Holmusk - Health Analytics Report

Shivaram Andiappan Selvaraj

Problem description:

To find insights from clinical data.

Data Engineering:

Four Datasets Bill amount, Bill id, Demographics and Clinical data are joined by left join query using bill id and patient id columns as join key.

Feature Engineering:

Variable	Formula
BMI	Weight in Kg / Height *Height in meters
Age	Date of Birth - Current Date
Days spent in hospital	Discharge Date - Admission Date

Data Imputation:

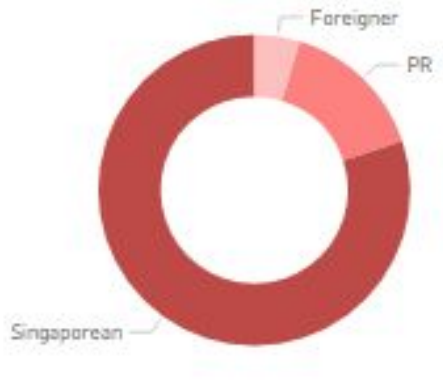
The missing data in Medical history 2 and medical history 5 variables are imputed using mode of those variables as they being categorical data type

Insights

Breakdown of Participant by Gender:

- 49.94% are male
- 50.06% are female

Breakdown of Participant by Resident Status:

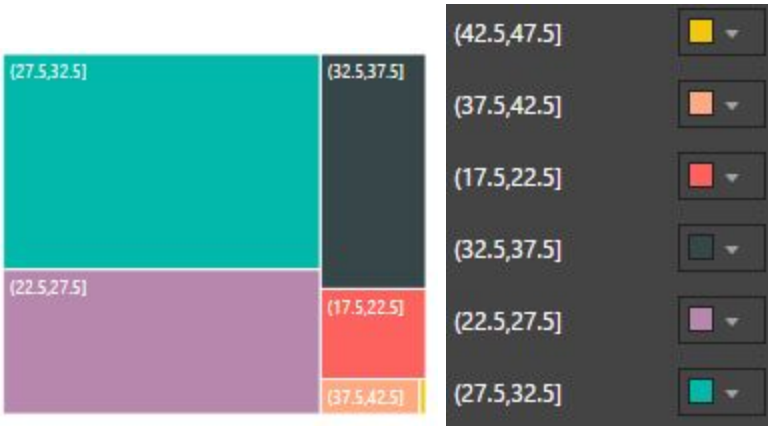


- 80.12% are Singaporean
- 15.5% are Singapore PR holder
- 4.74% are Foreigners

Breakdown of Participant by Race:

- 63.71% are Chinese
- 20.79% are Malay
- 10.12% are Indians
- 5.38% are Others

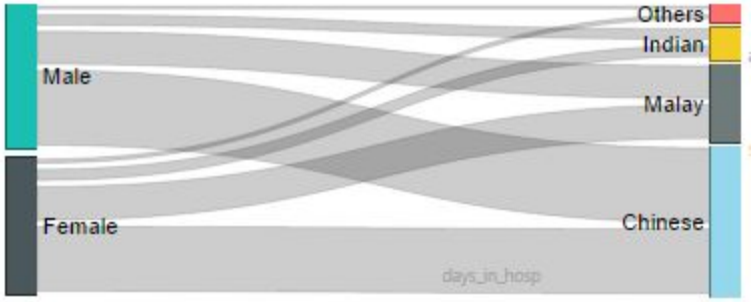
BMI Analysis:



BMI Index range	Count	Health Condition
17.5-22.5	212	Healthy Weight
22.5-27.5	1025	Healthy Weight
27.5-32.5	1527	Overweight
32.5-37.5	553	Obesity
37.5-42.5	78	Obesity
42.5-47.5	5	Obesity

More than 50% population are either in Overweight or Obesity category.

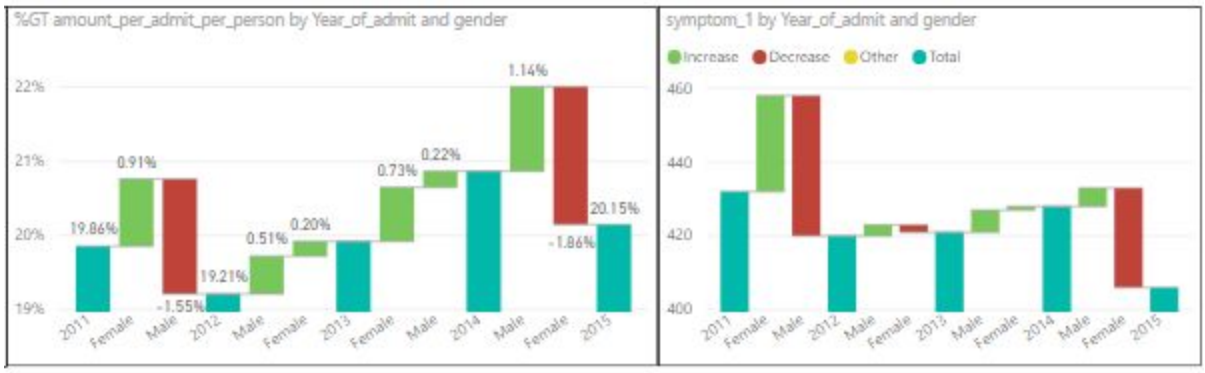
Flow of Medical Expense between Gender and Race:



The above Sankey diagram shows a clear visualization for drivers of cost between Gender and Race variable.

Gender	Race	Amount
Male	Chinese	22.11 M
	Malay	10.29 M
	Indian	03.98 M
	Others	01.73 M
Female	Chinese	19.30 M
	Malay	10.57 M
	Indian	04.16 M
	Others	02.18 M

Gender affects Medical expense:



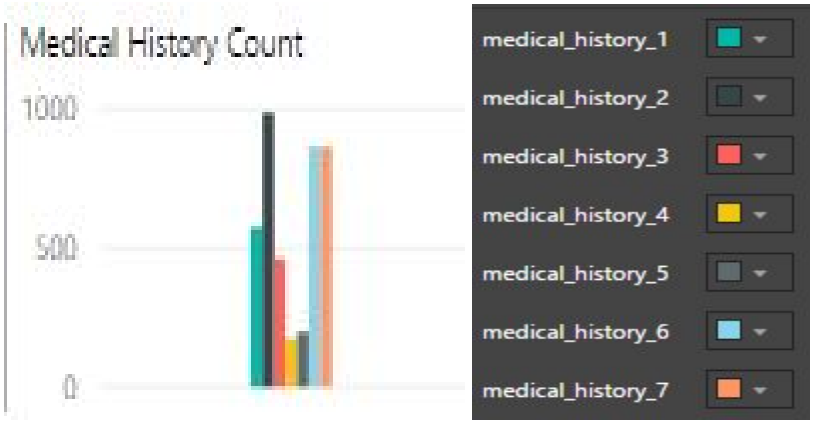
The cost of medication is more or less constant in all the given years with slight a slight upward trend. When it is analysed in depth based on a gender basis, we came to know that the amount incurred has been in constant rise for the male whereas there are some falls for the female. So in order to support this the second chart shows the count of Symptom1 based on a gender. It also shows that women category has three falls whereas men category does not have a fall at all. From this it is clear that Men is a driving factor for rising cost. A remedy for this would be to prevent Men from diseases prior to Women.

Age - a major factor for Patient's sickness

The Histogram above shows the count of all age groups of patients who got admitted so far. The maximum number of patients are from the Age group of 40-50.

Age	Count
45	117
43	107
44	106
42	100

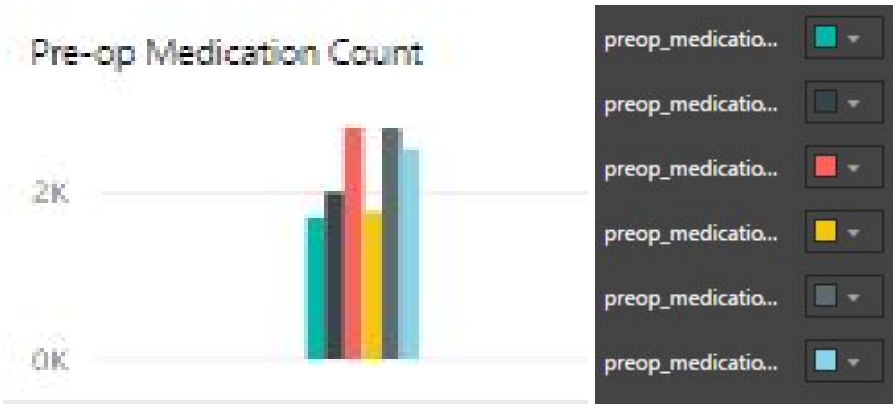
Medical History Analysis:



Medical History	Count
1	575
2	986
3	463
4	177
5	198
6	866
7	865

Pre-op Medication Analysis:

