

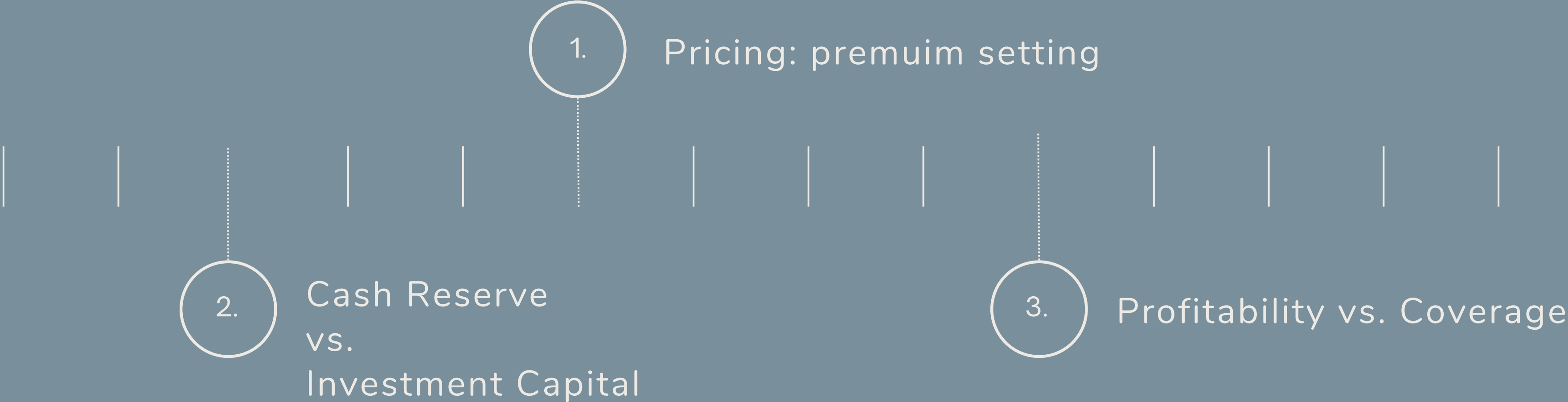
MSCI-623

Beilin Ye
Harneet Kaur
Lida Ghasemi



CLAIM AMOUNT PREDICTION ON U.S. CAR INSURANCE

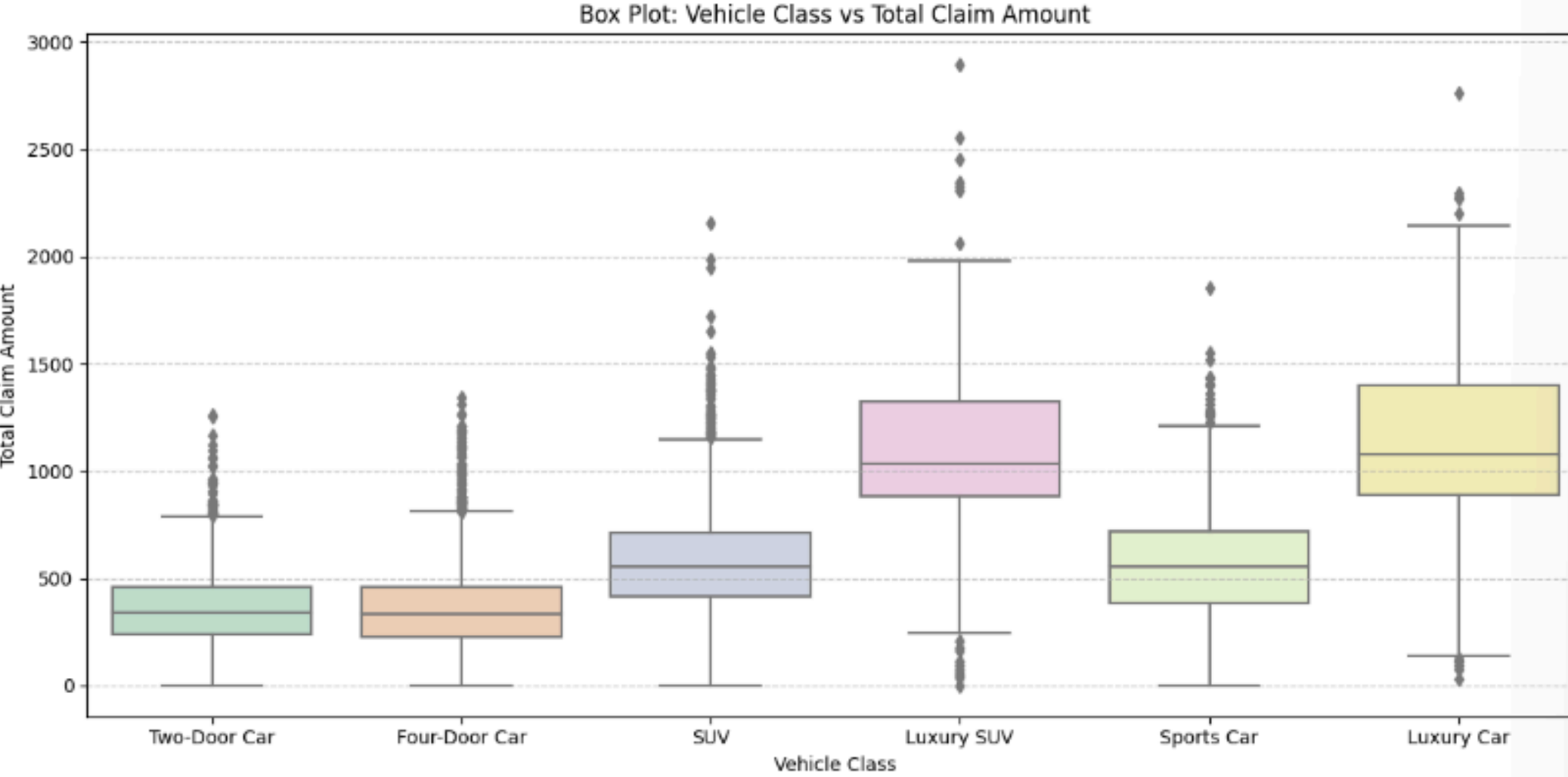
Introduction: why is it important:



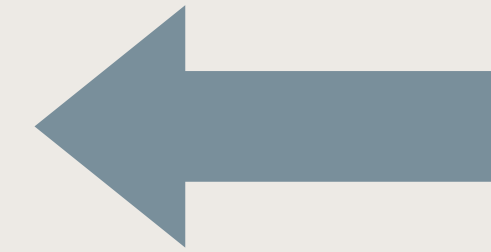
DATA

MAIN FEATURES IN OUR DATASETS:

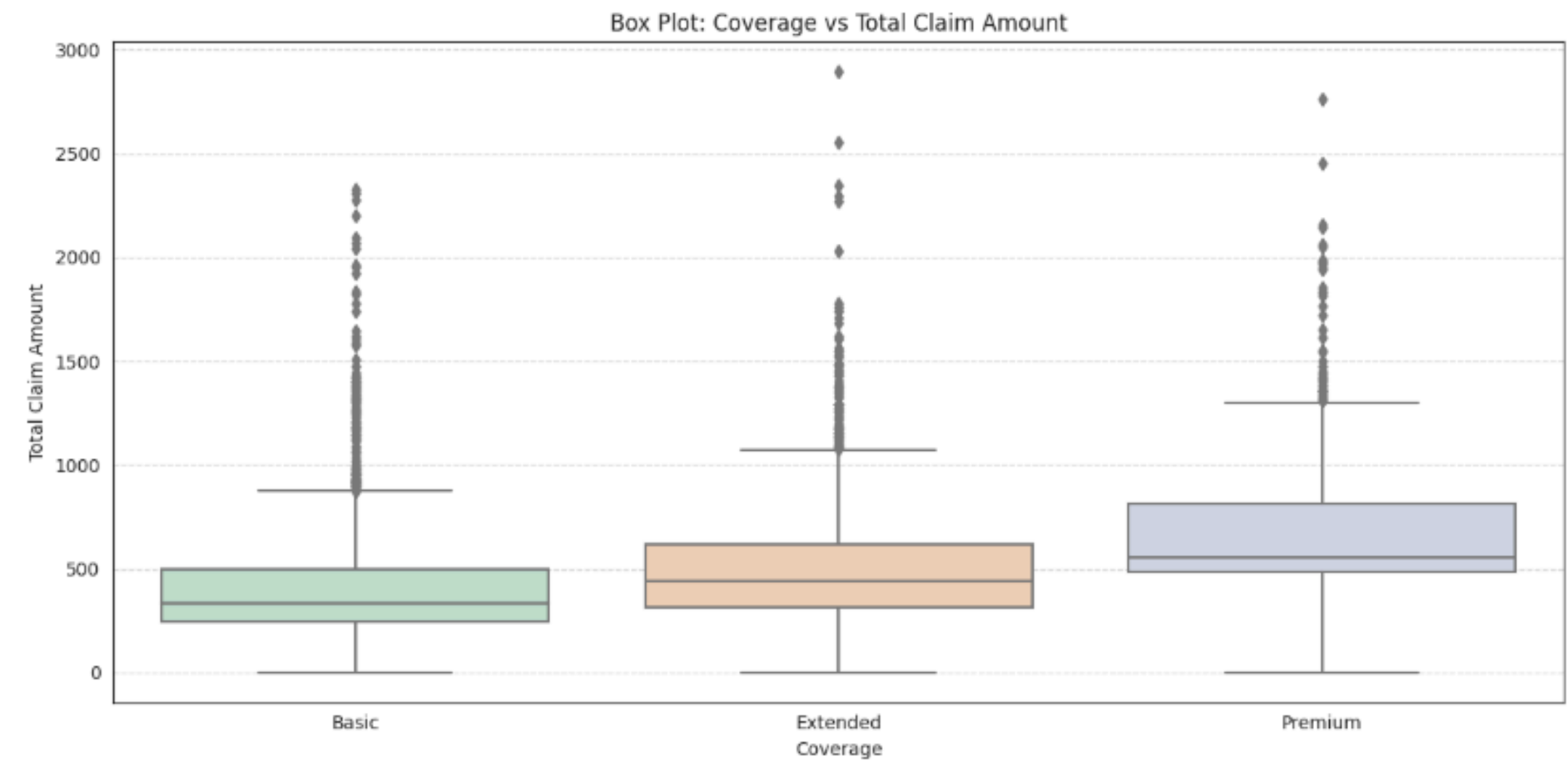
- Monthly Premium
- Age, Gender, Income
- Employment Status
- Vehicle Size, Car Make
- Number of Vehicles
- Location
- Coverage Type



