Project Report

Auto Insurance

DATA 321

Daniel Kozma

Gabriel Medeiros

Rafael Ennes

Abstract

To search for the fair price for insurance companies is not easy. As clients we never know if we are paying the right amount of money and if the companies are proving the right coverage. However, auto insurance companies are constantly trying to provide the best service for their customers with coverages that are suitable for their needs and their budget. There is usually multiple options to choose from which fit into different parameters that can be measured with a wide number of variables. Therefore, using the data below at "Data Description" our group decided to answer some normal questions. Such as, "Which policy generates the most revenue?", "Do People with higher income tend to pay for higher premiums?", "Which state has the highest coverage?", "Do women tend to get involved in more accidents?", "What are the most frequent claims, and which ones get covered through the company the most?". Which the dataset is about an auto insurance company that matches all the requirements, and provides data about Insurance claims including location, policy type, claim, gender, income, education, etc. Even though the company has operated in 5 states only, it is more than enough for us to make real and authentic analyses. We decided to take the approach of the insurance company, in order to visualize information that can be used for marketing. Through this dataset analysis, we can figure out whom to target with which coverage, where revenue is coming from, and expenses are going to, and what are needed in order to maximize profit. Our group believes that the information in this dataset could help a marketing department to target the right audience with the right coverage, based on their needs.

1 Data Description

The dataset used in this project can be found at <u>https://opendata.stack</u>

exchange.com/questions/7807/where-can-i-find-automobile-insurance-claims-data-set

and contains 9135 rows and 26 columns. Each row in the dataset represent a different claim

that contains the following information:

Column Name	Type	Description
Customer	Categorical	This column contains the system's ID for each customer
Country	Categorical/Geographical	Provides information about the country in which the customer lives
State Code	Categorical/Geographical	Provides customer's state code
State	Categorical/Geographical	Provides the name of the state in which the customer is living
Claim Amount	Numerical	Shows the amount in dollars for each claim

Response	Categorical and Binary	Shows whether the client had a response or not
Coverage	Categorical	Represents the type of coverage the customer is paying for
Education	Categorical	Represents the level of education for each customer
Effective to date	Numerical/Date	Represents the date when the claim was recorded as being effective
Employment Status	Categorical	Shows the employment status for each customer
Gender	Categorical and Binary	Represents the gender of the customer as Male or Female
Income	Numerical	Shows the income for each customer
Location Code	Categorical	Shows if the client lives in a rural/urban area

Marital Status	Categorical	Represents the marital status for each ID
Monthly Premium	Numerical	Represents the monthly payment for each coverage type
Months Since Last Claim	Numerical	Tells us when this same client reported another claim, if the value is zero it means that the customer reported the last claim in the same month but in a different day/week
Months Since Policy Inception	Numerical	This column represents how many months ago it was the last Policy Inception
Number of Open Complaints	Numerical	This column shows the number of open complaints per client
Number of Policies	Numerical	Represents the record of total number of policies for each client
Policy Type	Categorical	Represents the category of the vehicle owned by the customer (corporate, personal, special)

Policy	Categorical	Represents the division of the policy type (corporate L1, personal L3…)
Claim Reason	Categorical	Categorizes why the customer called the company
Sales Chanel	Categorical	Reports how the customer got in touch with the company
Total Claim Amount	Numerical	Shows the total amount in dollars of the claim
Vehicle Class	Categorical	Shows the vehicle's class (two-doors, four-doors…)
Vehicle Size	Categorical	Represents the vehicle's size as small, large, or midsize.

2 Project Proposal

It is common knowledge that says that companies are focused on profit, and that is why business analysts are important: To analyze and create solutions for an issue, or to create a method that would increase the profit on a company using tools that may

have never been used before to reach a profitable configuration in a company. Using the knowledge acquired in this class, our group was able to find a dataset from an insurance company for vehicles. After some work with exploratory analysis on the data, we decided that this information could be given to a marketing division of a company, which would be able to ensure that each customer could be reached in an efficient way offering a coverage that supplies their specific necessities.

