

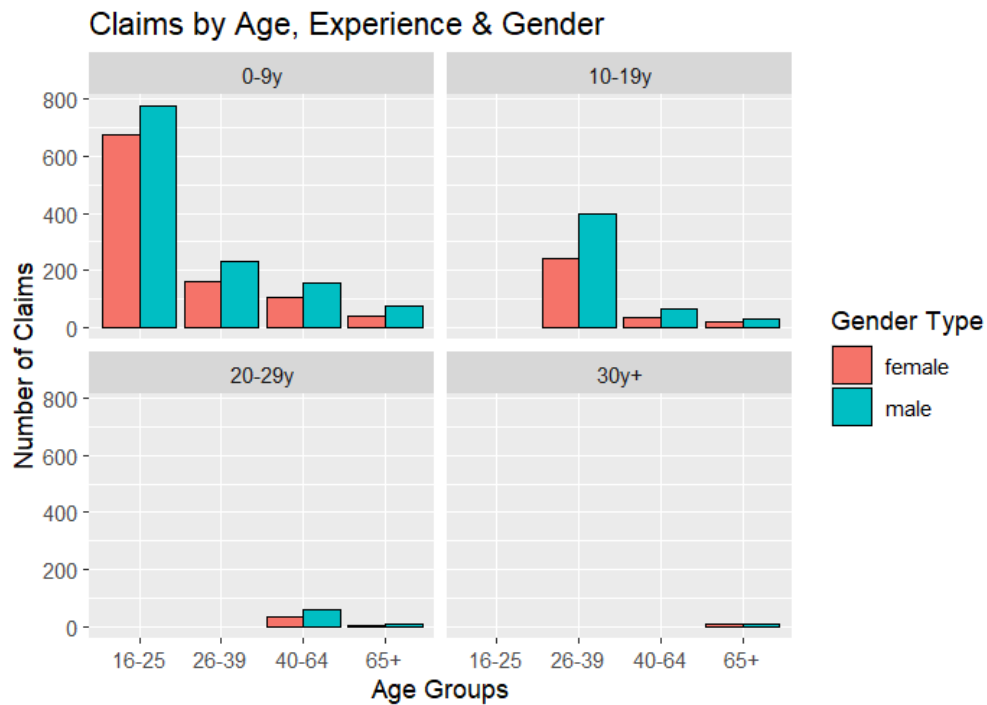
Mid-Term EDA Project

(Analytical questions + Explanations)

Thant Sin Tun
Dr. Nwe Nwe Htay Win
Data Communication and Ethics
Department of Mathematics & Science
Parami University
4.5.2025

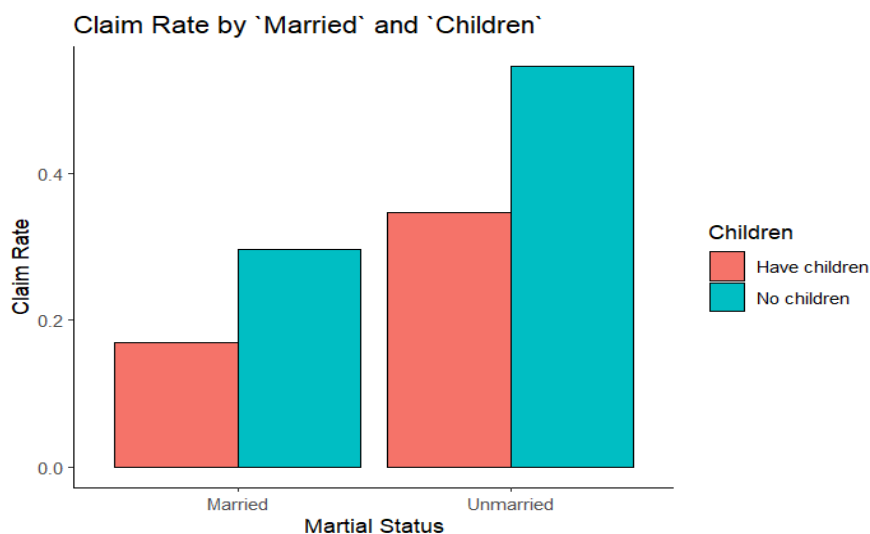
Customer & Vehicle Profile

Question 1: Which driver profiles (age, experience, gender) are most likely to result in claims (outcome = True)?