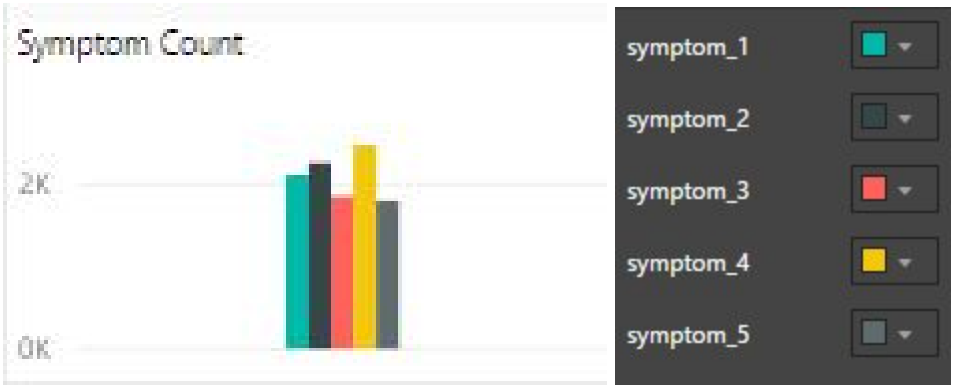
Patients are mostly having Medical History 2 followed by Medical History 6 and 7. In the future unaffected people should be screened or prevented from this particular disease.



Pre-op Medication	Count
1	1713
2	2010
3	2791
4	1779
5	2787
6	2530

Pre-op medication 3,5 and 6 are the top three medications done before any surgery. So corresponding doctors should be scaled according the requirement for undisrupted work. Consider Pre-op medication 3 as Anesthetist who gives Anesthesia before surgery. Then it is ideal to have more Anesthetist compared to other doctors.This would enable efficiency in medical workforce.

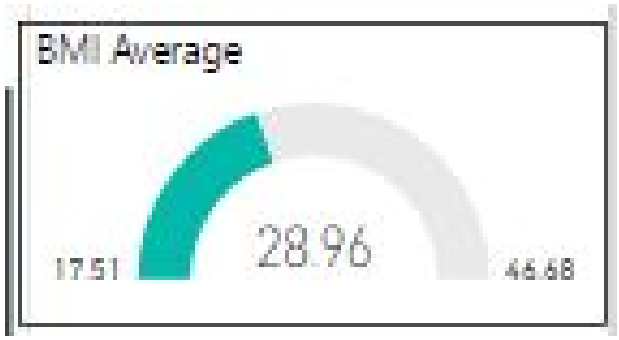
Symptom Analysis:



Symptom	Count
1	2107
2	2252
3	1852
4	2470
5	1791

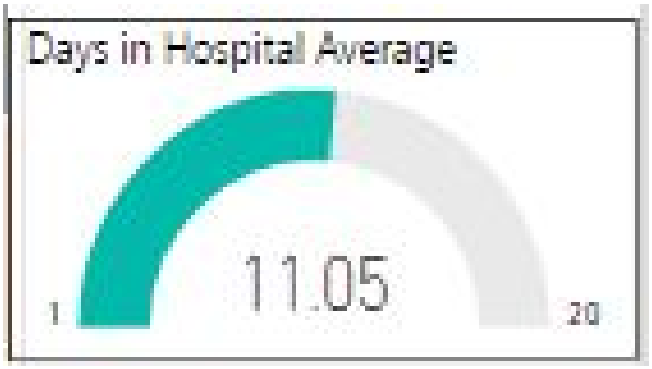
Most people are suffering from Symptom 4 and Symptom 2. These symptoms leads to certain diseases. For example let Symptom 4 be ‘Fatigueness’ and Symptom 2 be ‘sudden dizziness’. They both are symptoms for Heart attack. Patients who fall under both system should be monitored for Heart diseases.

BMI Average:



The Body Mass Index ranges from a minimum of 17.51 to a maximum of 46.68 with average being 28.96 which in turn falls into ‘Overweight’ category’. A BMI between 18.5 and 25 is a ‘Healthy Weight’ category. In order to reach the healthy BMI, patients should be asked to reduce their weight according to their height.

Days spent in Hospital Average:



A patient occupies a room in hospital with 1 day being the least and 20 days being the maximum. This information could be useful in finding the room occupation by patients and plan the allocation of rooms to patients efficiently in advance.

Market Basket Analysis:

Medical History, Symptoms and Pre op medication variables are used in MBA to find top relationships between these variables. Some of the rules are mentioned below.

LHS	RHS	Support	Confidence	Lift	Count
{medical_history_4, medical_history_5}	{preop_medication_6}	0.002941176	1	1.343874	10
{medical_history_4, medical_history_5, symptom_3}	{preop_medication_3}	0.001470588	1	1.218201	5

Association rule 1: So from above table it is clear that a patient has medical history 4 and medical history 5 should undergo pre op medication 6

Association rule 2: A patient with medical history 4, medical history 5 and symptom 3 will have to undergo pre op medication 3

Final Dashboard: