

CS459 – Business Intelligence

Final Project: Solving Business Problems using BI Methodology

Dataset: IBM Watson Marketing Customer Value Data

Team members:

Danish Badar Qureshi – 22890

Hamza Akbar – 09364

Introduction

The main aim of this project is to apply the Business Intelligence knowledge and the BI Methodology to solve real world business problem.

A dataset is chosen which is then cleaned and understood using Exploratory Data Analysis (EDA). It is then used to create a story in Power BI in the form of a multi-page report with multiple dashboards.

Dataset Used

IBM Watson Marketing Customer Value Data: <https://www.kaggle.com/datasets/pankajsh06/ibm-watson-marketing-customer-value-data>

The dataset above gives information about the different insurance plans of different customers and their expenditures, claims, and sales practices. It is used for marketing purposes to target customers more efficiently.

It has numerical attributes including: 'Customer Lifetime Value', 'Income', 'Monthly Premium Auto', 'Months Since Last Claim', 'Months Since Policy Inception', 'Number of Open Complaints', 'Number of Policies', and 'Total Claim Amount'.

Categorical attributes include: 'State', 'Response', 'Coverage', 'Education', 'Employment Status', 'Gender', 'Location Code', 'Marital Status', 'Policy Type', 'Policy', 'Renew Offer Type', 'Sales Channel', 'Vehicle Class', and 'Vehicle Size'

The dataset above gives information about the various apps on the google play store. It has attributes like category, genre, price, installs, last updated, etc.

factors affecting Employee Attrition. It analyzes how parameters like age, income, education, job satisfaction, time at company, etc affect each other and the employee attrition at a company.

Data Wrangling

The data obtained from the above Kaggle link was carefully observed and understood. The dataset was in excellent condition with almost no discrepancies found. There were:

- No missing values
- No data type errors
- No data entry errors

The only problem found was the inconsistency in the “Effective To Date” variable's formatting and data type, which was fixed.

Previous date format:

```
0    2/24/11
1    1/31/11
2    2/19/11
3    1/20/11
4     2/3/11
Name: Effective To Date, dtype: object
```

New date format:

```
0    2011-02-24
1    2011-01-31
2    2011-02-19
3    2011-01-20
4    2011-02-03
Name: Effective To Date, dtype: datetime64[ns]
```

