



smartcademy

CERTIFICATE OF COMPLETION

THIS IS AWARDED TO

NURSHARINAH BINTESOHAIMI

in recognition of having successfully passed the

DATA ANALYTICS COURSE
(IBF-STS: P210610VYZ)

COURSE TAKEN FROM

06/12/2021

to

13/12/2021

START DATE

END DATE

and has achieved a grade of

MERIT

GRADE

Jason Li

JASON LI

HEAD OF ACADEMIC

CASE STUDY 3: INSURANCE PREMIUM DATA

Nursharinah
Binte Sohaimi



INTRODUCTION

BUSINESS MODEL

Underwriting

- Charging a fee (i.e., premiums) for taking in financial risk, via insurance policy, and collectively deduct it from paid out claims & business operation fees.

Investment Income

- Reinvesting a portion of reserves from monthly premiums collected into other interest-generating assets in the financial markets, to increase revenue.

VALUE PROPOSITION

Underwriting

- No upfront of heavy investment needed, and only must pay out if customer has made legitimate claim when met with an adversity during the tenure of the policy.

Investment Income

- No cash down needed to produce a product & more funds can compliment an insurer's investment portfolio. Despite investing conservatively, the gains can also potentially offset business operation fees.

BUSINESS DILEMMA

Insurance companies face ongoing challenges in optimizing profits and taper losses through their business models. As the world generally progresses, trigger points for conditional payment over the length of policy may varies and insurers must be steadfast in retrofitting their financial products to stay relevant. This analysis aims to explore key matrices that will affect the content of policymaking by using the insurance premium dataset to identify insights on the public general wellbeing.

DATA MINING

Source

Downloaded from Kaggle.com

A huge repository of community published data & code

Dataset (metadata)

Represents insurance premium data (for basic health policy)

Sources: Census Bureau

Collection methodology: Publicly available dataset online

Maintainers (owner): Simran Jain

Date created: 2020-09-09 (last updated)

Data Content

Columns: Age, Sex, BMI, No. of Children, Smoker/Non,
Region, Charges

Rows: States details of one policyholder

EXPLORATORY DATA ANALYSIS

DATA QUALITY ASSURANCE:

- ❖ Data cleansing was performed with Python using Jupyter Notebook.
 - to improve data quality and prepare it for analysis and interpretation.
- ❖ Data quality was ensured against the 6 key standards:
 - Completeness/ Comprehensiveness
 - Consistency
 - Accuracy
 - Format
 - Timeframe
 - Validity/ Integrity

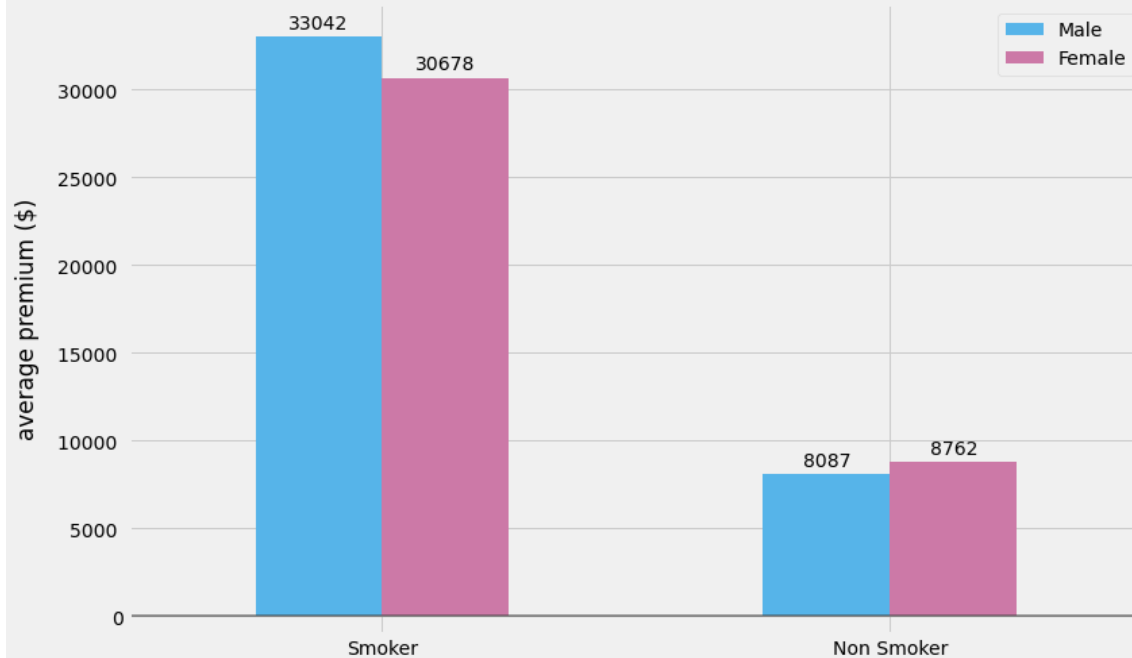
DATA CLEANING STEPS:

