**Columbia Asia Hospital Assignment**

**Newton School – By Arti Awasthi**

**Problem Statement**

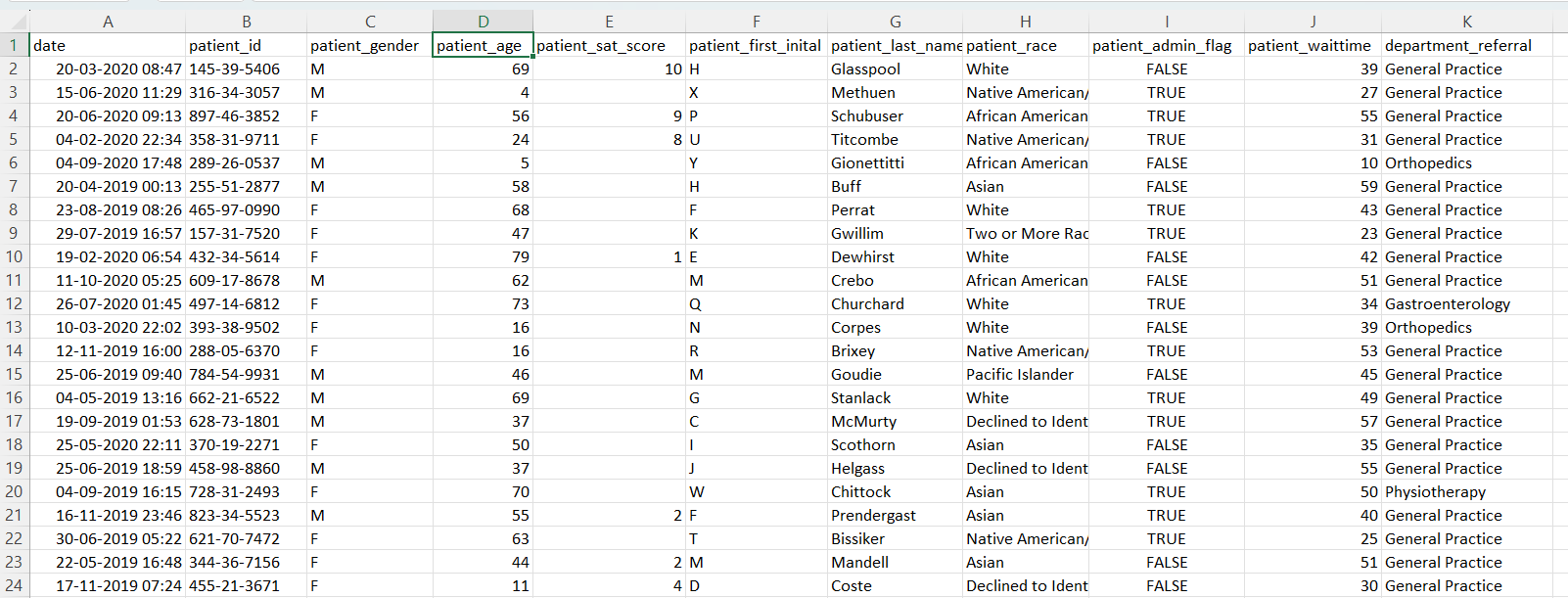
You have been hired as a consultant data analyst by Columbia Asia Hospital. The Hospital is looking for the following key insights for the following objectives:

Assess the hospital's revenue generation

Insights about suitable departments for new hires

Strategies suggestions for patient discounts

Your task is to perform data analysis and come up with a report in order to help the organisation with the mentioned objectives.



The image above displays details about Columbia Asia Hospital data, including:

* **Date**: This column contains date and time information without specifying AM or PM. The format is DD-MM-YYYY HH:MM.
* **Patient ID**: Each patient is assigned a unique identifier, which seems to be in the format 124-62-3289.
* **Patient Gender**: This column records the gender of the patient, denoted by 'M' for male and 'F' for female.
* **Patient Age**: The age of the patients is listed in years.
* **Patient Sat Score**: It seems to represent a satisfaction score given by or for the patient. However, the scores are single-digit, and it's not clear what the scale is.
* **Patient First Initial**: This column contains the first initial of the patient's first name.
* **Patient Last Name**: The surname of the patient is listed in this column.
* **Patient Race**: The racial or ethnic background of the patient is recorded here, with categories such as 'White', 'African American', 'Asian', 'Native American/Alaska Native', and 'Two or More Races'.
* **Patient Admin Flag**: This column contains boolean values ('TRUE' or 'FALSE') which might indicate whether the patient was admitted or some other administrative flag.
* **Patient Wait Time**: Appears to indicate the time the patient waited, possibly in minutes, before being seen or processed.
* **Department Referral**: This column lists the department to which the patient was referred, with entries such as 'General Practice', 'Orthopedics', 'Gastroenterology', or 'None' indicating no referral.
* **Doctor Name**: Identifies the doctor who attended each patient.
* **Appointment Fees**: The cost charged for a doctor's consultation.
* **Total Bill**: The overall amount billed to the patient, including all services and charges.

**Objective Questions**

1. In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis.

In the dataset shared, there was only one column ‘**Patient Sat Score’** where there were many entries in which the entries were missing. In this case, we just cannot just replace the missing entries with **‘NA’** as the entries in this column are in the integer format. Apart from this, if we replace the blank entries with 0, it will not be very helpful as this will also lead to the false indications in the Avg Patient Sat Score, so leaving it as it is the best course of action for the time being.

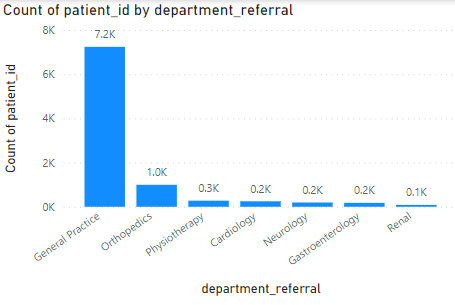
1. Assess the Average Waiting Time: Analyse the patient wait times to identify the average duration a patient spends before receiving care.

After performing the analysis on the dataset given, we have calculated that the average waiting to be 35.26 mins



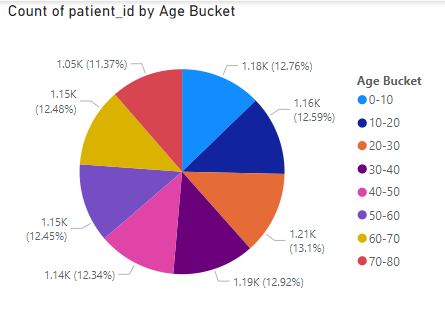
1. Visits by Department Referral: Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.