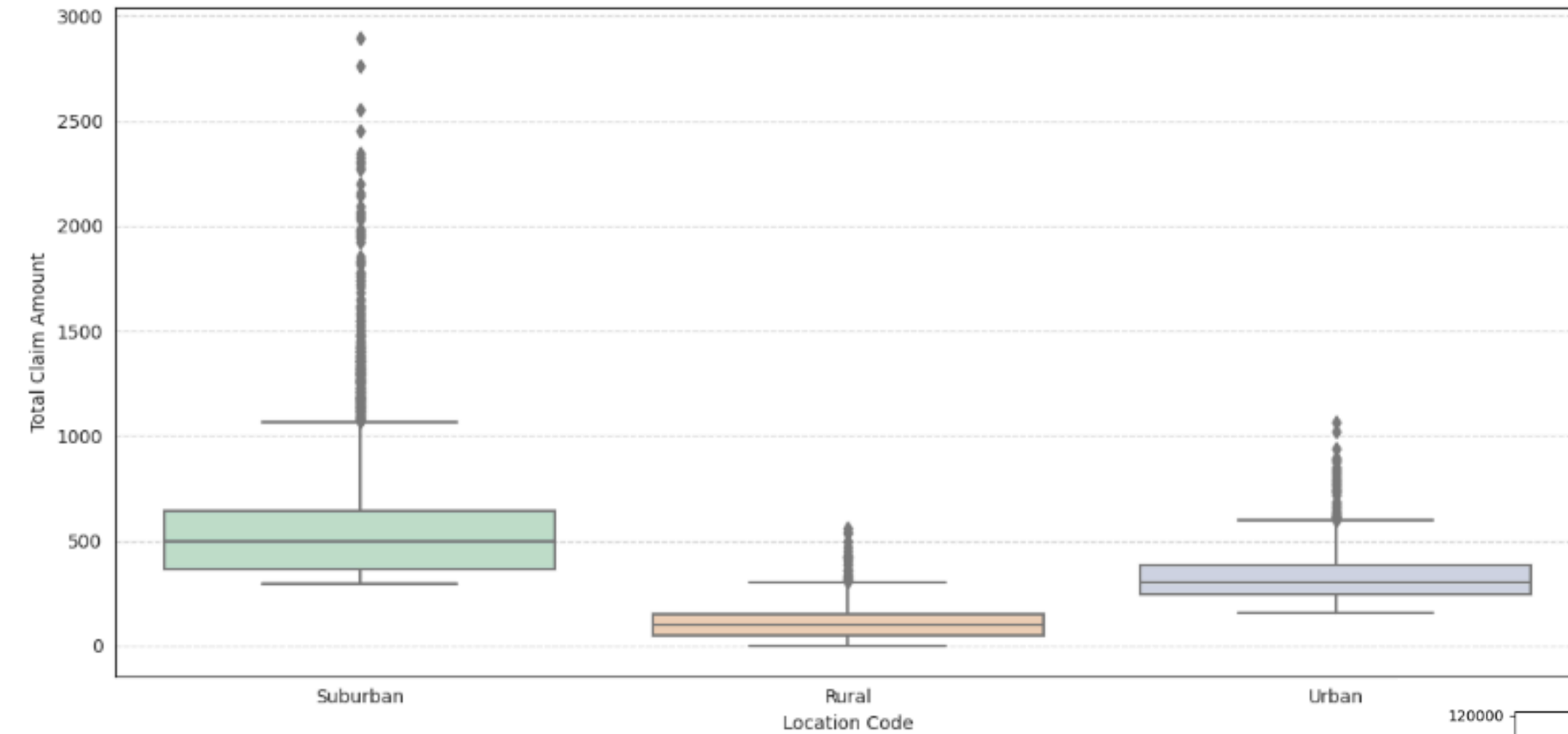
Vehicle Class vs Total Claim Amount



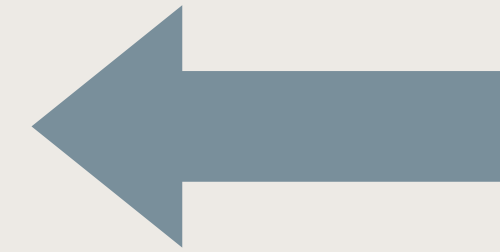
Coverage Type vs Total Claim Amount



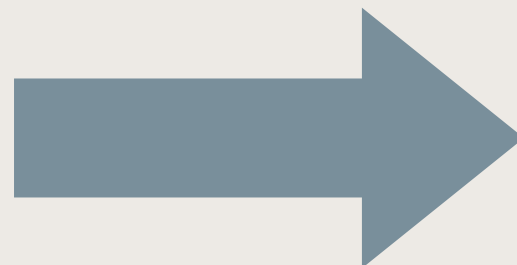
Box Plot: Location Code vs Total Claim Amount