- 1. Initial verification of database:**
 - a. Check first & last 5 rows of table (have general view of rows & columns)
 - b. Determine no. of rows & columns, Datatypes, and Non-null Count
- 2. Verify any “dirty data”:**
 - a. Use .describe() function as statistical computation for integer/float variables
 - b. Use For loop & .nunique() functions for object variables to find unique values and if the values match category of columns
- 3. Data manipulation:**
 - a. Change appropriate column name in dataframe (e.g.: sex→gender , charges→premium)

EXPLORATORY DATA ANALYSIS

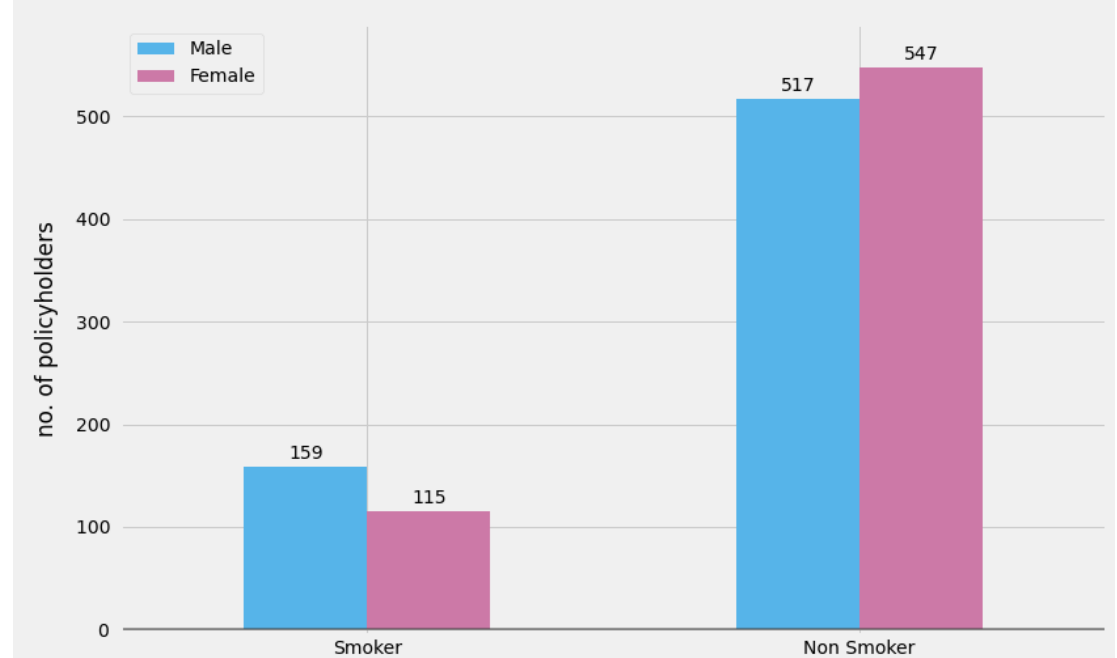
Smoke inhalation leads to greater expenses

On average, smokers are generally charged a higher premium (above \$30,000) than non-smokers (below \$9,000). The amount is almost thrice than those less at risk!



