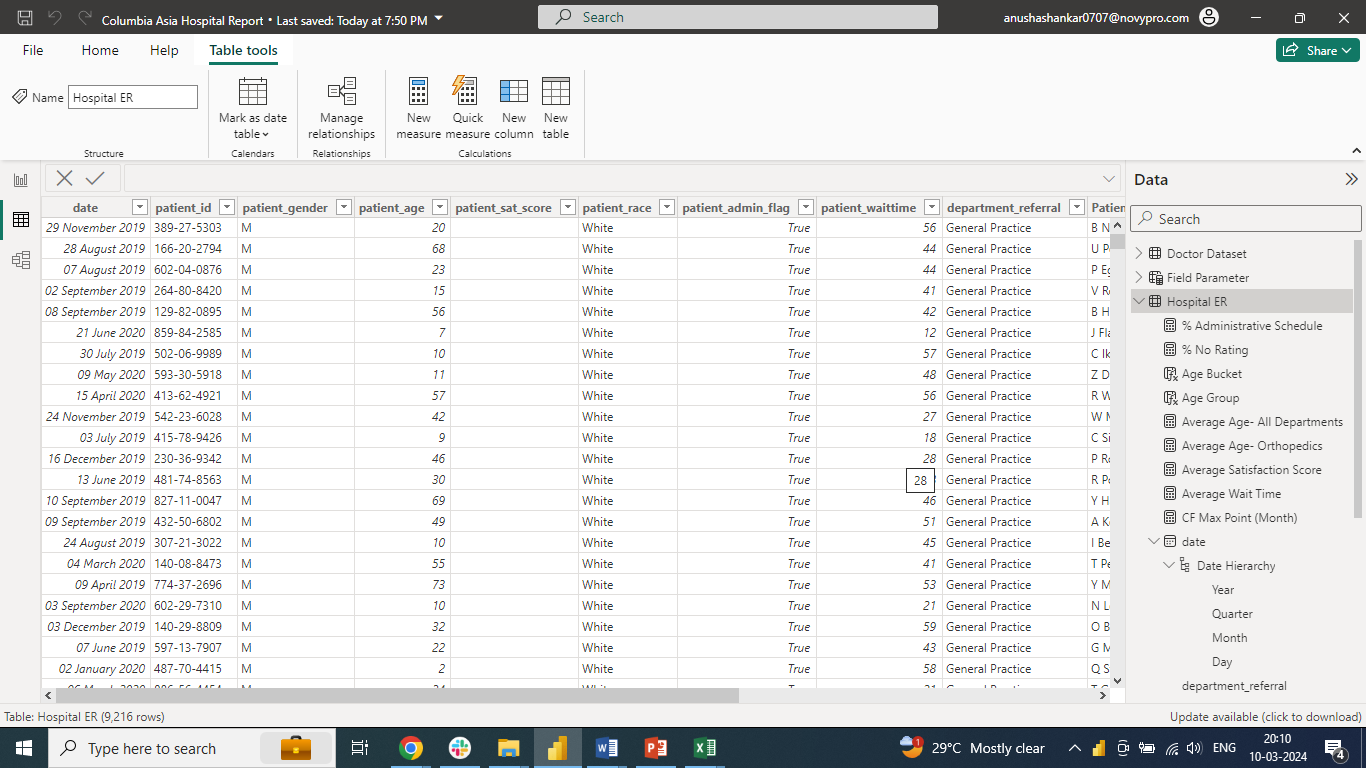
**Objective Questions:**

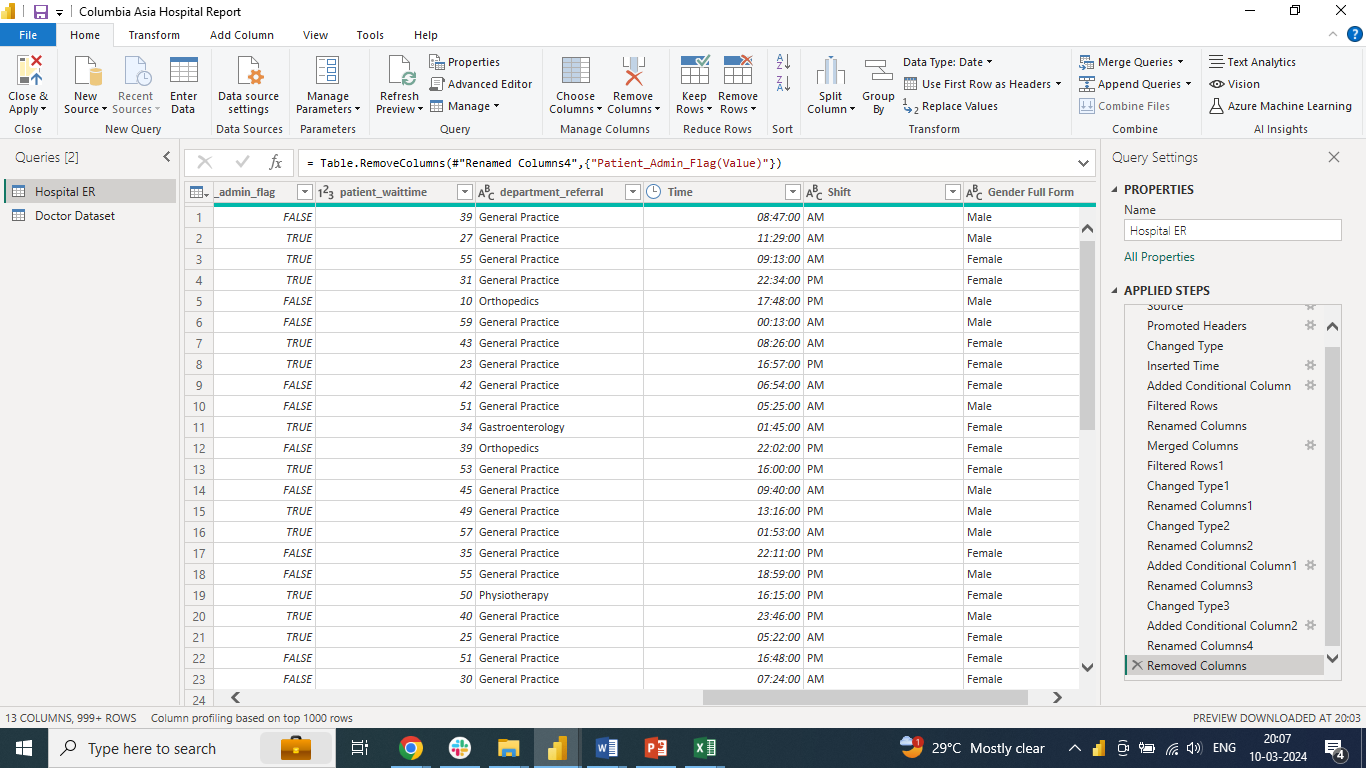
1. In analyzing the hospital dataset with Power BI, ensure data cleaning to address inconsistencies and missing values before further analysis

**In Power Query Editor, below changes has been done:**

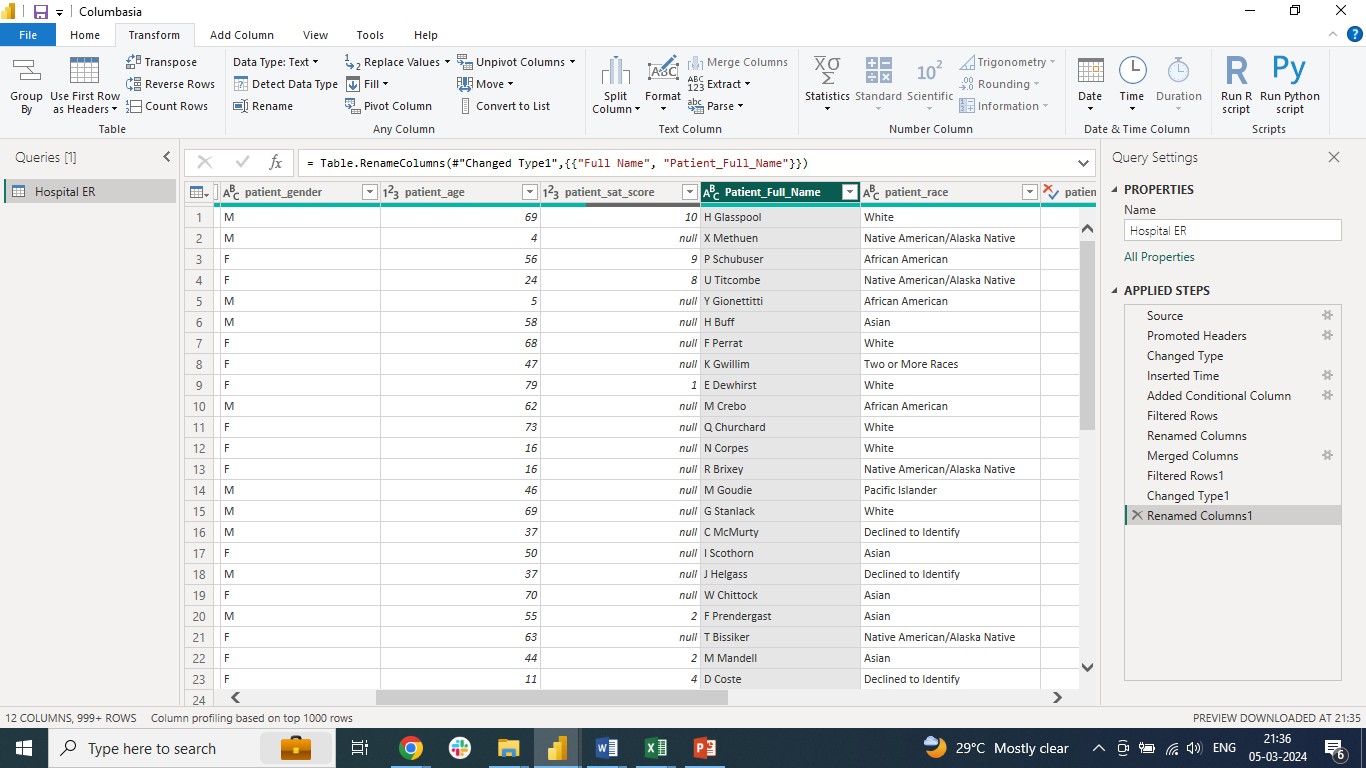
**The date format appears inconsistent. Some dates are in "dd-mm-yyyy" format (e.g., "20-03-2020"), while others are in "mm-dd-yyyy" format (e.g., "12-11-2019"). It's important to have a consistent date format for easier processing and analysis. Hence, the Date is converted to “dd-mmmm-yyyy” format**



