# Exploring Insurance Data: An In-Depth Analysis

# Introduction

In the realm of data science, exploratory data analysis (EDA) plays a pivotal role in understanding the intricacies of a dataset. This report delves into a vehicle insurance dataset, unraveling patterns, anomalies, and insights that can drive business decisions and strategies. Through various visualizations and statistical analyses, we aim to shed light on the key attributes that influence vehicle insurance policies, premiums, and claims.

## Data Overview

The dataset comprises 1,000,098 records and 52 columns, capturing various aspects of vehicle insurance, including policy details, vehicle specifications, and financial metrics. Initial observations reveal that certain columns have missing values, which we addressed through imputation techniques.

### Load the Data

```
import sys sys.path.append('../src')  # Add the src directory to the system path

from data_quality_check import DataQualityCheck
from data_clean_processing import
DataCleanProcessing

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot
as plt


# Set the file path for the data file file_path =
'../data/MachineLearningRating_v3.txt'
```

```
# Display the first few rows
data.head()
```

| | UnderwrittenCoverID | PolicyID | TransactionMonth | IsVATRegistered | Citizenship | LegalT |
|---|---|---|---|---|---|---|
| **0** | 145249 | 12827 | 2015-03-01 00:00:00 | True | | C Corpor |
| **1** | 145249 | 12827 | 2015-05-01 00:00:00 | True | | C Corpor |
| **2** | 145249 | 12827 | 2015-07-01 00:00:00 | True | | C Corpor |
| **3** | 145255 | 12827 | 2015-05-01 00:00:00 | True | | C Corpor |
| **4** | 145255 | 12827 | 2015-07-01 00:00:00 | True | | C Corpor |

5 rows × 52 columns

## Checking Missing Value

```
# Initialize and load data data_quality =
DataQualityCheck(file_path) data =
data_quality.load_data()
data_quality.basic_info()
```

```
   UnderwrittenCoverID  PolicyID   TransactionMonth  IsVATRegistered \
0                145249     12827  2015-03-01 00:00:00         True
1                145249     12827  2015-05-01 00:00:00         True
2                145249     12827  2015-07-01 00:00:00         True
3                145255     12827  2015-05-01 00:00:00         True
4                145255     12827  2015-07-01 00:00:00         True

    Citizenship        LegalType Title Language              Bank \
0                 Close Corporation    Mr  English  First National Bank
```

```
1                    Close Corporation   Mr  English  First National Bank
2                    Close Corporation   Mr  English  First National Bank
3                    Close Corporation   Mr  English  First National Bank
4                    Close Corporation   Mr  English  First National Bank

    AccountType  ...             ExcessSelected CoverCategory  \

0   Current account  ...         Mobility - Windscreen   Windscreen
1   Current account  ...         Mobility - Windscreen   Windscreen
2   Current account  ...         Mobility - Windscreen   Windscreen
3   Current account  ...  Mobility - Metered Taxis - R2000   Own damage
4   Current account  ...  Mobility - Metered Taxis - R2000   Own damage

  CoverType        CoverGroup          Section  \

0   Windscreen  Comprehensive - Taxi  Motor Comprehensive
1   Windscreen  Comprehensive - Taxi  Motor Comprehensive
2   Windscreen  Comprehensive - Taxi  Motor Comprehensive
3   Own Damage  Comprehensive - Taxi  Motor Comprehensive
4   Own Damage  Comprehensive - Taxi  Motor Comprehensive

                Product StatutoryClass StatutoryRiskType  \

0   Mobility Metered Taxis: Monthly    Commercial    IFRS Constant
1   Mobility Metered Taxis: Monthly    Commercial    IFRS Constant
2   Mobility Metered Taxis: Monthly    Commercial    IFRS Constant
3   Mobility Metered Taxis: Monthly    Commercial    IFRS Constant
4   Mobility Metered Taxis: Monthly    Commercial    IFRS Constant

  TotalPremium TotalClaims

0        21.929825       0.0
1        21.929825       0.0
2         0.000000       0.0
3       512.848070       0.0
4         0.000000       0.0

[5 rows x 52 columns]

(1000098, 52)

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1000098 entries, 0 to
1000097 Data columns (total 52 columns):

 #  Column              Non-Null Count    Dtype
```

```
--- ------              --------------   -----
0   UnderwrittenCoverID     1000098 non-null  int64
1   PolicyID                1000098 non-null  int64
2   TransactionMonth        1000098 non-null  object
3   IsVATRegistered         1000098 non-null  bool
4   Citizenship             1000098 non-null  object
5   LegalType               1000098 non-null  object
6   Title                   1000098 non-null  object
7   Language                1000098 non-null  object
8   Bank          854137 non null             object
```

Missing Values

```
9   Bank                          145961
10  AccountType                    40232
11  MaritalStatus                   8259
12  Gender                          9536
13  mmcode                           552
14  VehicleType                      552
15  make                             552
16  Model                            552
17  Cylinders                        552
18  cubiccapacity                    552
19  kilowatts                        552
20  bodytype                         552
21  NumberOfDoors                    552
22  VehicleIntroDate                 552
23  CustomValueEstimate           779642
24  CapitalOutstanding                 2
25  NewVehicle                    153295
26  WrittenOff                    641901
27  Rebuilt                       641901
28  Converted                     641901
29  CrossBorder                   999400
30  NumberOfVehiclesInFleet      1000098
dtype: int64
```

## Data Clean Processing

### Data Cleaning

To ensure robust analysis, missing values were handled using appropriate imputation methods:

Categorical columns were filled with the mode.

Numerical columns were filled with the mean.

```
# Initialize the DataCleanProcessing class with the loaded data data_cleaner =
DataCleanProcessing(data)

# Clean the data cleaned_data =
data_cleaner.clean_missing_values()

# Verify no missing values remain if
data_cleaner.verify_no_missing_values():

   print("Data cleaned successfully with no missing values remaining.") else:
print("There are still missing values in the data.")
```

```
d:\Insurance_Claims_Analysis\notebooks\../src\data_clean_processing.py:13: FutureWarn
 The behavior will change in pandas 3.0. This inplace method will never work because t For
      example, when doing 'df[col].method(value, inplace=True)', try using 'df.method({
```

```
# Display the shape of the dataframe after cleaning, to check any drop column. data.shape
```

```
(1000098, 52)
```

```
# Display information about the dataframe
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1000098 entries, 0 to
1000097 Data columns (total 52 columns):

 #  Column                  Non-Null Count    Dtype

0    UnderwrittenCoverID  1000098        non-null int64
1    PolicyID             1000098        non-null int64
2    TransactionMonth     1000098        non-null object
3    IsVATRegistered      1000098        non-null bool
4    Citizenship          1000098        non-null object
5    LegalType            1000098        non-null object
6    Title                1000098        non-null object
7    Language             1000098        non-null object
8    Bank                 1000098        non-null object
9    AccountType          1000098        non-null object
```

```
10  MaritalStatus          1000098       non-null object
11  Gender                 1000098       non-null object
12  Country                1000098       non-null object
13  Province               1000098       non-null object
14  PostalCode             1000098       non-null int64
15  MainCrestaZone         1000098       non-null object
16  SubCrestaZone          1000098       non-null object
17  ItemType               1000098       non-null object
18  mmcode                 1000098       non-null float64
19  VehicleType            1000098       non-null object
20  RegistrationYear       1000098       non-null int64
21  make                   1000098       non-null object
22  Model                  1000098       non-null object
23  Cylinders              1000098       non-null float64
24  cubiccapacity          1000098       non-null float64
25  kilowatts              1000098       non-null float64
```

## Exploratory Data Analysis(EDA)

### Distribution of Key Attributes

The visualizations above depict the distribution of several important columns, including `PolicyID`, `PostalCode`, `RegistrationYear`, `Cylinders`, `cubiccapacity`, `kilowatts`, `NumberOfDoors`, `CustomValueEstimate`, `SumInsured`, `CalculatedPremiumPerTerm`, `TotalPremium`, and `TotalClaims`. Here are some notable observations:

1. PolicyID and UnderwrittenCoverID: These IDs show a right-skewed distribution, indicating that most policies fall within a certain range, with fewer policies having very high IDs.

2. PostalCode: There is a varied distribution across postal codes, suggesting a wide geographical spread of policyholders.

3. RegistrationYear: The majority of vehicles were registered between 2000 and 2015, reflecting a relatively new fleet.

4. Cylinders and NumberOfDoors: Both attributes exhibit significant peaks at common values (e.g., 4 cylinders, 4 doors), typical of standard vehicle configurations.

5. cubiccapacity and kilowatts: These attributes, related to vehicle power, show that most vehicles fall within a common range of engine capacity and power output.

6. CustomValueEstimate and SumInsured: There are high-value outliers, but most values cluster at lower ranges, indicating standard valuation and insurance coverage for most vehicles.

7. CalculatedPremiumPerTerm, TotalPremium, and TotalClaims: Premiums and claims also show skewed distributions, with most policies having relatively low premiums and claims amounts.

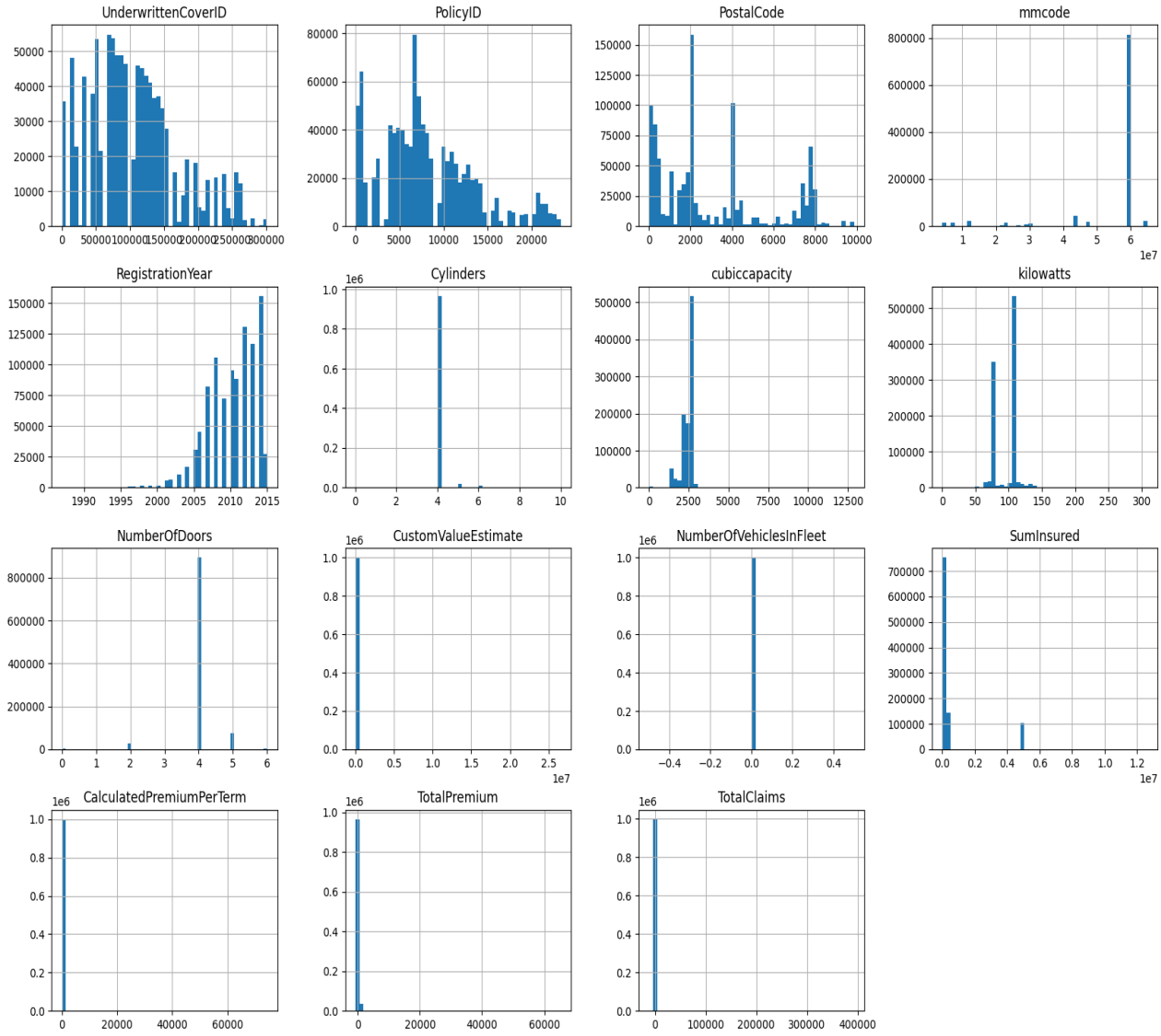
Notably, certain columns such as `CustomValueEstimate`, `WrittenOff`, `Rebuilt`, and `Converted` had a high proportion of missing values, which may require further investigation or consideration for exclusion from certain analyses.

## Univariate Analysis

### Histograms for numerical columns:

Insight: This plot reveals the most frequent premium ranges and highlights any outliers that may require further investigation.

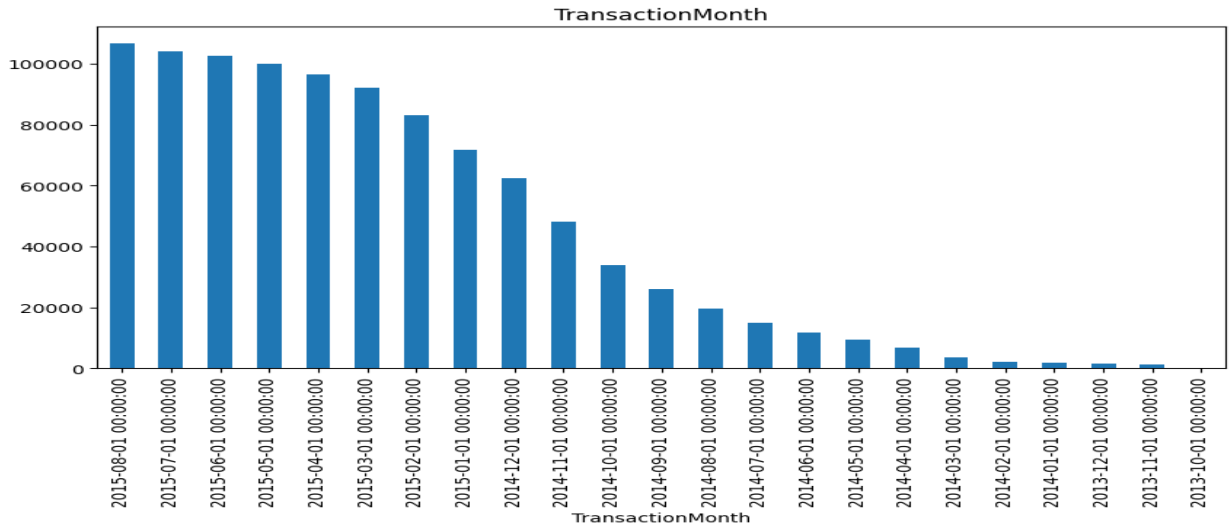
```
# Plot histograms for numerical columns
data.hist(bins=50, figsize=(20, 15))
plt.show()
```