The gap between Smoker & Non-smoker Demographic

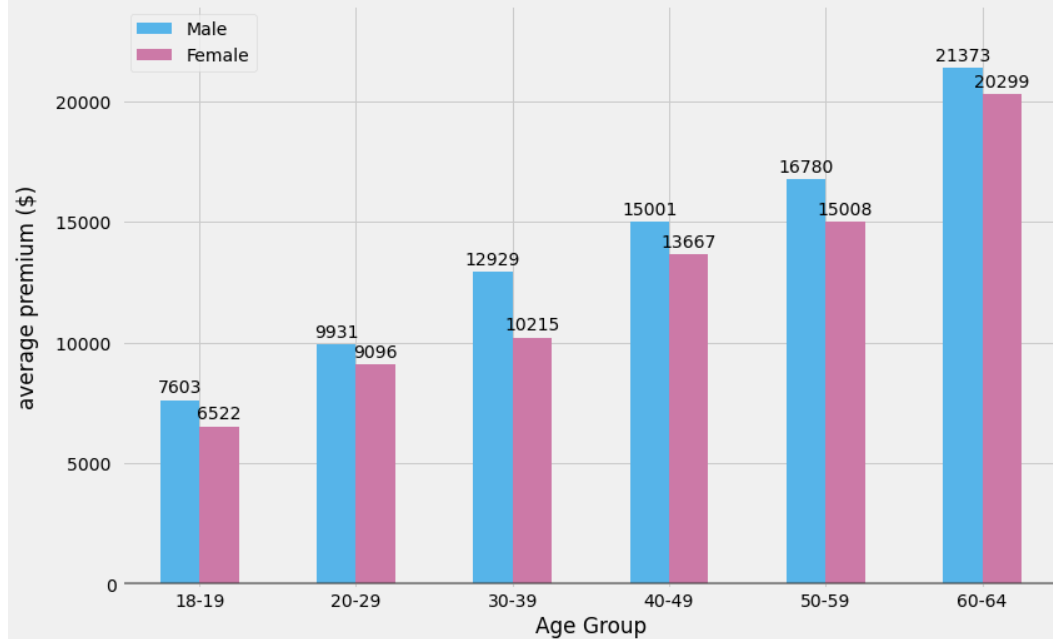
The sample size consists of 1338 individual policyholders. Of which, Smokers takes up 20% and Non-smokers are of 80% of the population.



EXPLORATORY DATA ANALYSIS

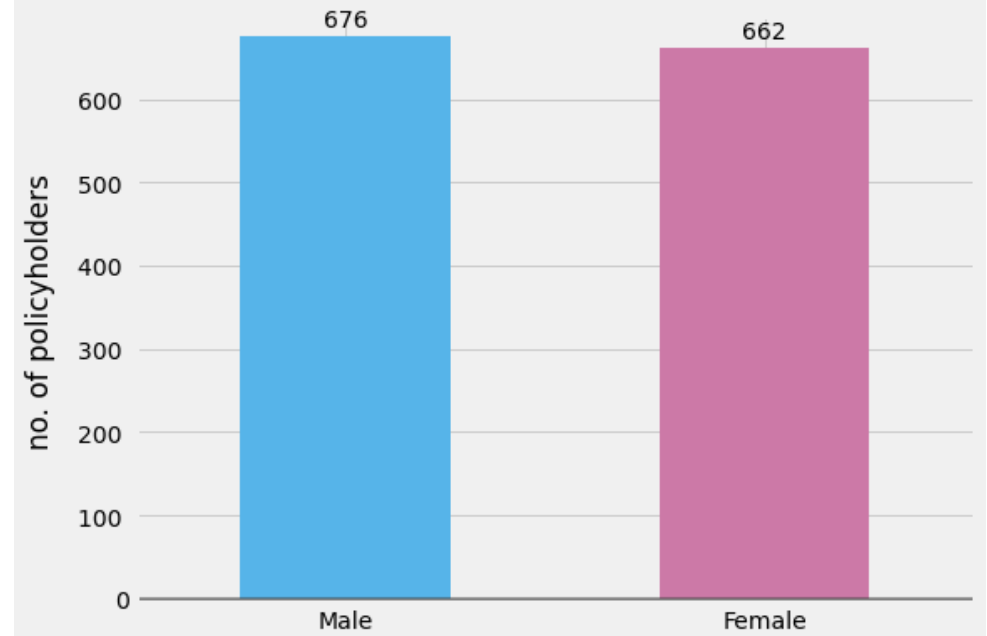
Shorter life expectancy but men pay higher fees

Premium charge is age dependant as the older the policyholder is from the time of getting insured, the higher the fees. Also generally males are charged slightly higher amount than females.



