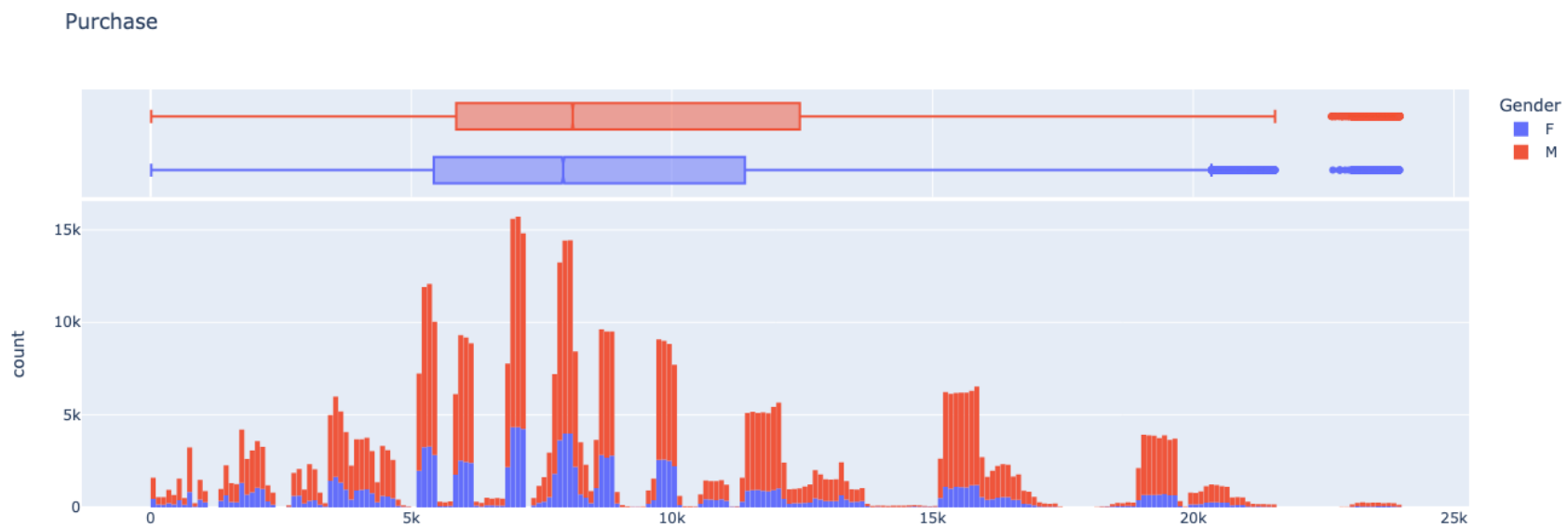
Report for Assignment-2

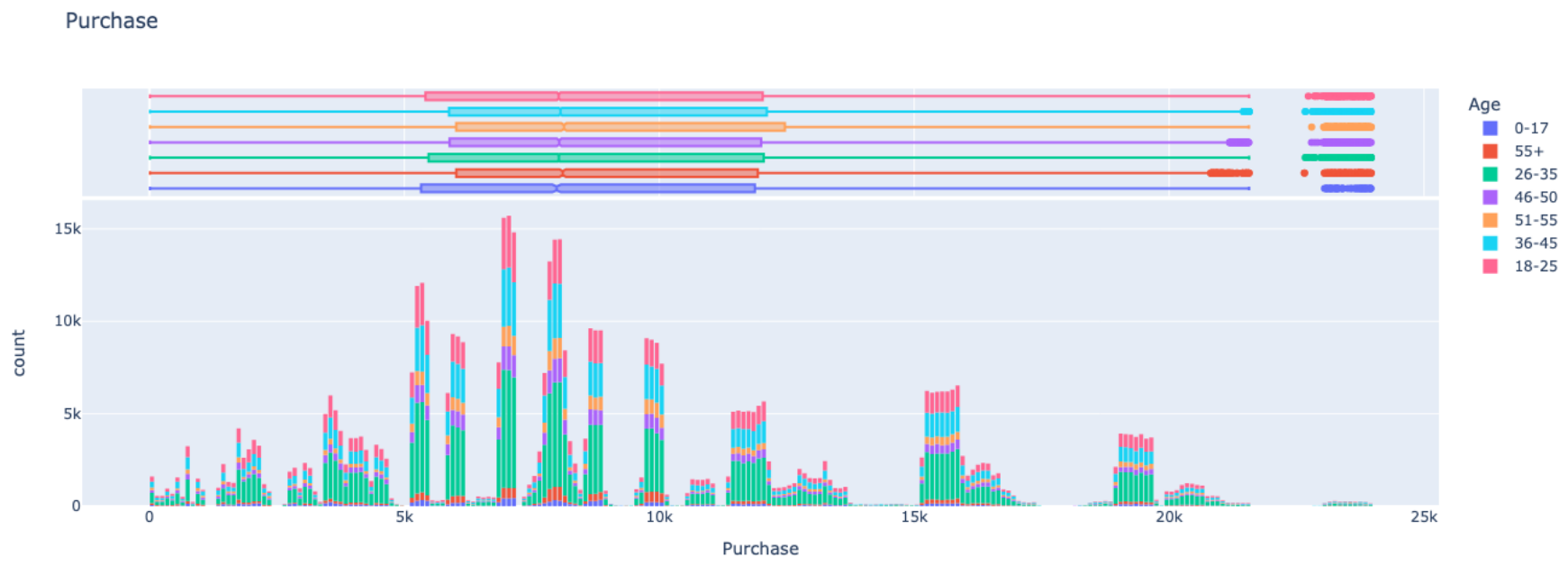
