Healthcare Dataset Challenge

Overview of the approach taken by Chua Ding Yuan

Overall Approach

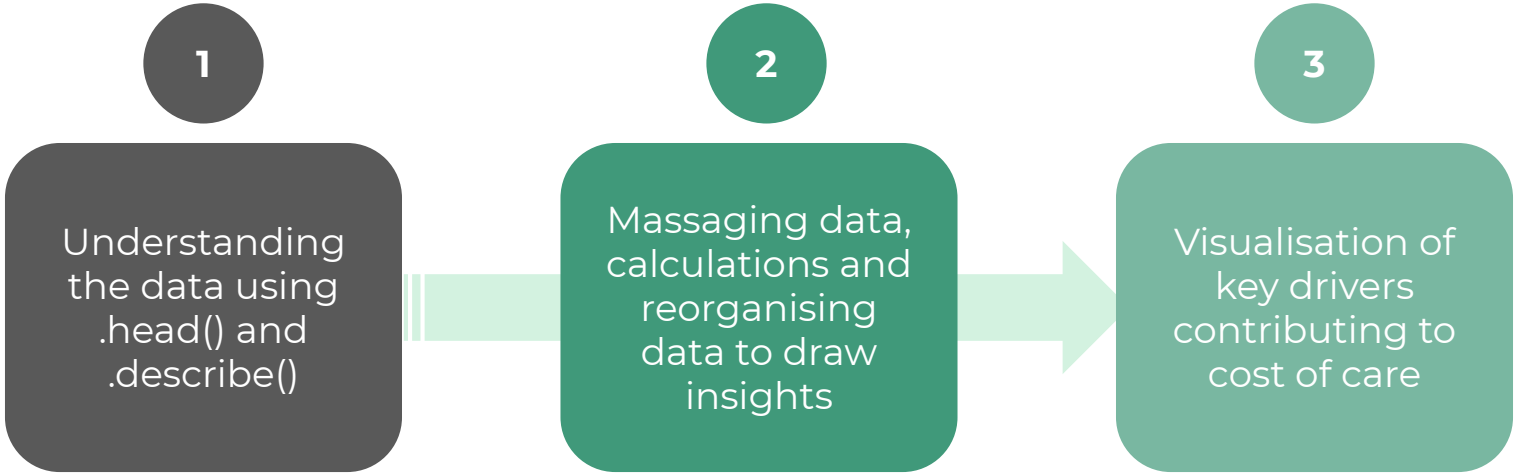
1. Using .describe() and .head() to understand data to be dealt with
2. Rearranging datasets to find basic statistical inferences of dataset
3. Finding further insights about the drivers of cost of care.

Age weight height vs cost
Medical condition vs cost
Resident status vs cost

Age medical condition vs cost (how to compare this)
Medical history symptom vs cost

Overall Approach

Problem Statement	"You are required to join the data given in different tables, and find insights about the drivers of cost of care."



Breakdown of given data sets

Given Data

Bill Amount

- 13600 rows
- Shows amount for each bill id
- Total of 13600 bills

Demographics

- 3000 rows
- Shows demographics of each patient
- Total of 3000 patients in record

Bill ID

- 13600 rows
- Shows patient id and date of admission
- Comparing with clinical data, there were 4 bills per admission

Clinical Data

- 3400 rows
- Shows clinical data for each admission
- Patients, on average, were admitted 1 time.
- But some were admitted more, up to 4 times

Self-added Data

Extra columns of data were calculated in order to try and draw more insights

1. Age of patients
2. Length of Stay
3. Number of admissions
4. Average cost of care of each admission for each patient

The main working dataframe is a combination of all 4 datasets showcasing the 3400 admissions of the 3000 patients as well as the extra columns added for further analysis.