The Gender Demographic

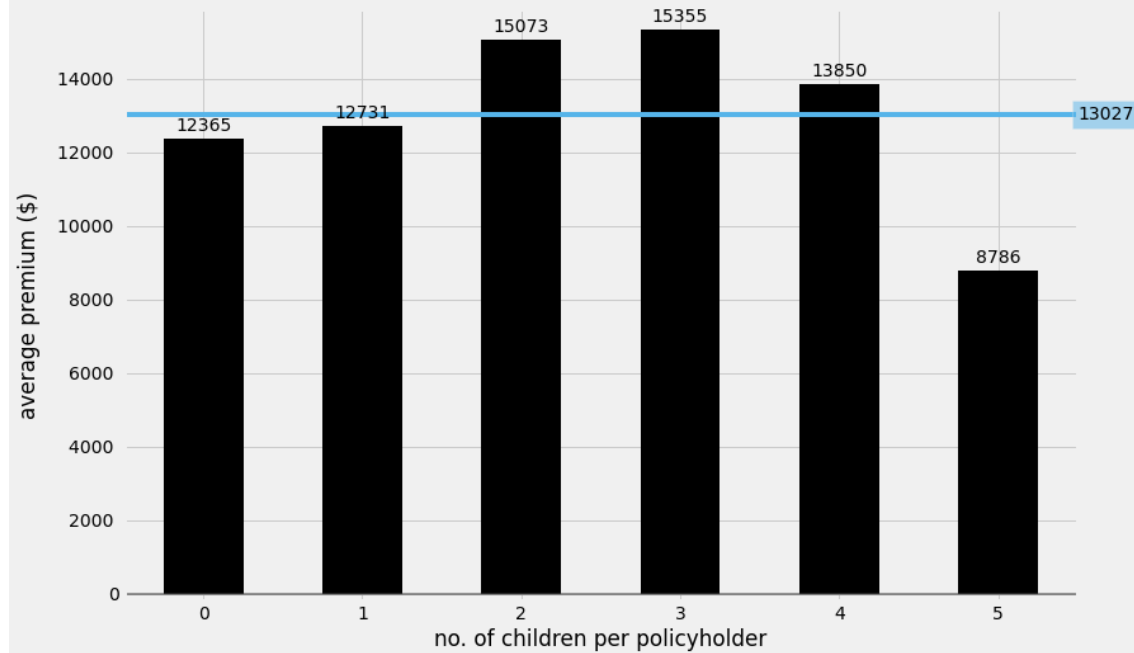
The sample size consists of 1338 individual policyholders. Of which, the gender distribution is about 50%.