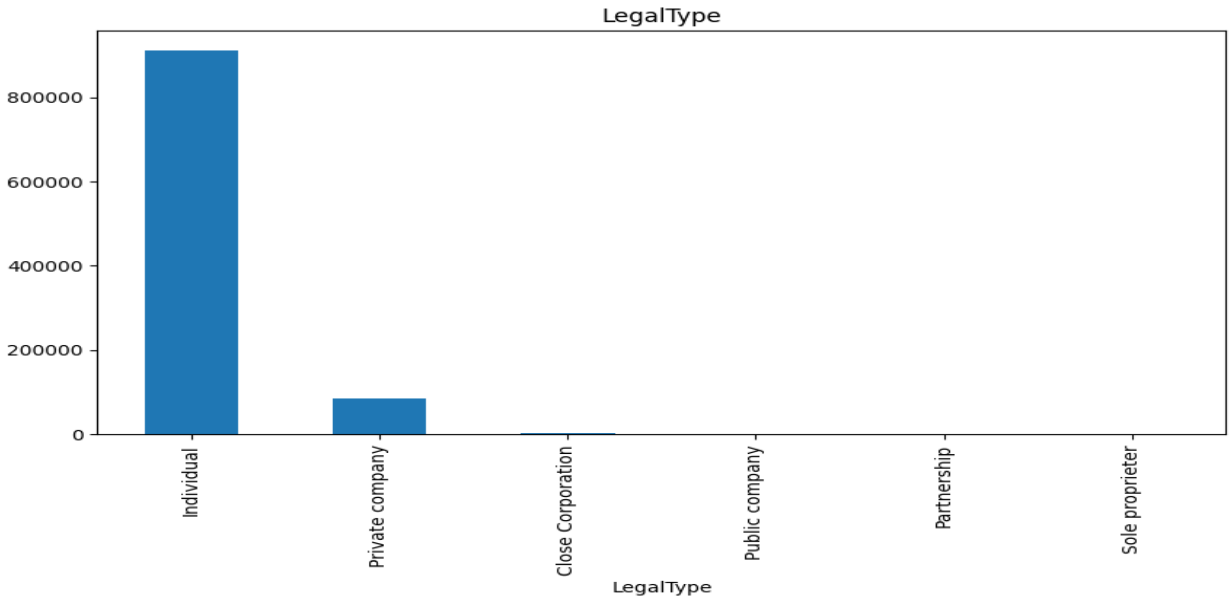


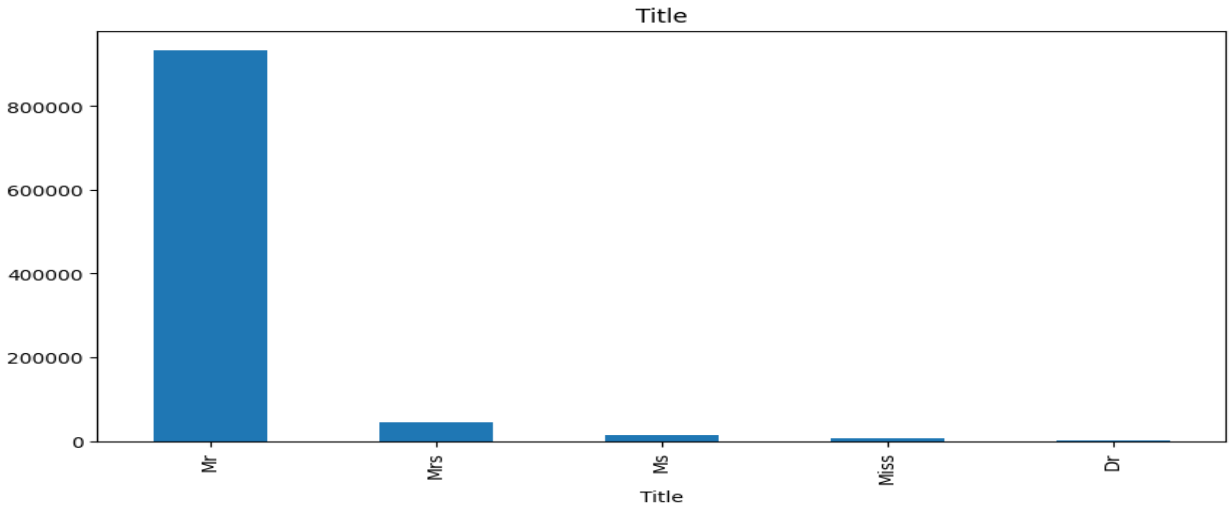

Bar charts for categorical columns:

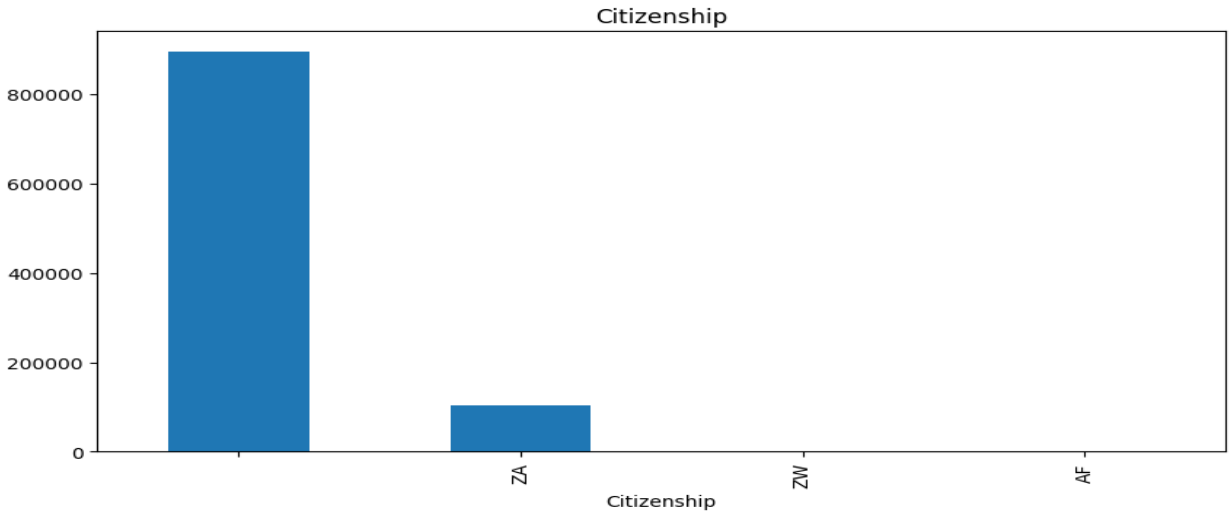
```
# Plot bar charts for categorical columns
for column in data.select_dtypes(include=['object']).columns:
    data[column].value_counts().plot(kind='bar', figsize=(10, 5))
    plt.title(column)
    plt.show()
```

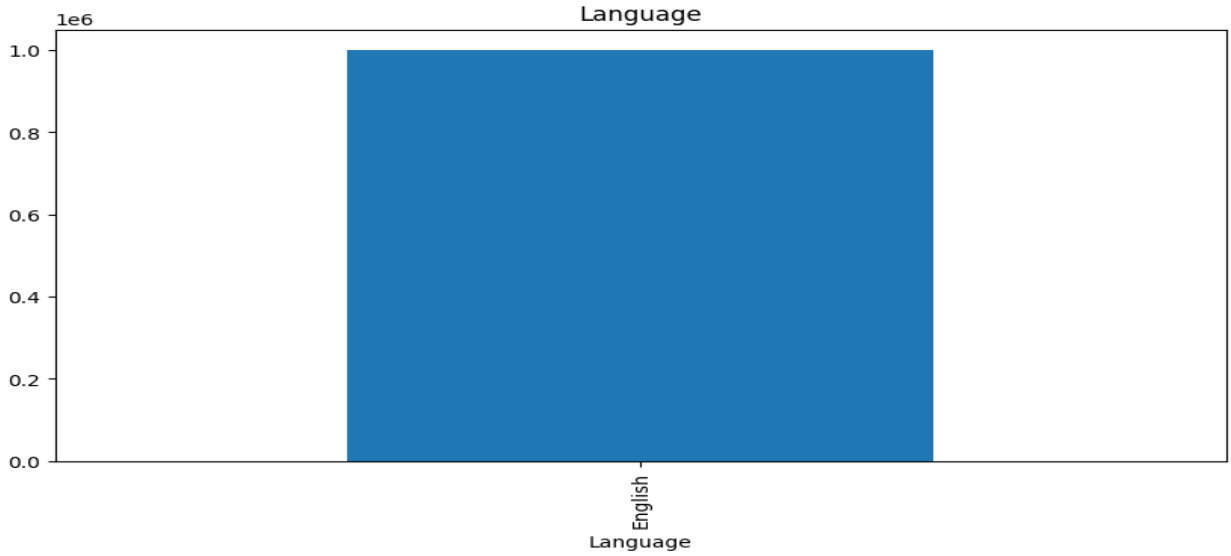
Insight: This plot helps in understanding whether higher premiums correlate with higher claims, providing insights into customer behavior and risk assessment.

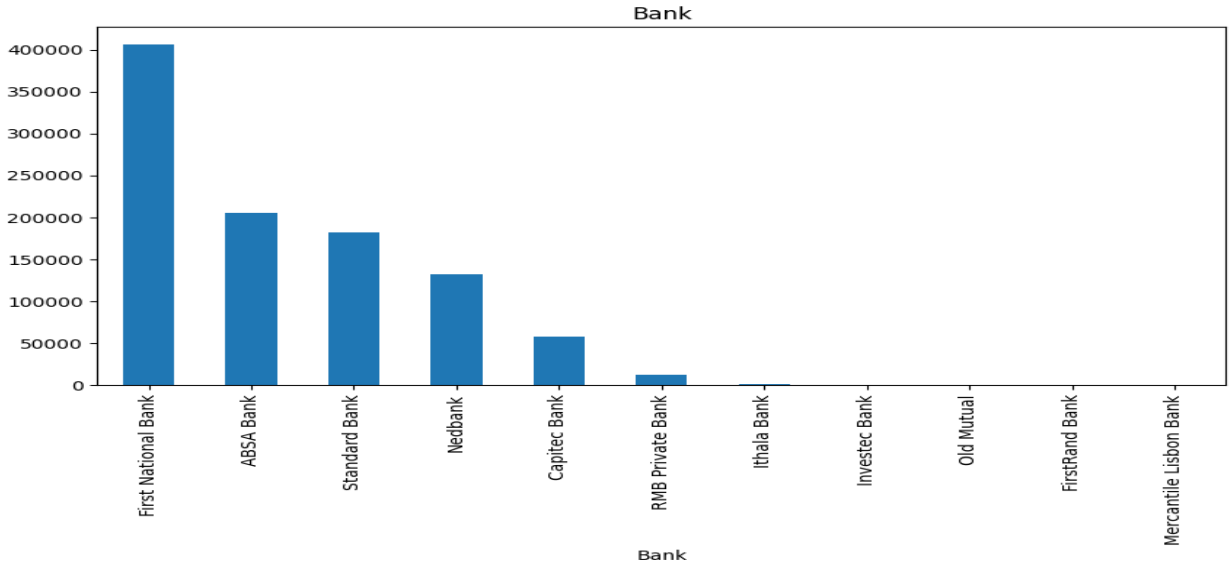

TransactionMonth
100000
80000
60000
40000
20000
0
2015-08-01 00:00:00
2015-07-01 00:00:00
2015-06-01 00:00:00
2015-05-01 00:00:00
2015-04-01 00:00:00
2015-03-01 00:00:00
2015-02-01 00:00:00
2015-01-01 00:00:00
2014-12-01 00:00:00
2014-11-01 00:00:00
2014-10-01 00:00:00
2014-09-01 00:00:00
2014-08-01 00:00:00
2014-07-01 00:00:00
2014-06-01 00:00:00
2014-05-01 00:00:00
2014-04-01 00:00:00
2014-03-01 00:00:00
2014-02-01 00:00:00
2014-01-01 00:00:00
2013-12-01 00:00:00
2013-11-01 00:00:00
2013-10-01 00:00:00
TransactionMonth

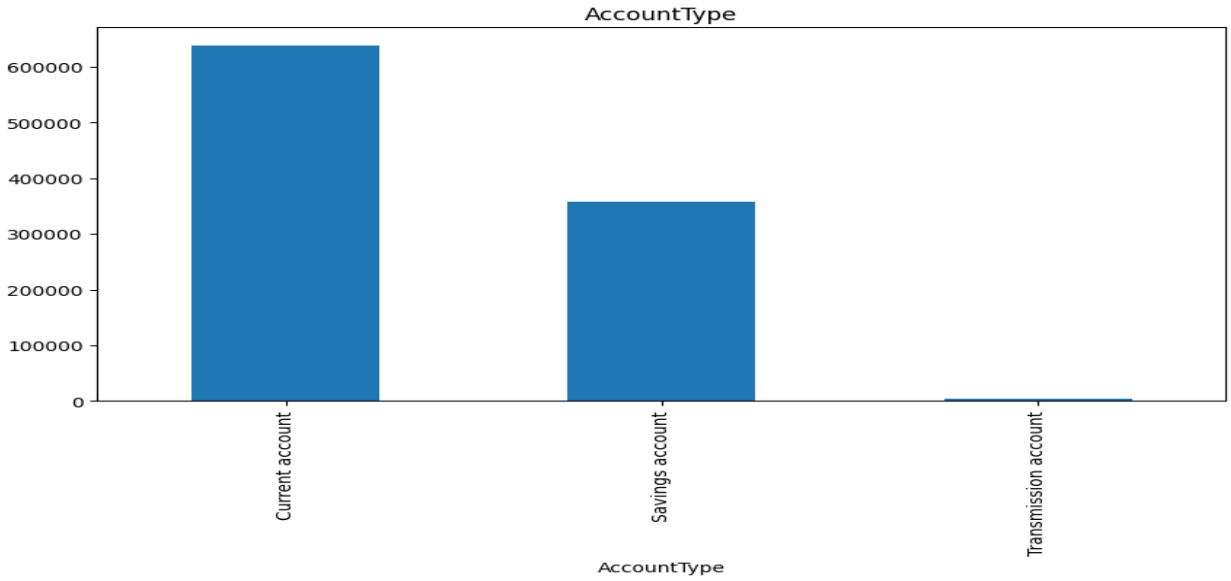

LegalType
800000
600000
400000
200000
0
Individual
Private company
Close Corporation
Public company
Partnership
Sole proprieter
LegalType