3 Methodology

In order to use the dataset to generate useful visualizations to reach our estimations with the proposal of the project, we made use of Tableau Desktop as we learned throughout the semester. Tableau Desktop was used to create visualizations such as pie charts, lines (discrete), filled maps, tables, bar charts, tree-maps, and dashboards that compiled graphs to better visualize and compare the result obtained through combination of different variables from the dataset.

## 4 Implications through the analysis process

At first sight our group assumed that the data did not present any issue, but when we dug deeper into the analysis process we came across different situations with the data. First, we did not receive enough information about the columns, but with group debates and discussions we could describe the role of each column.

5 Project findings

This section of the project is destined to present what the group could find out through the dataset analysis. Here it will be presented the graphs and visualizations

followed by comments explaining what it is possible to conclude when looking at the visual representations the group created when trying to answer the research questions the group came up in the brainstorming process.

The following dashboard will help is to visualize accurately what are the states with most cases and the states with the highest average of prices.
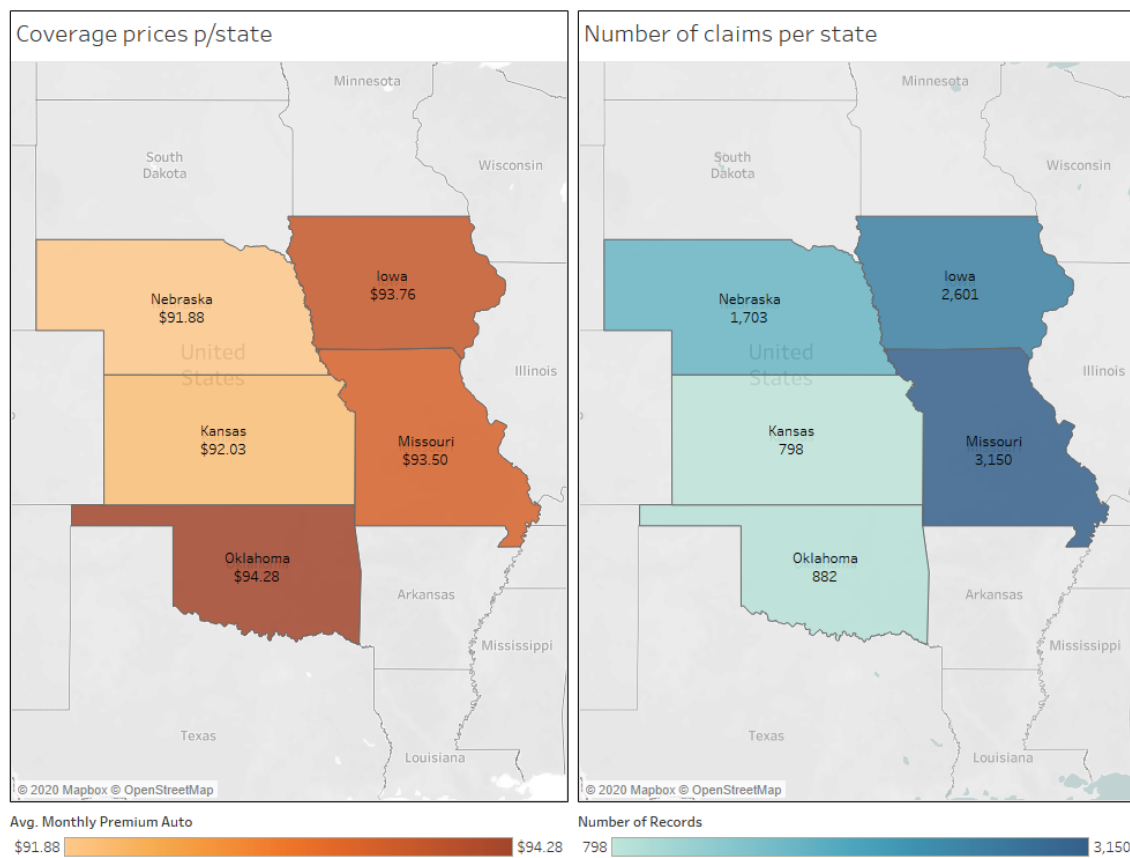
1)   **What are the states with the most accidents?**

