**Columbia Asia Hospital analysis**

**Objective Questions:**

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

To ensure data cleanliness in the hospital dataset with Power BI:

* Identify Missing Values:
  + Inspected the dataset for null or inconsistent values using Power Query Editor.
* Handle Missing Data:
  + Replaced null values, such as in the patient\_sat\_score column, with the average value to maintain data completeness.
* Ensure Consistency:
  + Corrected any inconsistencies or erroneous data to prepare it for analysis.
* This cleaning process ensures the dataset is accurate and reliable for further analysis.

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

|  |  |  |
| --- | --- | --- |
| Average Waiting Time Analysis:   * Formula Used: DAX * Result: The average patient waiting time is 35.26 minutes. | | A number of numbers on a white background  Description automatically generated |
| Average\_Waiting\_time = AVERAGE('Hospital ER (1)'[patient\_waittime]) | | |

1. Significance: Highlights operational efficiency and helps identify areas to improve patient care delivery.**Visits by Department Referral:** Calculate the total number of visits to each department based on referrals to understand which departments are most frequently visited.

* Formula Used:

|  |
| --- |
| Count\_of\_patients = COUNT('Hospital ER (1)'[patient\_id]) |

* Visualization: Created a chart with:
  + X-axis: Department
  + Y-axis: Count of patients (measure).

**A graph showing patients by referrals

Description automatically generated with medium confidence**

* Result:
  + Identified the total number of visitors per department based on referrals.

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

* Approach :
  + To better understand patient demographics and how different age groups utilize healthcare services
* Formula Used:
  + Created a new column using SWITCH to categorize patients by age group:

Patient\_age\_bucket = SWITCH(TRUE(),

'Hospital ER (1)'[patient\_age] >= 0 && 'Hospital ER (1)'[patient\_age] <= 18, "0-18",

'Hospital ER (1)'[patient\_age] >= 19 && 'Hospital ER (1)'[patient\_age] <= 35, "19-35",

'Hospital ER (1)'[patient\_age] >= 36 && 'Hospital ER (1)'[patient\_age] <= 50, "36-50",

'Hospital ER (1)'[patient\_age] >= 51 && 'Hospital ER (1)'[patient\_age] <= 65, "51-65",

'Hospital ER (1)'[patient\_age] > 65, "65+")

* + Visualization:
    - Added the new Patient\_age\_bucket column to the chart's X-axis and used a measure to count the number of visitors for each age group.

A graph showing patients by age group

Description automatically generated

* + Result:
    - You successfully created age buckets and visualized the number of visits for each demographic group.

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?

Null values were present in the patient satisfaction scores, which were subsequently filled with the average satisfaction score to ensure data completeness.

* + Identified Null Values:
    - Inspected the patient\_sat\_score column in Power Query to locate rows with missing values.
  + Calculated the Average:
    - Used Transform > Statistics > Average to calculate the column's average (4.9978298611111107)
  + Replaced Nulls with the Average:
    - Rounded the average to 5 and replaced all null values using Transform > Replace Values.

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

Analysis of Visits by Gender

* + A breakdown of patient visits reveals the following gender distribution:
    - Male: 51.05%
    - Female: 48.69%
    - Not Categorized (NC): 0.26%
  + Visualization:
    - This indicates that the number of visits is almost evenly distributed between males and females, with a slightly higher proportion of male patients. The NC category is negligible at 0.26%.

A yellow circle with numbers and a number of percentages

Description automatically generated

* + Result:
    - The donut chart visually represents this distribution, highlighting the near parity in visits across genders.

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.