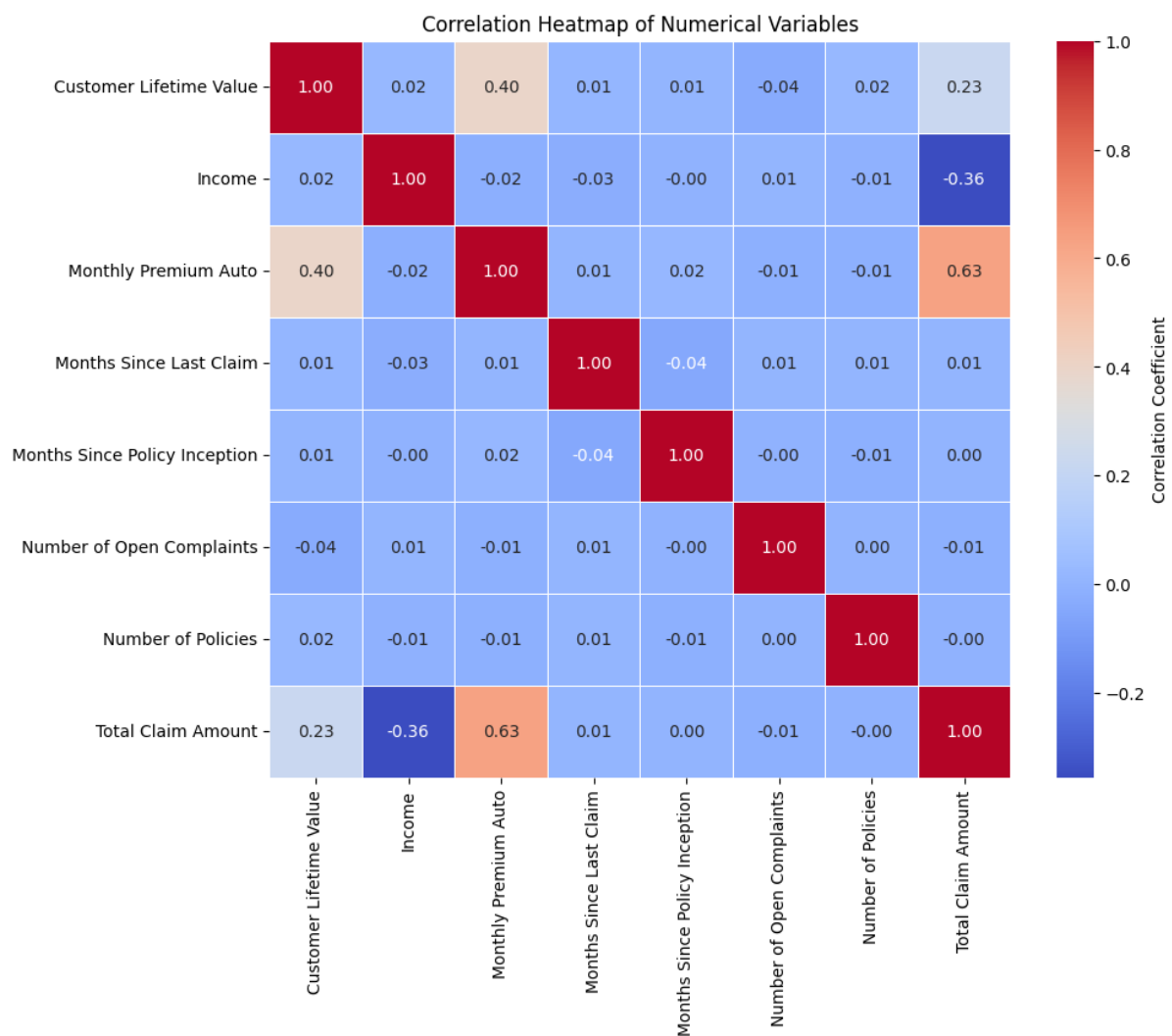
Exploratory Data Analysis (EDA)

To fully understand the dataset, an extensive EDA was performed comprising of the following elements:

1. Univariate Analysis of Numerical Data: Histograms, boxplots, density plots
2. Univariate Analysis of Categorical Data: Frequency Histograms or Bar Plots
3. Bi-variate Analysis - Correlation Heatmaps of Numerical Variables
4. ANOVA for Comparing Numerical Variables Across Different Categorical Groups
5. Chi-squared Tests for Independence Between Categorical Variables

Important findings are listed below:

1.