2)   **In what state it is more expensive to have a Premium coverage?**

Using the states provided by the dataset, we can analyze that the state where the average of prices is higher than the other states is in Oklahoma, with an average of $94.28 dollars with a register of 882 claims reported. Also, we can analyze that the state with the highest number of claims is Missouri with 3,150 cases with an average of $93.50 dollars. The difference of average price between Missouri and Oklahoma is of 78 cents with a difference of 2,268 cases, what brings out that where there are more demand the prices are lower; but when analyzing Kansas as an example, the difference in the average coverage prices are of 1.47 cents less than in Missouri even though in Kansans there are 2,352 less claims registered, what made the group to generate the following dashboard. Therefore;

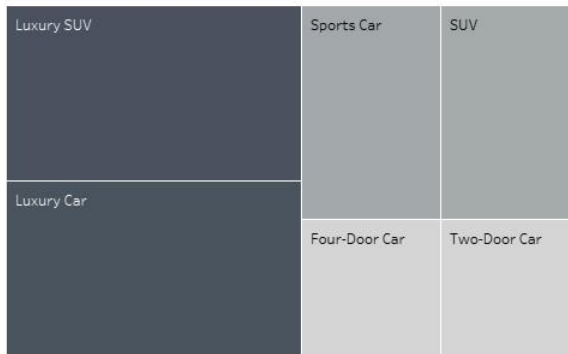1)**What are the states with the most accidents?**

Missouri with 3,150 cases with an average of $93.50 dollars

2)**In what state it is more expensive to have a Premium coverage?**

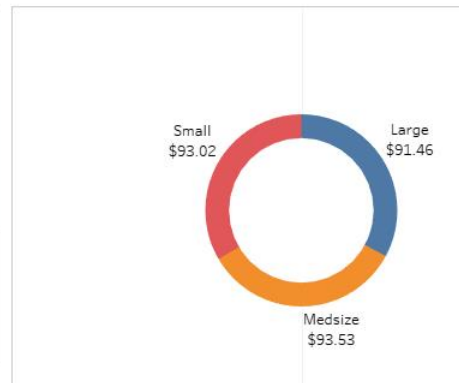Oklahoma, with $94.28 dollars with a register of 882 claims reported
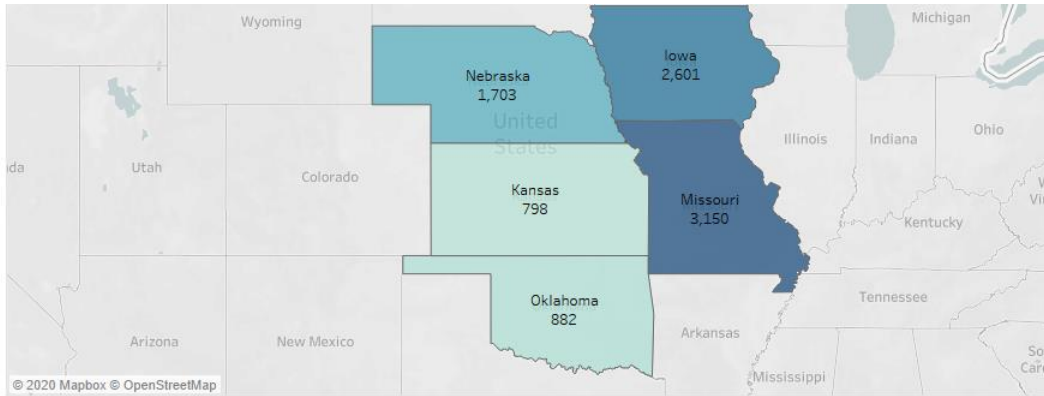
Prices per Vehicle Type

Vehicle size prices

Appling filters and interacting with the dashboard made it possible to the group to visualize that where the average price is higher, there are more cases registered for medium sized vehicles. Moreover, Oklahoma is where the average price for four–door, SUV, and two-door is higher than in the other states, and those are the type of cars with the highest number of records in every state what makes the average coverage prices to go up in Oklahoma.