To analyze this, we’ve created a bar chart in which we placed the departments in x-axis and number of patients on y-axis in Power-BI. Through this we can see the General Practice department is getting the most number of patients (around 7.2K patients), followed by Orthopedics (1.0K), Physiotherapy (0.3K). After that rest of the departments are getting less than 300 patients respectively in each department. Graph chart is attached below.



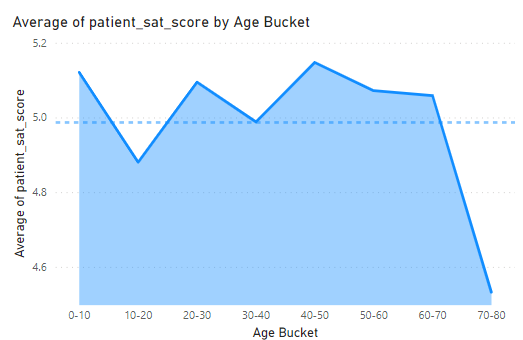
1. Patient Visits by Age Group: Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

To analyze this, we have created a new column using the transform data section in the Dataset where we divided the ages in the bucket of 10s i.e., 0-10, 10-20, 20-30, etc. in order for us to understand the distribution of patients. After that, we created a pie chart to visualize the distribution of users over these defined age buckets. In this we saw there were no specific age buckets which is having a significant high number of patients, almost all age buckets are having more or less same number patients. Pie chart for the same is attached below.

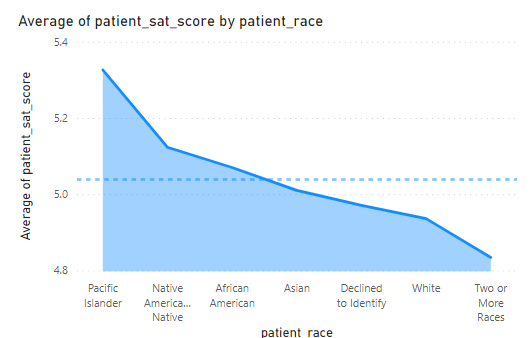


1. Average Satisfaction by Demographics: Determine the relationship between patient satisfaction scores, their age groups, and racial backgrounds to pinpoint areas for improvement in patient experience.

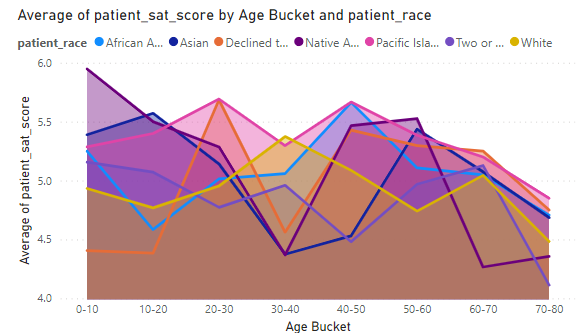
To analyze this question, we created a line chart of average patient SAT Score (patient\_sat\_score column) and the Age bucket which we created earlier in the Power BI. Apart from this, I’ve also added an average line in the graph to understand which particular section of the age bucket is falling below the average, so that the hospital can start focusing on how to increase the hospital SAT score. In this graphs, we can clearly see the two age buckets which are falling below the Average SAT score, 10-20 age bucket, 70-80 age bucket. The avg SAT score of 70-80 group is 4.53 which is mainly due to the high waittime of the patients. Apart from this the avg SAT score of 10-20 age bucket is 4.88 which is slightly below the avg score of hospital.



After this we also analyzed the data of how the different races are having the avg SAT scores. For this we created a line chart, where we placed the patient race in x-axis and avg patient\_sat\_score in the y-axis. Also similar to above chart, I’ve added the avg SAT score line to find out the races which are falling below the avg SAT score. In this chart we can see that the patients who are having two or more races are having the lowest avg satisfaction score, followed by white, declined to identify, and asians. To resolve this, the hospital needs to understand why this is happening,is there some form of racial discrimination happening, what are the different things which these particular races looks for in th medical amenities, etc..

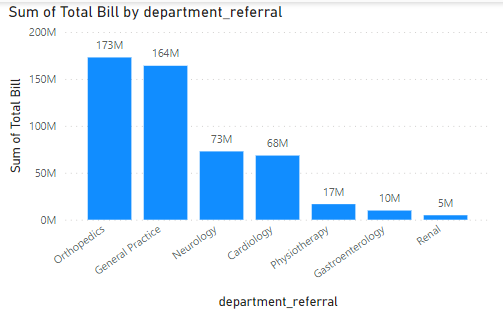


In order to fully understand the avg SAT score across all demographics (age and race), we have created a area chart (it can be line chart or bar chart as well) with where x-axis is age bucket and y-axis is avg patient sat score. Along with this, patient race is added in legend so that we can identify how different races are rating the hospital visit experience for different age group. This can help in pinpointing the areas where the hospital attention is needed for better patient SAT score.



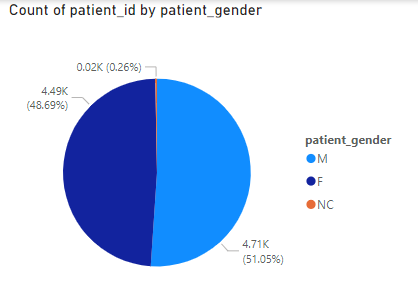
1. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.

For this, we have created the Bar chart, where x-axis is department Name and y-Axis is the sum of Total bill amount. Department Orthopedics is generating the highest revenue of around 173M, whereas the Department Renal is generating the lowest revenue of around 5M. There are 2 department who are generating more than 100M in revenue which are Orthopedics and General Practice. There are 2 department who are generating more than 20M but less than 100M which are Neurology and Cardiology. Rest of the department are generating less than 20M in revenue. Chart for the same is attached here.



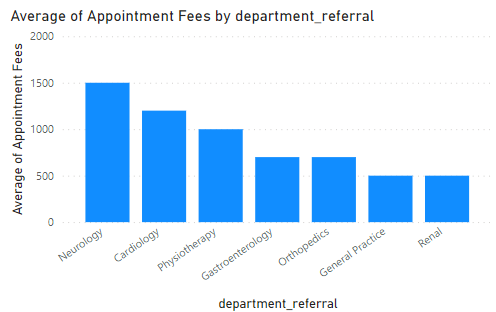
1. Is there any relation between the number of visits and the Gender of the patients?