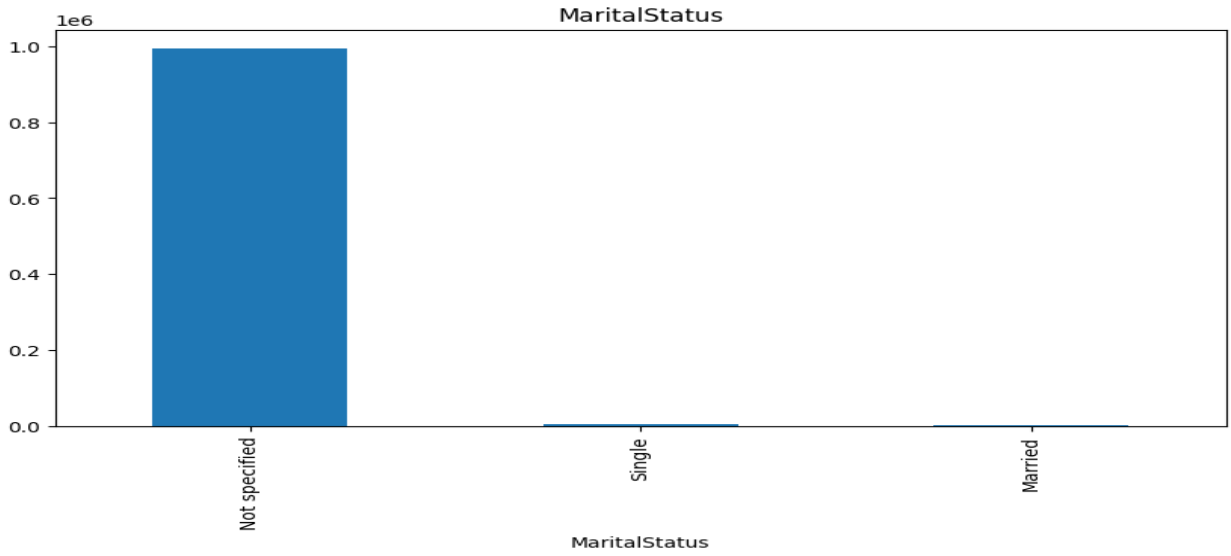

Title
800000
600000
400000
200000
0
Mr
Mrs
Ms
Miss
Dr
Title

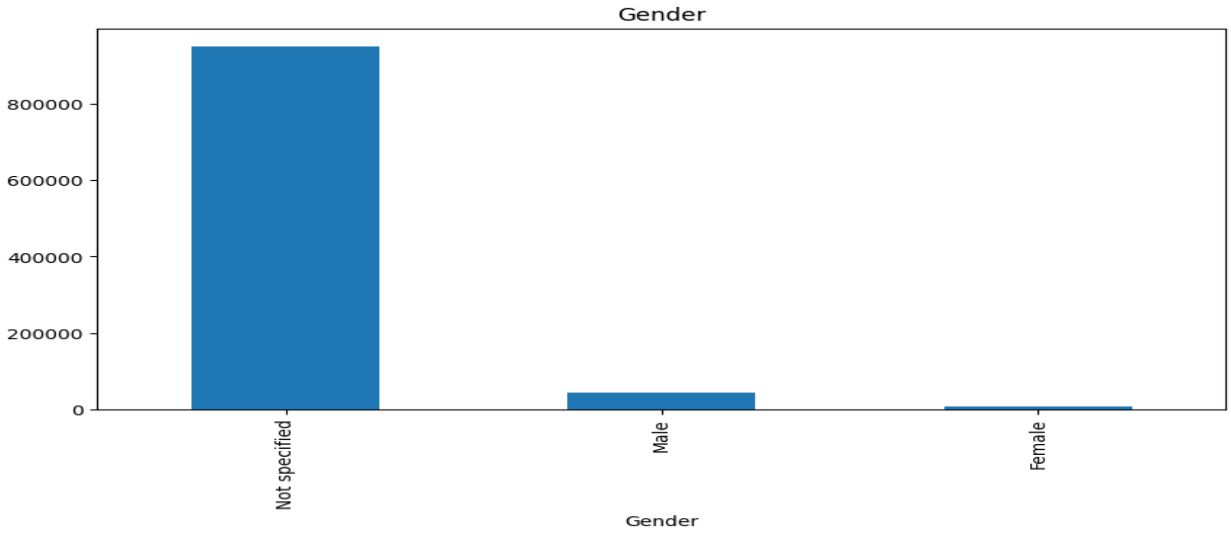

Citizenship
800000
600000
400000
200000
0
ZA
ZW
AF
Citizenship

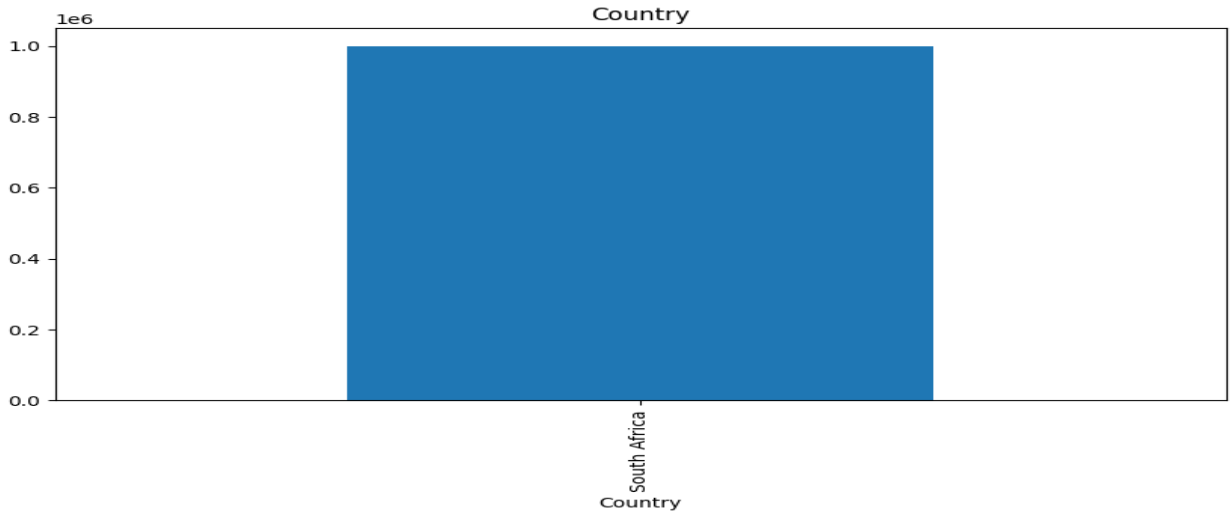

Language
1e6
1.0
0.8
0.6
0.4
0.2
0.0
English
Language

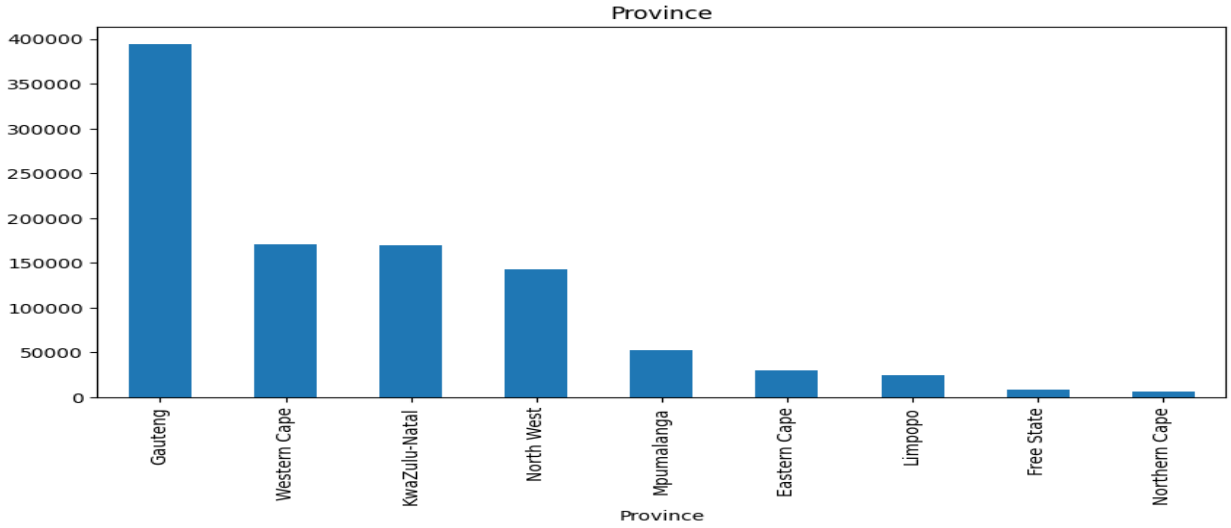

Bank
400000
350000
300000
250000
200000
150000
100000
50000
0
First National Bank
ABSA Bank
Standard Bank
Nedbank
Capitec Bank
RMB Private Bank
Ithala Bank
Investec Bank
Old Mutual
FirstRand Bank
Mercantile Lisbon Bank
Bank

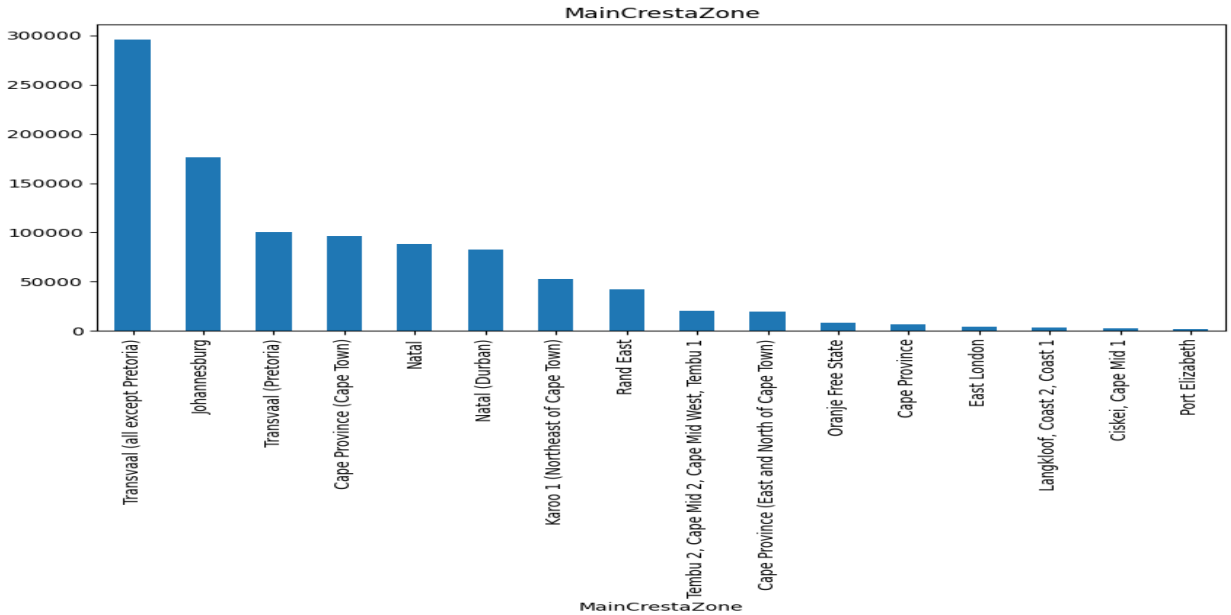

AccountType
600000
500000
400000
300000
200000
100000
0
Current account
Savings account
Transmission account
AccountType

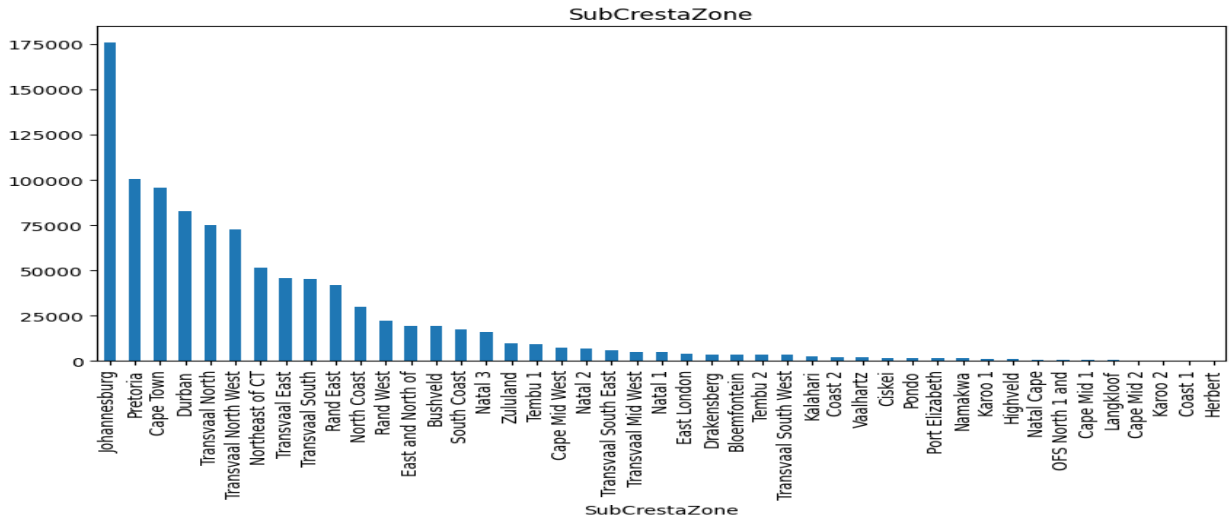

MaritalStatus
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Not specified
Single
Married
MaritalStatus