Claim Reasons

Scratch/Dent
15.68%
$423.39

Collision
41.08%
$469.17

Hail
32.03%
$408.78

Education

Master
8.11%
$94.04

Bachelor
30.09%
$92.50

High School or Below
28.71%
$94.64

College
29.35%
$92.67

Marital Status

Single
27.01%
$93.88

Divorced
14.99%
$92.60

Married
58.00%
$93.07

Days

Effective To Date

(Bar chart: Number of Records by day)
Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
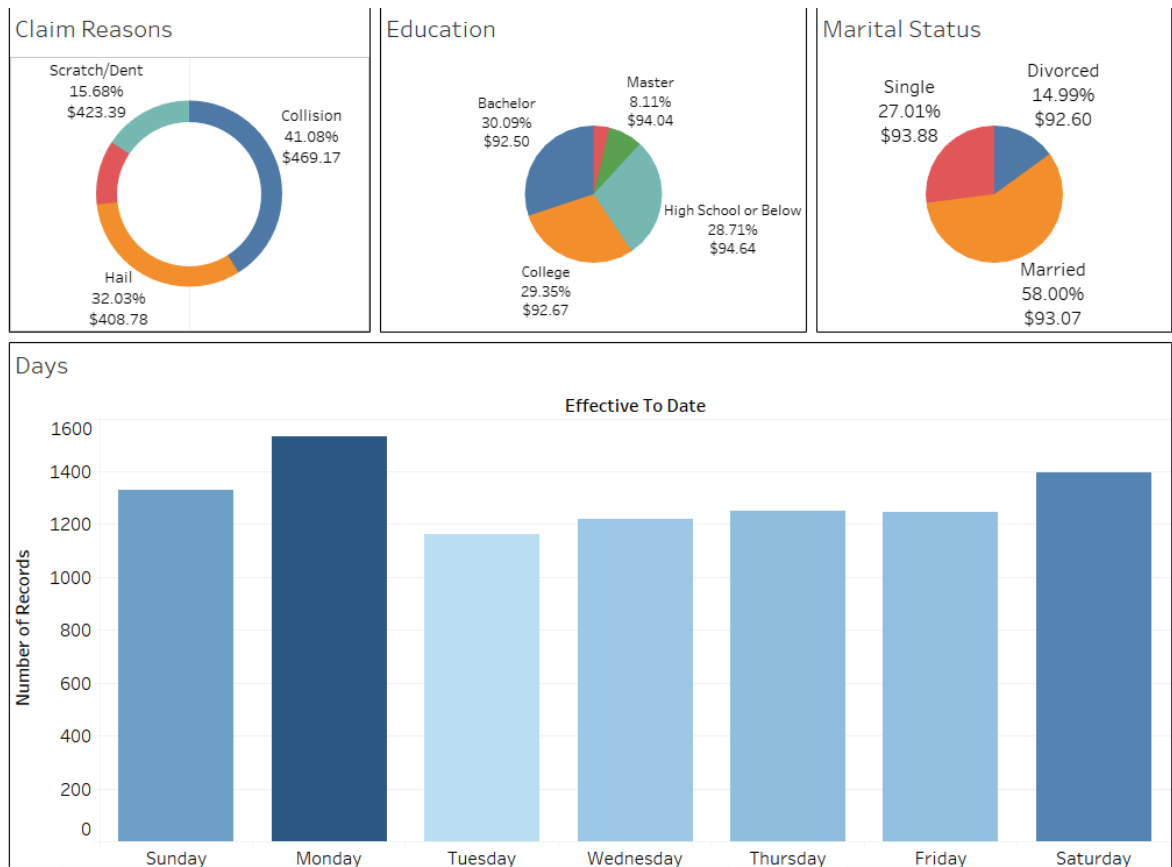
3) **What are the main accident causes?**

4) **What are the days with the biggest number of accidents?**

5) **Does married people get involved in accidents more than single/divorced people?**

