Hackathon: Restaurant Franchise Revenue

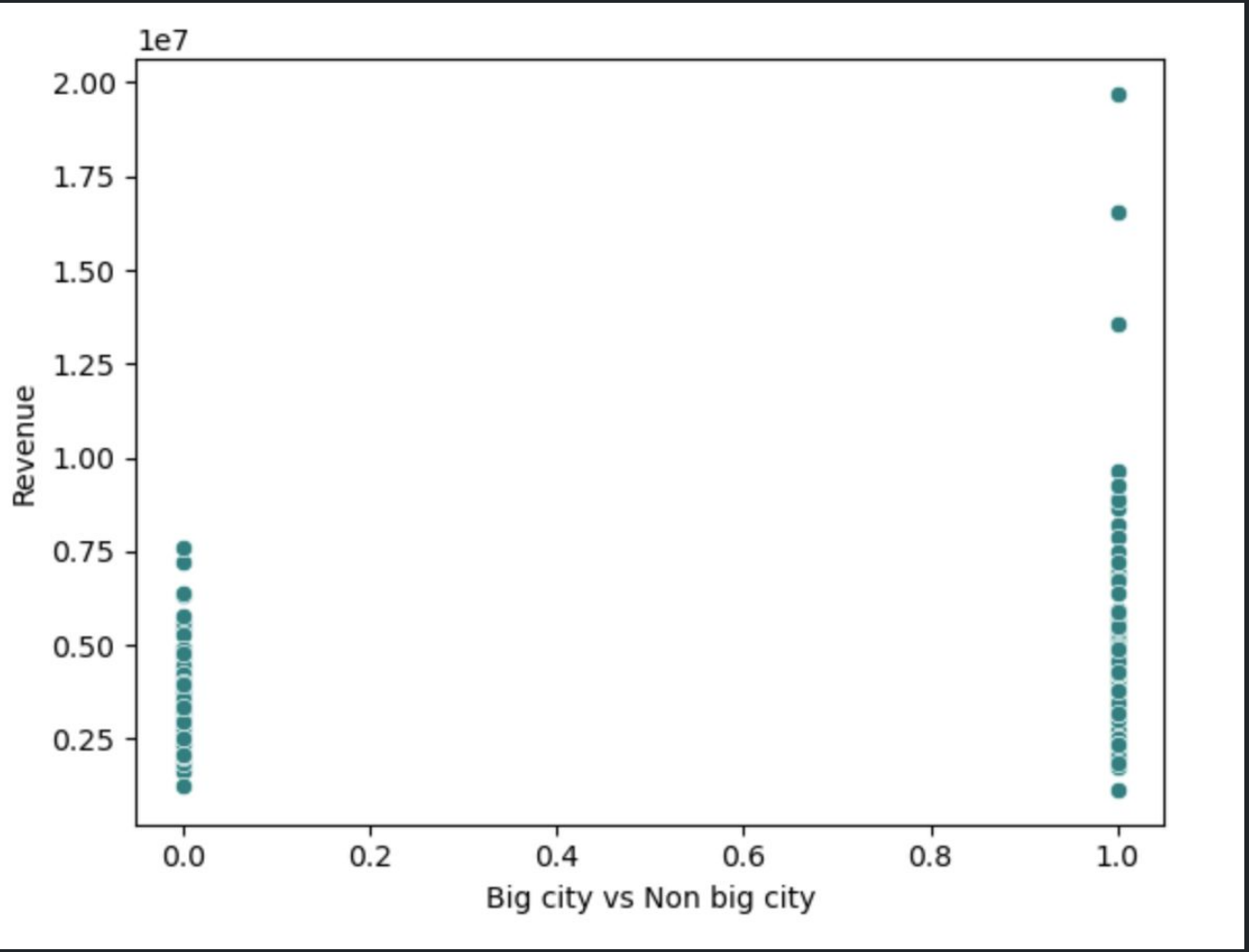
Group Project

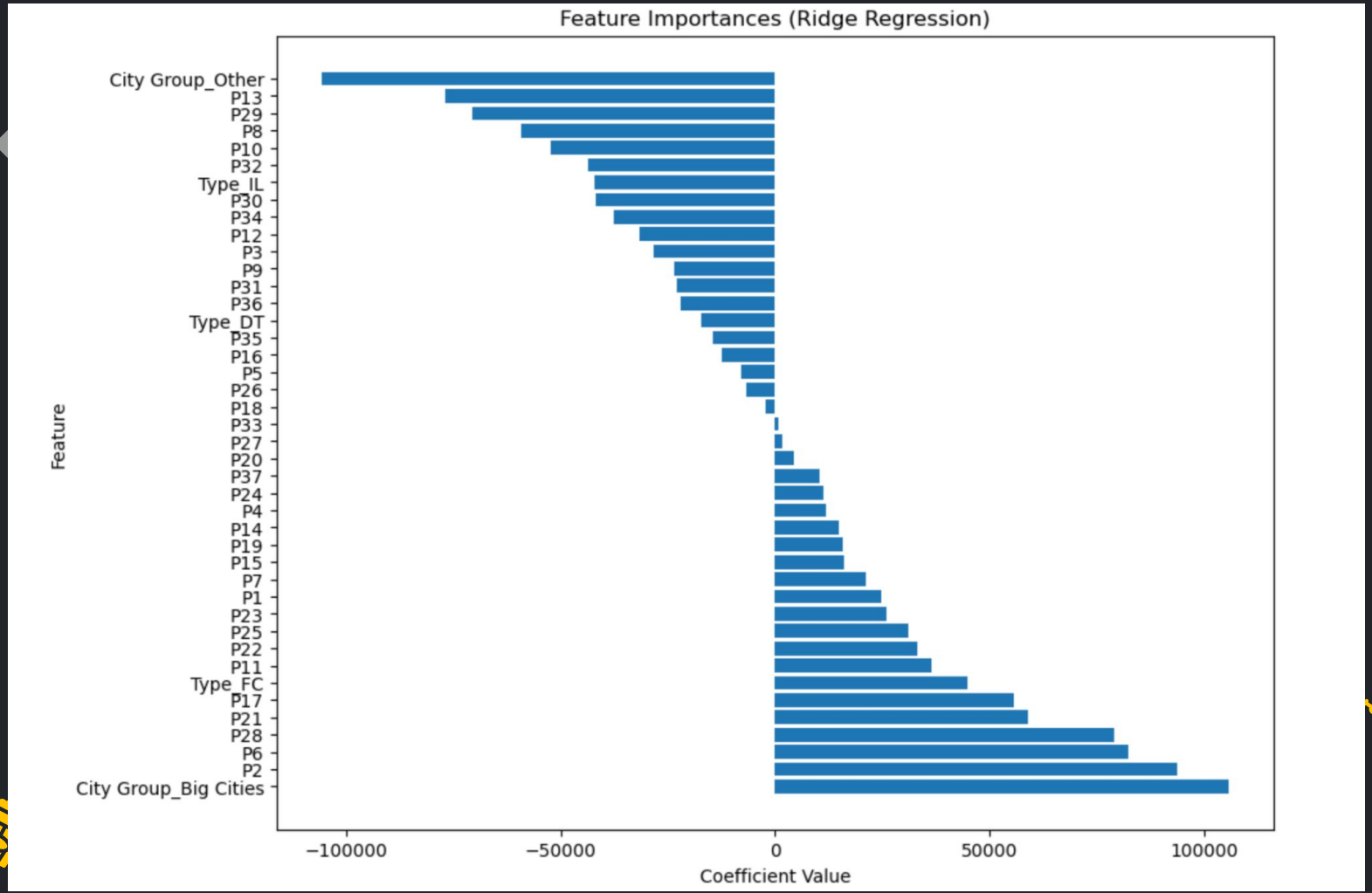
Problem Statement

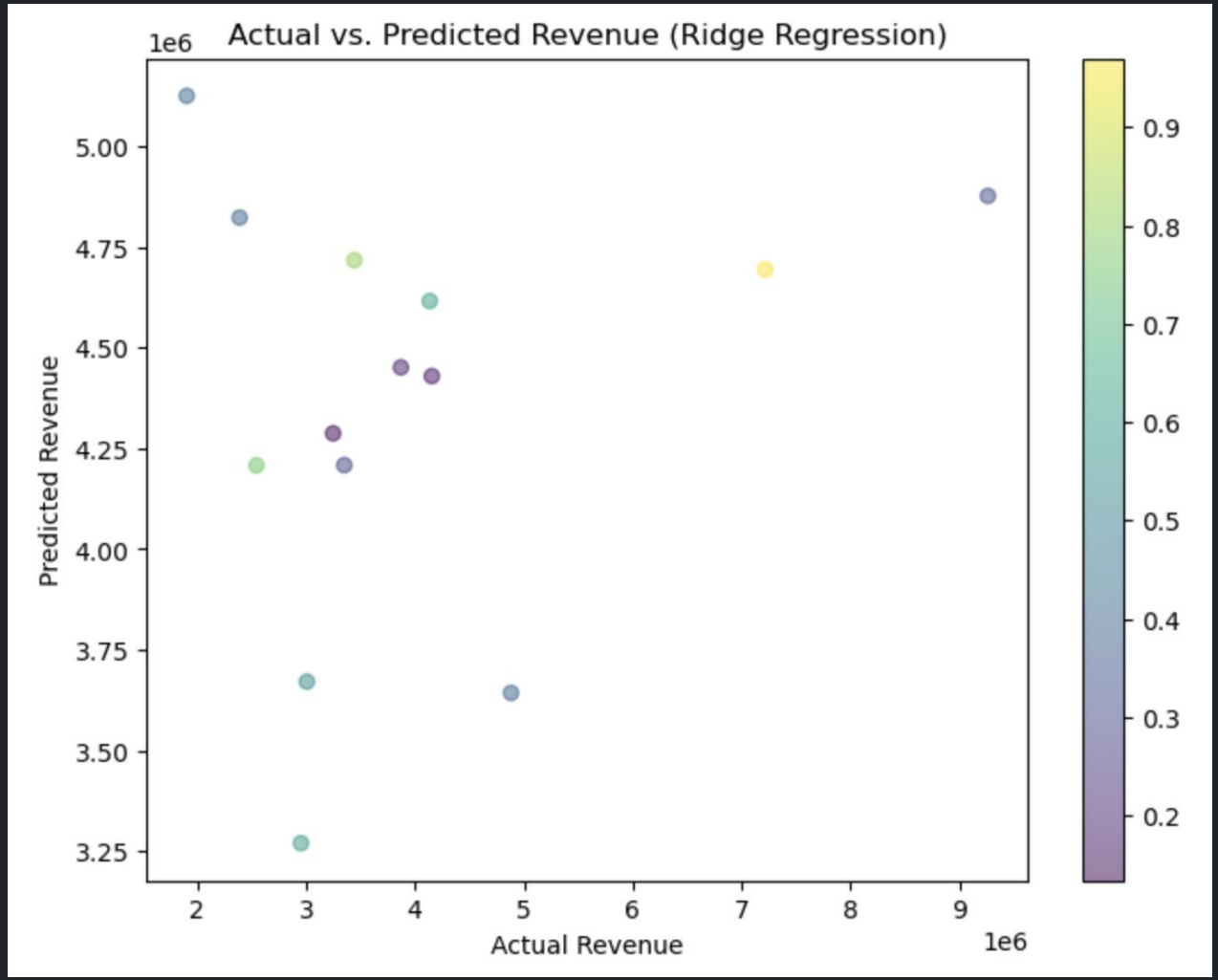
Our group chose a regression project on restaurant revenue prediction. The dataset came with a lot of different features, such as the types of cities, types of restaurants, and various demographic information including, but not limited to, population in any given area, age, gender distribution, and local commercial real estate.

Methods

- Imported libraries and datasets