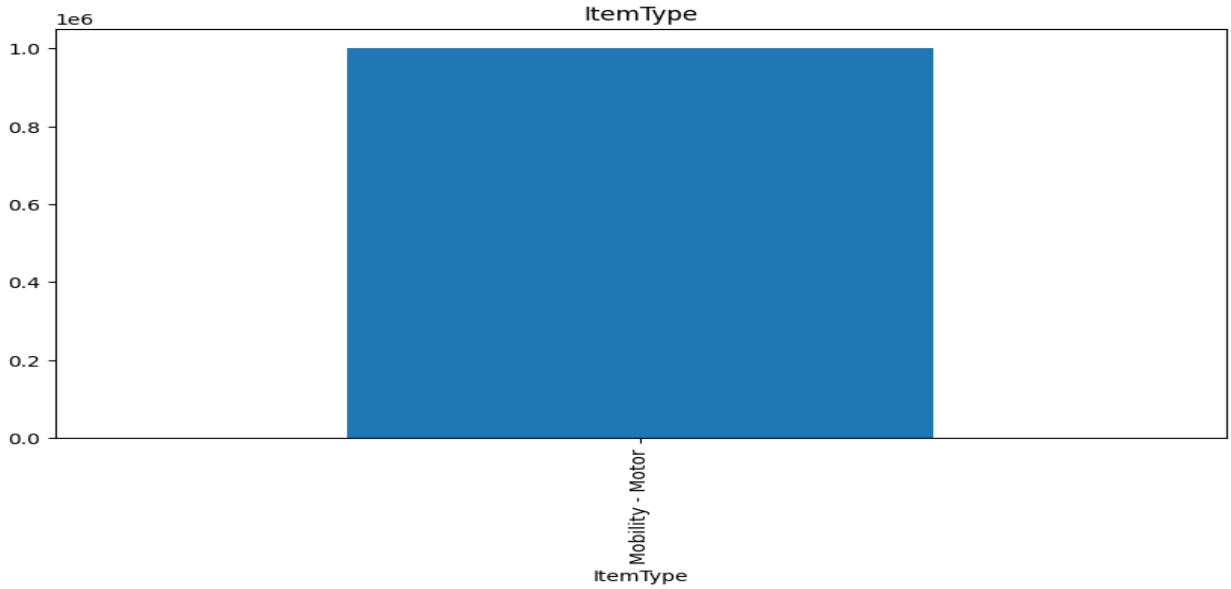

Gender
800000
600000
400000
200000
0
Not specified
Male
Female
Gender

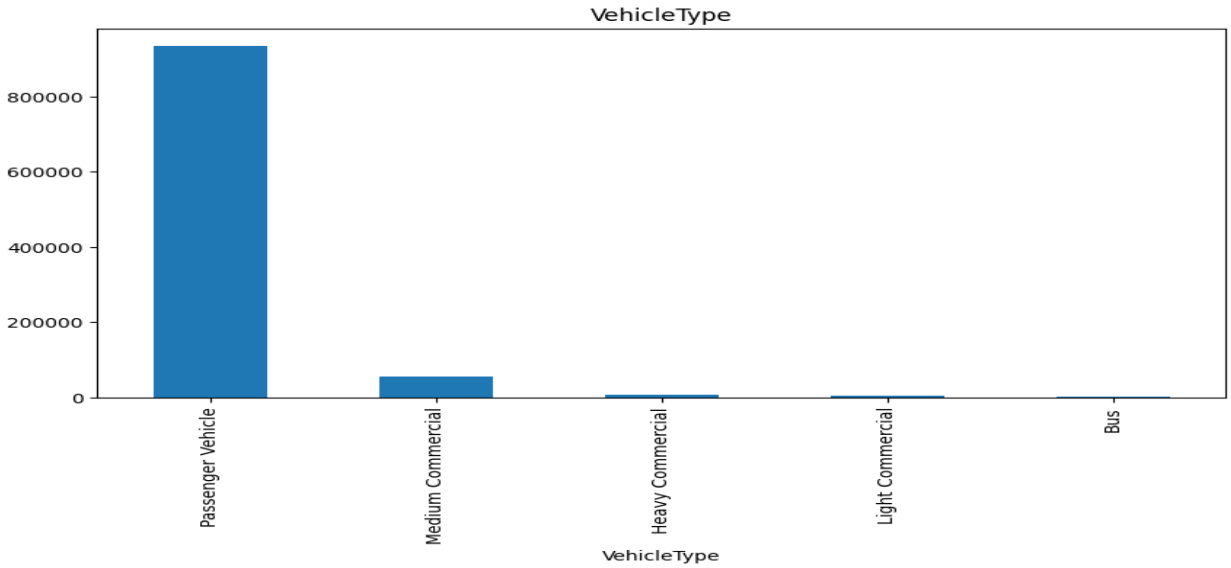

Country
1e6
1.0
0.8
0.6
0.4
0.2
0.0
South Africa
Country

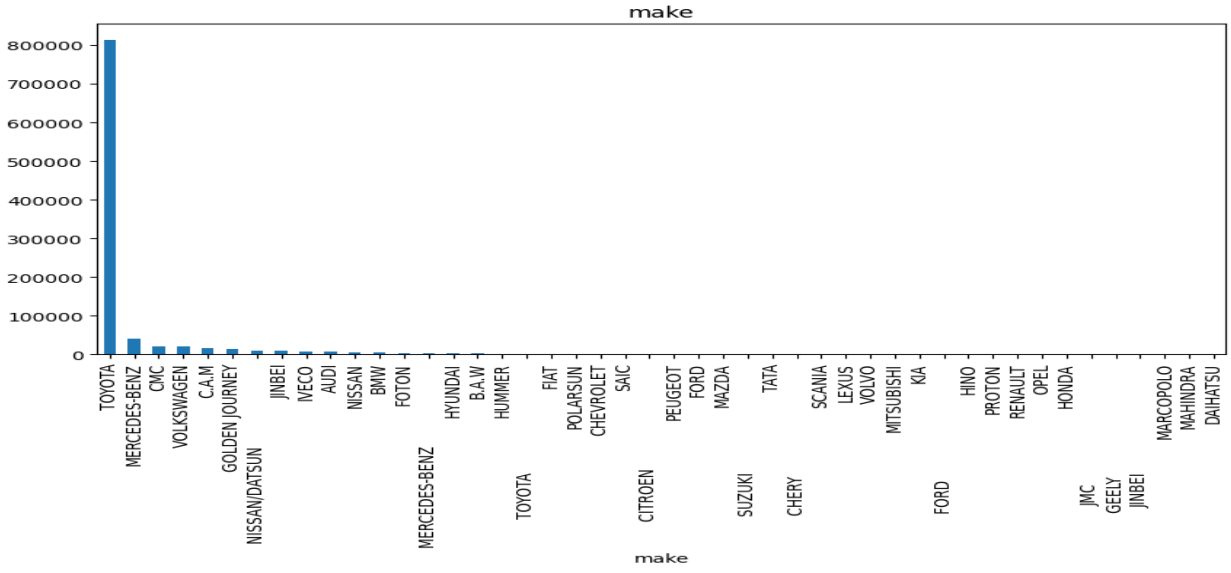

Province
400000
350000
300000
250000
200000
150000
100000
50000
0
Gauteng
Western Cape
KwaZulu-Natal
North West
Mpumalanga
Eastern Cape
Limpopo
Free State
Northern Cape
Province

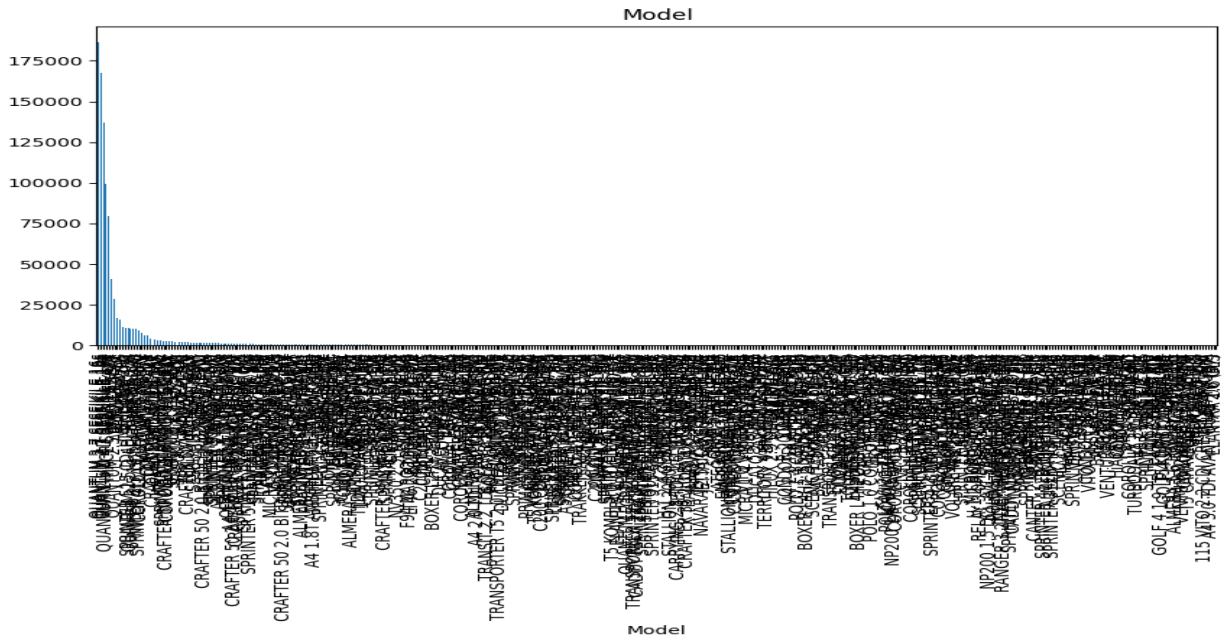

MainCrestaZone
300000
250000
200000
150000
100000
50000
0
Transvaal (all except Pretoria)
Johannesburg
Transvaal (Pretoria)
Cape Province (Cape Town)
Natal
Natal (Durban)
Karoo 1 (Northeast of Cape Town)
Rand East
Tembu 2, Cape Mid 2, Cape Mid West, Tembu 1
Cape Province (East and North of Cape Town)
Oranje Free State
Cape Province
East London
Langkloof, Coast 2, Coast 1
Ciskei, Cape Mid 1
Port Elizabeth
MainCrestaZone

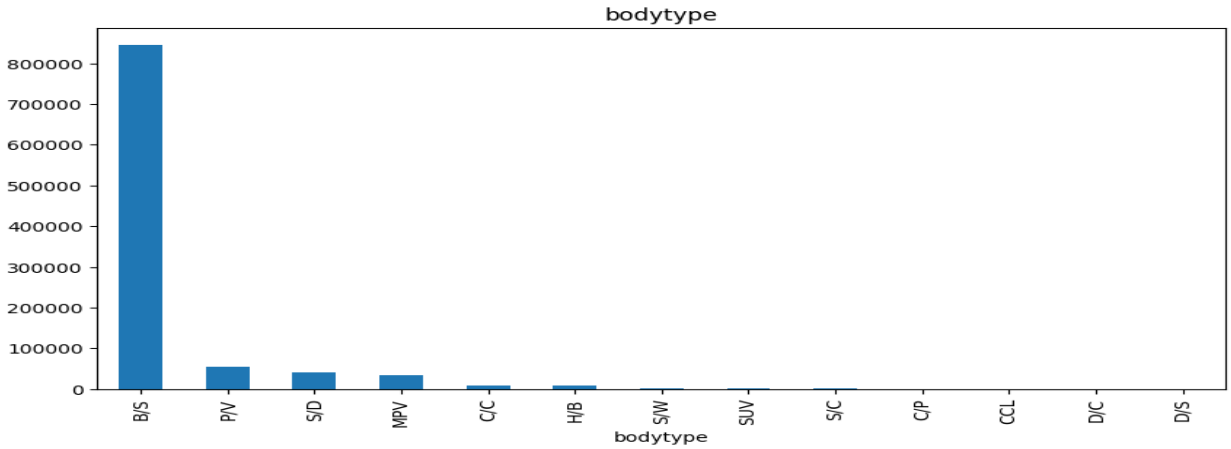

SubCrestaZone
175000
150000
125000
100000
75000
50000
25000
0
Johannesburg
Pretoria
Cape Town
Durban
Transvaal North
Transvaal North West
Northeast of CT
Transvaal East
Transvaal South
Rand East
North Coast
Rand West
East and North of
Bushveld
South Coast
Natal 3
Zululand
Tembu 1
Cape Mid West
Natal 2
Transvaal South East
Transvaal Mid West
Natal 1
East London
Drakensberg
Bloemfontein
Tembu 2
Transvaal South West
Kalahari
Coast 2
Vaalhartz
Ciskei
Pondo
Port Elizabeth
Namakwa
Karoo 1
Highveld
Natal Cape
OFS North 1 and
Cape Mid 1
Langkloof
Cape Mid 2
Karoo 2
Coast 1
Herbert
SubCrestaZone

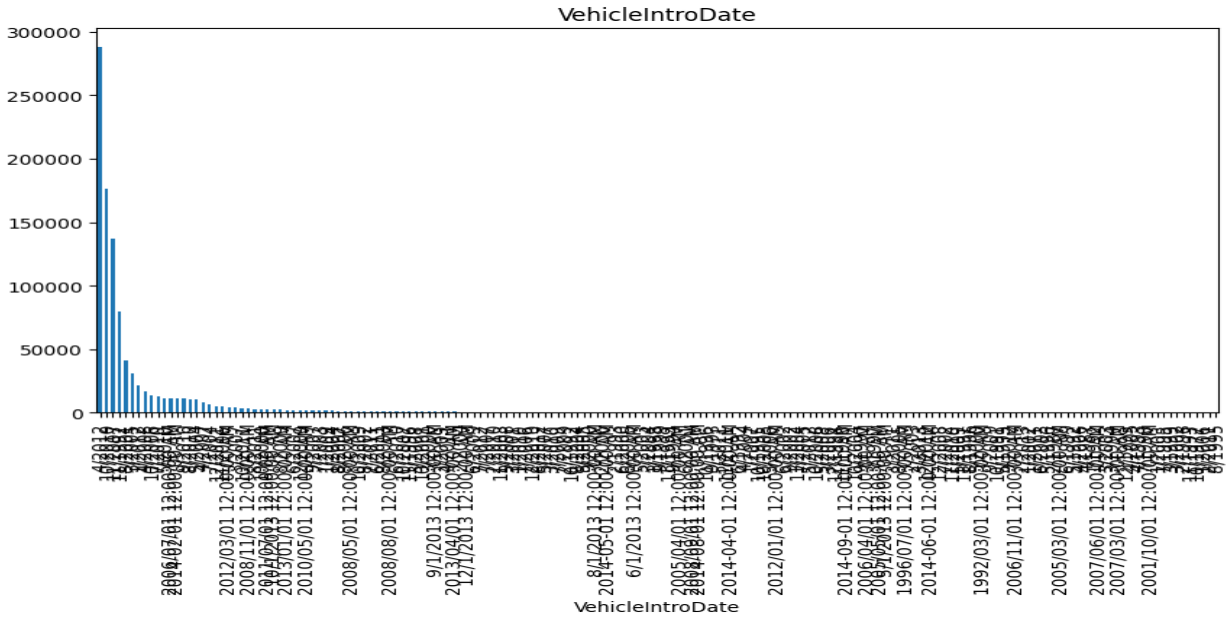

ItemType
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Mobility - Motor
ItemType

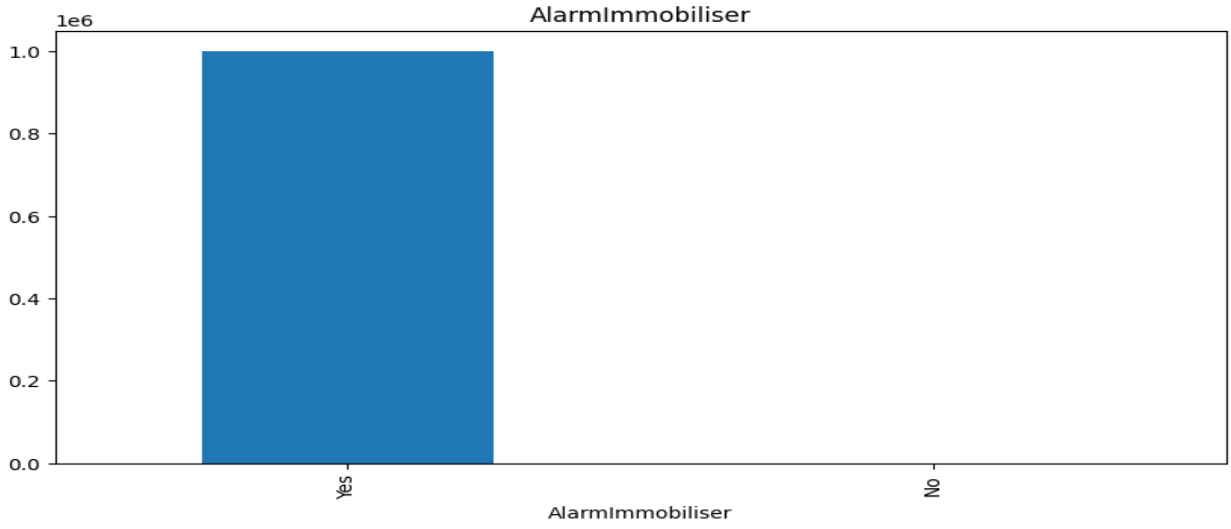

VehicleType
800000
600000
400000
200000
0
Passenger Vehicle
Medium Commercial
Heavy Commercial
Light Commercial
Bus
VehicleType

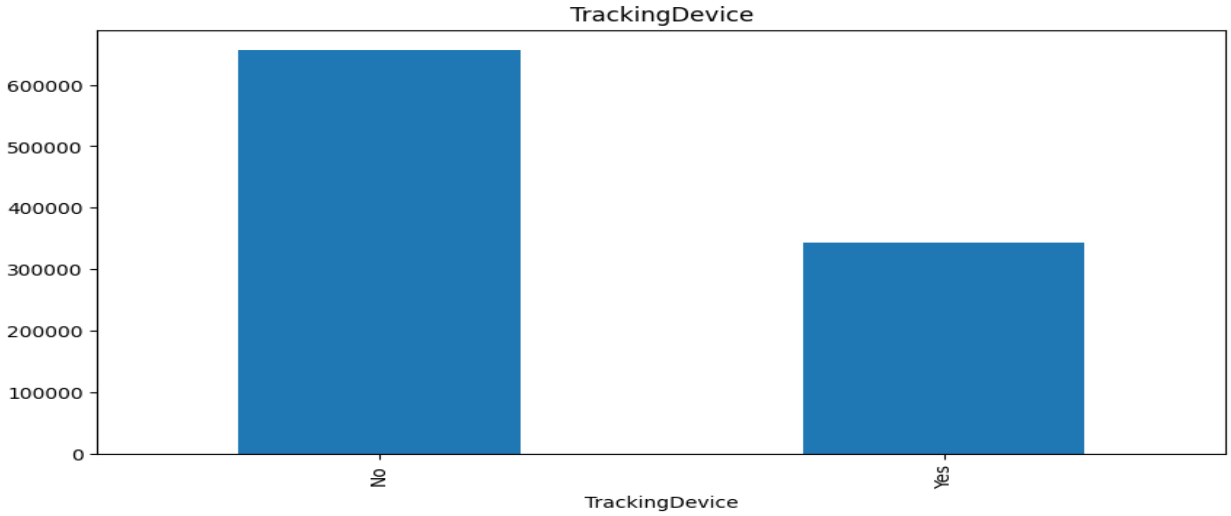

make
800000
700000
600000
500000
400000
300000
200000
100000
0
TOYOTA
MERCEDES-BENZ
CMC
VOLKSWAGEN
C.A.M
GOLDEN JOURNEY
NISSAN/DATSUN
JINBEI
IVECO
AUDI
NISSAN
BMW
FOTON
MERCEDES-BENZ
HYUNDAI
B.A.W
HUMMER
TOYOTA
FIAT
POLARSUN
CHEVROLET
SAIC
CITROEN
PEUGEOT
FORD
MAZDA
SUZUKI
TATA
CHERY
SCANIA
LEXUS
VOLVO
MITSUBISHI
KIA
FORD
HINO
PROTON
RENAULT
OPEL
HONDA
JMC
GEELY
JINBEI
MARCOPOLO
MAHINDRA
DAIHATSU
make

