InsureWise:

Risk and Premium Management

Leveraging Machine Learning to Improve Premium Assessment and Minimize Claim Losses

Introduction

Background and Context

- Car insurance claims significantly impact the profitability and operational efficiency of insurance companies
- Understanding the factors influencing claim probability is essential for effective risk management and pricing strategies

Significance of Predicting Insurance Claim Probability

- Predicting claim probability allows for more accurate premium setting, reducing financial risk for insurance companies.
- It enhances the ability to offer personalized policies, aligning premiums with the actual risk profiles of policyholders.
- Early identification of high-risk policies can lead to proactive measures, improving overall customer satisfaction and retention.

Overview of the Project

- Develop a predictive model to assess the claim probability for car insurance policies.
- Utilize comprehensive car policy features and safety ratings to build the model.

Objectives

Primary Goals

- Develop a Predictive Model: Our primary objective is to create a robust predictive model to assess the probability of car insurance claims.
- Risk Assessment: This model will help estimate the likelihood of a claim being made within a specified period, such as six months, enabling more accurate premium setting and risk management.

Secondary Goals

- Identify Key Factors: Beyond prediction, we aim to identify and analyze the critical factors that influence claim probability.
- Data Exploration: We will delve into the dataset to uncover insights into the vehicle, policyholder, and policy characteristics that contribute to higher claim risks.
- Improve Risk Management: By understanding these factors, we can provide actionable insights to insurance companies to refine their risk assessment and mitigation strategies.
- Enhance Customer Satisfaction: Offering personalized premiums and preventive measures will ultimately lead to improved customer satisfaction and retention.

Business Problem

Insurance Industry Challenge

- Insurance Claim Risk: Accurately predicting car insurance claims is a significant challenge in the insurance industry.
- Financial Impact: Claims have a considerable financial impact on insurance companies, affecting profitability and operational efficiency.
- Complex Risk Factors: Various factors, including policyholder behavior, vehicle characteristics, and external conditions, contribute to claim probability, making risk assessment complex.

Project's Role

- Leveraging Data Analytics: Our project aims to address this challenge through advanced data analytics and machine learning techniques.
- Predictive Modeling: By developing a predictive model, we strive to:
    - Assess Claim Probability: Provide accurate predictions of claim likelihood within a specified period.
    - Optimize Premiums: Enable insurance companies to set more precise and fair premiums based on individual risk profiles.
    - Enhance Risk Management: Offer insights into key risk factors, facilitating better risk management strategies.
    - Improve Customer Experience: Contribute to increased customer satisfaction by offering personalized and transparent insurance policies.

Assumptions

- Representative Dataset: The dataset accurately represents the population of car insurance policyholders under study.
- Sufficient Variables: The provided variables, including demographic, vehicle, and policy details, are sufficient to make accurate predictions about insurance claim probability.
- Data Integrity: The data is free from significant errors or biases that could skew the predictive model's accuracy.
- Stable Relationships: The relationship between the variables and insurance claim probability remains stable over time, ensuring the model's long-term reliability.

Data Description

Dataset Overview

- It comprises 44 columns, providing detailed insights into various factors influencing insurance claim probability.

Feature Descriptions

- **Demographic Attributes:** Information about policyholders, such as age and area cluster.
- **Vehicle Attributes:** Details about the insured vehicles, including age, make, model, segment, fuel type, and safety features.
- **Policy Attributes:** Data on policy specifics, such as policy tenure, population density, and coverage details.

Target Variable

- The focal point of our analysis is the "is_claim" variable.
- It indicates whether a claim has been made (1) or not (0) within the policy period.

Name	DType	Description
policy_id	Text	The unique identifier for each insurance policy.
policy_tenure	Number	The length of time (in years) that the policy has been active.
age_of_car	Number	The age of the insured car (in years) at the time the policy was taken.
age_of_policyholder	Number	The age of the policyholder (in years) at the time the policy was taken.
area_cluster	Text	A categorical variable representing the cluster or category to which the area of residence belongs.
population_density	Number	A measure of the population density of the area where the policyholder resides.
make	Text	The make or manufacturer of the insured car.
segment	Text	The segment or category to which the insured car belongs (e.g., compact, sedan, SUV).
model	Text	The specific model or variant of the insured car.
fuel_type	Text	The type of fuel used by the insured car (e.g., petrol, diesel, electric).
max_torque	Text	The maximum torque output of the car's engine.

Data Description

Dataset Overview

- It comprises 44 columns, providing detailed insights into various factors influencing insurance claim probability.

Feature Descriptions

- **Demographic Attributes:** Information about policyholders, such as age and area cluster.
- **Vehicle Attributes:** Details about the insured vehicles, including age, make, model, segment, fuel type, and safety features.
- **Policy Attributes:** Data on policy specifics, such as policy tenure, population density, and coverage details.

Target Variable

- The focal point of our analysis is the "is_claim" variable.
- It indicates whether a claim has been made (1) or not (0) within the policy period.

Name	DType	Description
max_power	Text	The maximum power output of the car's engine.
engine_type	Text	The type of engine used in the insured car (e.g., inline, V-type).
airbags	Number	The number of airbags installed in the car.
is_esc	Text	A binary variable indicating whether the car has an electronic stability control (ESC) system.
is_adjustable_steering	Text	A binary variable indicating whether the car has adjustable steering.
is_tpms	Text	A binary variable indicating whether the car has a tire pressure monitoring system (TPMS).
is_parking_sensors	Text	A binary variable indicating whether the car has parking sensors.
is_parking_camera	Text	A binary variable indicating whether the car has a parking camera.
rear_brakes_type	Text	The type of rear brakes used in the car.
cylinder	Number	The number of cylinders in the car's engine.

Data Description

Dataset Overview

- It comprises 44 columns, providing detailed insights into various factors influencing insurance claim probability.

Feature Descriptions

- **Demographic Attributes:** Information about policyholders, such as age and area cluster.
- **Vehicle Attributes:** Details about the insured vehicles, including age, make, model, segment, fuel type, and safety features.
- **Policy Attributes:** Data on policy specifics, such as policy tenure, population density, and coverage details.

Target Variable

- The focal point of our analysis is the "is_claim" variable.
- It indicates whether a claim has been made (1) or not (0) within the policy period.

Name	DType	Description
transmission_type	Text	The type of transmission used in the car (e.g., manual, automatic).
gear_box	Number	The number of gears in the car's gearbox.
steering_type	Text	The type of steering system used in the car.
turning_radius	Number	The minimum radius of the circular path that the car can make.
length	Number	The length of the car.
width	Number	The width of the car.
height	Number	The height of the car.
gross_weight	Number	The gross weight or total weight of the car.
is_front_fog_lights	Text	A binary variable indicating whether the car has front fog lights.
is_rear_window_wiper	Text	A binary variable indicating whether the car has a rear window wiper.

Data Description

Dataset Overview

- It comprises 44 columns, providing detailed insights into various factors influencing insurance claim probability.

Feature Descriptions

- **Demographic Attributes:** Information about policyholders, such as age and area cluster.
- **Vehicle Attributes:** Details about the insured vehicles, including age, make, model, segment, fuel type, and safety features.
- **Policy Attributes:** Data on policy specifics, such as policy tenure, population density, and coverage details.

Target Variable

- The focal point of our analysis is the "is_claim" variable.
- It indicates whether a claim has been made (1) or not (0) within the policy period.

Name	DType	Description
is_rear_window_defogger	Text	A binary variable indicating whether the car has a rear window defogger.
is_brake_assist	Text	A binary variable indicating whether the car has a brake assist system.
is_power_door_locks	Text	A binary variable indicating whether the car has power door locks.
is_central_locking	Text	A binary variable indicating whether the car has central locking.
is_power_steering	Text	A binary variable indicating whether the car has power steering.
is_driver_seat_height_adjustable	Text	A binary variable indicating whether the driver's seat height is adjustable.
is_day_night_rear_view_mirror	Text	A binary variable indicating whether the car has a day/night rearview mirror.
is_ecw	Text	A binary variable indicating whether the car has an electronic crash warning (ECW) system.
is_speed_alert	Text	A binary variable indicating whether the car has a speed alert system.
ncap_rating	Number	The safety rating of the car according to the New Car Assessment Program (NCAP).

Data Cleaning

Brewing the Data

Extract Values from Columns:

- Extracted numeric values from the 'max_torque' and 'max_power' columns and created new columns 'torque_value' and 'power_value' respectively

Drop Irrelevant Columns:

- Dropped the redundant 'policy_id' column along with the original 'max_torque' and 'max_power' columns

Column Identification:

- Binary Columns: Identified binary columns that contain 'Yes' or 'No' values
- Categorical Columns: Identified categorical columns with unique values less than or equal to 22
- Numerical Columns: Identified the remaining columns as numerical
- Target Column: Defined 'is_claim' as the target column for the predictive model

Categorical Variables
1. 'area_cluster'
2. 'population_density'
3. 'make'
4. 'segment'
5. 'model'
6. 'fuel_type'
7. 'engine_type'
8. 'airbags'
9. 'rear_brakes_type'
10. 'displacement'
11. 'cylinder'
12. 'transmission_type'
13. 'gear_box'
14. 'steering_type'
15. 'turning_radius'
16. 'length'
17. 'width',
18. height'
19. 'gross_weight'
20. 'ncap_rating'
21. 'torque_value'
22. 'power_value'

Binary Variables
1. 'is_esc'
2. 'is_adjustable_steering'
3. 'is_tpms'
4. 'is_parking_sensors'
5. 'is_parking_camera'
6. 'is_front_fog_lights'
7. 'is_rear_window_wiper'
8. 'is_rear_window_washer'
9. 'is_rear_window_defogger'
10. 'is_brake_assist'
11. 'is_power_door_locks'
12. 'is_central_locking'
13. 'is_power_steering'
14. 'is_driver_seat_height_adjustable'
15. 'is_day_night_rear_view_mirror'
16. 'is_ecw'
17. 'is_speed_alert'

Numerical Variables
1. 'policy_tenure'
2. 'age_of_car'
3. 'age_of_policyholder'

Exploratory Data Analysis (EDA)

EDA – Summary Statistics

Unearthing Insights

- Policy Tenure: Most policy tenures are relatively short, as indicated by the mean value being close to the minimum value.
- Age of Car: Cars in the dataset are mostly quite new, with the average age being around 0.07.
- Age of Policyholder: The policyholders have a diverse age range, but most are relatively young.
- Population Density: There is a high variability in population density, with a mean value suggesting urban areas but a large standard deviation indicating a mix of rural and urban areas.
- Make: The make of cars varies, with an average value suggesting a variety of car brands.
- Airbags: The average number of airbags is around 3, indicating a moderate safety level.
- Displacement, Cylinder, Gear Box, Turning Radius, Length, Width, Height, Gross Weight*: These variables describe the physical characteristics of the cars, with average values indicating typical car dimensions and features.
- NCAP Rating: The safety ratings vary, with an average rating suggesting moderate safety levels.
- Is Claim: The percentage of claims is relatively low, with most entries indicating no claim made.

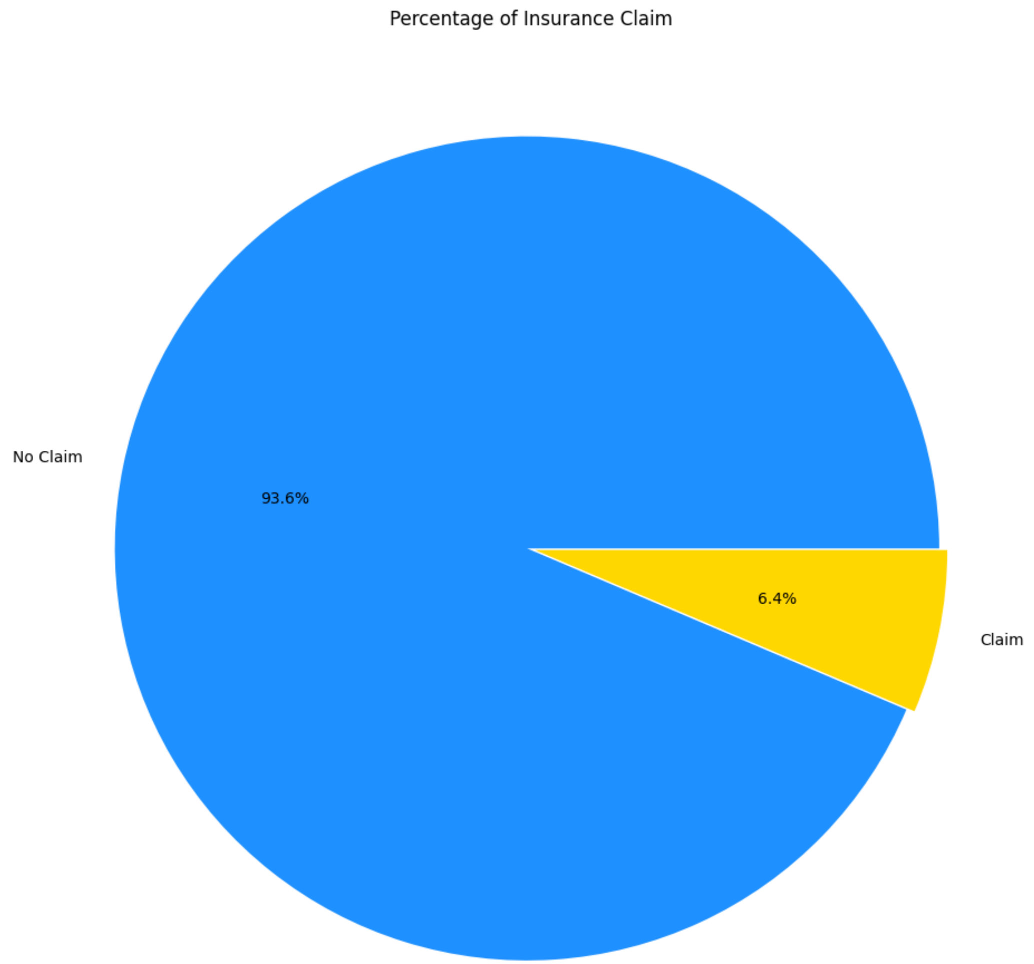
Variable	min	mean	max
policy_tenure	0.002735	0.611246	1.396641
age_of_car	0	0.069424	1
age_of_policyholder	0.288462	0.46942	1
population_density	290	18826.85867	73430
make	1	1.763722	5
airbags	1	3.137066	6
displacement	796	1162.355851	1498
cylinder	3	3.626963	4
gear_box	5	5.245443	6
turning_radius	4.5	4.852893	5.2
length	3445	3850.476891	4300
width	1475	1672.233667	1811
height	1475	1553.33537	1825
gross_weight	1051	1385.276813	1720
ncap_rating	0	1.75995	5
is_claim	0	0.063968	1

EDA – Distribution Analysis

Unearthing Insights

- Our target variable is highly imbalanced with 93.6% records belonging to 'No Claim' class while 6.4% belonging to 'Claim' class.