To analyze this, we have created the Pie chart to visualize the question. In the Legend we have placed the patient gender and in value we have placed the count of patient\_id, with this we can understand how many entries are happening for different genders. Here we can see that the Male and the Females are the 2 most prominent genders coming on top, with the difference of 2.35% in the cohort size, which not much significant as the gender gap in the world is around or more the same. This tells us that the both the genders are having equal number of visits in the hospital. Graph for the same is attached herewith.



1. Which department is charging the highest appointment fees in general?

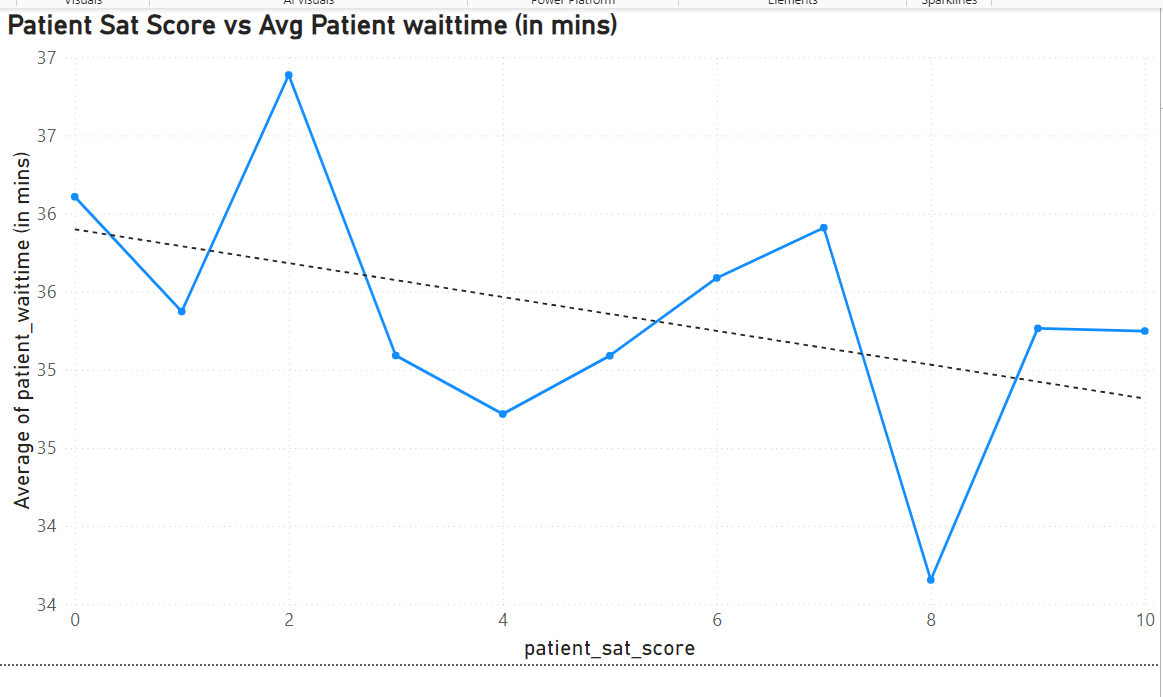
To analyze this, we have created a Bar chart, where x-axis is Department and y-axis is average of Appointment Fees. From this, we can see that the Neurology department is having the highest appointment fees as compared to all departments (1500), followed by Cardiology (1200).