Analysis of Average Satisfaction by Demographics

* + Age Group Analysis
    - Patient satisfaction scores across different age groups range between 4.92 and 5.05.
    - The 36-50 age group reported the highest satisfaction, with an average score of 5.05.
    - The 65 and above age group showed the lowest satisfaction, with an average score of 4.92.
  + Visualization:

A graph showing the age group satisfaction rate

Description automatically generated

* + Insight:
    - This suggests that younger and middle-aged patients are slightly more satisfied compared to older patients, indicating potential areas for improvement in services for senior patients.
  + Race Analysis
    - Patient satisfaction scores by racial background range from 4.96 to 5.09.
    - The highest average satisfaction score of 5.09 was reported by Pacific Islanders, while other racial groups fell slightly below this mark.
    - The lowest satisfaction score of 4.96 was observed among a specific racial group, highlighting an opportunity to focus on enhancing their experience.
  + Visualization

A graph showing the average satisfaction rate by patient

Description automatically generated

* + Insight:
    - These findings underline the importance of tailoring healthcare services to meet the unique needs of older patients and certain racial groups to improve overall satisfaction.

1. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each.
   * Department-wise Revenue Analysis
     + To evaluate the revenue generated by each department, I created a measure: Total\_Revenue = SUM(Sheet1[Total Bill]), which calculates the total billing amount across all departments.
     + Using this measure, I developed a column chart by plotting Department on the axis and Total\_Revenue on the values.
   * Visualization

A graph showing the amount of revenue

Description automatically generated

* + Insights
    - Orthopedics emerged as the highest revenue-generating department, indicating its substantial contribution to the hospital’s financial performance.
    - Renal recorded the lowest revenue, suggesting a need for further investigation into potential factors such as service offerings, patient inflow, or operational challenges.
  + Result
    - This department-wise revenue breakdown provides critical insights for decision-making, enabling management to optimize resources, improve low-performing areas, and capitalize on high-performing departments.

1. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question.
   * Highest Appointment Fee by Department
     + To identify which department is charging the highest appointment fees in general.
     + I used the LOOKUPVALUE DAX function to find the department corresponding to the maximum appointment fee. The measure created is:

|  |  |
| --- | --- |
| =LOOKUPVALUE(  Sheet1[department\_referral],  Sheet1[Appointment Fees],  max(Sheet1[Appointment Fees])  ) | A close up of a sign  Description automatically generated |

* + - Using this measure, I visualized the data by displaying the department charging the highest appointment fees in a card.
  + Insights
    - The analysis revealed that the Neurology department charges the highest appointment fee among all departments.
    - This insight highlights Neurology’s premium positioning, possibly due to specialized expertise, advanced equipment, or demand for its services.
  + Result
    - This analysis helps management better understand pricing strategies across departments and evaluate their alignment with the hospital's revenue goals and market positioning.

1. Create a tabular visualization in the Report view which consists of Month-wise total visits in the hospital. Add a third column in the table that consists of the previous month’s total visits for each month’s row. Also, include a column that states whether the visits in a month are greater than that of the previous month's visits.

Month-wise Total Visits with Previous Month Comparison Table

* + To create a tabular visualization displaying Month-wise Total Visits, Previous Month's Visits, and a comparison column (Visit Increased?)

a. Create a Calendar Table

* + You created a Calendar Table using the CALENDAR function to cover the date range in the Hospital ER Table:

|  |  |
| --- | --- |
| Calendar = CALENDAR(MIN('Hospital ER (1)'[date]), MAX('Hospital ER (1)'[date])) | Purpose: Ensures the date table spans the entire range of dates in the hospital data. |

* + Mark as Date Table: Mark this table as the official Date Table under Modeling > Mark as Date Table.

b. Define Year, Month, and Month Name Columns

* + To organize data, you added calculated columns in the Calendar Table:
  + Year Column:
    - Year = YEAR('Calendar'[Date])
  + Month Column:
    - Month = MONTH('Calendar'[Date])
  + Month Name Column:
    - MonthName = FORMAT('Calendar'[Date], "MMMM")
  + These columns help structure and sort data chronologically.

c. Establish Relationships

* + Create a One-to-Many relationship between the Hospital ER Table (Date Column) and the Calendar Table (Date Column).
  + Set the cross-filter direction to Single.