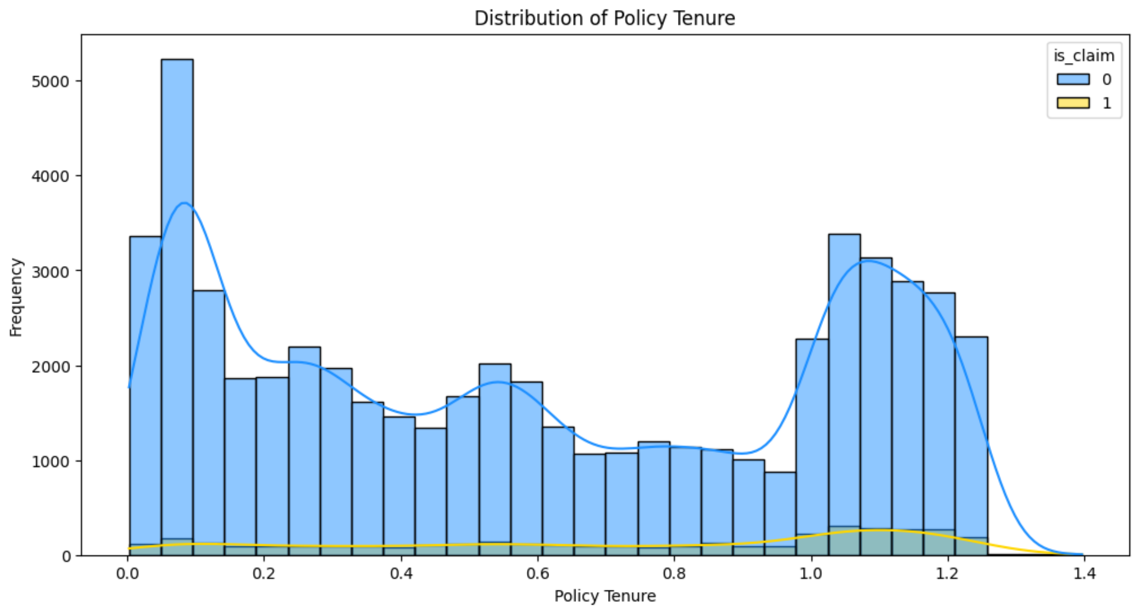
- The number of 'No Risk' cases outweighs the number of 'Risk' cases.

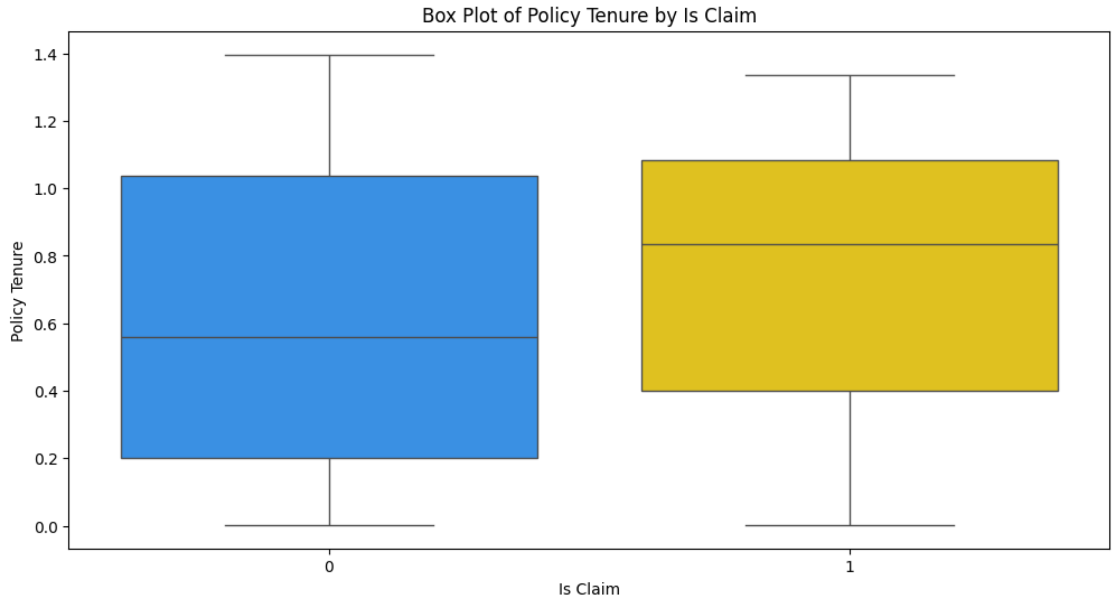


EDA – Distribution Analysis

Unearthing Insights

- Both claim and no-claim groups have a wide range of policy tenures.
- Claims occur across various policy tenures, but the density of claims seems to be higher at certain ranges.
- There is no significant difference in the median policy tenure between the two groups.

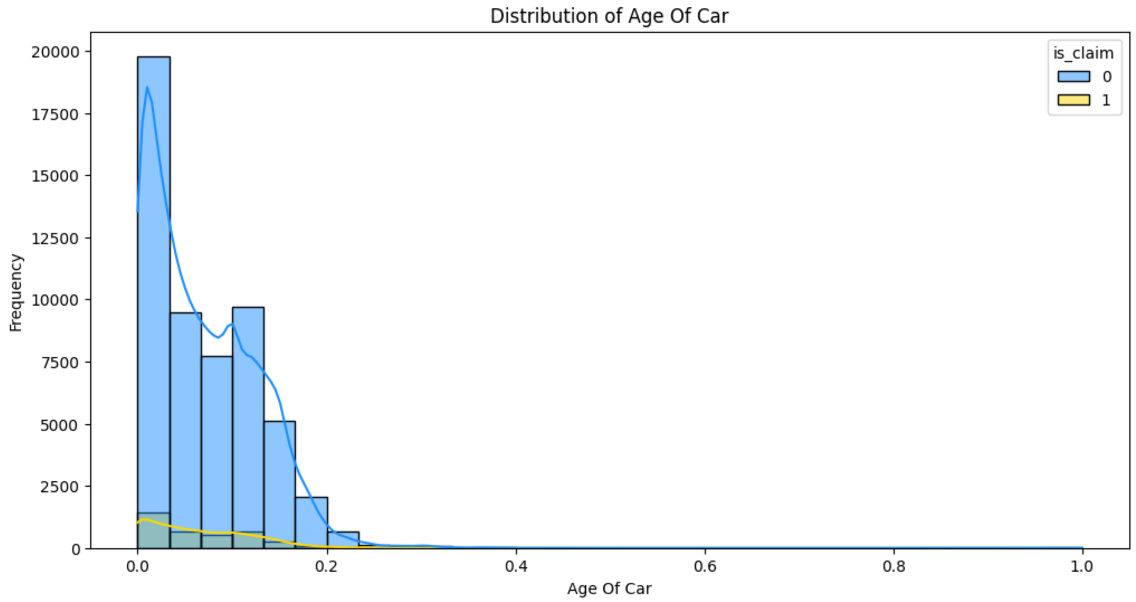


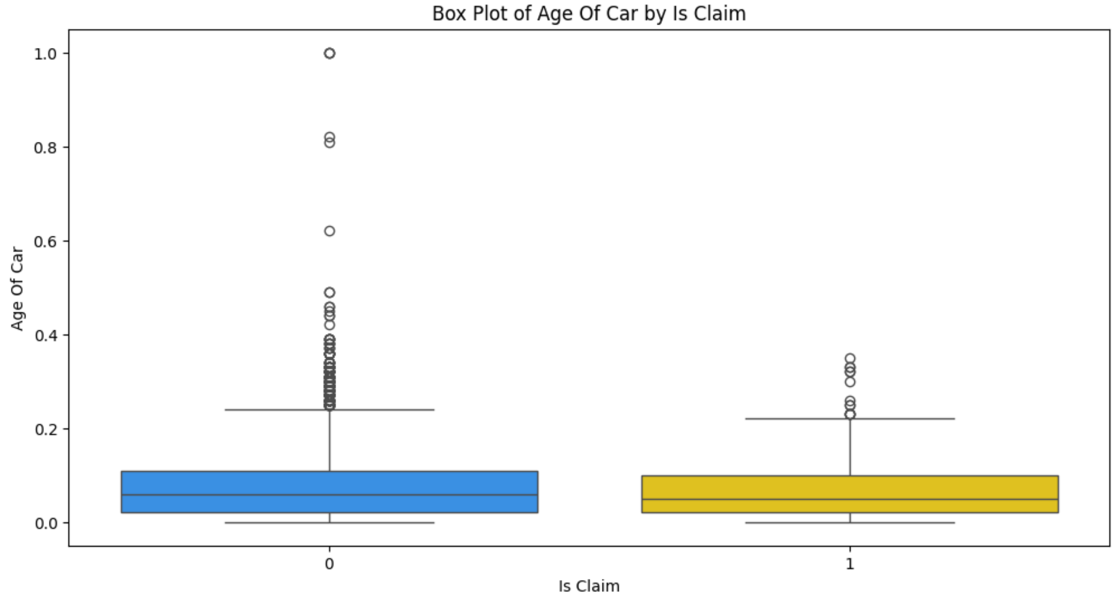


EDA – Distribution Analysis

Unearthing Insights

- The distribution indicates a higher concentration of newer cars in the dataset.
- Claims are spread across different ages of cars, with a noticeable density in newer cars.
- There is no noticeable difference in the median age of cars between the two groups.

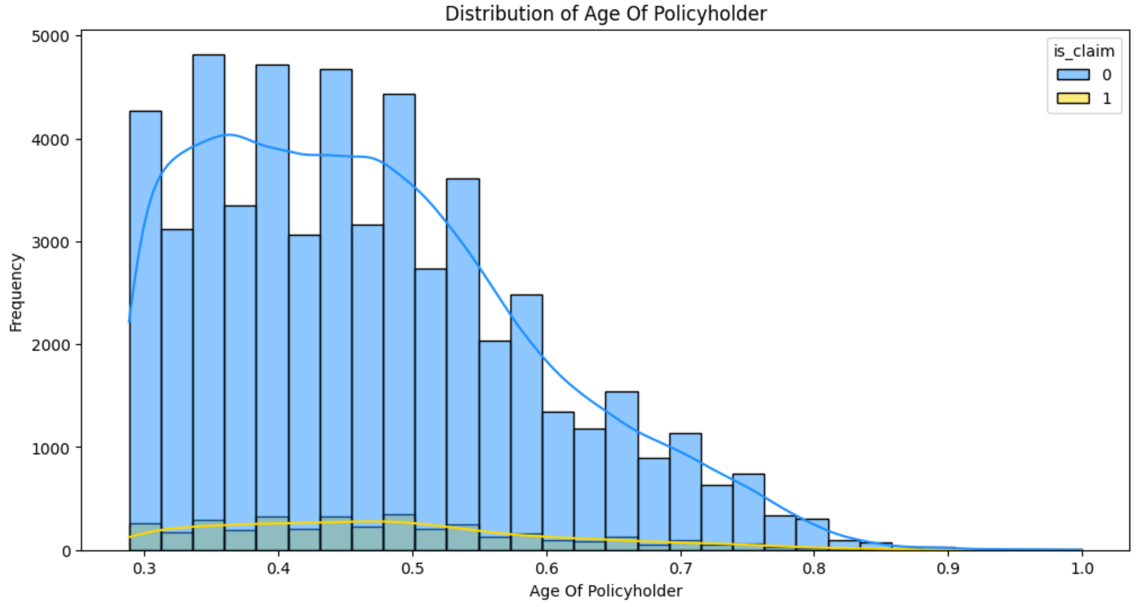


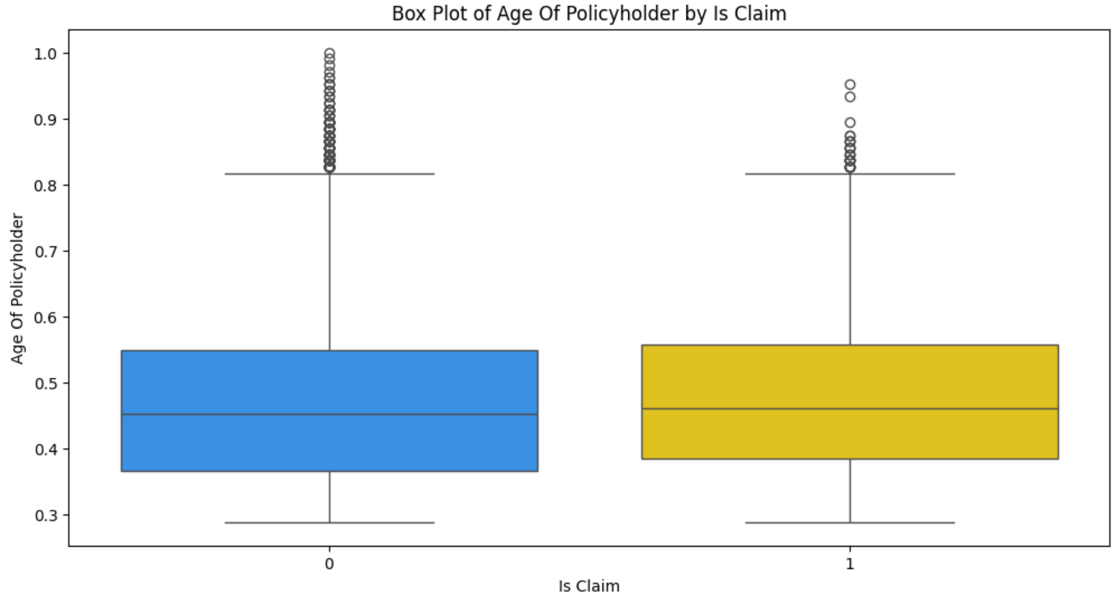


EDA – Distribution Analysis

Unearthing Insights

- The distribution shows that the age of policyholders is fairly uniform across a certain range.
- Claims are observed across various age groups of policyholders, indicating no strong bias towards a particular age group.
- The age of policyholders is similarly distributed for both claim and no-claim groups.
- The median age of policyholders does not show a significant difference between the two groups.





