N. SAI DHANUSH (DEC-23 BATCH)

POWER BI PROJECT REPORT

OBJECTIVE QUESTIONS:

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

Missing data values and data inconsistencies are handled to ensure and extract actionable insights from raw data.

1. Analyze the patient wait times to identify the average duration a patient spends before receiving care.

This is done by utilizing average DAX Function to calculate the average waiting time provided the patient waiting time.

DAX FUNCTION: Average Wait time (in mins) = AVERAGE('Hospital ER'[patient\_waittime])

After Calculation, average waiting time obtained is 35.26 minutes.

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

In order to answer this, we created a visualization in POWER BI named “ Bar Chart”, which aids us to understand the count of patient visits to each department.

A screenshot of a computer

Description automatically generated

From the above bar chart, we can observe that, “General Practice” is the department with highest number of visits ~ 7.2k (78%)

1. Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most.

In our data, we don’t have patient’s age segregated, but grouping/clustering patient’s age let us to know which demographic most utilizes the health care services.

So, this is done by using a DAX Function named “SWITCH”.

And the syntax for that is:

Age Bracket = SWITCH(

                    TRUE(),

                    'Hospital ER'[patient\_age] < 15, "0-15",

                    'Hospital ER'[patient\_age] < 25, "15-25",

                    'Hospital ER'[patient\_age] < 40, "25-40",

                    'Hospital ER'[patient\_age] < 60 , "40-60",

                    'Hospital ER'[patient\_age] >= 60, "60+"

)

Here, I have segregated age groups into 5 groups.

A graph of blue rectangular shapes

Description automatically generated with medium confidence

From the above figure, we can easily understand the demographic visits of patients.

1. Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for?

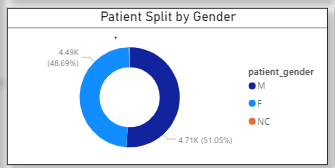
In our data, we had null values in satisfaction rating named “patient\_sat\_score”, where out of 9216 patients only 2295 patients gave their feedback as a satisfaction score.

There might be many reasons the patient hasn’t gave feedback. But considering their emergencies, time constraints and other factors, even though 75% patients haven’t rated.

I have not dropped any null values, because if we drop null values then almost 75% of data is lost, and also didn’t try to extrapolate using mean or median values of satisfaction score, because satisfaction scores do change very dynamically from individual to individual based on their beliefs and service obtained.

Thus we retained all null values in our case!!

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

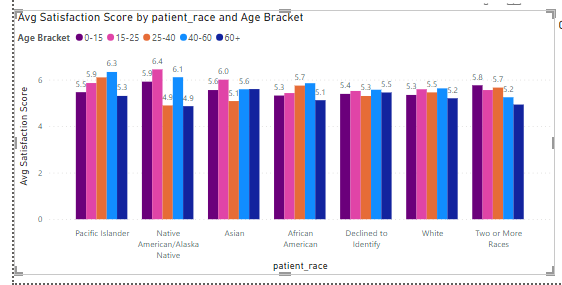


In our case, we have Male patients count as 4705 (51.05%) and female patients count as 4487 (48.7%) and 24 patients who hasn’t mentioned their gender.

So, we can infer that there is no relation between number of visits based on gender, as both male and female almost equally availed the healthcare services.

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

In order to answer this question, we have utilized clustered column chart to visualize satisfaction rates of patients from different races clustered by different age groups.



From the above chart, we can infer that highest satisfaction rating is given by “Native American/Alaska” region people, whereas least satisfied patients are also found in this region within age bracket of “25-40” and “60+”. There is a lot of variance among this region people in terms of satisfaction rates, and other regions’ satisfaction rates seems almost equal among all age groups.

Hence, Native America/Alaska region patients need to be analyzed properly or else other feedback factors need to be considered.

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

This can be easily addressed with the aid of Column/Bar chart, in our case we have plotted department on X-axis and Revenue on Y-axis on a column chart to understand the revenue generated by each department.

“Revenue” generated is not in our data, this is calculated as the sum of total bill for each patient by using a simple DAX Function known as “SUM”.

A screenshot of a graph

Description automatically generated

We can infer that “Orthopedics” department generated highest revenue of 170 Million and “Renal” department generated least revenue with an amount of 4.75 Million.

1. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.

