Train and test data split:

X\_train, X\_test, y\_train, y\_test=train\_test\_split(X, y,

test\_size = 0.25,

random\_state = 246)

K-means results:

Prediction for XYZ:

Revenue of 116M with 2.56 M subscribers

Silhouette Score was 86% which is high, saying that this is a good model which shows clustering around the means

Linear regression model prediction:

RMSE is about 50 i.e if we use subscribers to predict revenue we may be right 50% of the times, which is as good as taking a random guess or a coin toss.

Number of subscribers will not be a good predictor for Ad Revenue