Exploratory Data Analysis:

Purchase trends:

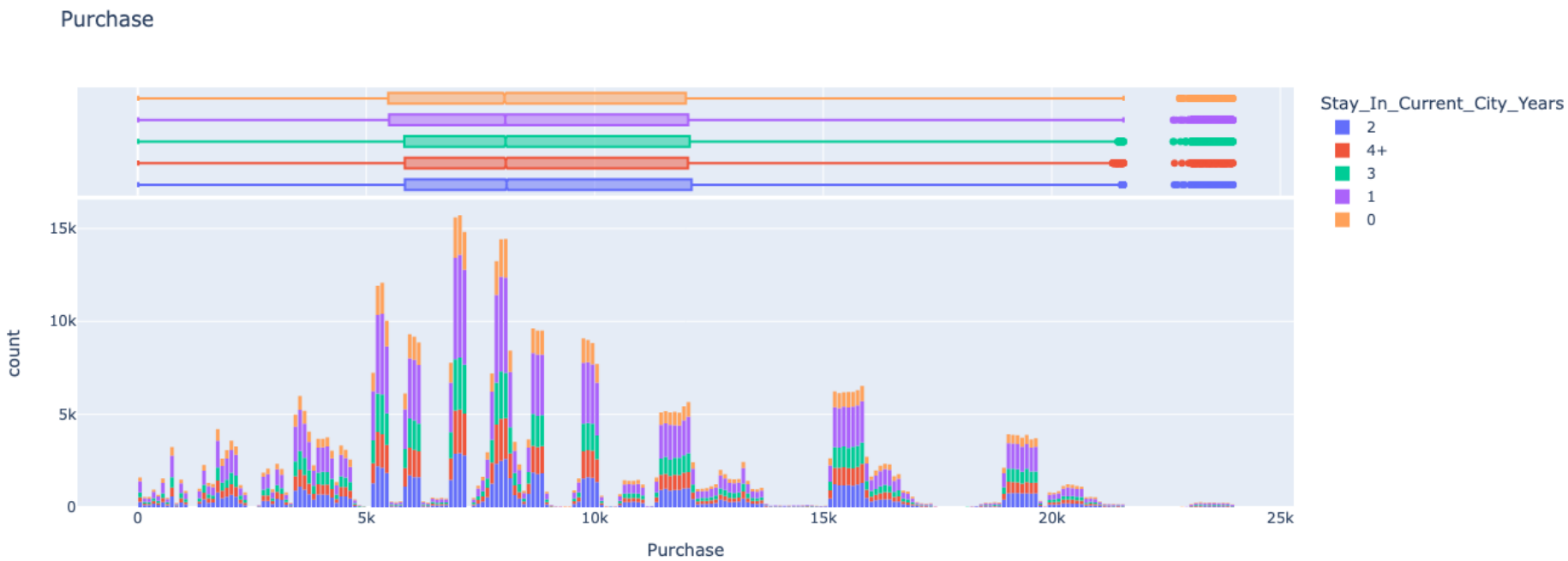
Gender:



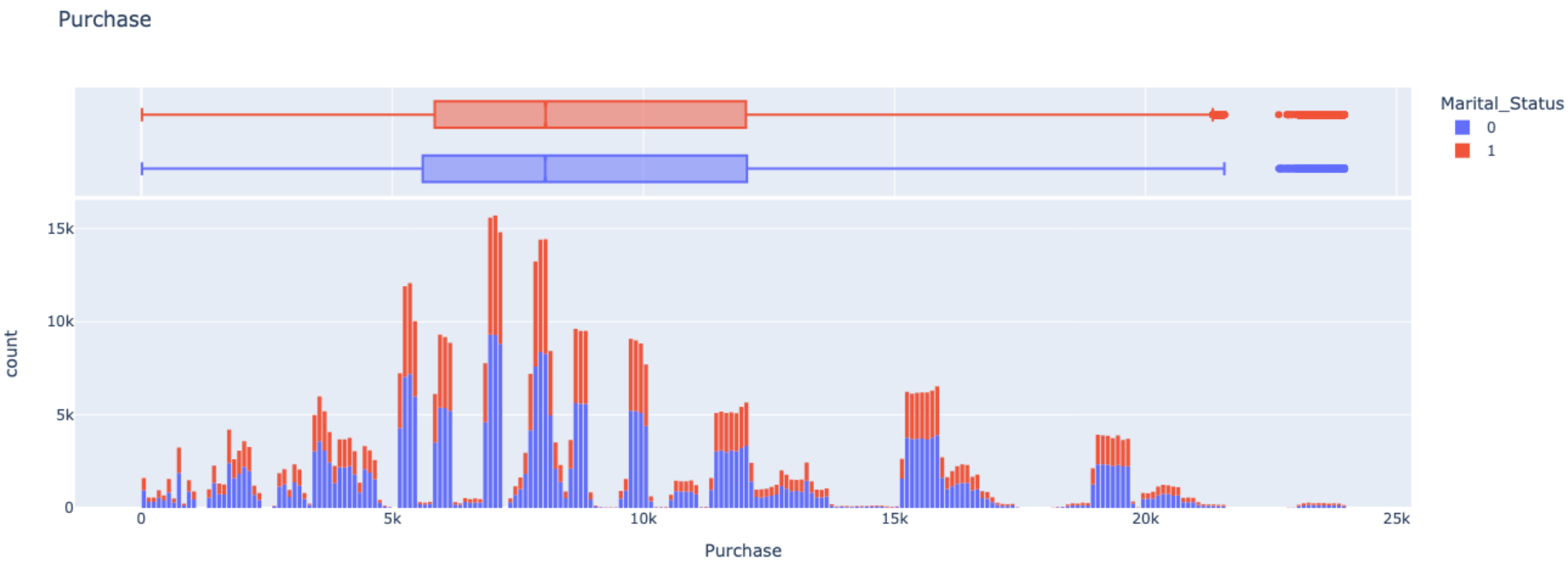
Age:



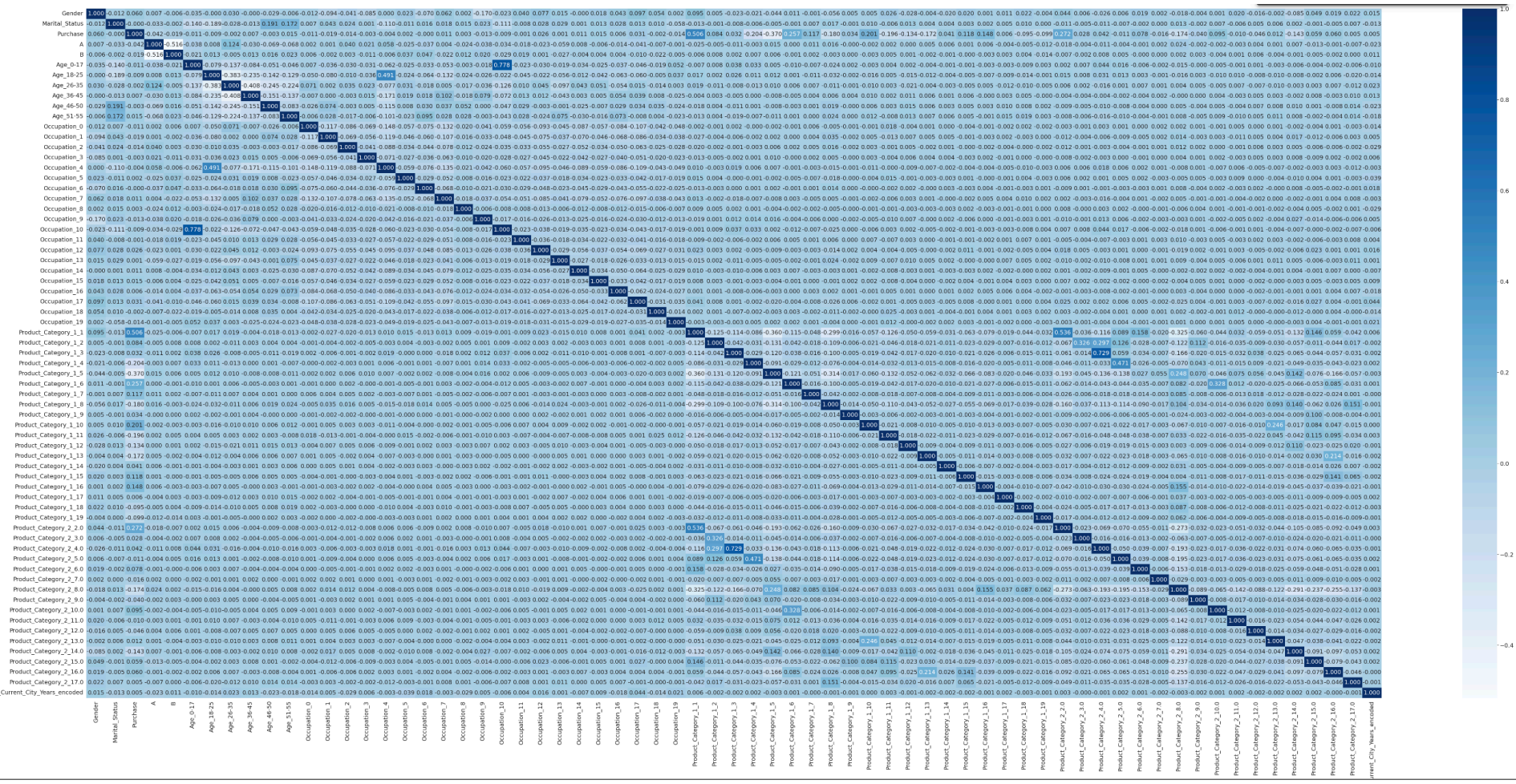
Duration of Stay in current city (years)



Marital Status



Correlation Matrix



Experiment 2:(Linear regression closed form solution approach)