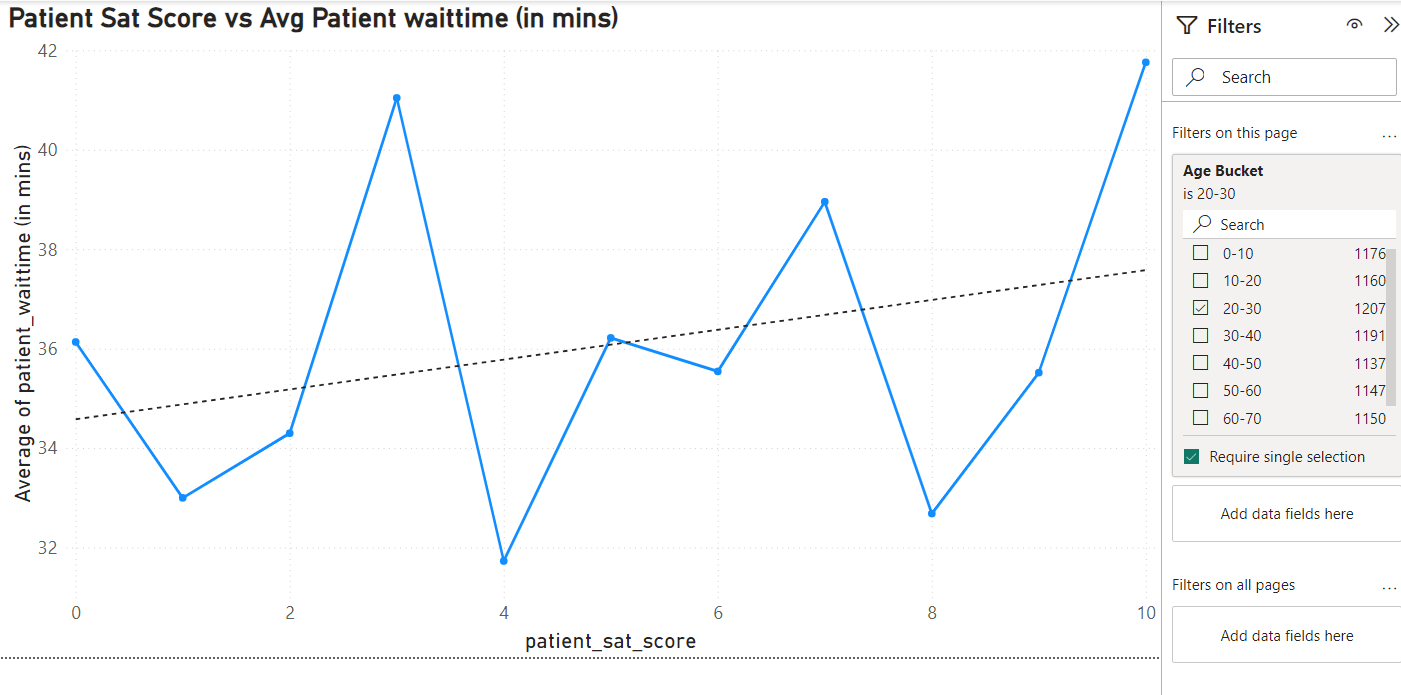


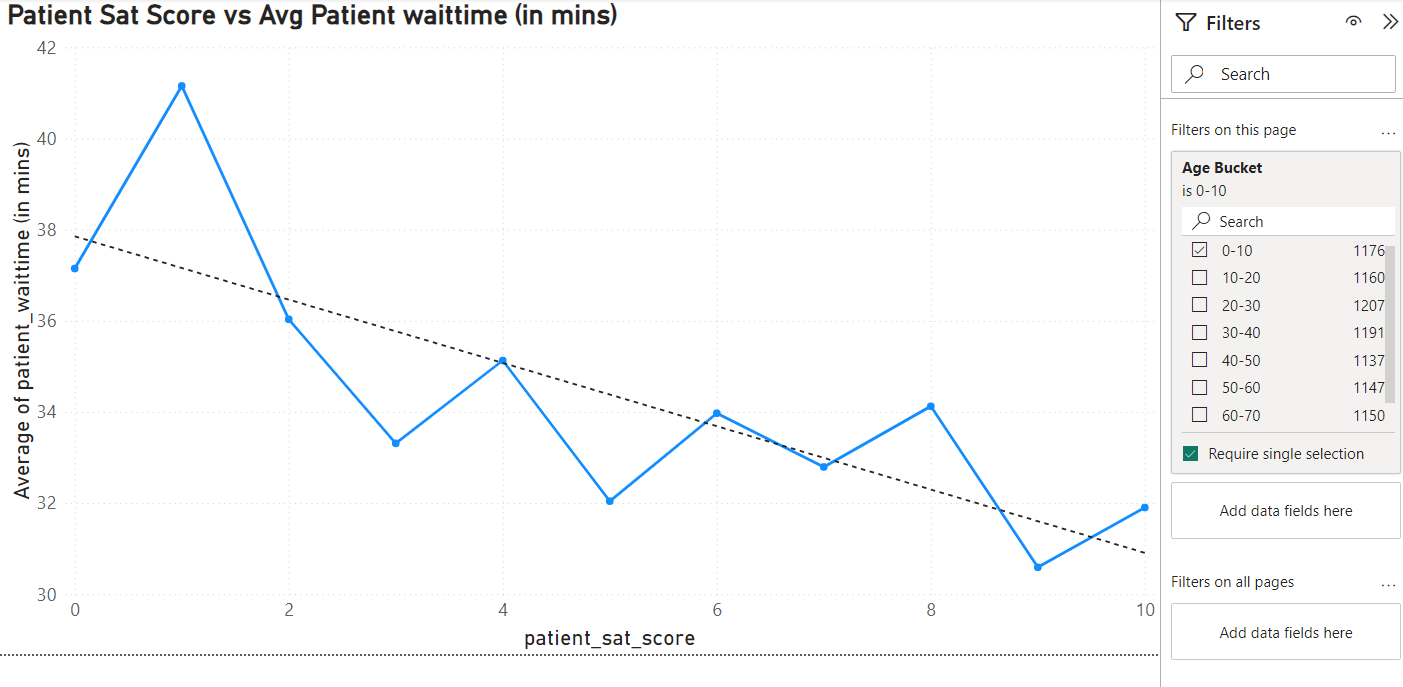
**Subjective Questions:**

1. What is the relation between patient wait time and satisfaction scores?

To analyze this, we have created a line chart where x-axis is the different SAT scores and y-axis contains the avg of patients waittime. Along with this, I’ve also plotted the trendline in the graph which tells us how the data is moving. This graph and trend line shows us that as the waittime of patients decreases the patients SAT score also decreases. This statements holds true for the overall dataset which is given to us, but not for all age groups, the behaviour of the different age group changes differently. For some age groups like 50-60 and 20-30 shows opposite behaviour as compared to the rest of the age groups. Graphs for the analysis is attached below.

