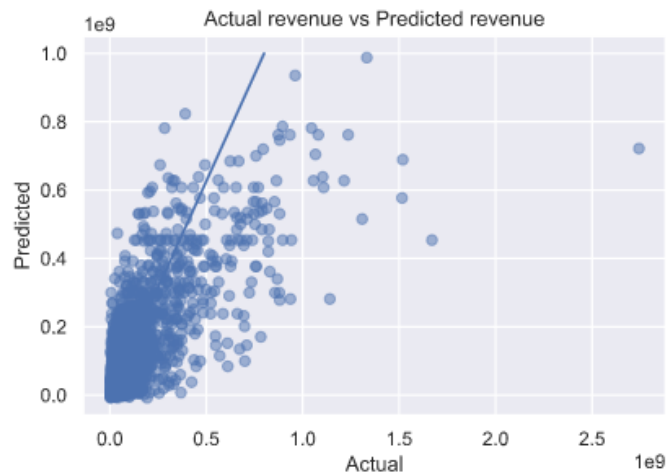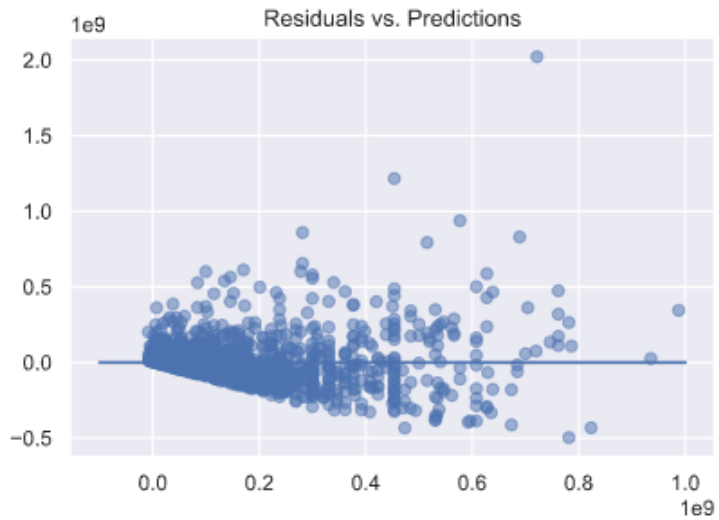## Baseline Model

- Target = revenue
- Features  = Budget & distributor
- Model = Linear regression model
- Cross validation done by 5 Kfold cv
- Avg. R2 = 0.52

## Visual analysis



Actual revenue vs Predicted revenue



Actual vs Predicted on test_data_set

Residuals vs. Predictions

- The model seems to perform on the test_data as it does on the train _data; but  a few data points are scattered away from.
- Residuals doesn't seem to have a normal distribution
- Needs a look and improvement on the distribution of the target variable and more tuning on the features.