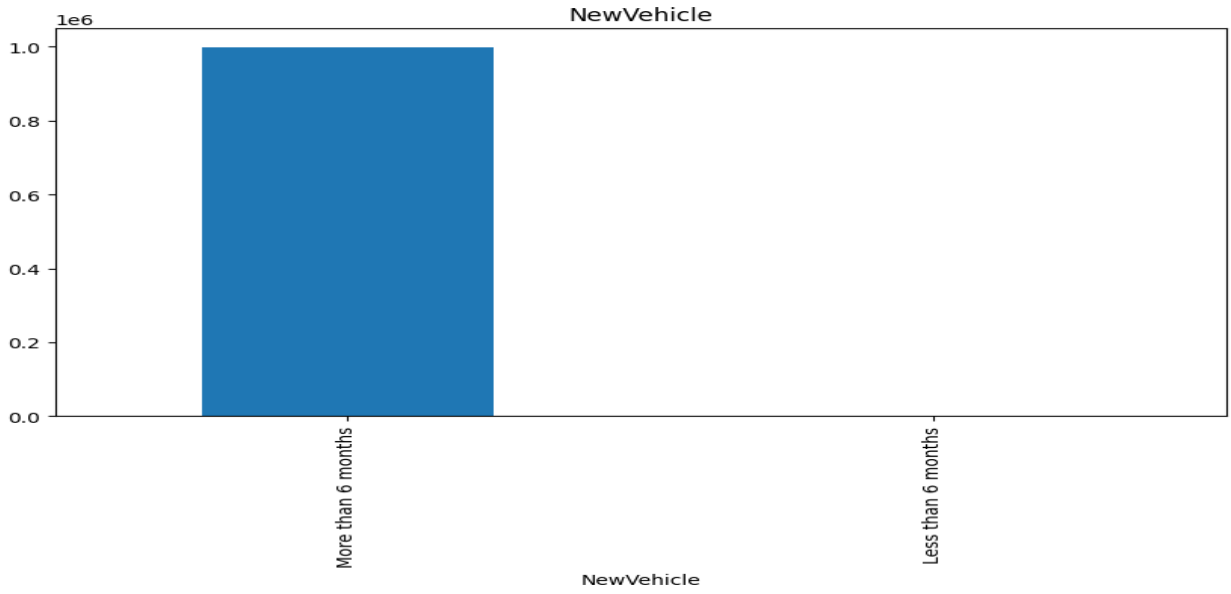

Model
175000
150000
125000
100000
75000
50000
25000
0
Model

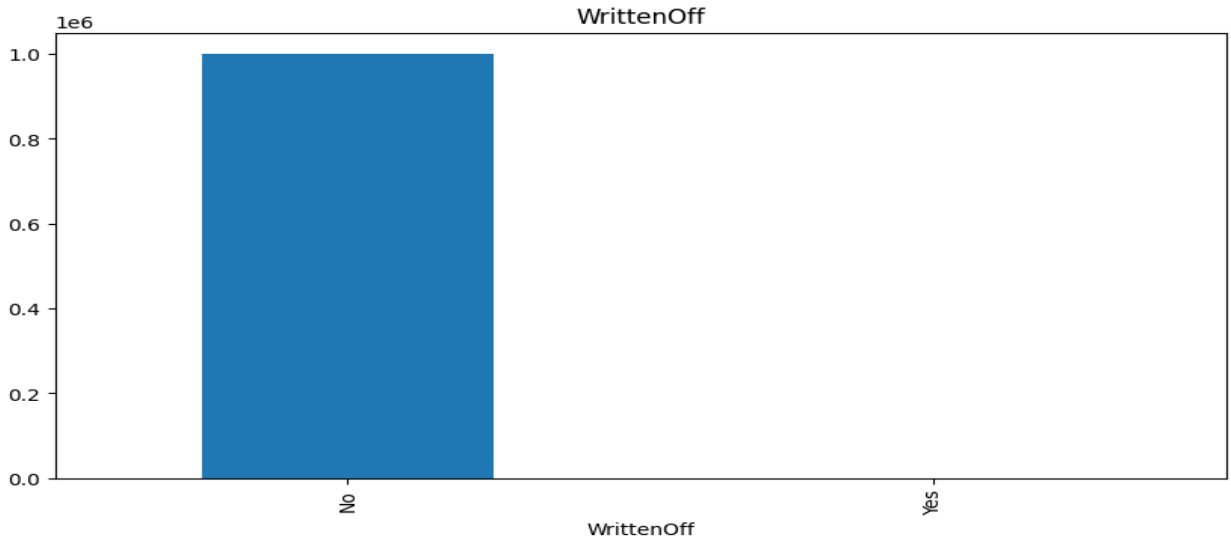

bodytype
800000
700000
600000
500000
400000
300000
200000
100000
0
B/S
P/V
S/D
MPV
C/C
H/B
S/W
SUV
S/C
C/P
CCL
D/C
D/S
bodytype

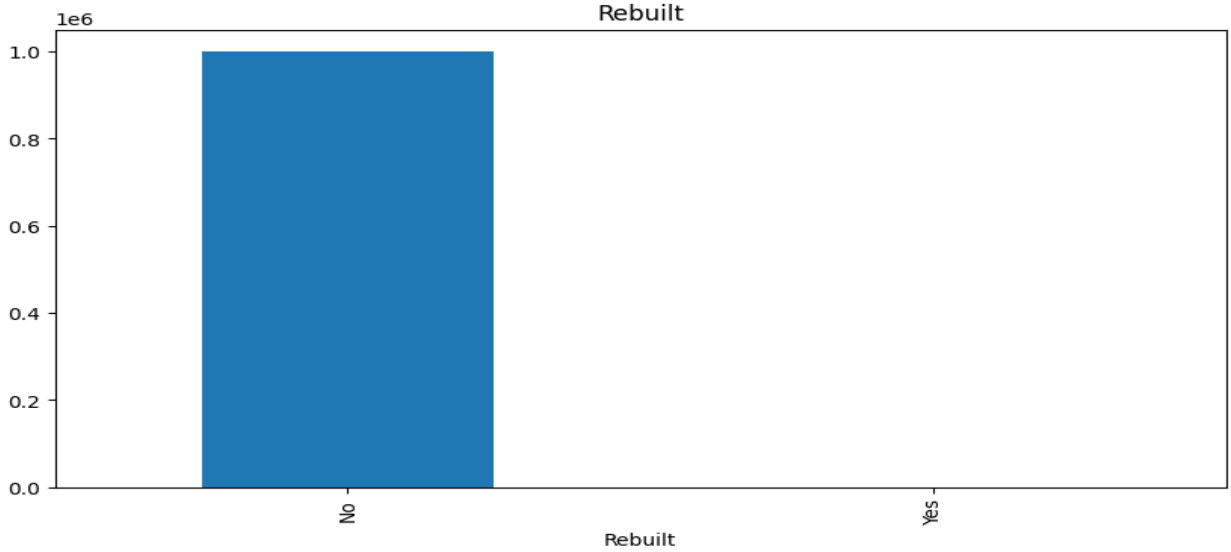

VehicleIntroDate
300000
250000
200000
150000
100000
50000
0
VehicleIntroDate

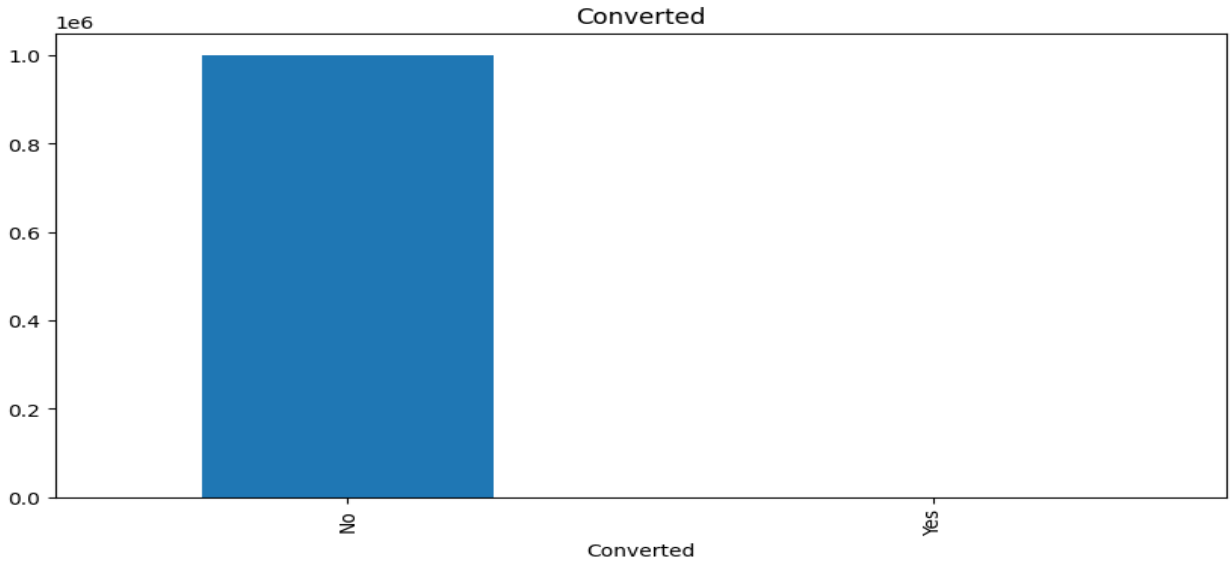

AlarmImmobiliser
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Yes
No
AlarmImmobiliser

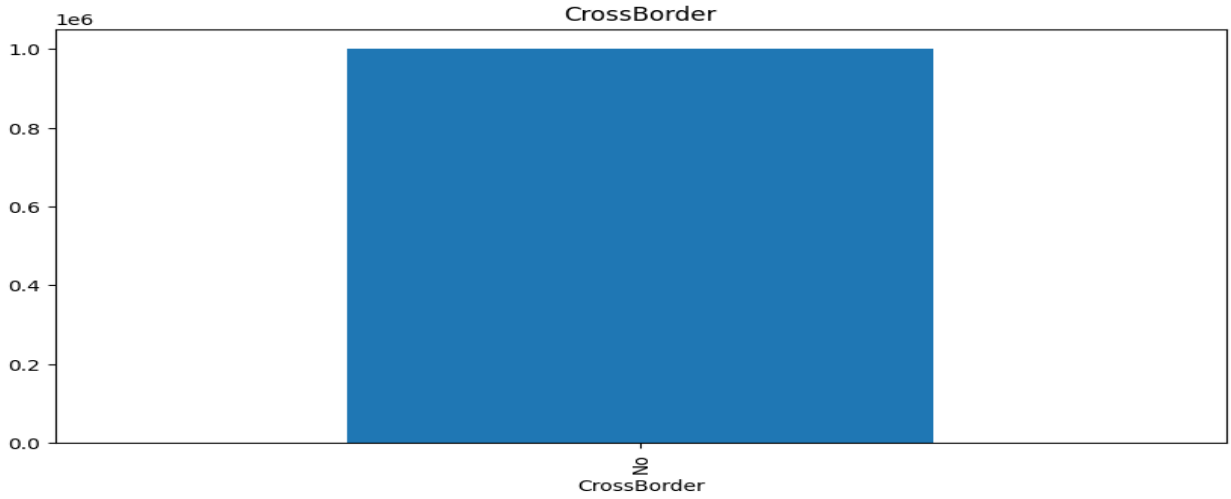

TrackingDevice
600000
500000
400000
300000
200000
100000
0
No
Yes
TrackingDevice

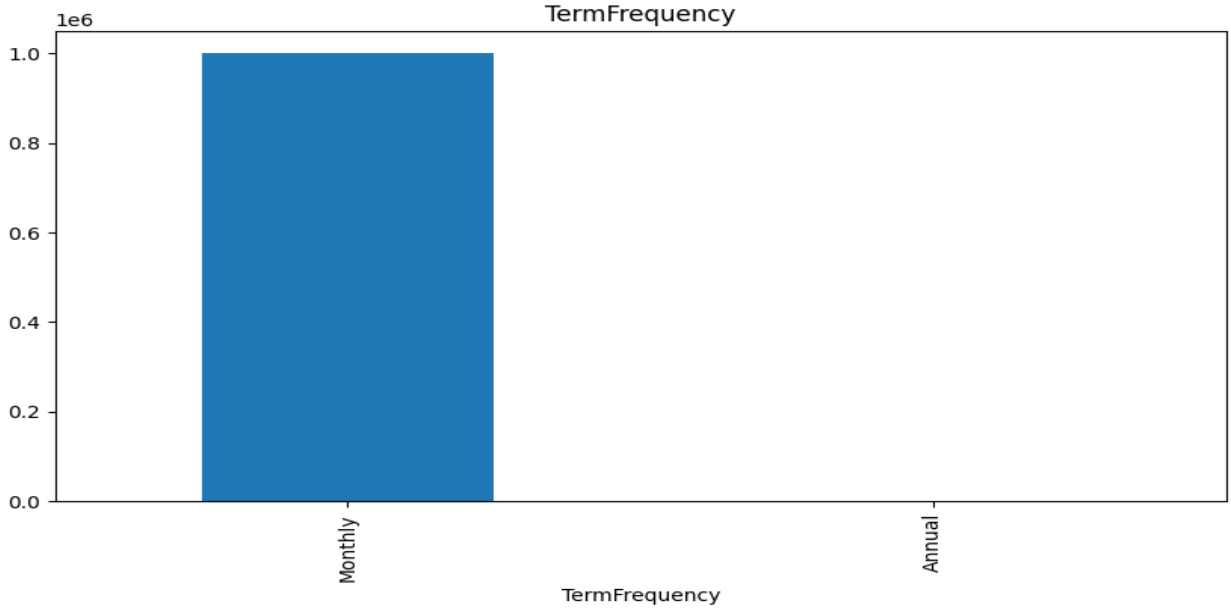

NewVehicle
1e6
1.0
0.8
0.6
0.4
0.2
0.0
More than 6 months
Less than 6 months
NewVehicle