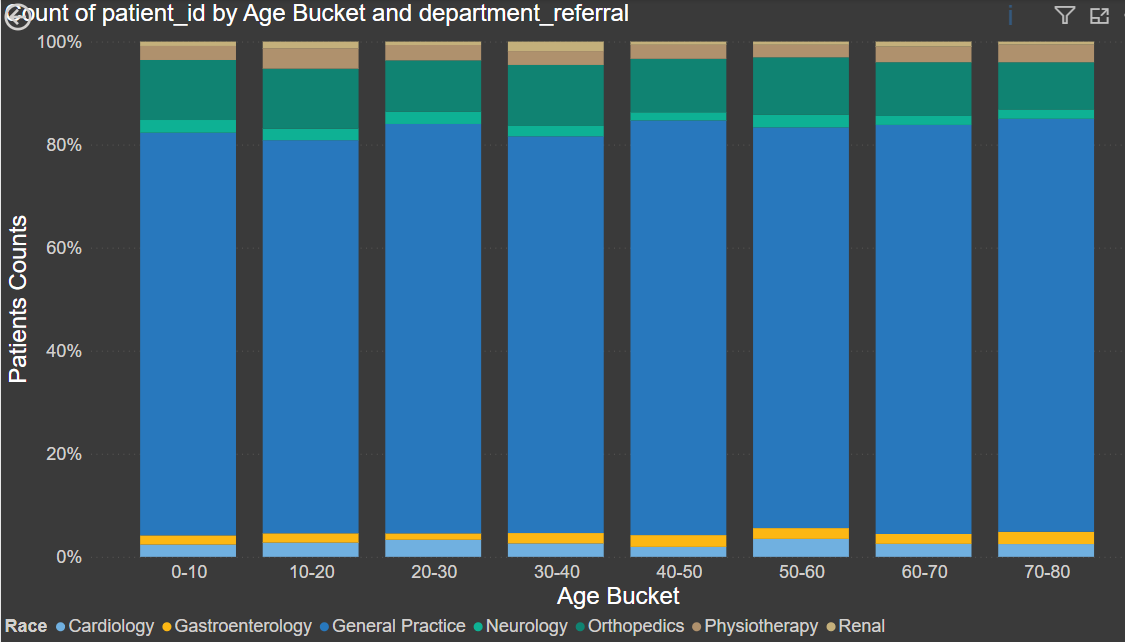


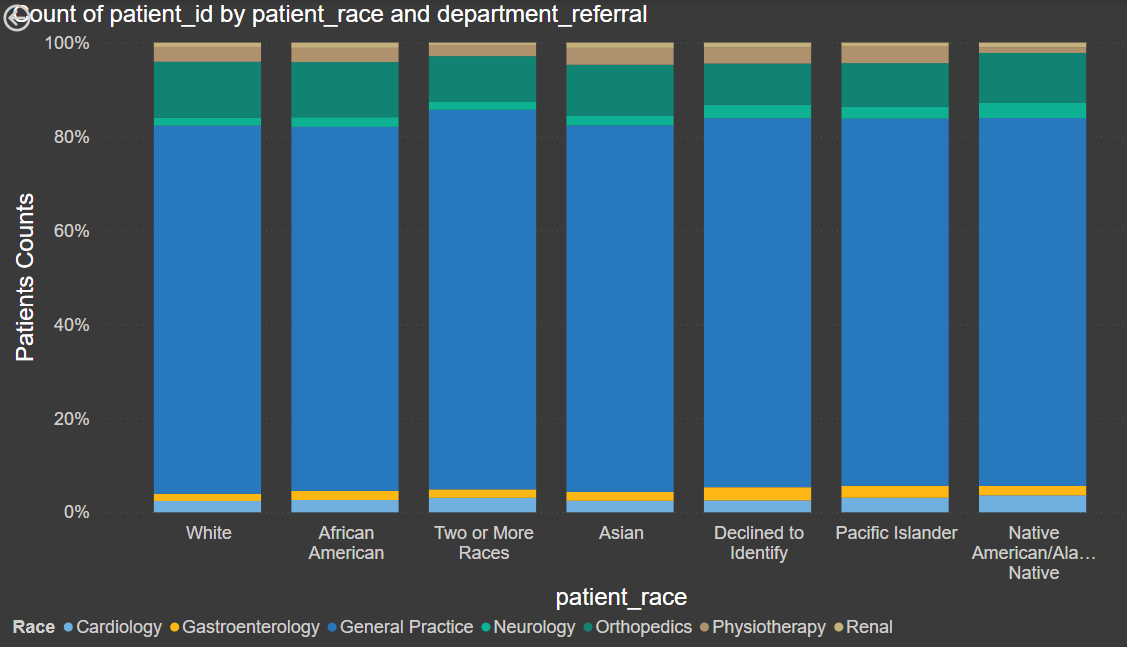




2. How do patient demographics affect the frequency of visits to different departments?

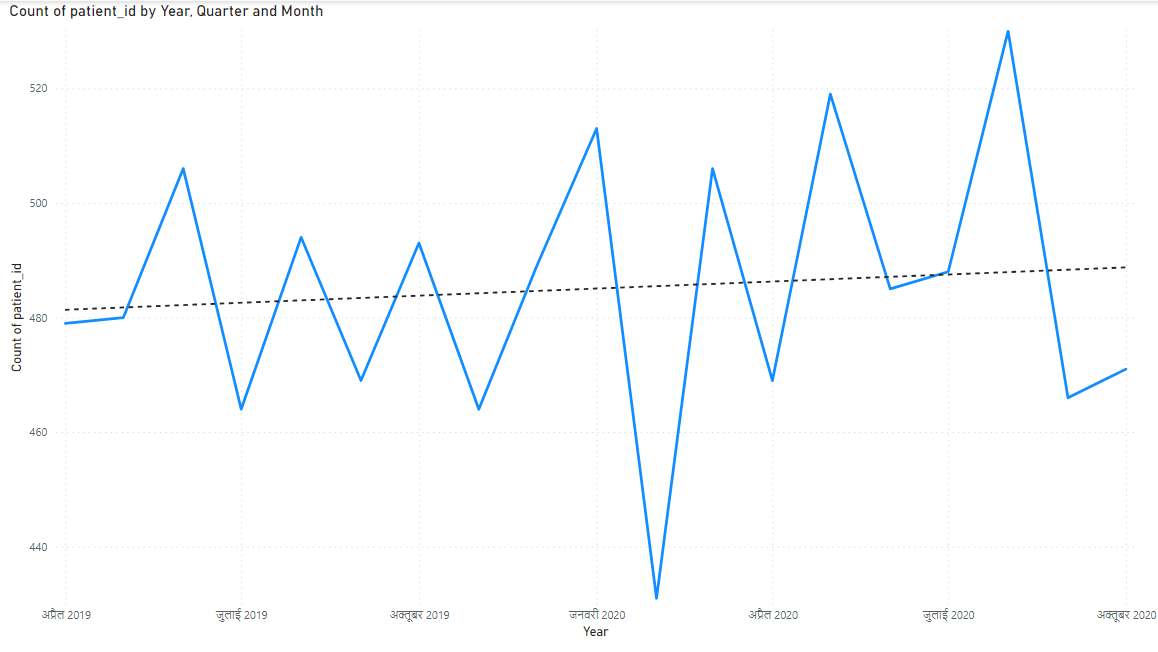
For this analysis, we need to create two stacked charts, where x-axis is age bucket in one chart and patient race in another chart, on y-axis of the both chart will contain the patient counts, in the legend, it will contains the department name in both charts. After visualizing the both charts we can see that there is no considerable difference in any particular age group and patient race.





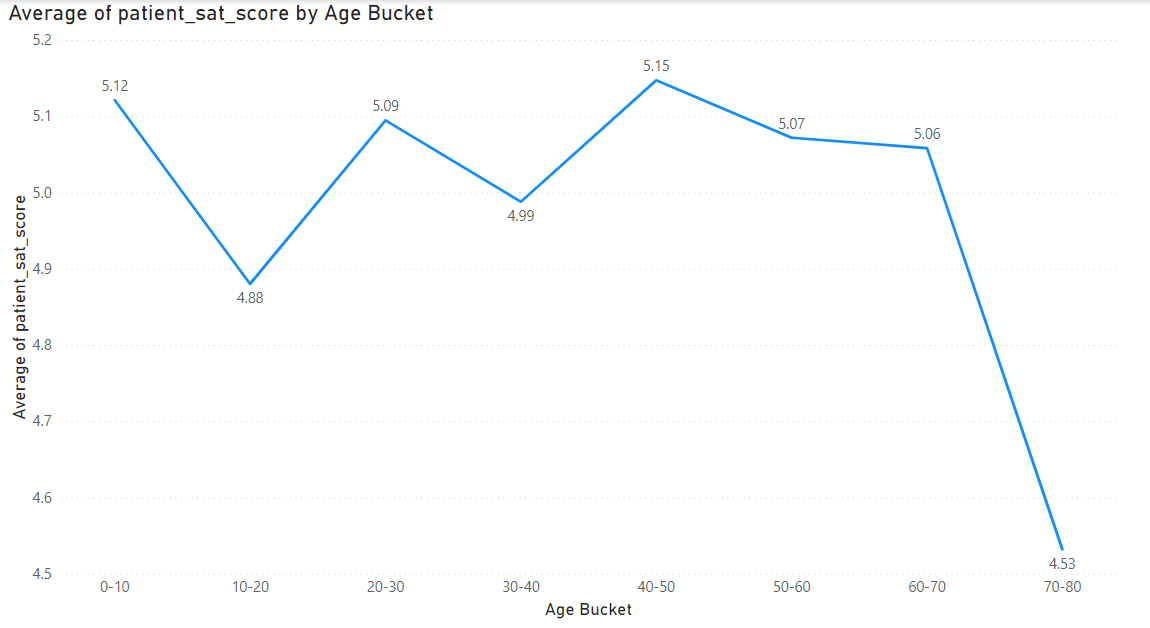
3. Is there a noticeable trend in the volume of patient visits throughout the year?