d. Create Measures

* + You defined the following measures:

|  |  |
| --- | --- |
| Current Month Visits: | Month\_Wise\_Total\_Visits = COUNT('Hospital ER (1)'[patient\_id]) |
| Previous Month Visits: | Pre\_Month\_Total\_visits = CALCULATE(  COUNT('Hospital ER (1)'[patient\_id]),  PREVIOUSMONTH('Calendar'[Date])) |
| Visit\_Increased: | Visit\_Increased = IF(ISBLANK([Month\_Wise\_Total\_Visits]) || ISBLANK([Pre\_Month\_Total\_visits]), BLANK(),  IF([Month\_Wise\_Total\_Visits] > [Pre\_Month\_Total\_visits],  "Yes",  "No")) |

e. Tabular Visualization

A table with numbers and a number of months

Description automatically generated

* + Result

Add the following fields to your Table Visual:

* + - Year (from the Calendar Table)
    - Month Name (from the Calendar Table)
    - Month\_Wise\_Total\_Visits (measure)
    - Pre\_Month\_Total\_visits (measure)
    - Visit\_Increased (measure)

1. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith.
   * To calculate the total number of patients who have visited Dr. Smith, you can use the following DAX formula:

|  |  |
| --- | --- |
| Dr.Smith\_Patients = CALCULATE(  COUNT('Hospital ER (1)'[patient\_id]),  'Sheet1'[Doctor Name] = "Dr. Smith" | A close-up of numbers  Description automatically generated |

* + Explanation:
    - CALCULATE changes the context of the calculation, allowing you to apply filters.
    - COUNT('Hospital ER (1)'[patient\_id]) counts the number of patients.
    - 'Sheet1'[Doctor Name] = "Dr. Smith" filters the data to only include records where the Doctor Name is "Dr. Smith".
  + Results:
    - Using the DAX formula, we calculated that 5,986 patients have visited Dr. Smith.
    - This was achieved by filtering the Sheet1 table for entries where the Doctor Name is "Dr. Smith" and counting the number of unique patient\_id entries.

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?
   * Approach
     + Orthopedics: Create a measure with a filter to calculate the average age for Orthopedics and display it in a Card Visual.
     + All Departments: Create a general measure without filters, add it to a Table Visual with the department column to display averages for all departments.
   * Explanation
     + To calculate the average age for Orthopedics, a measure is created using the following formula
     + Average\_age\_Orthopedics =

CALCULATE(AVERAGE('Hospital ER (1)'[patient\_age]),

'Hospital ER (1)'[department\_referral] = "Orthopedics")

* + - The AVERAGE function then computes the average age, and the result is displayed in a Card Visual, as it represents a single value.
    - For all departments, a general measure is created without any filter:
    - Avg\_age\_of\_patient =

AVERAGE('Hospital ER (1)'[patient\_age])

* + - This measure calculates the overall average age for patients. Adding this measure to a Table Visual, along with the department\_referral column, groups the data by department and shows the average age for each.
    - Finally, the measures are formatted as whole numbers using the data format options in Power BI to ensure the results align with the realistic representation of age.
  + Visualization

A number on a white background

Description automatically generatedA table with text overlay

Description automatically generated

* + Result
    - A Card Visual shows the average age for Orthopedics.
    - A Table Visual displays the average age for each department.
    - These distinct approaches ensure accurate and meaningful insights tailored to each requirement.