EDA – Distribution Analysis

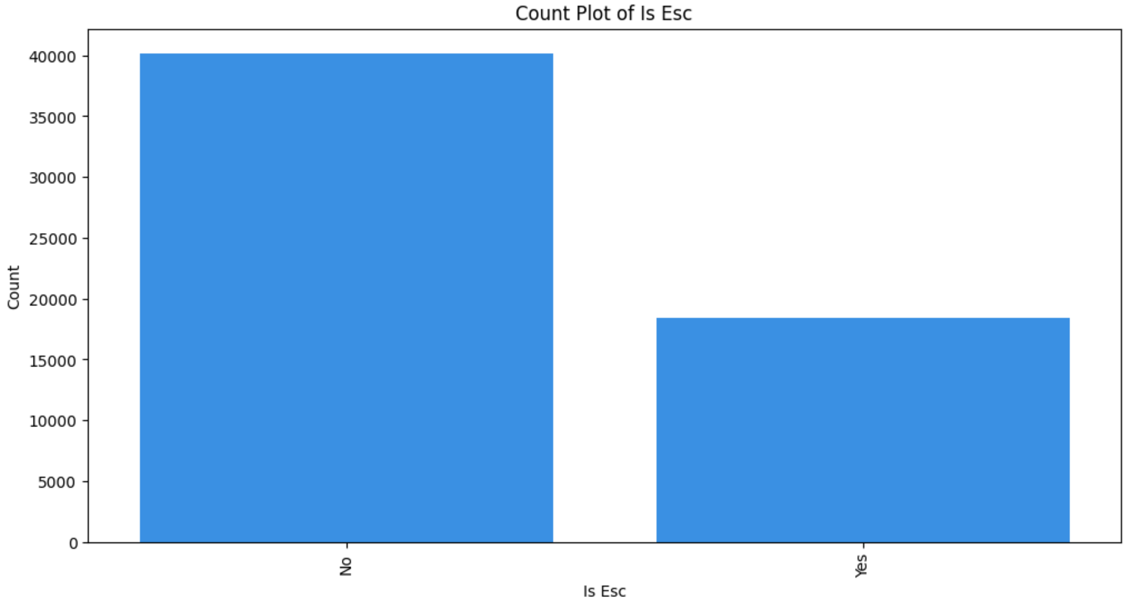
Unearthing Insights

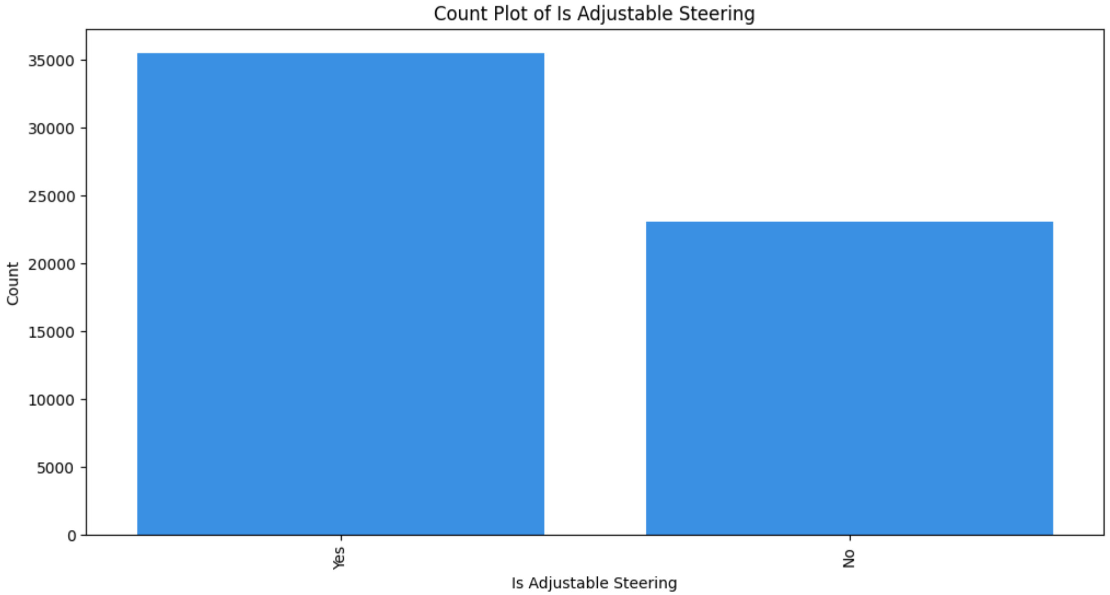
Electronic Stability Control

- About 40% of the cars have ESC
- 60% do not have this feature

Adjustable Steering

- Roughly 55% of the cars have adjustable steering
- 45% do not have this feature





EDA – Distribution Analysis

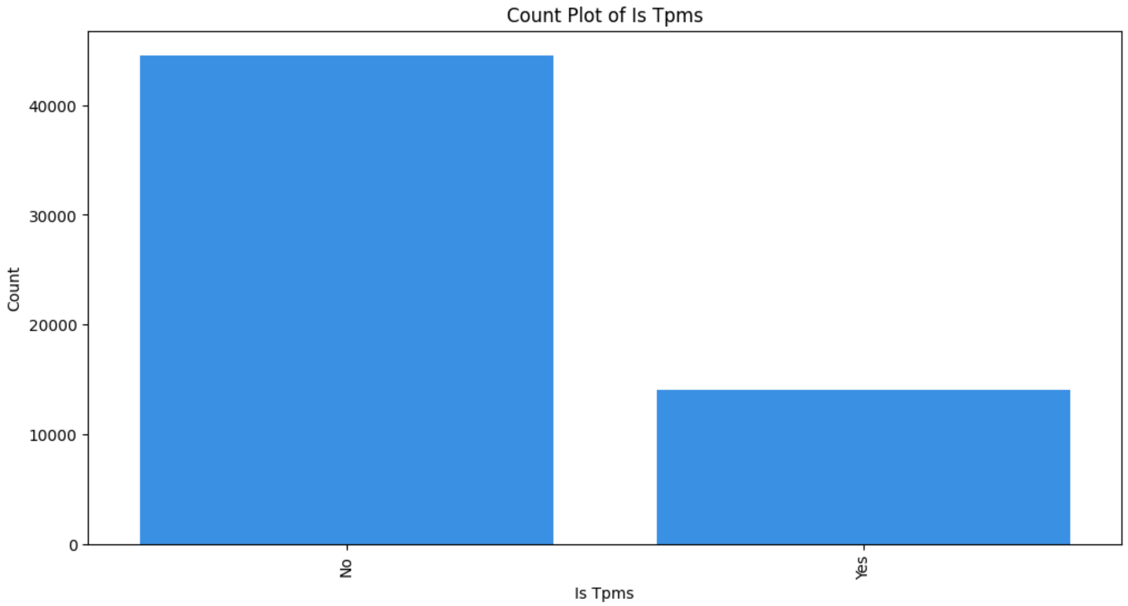
Unearthing Insights

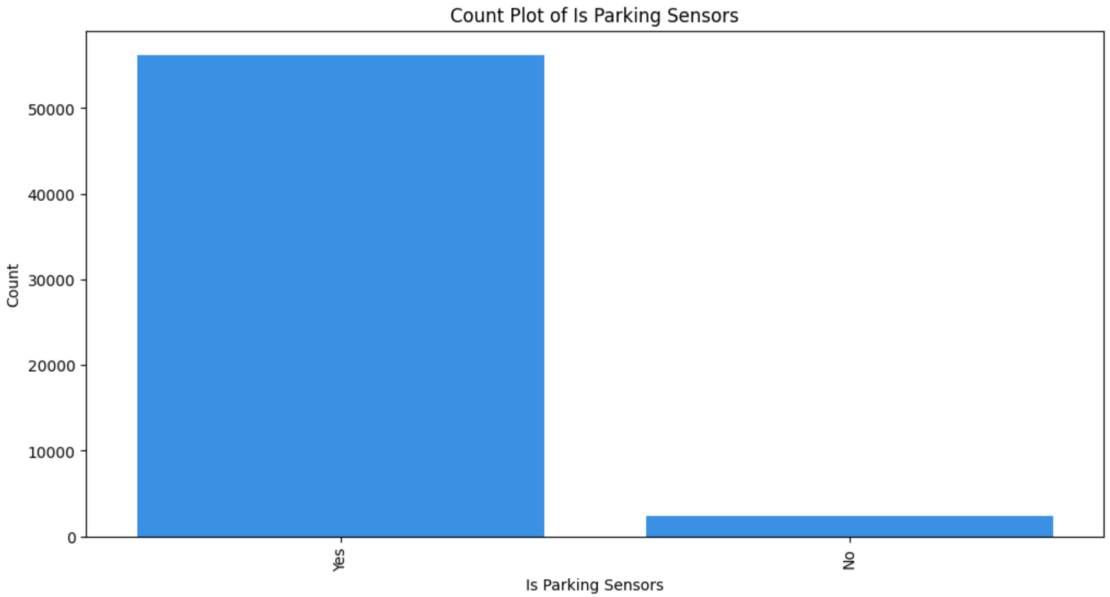
Tire Pressure Monitoring System

- Approximately 60% of the cars are equipped with TPMS
- 40% do not have this feature

Parking Sensors

- Around 55% of the cars have parking sensors
- 45% do not have this feature





EDA – Distribution Analysis

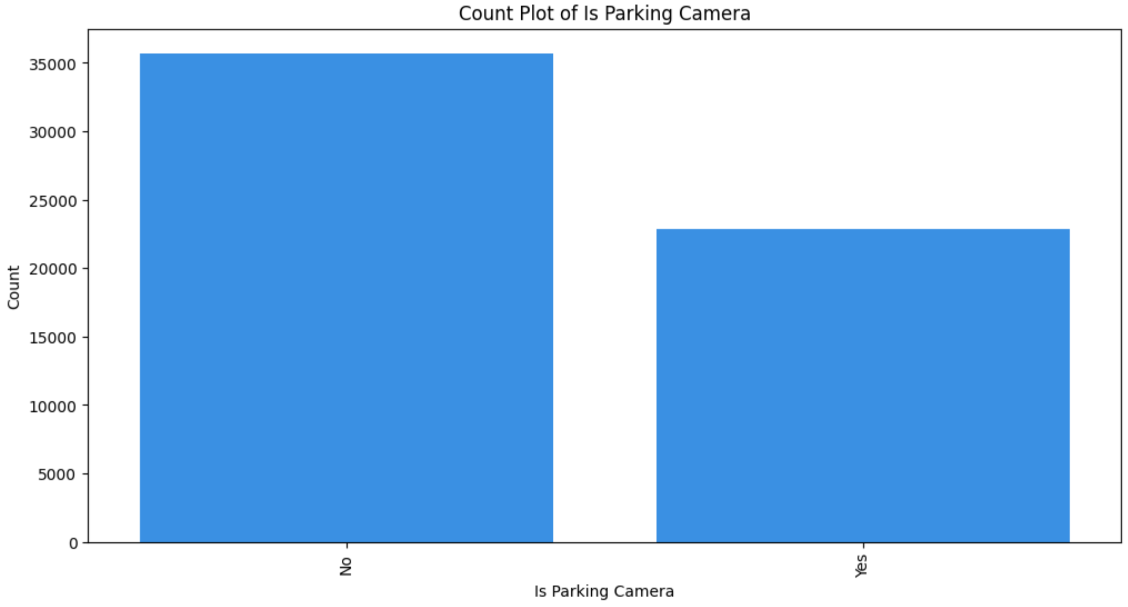
Unearthing Insights

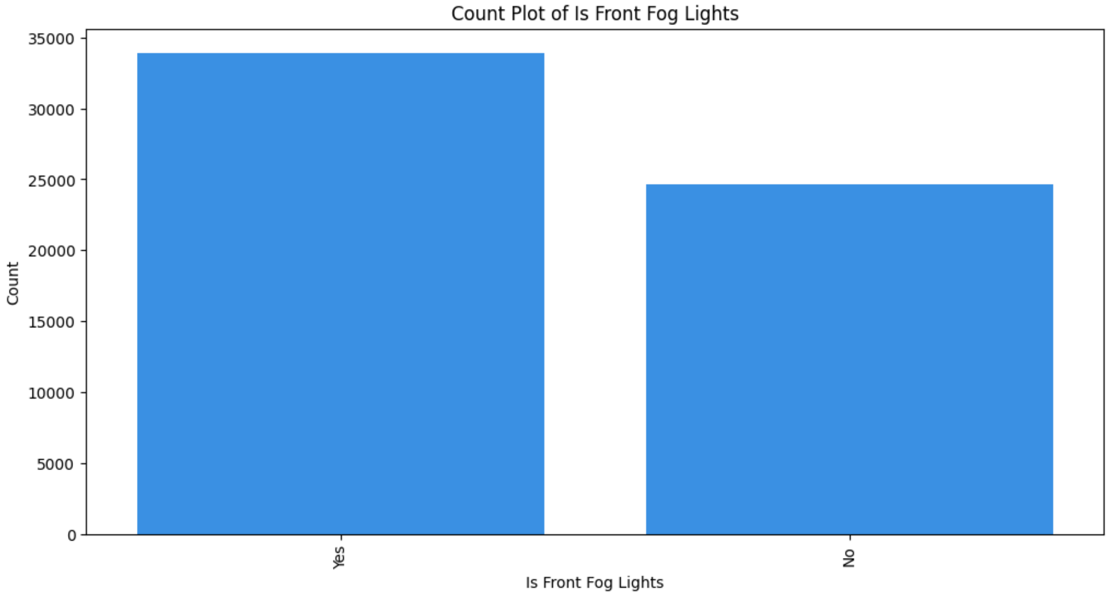
Parking Camera

- A smaller proportion (around 45%) have parking cameras
- The majority (55%) do not have this feature