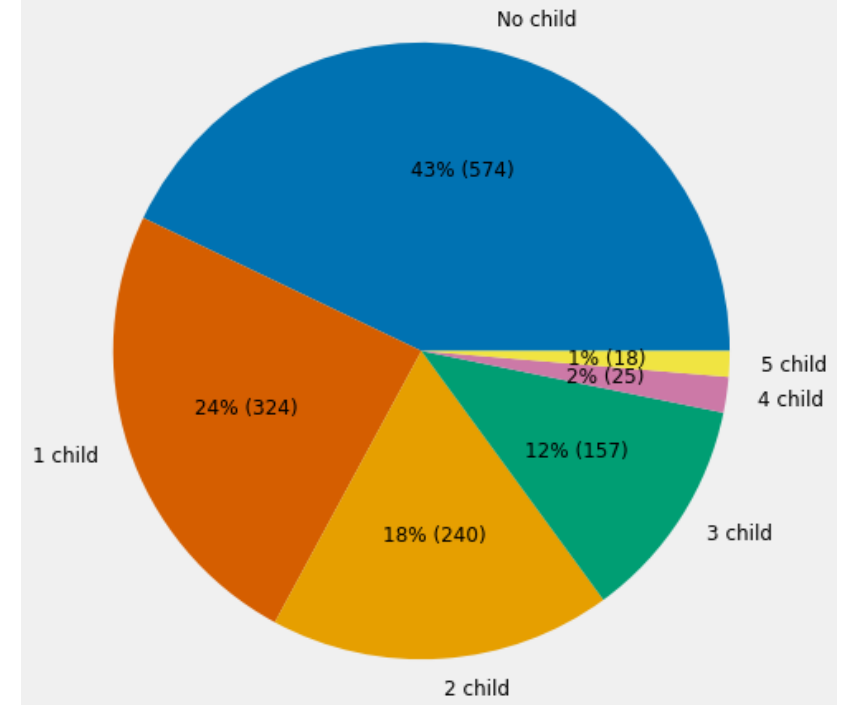
EXPLORATORY DATA ANALYSIS

More children, More liabilities

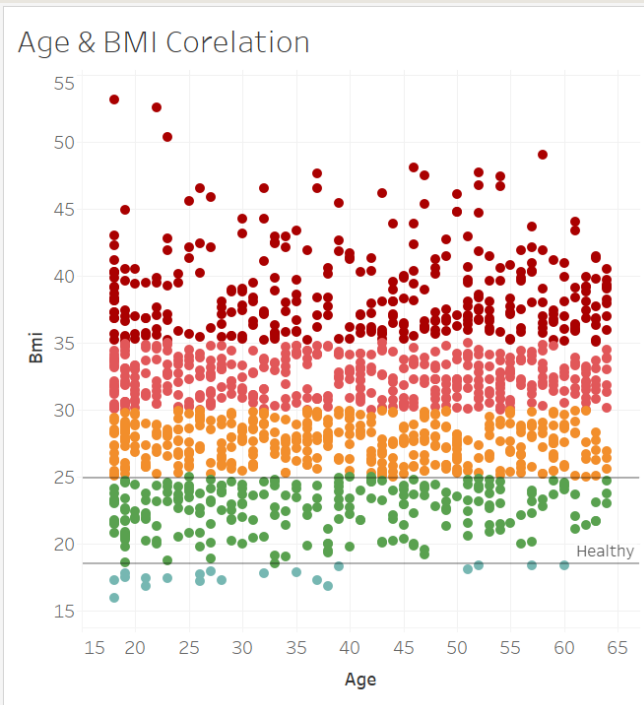
As a policyholder has more children, the average premium increases gradually.
The general average charges is \$13,027 among the child dependant distribution.