This is achieved by using a DAX Function named “LOOKUP”, which takes 3 arguments

Department with Highest Fee = LOOKUPVALUE(Sheet1[department\_referral], Sheet1[Appointment Fees], MAX(Sheet1[Appointment Fees]))

By using above DAX Formula, we concluded that “Neurology” department charges highest fee, which is 1500

1. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.

To calculate the number of patients who have visited/Consulted Dr. Smith we use “CALCULATE” function, which internally uses a filter which we need to mention upon the condition.

No: of Patients Visited Dr.Smith = CALCULATE(

                                            COUNTROWS(Sheet1),

                                            Sheet1[Doctor Name] = "Dr. Smith"

)

1. Calculate the average age of the patients who visit the Orthopedics department. Will the approach used to calculate this metric be different if the requirement had been all departments’ average age?

To calculate the average age of patients who visited “Orthopedics” department, can be achieved by Using DAX Formula:

AvgAge\_Orthopedics =

AVERAGEX (

    FILTER (

        'Hospital ER',

        'Hospital ER'[department\_referral] = "Orthopedics"

    ),

    'Hospital ER'[patient\_age]

)

Yes, In case of calculating average age for all departments we need not to do row iteration, just an aggregation is enough to obtain, whereas in case of calculating average age of orthopedics department, we need to filter rows and then aggregate. So the syntax differs as shown below:

AvgAge\_AllDepartments =

AVERAGE ( 'Hospital ER'[patient\_age] )

1. Were there any data format issues in the data, and if there were/are how you handle them?

Yes, there were data format issue. In Hospital ER data, where the date column was not in date format. Instead, it was in text format. So, this format doesn’t help us in making a slicer for date.

Thus, the format is changed to “Date” from text with the help of column functions in the

Power query editor.

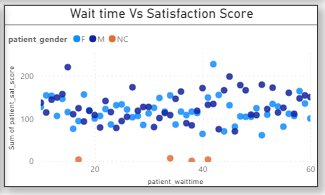
1. When we add a column in Power Query what’s the code that comes in M language in the formula bar? What do you know about M-query?

In Power Query, creating a new column results in M language code displayed in the formula bar, commonly formatted as `= Table.AddColumn(#"PreviousStepName", "NewColumnName", each [Expression])`. M-query, or Power Query Formula Language, empowers users to efficiently execute data transformation tasks within Power Query, aiding in activities such as data cleaning, reshaping, and preparation for loading into Excel or Power BI.

SUBJECTIVE QUESTIONS:

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

Since, both wait time and satisfaction scores are numerical values, scatter plot is best to explain the relationship between these two variables, if any exist



From the above figure, we can observe that there is no positive or negative correlation between these two variables. They are very random.

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

This can be achieved by plotting a clustered column chart, by plotting demographics on X-axis,

Visit count on Y-axis, and department as legend

A graph of different colored lines

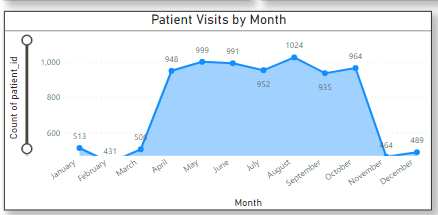
Description automatically generated with medium confidence

From this we can infer that patients from all demographic conditions visit General practice on a bigger scale.

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

To address this, we need to plot visit of patients throughout the year. So, Line chart/Area chart would better answer this. Since we are asked to find the trend throughout the year.

Let’s plot Months on X-axis and count of visits on Y-axis.



We can Observe that. During the months of January, February, November & December, the total number of visits is much less when compared to other months.

The peak visit count is observed during the months of April to October.

Highest Visit count as 1024 during the month of “August” and lowest visit count during the month of “February”.

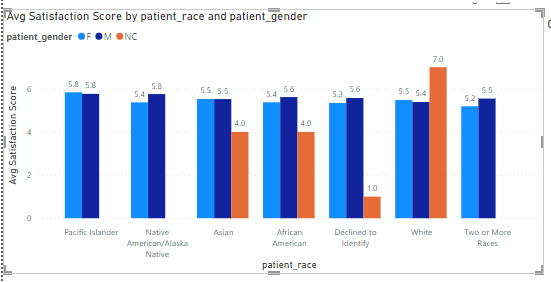
1. Which age groups report the highest and lowest satisfaction scores?

Let’s plot this on a column chart to find out which demographics contribute to highest and lowest satisfaction scores. “15-25” age group has higher satisfaction rate and “60+” group has lower satisfaction scores.

A graph with numbers and a number of bars

Description automatically generated with medium confidence

1. Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?