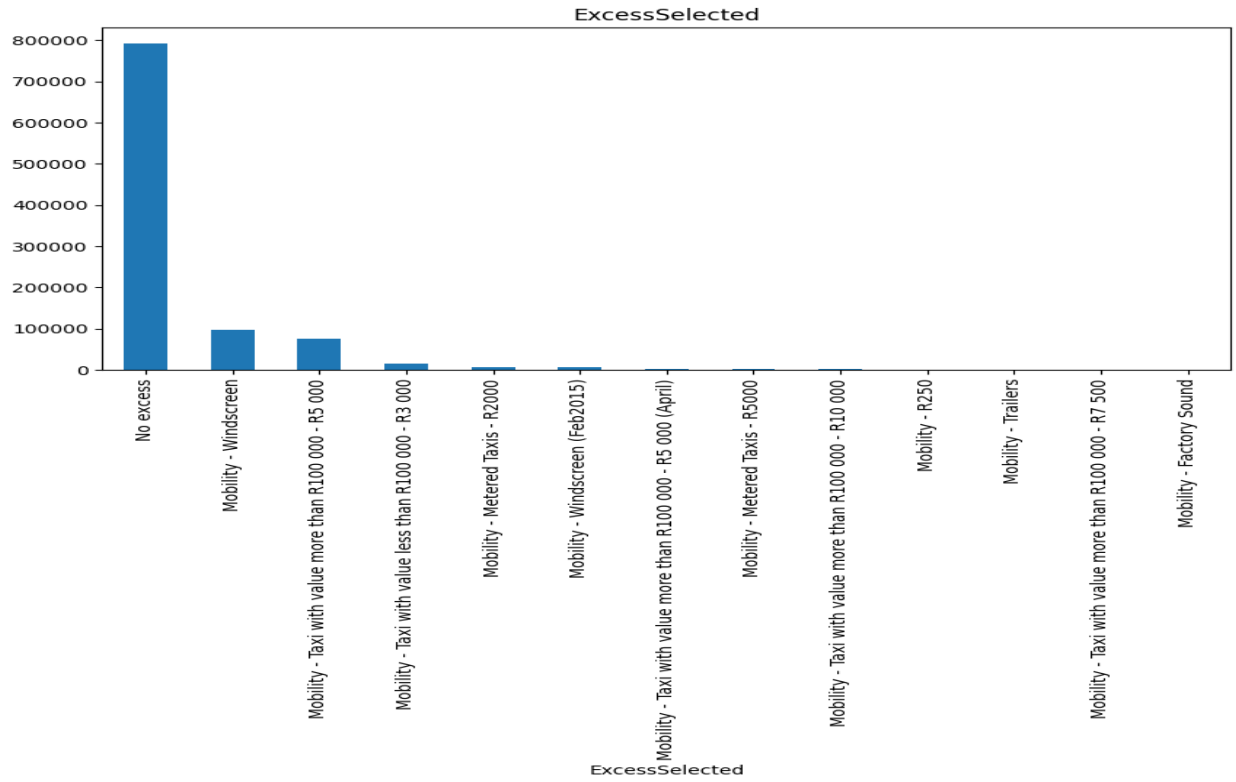

WrittenOff
1e6
1.0
0.8
0.6
0.4
0.2
0.0
No
Yes
WrittenOff

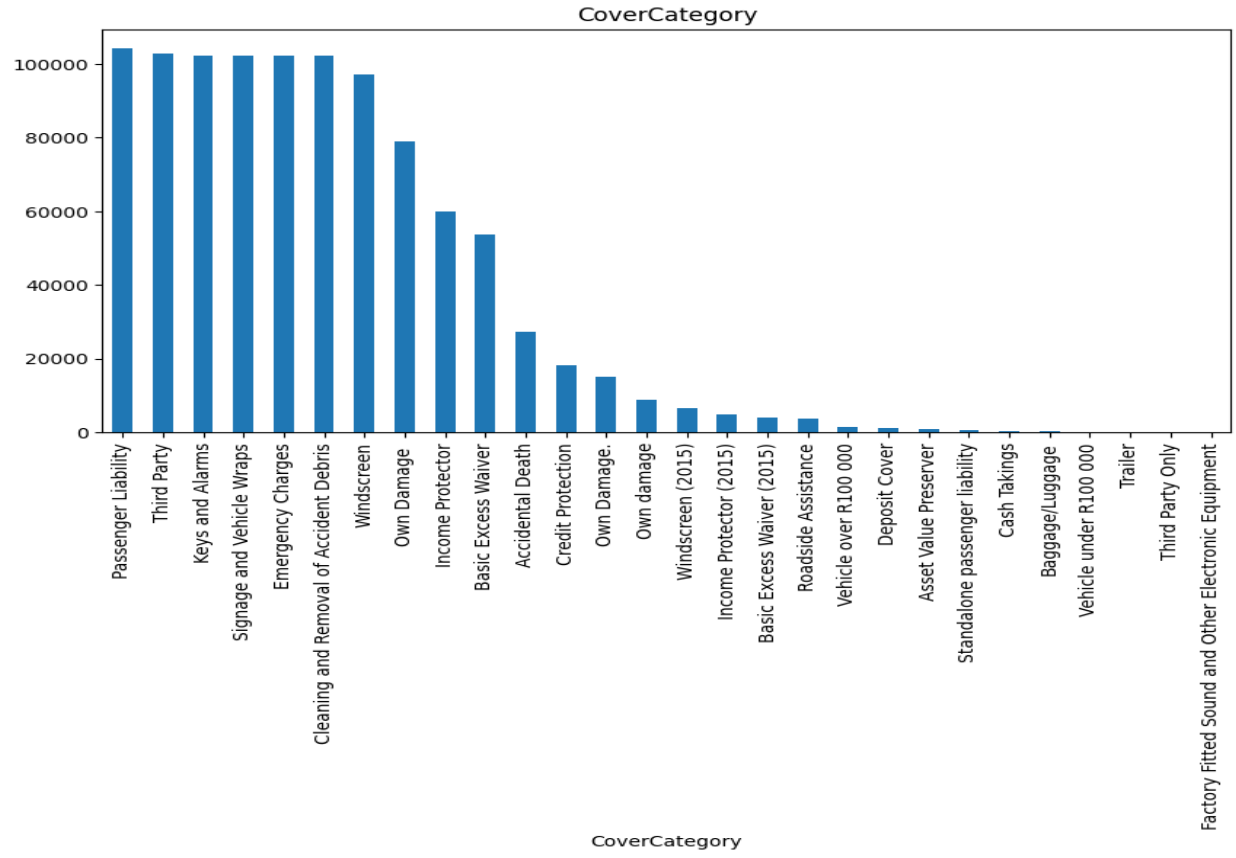

Rebuilt
1e6
1.0
0.8
0.6
0.4
0.2
0.0
No
Yes
Rebuilt

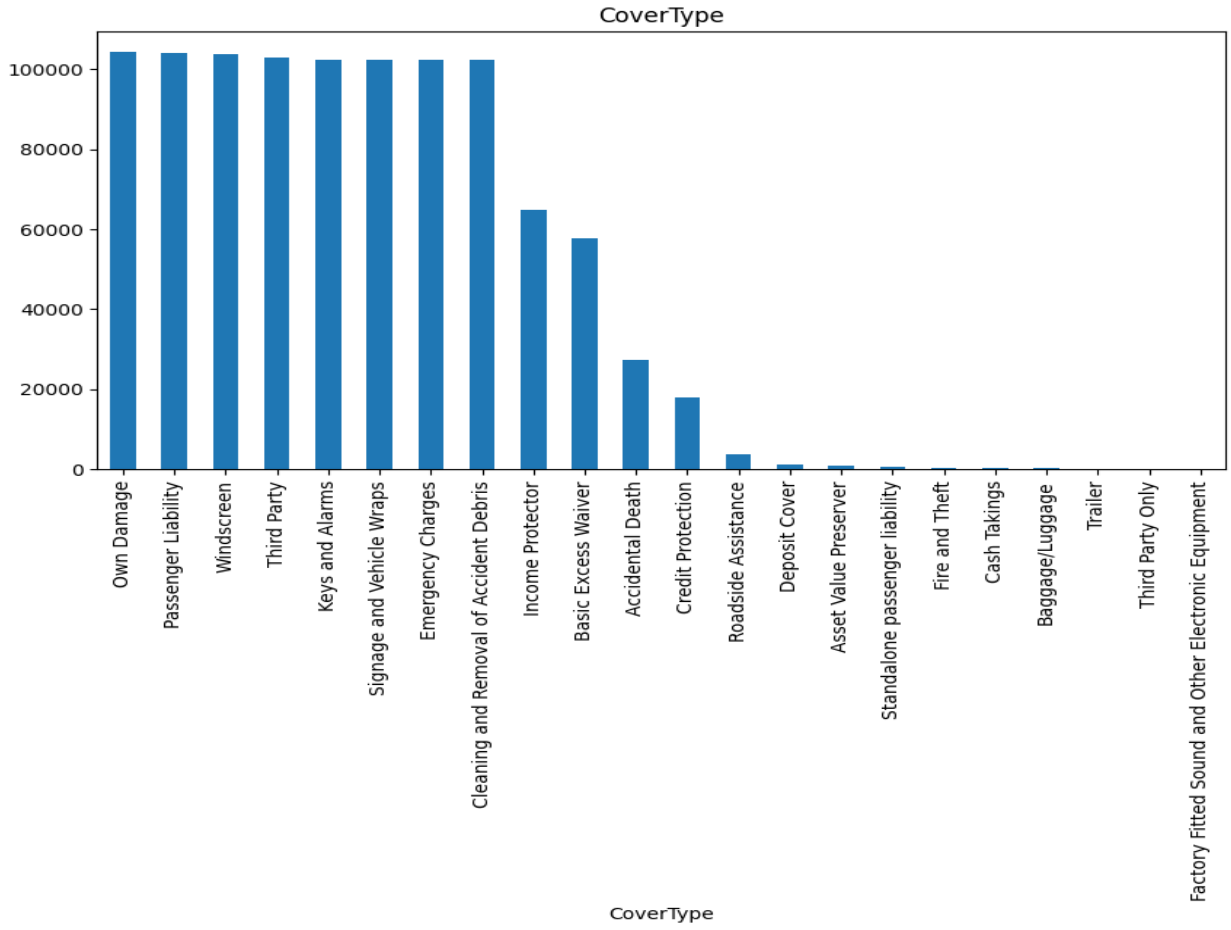

Converted
1e6
1.0
0.8
0.6
0.4
0.2
0.0
No
Yes
Converted

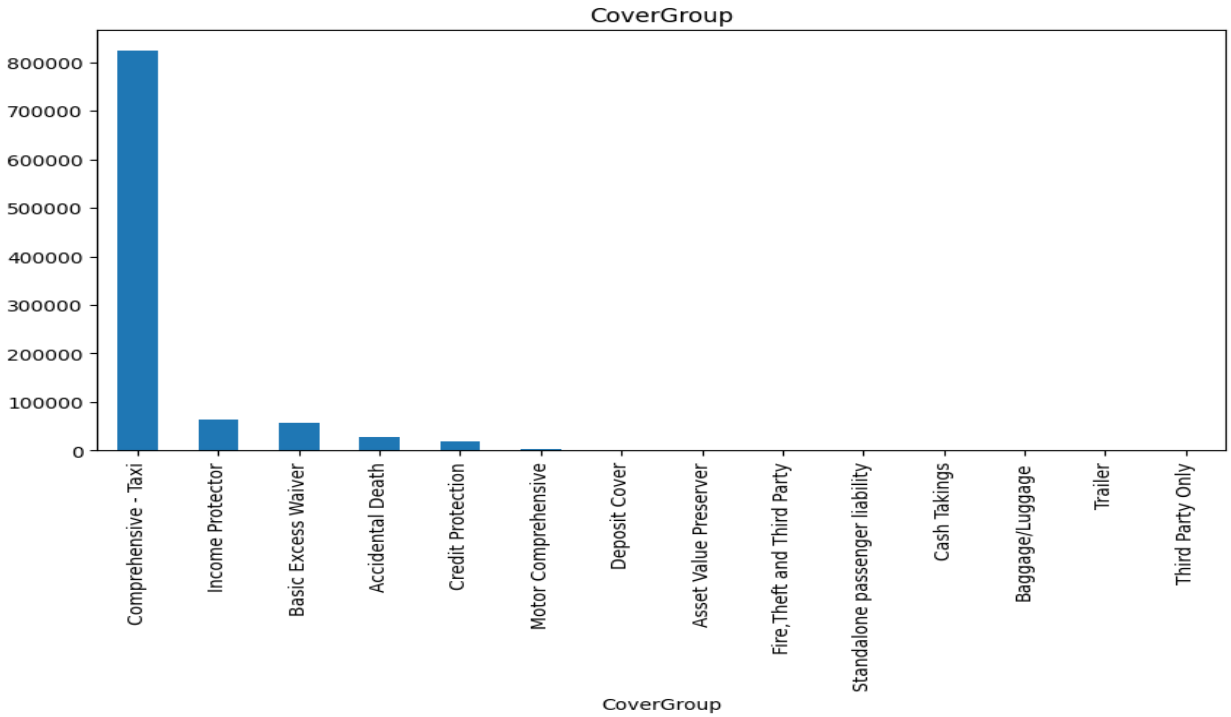

CrossBorder
1e6
1.0
0.8
0.6
0.4
0.2
0.0
No
CrossBorder

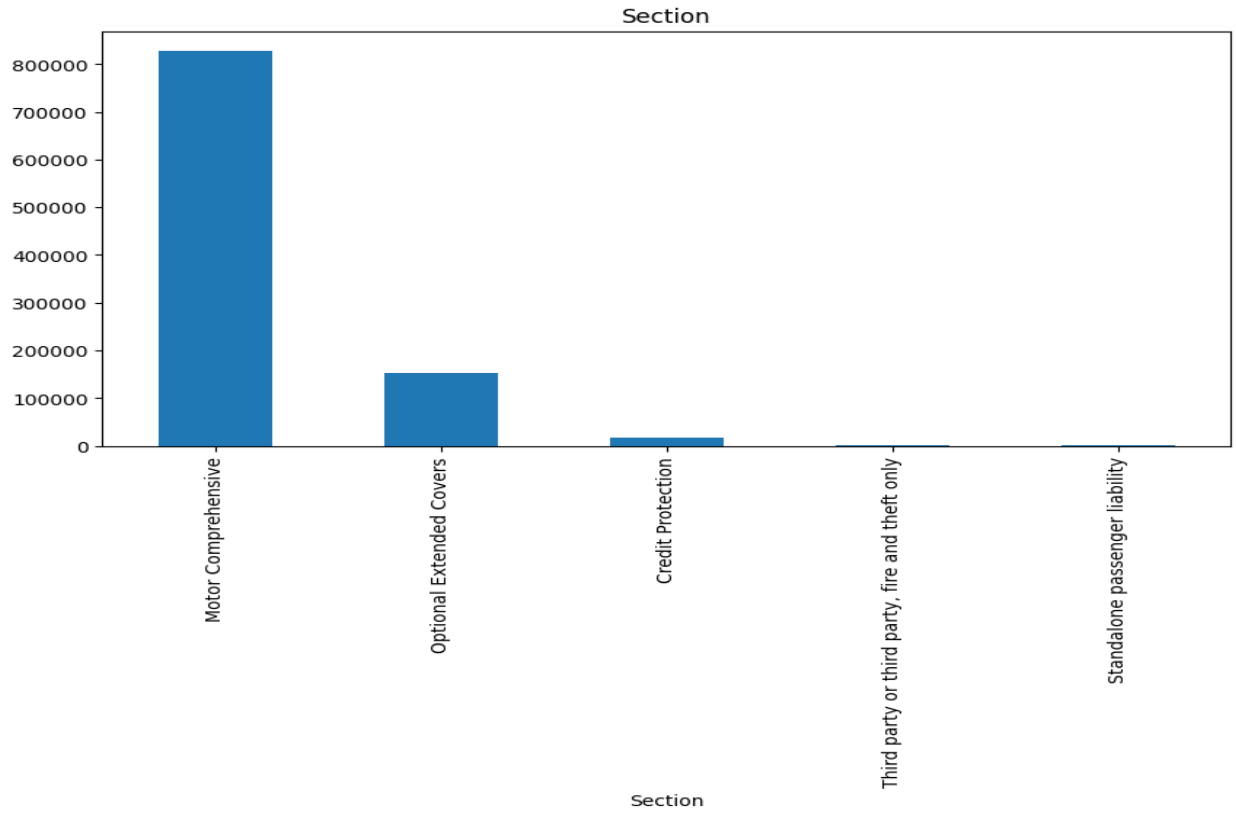

TermFrequency
1e6
1.0
0.8
0.6
0.4
0.2
0.0
Monthly
Annual
TermFrequency

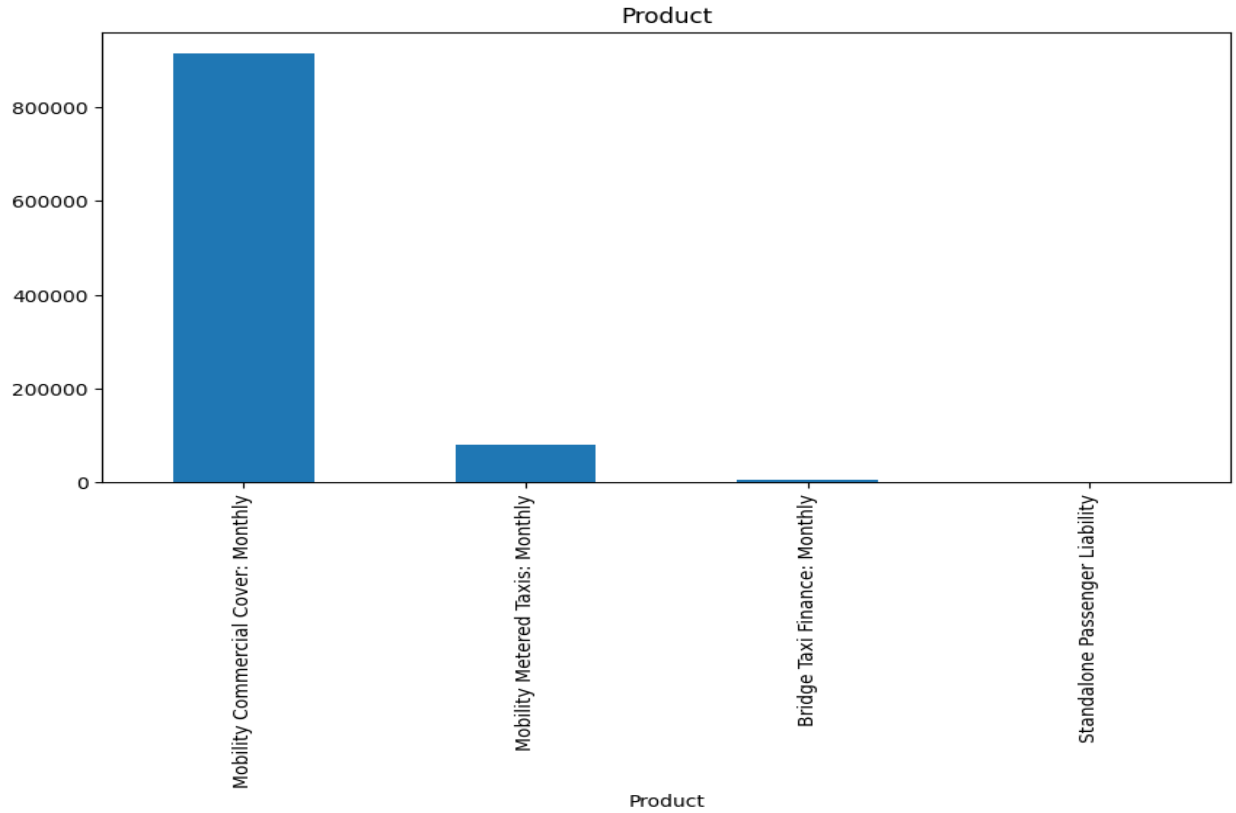

ExcessSelected
800000
700000
600000
500000
400000
300000
200000
100000
0
No excess
Mobility - Windscreen
Mobility - Taxi with value more than R100 000 - R5 000
Mobility - Taxi with value less than R100 000 - R3 000
Mobility - Metered Taxis - R2000
Mobility - Windscreen (Feb2015)
Mobility - Taxi with value more than R100 000 - R5 000 (April)
Mobility - Metered Taxis - R5000
Mobility - Taxi with value more than R100 000 - R10 000
Mobility - R250
Mobility - Trailers
Mobility - Taxi with value more than R100 000 - R7 500
Mobility - Factory Sound
ExcessSelected

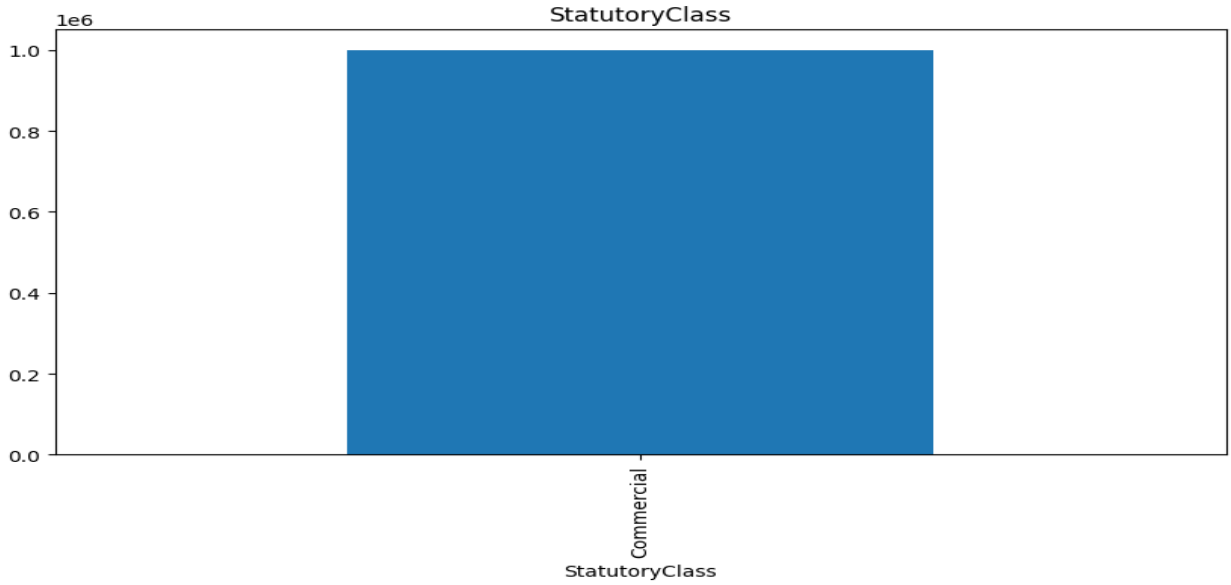

CoverCategory
100000
80000
60000
40000
20000
0
Passenger Liability
Third Party
Keys and Alarms
Signage and Vehicle Wraps
Emergency Charges
Cleaning and Removal of Accident Debris
Windscreen
Own Damage
Income Protector
Basic Excess Waiver
Accidental Death
Credit Protection
Own Damage.
Own damage
Windscreen (2015)
Income Protector (2015)
Basic Excess Waiver (2015)
Roadside Assistance
Vehicle over R100 000
Deposit Cover
Asset Value Preserver
Standalone passenger liability
Cash Takings
Baggage/Luggage
Vehicle under R100 000
Trailer
Third Party Only
Factory Fitted Sound and Other Electronic Equipment
CoverCategory

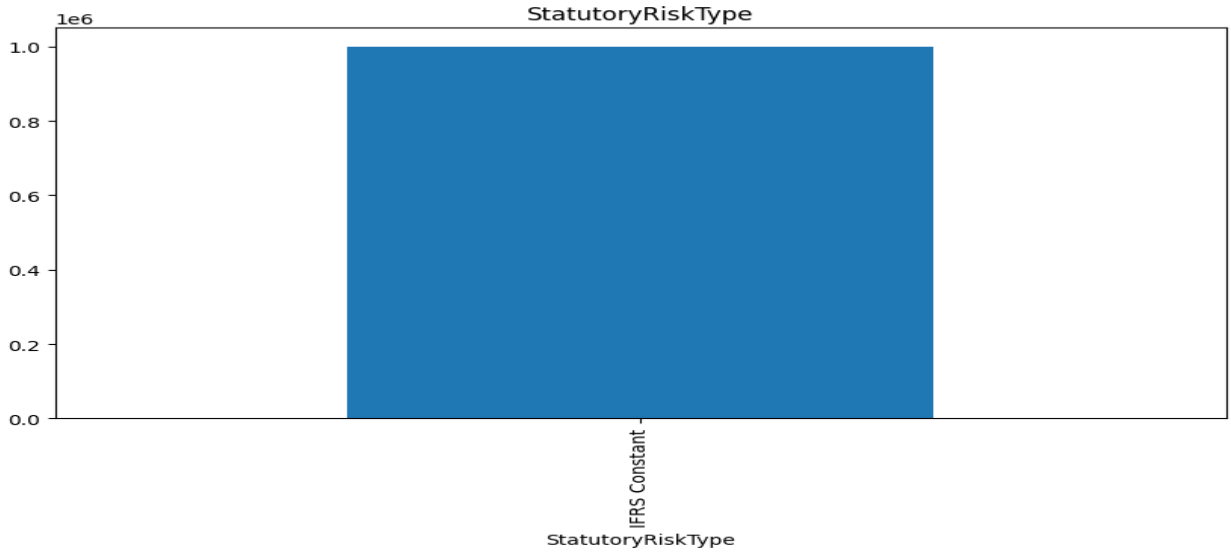

CoverType
100000
80000
60000
40000