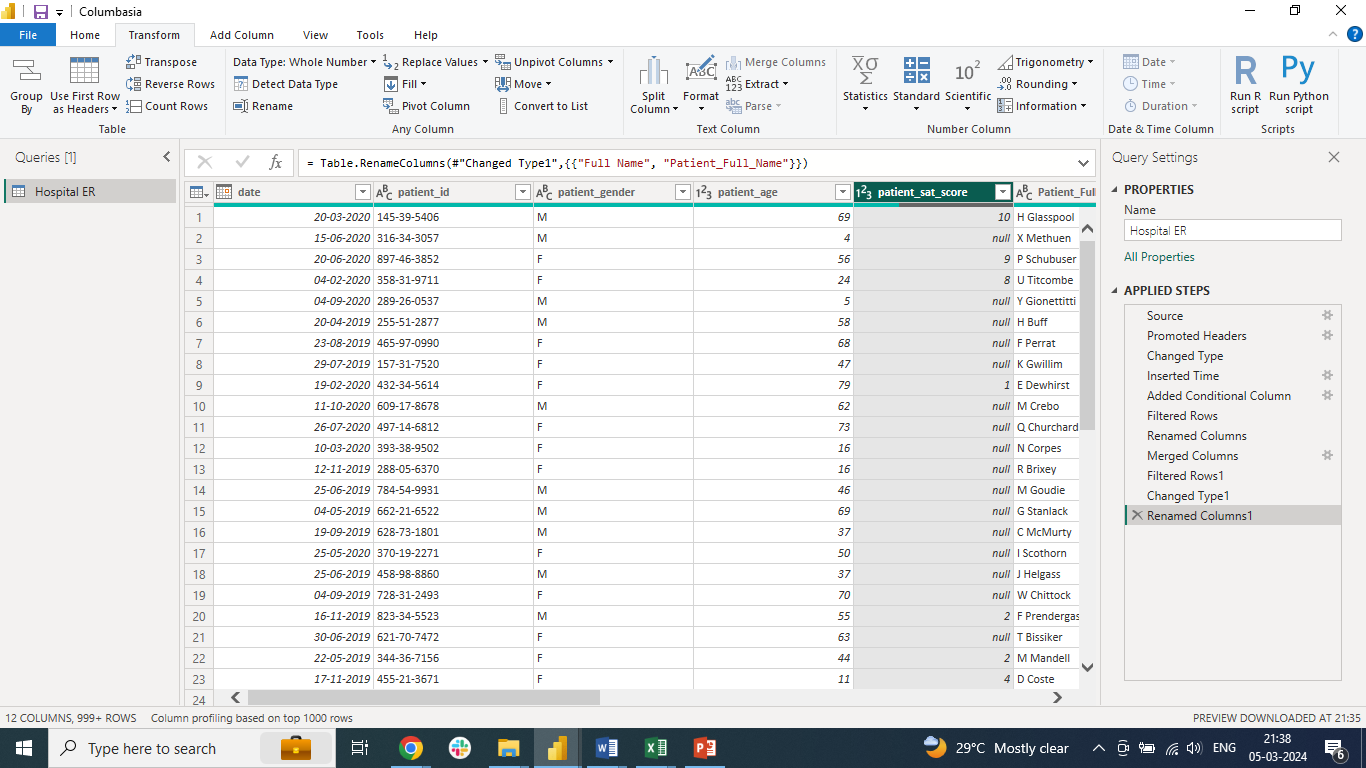
**From the Date Column, Add Column has been applied With Time and Shift- AM/PM information**



**Patient\_first\_initial and Patient\_last\_name has been combined to create “Patient\_Full\_Name” Column**



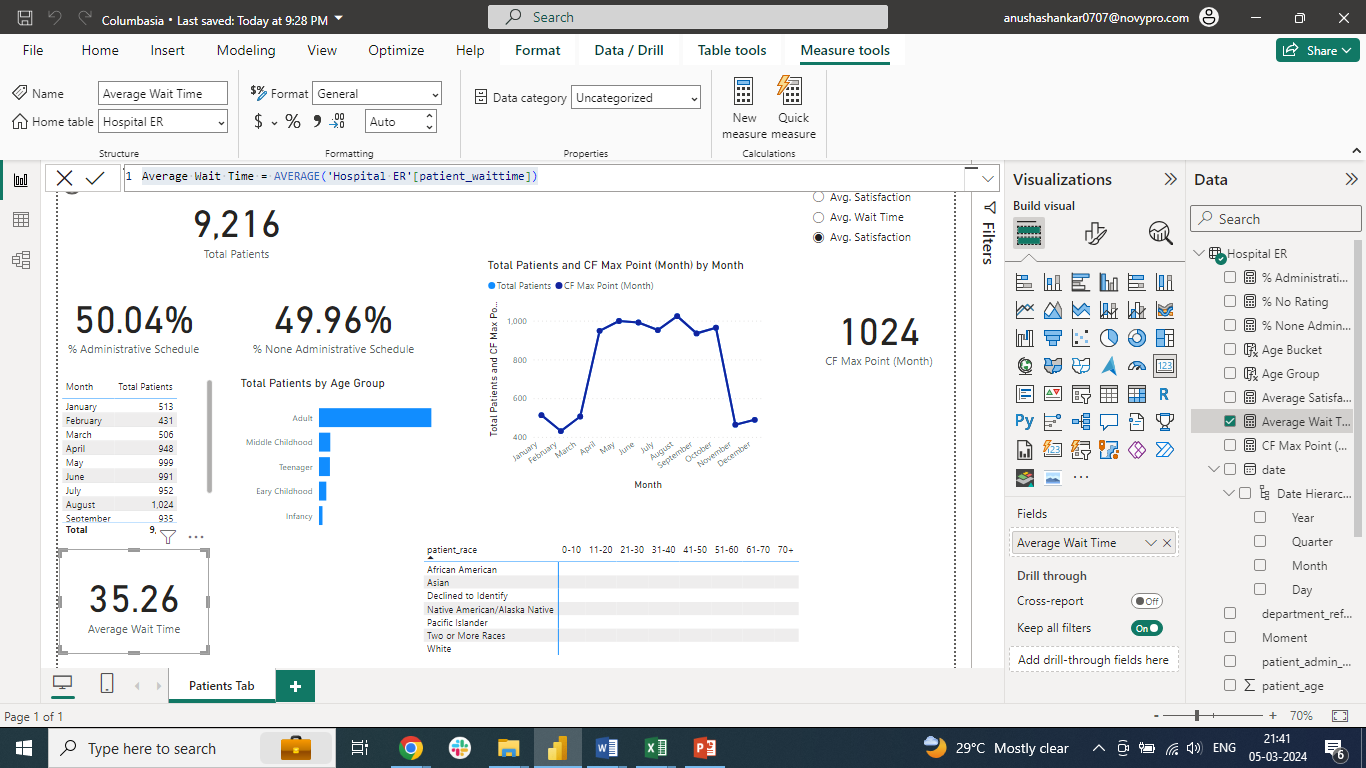
**“Null values” are retained in the patient\_sat\_score**



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

**Average Wait Time is: 35.26**

**Used DAX function:** Average Wait Time = AVERAGE('Hospital ER'[patient\_waittime])



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: **General Practice is the most visited department with 7240 visits**



1. **Patient Visits by Age Group:** Segregate patient visits according to different age groups to see which demographics utilize healthcare services the most- **“White” utilizes the service more**

**Note: Created different Age Buckets to summarize by using Dax function**

Age Bucket =

SWITCH(

True(),

'Hospital ER'[patient\_age]<=10, "0-10",

'Hospital ER'[patient\_age]<=20, "11-20",

'Hospital ER'[patient\_age]<=30, "21-30",

'Hospital ER'[patient\_age]<=40, "31-40",

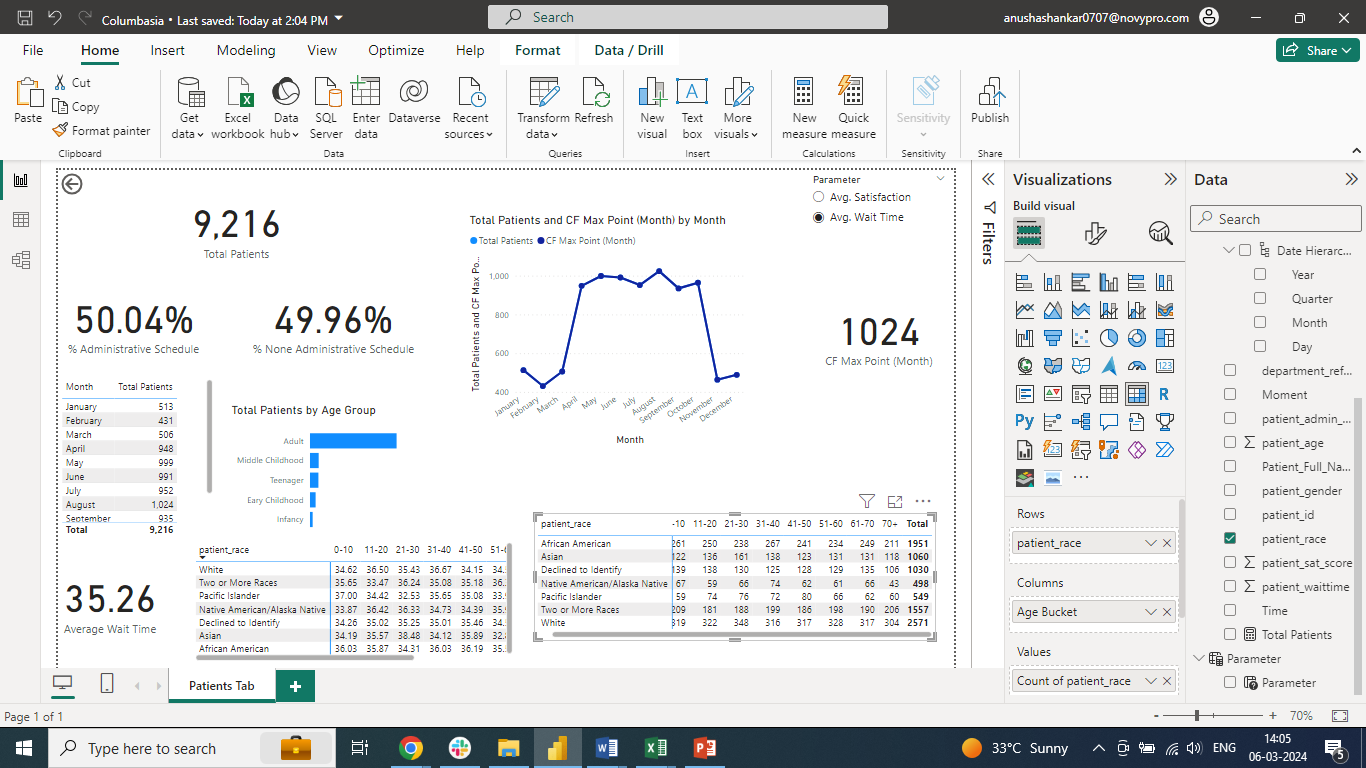
'Hospital ER'[patient\_age]<=50, "41-50",

'Hospital ER'[patient\_age]<=60, "51-60",

'Hospital ER'[patient\_age]<=70, "61-70",

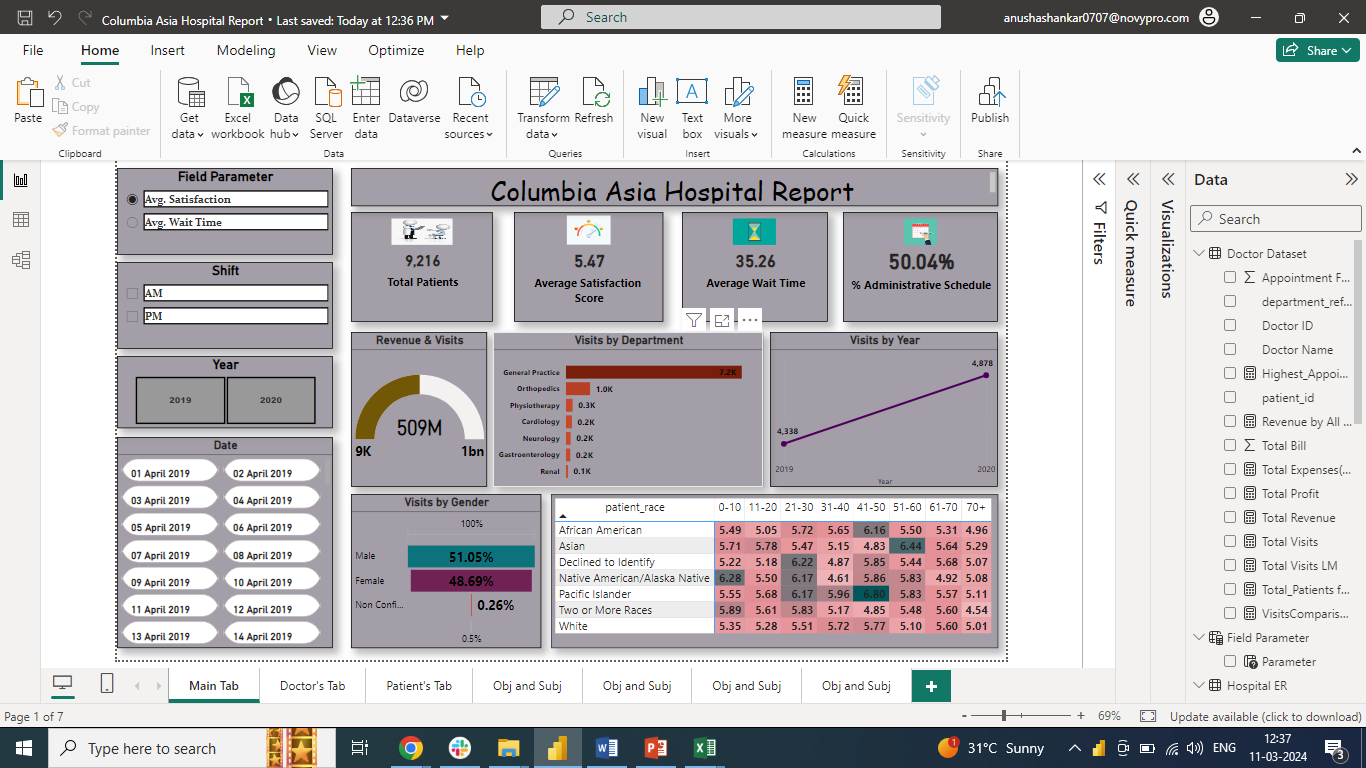
"70+"

)

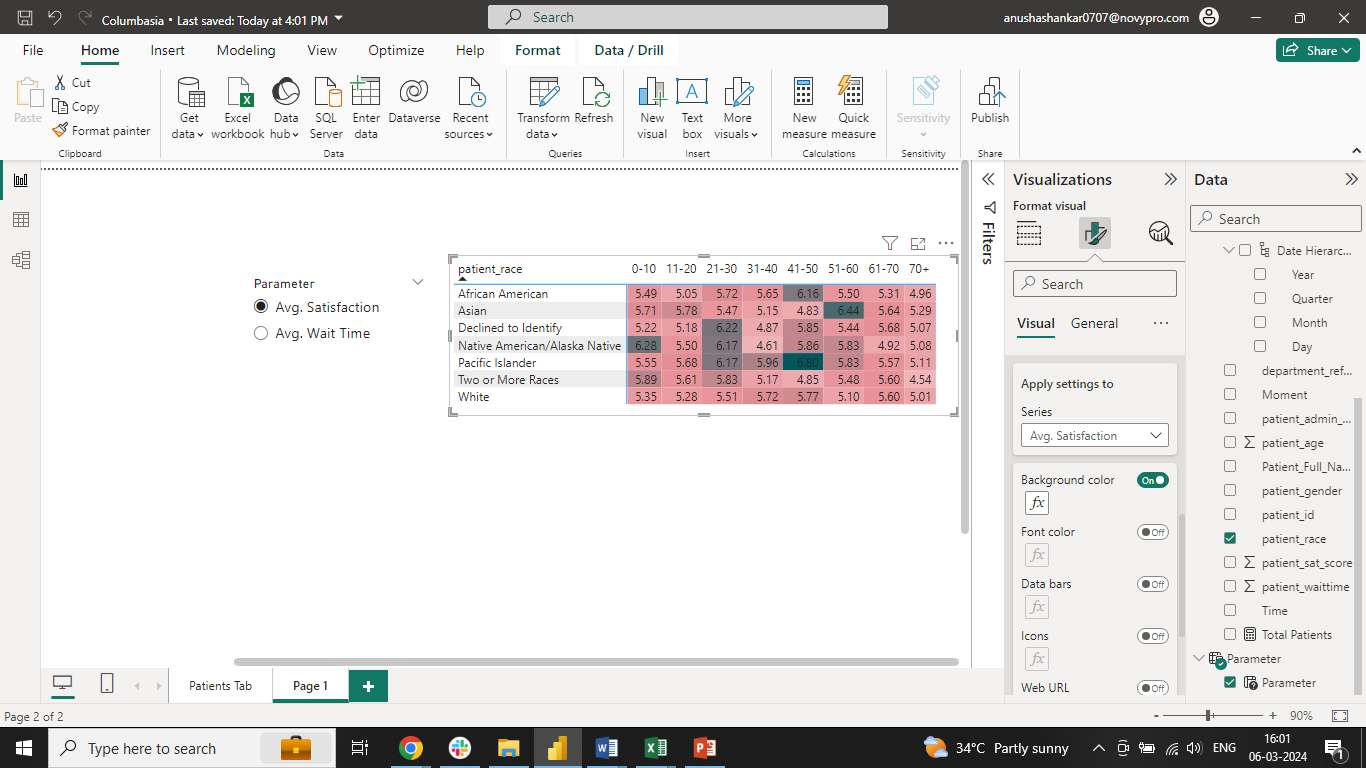


1. Were there any Null values in the data? What would be the best way to handle these Null values and which approach have you opted for? – “**6699 Null values” are present in the data. Null values are retained as it is 75.10% of the total dataset**
2. Is there any relation between the number of visits and the Gender of the patients?

**The dataset indicates a slight male predominance with 51.05% male, 48.69% female patients and 0.26% non-confirming**



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-

* **Pacific Islander are the most satisfied with an Average satisfaction score of 6.80 and they fall in the Age bucket of 41-50**
* **Two or More Races category has the least satisfaction score of 4.54 and they fall in the Age bucket of 70+**
* 

**Step 1: Created separate Table in Table View named “Field Parameter” to include “Average Satisfaction score” and “Average Wait time” as a slicer**