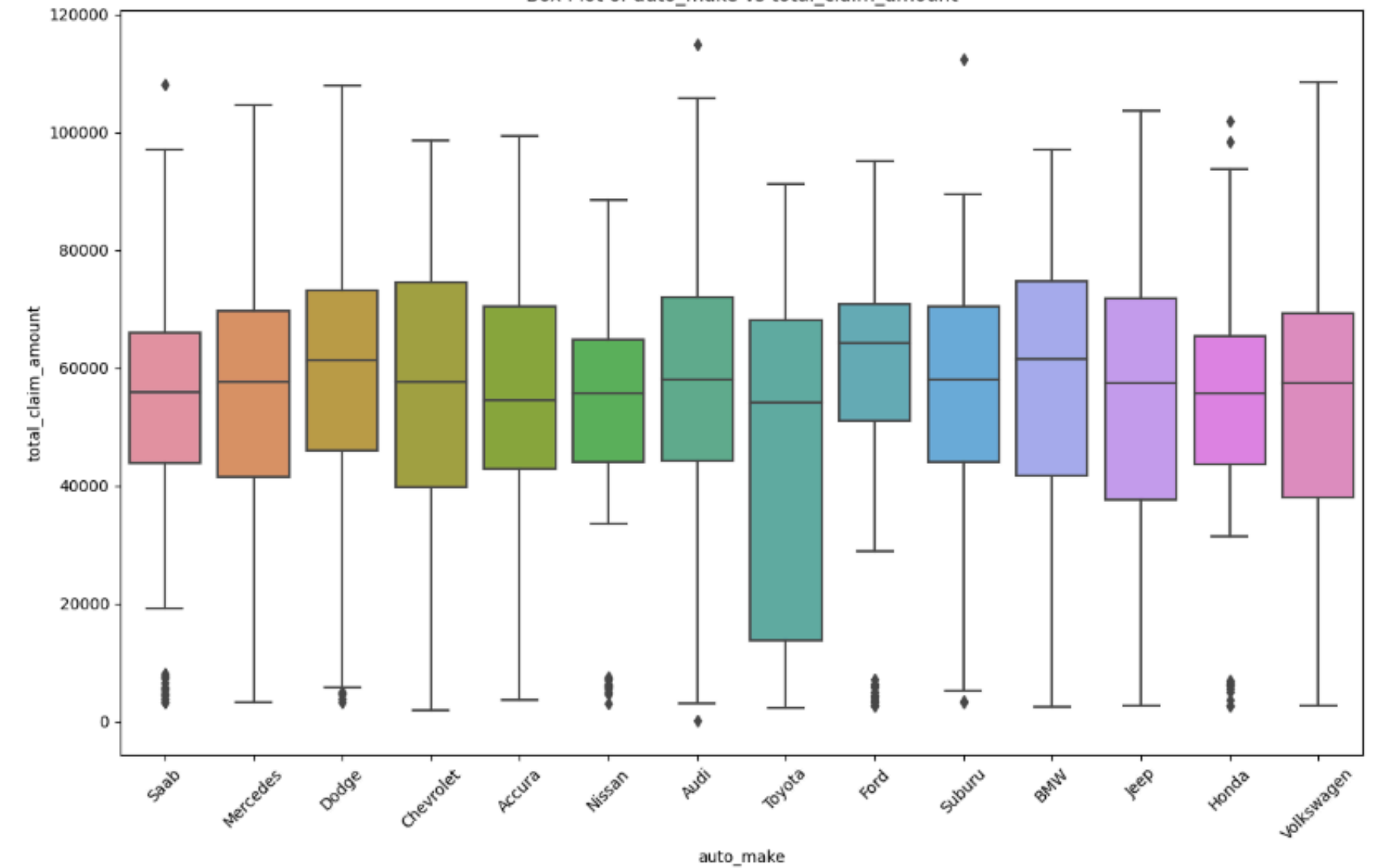
Location vs Total Claim Amount



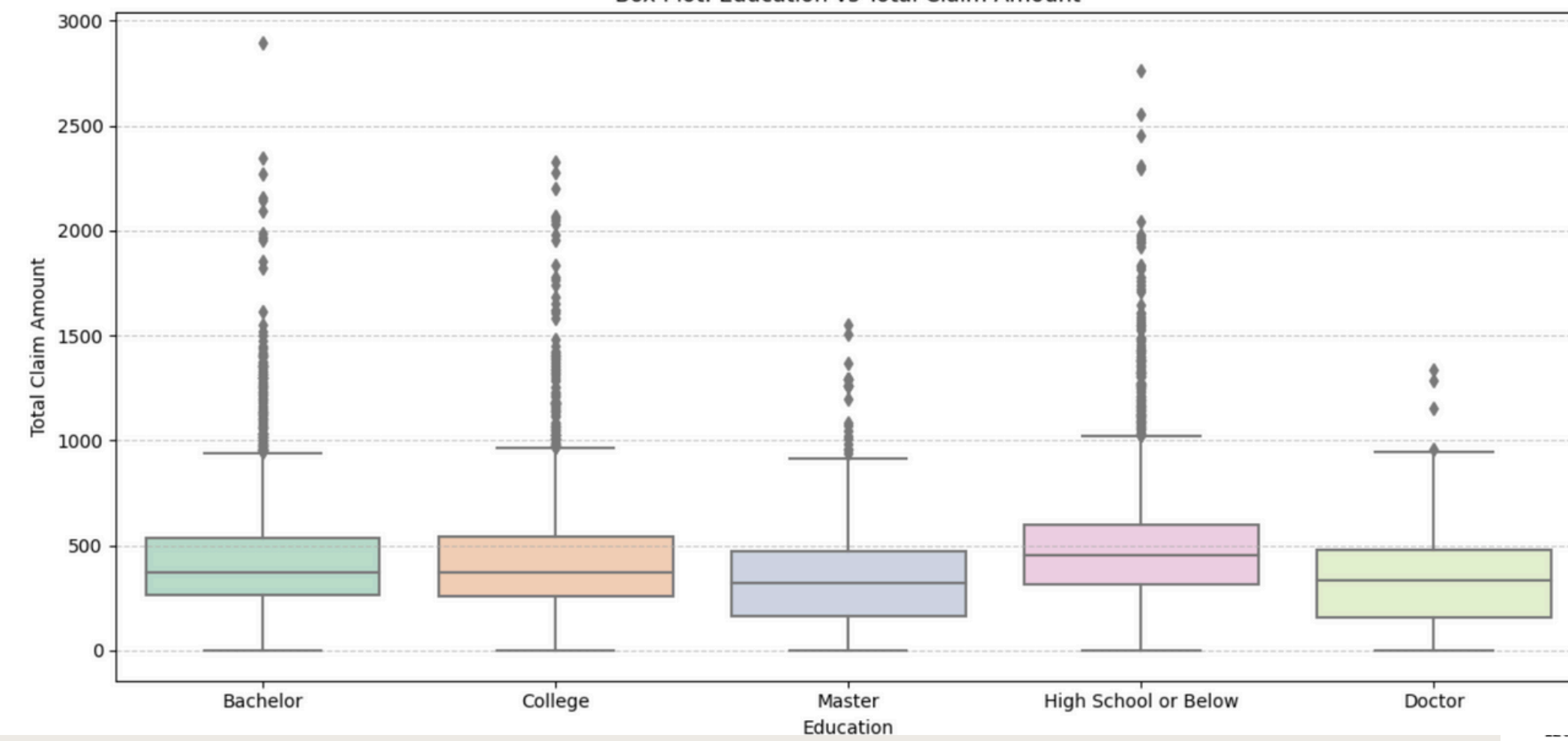
Car Make vs Total Claim Amount



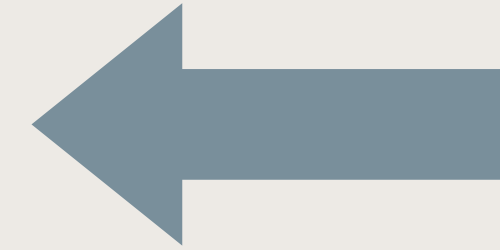
Box Plot of auto_make vs total_claim_amount