Medical History has slight impact on cost of care

The Process

Find out if having a certain type of **medical history** affects bill amount

Find out if number of medical history affects bill amount

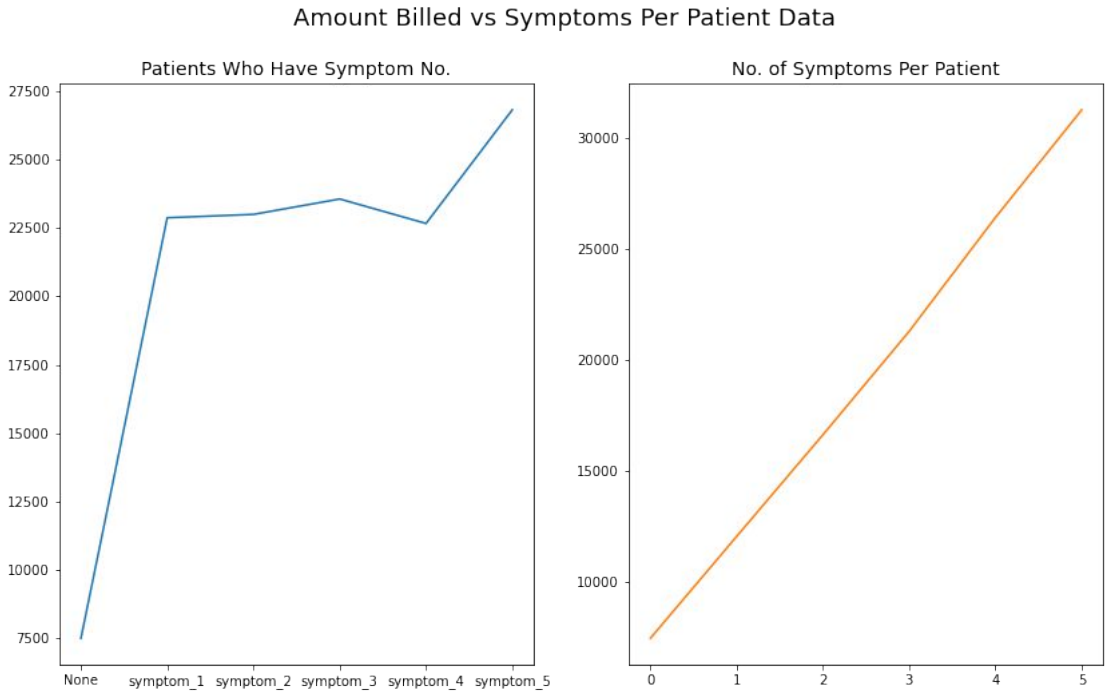
The Outcome

Patients that have **medical histories 1 and 6** experience slightly higher than the rest with medical history

Every patient with medical history incur **~20% higher** cost of care than those without

Lastly, as expected, higher number of medical history increases cost of care

Number of symptoms seen in a patient have a positive linear relation to cost of care



The Process

Find out if having a certain type of **symptoms** affects bill amount

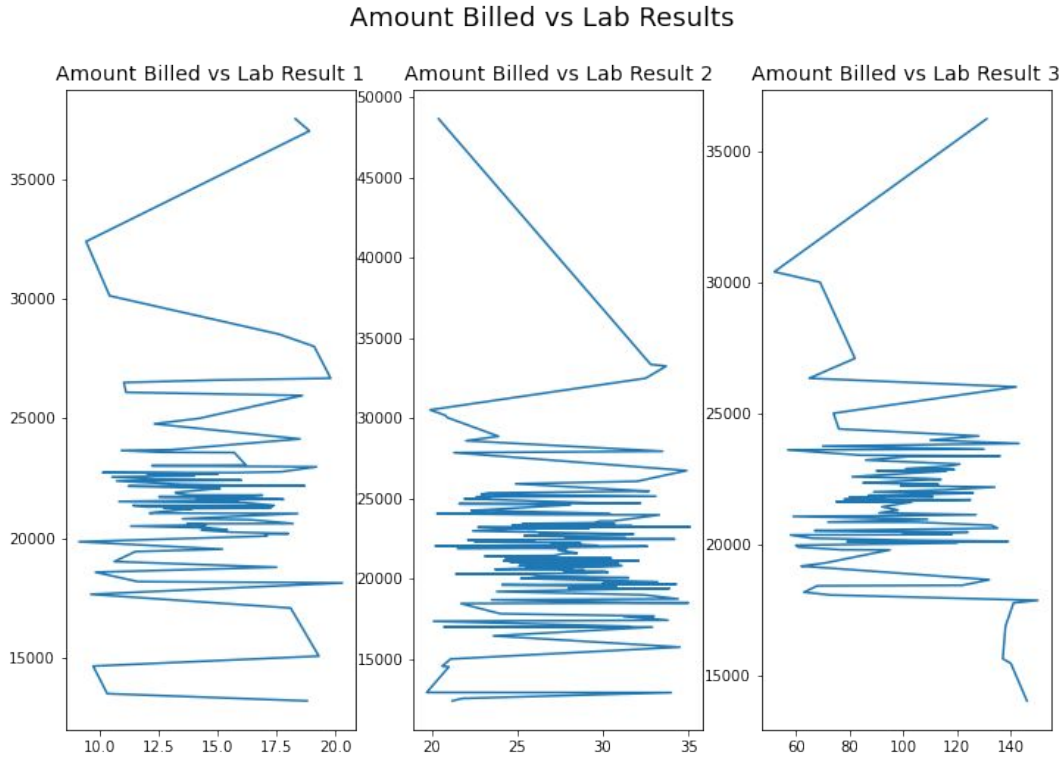
Find out if number of symptoms affects bill amount

The Outcome

There is a **positive linear relationship** between number of symptoms and amount billed.

Having **ANY Symptom**, increases cost of care **by 220%**

Lab results have no impact in determining cost of care



The Process

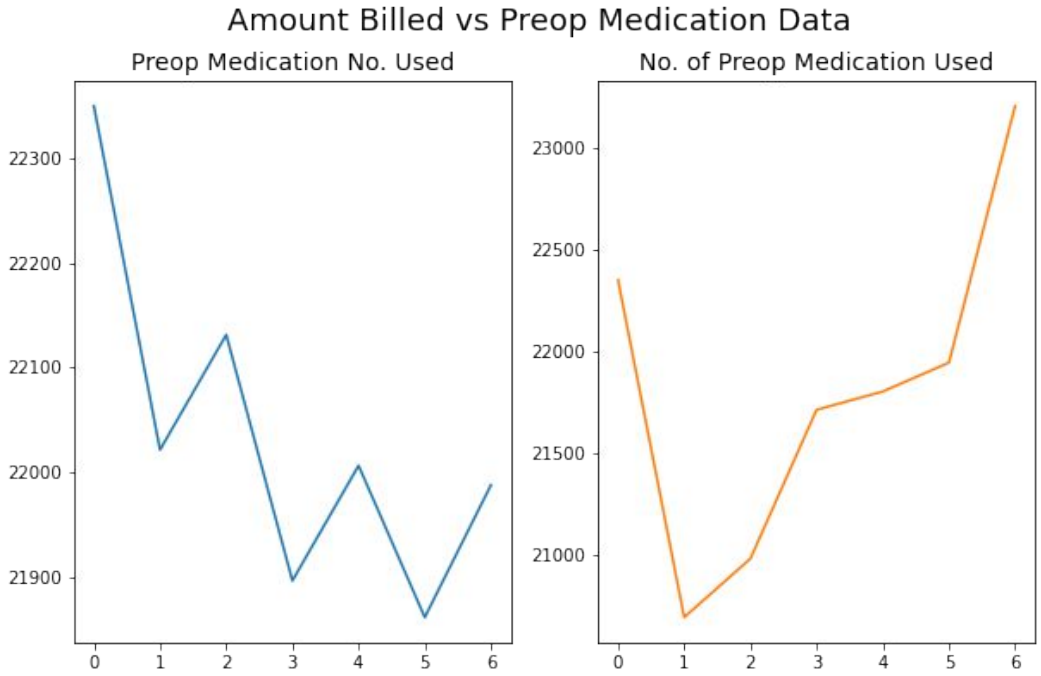
To determine a relationship between **lab results** and the amount billed (i.e cost of care).

The Outcome

There does not seem to be any relationship between lab results and the amount billed.

Each result for lab result 1, 2 and 3 have a wide range of amount billed.