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

Steps Taken to Resolve Data Format Issues:

* + Combining Columns for Patient Name:
    - The patient\_first\_initial and patient\_last\_name columns were stored separately, making it inconvenient for analysis.
    - Using Power Query, I merged these columns into a single patient\_name column, ensuring proper formatting for better usability in queries and reports.
  + Extracting Dates and Creating a Calendar Table:
    - The date column included both date and time, but only the date component was needed for analysis.
    - I created a custom column to extract the date and generated a calendar table to enable advanced date-based filtering and trend analysis.
  + Handling Blank Spaces in Scores:
    - The patient\_sat\_scores column had blank spaces, which could disrupt calculations and analysis.
    - Initially, the data type of patient\_sat\_scores was in text format, so I converted it to a number.
    - After conversion, I replaced the blank spaces (null values) with the column's mean value, ensuring data consistency and avoiding skewed results during computations.
  + Result:
    - The dataset was transformed into a clean and consistent format, making it ready for reliable analysis and accurate insights.

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?
   * When you add a column in Power Query, the M language automatically generates a formula in the formula bar. For example, if you create a column to concatenate the first and last names of patients, the following code is generated:

= Table.AddColumn(#"Removed Duplicates", "patient\_name", each [patient\_first\_inital] & " " & [patient\_last\_name])

* + Explanation of the Formula
    - Table.AddColumn: This function adds a new column to the existing table.
    - #"Removed Duplicates": Refers to the previous step in the query where duplicates were removed.
    - "patient\_name": Specifies the name of the new column being created.
    - each [patient\_first\_inital] & " " & [patient\_last\_name]: Combines the values in the patient\_first\_inital and patient\_last\_name columns with a space in between, row by row.
  + What is M-Query?
    - M-Query is a formula language used in Power Query to manipulate, transform, and shape data. It is case-sensitive and functions in a step-by-step manner where each step is recorded as code. Key characteristics of M-Query:
    - Designed for data transformation.
    - Operates in a functional programming style.
    - Is highly readable and editable through the Advanced Editor.
    - By leveraging M-Query, you can customize and automate complex data transformation tasks efficiently.

1. Identify the top 5 doctors who generated the most revenue but had the fewest patients. (SQL)
   * The top 5 doctors generating the highest revenue despite having the fewest patients are identified by calculating their total revenue and unique patient count. This highlights doctors who focus on high-value services, making them key contributors to revenue efficiency in the organization.

A computer screen shot of a computer screen

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Find the department where the average waiting time has decreased over three consecutive months. (SQL)
   * The departments with consistently decreasing average waiting times over three consecutive months were identified. This indicates improved efficiency and better patient flow management, contributing to enhanced service quality.

A computer screen with text

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Determine the ratio of male to female patients for each doctor and rank the doctors based on this ratio. (SQL)
   * The ratio of male to female patients for each doctor is calculated and used to rank doctors in descending order based on this ratio. This analysis provides insights into patient demographics for each doctor, helping to identify trends or biases in patient distribution.

A screenshot of a computer program

Description automatically generatedA table of numbers and names

Description automatically generated with medium confidence

1. Calculate the average satisfaction score of patients for each doctor based on their visits. (SQL)
   * The average patient satisfaction score for each doctor is calculated, considering visits where missing scores are replaced with a default value of 5. This provides an overall measure of how patients rate their experience with each doctor, ranked by the highest satisfaction scores.

A computer screen with text

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Find doctors who have treated patients from different races and calculate the diversity of their patient base. (SQL)
   * Doctors who have treated patients from multiple races are identified, and the diversity of their patient base is measured by counting the distinct races they have treated. This helps highlight doctors with a more diverse patient demographic

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Calculate the ratio of total bills generated by male patients to female patients for each department. (SQL)
   * The ratio of total bills generated by male patients to female patients is calculated for each department. This provides insights into the financial contribution from each gender, helping to identify trends in patient spending across different departments.

A computer screen with white text

Description automatically generatedA screenshot of a computer

Description automatically generated

1. Update the patient satisfaction score for all patients who visited the "General Practice" department and had a waiting time of more than 30 minutes. Increase their satisfaction score by 2 points, but ensure that the satisfaction score does not exceed 10. (SQL)
   * The satisfaction scores for patients in the "General Practice" department who had a waiting time greater than 30 minutes are updated by increasing their score by 2 points. If the new score exceeds 10, it is capped at 10. This ensures that patient satisfaction scores are adjusted fairly while maintaining the maximum allowable score.

