AUTO INSURANCE MONTHLY PREMIUM PREDICTION

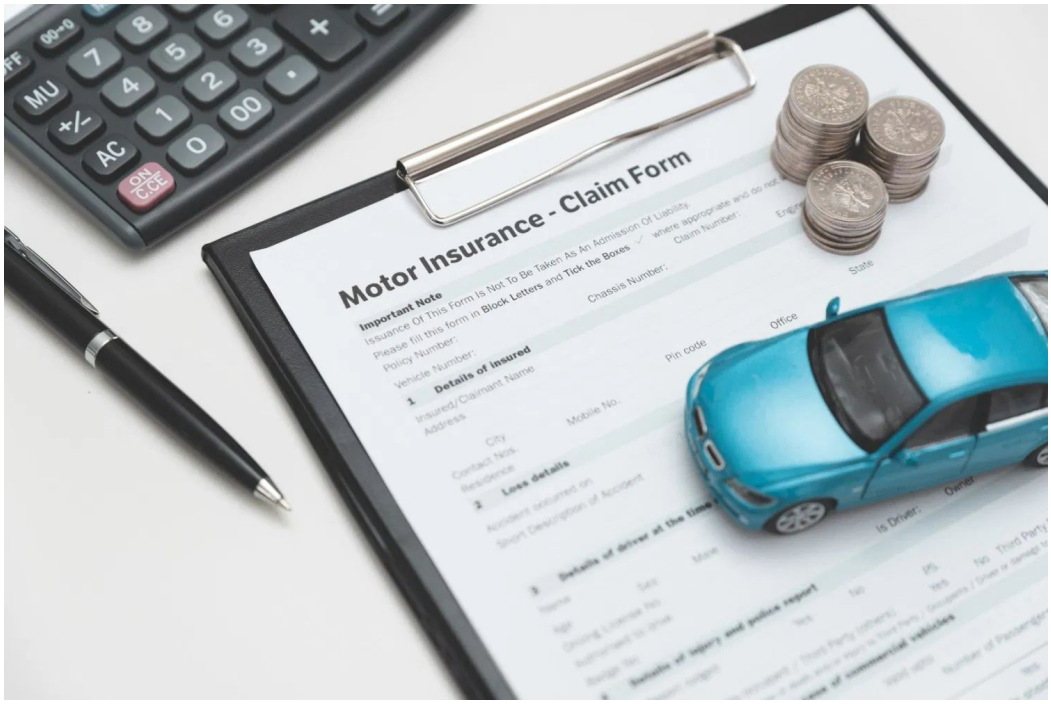

Figure 1 : Picture is from Auto insurance page

Auto insurance is a contract between a driver and the insurance company that protects the driver against the financial loss of an accident, theft or any unforeseen incident. In the US, all drivers must have a liability insurance that meets the legal requirements, which vary with the states' laws.

Policyholders pay monthly premiums. Various factors affect the insurance premiums of a customer. The list of factors is not limited to but includes the state, vehicle type, accident history, deductibles, age, gender, demographic information, credit score, and many different factors; for example, a lower the deductible amount results in a higher the premium amount.

Deductible is the amount of money that a policyholder pays if he/she is involved in an incident and files a claim. The process of claim filings is as follows. A policyholder fills out an initial claim to the insurance company; this is called an insurance claim,

which is a formal request by a policyholder to an insurance company for compensation of loss. After this, a validation of the insurance company is completed; if the complaint is approved, the company issues the payment to client or any other responsible party; the policyholder pays the deductible amount.

Problem Statement

An insurance premium is the amount of money that clients pay for the insurance policy that he/she purchases. Insurance premiums vary with various factors. Insurance companies do risk assessments to determine the insurance premium of a client. For example, a higher risk entails higher premiums. From the client's perspective, they look for the best coverage and meet the liability requirements with a minimum budget. This capstone project investigates the significance of factors that affect the insurance premiums of the clients; it provides a systemic analytical investigation of the data that can provide decision support for insurance companies to find the optimized affordable premiums.

Stakeholders

Insurance companies are the stakeholders of this project. Besides, the outcome of this project can be used for company websites and smartphone applications, which optimize & calculate the premiums for their customers.

Data Set

The data was obtained from the Emcien data repository data in a CSV file format. The data includes Auto Insurance claims data from 2011 to 2019; it has 9134 rows and each row represents a unique policyholder; its columns represents 25 different features, which excludes the customer ID, of a policyholder.

The liability auto insurance requirements for Oklahoma,Nebraska and Kansas drivers are:

Oklahoma

1. 25,000$ bodily injury per person per accident.
2. 50,000$ bodily injury for all persons per accident.
3. 25,000$ property damage liability.
4. 425,000/50,000$ uninsured motorist bodily injury.