For this, we analyzed the data found out that the data is starting from April, 2019 and ending in October, 2020, due to which we cannot analyze the data in the yearly manner as data size is not uniform. Also, we cannot analyze the data based on Month as there are certain months for which the data is not there in both year. Due to which the analysis cannot be performed on the month level or quarter level. Due to this, only month-year level analysis can be performed to check if there are any combined months in which the patients are visiting. To analyze this, we have created a line chart where x-axis is date and y-axis is count of patients. Initially it will give the year level breakdown in the x-axis, but after expanding on the date hierarchy level, we reached the month-year level. In that graph we can see a zig-zag pattern which doesn’t give any clear insight. It only shows that the number of patients increases in every next month. Apart from this, we’ve also applied the trend line on this graph in order to see if the number of patients are increasing or decreasing over the period. We can see that the trendline is in upward direction which says that the number of patients are increasing month on month. Graph for the same is attached below.



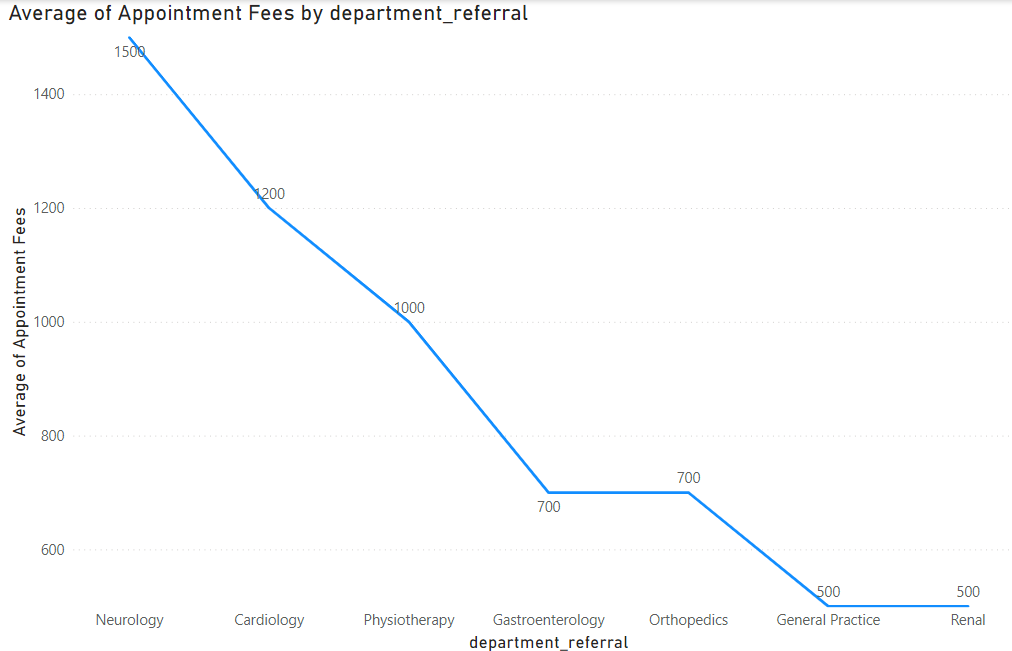
4. Which age groups report the highest and lowest satisfaction scores?

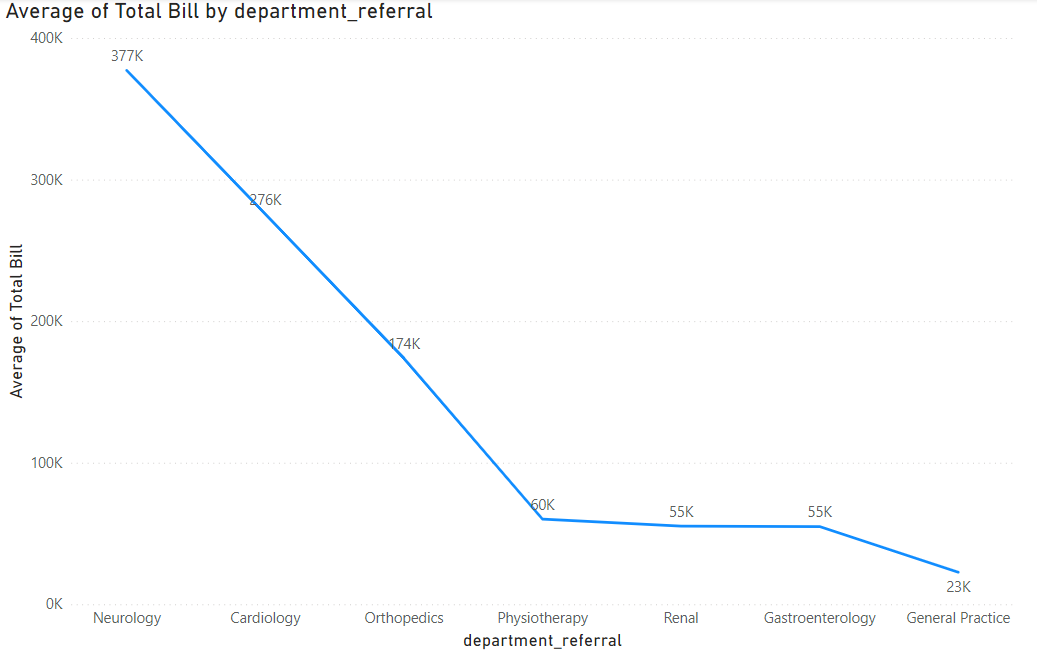
To analyze this, we’ve created a line graph where x-axis is age bucket and y-axis is the average of patient sat score. With this graph we can see that age group 40-50 is having the highest avg sat score, whereas the age group 70-80 has the lowest avg sat score.