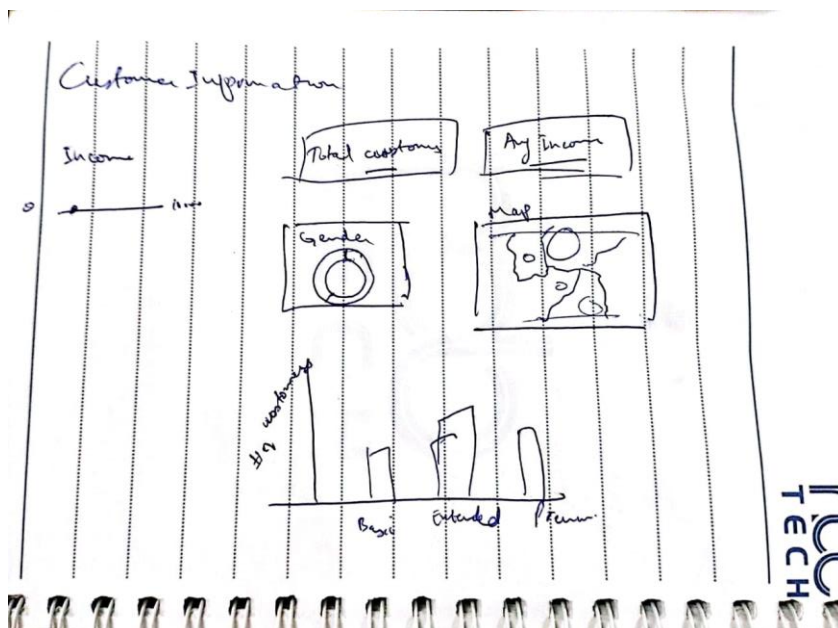


- For relationship between categorical and numerical values, it is noted that state, response, policy type, and policy do not significantly affect various customer metrics such as lifetime value, income, and claim amounts. However, factors like coverage, education, employment status, gender, location code, marital status, renew offer type, vehicle class, and vehicle size do have significant effects on these metrics. For instance, coverage, education, and marital status have significant impacts on income, while renew offer type and vehicle class significantly influence total claim amount.
- In this chi-square analysis, several variables were tested for independence. The results indicate that State is independent of most variables tested. However, it shows dependency on Policy Type and Policy. Response demonstrates dependence on several factors, including Education, EmploymentStatus, Location Code, Marital Status, Renew Offer Type, Sales Channel, Vehicle Class, and Vehicle Size. Coverage is dependent on Education, EmploymentStatus, Location Code, Renew Offer Type, and Sales Channel. Education shows dependence on EmploymentStatus, Location Code, Marital Status, Renew Offer Type, Sales Channel, Vehicle Class, and Vehicle Size. EmploymentStatus is influenced by Gender, Location Code, Marital Status, Renew Offer Type, Sales Channel, and Vehicle Size. Gender is dependent on Location Code and Marital Status. Location Code is dependent on Marital Status, Renew Offer Type, Vehicle Class, and Vehicle Size. Marital Status demonstrates dependency on Renew Offer Type and Vehicle Class. Policy Type is dependent on Policy. Renew Offer Type is dependent on Sales Channel and Vehicle Class. Sales Channel is dependent on Vehicle Class. Vehicle Class is dependent on Vehicle Size.

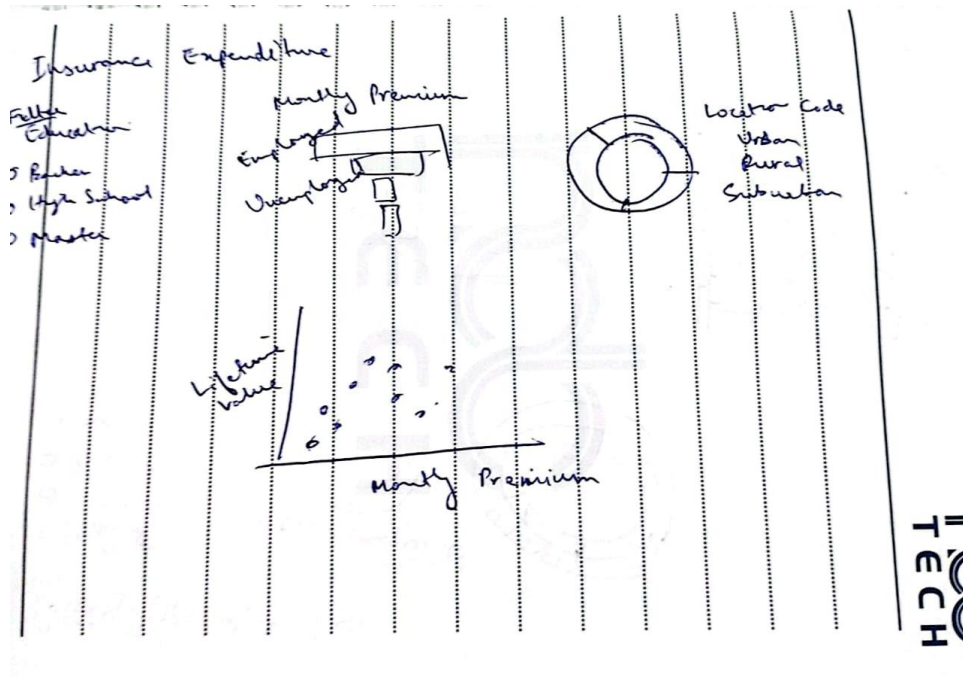
Data Visualization: Paper Charts

After thoroughly understanding the data, we started experimenting with various approaches to visualize our data. To do so, we constructed rough diagrams to understand how our dashboard would look like. The following paper charts were initial ideas put on paper to better understand where we were going and what our approach would be. Several charts were later changed, modified, or eliminated in the final staged of development.