Front Fog Lights

- About 60% of the cars are equipped with front fog lights
- 40% do not have this feature





EDA – Distribution Analysis

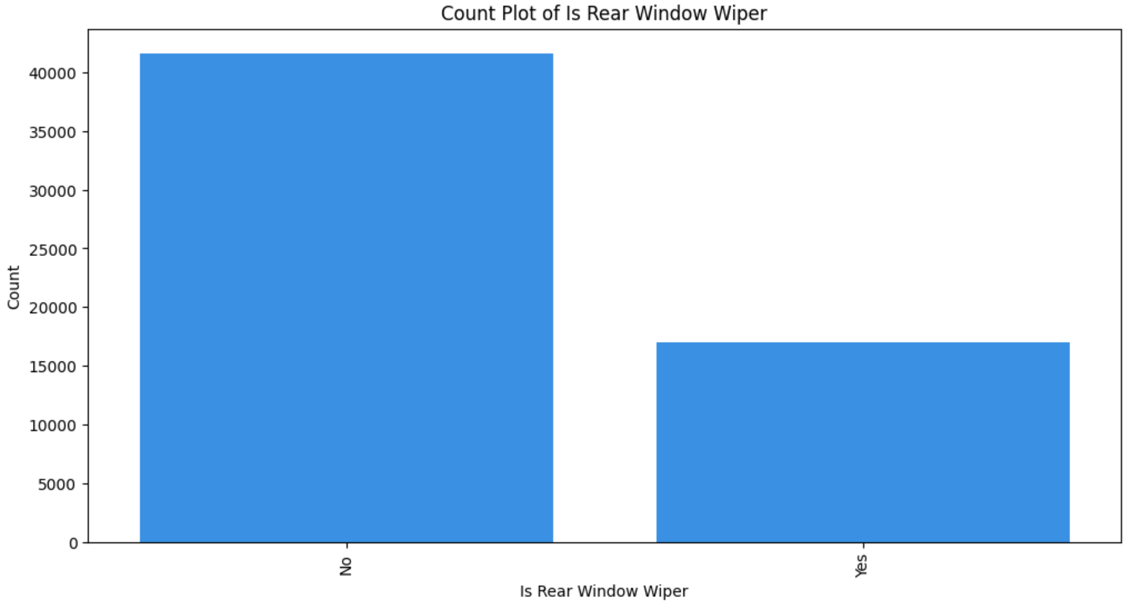
Unearthing Insights

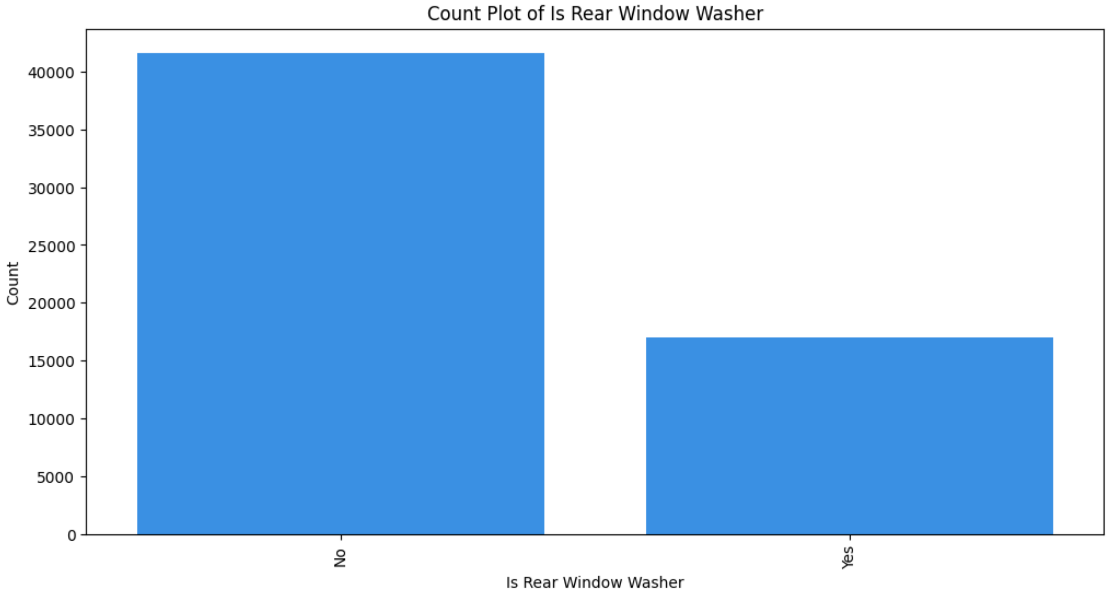
Rear Window Wiper

- Around 50% of the cars have rear window wipers
- 50% do not have this feature

Rear Window Washer

- Similar to rear window wipers, around 50% have rear window washers
- 50% do not have this feature





EDA – Distribution Analysis

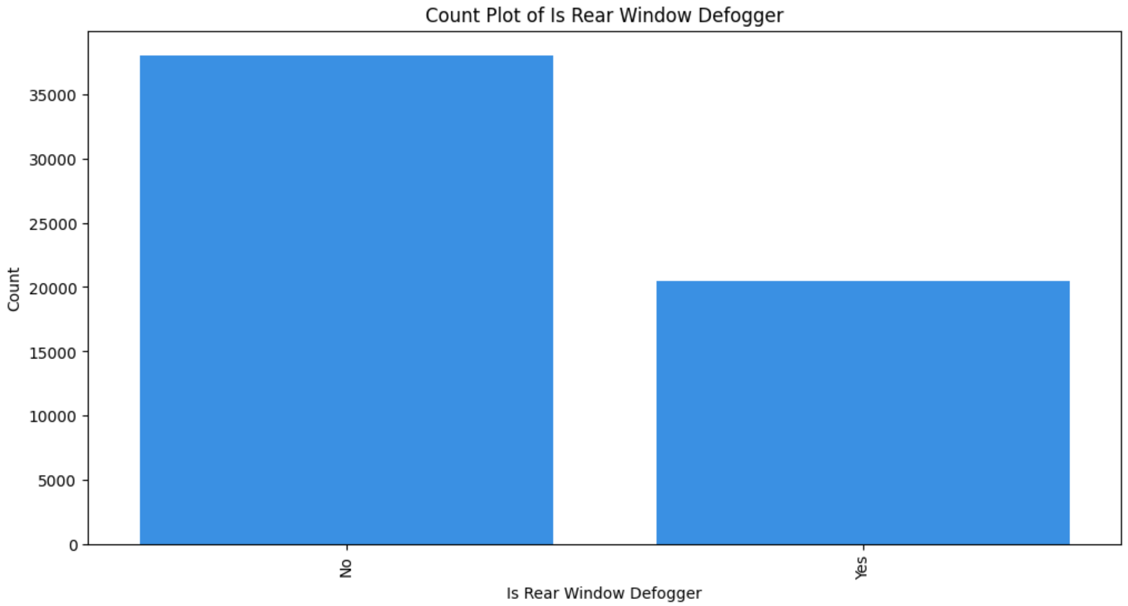
Unearthing Insights

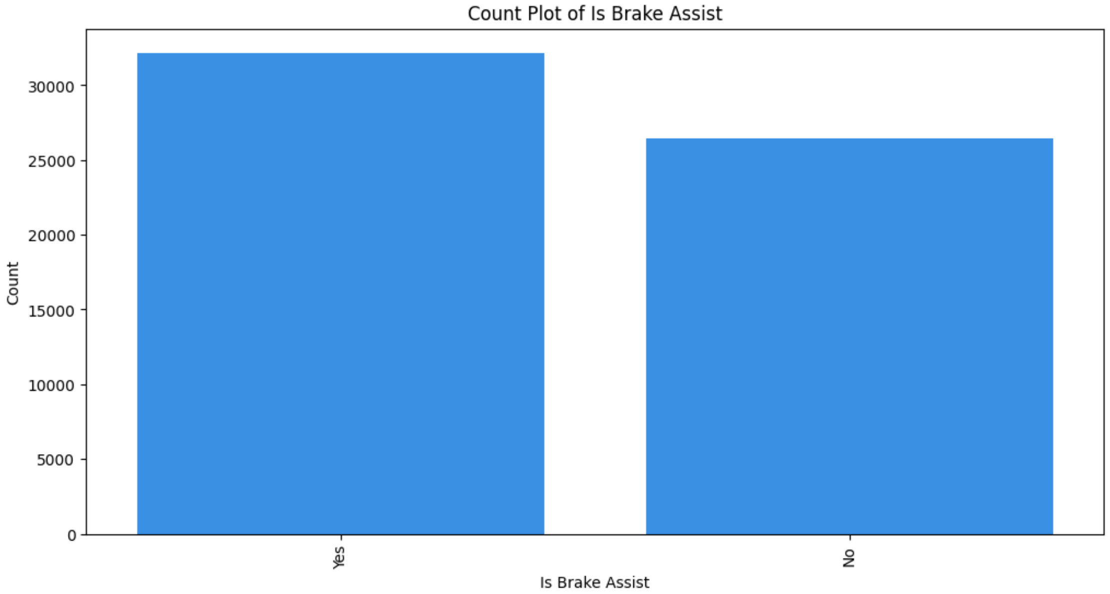
Rear Window Defogger

- A significant majority (around 75%) of the cars are equipped with rear window defoggers
- Only 25% do not have this feature

Brake Assist

- Majority (around 85%) of the cars do not have brake assist
- Only a small proportion (15%) have this feature



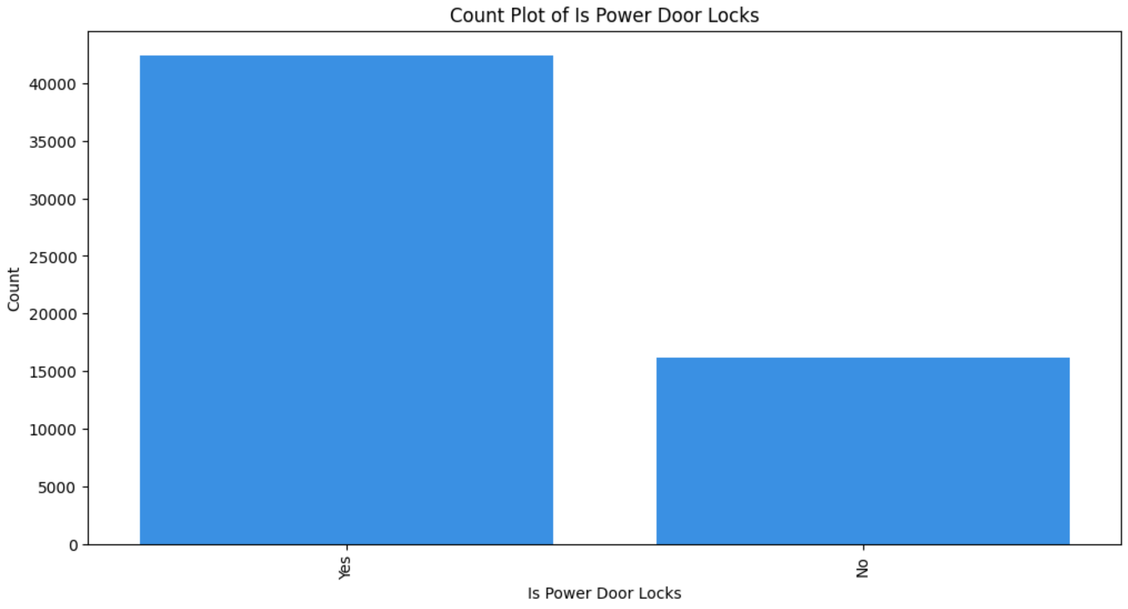


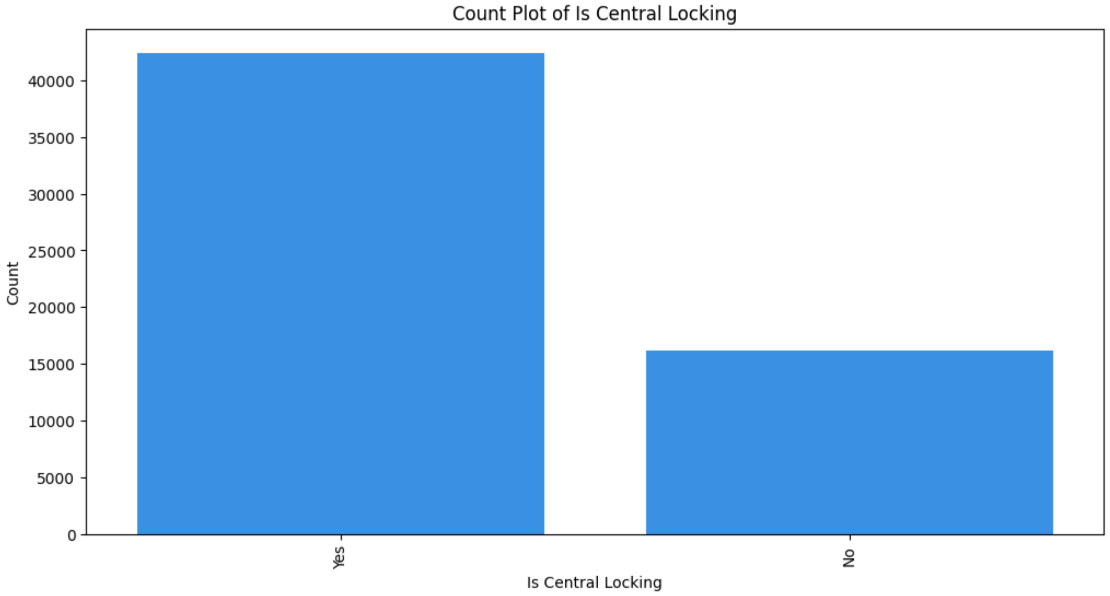
EDA – Distribution Analysis

Unearthing Insights

Power Door Locks
- About 55% of the cars have power door locks
- 45% do not have this feature

