8/3/2020

Data Visualization using Tableau:

Car Claims Insurance

Tahmid Bari
PGP-DSBA GREAT LEARNING – THE UNIVERSITY OF TEXAS AT AUSTIN (MCCOMBS SCHOOL OF BUSINESS)

Contents

Project Objective

This project requires to prepare Data Visualization Story of Car Insurance Key Insights and Finding based of Car Claim Insurance dataset provided.

Dataset Overview

Few of the details of Car Claim Insurance dataset provided for analysis is captured in the table below

ID	ID OF CUSTOMER
KIDSDRIV	Total No of kids the car driver has
BIRTH	Date of birth
HOMEKIDS	No of kids the driver has at home
YOJ	Year in which the car owner applied for the car insurance policy
INCOME	Income
PARENT1	If car owner has his parents
HOME_VAL	Value of the house owned
MSTATUS	Marital status
GENDER	Gender
EDUCATION	Education level
OCCUPATION	Occupation level
TRAVTIME	Travel time taken in minutes (On an average)
CAR_USE	Purpose of using the car
BLUEBOOK	What is the worth of the car
CAR_TYPE	Car type
OLDCLAIM	Previous claim
CLM_FREQ	How often claimed
CLM_AMT	Currently Claimed amount
CAR_AGE	Age of car
URBANICITY	Where the car is being driven primarily

Dashboard Link

Please click here to view the Tableau Dashboard.

https://public.tableau.com/profile/tahmid.bari#!/vizhome/DVTCar_Claims_InsuranceProjectbyTahmidBari/CarsClaimStory

Summary

The Tableau dashboards are intended to analyse Car Insurance Claims in Detail and to support further decision making in providing better Customer Care.

This project is an open-ended problem. We need to analyze different variables impacting high or low claim amount and claim frequency and build a story accordingly.

It is assumed that the year in which the car owner applied for the car insurance policy is current year minus the number of total years given in the data set.

Claim amount & Claim Frequency are dependent variables and other variables like gender, Mstatus , Homekids, Car Types , Education, Occupation, Income , Parents, Car Age, YOJ, Parents, Urbancity, Birth etc are independent variables.

We will try to analyse univariate and /or multivariate analysis to identify the relationship and impact of each variable on car insurance claim amount & claim frequency.

We are able to find female customer proportion is slightly higher than men as well as looking at the marital status also show higher proportion of female population, as for education trend there is relatively equal spread between both the Gender.

Women prefer Sports Car and SUVs in both Urban and Rural areas. Panel truck has been most less considered with majority of male driver. No. of different vehicles are much higher in Urban areas compared to the rural areas.

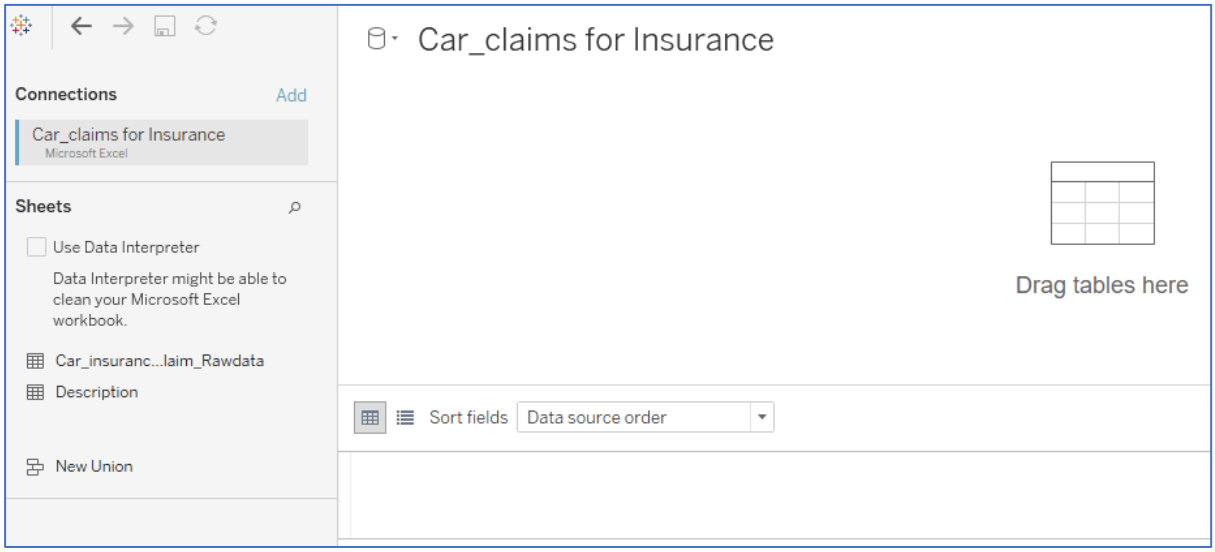
Most of insurance claimed car owners are females and their frequency of claim also high compare with male customers. Irrespective of gender. Also, among the occupation class, blue collars have claimed the maximum. second high are clerical and third high are professional.

SUV car type having one of the least average cost and has the highest claim frequency also majority of SUV owners are females.

Data Visualisation – Step by step Approach

Environment Set up and Data Import

- Tableau Public was used for the analysis of the data set given
- The Data in the excel file was imported to the tableau environment by connecting the source file.



Assumptions

- The Sample size is adequate to perform techniques applicable for Analysis.
- Dataset File to be used for this project is Car_claims for Insurance.xlsx

EDA

In this analysis, objective is to identify key characteristics of the dataset using descriptive analysis methods.

We first import the dataset into tableau from excel. On the data source tab, we notice a lot of columns with null values that may affect our visualizations

Missing Value Treatment

To remove the null values at the dataset level.

Data tab --> Edit data source filters ---> choose columns with null values --> select include non-null values only

Calculated Fields:

While visualizing the data on Tableau, below is our observation

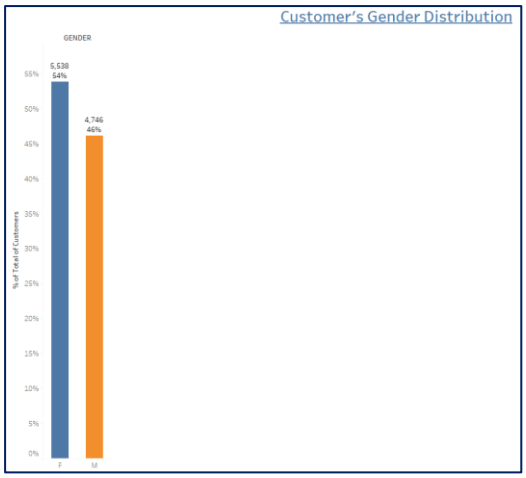
1) YOJ - Year of the person taking the insurance was in an integer - for ex: 8.

We should use the birth year which was given in DD/MM/YYYY format and added this integer to the year field to find out which year the customer has purchased the insurance.

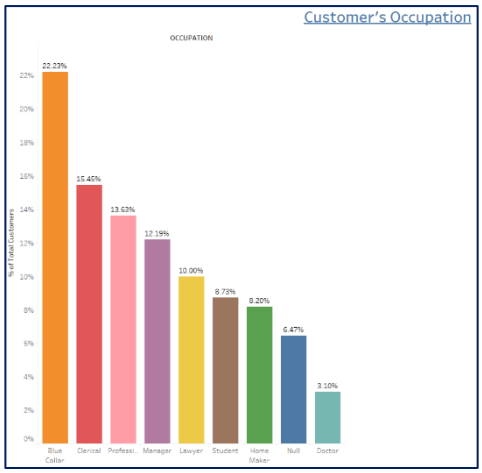
Below is the formula to calculate that.

DATEADD('year',[YOJ],[Birth])