6) **What level of education has the most accidents?**

Analyzing this dashboard made it possible to conclude that the principal claim reasons are Collision and Hail, furthermore the average prices assigned to those claims are not related to the number of claims since Hail has double the number of claims for Scratch/Dent, but Scratch/Dent has an average cost of $14.61 dollars less than the average for Hail claims. In addition, there are more claims for people that are at High School, College, and people that already have bachelor, but somehow the average

monthly prices for High School and Doctorate are higher than for any other education level. Married people tend to get involved in more accidents than Single and Divorced people, but the average prices are close for each one of the marital status, what could be done here is to increase the coverage plan price to married people since they are occupying more than 50% of the claims in the registered states. According to the data, it is expected than the accidents happens the most in Mondays and Saturdays. Therefore;

3) What are the main accident causes?

Collision and Hail.

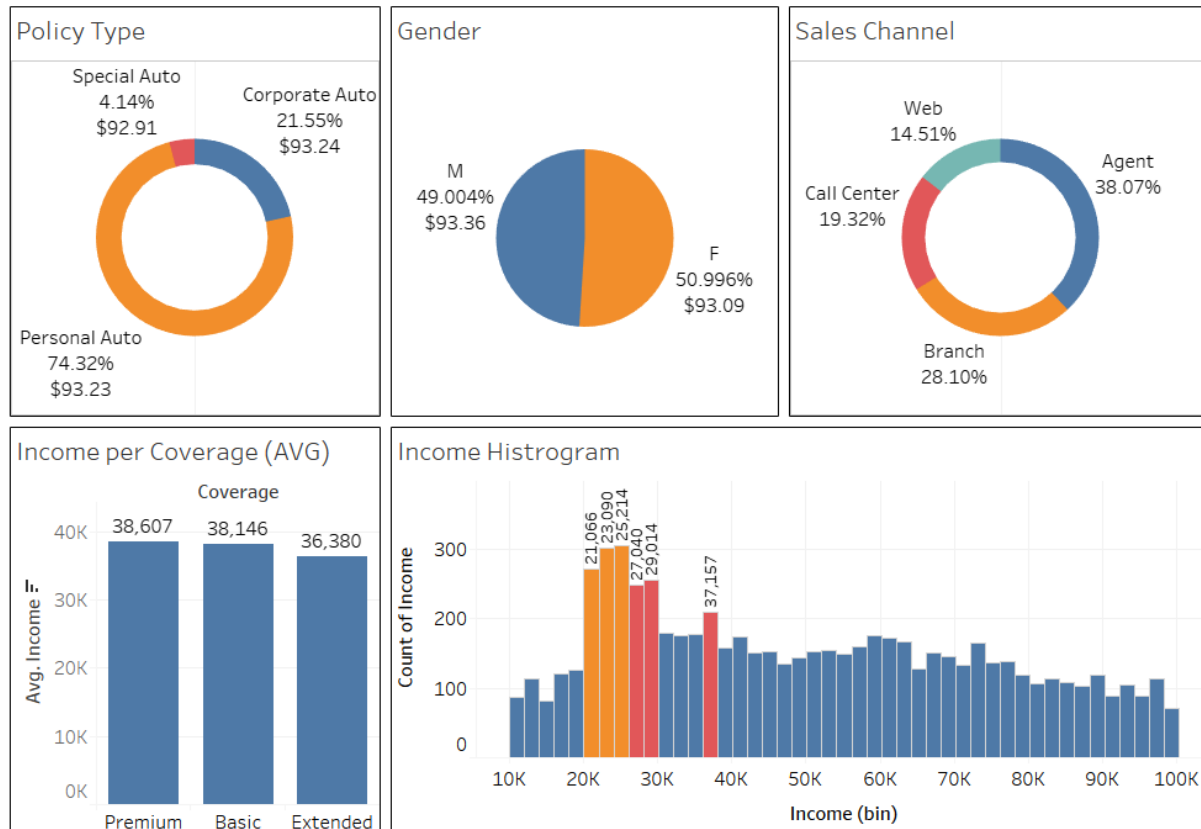**4) What are the days with the biggest number of accidents?**

Monday

5) Does married people get involved in accidents more than single/divorced people?