MSE value without feature scaling on Test Set: 25063873.172

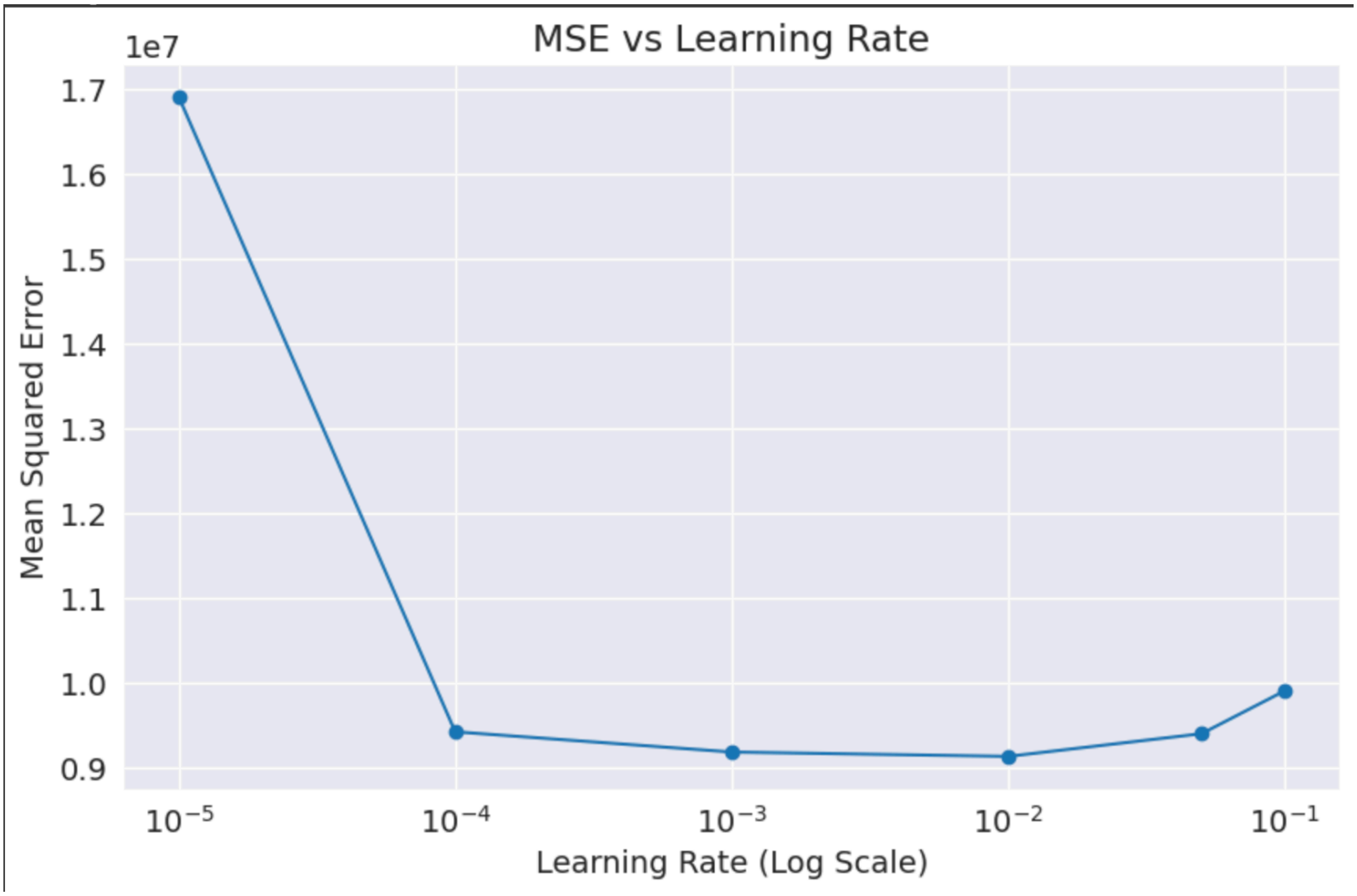
MSE value with feature scaling on Test Set: 9077460.414

Mean Absolute percentage error: 37.6%

Experiment 3:(Linear regression with gradient descent approach)

Variation of MSE values with Learning rate:

Learning Rate	MSE
0.00001	16913812.00000
0.0001	9419581.00000
0.001	9180863.00000
0.01	9129551.00000
0.05	9397110.00000
0.1	9903681.00000