A black screen with white text

Description automatically generated

**Subjective Questions**

1. What is the relation between patient wait time and satisfaction scores?
   * Approach
     + A quick measure was created in Power BI by assigning department\_referral as the category, patient\_sat\_score as Measure1, and patient\_waittime as Measure2.
     + The DAX formula was automatically generated to calculate the correlation between these two measures.
     + To visualize the correlation value, a Gauge Chart was used with a range set from 0 to 2, where the final correlation value observed was 1.
   * Explanation
     + Correlation analysis in this context helps determine the strength and direction of the relationship between patient wait times and satisfaction scores.
     + A correlation value of 1 represents a moderate positive relationship, meaning that as patient wait times increase, satisfaction scores tend to decrease, but not in a strong or direct manner.
     + This suggests that while wait times influence patient satisfaction, the relationship is not perfectly linear and may be influenced by other factors such as staff behavior, quality of care, or facility conditions.
     + The Gauge Chart effectively illustrates this result, making it easier to interpret the correlation score and its significance within the range.
   * Visualization

A yellow and white scale

Description automatically generated

* + Insight
    - The analysis revealed a moderate positive correlation of 1, implying that patient wait times do have an impact on satisfaction scores, though not overwhelmingly.
    - Departments with shorter wait times or better management of patient expectations might achieve higher satisfaction, but other factors such as service quality and communication likely play a crucial role in shaping patient perceptions.
  + DAX Formula

A screenshot of a computer code

Description automatically generated

* + Result
    - This analysis highlights the need to balance wait times and other quality-of-service factors to improve overall patient satisfaction.

1. How do patient demographics affect the frequency of visits to different departments?
   * Approach
     + Prepare the data with columns for Department, Patient Count, and demographics (Age Group, Race, Gender).
     + Create a Stacked Column Chart with Department on the x-axis and Patient Count on the y-axis.
     + Use Field Parameters to switch between demographic variables (Age Group, Race, Gender) to dynamically filter the data.
   * Explanation
     + The goal of this analysis is to explore how patient demographics influence the frequency of visits to various departments. By using a Stacked Column Chart, we can see how different demographic groups contribute to the overall patient count per department.
     + Field Parameters allow us to toggle between demographic categories (such as Age Group, Race, and Gender) to analyze the impact of each on department visit frequency.
     + This helps in understanding whether certain departments are preferred by specific demographic groups and whether there are patterns or imbalances in departmental utilization based on patient characteristics.
   * Visualizations

A graph of patients with different colors

Description automatically generated with medium confidenceA close-up of a list of words

Description automatically generated

* + Insight
    - Demographics do not affect the number of visitors to other departments because the number of visits depends entirely on the patient’s health and the location of the hospital.The demographics of different departments are unaffected directly.
  + Result
    - The stacked column chart reveals that demographics do not significantly affect the overall visit counts across departments.

1. Is there a noticeable trend in the volume of patient visits throughout the year?
   * Approach
     + Use a Line Chart to visualize the trend in the volume of patient visits over time.
     + Add Year and Month Name (in hierarchy) on the x-axis to allow drill-up and drill-down functionality for analyzing trends at different time granularities.
     + Use Count\_of\_Patients measure on the y-axis to track the number of patients per month.
   * Explanation
     + This analysis focuses on identifying trends in patient visit volumes over the course of the year. By using a Line Chart with a Year and Month Name hierarchy, we can observe the changes in patient volume both annually and monthly.
     + The Count\_of\_Patients measure provides insights into how the patient visit frequency evolves across time. The drill-up and drill-down features of the hierarchy allow us to examine trends at both a high-level (yearly) and more granular level (monthly).
     + This visualization helps in identifying seasonal patterns or fluctuations in patient visits.
   * Visualization

A graph showing the number of patients

Description automatically generatedA graph showing the number of patients

Description automatically generatedA graph showing the number of patients

Description automatically generated

* + Insight
    - In 2019, the total count of patient visits was 4,338, whereas in 2020, the count increased to 4,878. This indicates a 7.5% increase in patient visits between 2019 and 2020.
    - It suggests a possible trend or changes in healthcare demand, which could be driven by various factors such as increased healthcare awareness, new medical services, or external factors like the pandemic.
  + Result
    - The analysis of patient visit trends over the years reveals a noticeable upward trend in 2020 compared to 2019.