Central Locking
- Similar to power door locks, around 55% have central locking
- 45% do not have this feature





EDA – Distribution Analysis

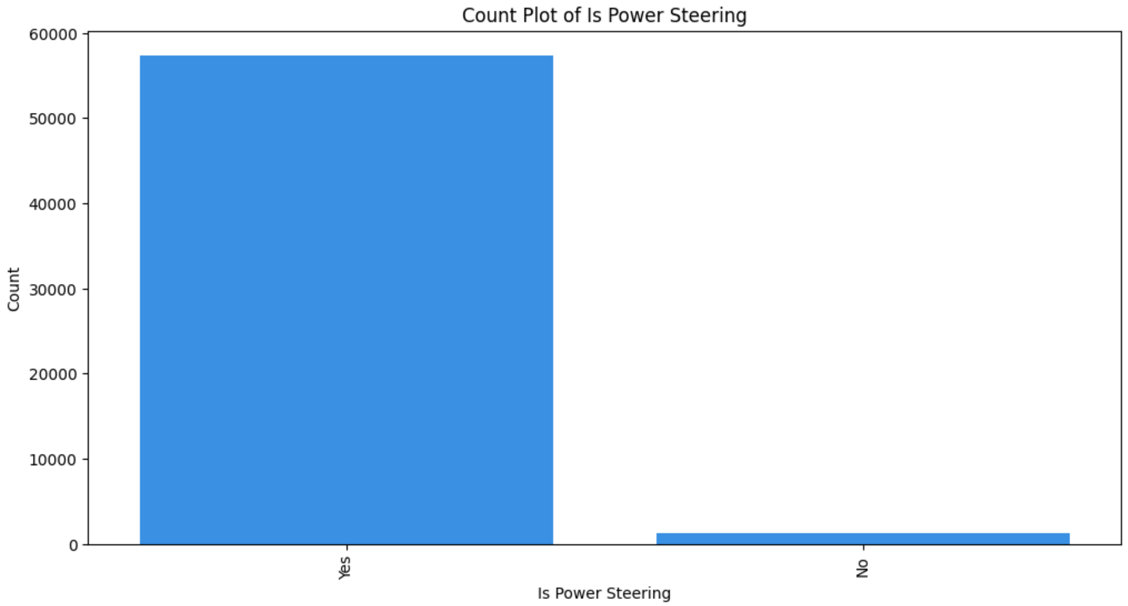
Unearthing Insights

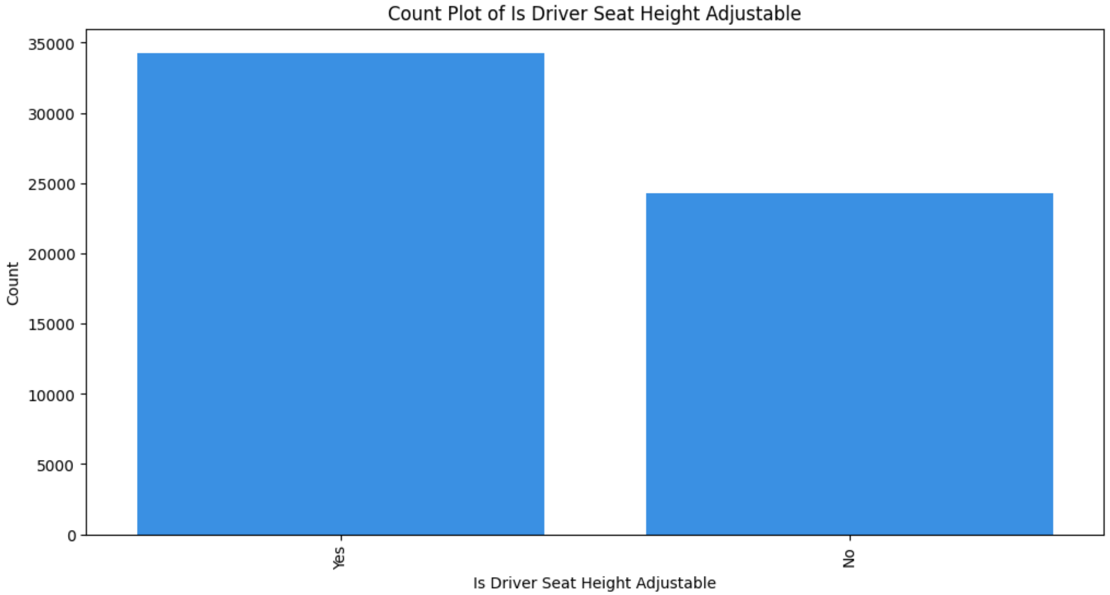
Power Steering

- A significant majority (around 90%) of the cars are equipped with power steering
- Only 10% do not have this feature

Driver Seat Height Adjustable

- A smaller proportion (around 40%) have adjustable driver seat height
- The majority (60%) do not have this feature





EDA – Distribution Analysis

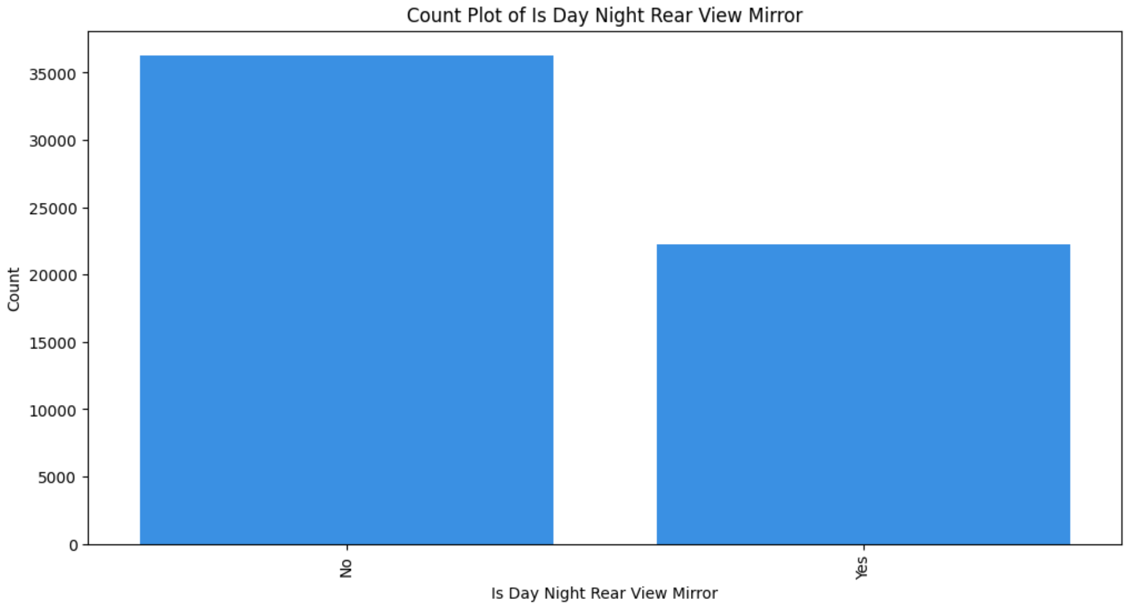
Unearthing Insights

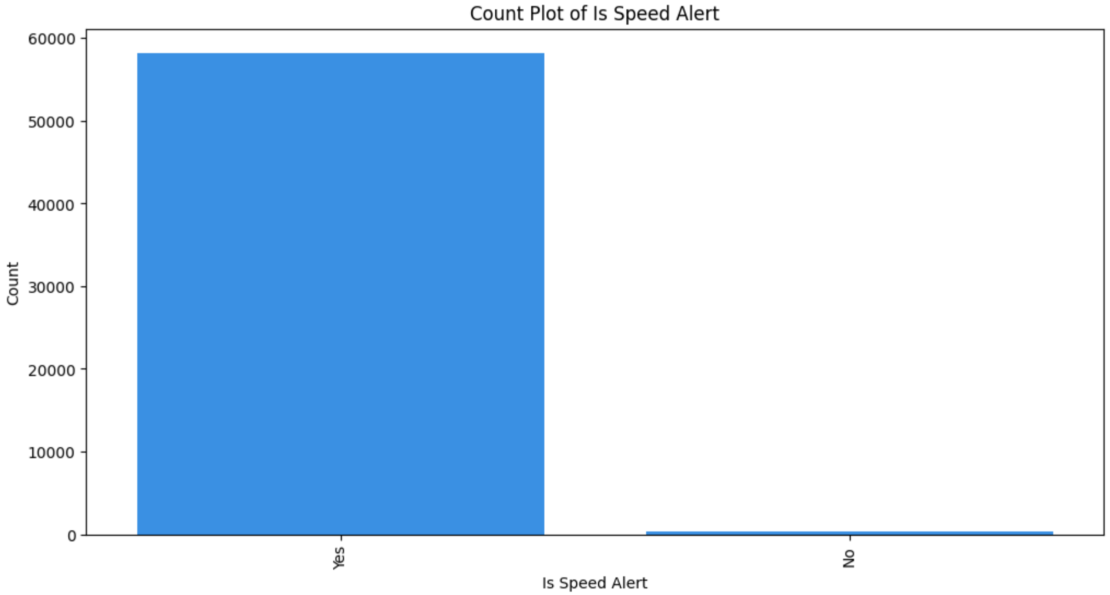
Rear View Mirror

- Around 55% of the cars have day-night rear view mirrors
- 45% do not have this feature

Speed Alert

- An overwhelming majority (95%) of the cars have speed alert systems
- Only 5% do not have this feature





EDA – Distribution Analysis

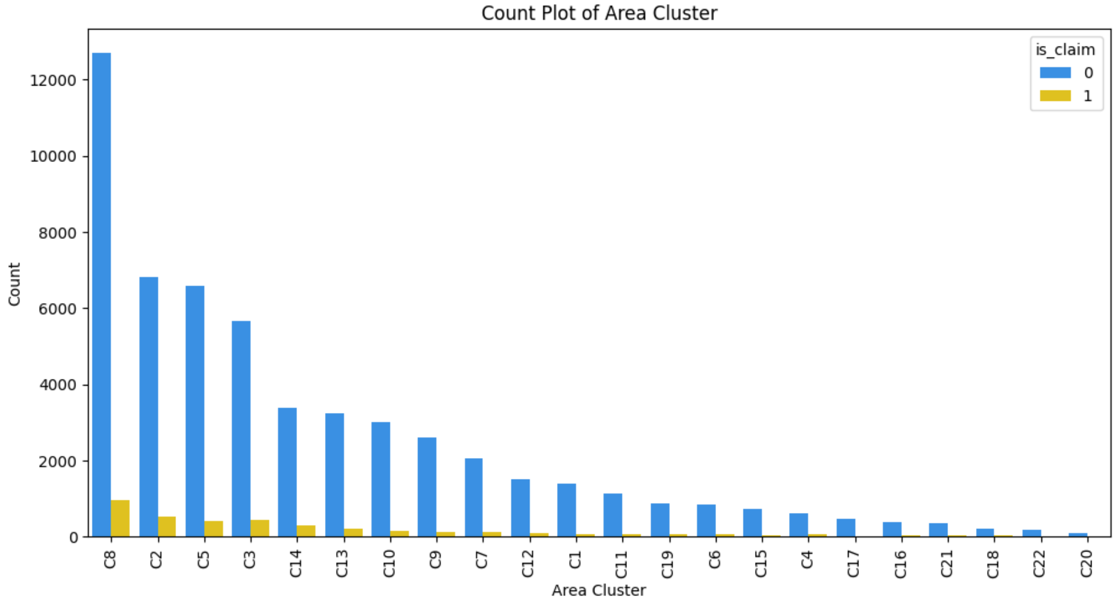
Unearthing Insights

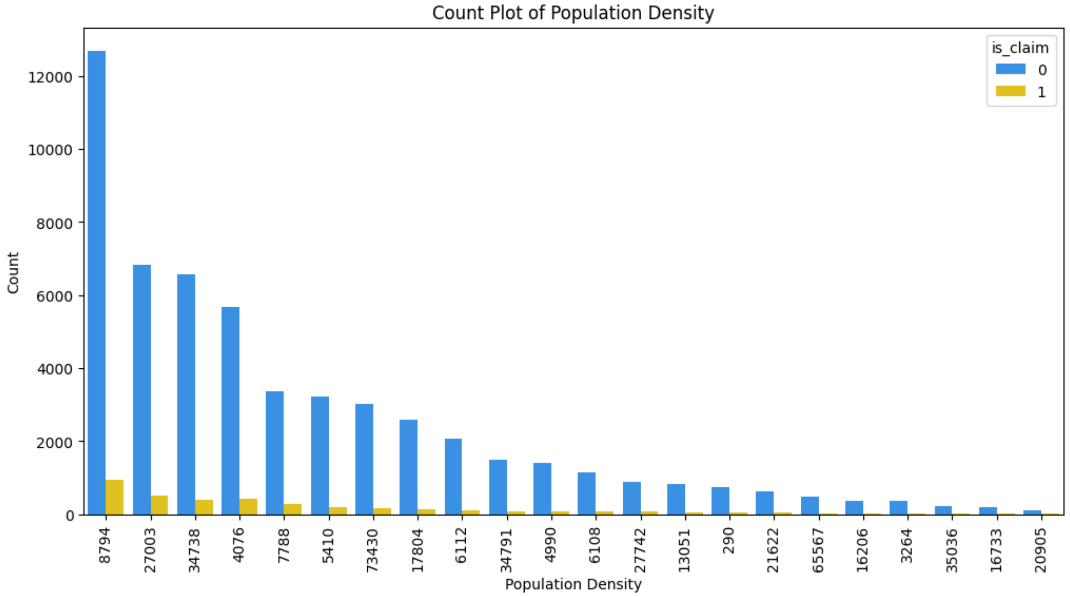
Area Cluster

- Claims occur across various clusters, but some clusters (e.g., C2, C5) show higher claims frequency

Population Density

- Higher population density areas tend to have more claims, indicating a possible correlation between density and claim likelihood