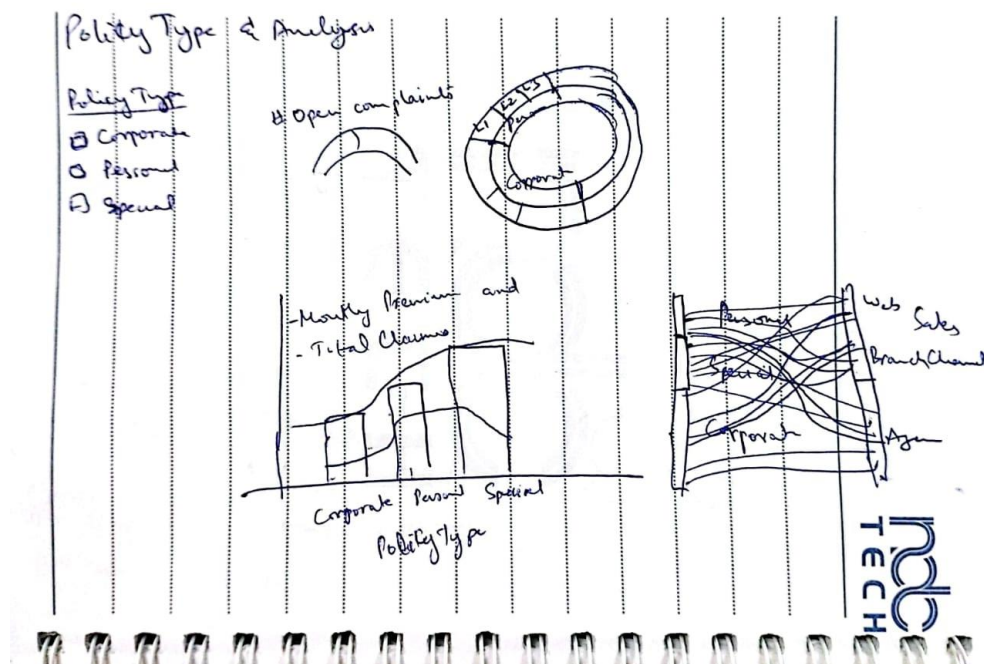
1. Customer Information



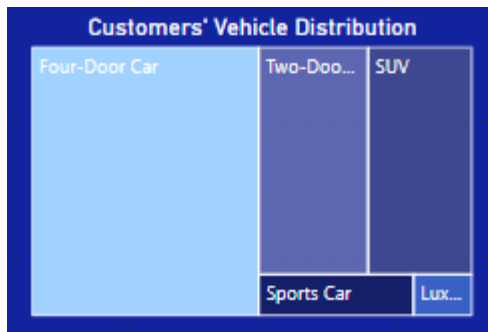
2. Insurance Expenditure



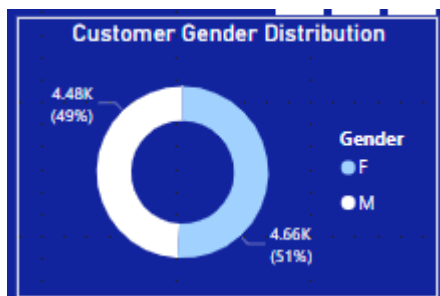
3. Policy Type and its Analysis



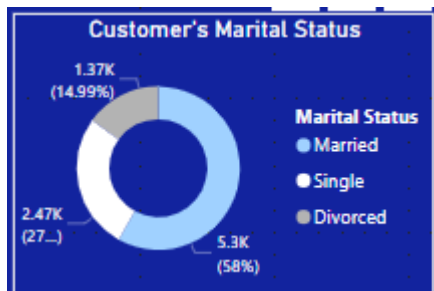
Power BI: Individual Charts



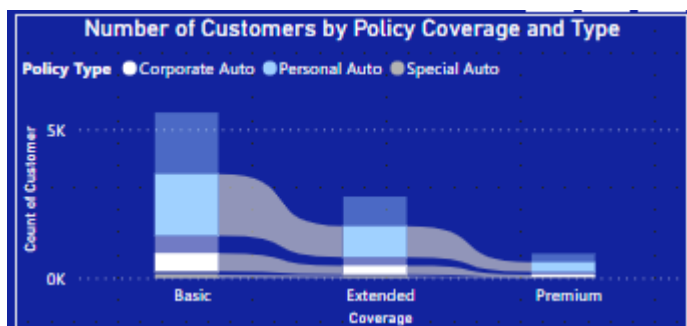
This chart shows the customers vehicle distribution, that is how many customers have what kind of cars.