Box Plot: Education vs Total Claim Amount



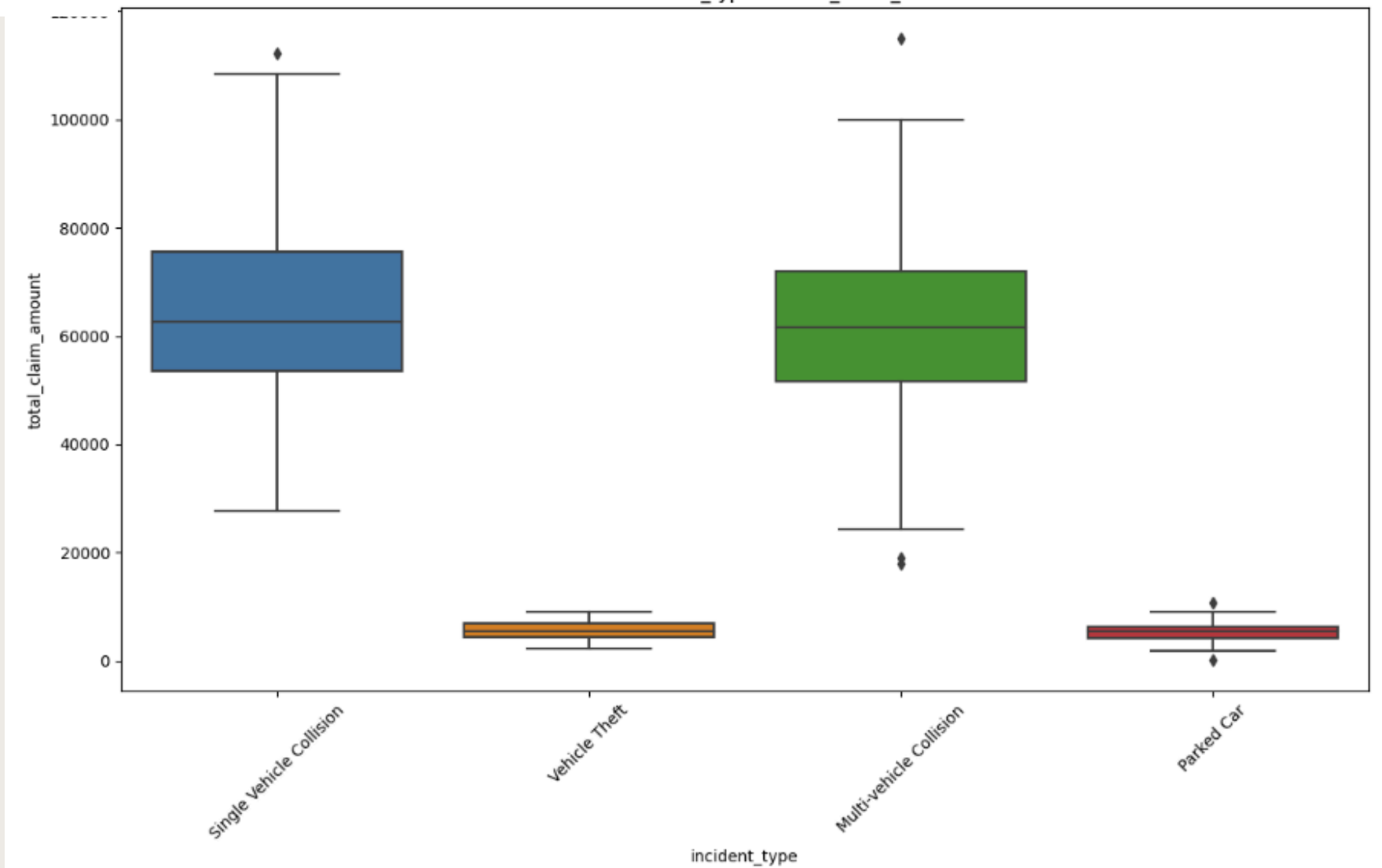
Education vs Total Claim Amount



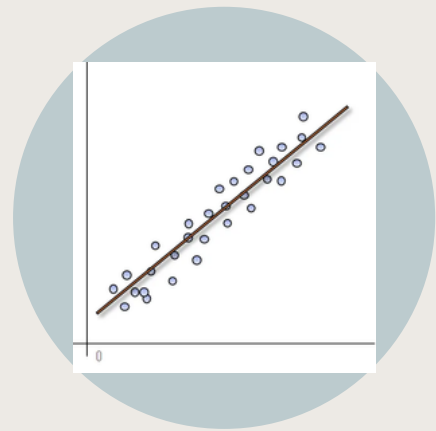
Incident Type vs Total Claim Amount



Box Plot of incident_type vs total_claim_amount

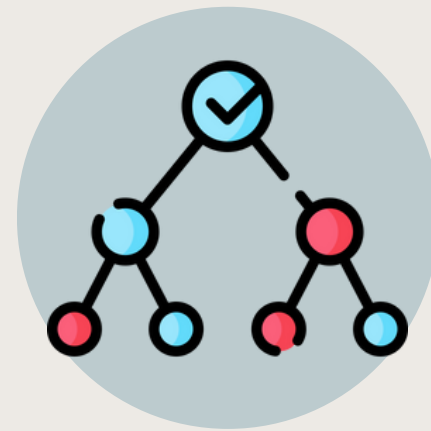


METHODOLOGIES & RESULTS



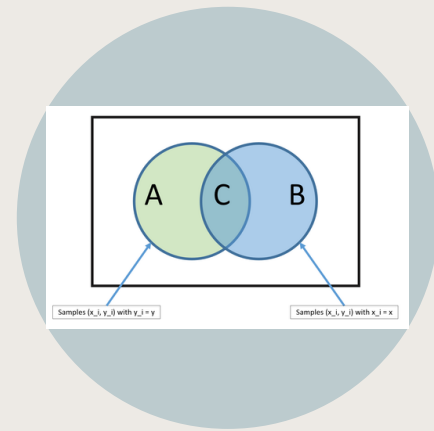
Linear Regression

We achieved 76% and 96% accuracy in our datasets using Linear Regression for predicting the total claim amount.