EDA – Distribution Analysis

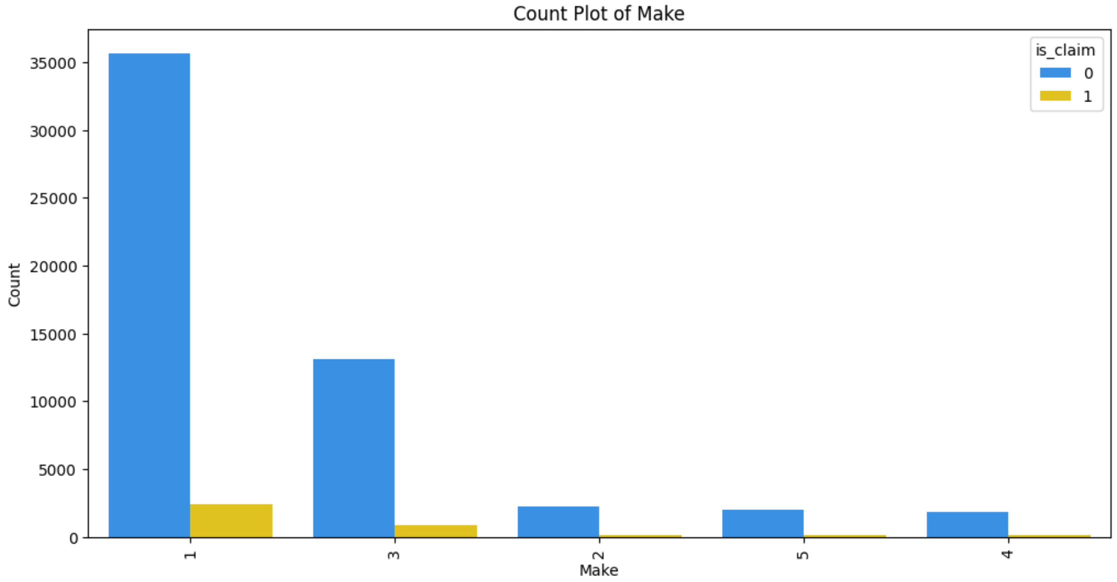
Unearthing Insights

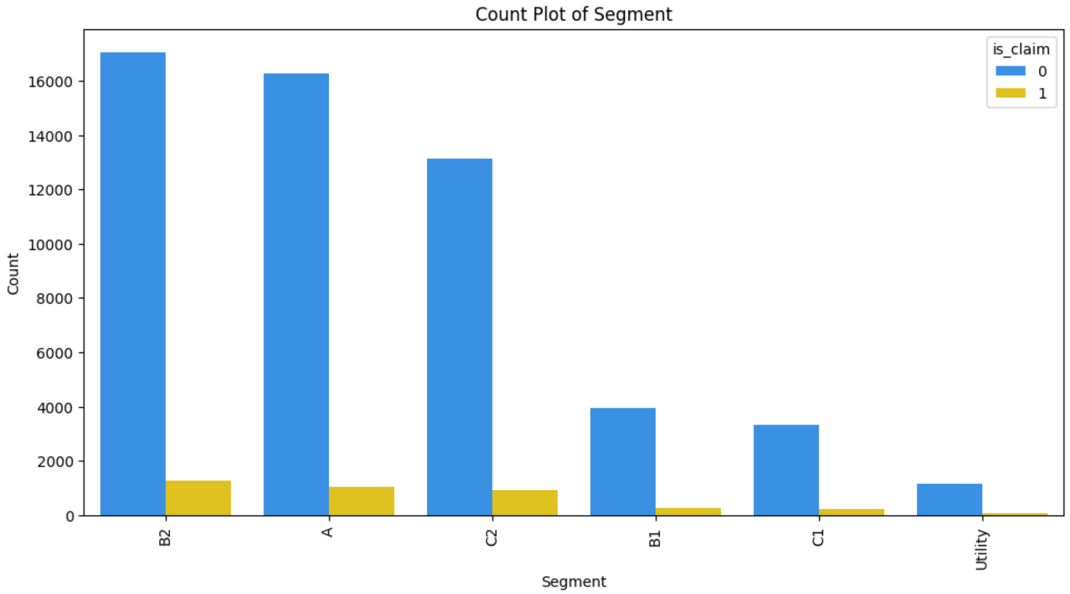
Make

- Some makes (e.g., 1, 2) have higher claims, suggesting potential reliability issues or higher exposure

Segment

- Certain segments (e.g., A, B2) have a higher number of claims, indicating different risk profiles for different car segments



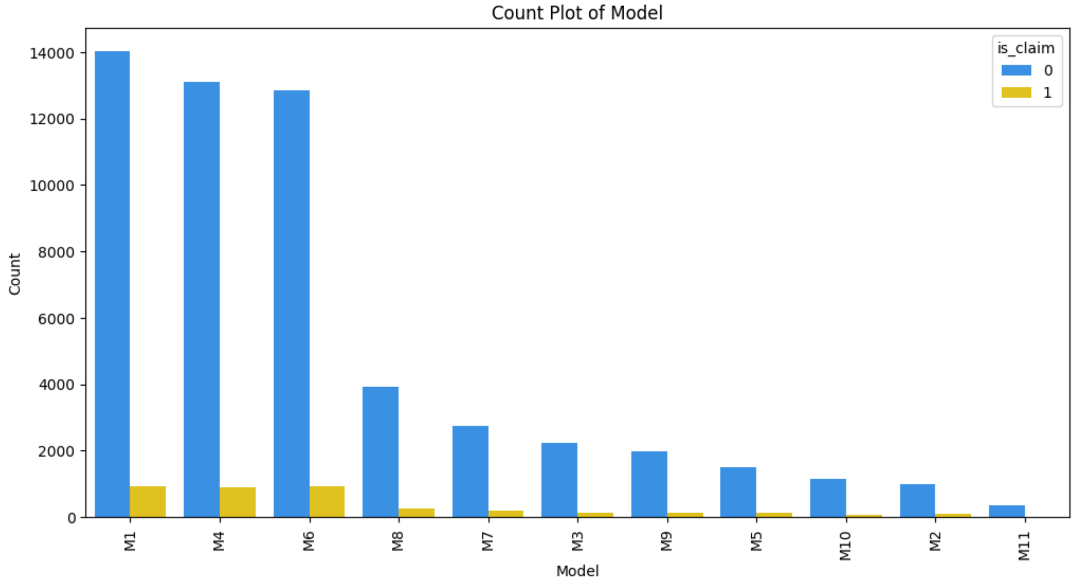


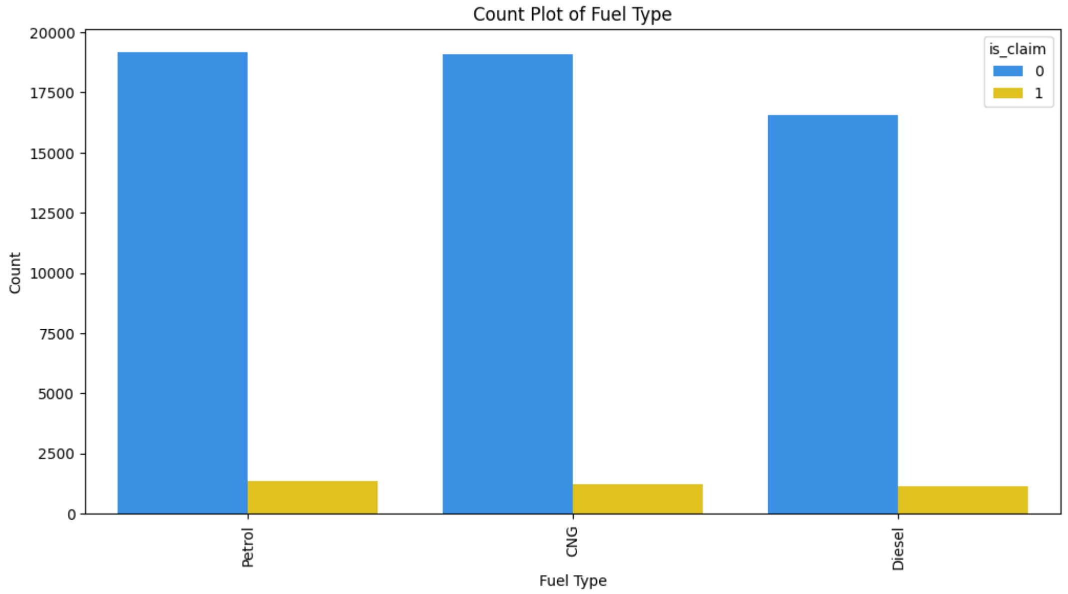
EDA – Distribution Analysis

Unearthing Insights

Model
- Some models show higher claims, highlighting potential model-specific risk factors

Fuel Type
- Cars with petrol engines have a higher number of claims compared to diesel and CNG



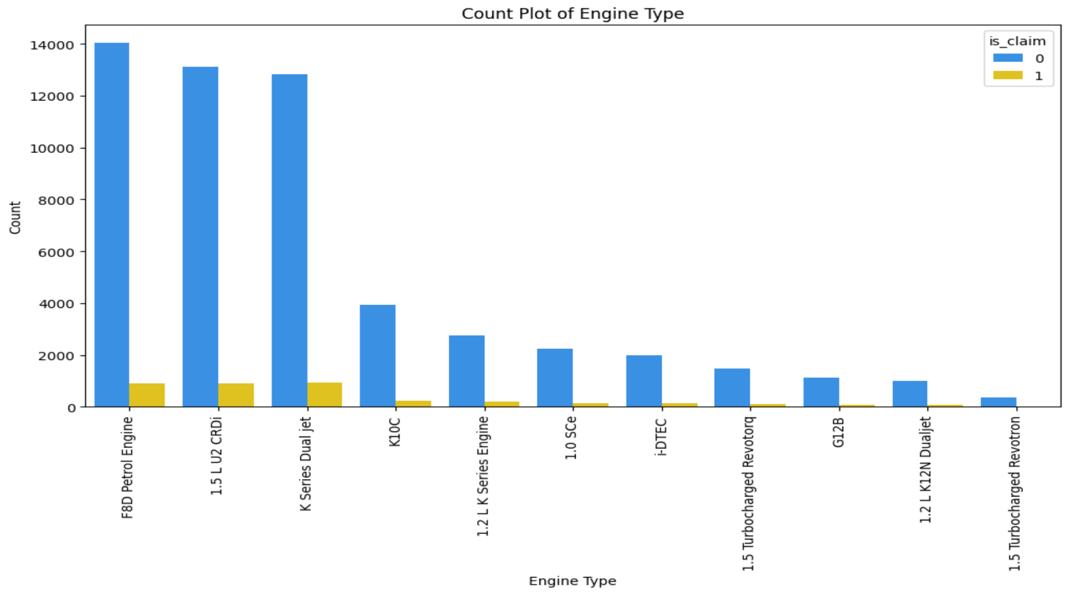


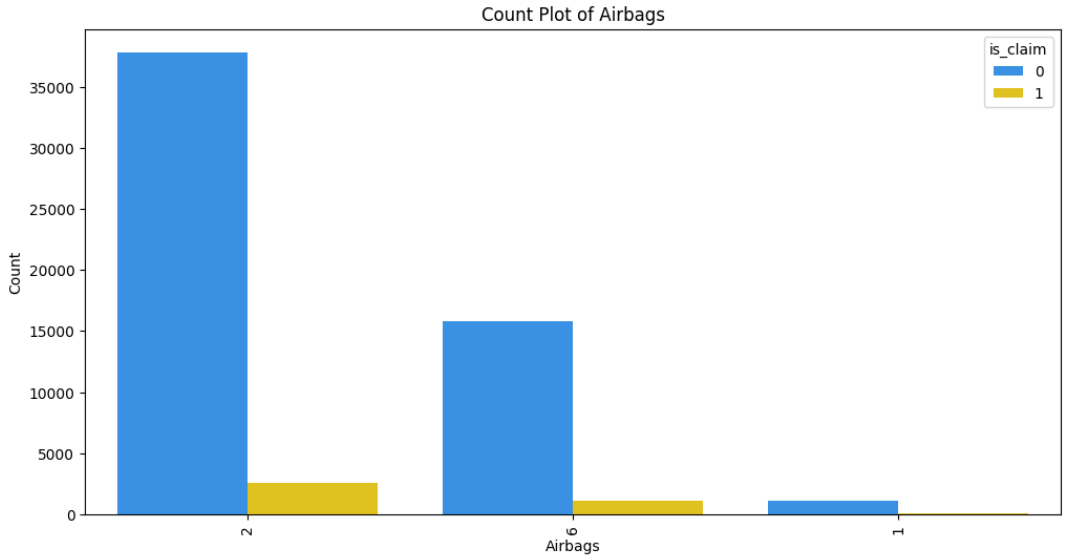
EDA – Distribution Analysis

Unearthing Insights

Engine Type
- Distribution varies with some engine types showing more claims

Airbags
- Higher number of claims for cars with fewer airbags, indicating higher risk for less equipped cars