**Syntax: Field Parameter = {**

**("Avg. Satisfaction", NAMEOF('Hospital ER'[Average Satisfaction Score]), 0),**

**("Avg. Wait Time", NAMEOF('Hospital ER'[Average Wait Time]), 1)**

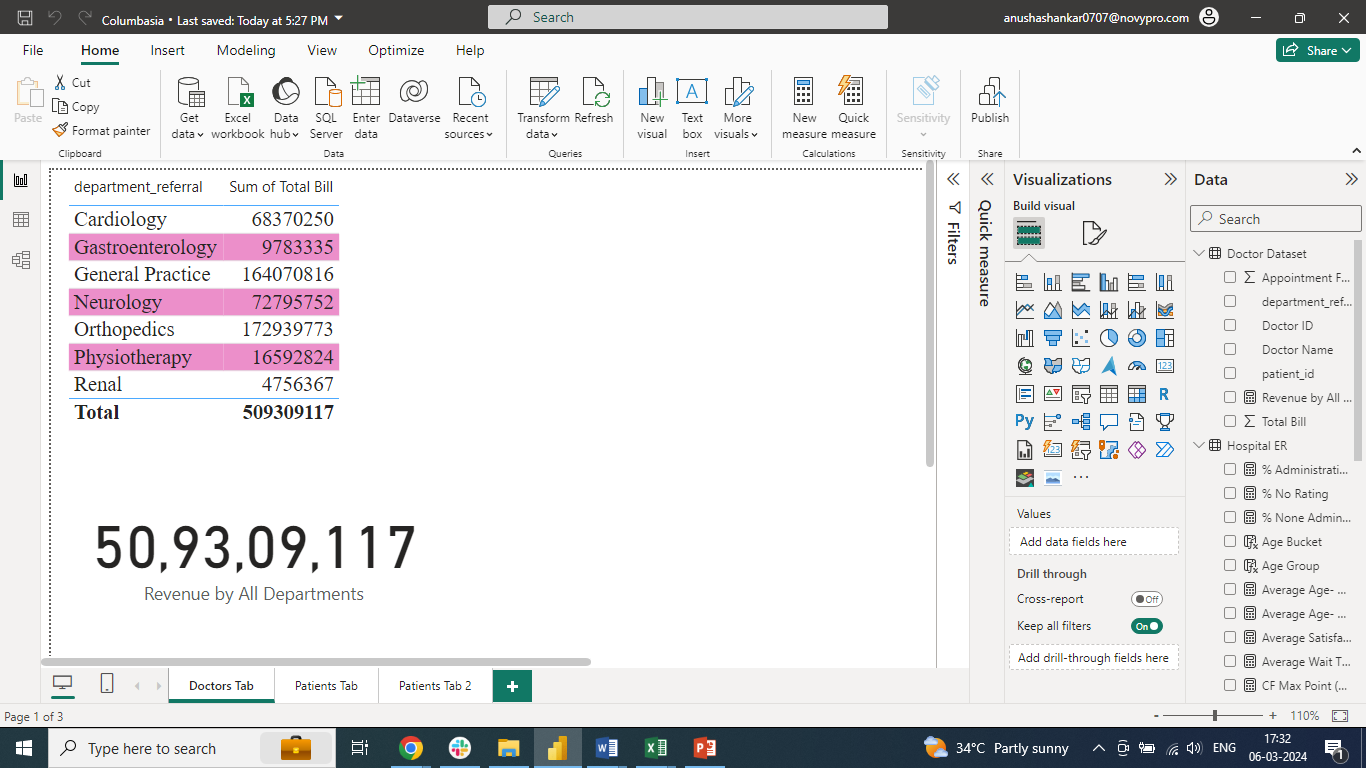
**}**

**Step 2: A Matrix has been created in report View to show the patient’s age group ( Age buckets), racial background and Average satisfaction score**

**Step 3 : Given conditional formatting for the background color- Gradient fill to explain the Average Satisfaction score (ranging from light to dark color- less to high score)**

1. The hospital's managing director seeks to evaluate the revenue of each department to understand how much revenue is generated by each- Total revenue by all the departments is: **50,93,09,117**

**Dax Function used: Revenue by All Departments = SUM('Doctor Dataset'[Total Bill])**



1. Which department is charging the highest appointment fees in general? Use an aggregation DAX function to solve this question- “**Neurology department” charges highest appointment fees**

**Dax Function: Highest\_Appointment\_Fee =**

**VAR MaxFee = MAX('Doctor Dataset'[Appointment Fees])**

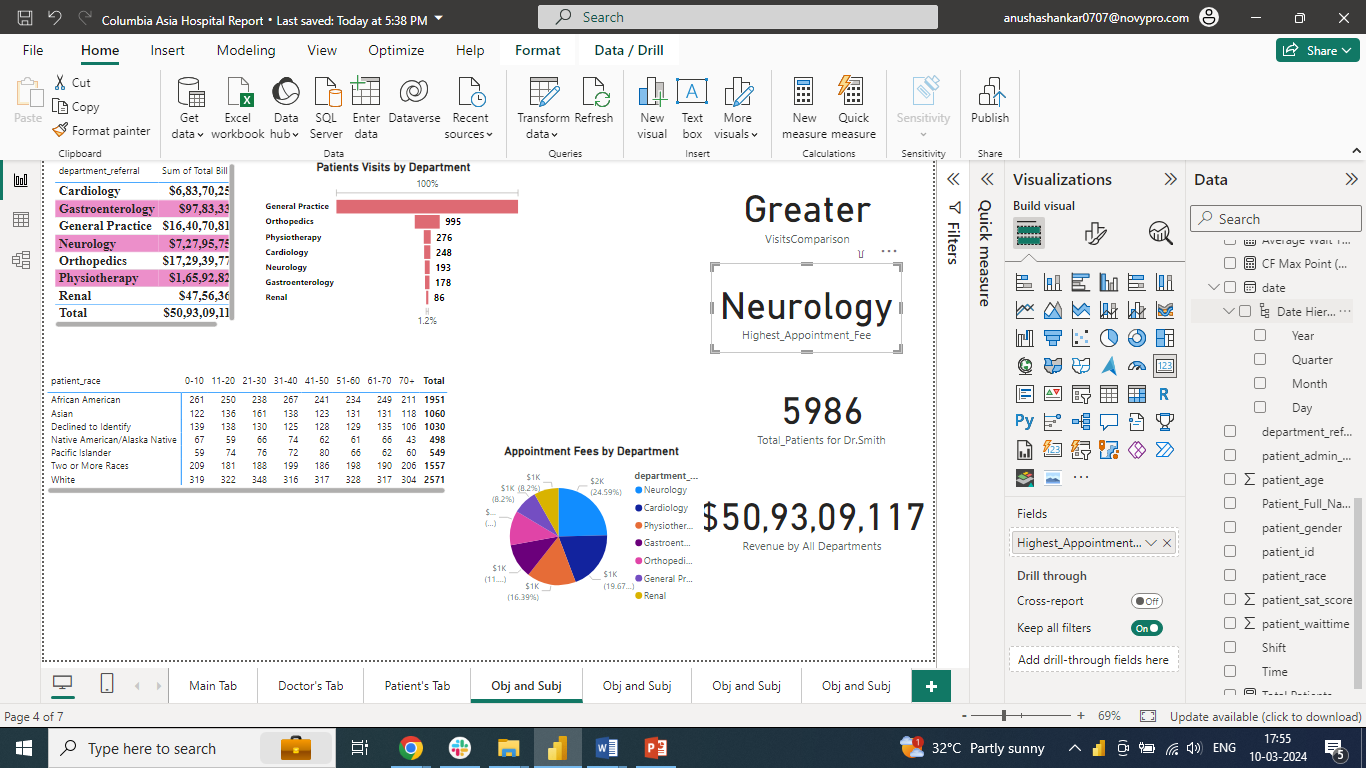
**RETURN**

**CALCULATE (**

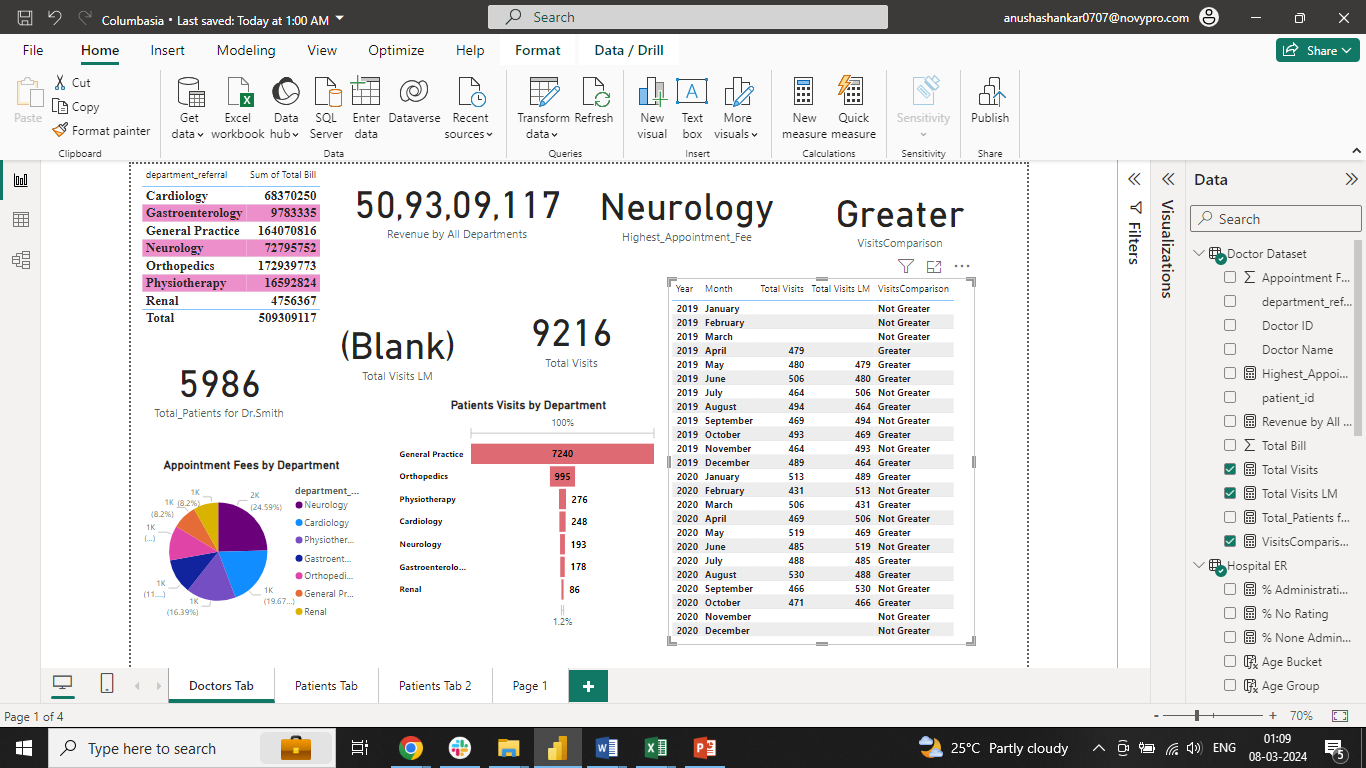
**MAX('Doctor Dataset'[department\_referral]),**

**FILTER ('Doctor Dataset', 'Doctor Dataset'[Appointment Fees] = MaxFee)**

**)**



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.