Pre op medication usage decreases cost of care



The Process

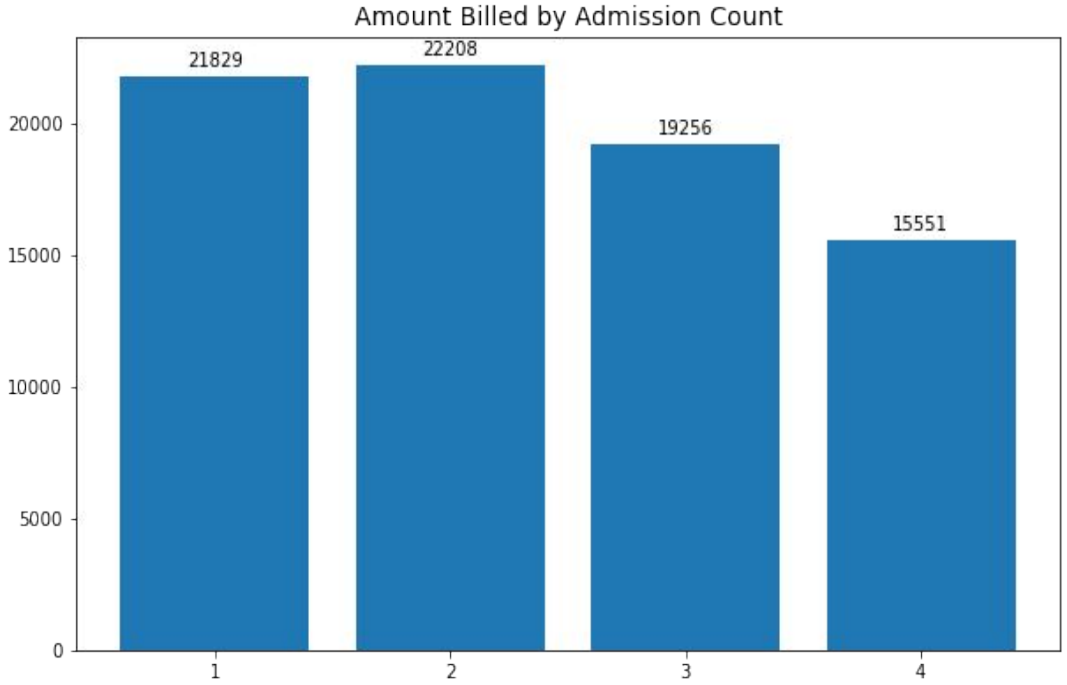
To determine a relationship between **Pre op Medication used** and the amount billed (i.e cost of care).

The Outcome

As expected, cost of care increases linearly when **1 or more** pre op medication is used. (orange chart)

However, the average cost of care when **no pre op medication** is used turned out to be **~2% higher** than if any one were used. (blue chart)

Having more admissions decreases average amount billed per admission



The Process

Taking the total amount billed per patient and dividing it by number of admissions to get the average amount billed per admission