5. The hospital management intends to offer discounts to patients.

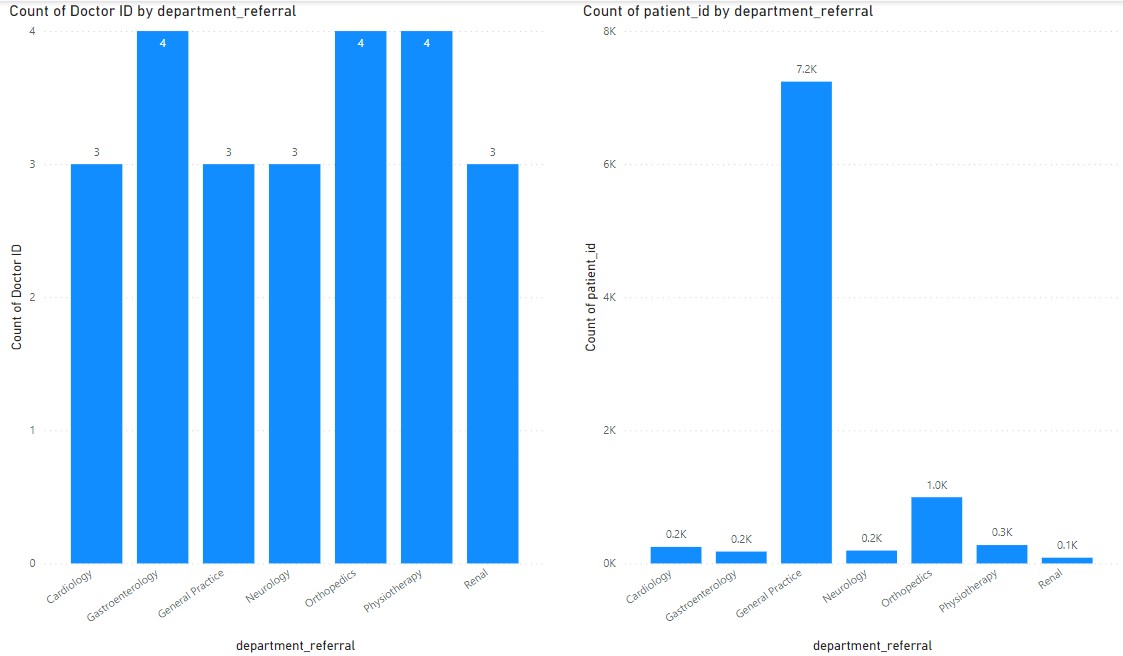
To analyze this question, we need to check how the different departments are charging the patients for the appointment as well as how big there total bill is going. To analyze this, we have created two line chart where x-axis is department and y-axis is avg of appointment fee in one chart and avg of total bill amount in another chart. With these charts, can clearly see that neurology and Cardiology are two departments where the appointment fee and total bill amount are both high. We can start giving the discount to patients from these department as per the operational cost of the department for per patients. In the economy on this doesn’t work, then we can start looking for other departments where we are making most profit per patient to give the discount.

****

****

6. The hospital has a budget to hire 2-3 new doctors.

To analyze this question, we need to analyze first the number of doctors each department are having and how many patients are coming in each department. For this, we need to create 2 bar chart, where on x-axis contains department and y-axis contains doctor unique count in one graph and patient count in another graph. With these two graphs we need to find the departments in which the patient to doctor ratio is high. In these graphs we can see that the general practice department currently only have 2 doctors, who are serving around 7.2k patients in the time period for which data is given. We need to hire more doctors in this department.

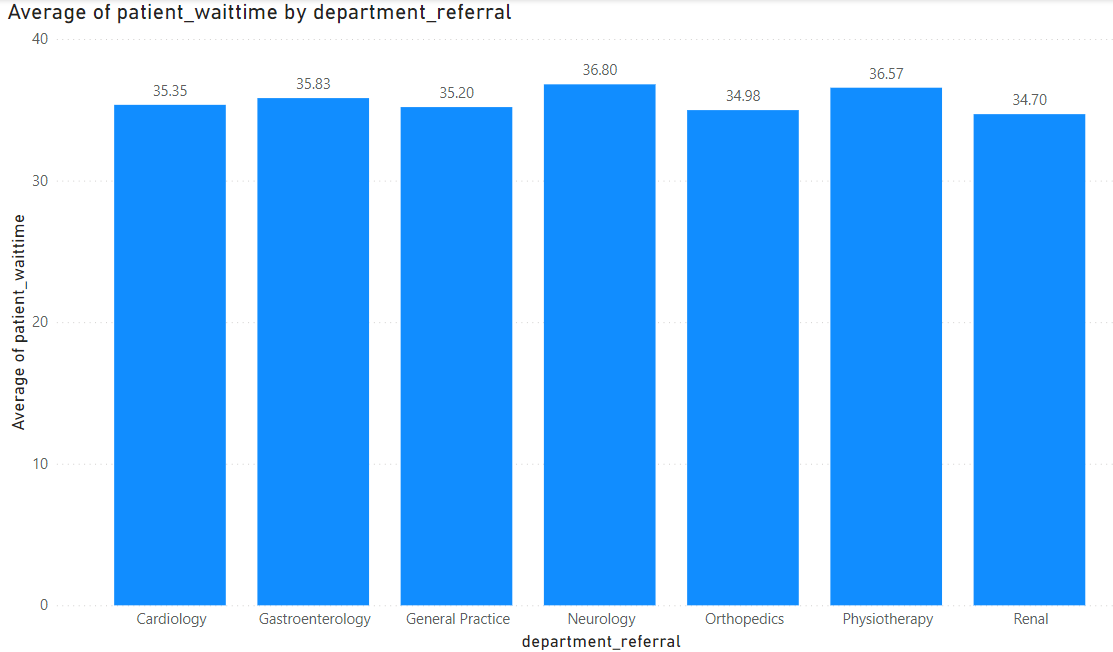


7. Is the hospital profitable?

We can only tell the revenue of the hospital, but we cannot calculate the cost of the hospital as the dataset around that is not provided, which is necessary for profitability.

8. Any Department for which the waiting time is oddly large?

To analyze this question, we need to create a bar chart where the x-axis is the department and the y-axis is the avg of patient waittime. With this graph, we can see that almost all department have near similar waittime. So there are no department in which the wait time is oddly high.