1. Using ‘Calculate’ and a row iteration DAX function calculate the total number of patients who have visited Dr. Smith- **Total Patients for Dr.Smith is 5986**

**Total\_Patients for Dr.Smith =**

**CALCULATE (**

**COUNTX(**

**FILTER('Doctor Dataset', 'Doctor Dataset'[Doctor Name] = "Dr. Smith"),**

**'Doctor Dataset'[patient\_id]**

**)**

**)**

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? –
   1. **Average age visiting Orthopedics department is 38.66**
   2. **Average age visiting All departments is 39.86**

**Average Age- Orthopedics = CALCULATE(AVERAGE('Hospital ER'[patient\_age]),'Hospital ER'[department\_referral]= "Orthopedics")**

**Average Age- All Departments = CALCULATE(AVERAGE('Hospital ER'[patient\_age]))**

**Note: Approach to calculate the metric is same. However, Filter will not be used if we have to calculate Average age of patients visiting All departments**

1. Were there any data format issues in the data, and if there were/are how you handle them?- **Based on the provided “Hospital ER” dataset, below changes are carried out:**

* **The date format appears inconsistent. Some dates are in "dd-mm-yyyy" format (e.g., "20-03-2020"), while others are in "mm-dd-yyyy" format (e.g., "12-11-2019"). It's important to have a consistent date format for easier processing and analysis. Hence, the Date is converted to “dd-mmmm-yyyy” format**
* **Instead of M, F and NC in “Patient\_Gender”, Full form are given as it will avoid any confusion**
* **Patient\_First\_Initial and Patient\_Last\_Name are combined to Patient\_Full\_Name**
* **Null values in Patient\_Sat\_Score are retained.** **Retaining null values ensures that no data is lost and allows for accurate representation of the dataset**

**No data format issues in “Doctor\_Patients\_data” dataset**

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?-

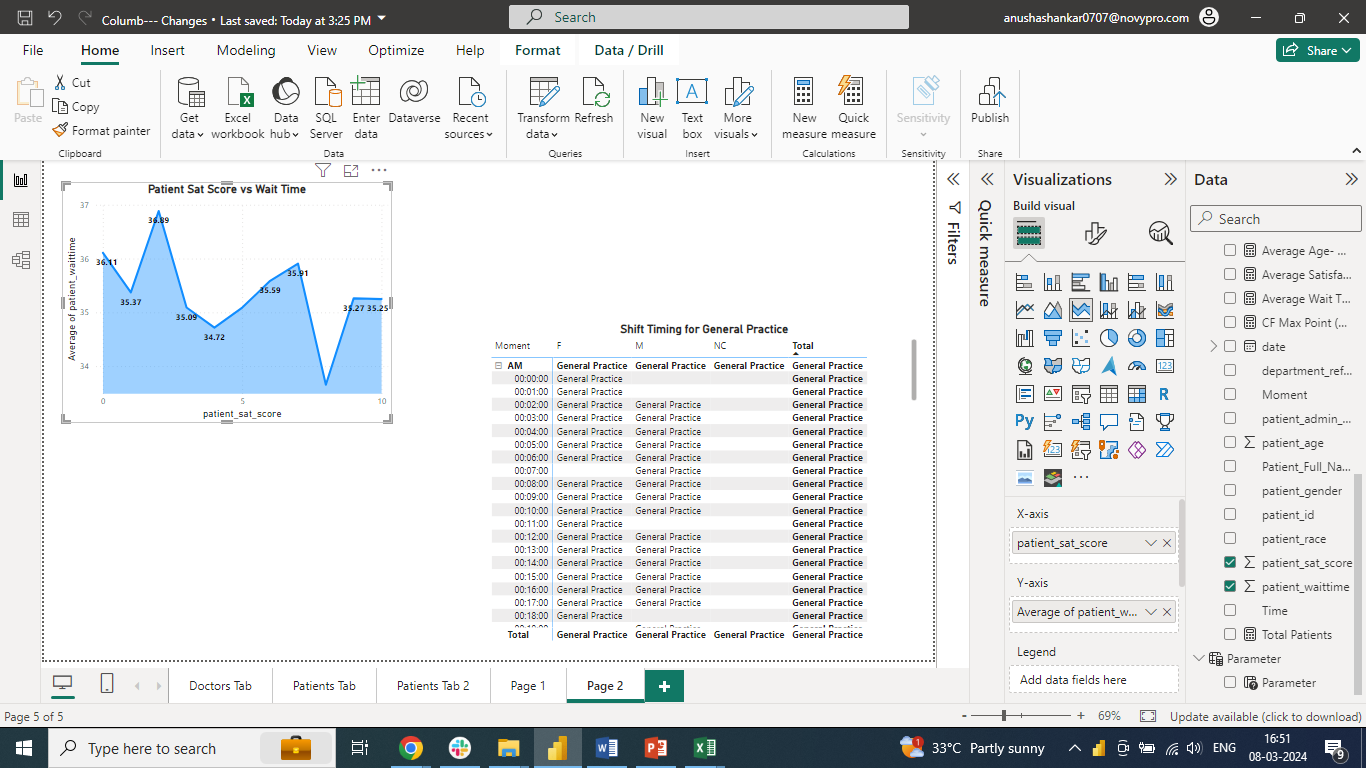
**=Table.AddColumn(#"PreviousStep", "NewColumnName", each [Column1] + [Column2])**

**M Query or M language, is a powerful scripting language used within Power Query to perform data transformation tasks. It is designed to be intuitive and expressive, making it easier for users to perform complex data manipulation operations without having to write extensive code. M Query is used extensively in Power Query for tasks such as filtering, sorting, merging, grouping, and transforming data from various sources**

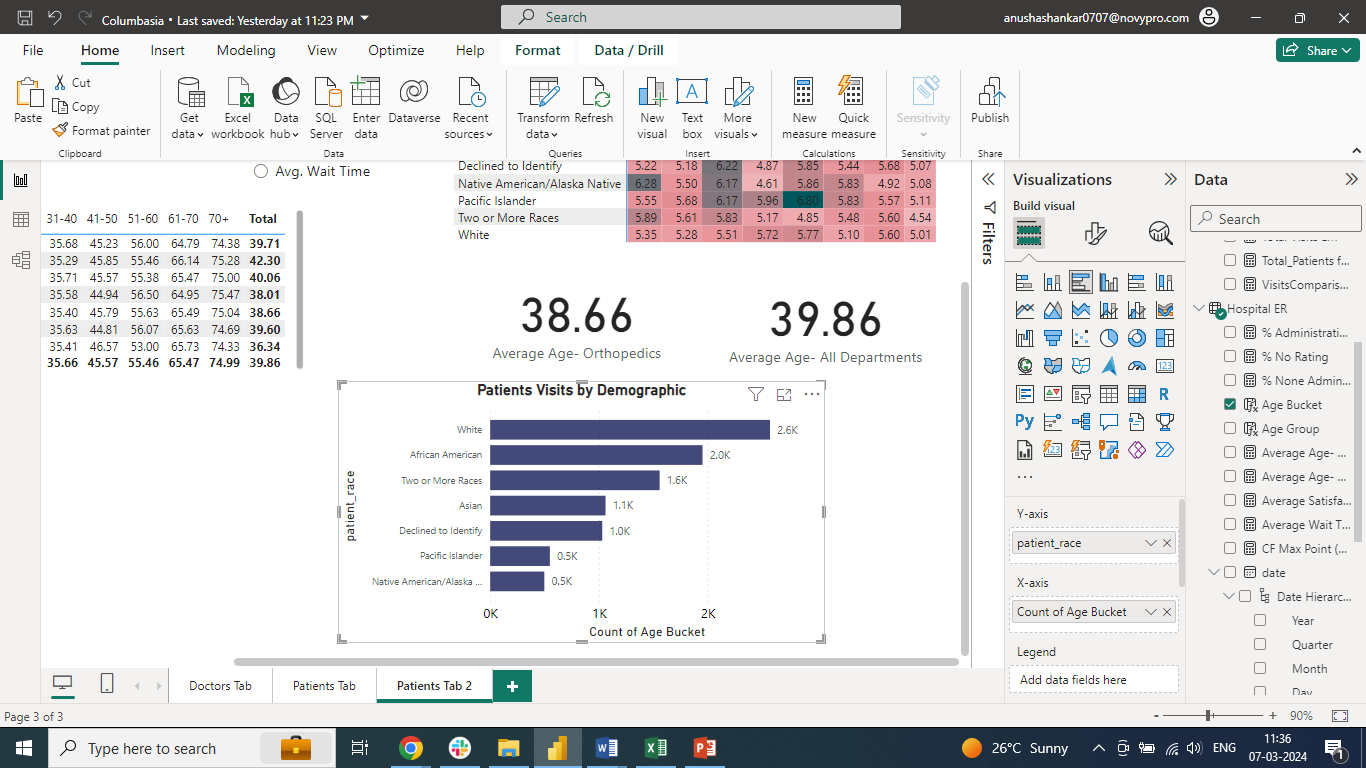
**Subjective Questions**

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

**No Significant Relationship: There might not be a significant relationship between patient SAT scores and wait times. As per the Stacked area chart, we see that the wait time has just close to 1 min difference between high (10) and lowest sat score (0)**

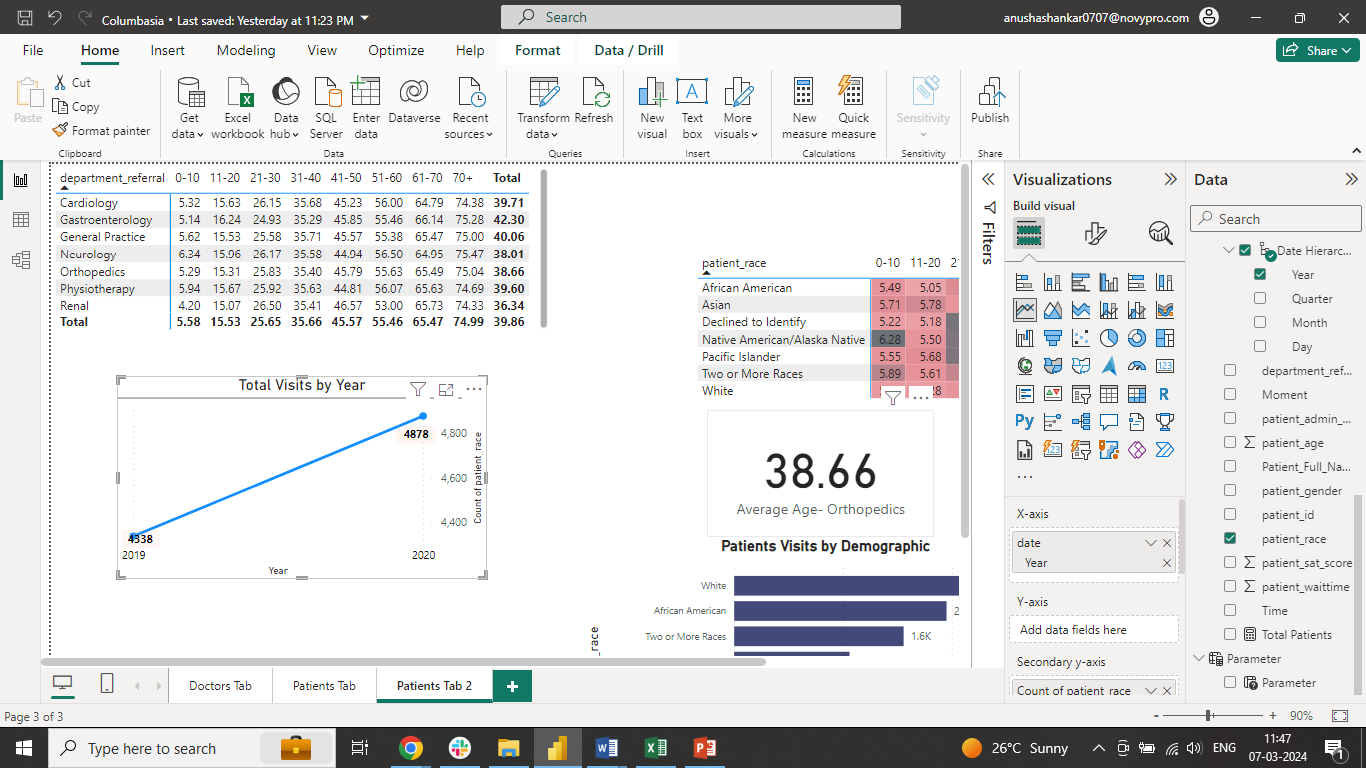


1. How do patient demographics affect the frequency of visits to different departments?



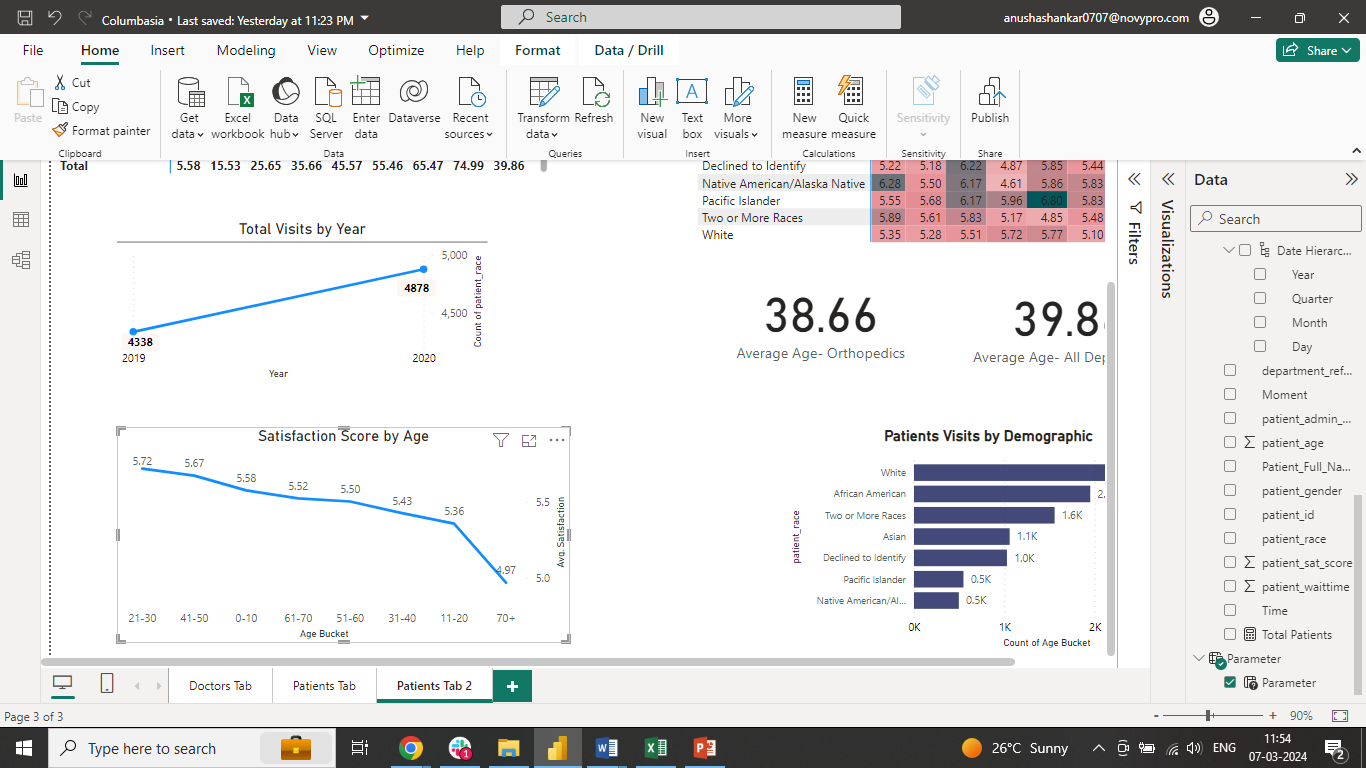
**As per the Clustered bar chart we can understand that White’s number of visits are the highest and Native American/ Alaska Native’s number of visits are the lowest indicating racial discrimination**

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



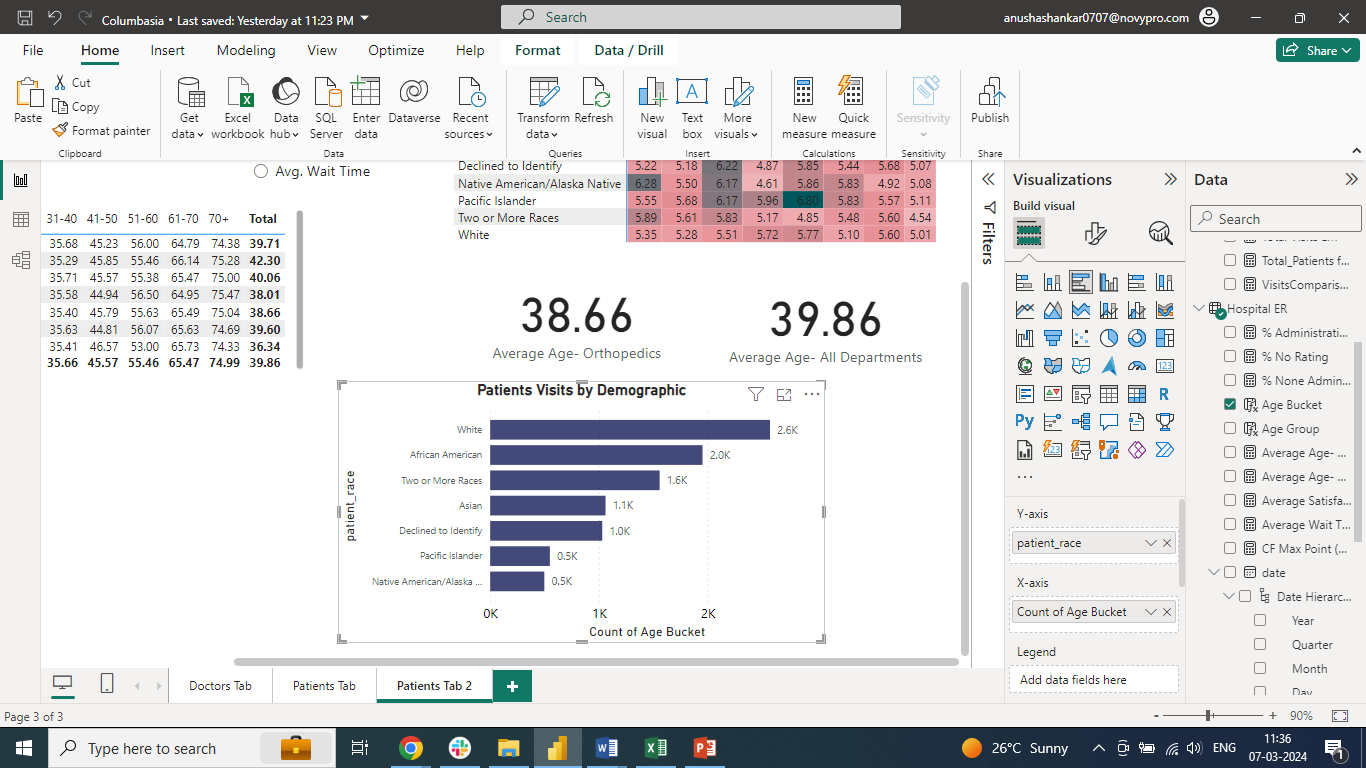
**From 2019 to 2020, we observe that the Patients visits has increased from 4338 to 4878**

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

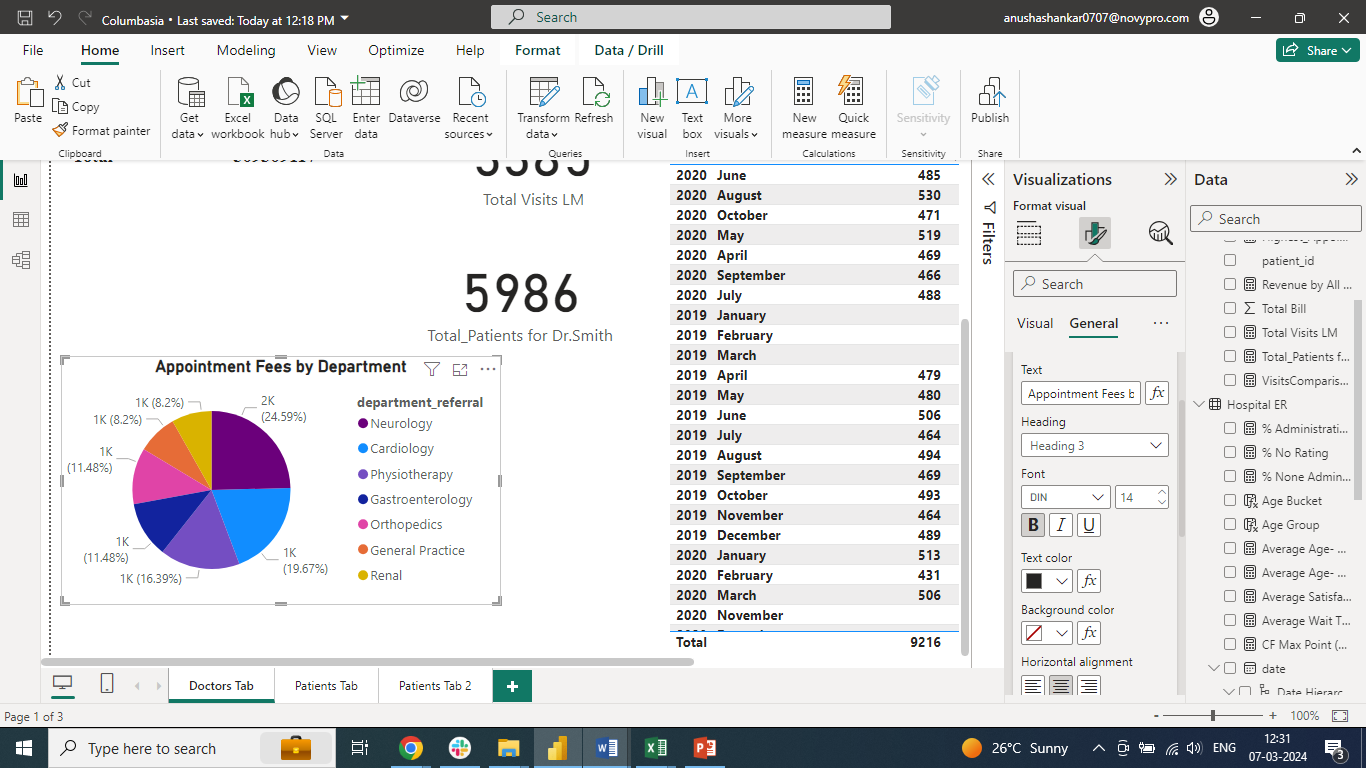


**21-30 Age groups has highest satisfaction scores vs 70+ has lowest satisfaction scores with 5.72 vs 4.97**

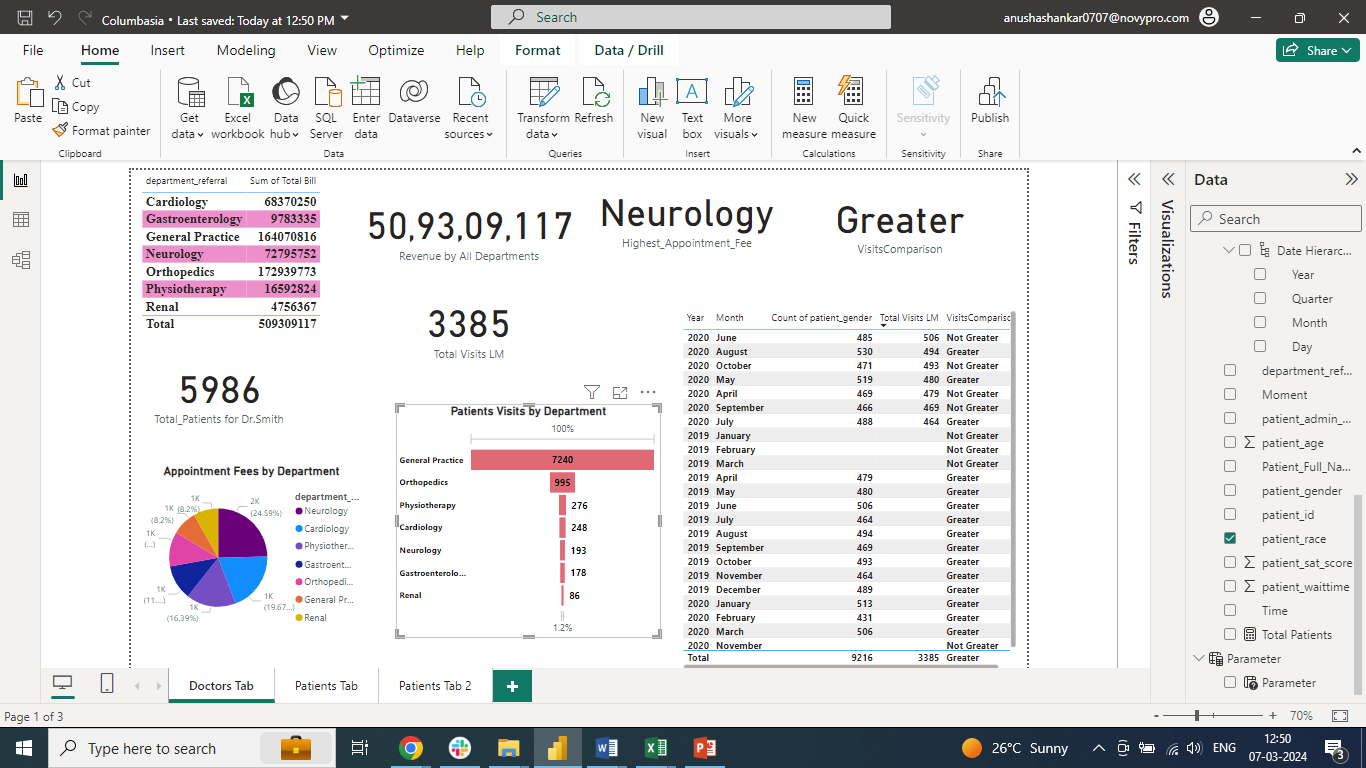
1. Say someone outside of the hospital claims that there is racial or gender-based discrimination in the hospital, how will you identify whether the claim was right or not?- **Yes, it is true that there is racial discrimination in the hospital. With the number of visits based on Demographics we find that the White race is given more priority compared to Native American/ Alaskan**



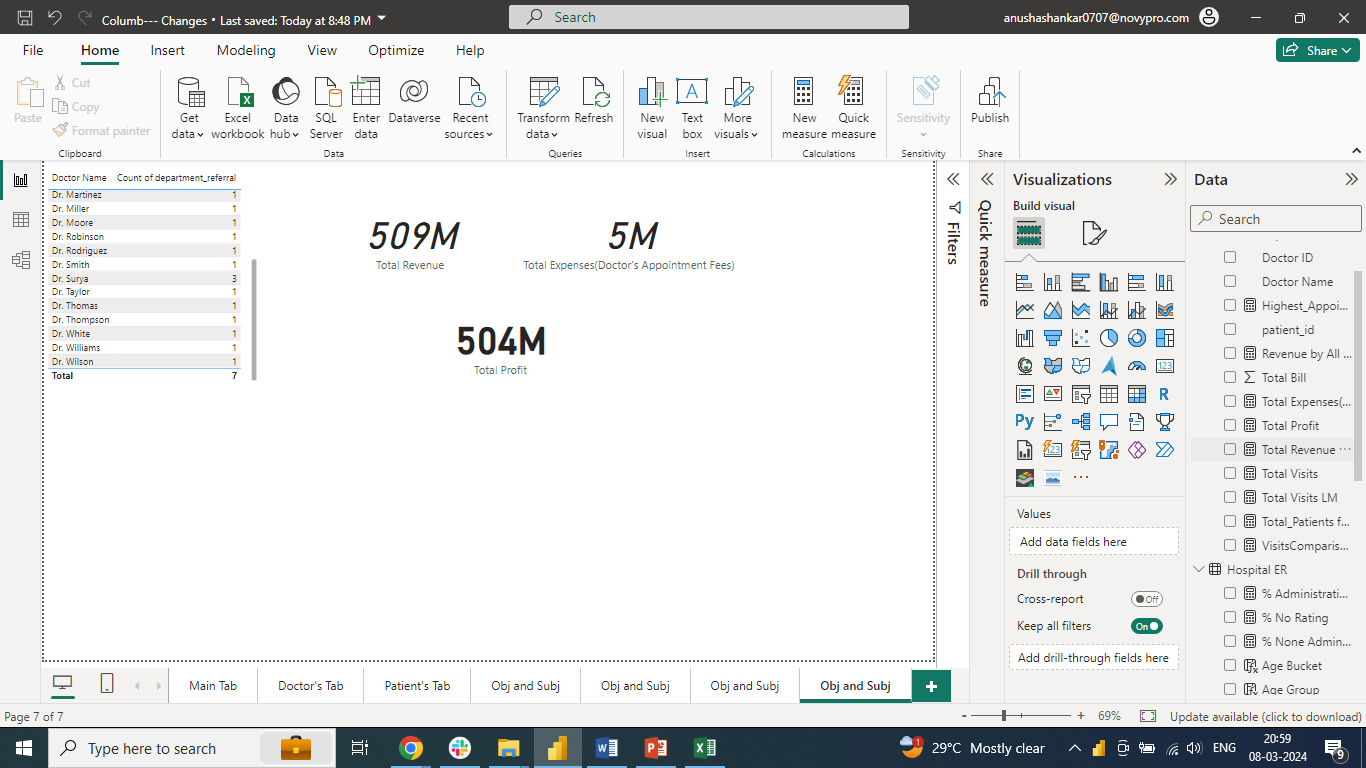
1. The hospital management intends to offer discounts to patients. How should these offers/discounts be assigned to patients, on what basis, and why? – **As we can observe from the Pie chart that the Neurology department charges the highest compared to others, the management can try and provide discounts to the patients visiting Neurology**