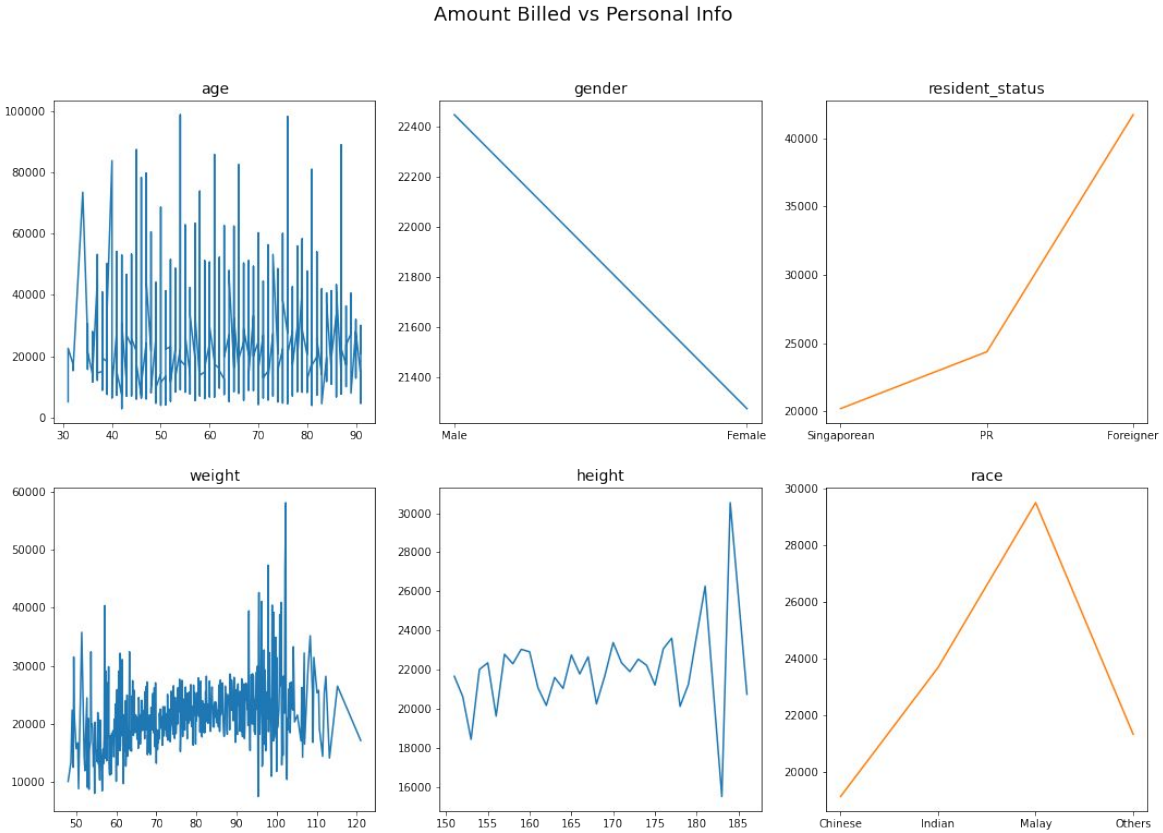
The Outcome

Patients admitted once or twice generally had to pay more per visit.

Generally **~40%** more than a patient with 4 visits.

This is expected because subsequent visits could be check ups that are not as costly.

Resident Status and Race is the largest driver of cost of care



The Process

Filter each variable under personal info (i.e. demographics)

Take the mean amount billed for each patient visit for each parameter

Plot the figures on the left

The Outcome

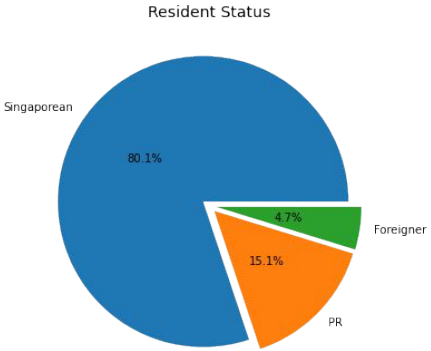
2 conclusions can be drawn

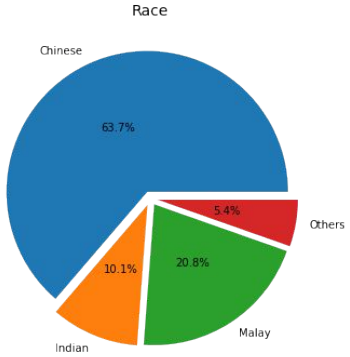
1. **Foreigners**, on average, pay **~87%** more than Singapore Citizens and PR
2. **Malays**, on average, pay **~40%** more than other races

Diving further into demographic drivers of cost of care for the Top 50 average amount billed

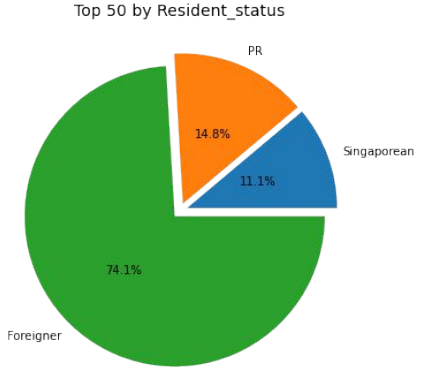
**The dataset was broken down into average amount billed per admission per patient.
Then sorted by average amount billed in descending order.**

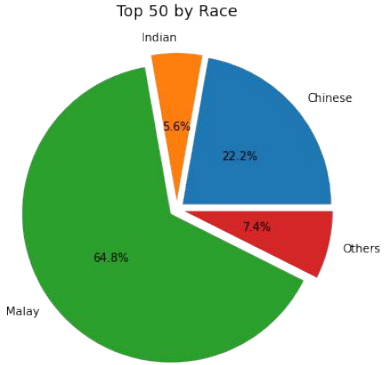
Demographics for all 3000 patients





Demographics for Top 50 amount billed





Malays and Foreigners make up a 20.8% and 4.7% respectively of their original demographics.
However, that increased to **64.8% Malays** and **74.1% Foreigners** in the Top 50 average amount billed.

Key Drivers of cost of care

1	Symptoms Shown	High Impact	Having any symptoms increase cost of care by ~220%	Largest driver for cost of care
2	Race and Resident Status	High Impact	Malay & Foreigners have ~40% and ~87% higher costs of care respectively	
3	Medical History	Med-High Impact	Having medical history increases cost of care by ~20%. Even more so for medical history 1 & 6.	
4	Number of admissions	Low Impact	Patients with 2 or less admissions are seeing ~40% higher average bills	
5	Pre Op Medication Usage	Low Impact	Usage of any pre op medication decreases cost of care by ~2%	

Thank You