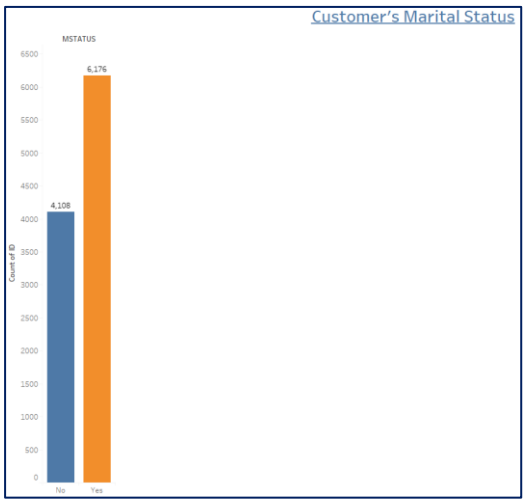
1. Customer's Gender Distribution



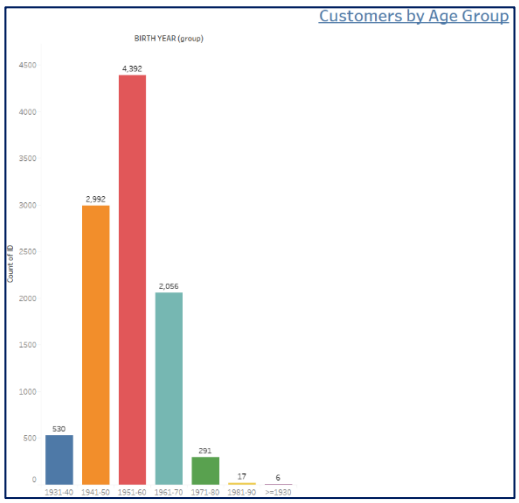
4. Customer's Occupation



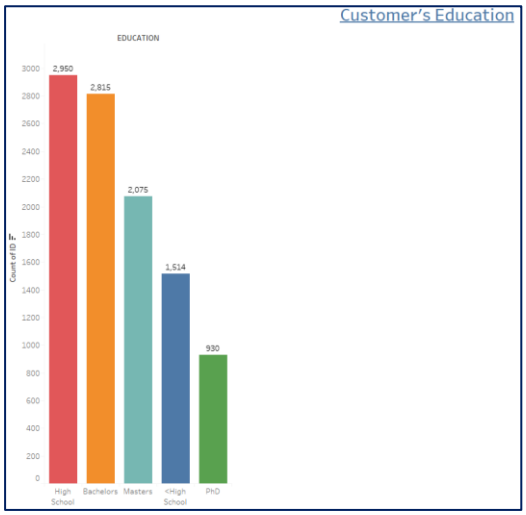
2. Customer's Marital Status



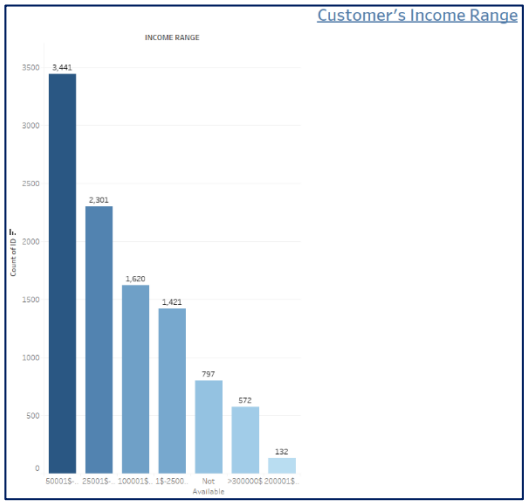
5. Customer's by Age Group



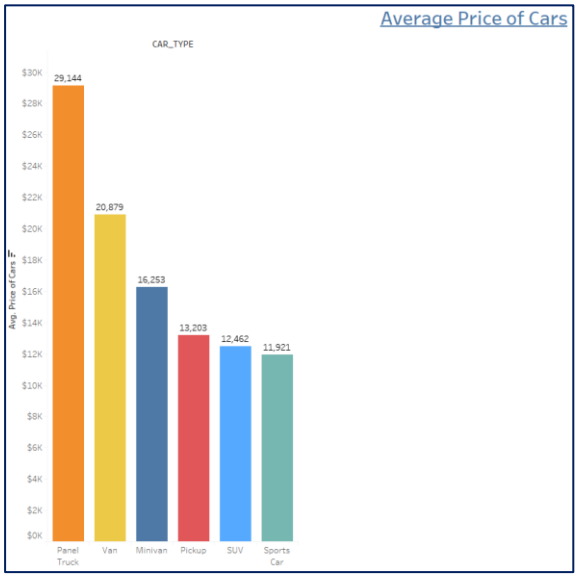
3. Customer's Education



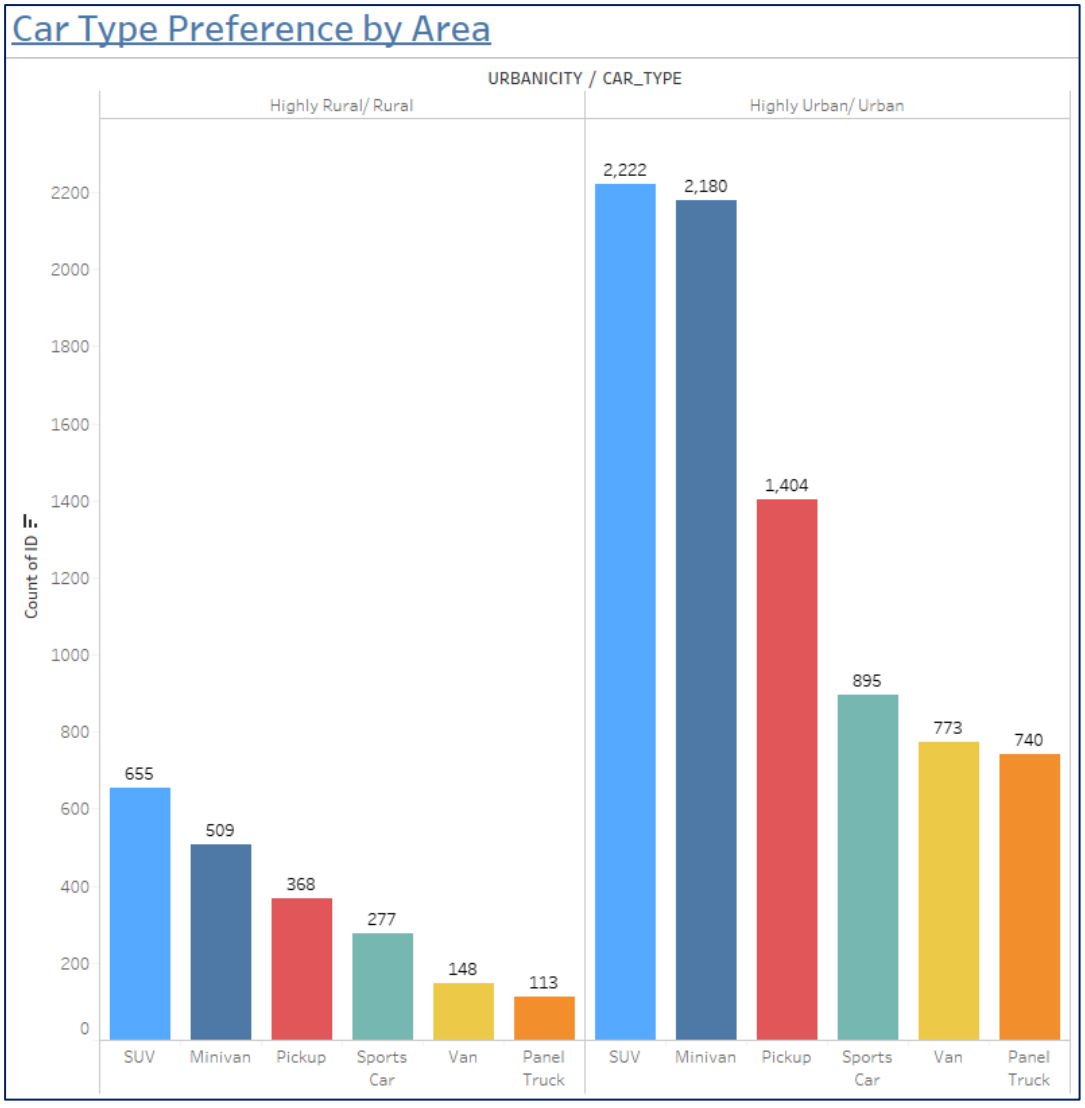
6. Customer's Income Range



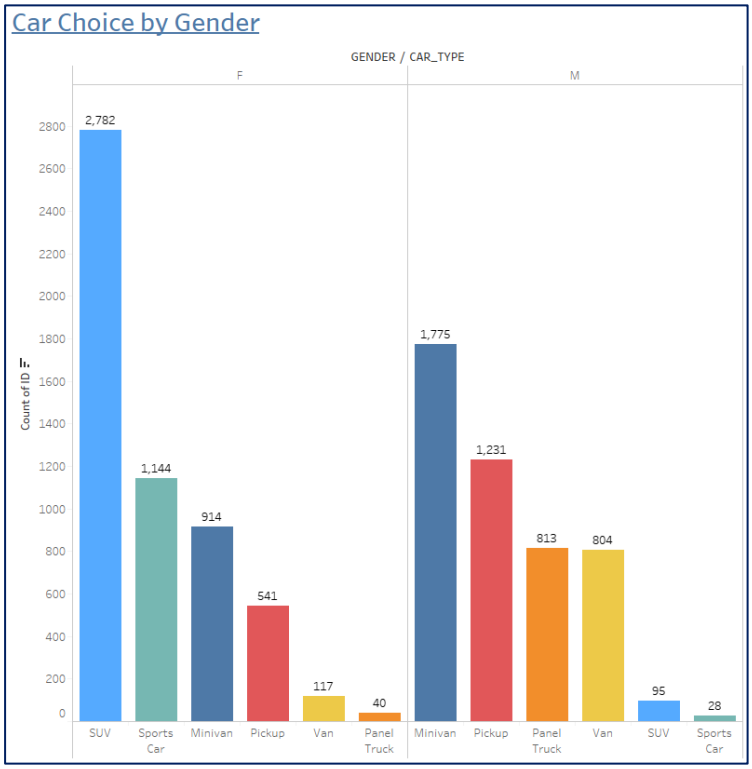
7. Average Price of Cars



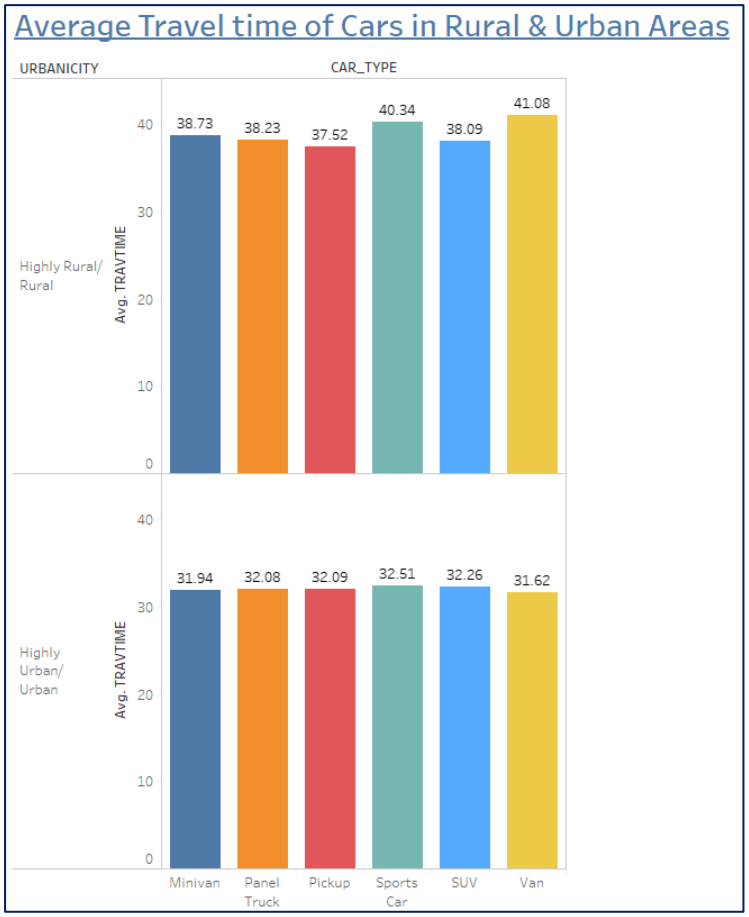
8. Car Type Preference by Area



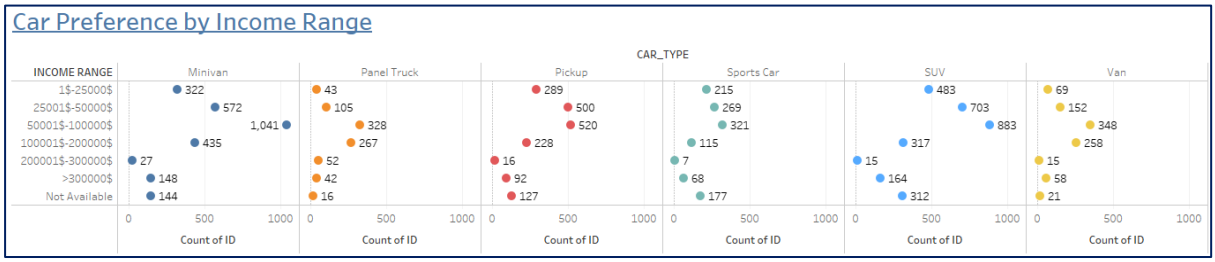
9. Car Choice by Gender



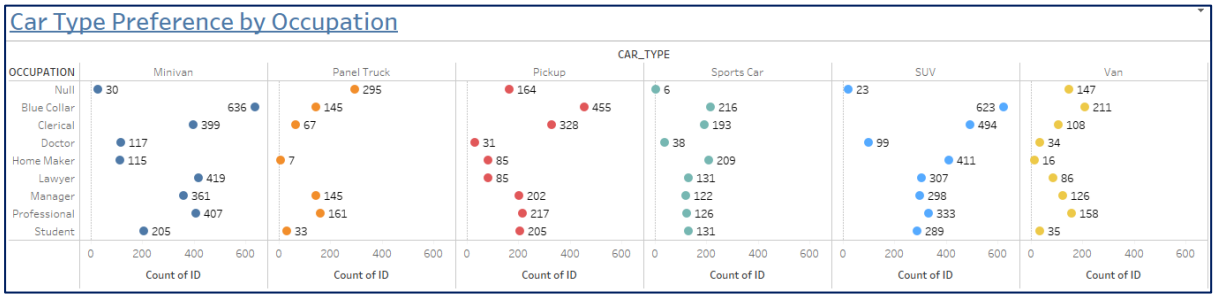
10. Average Travel time of Cars in Rural & Urban Areas



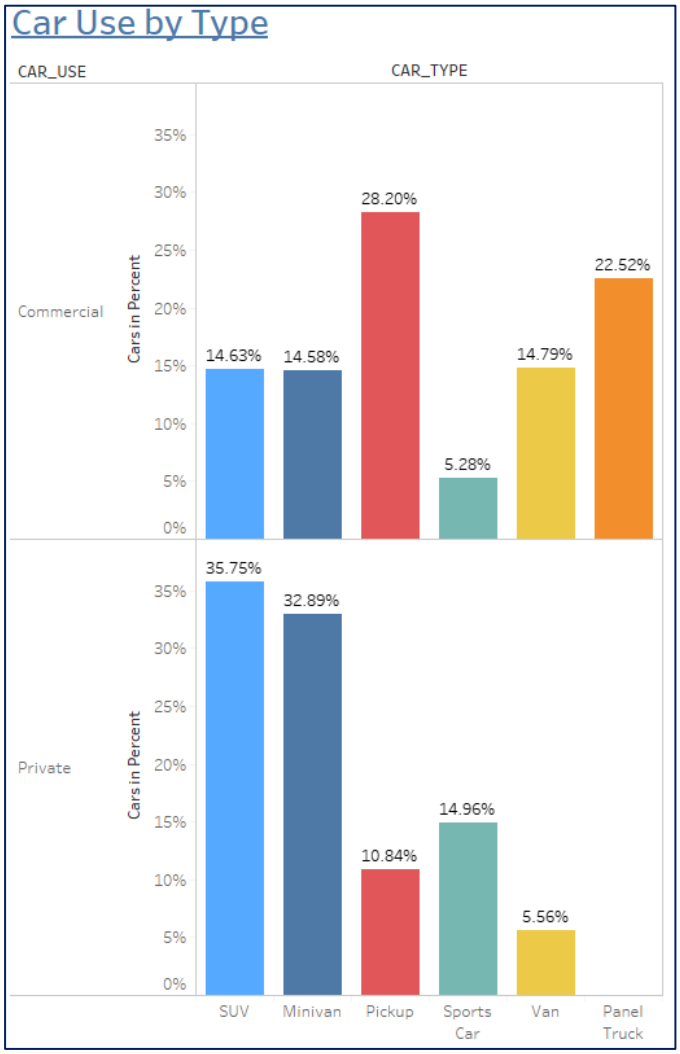
11. Car Preference by Income Range



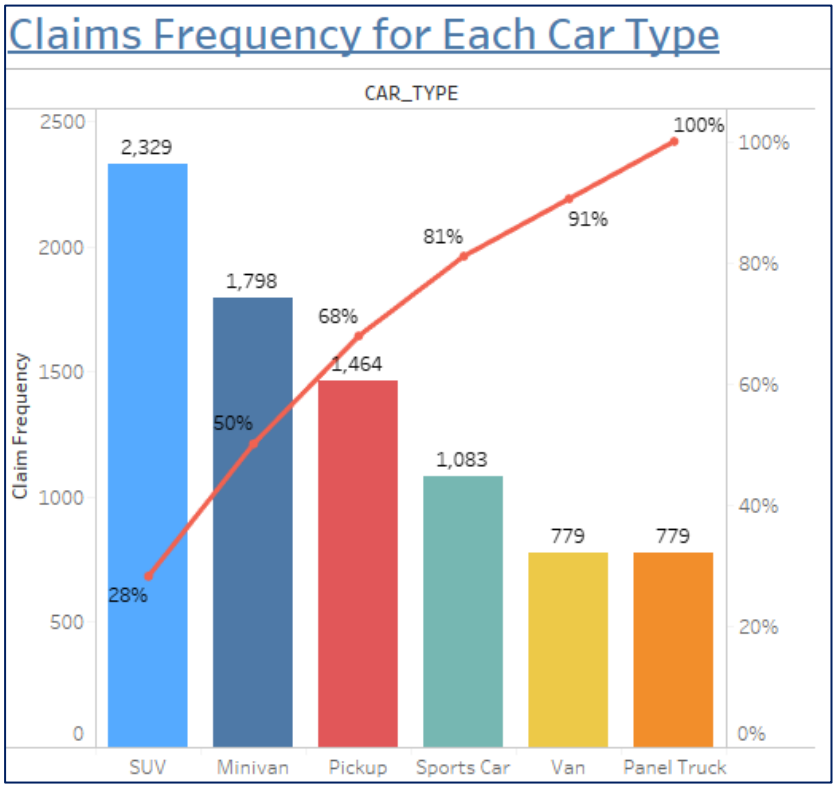
12. Car Type Preference by Occupation



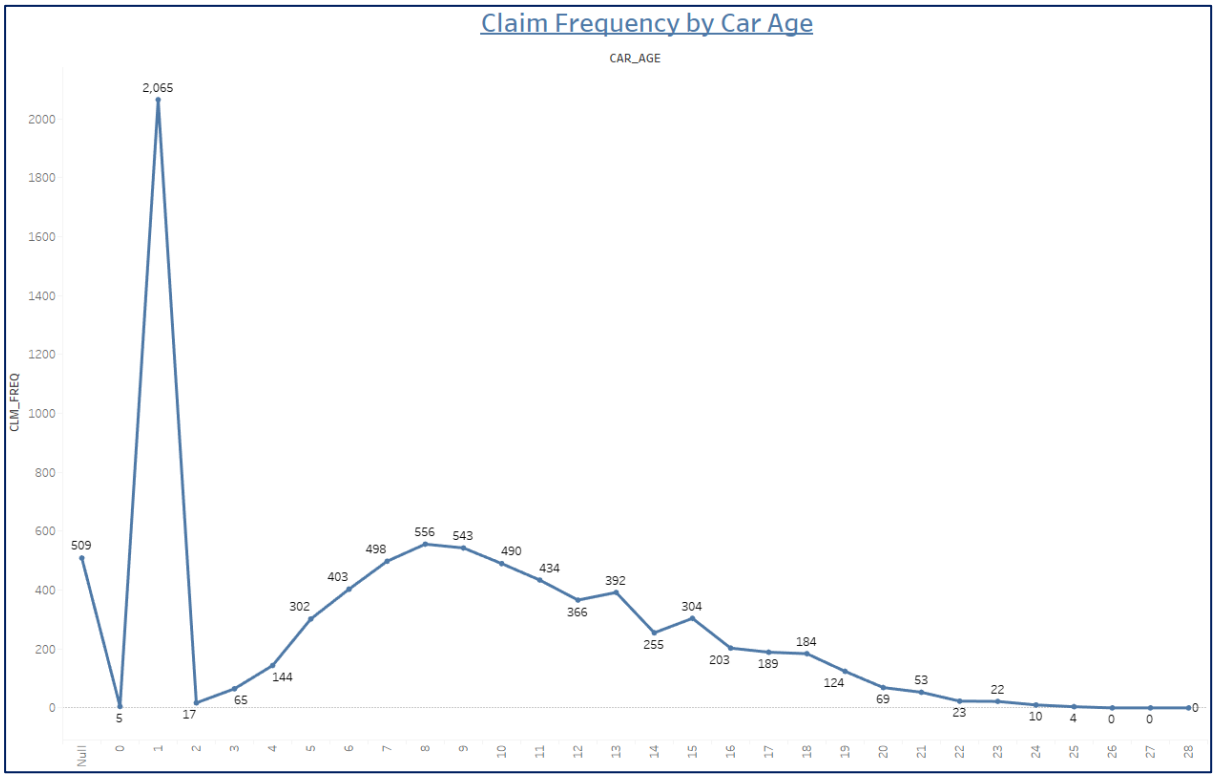
13. Car Use by Type



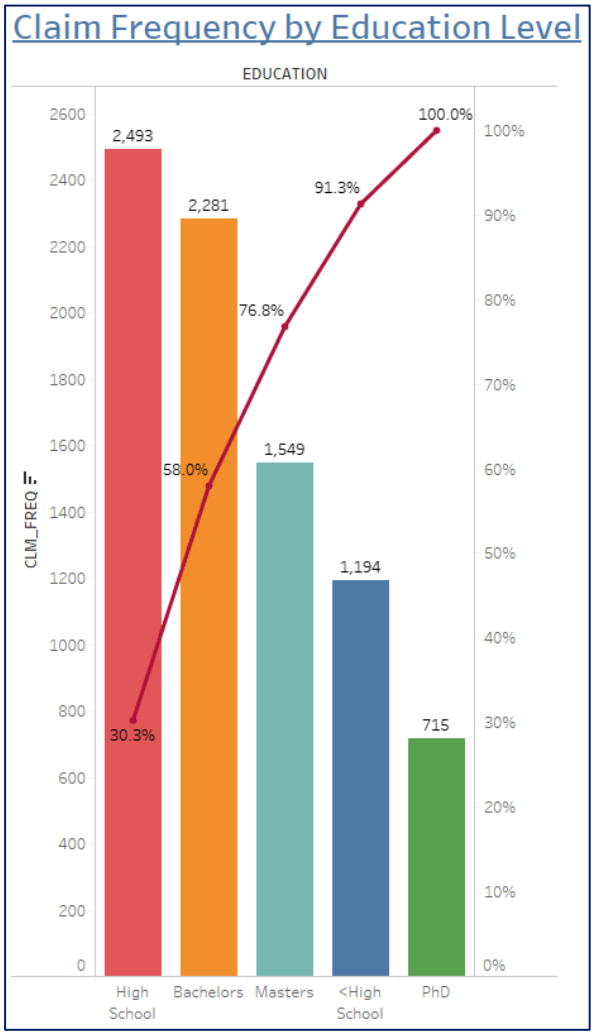
14. Claims Frequency for Each Car Type



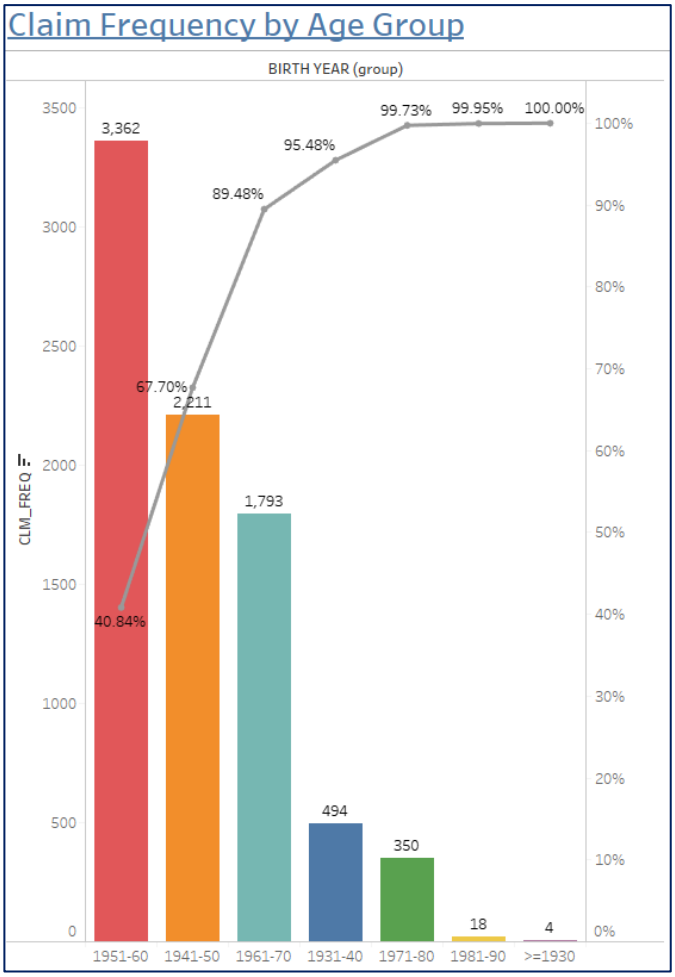
15. Claim Frequency by Car Age



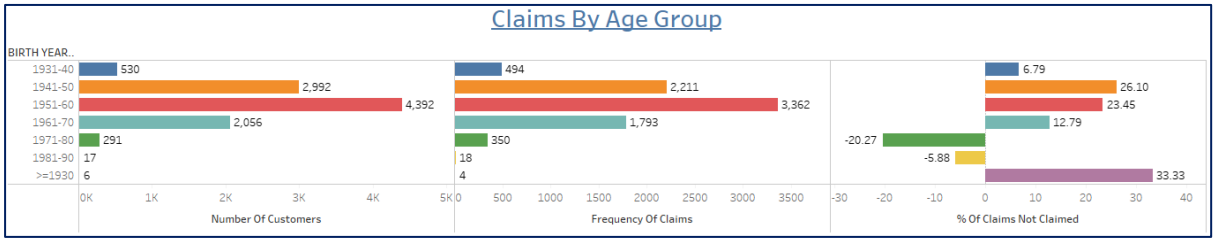
16. Claim Frequency by Education Level



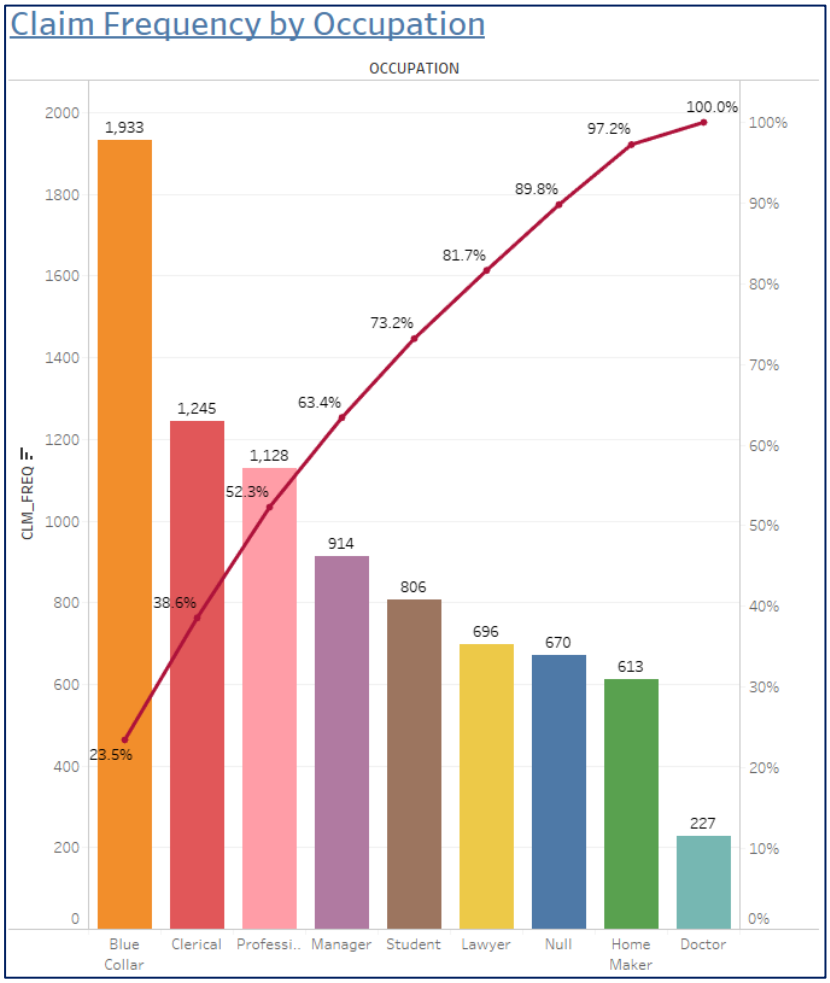
17. Claim Frequency by Age Group



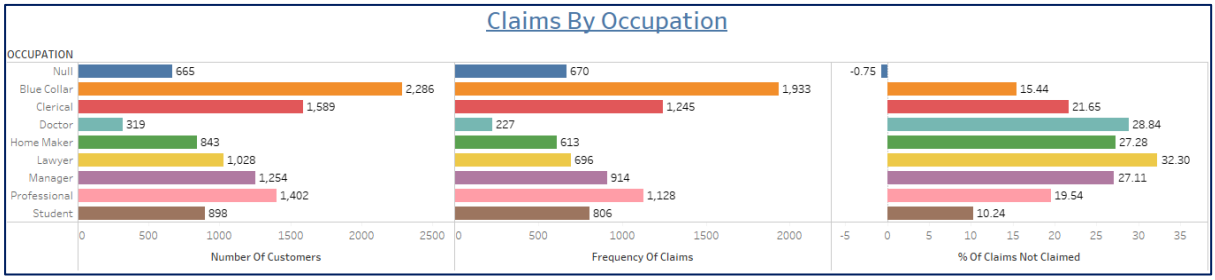
18. Claims by Age Group



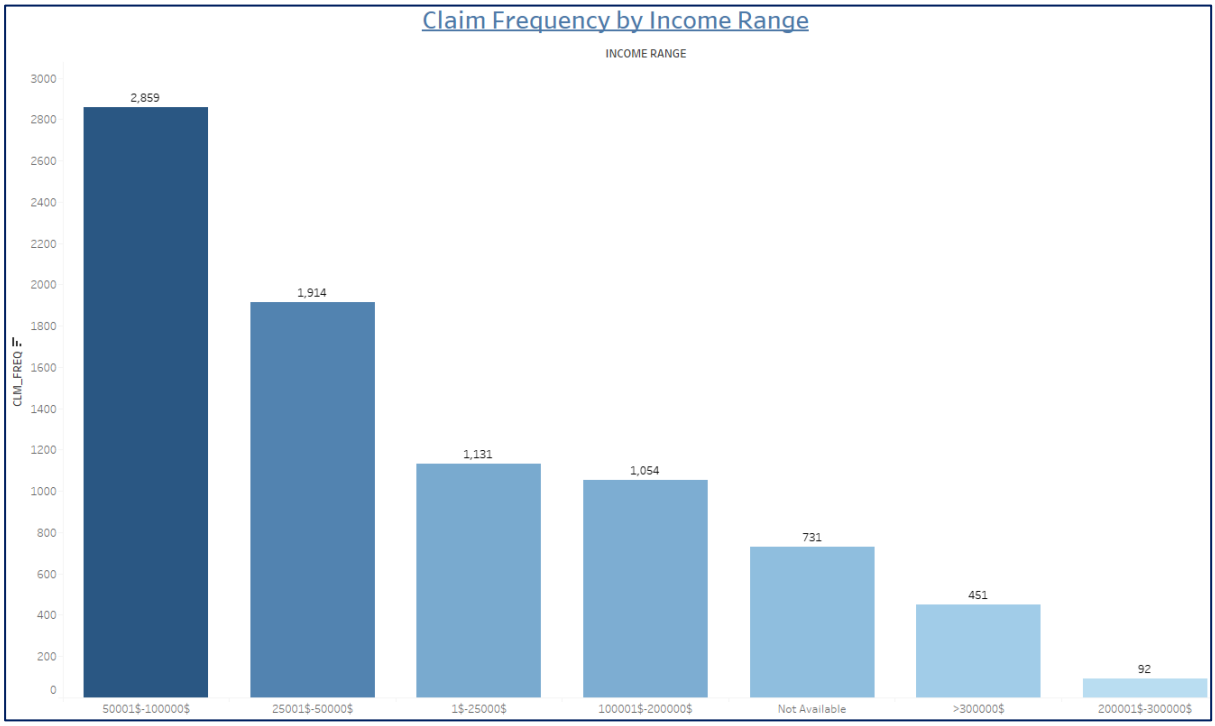
19. Claim Frequency by Occupation



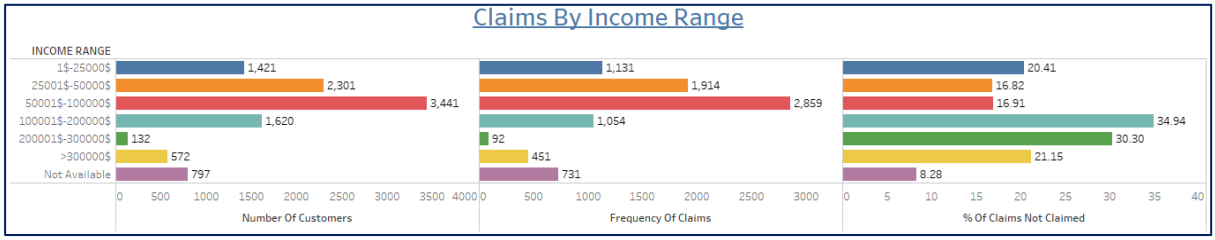
20. Claims by Occupation



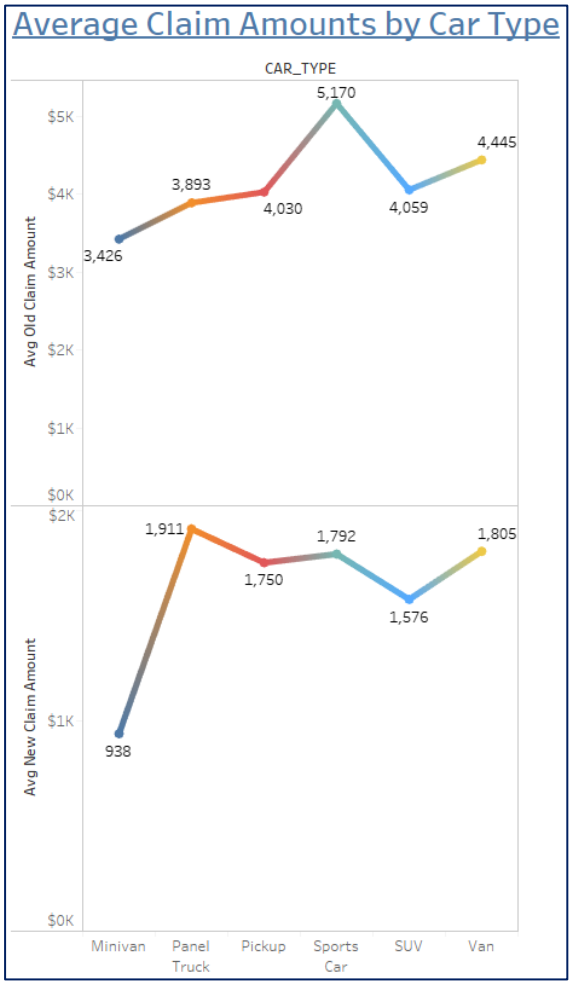
21. Claim Frequency by Income Range



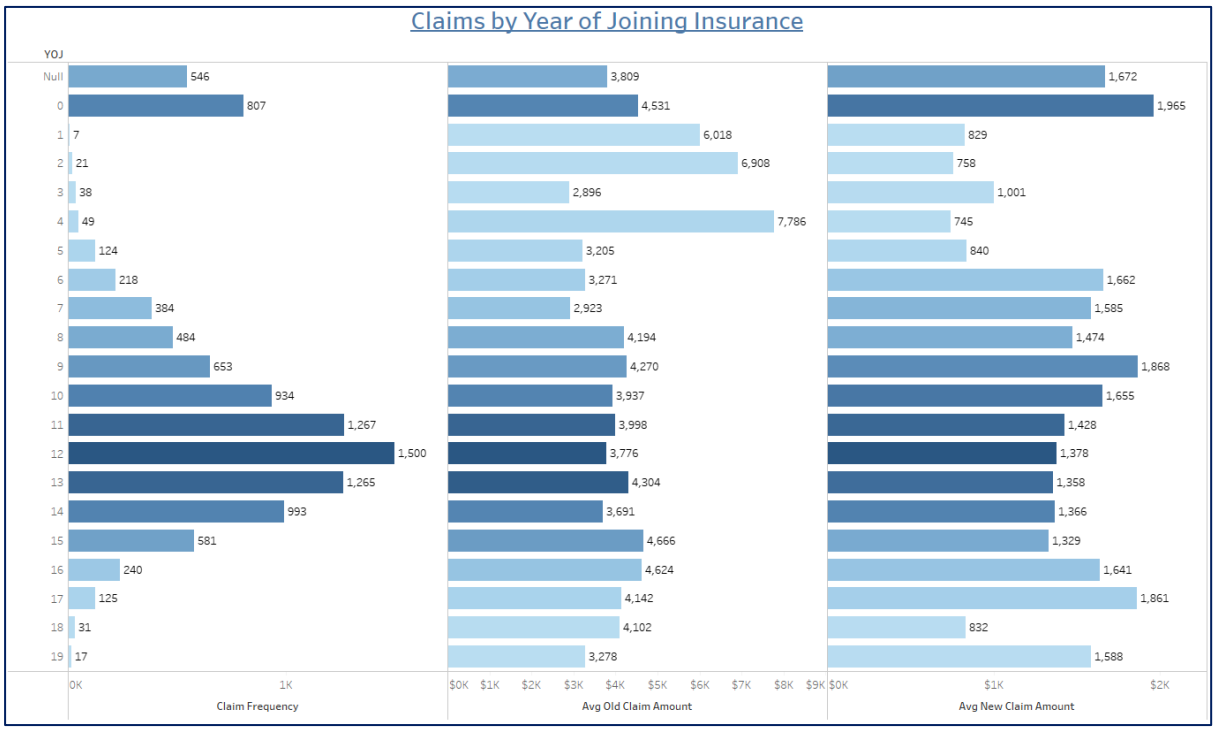
22. Claims by Income Range



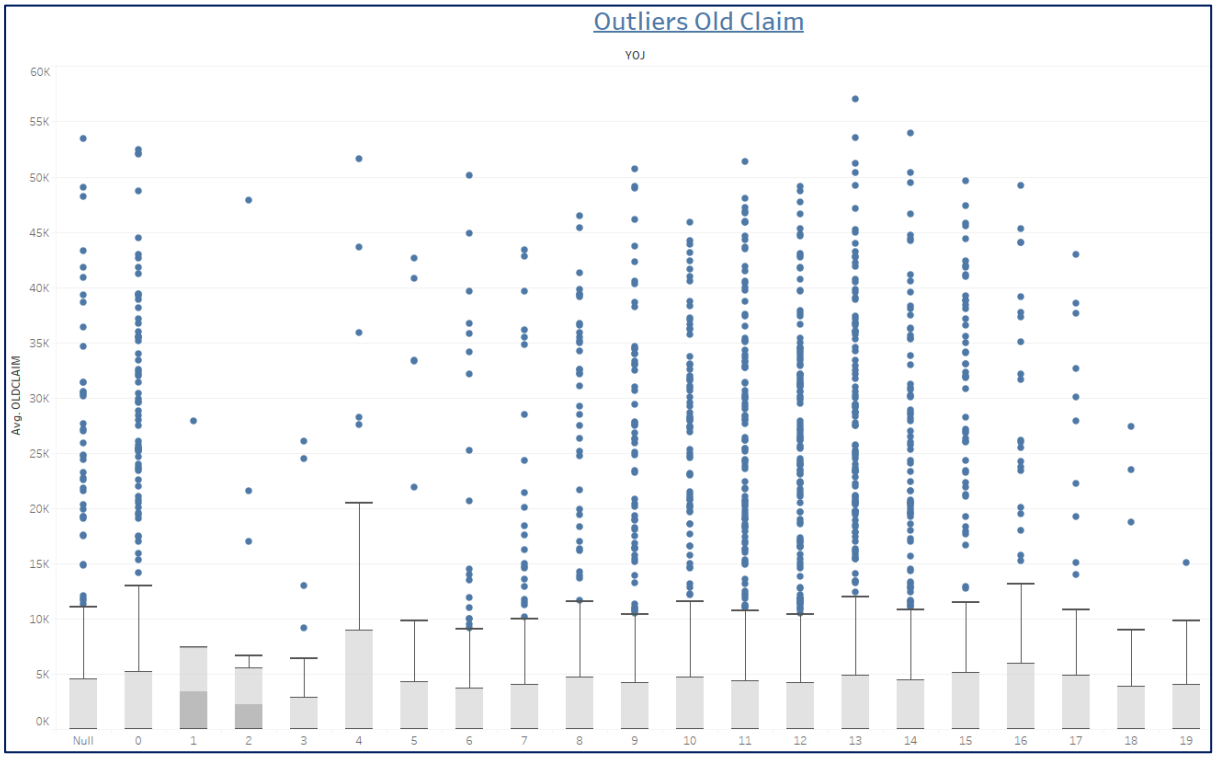
23. Average Claim Amounts by Car Type



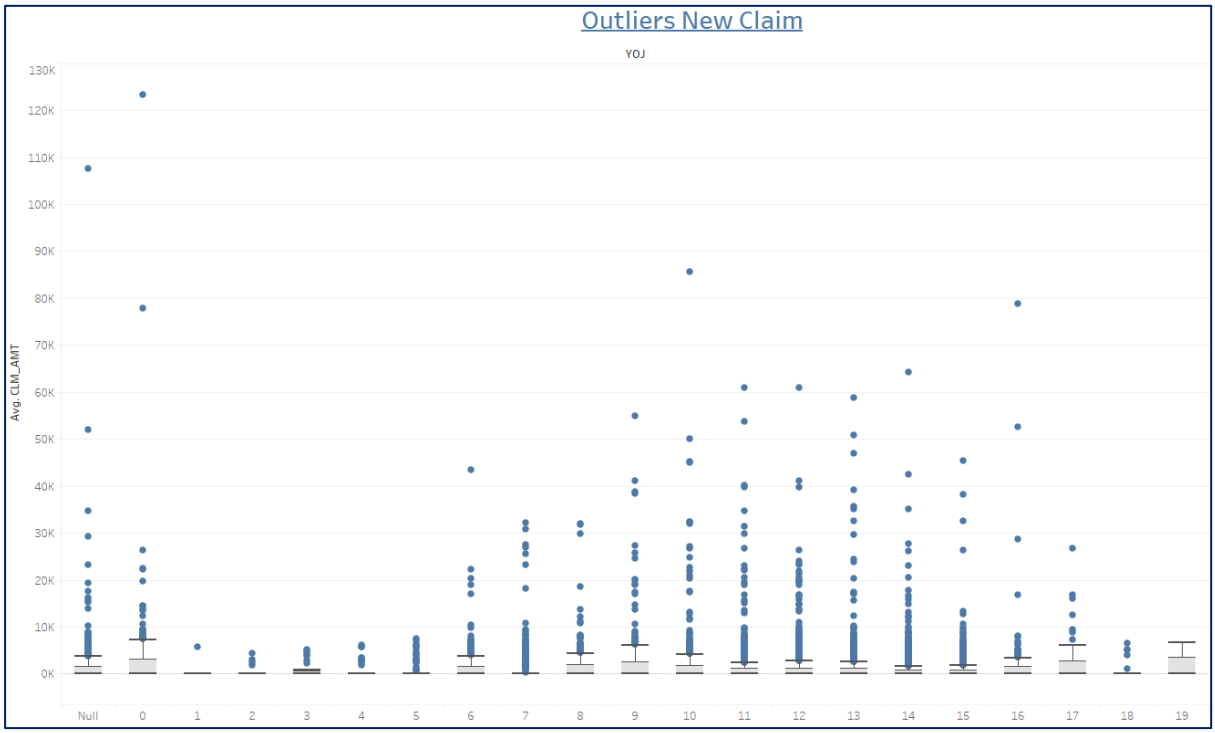
24. Claims by Year of Joining Insurance



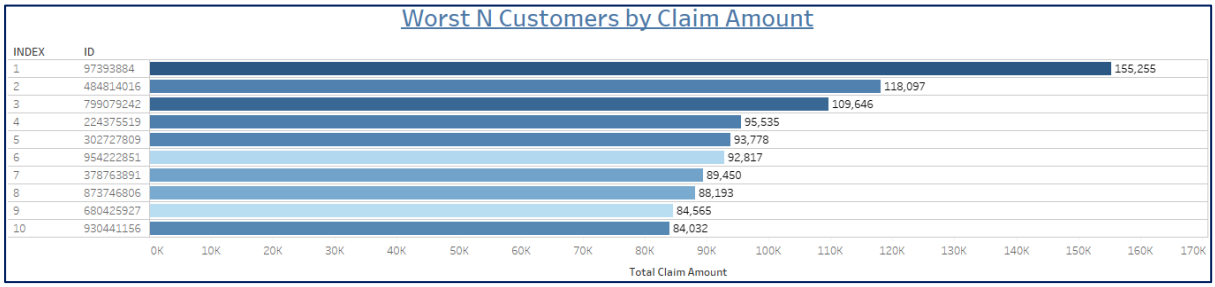
25. Outliers Old Claim



26. Outliers New Claim

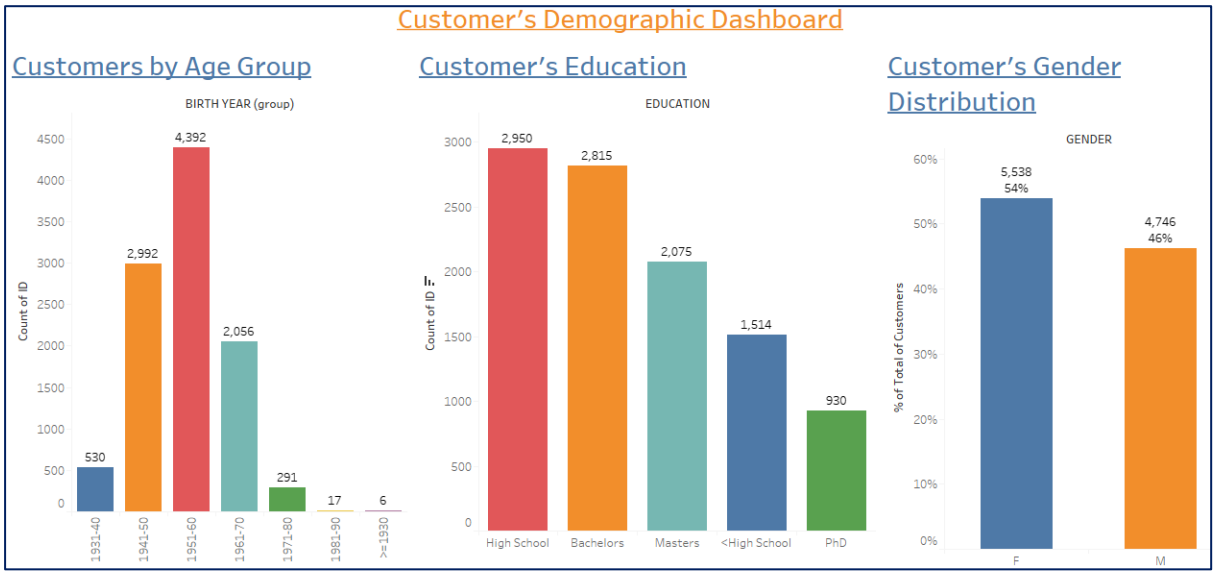


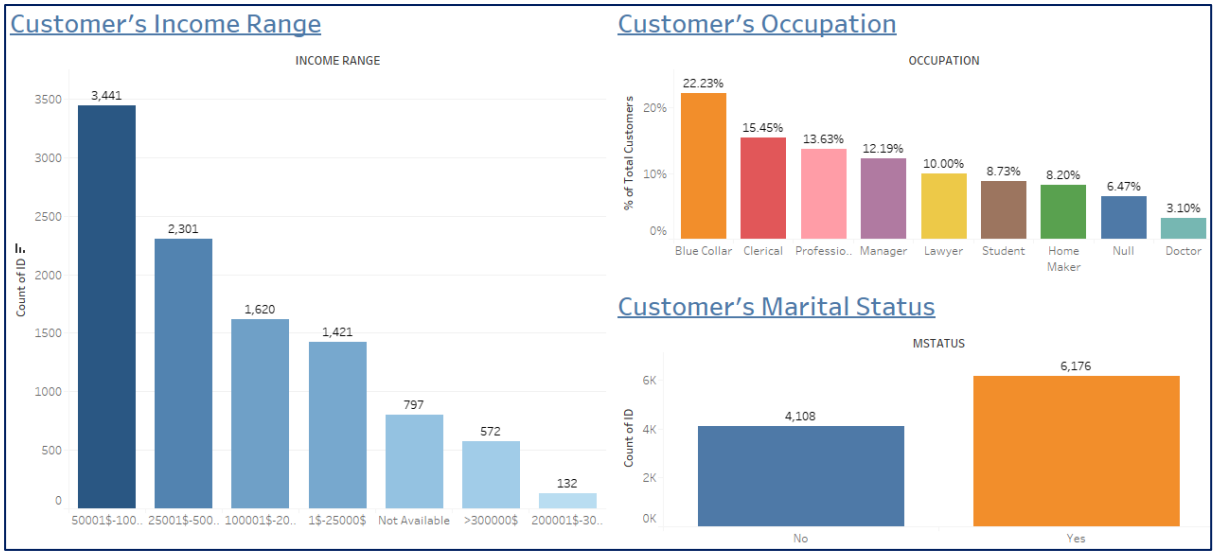
27. Worst n Customers by Claim Amount



Customer's Demographic Dashboard

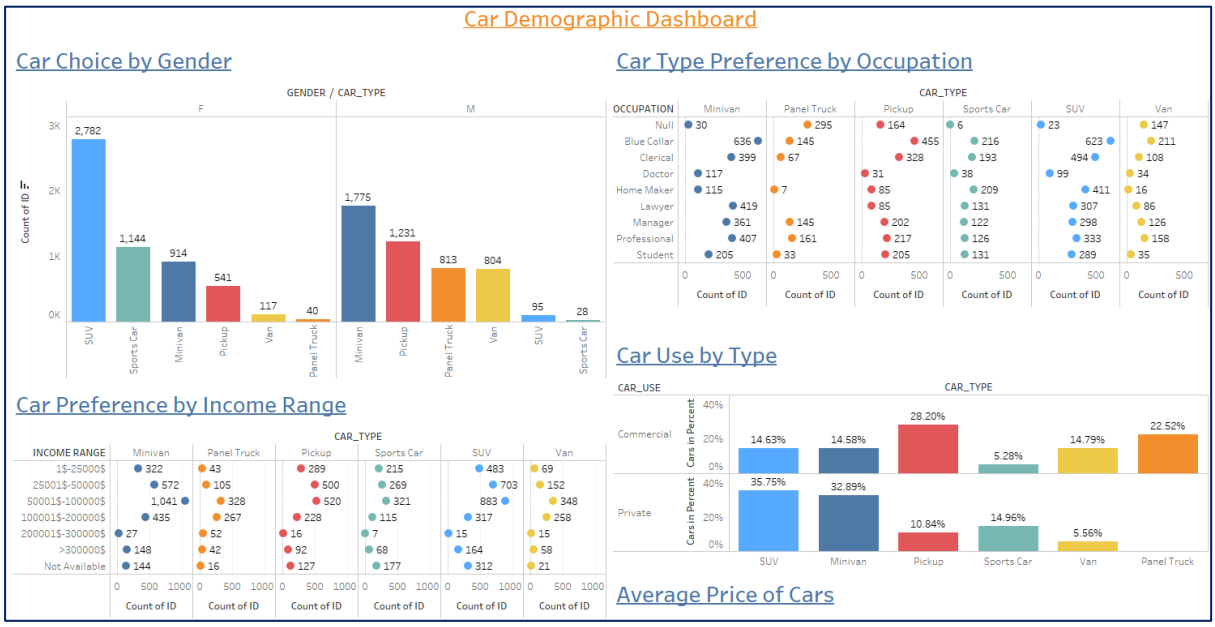
For any insurance company their major resource is Customers. Customers are broadly categorized by the following: Gender, Marital Status, Education, Occupation, Income Range and Age Group. From the available data, we can see the Female population is more than Males at 54% of the total. There are a greater number of customers who are married with majority of the