Decision Trees

We achieved 85% and 96% accuracy in our datasets using Decision Trees for predicting the total claim amount.



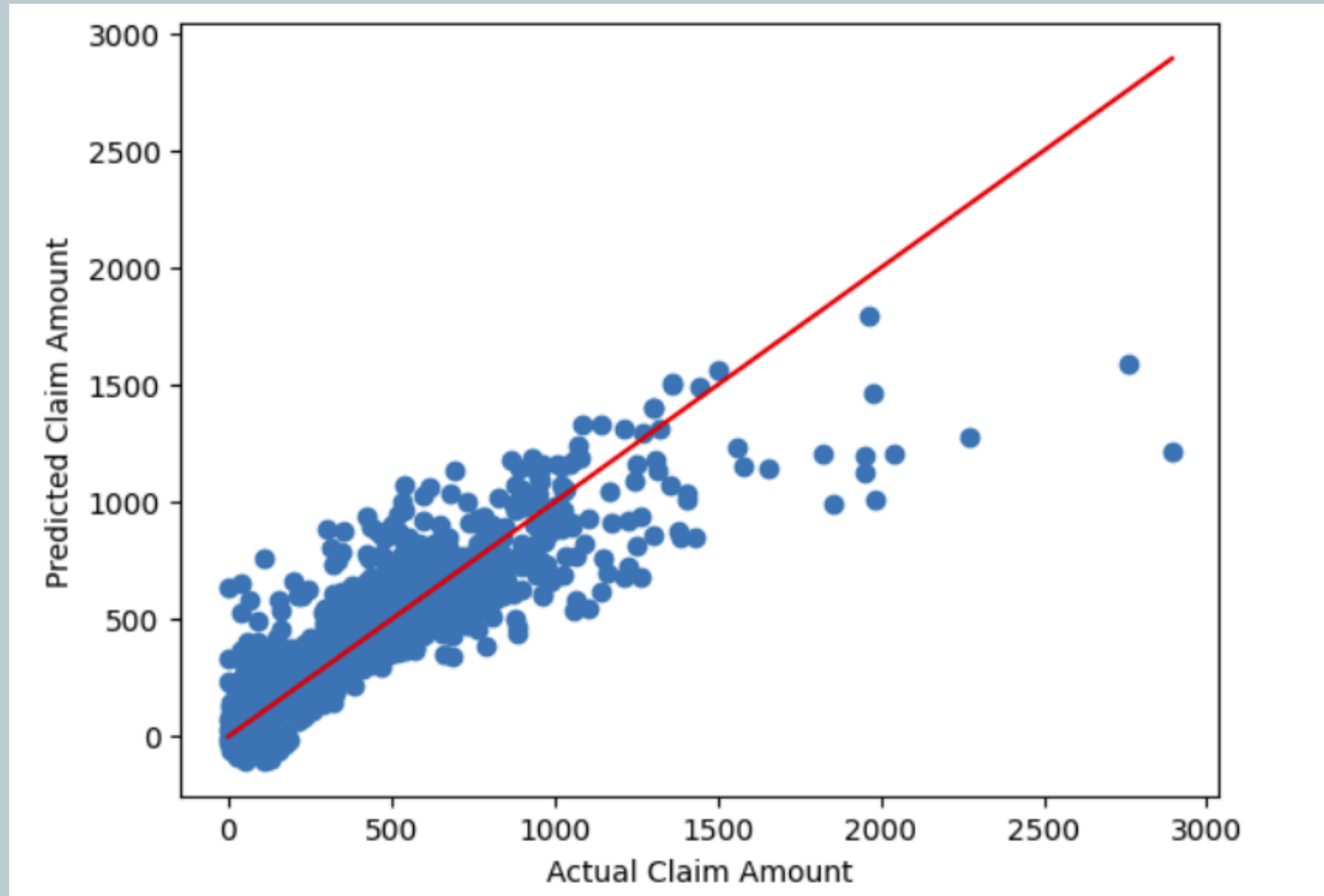
Naive Bayes

We achieved 96% and 90% accuracy in our datasets using Naive Bayes for predicting the total claim amount class of high or low.