1. Which age groups report the highest and lowest satisfaction scores?
   * Age Group Analysis
     + Patient satisfaction scores across different age groups range between 4.92 and 5.05.
     + The 36-50 age group reported the highest satisfaction, with an average score of 5.05.
     + The 65 and above age group showed the lowest satisfaction, with an average score of 4.92.
   * Visualization:

A graph showing the age group satisfaction rate

Description automatically generated

* + Insight:
    - This suggests that younger and middle-aged patients are slightly more satisfied compared to older patients, indicating potential areas for improvement in services for senior patients.

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?
   * Approach
     + Use a Clustered Column Chart with a Field Parameter in the X-axis, which allows switching between race and gender.
     + For the Y-axis, plot the average waiting time and average satisfaction score for each group (race and gender).
     + This chart helps in comparing how different race and gender groups experience their visits in terms of waiting time and satisfaction scores.
   * Explanation
     + To investigate whether there is any evidence of racial or gender-based discrimination in the hospital, a Clustered Column Chart was used.
     + The Field Parameter in the X-axis allows for a dynamic comparison between race and gender, while the Y-axis represents the average waiting time and average satisfaction score.
     + By comparing these metrics, we can analyze whether any demographic group experiences significantly longer wait times or lower satisfaction, which could indicate discriminatory practices.
   * Visualization

A close-up of a text

Description automatically generatedA graph of a number of people

Description automatically generated

* + Insight
    - By examining the average waiting times and patient satisfaction scores displayed in the chart, we conclude that there is no evidence of racial or gender discrimination occurring in the hospital.
    - The data shows consistent and equitable treatment across different race and gender groups, with no significant disparities in either waiting times or satisfaction.
  + Result
    - The analysis reveals that waiting times and satisfaction scores are similar across race and gender groups, suggesting that the hospital does not exhibit any discriminatory practices.

1. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why?
   * Approach:
     + To determine which patients are eligible for discounts, I used a DAX formula based on two conditions:
       - A Total Bill of 10,000 or more.
       - A Satisfaction Score of 6 or higher.
     + Using these conditions, I created a DiscountEligibility column that classifies patients as either "Eligible" or "Not Eligible." I then visualized the results in a donut chart showing the count of eligible and non-eligible patients.
   * Explanation:
     + The hospital management intended to offer discounts to high-spending and satisfied patients. I used a DAX formula to categorize patients based on their total bill and satisfaction score, creating a field to reflect whether a patient qualifies for the discount.
     + The formula used is:

|  |
| --- |
| DiscountEligibility = IF(SUM(Doctor\_Patients\_data[Total Bill]) >= 10000 && 'Hospital ER (1)'[patient\_sat\_score] >= 6, "Eligible", "Not Eligible") |

* + Visualization

A yellow circle with numbers and a number of percentages

Description automatically generated

* + Insight:
    - Out of the total 9,216 patients, 12.46% were eligible for the discount, and 87.54% were not.
  + Result:
    - The donut chart shows that a small proportion (12.46%) of patients are eligible for the discount, while the majority (87.54%) do not meet the eligibility criteria.

1. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire.
   * Approach:
     + I analyzed the hospital's departments using key metrics: Count of Patients, Sum of Appointment Fee, and Sum of Total Bill to recommend which departments should hire 2-3 new doctors. This was done using a table visualization.
   * Explanation:
     + Departments with high patient volumes, appointment fees, and total bills were prioritized.
     + Orthopedics had the highest patient count (995), appointment fees (696,500), and total bill (172,939,773).
     + General Practice had a large patient count (7,240) and significant revenue.
     + Neurology had fewer patients (193) but substantial fees and bills.
   * Visualization

A screenshot of a medical report

Description automatically generated

* + Insight:
    - Orthopedics and General Practice have the highest patient load and revenue, making them top choices for new hires.
    - Neurology, despite lower patient count, still shows significant demand.
  + Result:
    - I suggest hiring new doctors for Orthopedics, General Practice, and Neurology based on their high patient volume and revenue generation.

1. Is the hospital profitable? How will you determine the profitability?
   * Approach:
     + To determine the hospital's profitability, I created a visual table with Year and Month as columns.
     + I calculated the total revenue, total appointment fees, total profit, and profit percentage using DAX measures.
     + These metrics helped assess whether the hospital is generating a profit or not.
   * Explanation:
     + Total Appointment Fee: Calculated by summing the appointment fees for each patient.
     + Total Revenue: Sum of the total bill amounts for all patients.
     + Total Profit: Derived by subtracting the total appointment fee from the total revenue.
     + Profit Percentage: Calculated by dividing the total profit by the total revenue and multiplying by 100 to get the percentage.
     + The Profit Percentage values were consistent, showing around 98.96% and 98.87%, indicating a very high profit margin.
   * Insight:
     + The profit margin remains consistently high, well above 90%, which indicates the hospital is highly profitable.
     + The hospital generates substantial revenue compared to its costs, with only a small portion of the revenue going toward appointment fees.

A table with numbers and a few months

Description automatically generated

* + Result:
    - Based on the profit percentage consistently being above 98%, it is clear that the hospital is highly profitable.

1. Any Department for which the waiting time is oddly large?
   * Approach:
     + To identify departments with unusually large waiting times, I created a line chart with Department on the x-axis and Average Waiting Time on the y-axis. This visualization allowed for a clear comparison of waiting times across all departments.
   * Explanation:
     + The line chart highlighted variations in waiting times for different departments.
     + By analyzing the chart, I observed that the Neurology department had the highest average waiting time of 36.80 minutes, which stands out compared to other departments
   * Visualization

A graph with numbers and a line

Description automatically generated

* + Insight:
    - The Neurology department's high waiting time could indicate issues such as a lack of sufficient staff, high patient demand, or scheduling inefficiencies.
  + Result:
    - The Neurology department was identified as having the largest average waiting time of 36.80 minutes, which is significantly higher than other departments

1. Come up with strategies to provide discounts to the patients.
   * Provide discounts to patients with a total bill exceeding ₹10,000 to reward high spenders.
   * Offer discounts to patients with a satisfaction score of 6 or above to encourage loyalty and positive feedback.
   * Introduce tiered discount levels based on the total bill amount, e.g., 5% for ₹10,000–₹20,000 and 10% for bills above ₹20,000.
   * Grant compensatory discounts to patients with unusually long waiting times to improve their experience.
   * Implement seasonal or promotional discounts to attract more patients during low-demand periods.
2. 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?
   * Approach
     + Created a table visualization with Doctor Name and Department. Used DAX to assign doctors to two shifts: "8:00 AM - 2:00 PM" and "2:00 PM - 8:00 PM".
     + Applied conditional formatting to highlight the General Practice department.
   * Explanation
     + To assign doctors to shifts, the following DAX formula was used:

|  |
| --- |
| Shift = IF(MOD(RANKX(ALL('Doctor\_Patients\_data'),  'Doctor\_Patients\_data'[Doctor Name], , ASC), 2) = 0,  "8:00 AM - 2:00 PM", "2:00 PM - 8:00 PM") |

* + - This formula ranks doctors by name and alternates the shift assignment using the MOD function.
    - Dr. Johnson and Dr. Williams were assigned to the morning shift (8:00 AM - 2:00 PM) due to higher patient traffic, while Dr. Smith was assigned to the evening shift (2:00 PM - 8:00 PM) to handle lower patient volume.
  + Visualization

A table with a list of doctors

Description automatically generated

* + Result
    - The two-shift policy allows for even distribution of workload, ensures better patient care, and reduces wait times, particularly in the General Practice department.