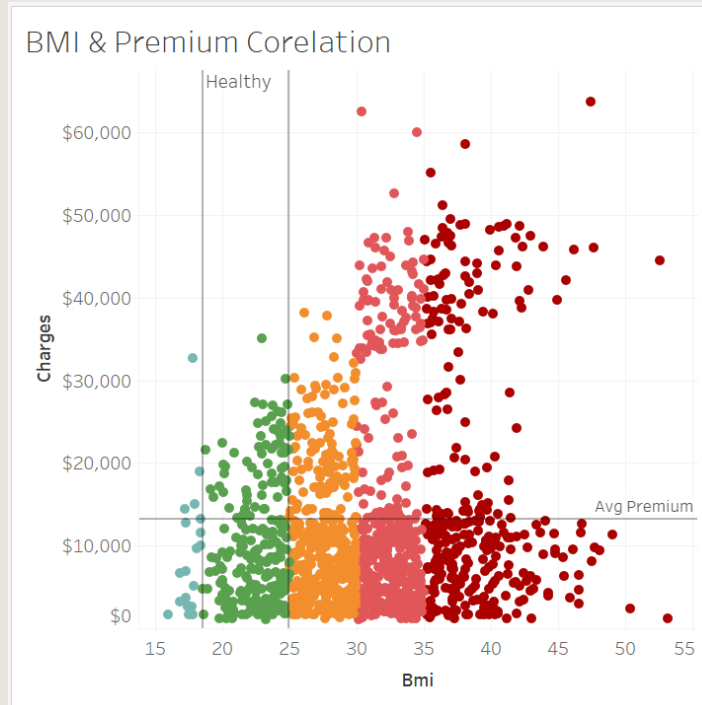
Child Bearing per Policyholder



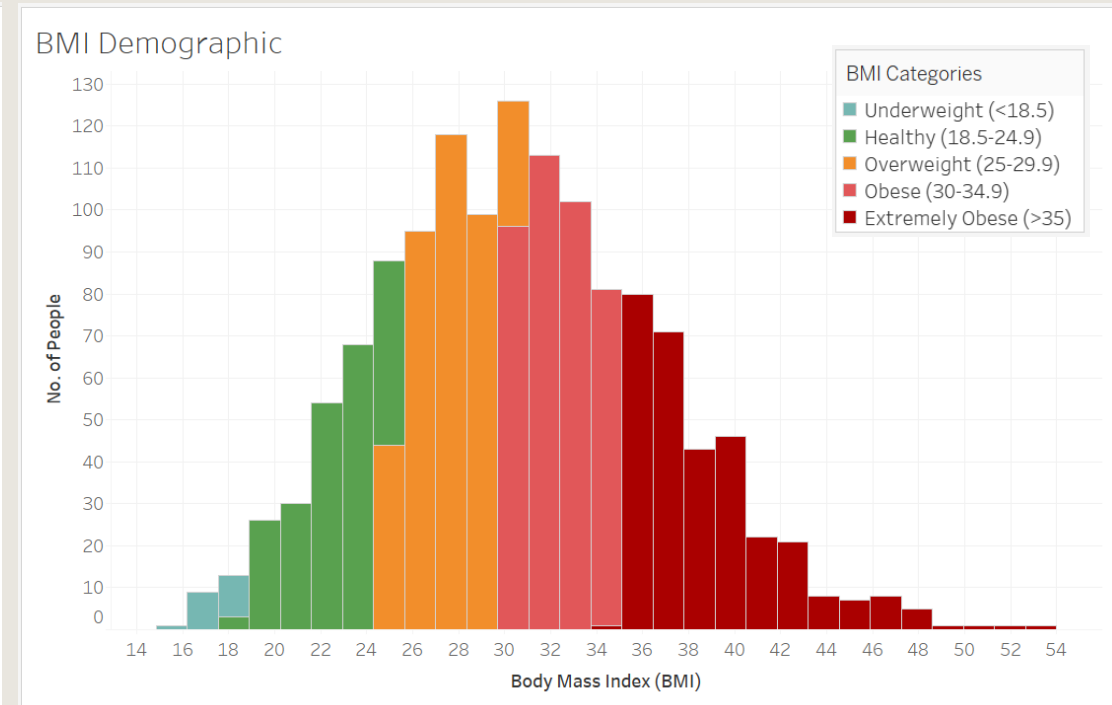
DASHBOARD



No matter the age, consistently there are trends of people who are highly at risk due to their unhealthy BMI. Also as the population gets older, the more unhealthy they become.



One factor that determines the average premium of \$13,270 throughout the pool of clients are due to groups of “Overweight” & “Obese” that skews the amount being charged in general.



From sample size of 1338 policyholders, most of the population has an average BMI of 30.7 and are within the unhealthy range. More alarmingly, the data shows how heavily weighted people are on being “Obese” & “Extremely Obese”.

DATA ANALYSIS FINDINGS

1. Premium rates of smokers (despite gender differences) is minimally 3.5 times than non-smokers. This is considering that their organs might be affected due to tobacco abuse.
2. Men are charged at a slight (~10%) higher than women. One possibility is due to observation trends that men has shorter lifespan for reasons largely related to genetics & hormones.
3. The older the age group, the higher the premium amount. Because insurance policy is determined at the time of purchase and set for the duration of the policy.
4. When a policyholder has more children, their premiums goes up as they are charged for each person covered in the plan.
5. BMI is used to signify health status of an individual's body. The higher the BMI, the higher chance of getting affected by obesity-related illnesses. Hence it affects rate of premium due to higher medical expenditure chances.

STRATEGY: KEY RECOMMENDATIONS

- Relook into the policy for those with more than 5 children. These group of people are charged below the average premium of \$13,027. Also, their premiums are way lesser compared to those with no children.
- It is debatable to use BMI as being the best health indicator. Because multiple factors in the human body affects overall health. Examples includes, being a smoker and/or family history of health issues. Thus, solely adding high weightage to BMI parameters for average premium charged would lead to potential major losses for the insurance company.