Linear Regression

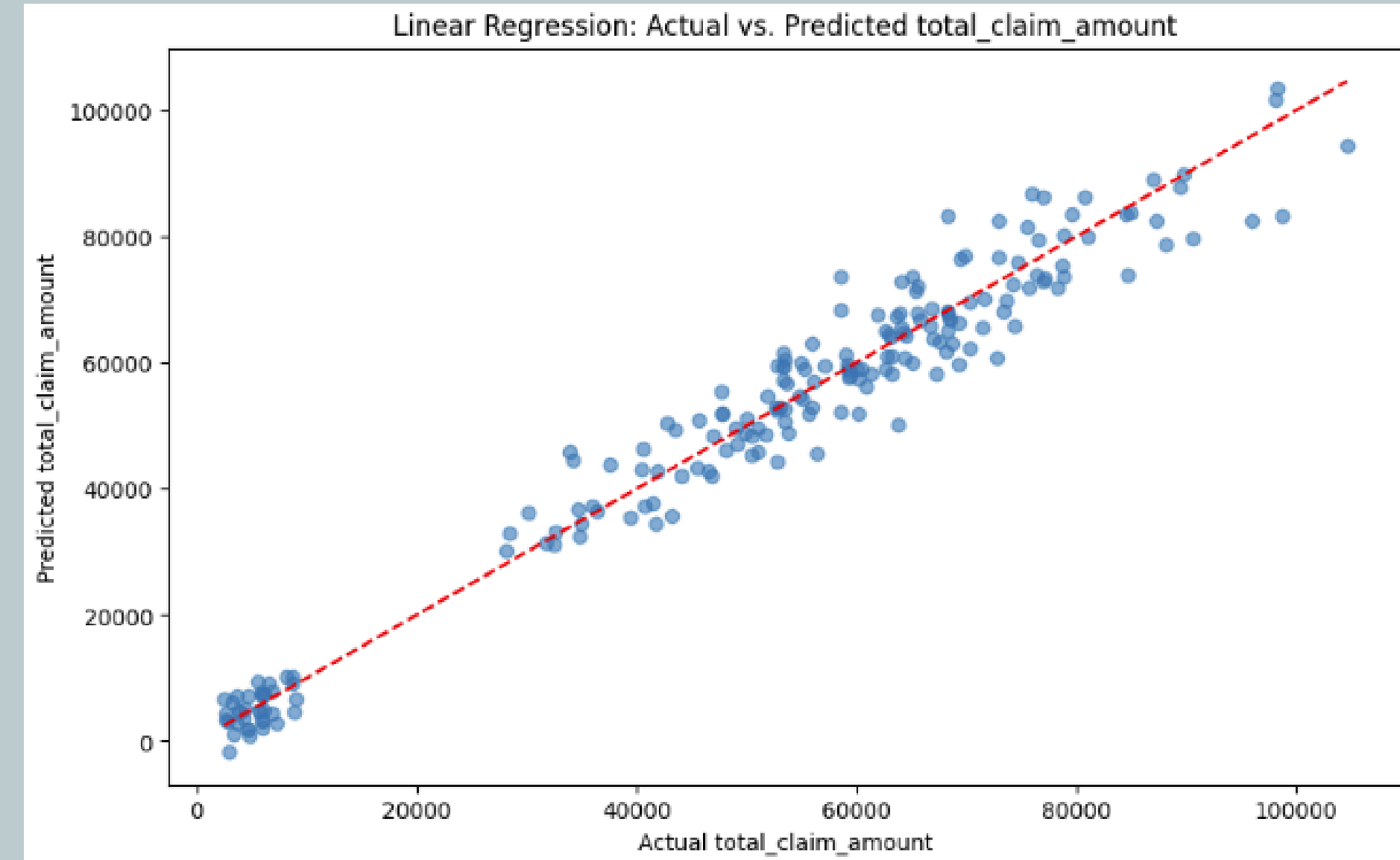
Actual vs Predicted Claim Amount

Dataset-1



R-squared: 0.7241172999836459

Dataset-2



R-squared : 0.9595324511314519

Decision Trees

Before
Hyperparameterism

R-Squared: ~50%

After
Hyperparameterism

R-Squared: ~81%

Dataset-1

Before
Hyperparameterism

R-Squared: ~94%

After
Hyperparameterism

R-Squared: ~96%

Dataset-2

Naive Bayes

```
Confusion Matrix:  
[[844  60]  
 [  0 922]]
```

Precision Score: 0.9398

Dataset-1

```
Confusion Matrix:  
[[90 15]  
 [ 5 90]]
```

Precision Score: 0.9581

Dataset-2

FINAL CONCLUSION & BUSINESS INSIGHTS

Major Features

Dataset 1

- Monthly Premium Auto
- Coverage
- Vehicle class
- Location

Dataset 2

- Number of vehicles in accident
- Location
- Incident date
- Car make

Minor Features

Dataset 1

- Age
- Education level
- Gender

Dataset 2

- Age
- Education level
- Gender

Thank You!