EDA – Distribution Analysis

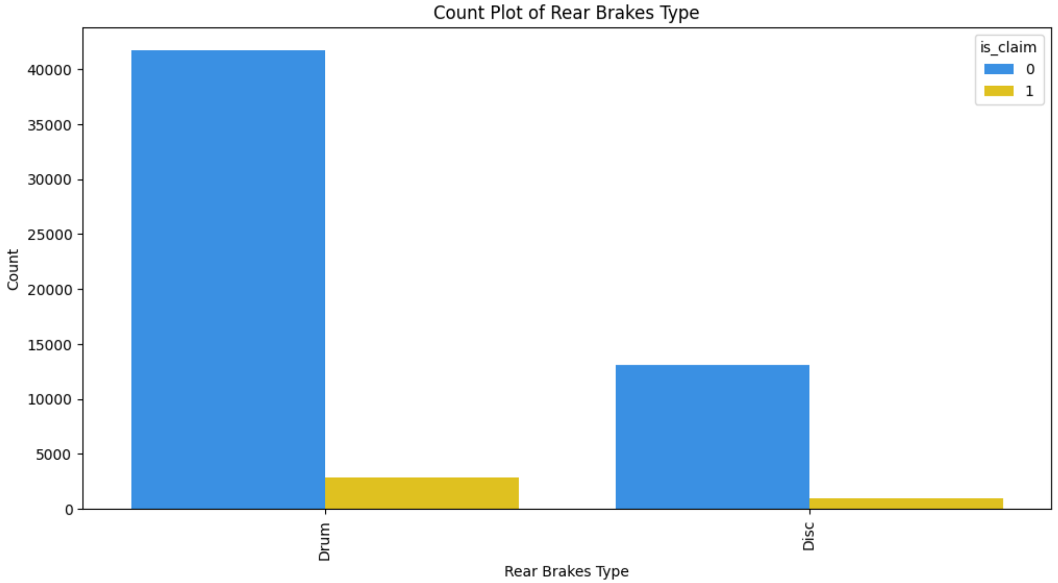
Unearthing Insights

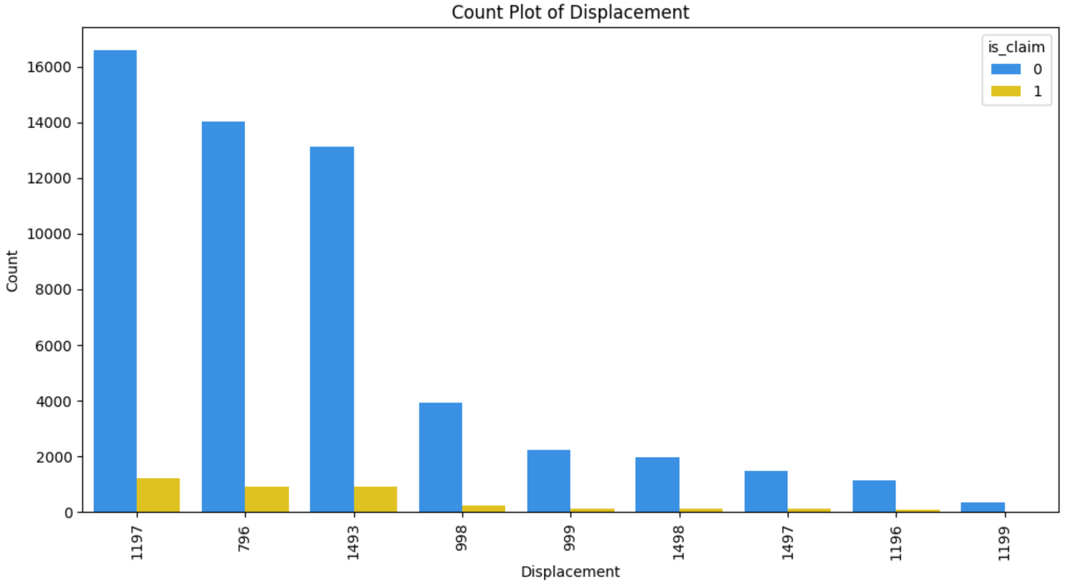
Rear Brakes Type

- Distribution varies, with some brake types showing higher claims

Displacement

- Higher displacement cars tend to have more claims





EDA – Distribution Analysis

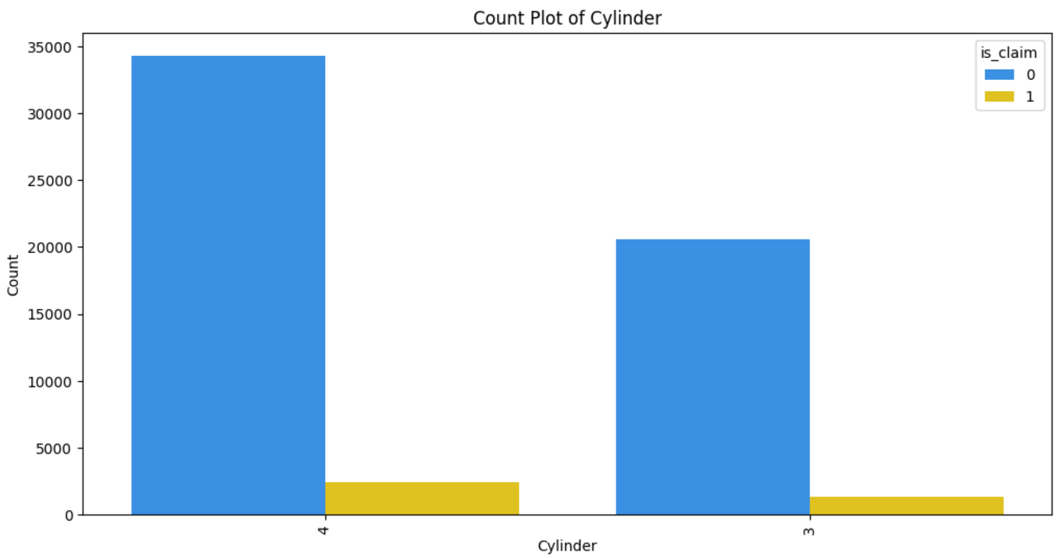
Unearthing Insights

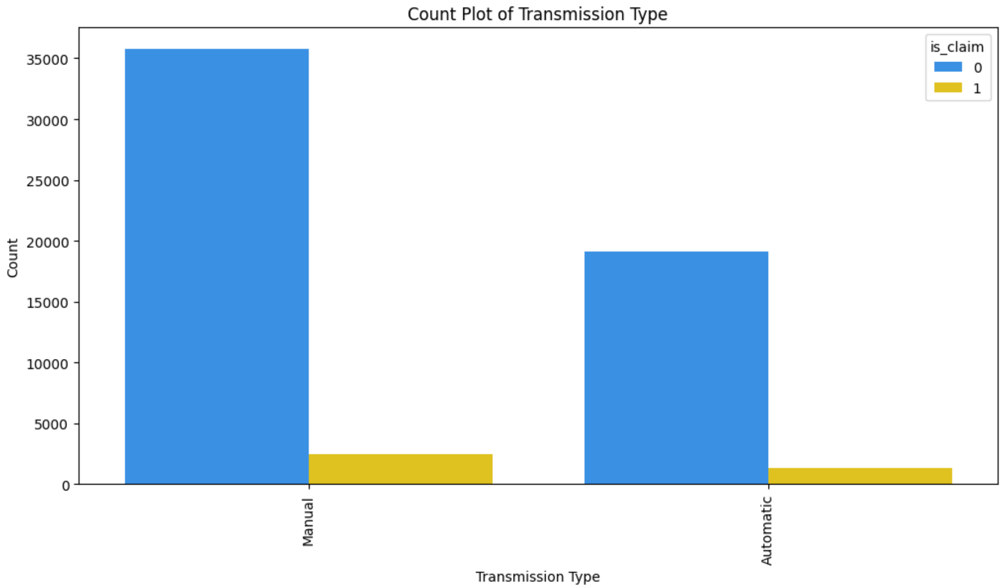
Cylinder

- Distribution varies with different cylinder configurations

Transmission Type

- Manual transmission cars have a higher number of claims compared to automatic



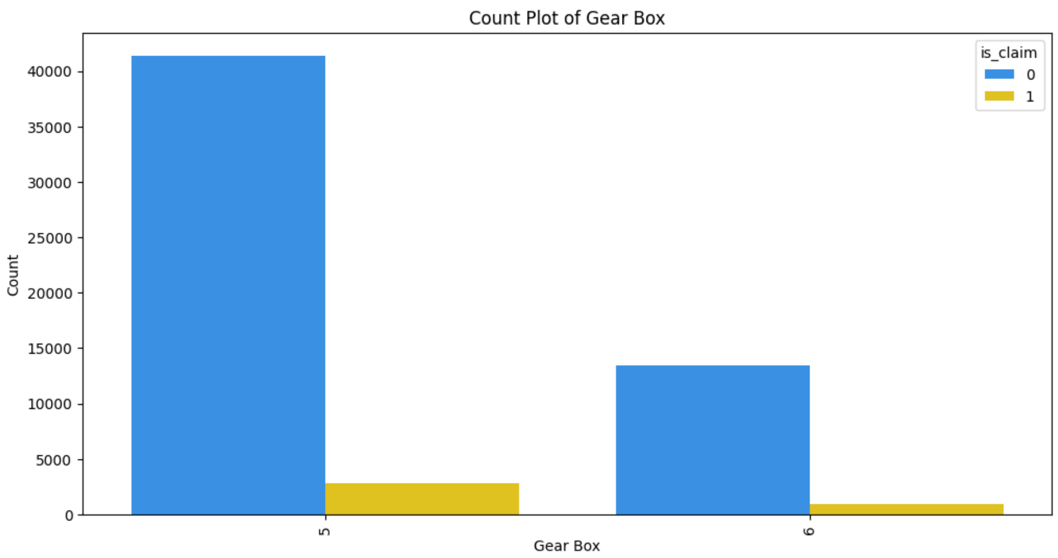


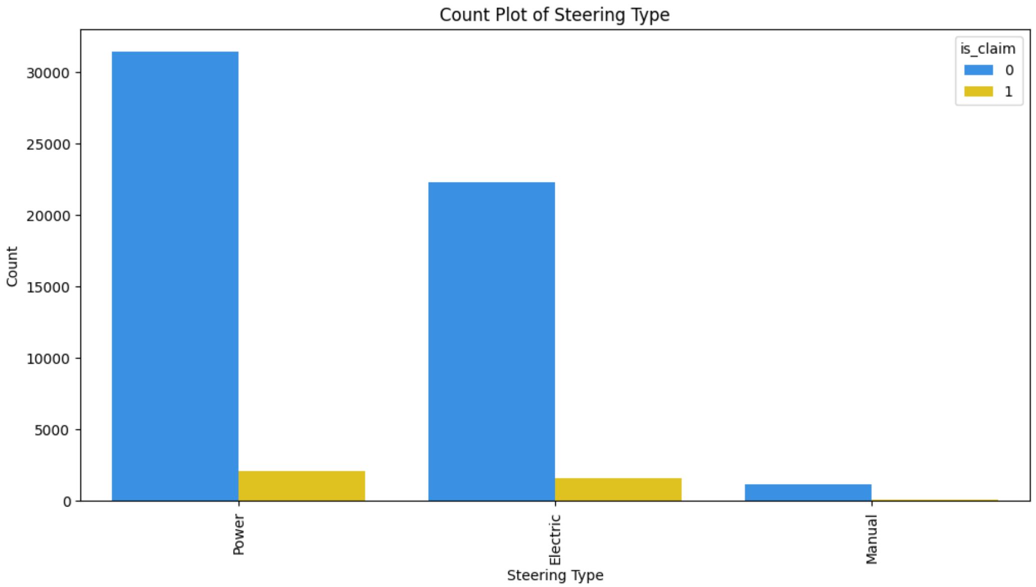
EDA – Distribution Analysis

Unearthing Insights

Gear Box
- Distribution across different gearboxes varies
- Some gearboxes show higher claims

Steering Type
- Distribution varies with different steering types





EDA – Distribution Analysis

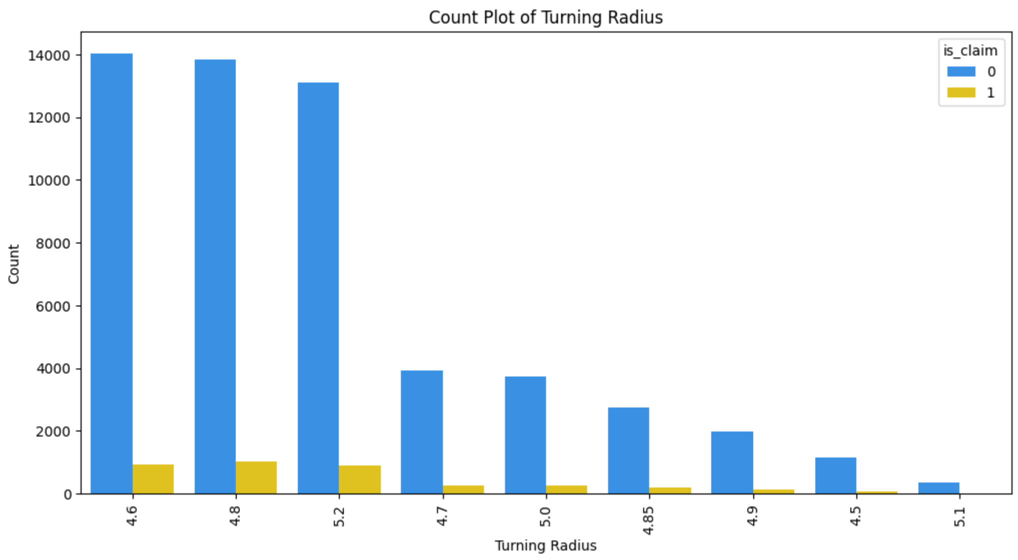
Unearthing Insights

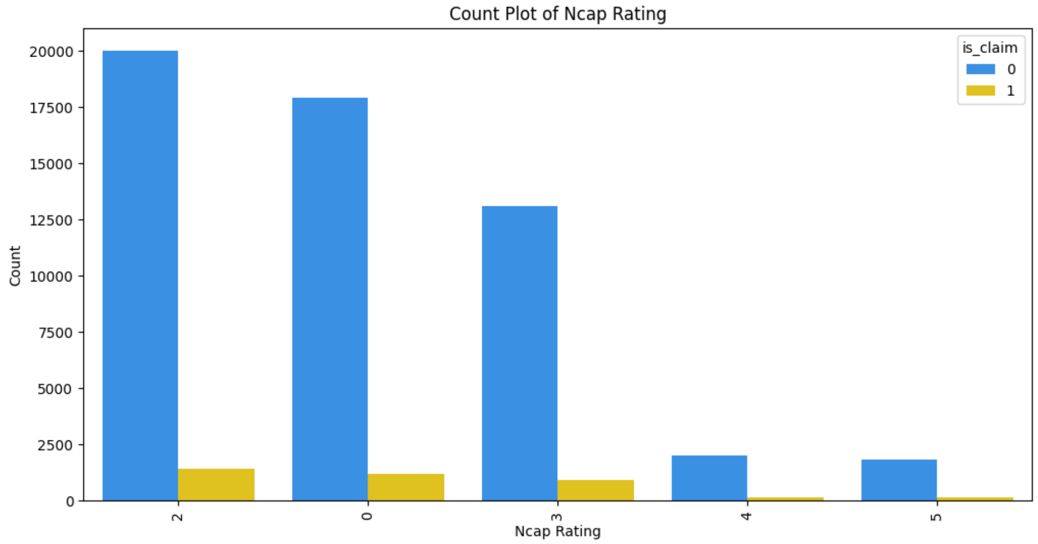
Turning Radius

- Distribution across different turning radii varies

NCAP Rating

- Cars with lower NCAP ratings tend to have higher claims, indicating a correlation between safety ratings and claim likelihood