To better understand this claim or allegation, Average satisfaction rate is the only metric which can give more explanatory answer. Let’s plot average satisfaction rate for patient race and gender as hue/legend.

From above chart, we can infer that the claim is False, because satisfactory score for both the genders is almost same across all the races.

1. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?

To understand the importance of giving discounts to patients is a better strategy and also it will empower patients to avail the healthcare services. So visit count of Patients would be a better strategy to answer this.

A graph of a patient count

Description automatically generated

From the above figure, it is very obvious that Pacific Islander and Alaska patient visit counts are less compared to other regions.

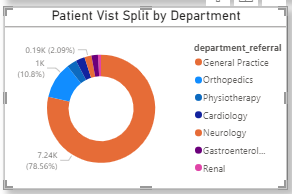
And also discount should be given to senior citizens, who are aged “60+” as their satisfaction scores are also less.

I believe these two measures will boost Health consciousness among the above mentioned two groups.

1. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.

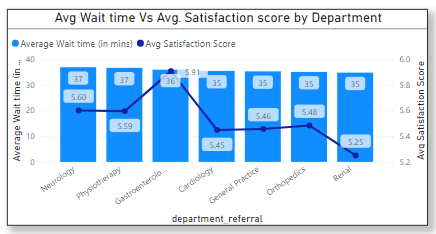
Let’s have a look at the doctor count and revenue created by each department.

This helps us better understand the areas of improvement in hiring new doctors.

From the below figure we can see that “Renal” Department has least number of doctors and also revenue created by the department is less than 1%. I would suggest to hire doctors from “Renal” department.

1. Is the hospital profitable? How will you determine the profitability?
2. Any Department for which the waiting time is oddly large?

This can be addressed easily with the help of a column chart or bar chart. In our we’ve tried to plot a Line & stacked column chart. Where department is on X-axis, Average waiting time on Y-axis & Avg satisfaction score on secondary Y-axis.



From above figure, we can observe that average waiting time for all department varies only in 2 minutes. ie: Highest average waiting time is 37 minutes and lowest average waiting time is 35 minutes. So, there is nothing odd about average waiting time across all departments.

1. Come up with strategies to provide discounts to the patients.

As we already discussed about the groups to whom discount should be made available.

Ie: Pacific Islander & Alaska region. And also patients who are aged “60+”.

By giving them a discount of 25% would be encouraging them to avail hospital services.

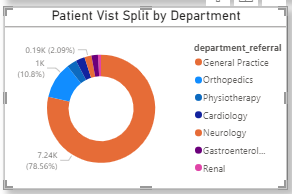
1. Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital?

Yes!! Aligning doctors from “General Practice” department would be a great strategy.

Because it is pretty obvious from the charts that people from different age groups have most visited this department, when compared to other departments.

A graph of different colored lines

Description automatically generated with medium confidence

From the above two charts, we can conclude “General Practice” Department has more count in patient visits across all age groups.

So, its better idea to avail doctors in two shifts, as patients can arrive at any time due to emergency, its always better to make the services available 24 hours a day.

Thus 7240 doctors from “General Practice” department can be split into 2 shifts on a 12-hour basis. That is 3620 doctors on the day shift and 3620 doctors in night shift.

1. What do you understand by PowerBI gateway? What are its use cases?

PowerBI Gateway serves as a secure bridge between on-premises data sources and the Power BI cloud service, enabling seamless data connectivity. It facilitates scheduled data refreshes, real-time querying, and data movement between on-premises and cloud environments. With its capabilities, organizations can leverage their existing on-premises data infrastructure while harnessing the analytics and visualization power of Power BI in the cloud. It finds applications in financial reporting, operational analytics, and business intelligence, offering a comprehensive solution for hybrid data scenarios without compromising on security or compliance.

1. How would you approach this problem, if the objective and subjective questions weren't given?

Given the task's ambiguity, I'd begin by thoroughly grasping the context of the data analysis concerning Columbia Asia Hospitals. Then, I'd methodically dissect the data, seeking key insights and patterns. Next, I'd explore diverse analytical methods and construct an adaptable action plan. Continuously open to fresh perspectives, I'd refine strategies as necessary to ensure optimal outcomes for the hospital analysis.

1. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?

From the Data, we have observed that every doctor is assigned to a department. There can be many doctors in a particular department, therefore many Ids in a department.

So, this can be named as Many to One type of cardinality Relationship.