customers with an income of more than $100K. Majority of the customers have a High School degree. Most of the customers have birth years between 1941-1970. Customers with Blue Collar jobs are the most equipped with cars.

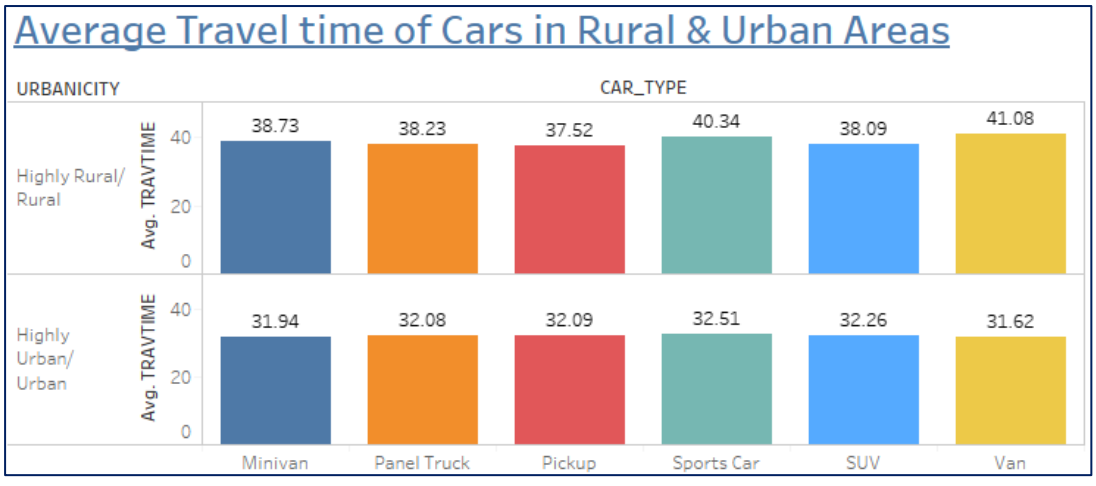


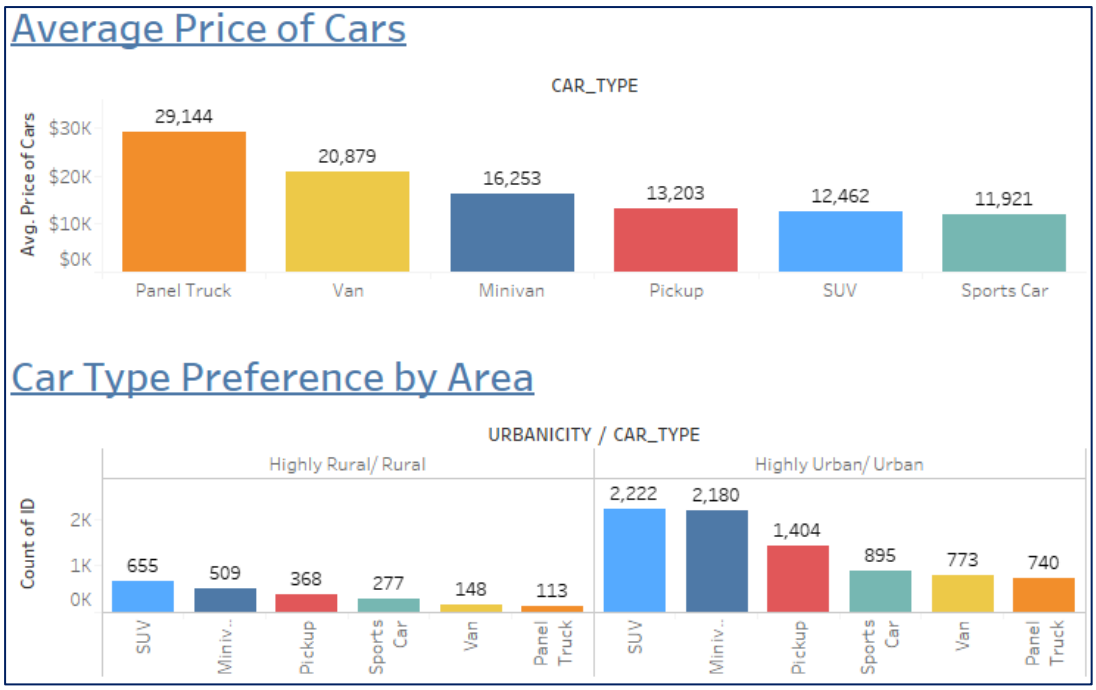


Car Demographic Dashboard

The properties, usage and market value of the car are the most important aspects that any insurance company has to look out for. We see that the average price of Panel Trucks is more which is priced at $29K followed by Vans whereas the most preferred car among both the genders is SUV and Minivan. Interestingly more no. of females prefers Sports Car over males. In terms of car travel time, all the cars are equally preferred in both Highly Rural/Rural and Highly Urban/Urban areas. For private usage, SUV and Minivan are preferred the most while for commercial purpose Pickup trucks are used the most. Most of the customers with high income range prefer SUV and Minivans. Irrespective of job type, SUV and Minivans are preferred most. So overall, we can see that SUV and Minivans are most popular among customers.