20000
0
Own Damage
Passenger Liability
Windscreen
Third Party
Keys and Alarms
Signage and Vehicle Wraps
Emergency Charges
Cleaning and Removal of Accident Debris
Income Protector
Basic Excess Waiver
Accidental Death
Credit Protection
Roadside Assistance
Deposit Cover
Asset Value Preserver
Standalone passenger liability
Fire and Theft
Cash Takings
Baggage/Luggage
Trailer
Third Party Only
Factory Fitted Sound and Other Electronic Equipment
CoverType


CoverGroup
800000
700000
600000
500000
400000
300000
200000
100000
0
Comprehensive - Taxi
Income Protector
Basic Excess Waiver
Accidental Death
Credit Protection
Motor Comprehensive
Deposit Cover
Asset Value Preserver
Fire,Theft and Third Party
Standalone passenger liability
Cash Takings
Baggage/Luggage
Trailer
Third Party Only
CoverGroup


Section
800000
700000
600000
500000
400000
300000
200000
100000
0
Motor Comprehensive
Optional Extended Covers
Credit Protection
Third party or third party, fire and theft only
Standalone passenger liability
Section


Product
800000
600000
400000
200000
0
Mobility Commercial Cover: Monthly
Mobility Metered Taxis: Monthly
Bridge Taxi Finance: Monthly
Standalone Passenger Liability
Product

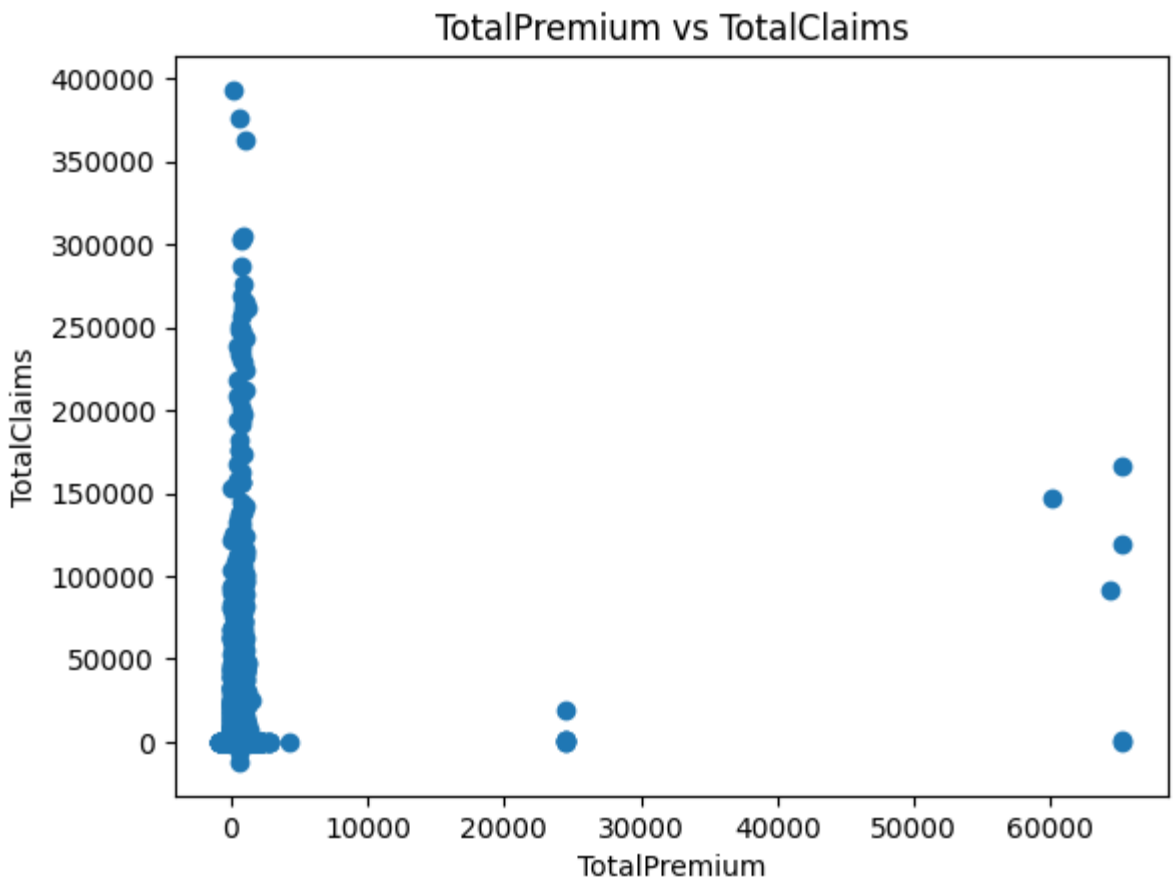



## Bivariate Analysis

**Scatter plots for relationships between numerical columns:**

```
# scatter plot: TotalPremium vs TotalClaims
plt.scatter(data['TotalPremium'], data['TotalClaims'])
plt.xlabel('TotalPremium')
plt.ylabel('TotalClaims')
plt.title('TotalPremium vs TotalClaims')
plt.show()
```

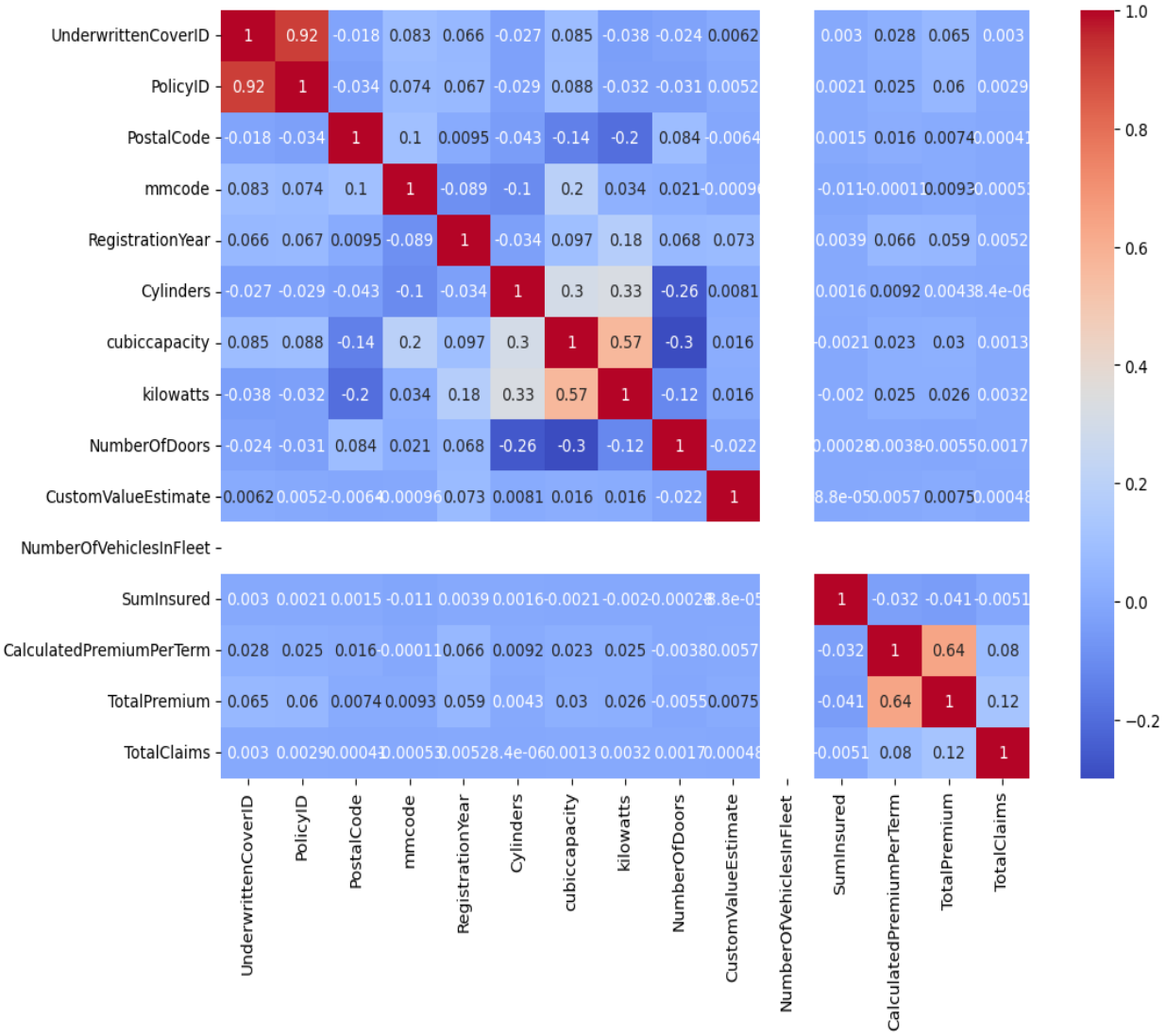



**Correlation matrix:**

```
# Select only numerical columns
numerical_data = data.select_dtypes(include=[float, int])

# Compute the correlation matrix
corr_matrix = numerical_data.corr()

# Plot the heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm')
plt.show()
```