According to the graphs, **young and less-experienced(0-9 years) drivers** (both male and female) are most likely to file claims. Moreover, it is notable that elderly people are less likely to file claims. Therefore, targeting Young and Less-Experienced Drivers (0-9 years) could reduce claim risk.

Question 2: Are married customers with children less likely to file claims?

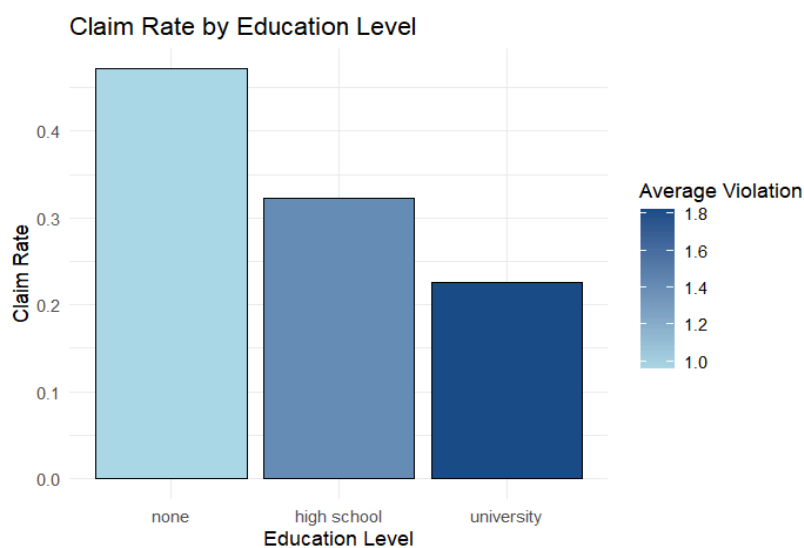


According to the bar chart, married customers with children are less likely to file claims.

- Furthermore, unmarried customers with no children have a high claim rate.
- Customers with no children have a higher claim rate than those with children, regardless of marital status.

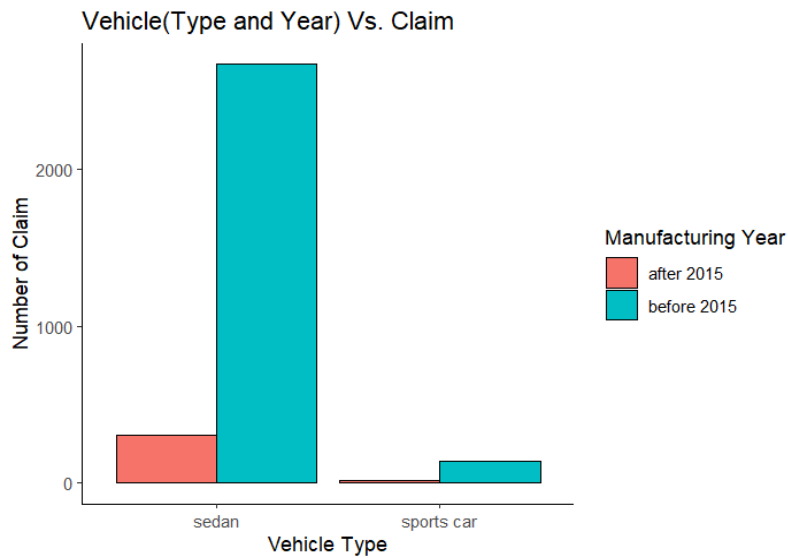
Hence, promoting family-focused insurance to Married Customers with Children could enhance a stable, lower-risk demographic.

Question 3: Is there any claim pattern difference across education levels?



- According to the bar chart, the **Claim Rate Decreases with Higher Education Level**. Drivers with higher education levels tend to file fewer claims.
- People with no education get fewer average violations, but have higher claims. Moreover, university-educated drivers have slightly **more violations** (darker bar), but fewer insurance claims.
- Afterward, it is notable that **Not all violations lead to claims**. In some cases, drivers with a certain level of education might be involved in less serious incidents.