EDA – Correlation Analysis

Unearthing Insights

Strong Correlations

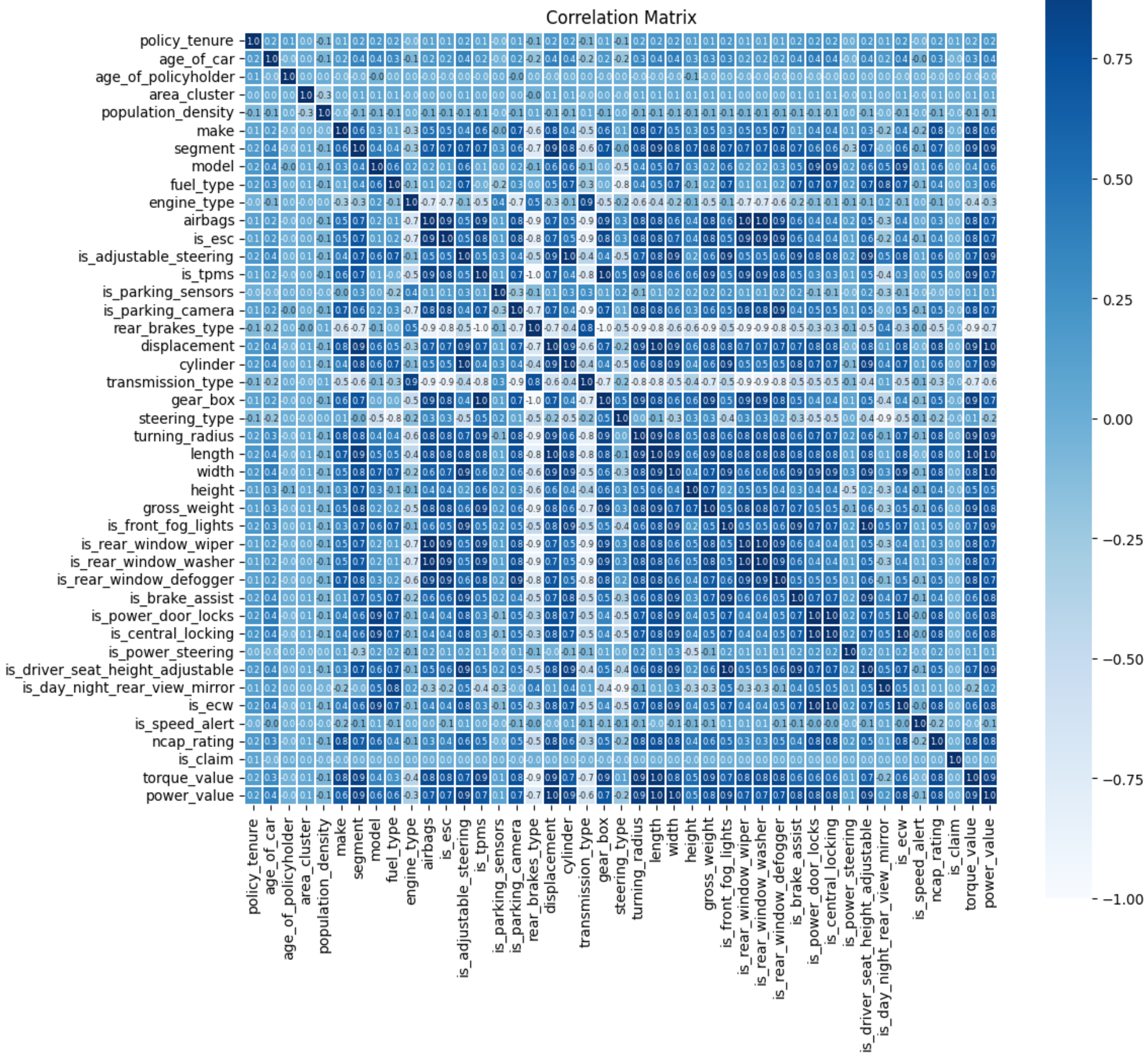
- displacement and power_value: High positive correlation (0.85), indicating that cars with larger engine displacements tend to have higher power values
- segment and make: Strong positive correlation (0.44), suggesting a relationship between the car segment and the manufacturer

Moderate Correlations

- ncap_rating and is_ecw: Moderate correlation (0.78), indicating that cars with higher safety ratings tend to have electronic stability control (ESC)
- width and length: Moderate positive correlation (0.77), suggesting that larger cars in terms of width also tend to be longer

Weak Correlations with Target Variable (is_claim)

- Most features show weak correlations with is_claim, indicating no single feature strongly predicts the likelihood of a claim
- policy_tenure and is_claim: Weak correlation (0.01), suggesting tenure has little impact on claim likelihood
- age_of_car and is_claim: Weak negative correlation (-0.02), indicating that older cars have a slightly lower likelihood of claims



Identified Unique Models with Features

Unearthing Insights

model	make	segment	fuel_type	max_torque	max_power	engine_type	airbags	is_esc	is_tpms	rear_brakes_type	displacement	cylinder	transmission_type	gear_box	steering_type	ncap_rating
M1	1	A	CNG	60Nm@3500rpm	40.36bhp@6000rpm	F8D Petrol Engine	2	No	No	Drum	796	3	Manual	5	Power	0
M10	1	Utility	CNG	85Nm@3000rpm	61.68bhp@6000rpm	G12B	1	No	No	Drum	1196	4	Manual	5	Manual	0
M11	4	C1	Petrol	170Nm@4000rpm	118.36bhp@5500rpm	1.5 Turbocharged Revotron	2	Yes	No	Drum	1199	3	Manual	6	Power	5
M2	1	C1	Petrol	113Nm@4400rpm	88.50bhp@6000rpm	1.2 L K12N Dualjet	2	Yes	No	Drum	1197	4	Automatic	5	Electric	2
M3	2	A	Petrol	91Nm@4250rpm	67.06bhp@5500rpm	1.0 SCe	2	No	No	Drum	999	3	Automatic	5	Electric	2
M4	3	C2	Diesel	250Nm@2750rpm	113.45bhp@4000rpm	1.5 L U2 CRDi	6	Yes	Yes	Disc	1493	4	Automatic	6	Power	3
M5	4	B2	Diesel	200Nm@3000rpm	88.77bhp@4000rpm	1.5 Turbocharged Revotorq	2	No	No	Drum	1497	4	Manual	5	Electric	5
M6	1	B2	Petrol	113Nm@4400rpm	88.50bhp@6000rpm	K Series Dual jet	2	No	No	Drum	1197	4	Manual	5	Electric	2
M7	1	B2	Petrol	113Nm@4400rpm	88.50bhp@6000rpm	1.2 L K Series	6	Yes	No	Drum	1197	4	Automatic	5	Electric	0
M8	1	B1	CNG	82.1Nm@3400rpm	55.92bhp@5300rpm	K10C	2	No	No	Drum	998	3	Manual	5	Power	2
M9	5	C1	Diesel	200Nm@1750rpm	97.89bhp@3600rpm	i-DTEC	2	No	No	Drum	1498	4	Manual	5	Electric	4

Preprocessing & Baseline Modelling

Baseline Modeling with Logistic Regression

Transparency

- Interpretability: Examine coefficients to understand feature impact
- Coefficient Representation: Log odds show feature influence on outcomes
- Clear Results: Provides interpretable results, unlike complex models
- Trust and Decision-Making: Transparency builds trust and aids decisions

Efficiency

- Less Intensive: Requires fewer resources than complex models
- Faster Analysis: Quick data analysis and model training
- Iterative Improvement: Enables rapid iteration and model selection

Handling Outliers

- Sensitive to Outliers: Can lead to overfitting/underfitting
- Boundary Influence: Skews decision boundaries, affecting probabilities

Mitigating Outliers

- Data Preprocessing: Detect and manage outliers
- Penalized Models: Use L1/L2 penalties to reduce outlier impact
- Cross-Validation: Ensures robust model performance evaluation

Conclusion

- Choice: Logistic Regression is transparent, efficient, and interpretable
- Outliers: Mitigated through preprocessing and regularization techniques

Metric	0 (Majority)	1 (Minority)
Accuracy		0.58
Precision	0.95	0.08
Recall	0.58	0.53
F1 Score	0.72	0.14
F2 Score		0.2475

Model Building and Evaluation

Scaling:

- To ensure uniformity in feature scales, we applied StandardScaler to the features. This preprocessing step is vital for algorithms sensitive to feature magnitudes.

Handling Imbalanced Data:

- Given the class imbalance in our dataset, we applied the Synthetic Minority Over-sampling Technique (SMOTE) to the training data. SMOTE generates synthetic samples for the minority class, helping to balance the class distribution and improve the model's ability to learn from minority class examples.

Classifier Initialization:

- We initialized several classifiers to explore a variety of algorithmic approaches for our classification task.

Classifier	Accuracy	Precision (1)	Recall (1)	F1 Score (1)	F2 Score (1)
Decision Tree	0.86	0.09	0.13	0.10	0.1167
Random Forest	0.87	0.09	0.12	0.10	0.1094
Balanced Random Forest	0.86	0.09	0.11	0.10	0.1073
Adaboost	0.65	0.09	0.49	0.15	0.2586
Gradient Boosting	0.79	0.11	0.29	0.15	0.2159
XGBoost	0.92	0.13	0.05	0.08	0.0593
LightGBM	0.93	0.12	0.02	0.03	0.0194
CatBoost	0.93	0.14	0.01	0.02	0.0099
CatBoost (tuned)	0.48	0.08	0.70	0.15	0.2796
Adaboost (tuned)	0.93	0.08	0.01	0.02	0.0128
Gradient Boosting (tuned)	0.93	0.12	0.02	0.04	0.0255
XGBoost (tuned)	0.85	0.10	0.16	0.12	0.1403

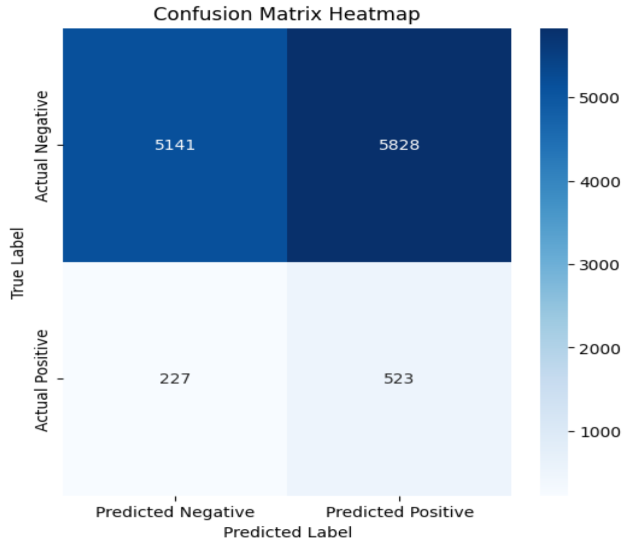
Model Building and Evaluation

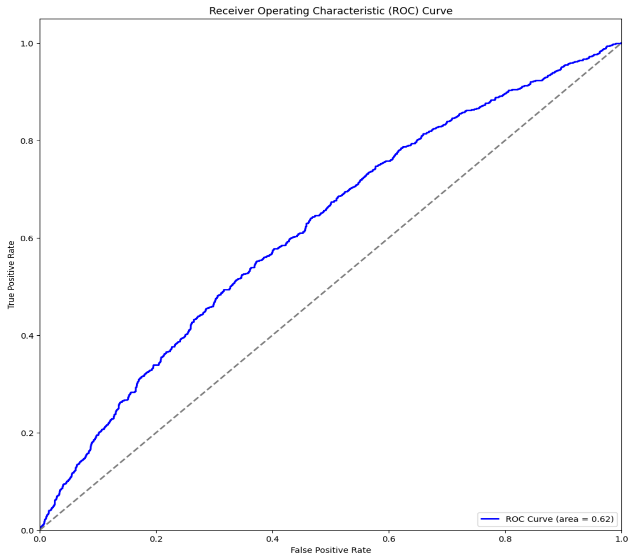
Selected Model (CatBoost)
Based on the evaluation results, CatBoost emerged as the most promising model for our classification task, achieving the highest F1 Score and F2 Score among the classifiers after tuning.

Performance Metrics for Selected Model (CatBoost)
Recall: 0.70
F1 Score: 0.15
F2 Score: 0.2796

Comparison with Baseline
Significant improvement in recall, F1 Score, and F2 Score over baseline models.
CatBoost's tuned version demonstrated superior performance in capturing minority class instances, which is crucial for imbalanced datasets.

Justification
Our choice of CatBoost as the selected model is validated by its superior performance in recall and F2 Score, highlighting its potential to excel in BFSI classification tasks where identifying minority class instances (frauds, defaults, etc.) is critical.





Recommendations

Insurance Interventions

- Based on these findings, we suggest targeted interventions to improve claim management and customer satisfaction
- Insurance companies can use this information to tailor policies, enhance risk assessments, and implement preventive measures

Targeted Interventions for High-Risk Customers

- High-risk customers need specialized attention
- We recommend personalized interventions to address their unique needs, such as tailored premium rates, additional coverage options, and proactive customer support
- This ensures that those most likely to file claims receive the care and support they require, ultimately benefiting both the customer and the insurance company

Key Insights

- Risk Factors Identification: Our models have identified key factors that increase the likelihood of claims
- Predictive Modeling: Utilizing advanced algorithms like CatBoost and XGBoost, we can accurately predict the probability of claims, allowing for better risk management
- Customer Segmentation: By segmenting customers based on risk levels, insurance companies can offer customized policies and interventions that address specific needs and behaviors

Conclusion

- The insights gained from our analysis of automobile insurance data provide a robust foundation for developing targeted interventions and policy adjustments. These strategies are designed to reduce claim risks, enhance customer satisfaction, and improve overall efficiency in claim management for insurance companies

"Thank you for joining us on this journey. Together, we can make a difference."