Yes, 58% of the accidents are from married

6) What level of education has the most accidents?

The Bachelors with 30.09%

Policy Type

Special Auto
4.14%
$92.91

Corporate Auto
21.55%
$93.24

Personal Auto
74.32%
$93.23

Gender

M
49.004%
$93.36

F
50.996%
$93.09

Sales Channel

Web
14.51%

Call Center
19.32%

Agent
38.07%

Branch
28.10%

Income per Coverage (AVG)

Coverage

Premium	Basic	Extended
38,607	38,146	36,380

Avg. Income

Income Histrogram

Count of Income

21,066 23,090 25,214 27,040 29,014 37,157

Income (bin)

7)   Is it true that Females get involved in more accidents than Males?

8)   What are the main means of communication that the customers use when they need to report an accident?

9)   Does corporation vehicles get less involved in accidents than personal vehicles?

10)   People with a higher income tends to pay for premium coverage?

Looking at this dashboard, it is possible to visualize that when comparing genders, female get involved in about 51% of the reported accidents, while men get involved in 49% of the reported accidents, but the monthly coverage prices still almost the same for both genders because the difference in cases is not significant to change the price based on number on accidents. The most used ways to report a claim are by calling the Agent or the Brench, and if the purpose of the company is to use the website as a main tool to the clients to report an accident, it is not being used as much as the other options. It means that the website or isn't divulgated enough or it's not good enough in interactivity and functionality for customers. Looking at the income histogram, most of the customers have an income of 21k to 25k, and the second biggest group of incomes are 27k, 29k, and 31k; somehow, those numbers are below the average of income for each one of the coverage types, what means that there is a wide range of people with different incomes that are choosing the same coverage plan, and it is not possible to predict somebody's plan choice judging by the income.

7) **Is it true that Females get involved in more accidents than Males?**

Yes, by 50.996% of the woman are involved in accidents.

8) **What are the main means of communication that the customers use when they need to report an accident?**

38.07% of the clients call their agent whatever they are involved in an accident.

9) **Does corporation vehicles get less involved in accidents than personal vehicles?**

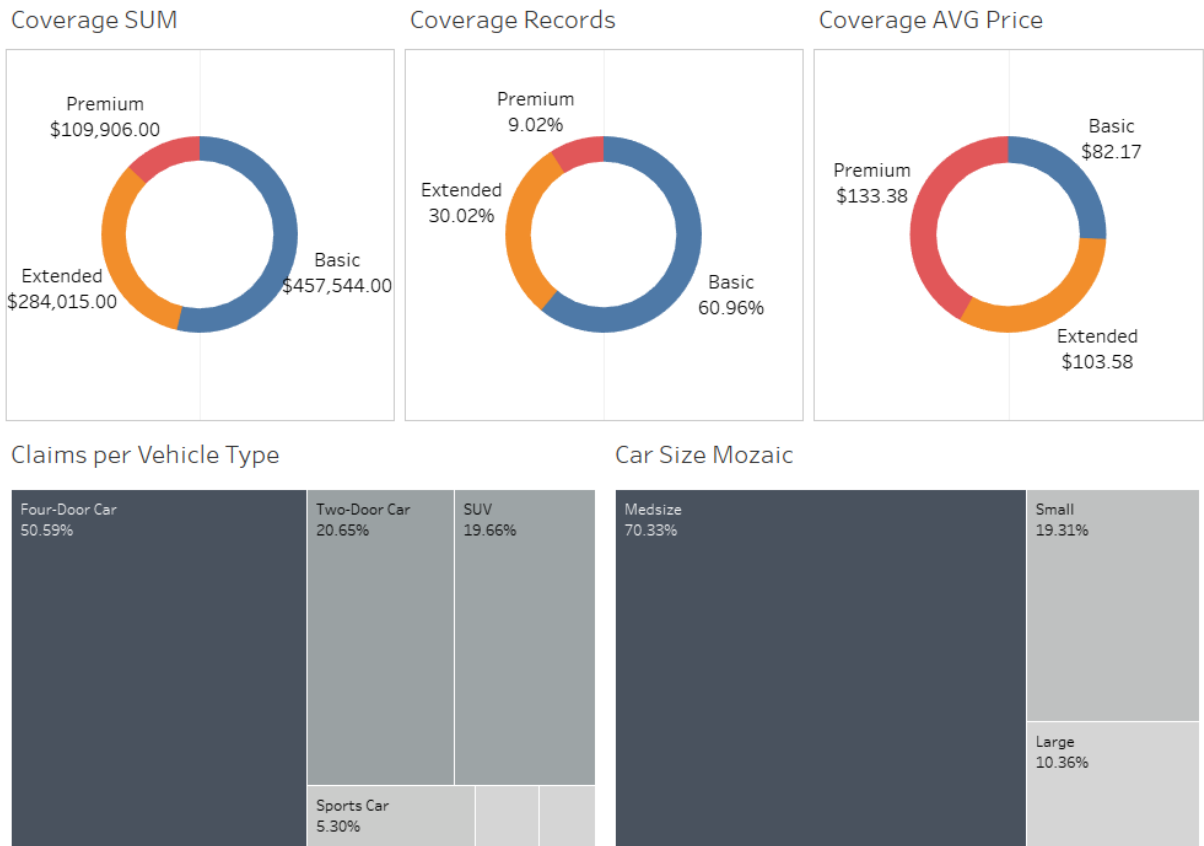Yes, they do get less involved into accident by 21.55%