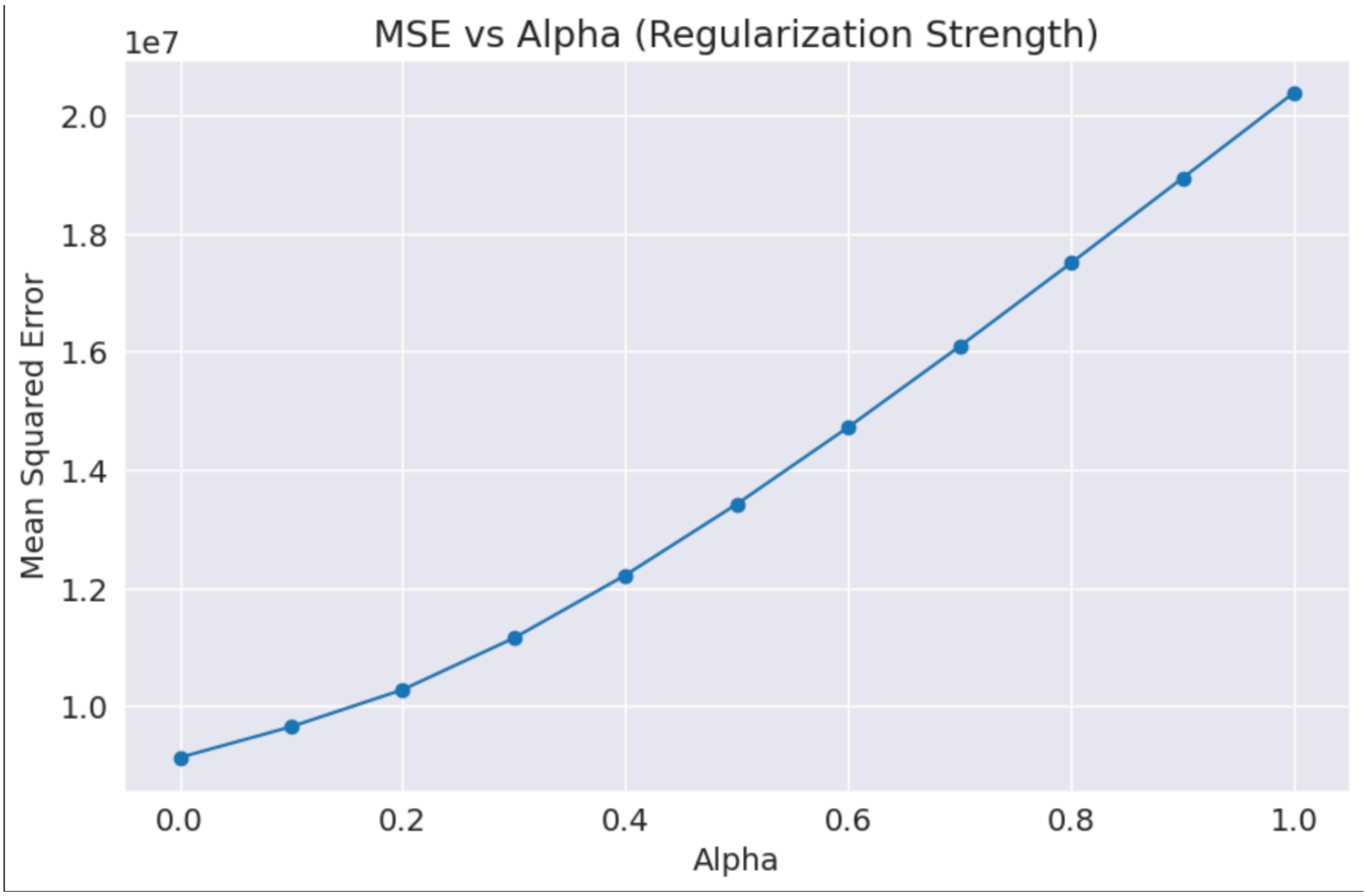


Optimal Value of Learning Rate=0.01

Experiment 4:(Linear regression using regularization(ridge regression)

Variation of MSE values with Hyper parameter alpha at learning rate=0.01:

Alpha	MSE
0.0	9129551.00000
0.1	9650933.00000
0.2	10276922.00000
0.3	11151564.00000
0.4	12216141.00000
0.5	13419016.00000
0.6	14719873.00000
0.7	16087687.00000
0.8	17498630.00000
0.9	18934422.00000
1.0	20380984.00000



Optimal Value of alpha=0.0

Experiment 5:

Model	Learning Rate	Alpha	MSE on test set
LIN_MODEL_CLOSED	-	-	9077460.414
LIN_MODEL_GRAD	0.01	-	9168969.00
LIN_MODEL_RIDGE	0.01	0.00	9107635.00

Observation:

On calculating mean squared error on at optimal values of hyper parameters alpha and learning rate for different models, we find that LIN_MODEL_CLOSED has lowest MSE, then LIN_MODEL_RIDGE and LIN_MODEL_GRAD has the highest value.