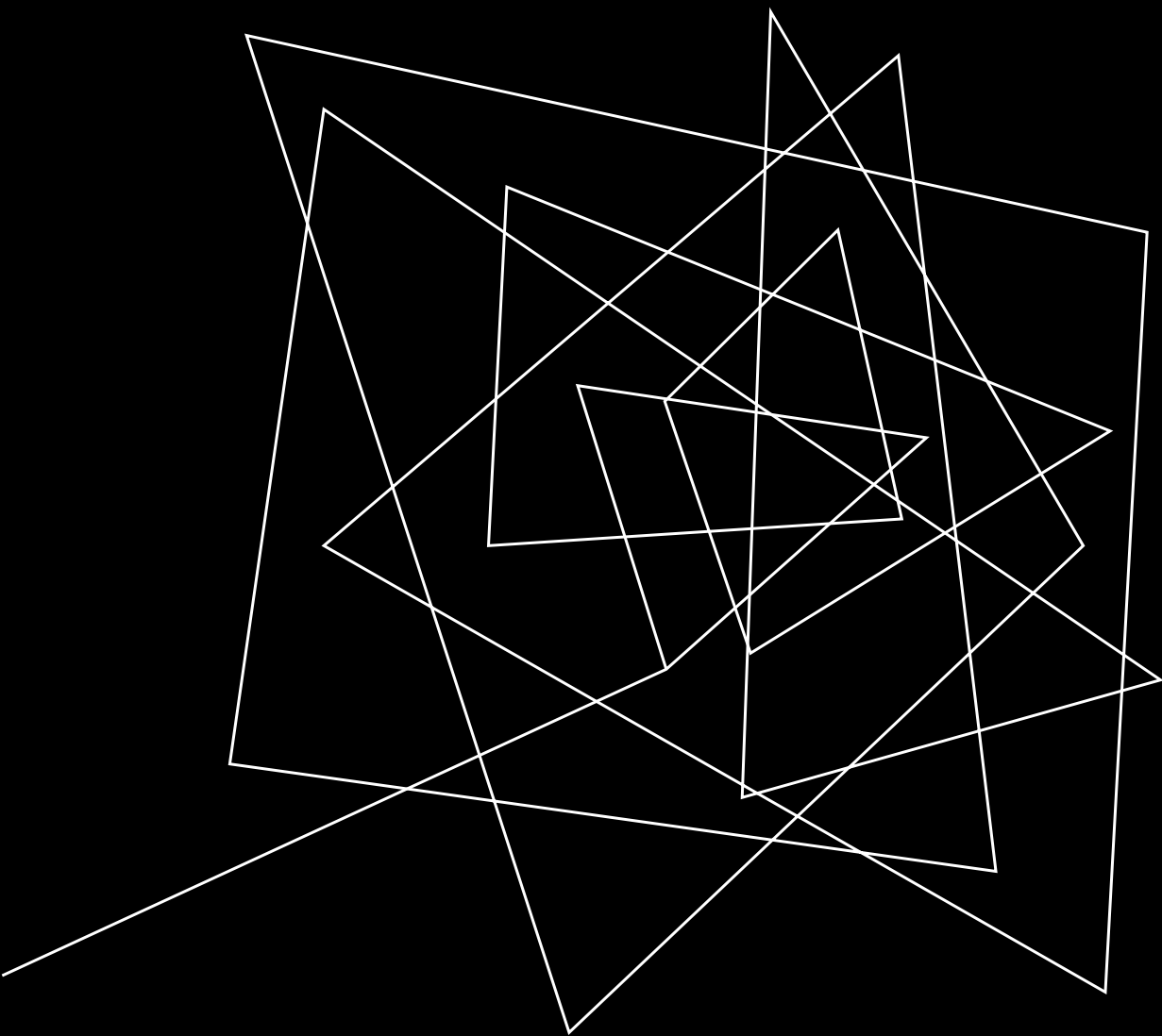


CONCLUSION & FUTURE WORKS

Insurance companies base their business models around assuming & diversifying risks. It is essential that the model involves pooling risk from individual payers & redistributing it across a larger portfolio. Hence it is important for insurance companies to try market effectively.

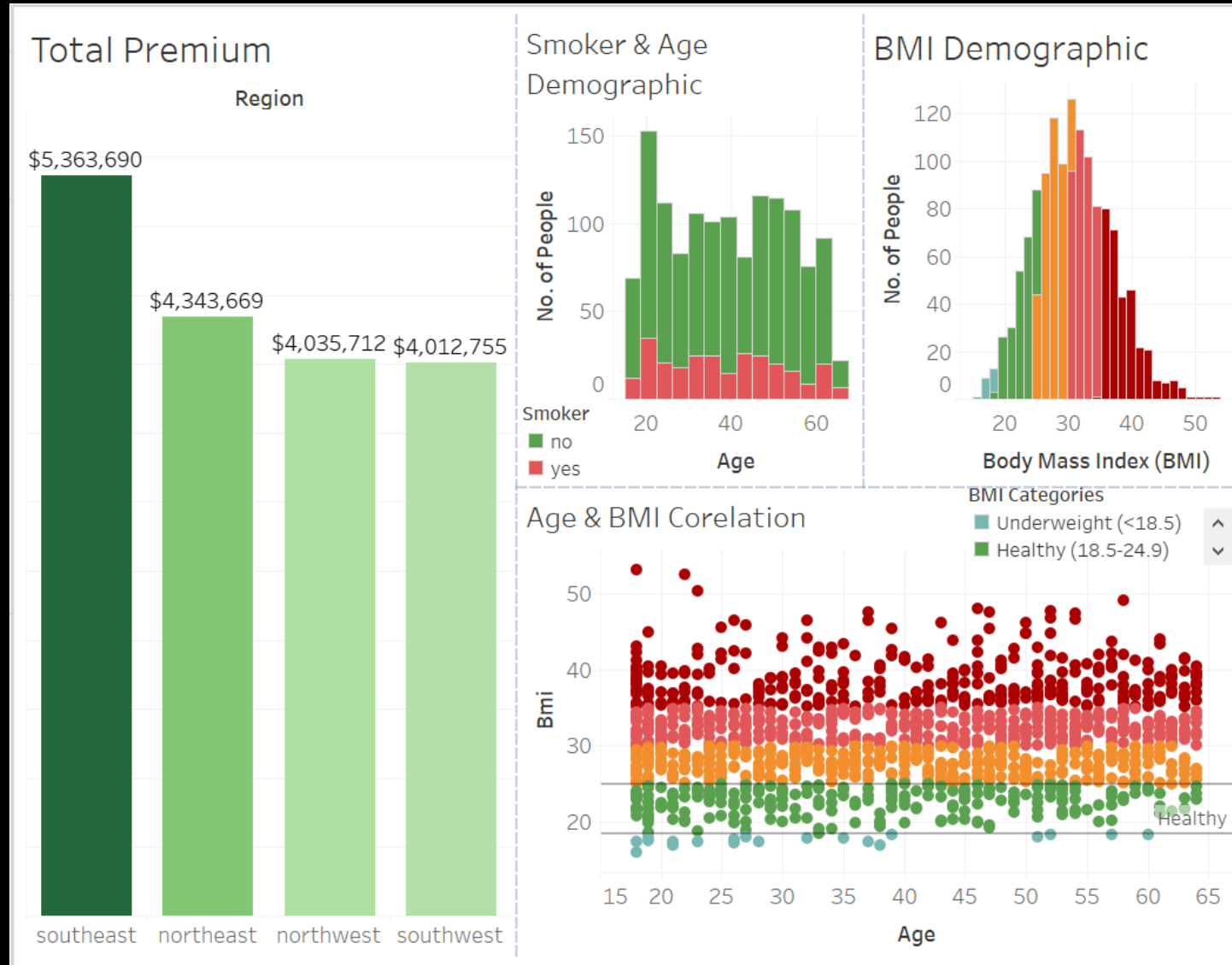
This is achieved by assigning applicants to various risk classes, based on factors such as age, gender, overall health & smoking status. However, improvements can be made to include aspects like family history, presence of preexisting conditions & occupation into consideration.