Question 4: How does the combination of vehicle type and manufacturing year influence the likelihood of insurance claims?

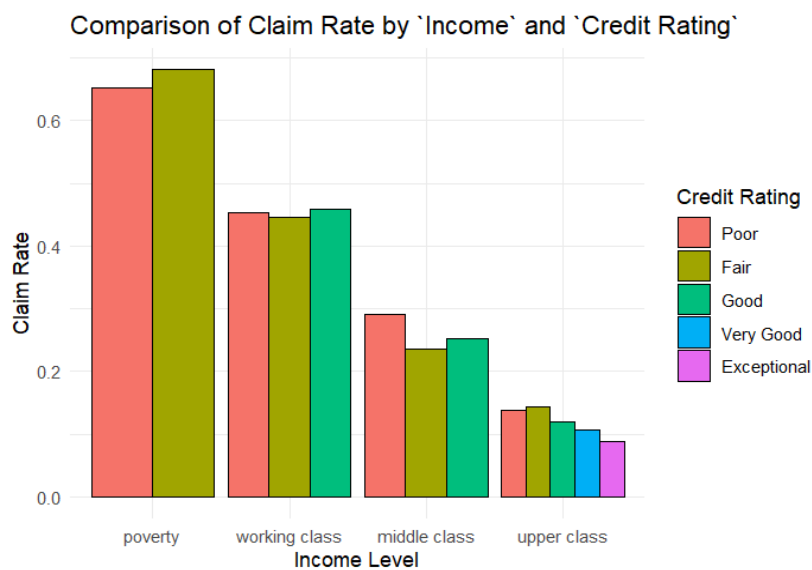


- Sedans dominate the claim counts, especially those manufactured **before 2015** have the highest claims. But, **after 2015**, newer sedan cars have fewer claims.
- Similarly, older sports cars have higher claims than the newer ones.
- Overall, older vehicles(**before 2015**) tend to be associated with **higher claim rates**.

Focusing on Older Vehicles (pre-2015) and specific car types like sedans and sports cars could improve risk management strategies in the company.

Socioeconomic Factors

Question 5: How does the claim rate differ by income and credit score?



- Notably, lower-income individuals with poor or fair credit scores are more likely to file claims. Customers with **Very Good** or **Exceptional** credit show lower claim rates, particularly noticeable in the "upper" class.
- However, there is one abnormal case: **Good** credit ratings have a slightly higher claim rate than lower credit ratings in the working class.

Overall, higher income and better credit generally indicate lower risk.

Question 6: Which region has the most stable-income customers and better credit scores?

Regions with the highest average credit

| location | avg_credit_score |
|-------------|------------------|
| <chr> | <dbl> |
| 1 Maryland | 0.540 |
| 2 San Diego | 0.523 |
| 3 Florida | 0.515 |
| 4 Vilnius | 0.515 |