1. What do you understand by PowerBI gateway? What are its use cases?
   * Power BI Gateway
     + A Power BI Gateway is a tool that enables secure data transfer between on-premises data sources and the Power BI cloud service, allowing you to refresh and access data stored locally in reports and dashboards.
   * Use Cases
     + Data Refresh: Automatically refreshes on-premises data for up-to-date reports.
     + Secure Data Transfer: Ensures encrypted, secure data transfer between on-premises sources and Power BI.
     + Access On-Premises Data: Allows on-premises data (e.g., SQL Server, Excel) to be used in Power BI.
     + Hybrid Data Integration: Combines on-premises and cloud data for unified reports.
     + DirectQuery: Enables real-time querying of live data sources in Power BI.
   * Types of Gateways
     + Personal Gateway: For individual use, refreshing personal data.
     + Enterprise Gateway: For organizations, supporting multiple users and data sources.
2. How would you approach this problem, if the objective and subjective questions weren't given?
   * Import the data into Power BI to begin the analysis process.
   * Use Power Query to transform the data by filtering, merging, and shaping it to fit the required structure.
   * Clean the data by addressing any inconsistencies, handling missing values, correcting data types, and ensuring the format is standardized across the dataset.
   * Identify the key performance indicators (KPIs) that are most relevant to the objectives of the analysis.
   * Analyze each KPI in detail, exploring the trends, relationships, and patterns to derive meaningful insights.
   * Create comprehensive reports and dashboards using various visualizations such as bar charts, line graphs, and tables to present the findings.
   * Add interactive slicers and filters to the reports, enabling users to dynamically explore the data based on different dimensions.
   * Review the final report for accuracy, clarity, and relevance, and provide actionable recommendations based on the insights gathered from the analysis.
3. Can you analyze and write the type of relationship between the doctor id and department, is it one-to-one?
   * Each doctor\_id is associated with one department.
   * Each department can have multiple doctors assigned to it.
   * This makes it a one-to-many relationship from the perspective of the doctors to the department. Alternatively, you could describe it as a many-to one relationship from the department's perspective to the doctors.
   * The relationship between doctor\_id and department is one-to-many because many doctors can belong to a single department, but each doctor is assigned to only one department.

**Report**

The hospital has asked for a report with three tabs:

* Main Tab
* Doctors’ Tab
* Patients’ Tab
* **Using the Main tab in the report,** the hospital should be able to look at the overall metrics like the number of daily visits, revenue produced on that day, customer satisfaction, how busy are different departments on that day, and general waiting time on that day. This tab should have a slicer of date.
* **Using the Doctors’ Tab,** the Chief Of Staff at the hospital should be able to look at the individual doctor’s performance metrics like customer satisfaction, the number of patients he was visited by, the revenue he has generated, and his appointment fees. This tab should have a slicer of the Doctor's Name or ID.
* **Using the Patients’ Tab,** the Patient’s Care Chief at the hospital wants to look at a customer’s profile which would involve metrics like the most frequently visited department, their age, their race, their waiting time, number of visits, the total amount that they have paid to the hospital, etc. All the metrics using which they can address the patient very carefully in their visits. This tab should have a slicer of the Patient's Name or ID.

**Make sure that all the visualizations look decent and are placed in a proper order. Each tab has different POCs (Point Of Contact), so make sure you involve all the metrics that POC may look at in that tab along with those mentioned in the tab description.**

**After making the report on the Desktop ensure that it is hosted on PowerBI service and use the hosted link for submission of the dashboard and mentioning on the resume.**