More exploration in the data should also be done to find more concrete relations between the parameters that would affect the amount of premium charged. For example, doing a deeper study on correlations of health factors (e.g., BMI, smoking status, body fat, personal & family health history) and what/how it affects them. This could potentially lead to a change in policymaking to reduce unnecessary losses.



THANK YOU.
APPENDICES

TABLEAU DASHBOARD



REFERENCES

- [1] O'Connell B. (2019, May 30). How Do Insurance Companies Make Money? TheStreet. Retrieved from: <https://www.thestreet.com/how-to/how-do-insurance-companies-make-money-14971728>
- [2] Ross S. (2021, July 28). What is the Main Business Model for Insurance Companies? Investopedia. Retrieved from: <https://www.investopedia.com/ask/answers/052015/what-main-business-model-insurance-companies.asp>
- [3] Speights K. (2021, October 27). How Do Insurance Companies Make Money? The Motley Fool. Retrieved from: <https://www.fool.com/investing/stock-market/market-sectors/financials/insurance-stocks/how-insurance-companies-make-money/>

SELF-REFLECTION

With past experiences of using Cprogramming & Arduino during my diploma & university degree studies has helped me to easily understand Python language. Although I do find some challenges trying to conform to how Python code structures are written. Thankfully through this project, I was able to practice more and get accustom through trial and errors. Critical thinking and self-help skills that I gained while doing small projects or Final Year Project during my academic studies, supported me to be resourceful in looking for help online.

Also, another beneficial skills I got to practice was reverse engineering. Which is to look at the codes available, understand what they do and try to replicate and/or tweak the code accordingly to how I wanted. Advice from Dr Alvin on envisioning the outcome and then breakdown into small steps in achieving them has helped me greatly in this project as I struggled on how to go on about doing this initially.

With 7 years of engineering academic background (electrical & mechanical), I find it tough in trying to understand how the finance industry works. But researching on how the business works and understanding the critical pain points of the potential client helped me to have a better view of their perspective and cater the data analysis to their needs.

Lastly, I do find some frustrations in trying to group the data and displaying it into graphical format. It was challenging at first since I was a novice, but now I got to understand the pains of dealing with unstructured (strings, words) data which is tougher than structured (numbers) data. However, these challenges didn't deter me but yet fuel my curiosity on how I can resolve them. It gives me a sense of exhilarating accomplishment when I got to see my work done exactly to how I envision them. In all, this has been truly a great experience for me to practice on my newfound skills and I am excited to continue my journey into data science, starting off as a data analyst and building up my expertise!