- Cleaned data on restaurant features
- Conducted exploratory data analysis
- Visualized data on variables that can affect restaurant revenue
- Created models to predict restaurant revenue
- Evaluated and compared different models and tuned hyperparameters


1e7
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
Revenue
0.0
0.2
0.4
0.6
0.8
1.0
Big city vs Non big city


Feature Importances (Ridge Regression)
City Group_Other
P13
P29
P8
P10
P32
Type_IL
P30
P34
P12
P3
P9
P31
P36
Type_DT
P35
P16
P5
P26
P18
P33
P27
P20
P37
P24
P4
P14
P19
P15
P7
P1
P23
P25
P22
P11
Type_FC
P17
P21
P28
P6
P2
City Group_Big Cities
Feature
−100000
−50000
0
50000
100000
Coefficient Value


Actual vs. Predicted Revenue (Ridge Regression)
1e6
5.00
4.75
4.50
4.25
4.00
3.75
3.50
3.25
Predicted Revenue
2
3
4
5
6
7
8
9
1e6
Actual Revenue
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2

Results

- Our best score is ~1.80 million (on Kaggle) using the Ridge model.
- The top score is ~1.73 million.
- Our score puts us in the top 225 out of 2,259 submissions or approximately the top 10%.
- BONUS: we were able to obtain a bronze medal.

Thank You!