Regions with the most stable income customers

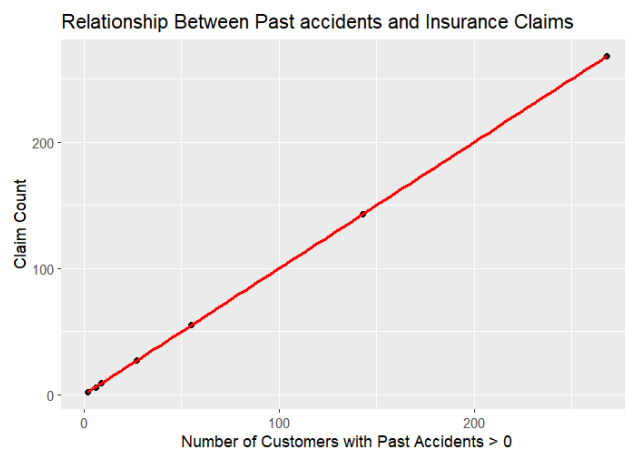
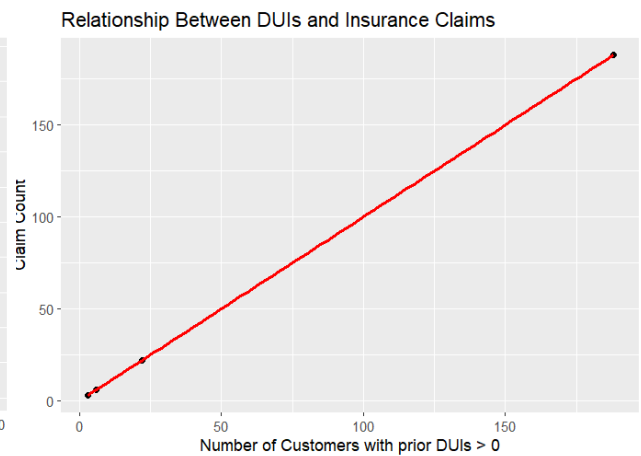
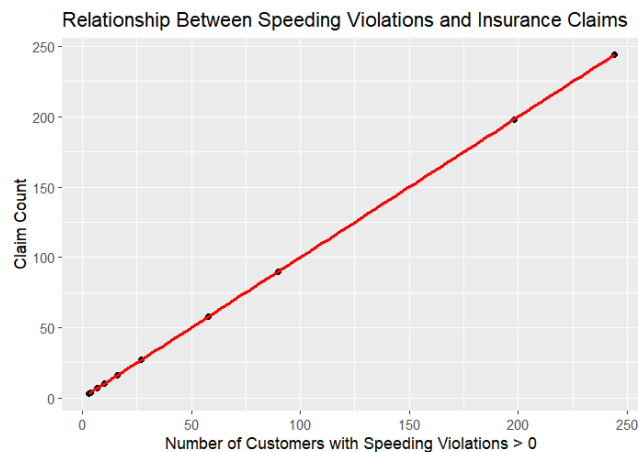
| location | stable_income_customer |
|-------------|------------------------|
| <chr> | <int> |
| 1 Vilnius | 4467 |
| 2 Florida | 1598 |
| 3 San Diego | 326 |
| 4 Maryland | 83 |

- While the average credit scores across regions are not significantly different, Maryland has the highest average credit score at 0.542.
 - On the other hand, Vilnius has the most customers in stable income categories.
 - Since upper classes are less likely to file claims, the company should focus on upselling services to a large upper-class market in Vilnius.
-

Driving Behavior and Risk

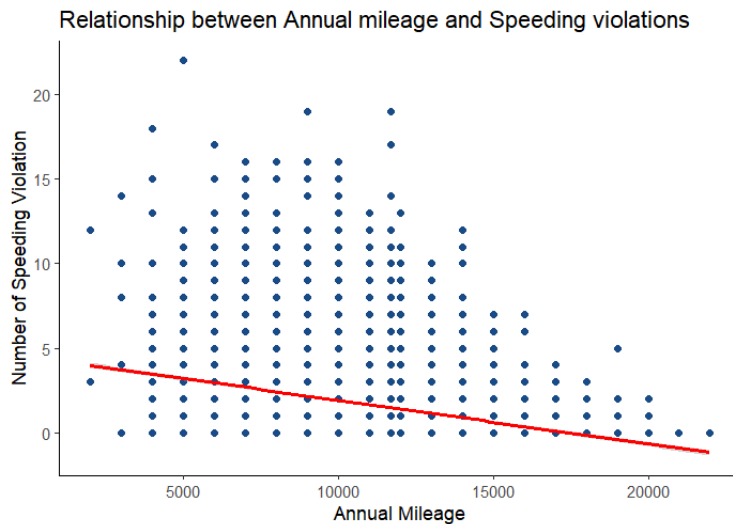
Question 7: What is the relationship between prior violations (speeding, DUIs, accidents) and claim occurrence?

Three scatterplots highlighting the relationship between insurance claims & prior violations



- There is **a strong positive relationship between prior violations and insurance claims** ~ linear and proportional.
- Drivers with more speeding violations, DUIs, and past accidents are significantly more likely to file claims.
- These trends are visually evident: as the number of customers with prior violations increases, so does the claim count.

Question 8: How does annual mileage relate to prior offenses (speeding violations, DUIs, past accidents)?