Iowa

1. 20,000$ bodily injury per person per accident.
2. 40,000$ bodily injury for all persons per accident.
3. 20,000/40,000$ uninsured motorist bodily injury.

Missouri

1. 25,000$ bodily injury per person per accident.
2. 50,000$ bodily injury for all persons per accident.
3. 10,000$ property damage liability.

The variables that are used in this study are:

- **Customer:** A unique customer identifier.
- **State Code:** 2 letter state code.
- **State:** 5 issuing states; Oklahoma, Nebraska, Kansas, Missouri, and Iowa.

- **Claim Amount**: Initially filled claim amount.
- **Response:** Yes or No respond to claim honor status.
- **Coverage:** Coverage type of the policy.
- **Education**: Level of education of the policy holder.
- **Effective Date:** The date the policy was issued.
- **Employment Status:** Employment status of the policyholder.
- **Gender:** Gender of policyholder; male or female.
- **Income:** Annual Income of the policyholder
- **Location Code:** Local code such as rural or urban.
- **Marital Status**: Marital Status of policyholder.
- **Monthly Premium Auto:** Monthly premium paid by policyholder.
- **Months Since Last Claim:** Number of months since last claim was filed.
- **Months Since Policy Inception**: Number of months since policy inception.
- **Number of Open Complaints:** Number of complaints filed by policyholder.
- **Number of Policies:** Policies currently held by the policyholder.
- **Policy Types**: Policy Type..
- **Policy:** Policy specific category.
- **Claim Reason:** Reason for claim.
- **Sales Channel**: Sales channel.
- **Total Claim Amount:** Final amount settled on claim.
- **Vehicle Class:** Vehicle Class.
- **Vehicle Size**: Vehicle Size.

Data Wrangling

1. **Unnecessary information columns:** At this stage, I filtered out unnecessary columns and dropped them.

```
df=df(columns=['column_name'])
```

2. **Incompatible types of columns:** The variables of 'EffecticeToDate' column were not datetime object. Therefore, I convert them to datetime object.

```
data['EffectiveToDate'] = pd.to_datetime(data['EffectiveToDate'], errors='coerce')
```

3. **White spaces in column names:** There were some spaces between the word of column names. I removed these white spaces.

```
df.columns = df.columns.str.replace(' ', '')
```

Key Findings

During exploratory data analysis, we ask the following questions:

1. What are the factors that significantly impact the monthly premiums?
2. How claim reasons correlate with other factors? (are you sure about this question?)

Initial findings are:

- The claim amount, which is the amount that was claimed by a policyholder, is higher than the total claim amount, which is the amount that was approved by the insurance company. According to data, %80 of the claim amount is less than 1000 and %80 percent of the total claim amount is less than 600.
- 314 customers do not have any claims.
- 6817 customers have an income and 2317 customers do not have an income.
- Total claim amount is higher between $20,000.00-$40,000.00 annual income, and total claim amount is between $200.00-$600.00. The higher total claim amounts mostly overlap with the lower-income range.
- 1308 students have honor status. They must be high school students and 198 of these students have no income.
- %70 of the students pay less than 100 $ monthly premium.
- The majority of customers buy low monthly premium policies, which have high deductibles.
- There is a correlation between the total claim amount and monthly premiums.

States	Number of Customers	Paid by customers	Sum of Total Claim Amounts	Percentages
Missouri	3150	$14,015,820	$1,379,130.69	%9.8
Iowa	2601	$11,867,224	$1,126,265.62	%9.5
Nebraska	1703	$7,504,705	$724,390.54	%9.7
Oklahama	882	$4,043,462	$386,977.99	%9.6
Kansas	798	$3,671,747	$348,202.18	%9.5

· The Table shows the number of policyholders in each of the five states; it shows the total premium paid by customers to the insurance company; it shows the amount of money paid by the insurance company for the claims. An interesting finding is that the total claim amount is %10 of the total premium in every state..

· There is no significant relationship between the monthly premiums of a policyholder and the number and the type of policies.

· The type of coverage affects the monthly premiums. Basic coverages have lower premiums, whereas the premium coverages have higher premiums.

· The education level of a policyholder does not significantly impact the monthly premiums, the P.hD. holders have slightly lower monthly premiums than the high school and lower grades holders.

· Gender, employment status, marital status,vehicle size,sales channel do not have a significant impact on monthly premiums.

· Urban and rural communities policyholders have lower monthly premiums than suburban ones.

· The lowest monthly premium belongs to two-door and four-door car types. Luxury cars have the highest monthly premiums.

· Collision is the primary claim reason in each state. The next most frequent claim reason is hail. This is normal because Nebraska, Missouri, and Kansas are in the top 5 states in the U.S. for hail loss (2019).

· Employed policyholders constitute the highest number of policyholders who have the claim reason for hail. The reason can be that these policyholders drive their cars more frequently than the others.