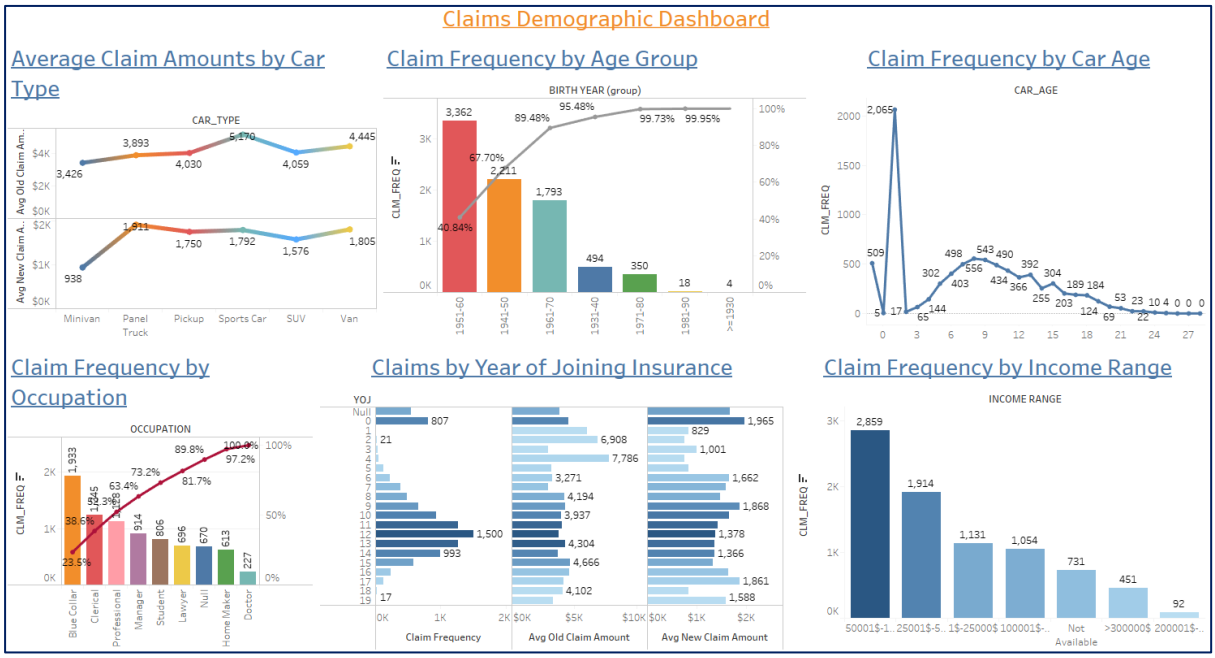


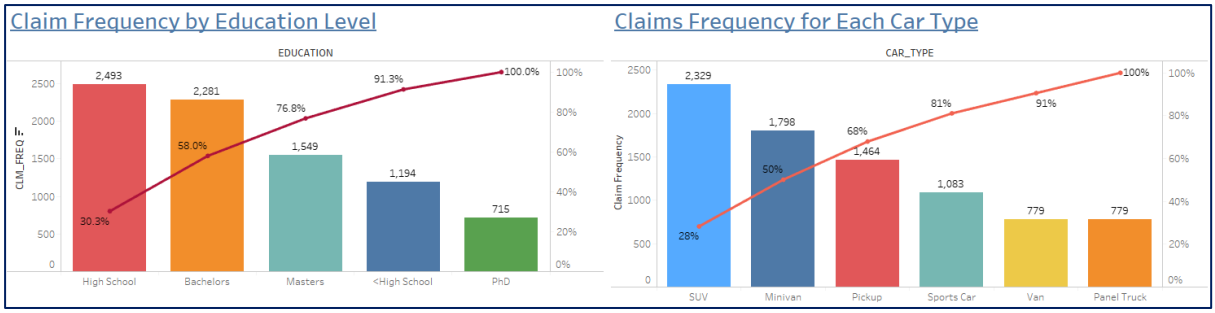




Claims Demographic Dashboard

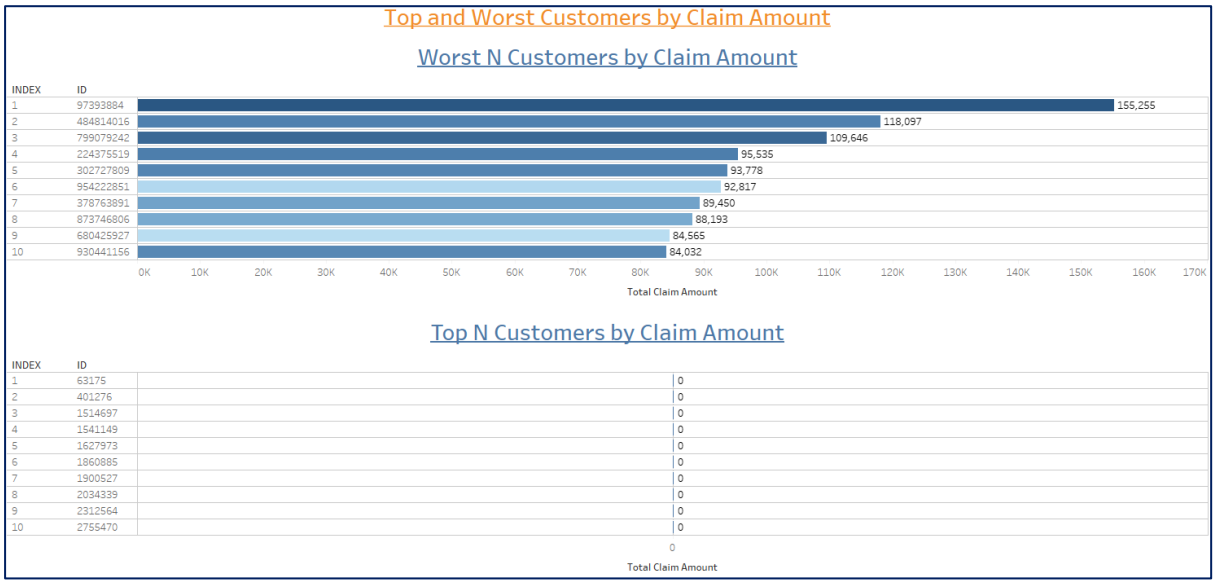
For any insurance company the percentage of claims and claims amount are the most important factors. We see that insurance has been claimed the most for SUV's and Minivans with both accounting for 50% of the claims. This can also due to the fact these are most preferred cars and are sold the most. Historically, Sports Car have the highest average of claim amounts. When it comes to applying the insurance the average old claim amount and new claim amount seems to very high compared to the no. of customers. This gives us an insight that there can be many outliers which are evident in the below plots.





Top and Worst Customers by Claim Amount

It is also important for an insurance company to figure out their top performing customers. Top customers can be categorized as who has less frequency of claims and amounts. Worst customers can be categorized as who frequently claims hefty amounts for which the insurance company has to pay. In this regard, customers can be categorized by their claim frequency, old claim amount and new claim amount and categorized as top or worst performers. The company might rollout new reduced offers or advantages to the top performing customers while for the worst customers, they might plan to increase the premium of insurance.



Executive Summary

- The Tableau dashboards are intended to analyse Car Insurance Claims in Detail and to support further Urban area noted high number of claims than rural area across all the Vehicle Type.
- SUV, Pickup and Mini Van are top three preferred cars by blue collar Occupation individuals.
- Blue collar drivers are the topmost who has highest income and high claim amount.
- Doctors are the safest drivers who has less claim amount with less frequency.
- Among Commercial Vehicle Pick Up Van has the highest Claim Amount and it is SUV for the Private Owners which has the highest Claim Amount.
- Among Commercial Vehicle Sports Car has the least Claim Amount and in case of Private owners it is Van that has the least claim.
- Claim Frequency by Female is higher than that of the Male Counter Part.
- When it comes to Commercial Vehicle it is Male has highest claim and in case of Private Vehicle the females have higher claim.
- Those who have Mini Van have higher Income indicating the People with high income prefer to won Mini Van.
- Also, a Male prefers a Mini Van and a female prefers a SUV
- It is observed that the 4th Quarter has seen higher claim across years then the other quarter.
- In Age group of 54 to 78 the claim frequency and amount has been higher indicating with Age higher claim.
- A married female with Phd has the least claim frequency.
- A married female with high school Education and married status has highest claim also, indicating education is also an important contributor for ensuring safe driving.
- Panel truck is the safest car with kids driving and SUV the most unsafe this is highlighted by the least claim and highest claim respectively.
- Sports car need to be carefully considered before finalizing as the average amount per car is very high and the frequency of Claims is also very high. The total number of Sports cars out of total types of cars should not exceed 20% otherwise there are chances of incurring losses to the Insurance company.
- Special plans can be designed for Students and Low-Income range users who are just starting out as their Claim Amount is less as well as frequency of Claims. Hence this is good for publicity.
- Doctors and Lawyers seem to have the lowest Claim amount and hence can be considered for any special bonuses/schemes as well.
- SUV owners who drive private vehicles - either new vehicles or 8 -10 years.
- Target unmarried people primarily, also target blue collar workers and unemployed people with a high school or bachelor's degree.
- Investigate the spike in people opting for car insurance between 1951 to 1960 and try to recreate similar marketing campaigns. Target pricing of insurance for people taking more claims in the year of 1955 to 1960 and later as they have the highest number of claims.