This chart shows the gender distribution of customers



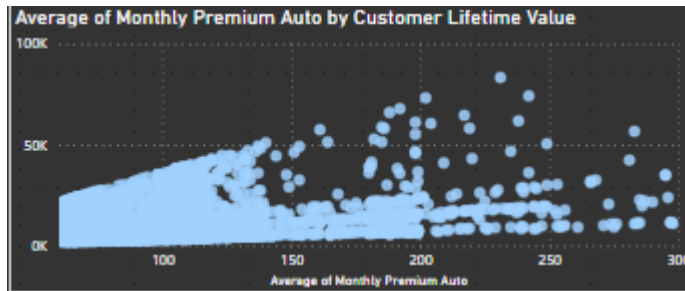
Above shows marital status and their percentages of the total customers



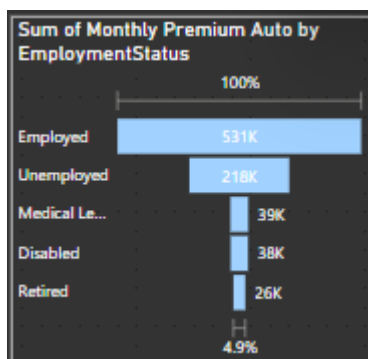
Above shows insurance policy coverage and type in a ribbon chart.



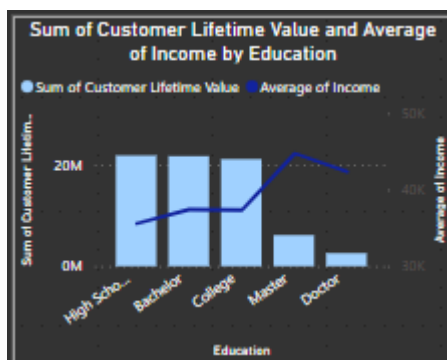
Above shows a map with customer's geographic distribution.



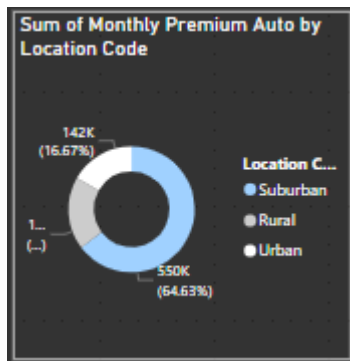
This is the monthly premium's average over customer's worth which is measured as its Lifetime value.



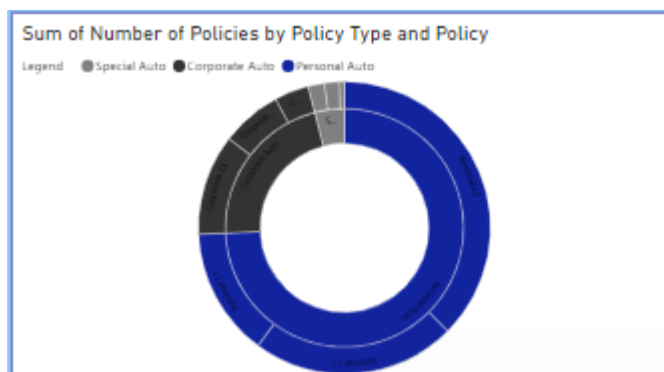
Above funnel chart shows premium based on employment status of the customers in descending order.



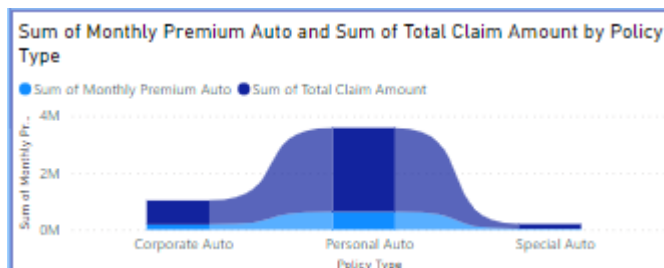
This chart shows how income and value of customers to affected by their education.