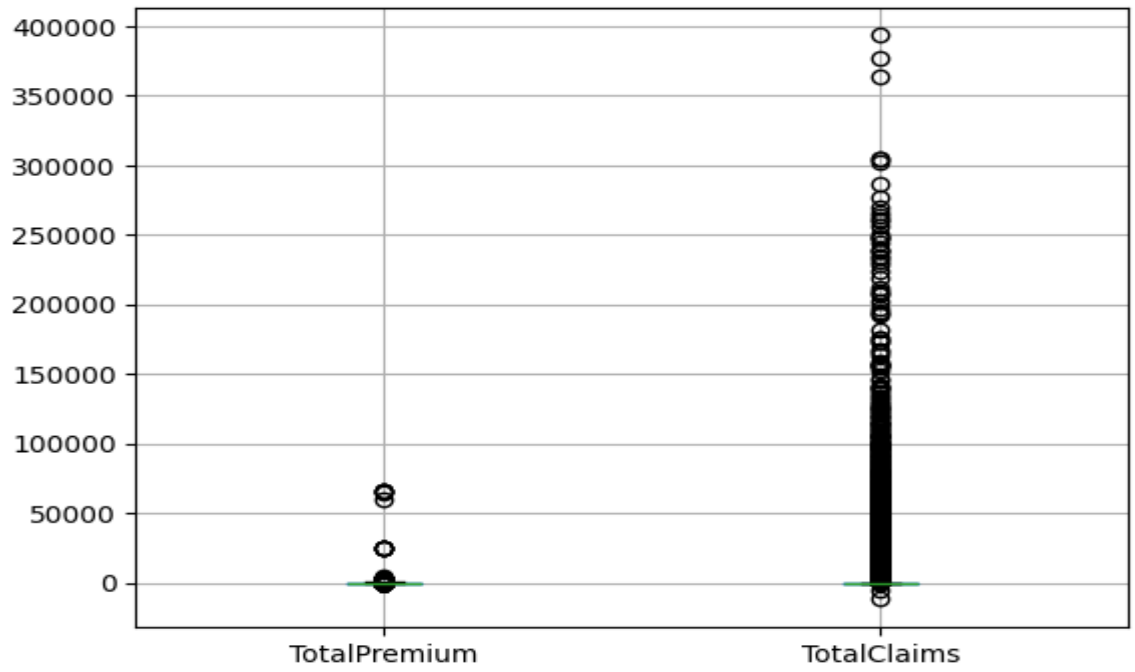



Insight: This heatmap provides a comprehensive view of how different features relate to each other, identifying key drivers for insurance claims and premiums.

## Outlier Detection

**Box plots for outlier detection:**

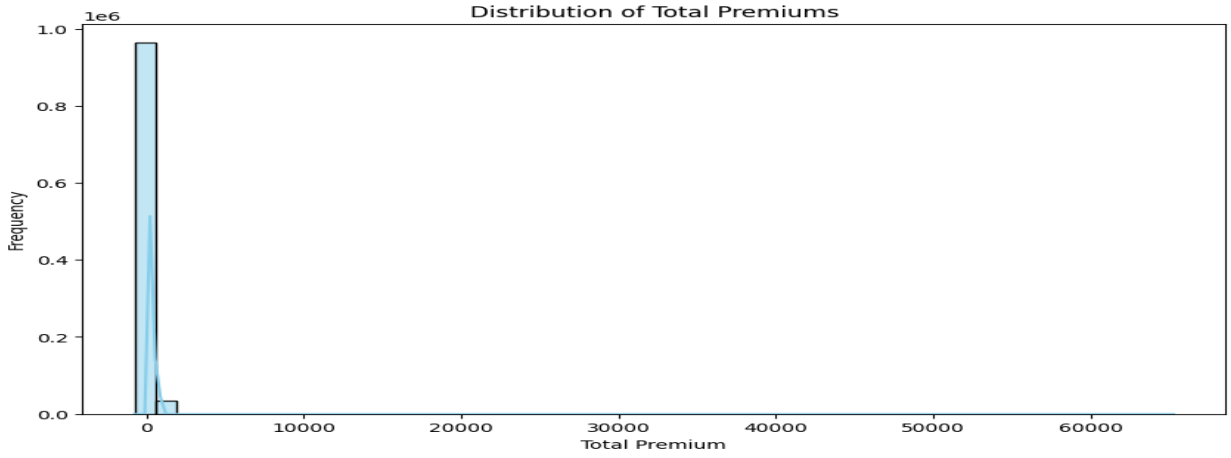
```
# Plot box plots for numerical columns to detect outliers
numerical_data.boxplot(column=['TotalPremium', 'TotalClaims'])
plt.show()
```




## Creating insightful and aesthetically pleasing visualizations

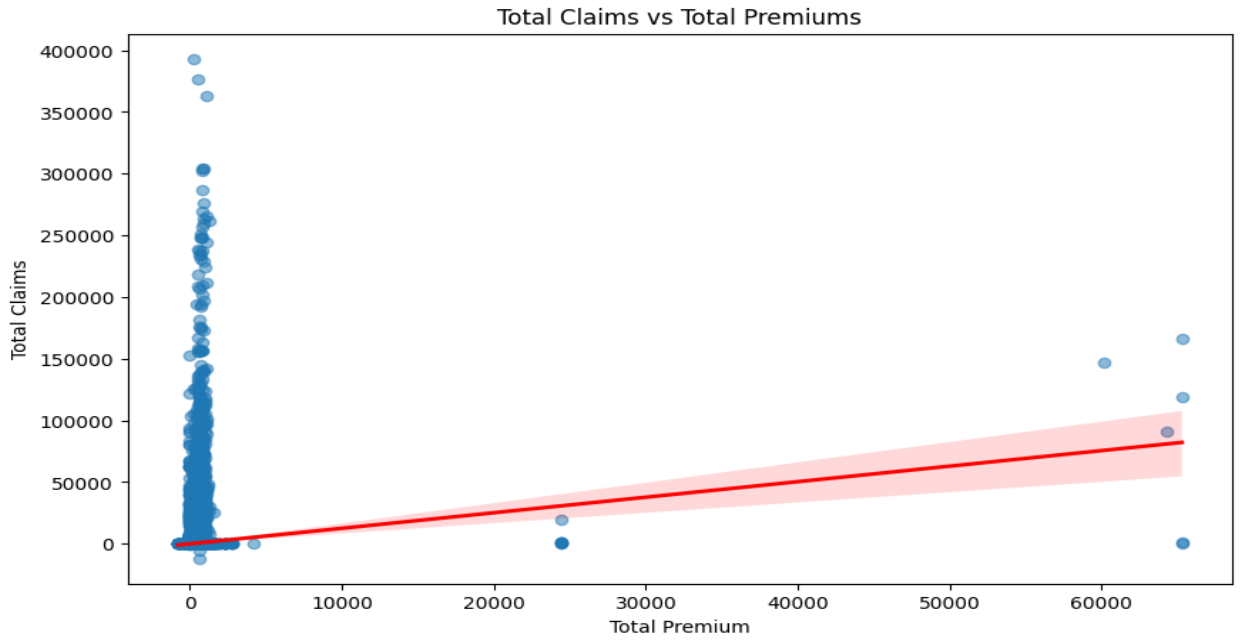
### Distribution of Total Premiums

```
# Distribution plot of Total Premium
plt.figure(figsize=(10, 6))
sns.histplot(data['TotalPremium'], bins=50, kde=True, color='skyblue')
plt.title('Distribution of Total Premiums')
plt.xlabel('Total Premium')
plt.ylabel('Frequency')
plt.show()
```



## Relationship Between Total Claims and Total Premiums

```
# Scatter plot with regression line for Total Claims vs Total Premiums
plt.figure(figsize=(10, 6))
sns.regplot(x='TotalPremium', y='TotalClaims', data=data,
scatter_kws={'alpha':0.5}, line_kws={'color':'red'})
plt.title('Total Claims vs Total Premiums')
plt.xlabel('Total Premium')
plt.ylabel('Total Claims')
plt.show()
```

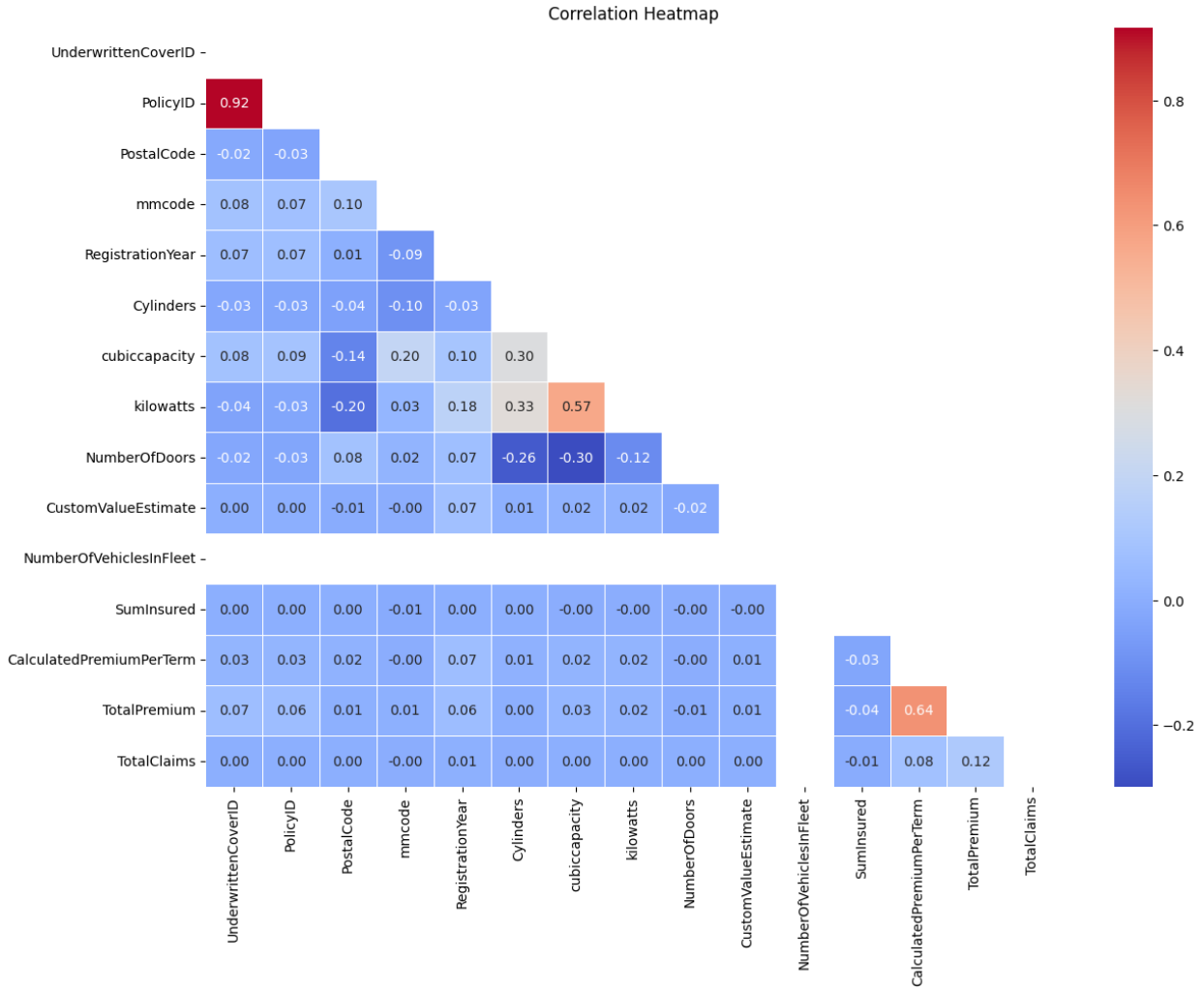



## Correlation Heatmap with Highlighted Key Variables

```
# Compute the correlation matrix
corr_matrix = data.select_dtypes(include=[float, int]).corr()

# Create a mask to display only the lower triangle of the heatmap
mask = np.triu(np.ones_like(corr_matrix, dtype=bool))

# Plot the heatmap with the mask
plt.figure(figsize=(14, 10))
sns.heatmap(corr_matrix, mask=mask, annot=True, fmt='.2f', cmap='coolwarm',
linewidths=0.5)
plt.title('Correlation Heatmap')
plt.show()
```




Insight: This heatmap provides a comprehensive view of how different features relate to each other, identifying key drivers for insurance claims and premiums.

## Key Insights

**1. Policy Characteristics:**

The dataset includes a variety of policy types and coverages, predominantly for newer vehicles.

A significant number of policies are related to comprehensive coverage for commercial vehicles like taxis.

**2. Vehicle Specifications:**

Most vehicles are standard configurations with common attributes like 4 cylinders and 4 doors.

High-value vehicles are less common but have a significant impact on the overall premium and claims amounts.

**3. Geographical Distribution:**

Policies are spread across various postal codes, indicating a wide customer base.

Certain regions may have higher concentrations of policyholders, which could be targeted for specific marketing strategies.

**4. Financial Metrics:**

Premiums and claims data suggest that while most policies are low-risk with minimal claims, there are outliers with high premiums and claims, indicating potential high-risk segments.

## Conclusion

Through this exploratory data analysis, we have gained several key insights:

- The distribution of total premiums reveals common premium amounts and potential outliers.
- There is a noticeable relationship between total premiums and total claims, indicating that higher premiums might lead to higher claims.
- The correlation heatmap uncovers hidden relationships between various features, which can guide further analysis and model building.

This analysis not only helps in understanding the current data but also sets the stage for predictive modeling and strategic decision-making to improve marketing strategies and optimize insurance products.

The EDA provided valuable insights into the vehicle insurance dataset. Key findings include:

The predominance of newer vehicles and comprehensive coverage policies.

Common vehicle configurations leading to predictable premium and claim amounts.

Geographical diversity among policyholders.

## Recommendations

**1. Targeted Marketing:**

Focus marketing efforts on regions with high concentrations of policyholders to maximize outreach and engagement.

Consider specialized campaigns for high-value vehicle owners to address their specific insurance needs.

**2. Risk Management:**

Develop strategies to mitigate risks associated with high-premium and high-claim policies.

Implement more stringent underwriting criteria for high-risk segments identified in the dataset.

**3. Product Development:**

Explore opportunities to create new insurance products tailored to common vehicle configurations and customer profiles.

Enhance coverage options for commercial vehicles, particularly in regions with high taxi densities.

**4. Data Quality Improvement:**

Address data quality issues, particularly the high proportion of missing values in critical columns like `NumberOfVehiclesInFleet` and `CustomValueEstimate`.

Implement better data collection and validation processes to ensure completeness and accuracy.

Through this analysis, we have laid the groundwork for deeper investigations and strategic initiatives that can drive business growth and improve customer satisfaction in the vehicle insurance sector.

## Next Steps

To build on these insights, future steps could include:

- Developing predictive models to forecast claims based on premiums and other features.
- Conducting deeper analysis on customer demographics and their impact on insurance claims.
- Implementing advanced feature engineering techniques to enhance model performance.

By continually refining our data analysis techniques and leveraging advanced machine learning models, we can transform raw data into valuable business insights.