10)People with a higher income tends to pay for premium coverage?

There are a wide range of people with different incomes that are choosing the same coverage plan, and it is not possible to predict somebody's plan choice judging by the income.

11)   What type of vehicle has the biggest number of accident records?

12)   What size of car has the biggest number of accident numbers?

13)   What is the Coverage type that gives more money for the company?

Coverage SUM	Coverage Records	Coverage AVG Price

Coverage SUM:
- Premium $109,906.00
- Extended $284,015.00
- Basic $457,544.00

Coverage Records:
- Premium 9.02%
- Extended 30.02%
- Basic 60.96%

Coverage AVG Price:
- Basic $82.17
- Premium $133.38
- Extended $103.58

Claims per Vehicle Type:
- Four-Door Car 50.59%
- Two-Door Car 20.65%
- SUV 19.66%
- Sports Car 5.30%

Car Size Mozaic:
- Medsize 70.33%
- Small 19.31%
- Large 10.36%

Through this dashboard, it is possible to analyze that the coverage that gave more money to the company is the Basic, what is directly connected with the number of records for each coverage type. The average price for each coverage varies according to the features each one offers, and it makes sense that clients are more interested in

the cheaper choice available, consequently there are more clients paying for the basic option, and there are more records of claims for the clients under basic coverage plans. The vehicle size with the biggest number of records are median sized cars, while the vehicle types that get involved in the biggest number of accidents are four door cars. Comparing to the previous dashboard that shows the price per vehicle type and size, median sized cars pays a higher average on monthly payments, and luxury cars are the type of cars that pays the biggest average of price. It is directly connected as median sized cars includes the most of the luxury cars that pays for higher monthly coverage prices.

11) What type of vehicle has the biggest number of accident records?

50.59% Four-Door car.

12) What size of car has the biggest number of accident numbers?

Medium size with 70.33%.

13) What is the Coverage type that gives more money for the company?

Basic one with $457,544.00 sum profit.

6 Conclusion

The provided visualizations help to answer all of our research questions and help to get to conclusions in different debates. Tableau was a very efficient tool to use in order to emphasize our findings from the data. We believe that a marketing department could effectively incorporate our findings into their marketing plan, and target the right leads with the desired, and affordable coverage options. As a group, we really enjoyed working together on this project. Communication was flowing very well while we tried to figure the answers to our questions. We also figured out each other's strengths, so we divided the work based on that. Overall, it was a great learning experience, and we know that this project helped to prepare us for work after college.