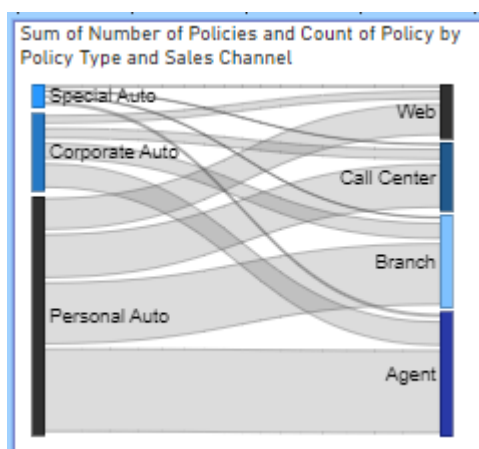
It shows the percentage of distribution of monthly premium based on rural, suburban, and urban locations in donut chart.



For policy analysis, above donut shows the percentage distribution of each type of policy.



The above ribbon chart compares how monthly premium and total claim is affected by policy types

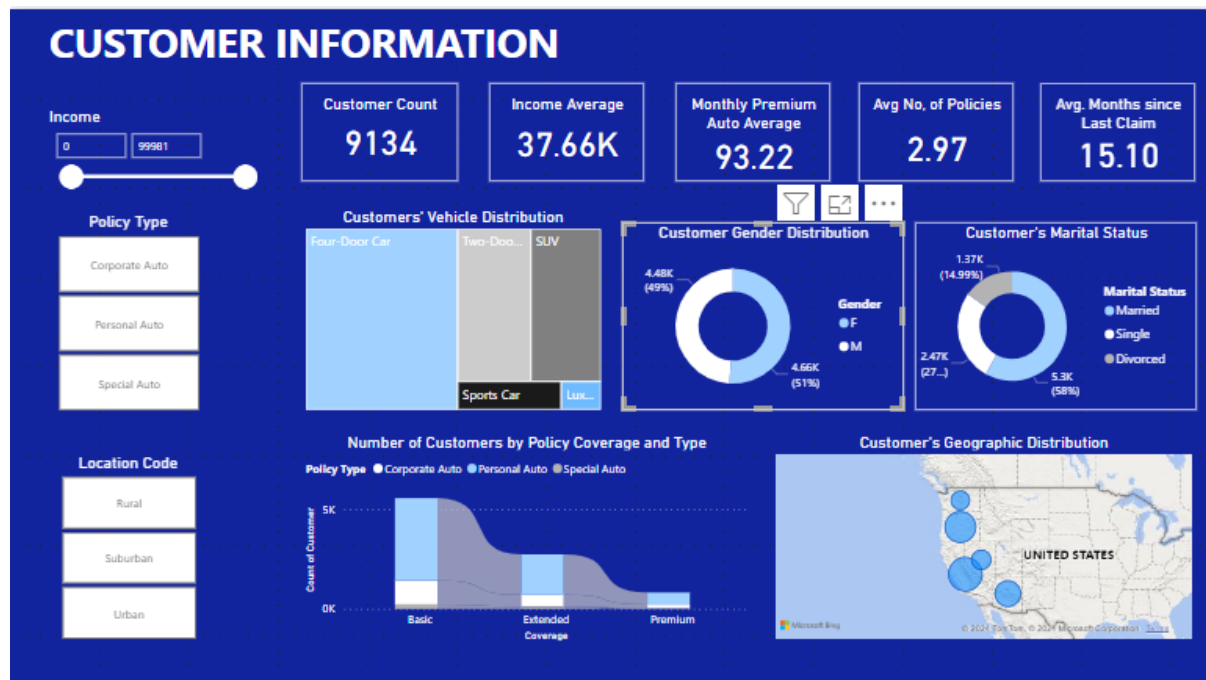


This Sankey chart depicts how different policy types are linked with their sales channel based on number of policies.