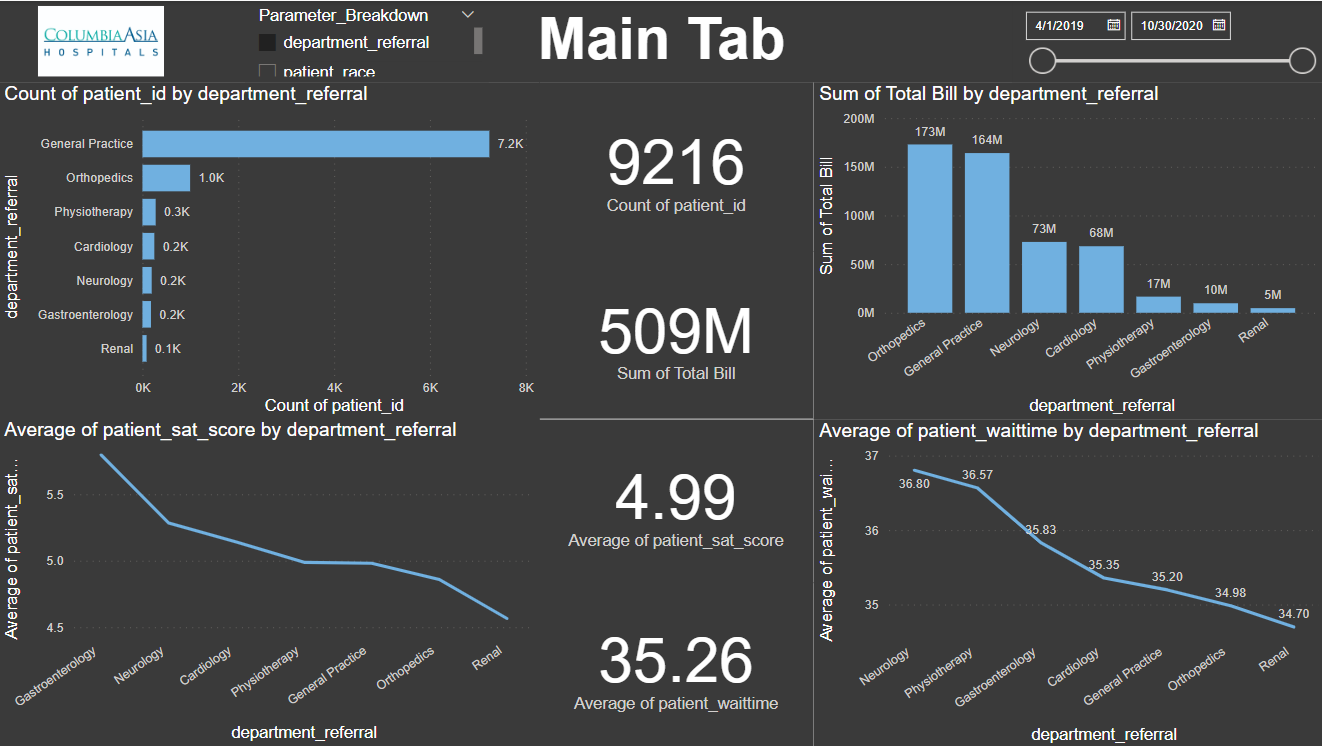


**Dashboard**

1. Main Tab

In the main tab of the dashboard, i have added the following items for the better visibility of patients and revenue and other items

1. Department level patient count - This was created via bar chart
2. Average of Satisfaction score via department name - This was done via line chart
3. Department level total revenue- This was done via bar chart
4. Avg patient waittime for each department - this was done via line chart
5. Total patient count
6. Total revenue
7. Average patient sat score
8. Average patient waittime

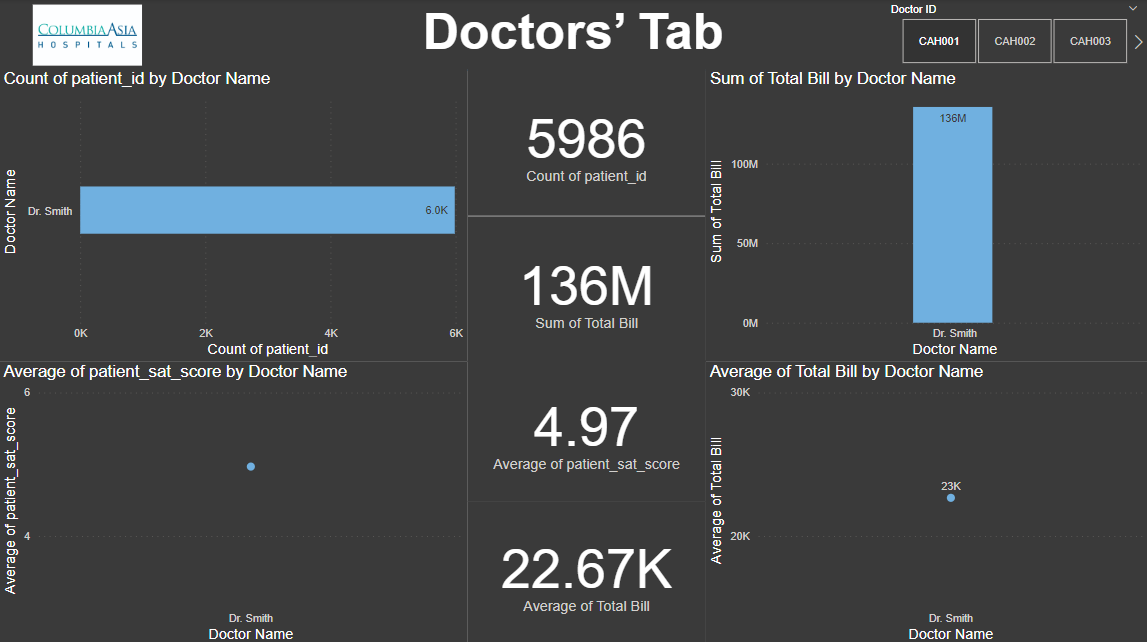


Also a Field Parameter has also been added to view this dashboard in either the department level or Patient race level or Patient Age bucket level.

1. Doctors’ Tab

In the doctor’s tab of the dashboard, i have added the following items for the better visibility for Chief Of Staff on how doctors are performing

1. Count of patient via doctor name
2. Average of patient sat score
3. Total revenue by doctor
4. Average bill amount per patient by doctor
5. Total patient count
6. Total revenue
7. Average sat score
8. Average bill amount per patient



1. Patients’ Tab

In the Patients’ Tab, i have added the following items for the better visibility for Patient’s Care Chief on the basic details of the patients:

1. Number of patient visits
2. Patients spending on various department
3. Patients number of visits
4. Total amount paid to the hospital
5. Average Patient waittime
6. Patients’ Gender
7. Patients’ Race
8. Patients’ average sat score
9. Average of Patients’ spending on each visits