There is **a negative correlation** between annual mileage and the number of speeding violations, despite a slight flat line.

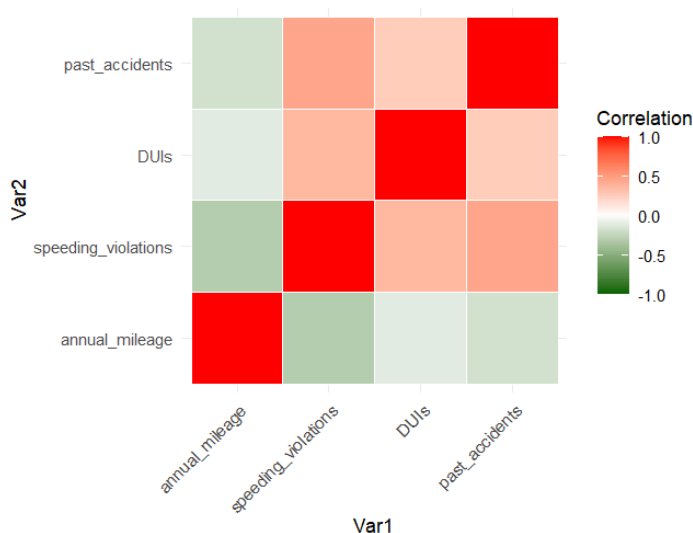
Drivers who accumulate more miles annually tend to commit fewer speeding violations.

“Like experiences talk”

Drivers with low annual mileage may be younger and less experienced, leading to a higher likelihood of speeding violations.

The following **“correlation matrix” and “heatmap”** show the relationship between speeding_violation, DUIs, past_accident, and annual mileage.

| | annual_mileage | speeding_violations | DUIs | past_accidents |
|---------------------|----------------|---------------------|------------|----------------|
| annual_mileage | 1.0000000 | -0.3081252 | -0.1112322 | -0.1871803 |
| speeding_violations | -0.3081252 | 1.0000000 | 0.3598378 | 0.4430742 |
| DUIs | -0.1112322 | 0.3598378 | 1.0000000 | 0.2593588 |
| past_accidents | -0.1871803 | 0.4430742 | 0.2593588 | 1.0000000 |

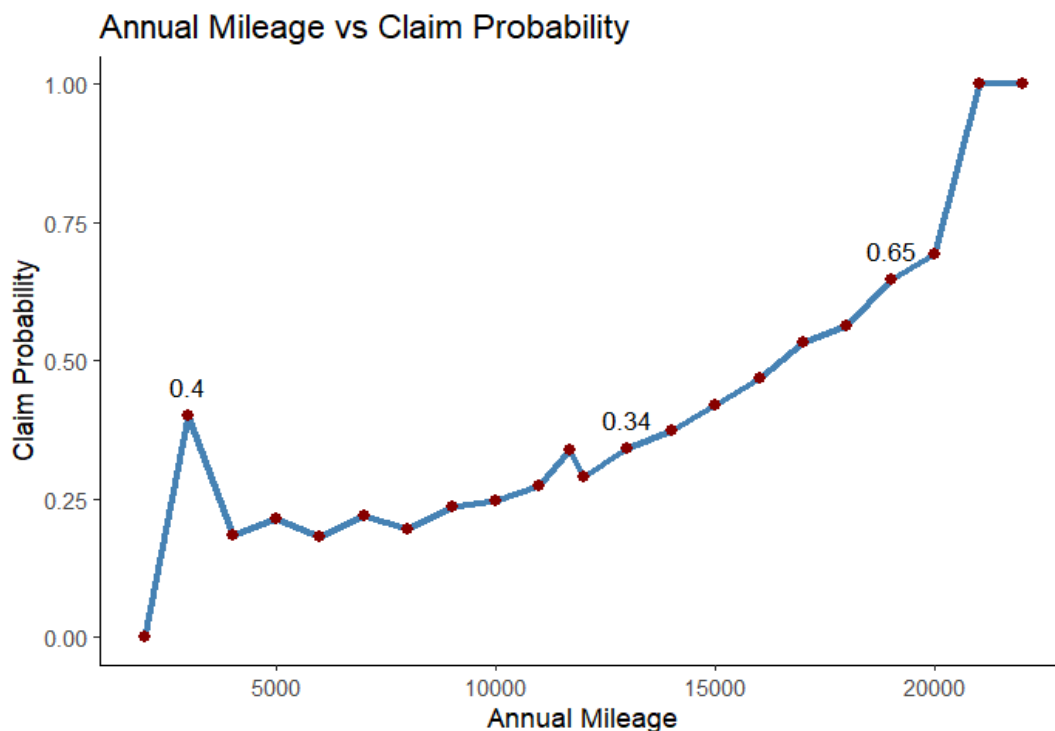


- High speeding violations are a strong risk indicator (linked to both DUIs and accidents)
- Interestingly, annual mileage shows a moderate negative correlation with all types of prior violations.

- Despite no signs of strong multicollinearity, it might be considered that more experienced drivers tend to commit fewer violations and be safer.

For business decisions, addressing high-risk drivers with prior violations (speeding, DUIs, accidents) is crucial, as they have a higher likelihood of filing claims. Implementing awareness campaigns and creating warning signs on highways might be popular solutions.

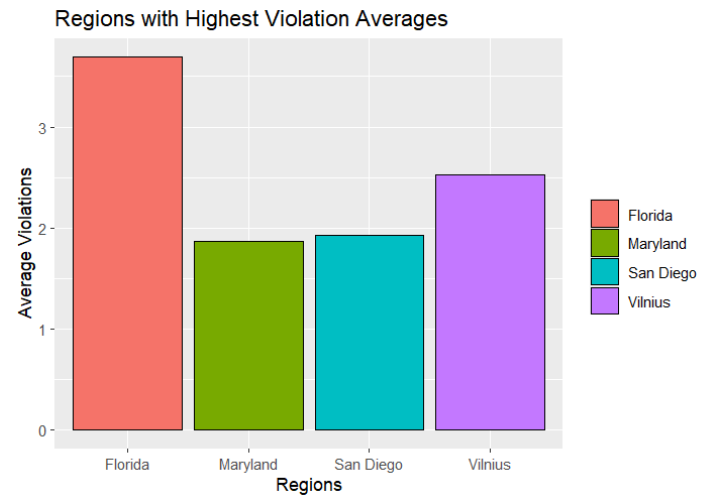
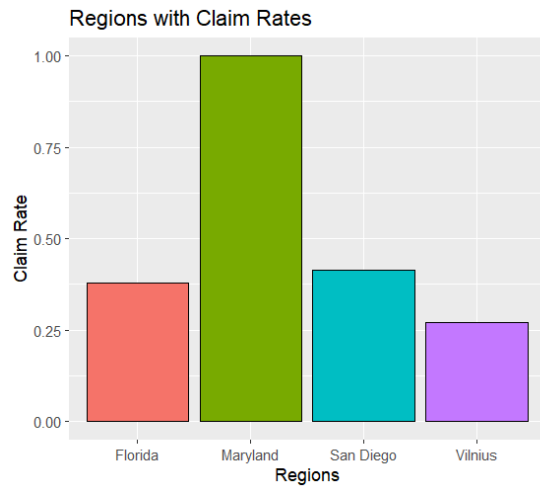
Question 9: How does annual mileage relate to claim probability?



According to the line chart, **positive relationship** between annual mileage and claim probability, but it's non-linear.

- Low mileage drivers (under 3000 miles) might include less-experienced drivers, as claim_prob increases sharply to 0.4.
 - Claim probability steadily increases from 0.34 to 0.65 for drivers with between 12000 and 20000 annual mileage.
 - On the other hand, very high mileage (above 18000 miles) is associated with the highest risk of filing a claim.
-

Question 10: Which regions (postal_code or location) have the highest claim rates and violation averages?



- Accordingly, **Maryland** has the highest claim rate.
 - In terms of average violations, **Florida** is at the top (3.68), but has a lower claim rate (0.38).
 - Despite lower claims, they violate the rules a lot, which may lead to future claims in Florida. Therefore, it might require preparing for future risk.
-