Power BI: Dashboards & Stories

Dashboard 1: Customer Information

Problem Statement: *Identifying and observing key characteristics of our existing customers.*

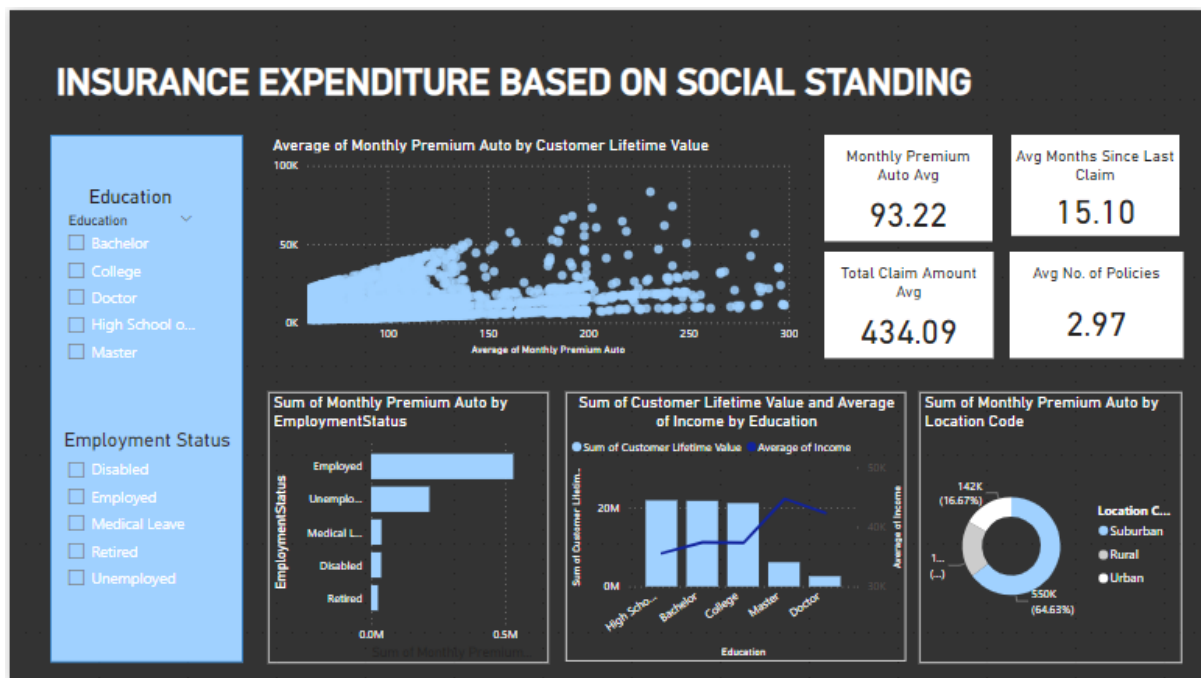


The above dashboard identifies key characteristics and features of our customers. We observe the total number of customers, their average income, how much they spend on policies and their geographical and social demographics to understand their personality. This dashboard allows us to better target our customers by helping us understand them thoroughly.

The dashboard contains slicers to allow selecting policy types and location codes so we may filter the customer analytics.

Dashboard 2: Insurance Expenditure Based on Social Standing

Problem Statement: *Observing the insurance expenditure (premium) and how it depends on the social and financial standing of the customer.*



This dashboard elaborately shows how a customer can become valuable depending on their social standing, by considering parameters like education, employment status, vehicle size and class, etc.

It contains slicers to filter out specific employment status of customers, or their educational background.

Dashboard 3: Policy Type Analysis

Problem Statement: *Understanding how different policy types have different revenue, expenditures and operations.*



The above dashboard allows closer inspection over the different kinds of policies available and how each type could have different costs, or revenues. We can observe that each type has different

monthly premiums and total claims. In addition, we also see how each policy type can have different number of complaints and how each one can be sold in the most efficient way.

The slicer to filter out individual policy types is important to study all three types distinctively.

Team Member Contributions

Danish Badar:

1. Understanding the data and business problem
2. Data Wrangling and EDA
3. Led the creation of business problem statements
4. Led the creation of dashboards on Power BI
5. Compiled this report.

Hamza Akbar:

1. Understanding data and what it represents
2. Worked on identifying three problem statement
3. Assisted in creating dashboards
4. Filled the Background.Data.Knowledge.Project.xlsx sheet.