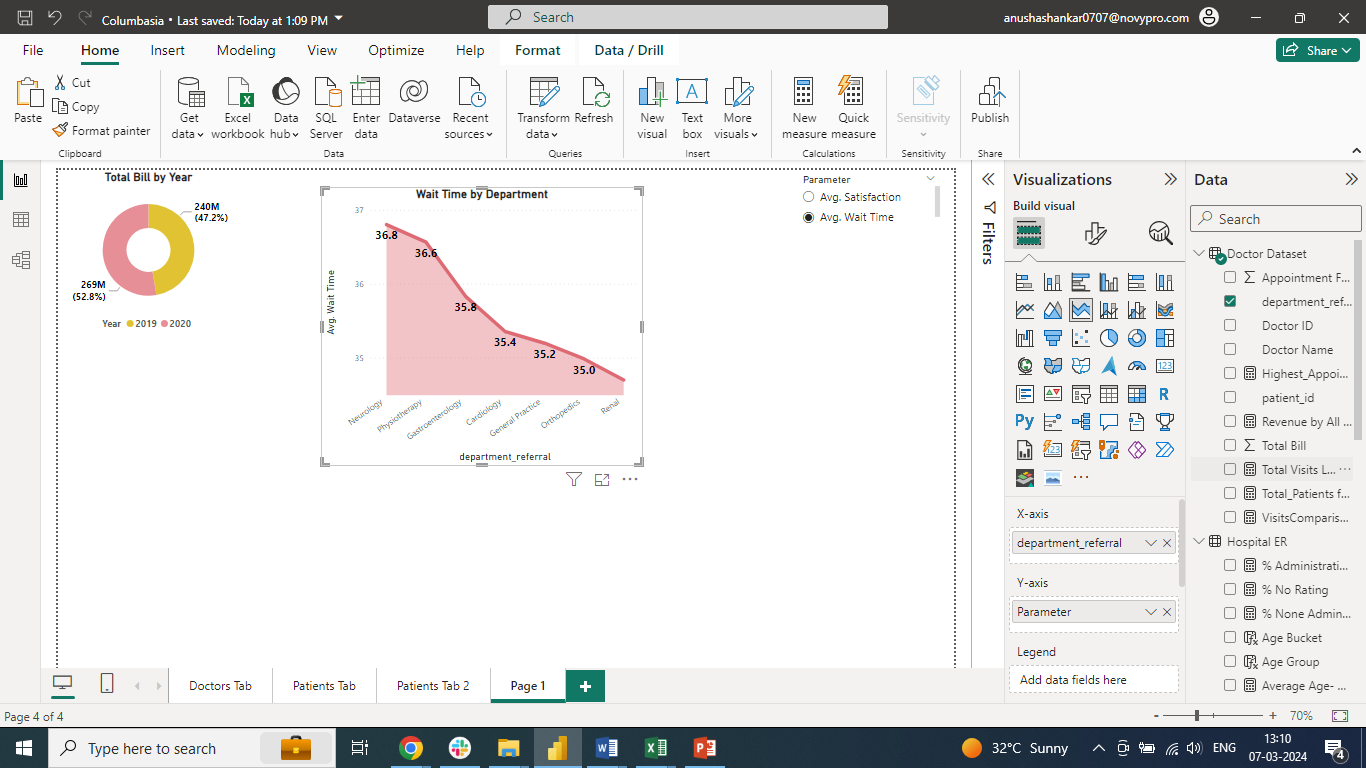
1. The hospital has a budget to hire 2-3 new doctors. They have asked for your suggestions on which departments they should hire- **We can hire 2-3 doctors in General Practice department as we can understand from the Funnel chart that highest patients visit (7240 visits) this department more than others**



1. Is the hospital profitable? How will you determine the profitability?- **Yes, the Hospital is profitable. We can determine this by Subtracting Doctor’s Appointment fees (considering this as expense) - Total Revenue**



1. Any Department for which the waiting time is oddly large? – **No, All departments has almost same Wait time**

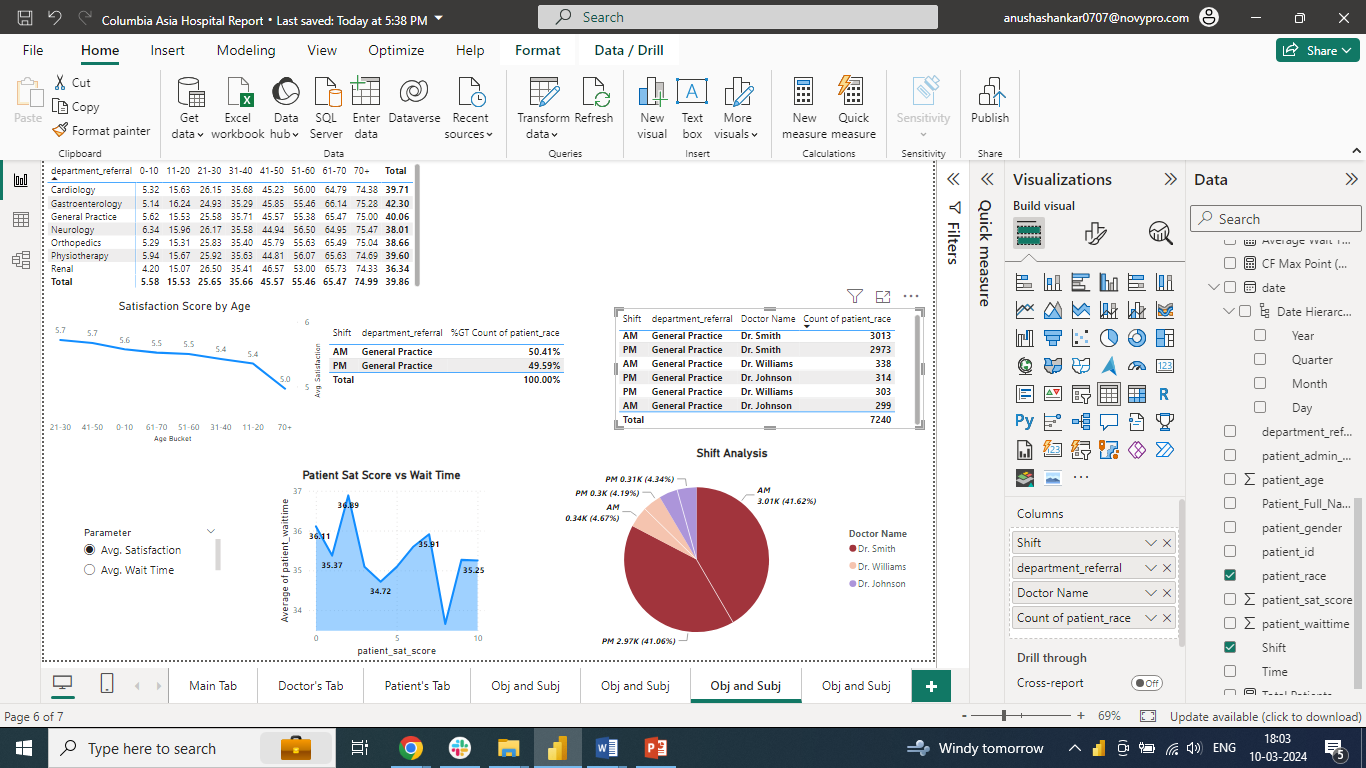


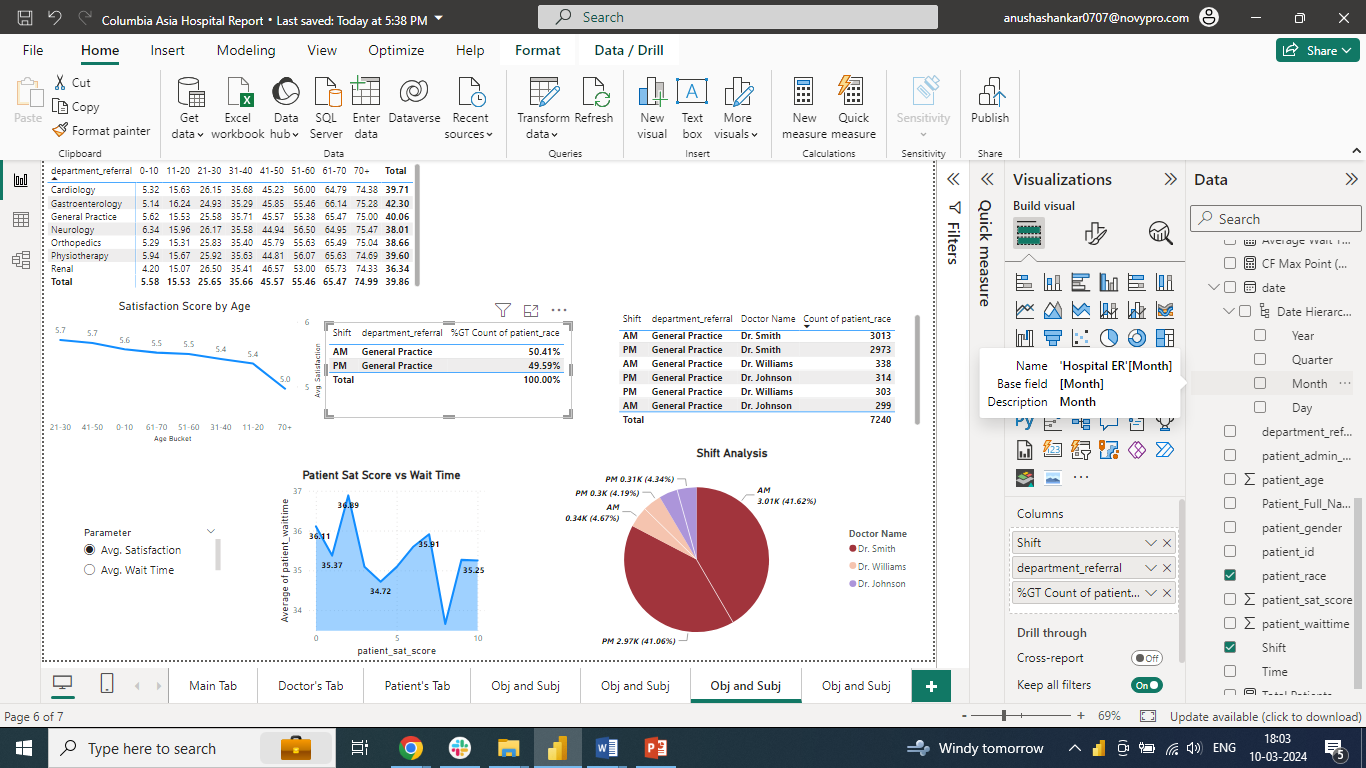
1. Come up with strategies to provide discounts to the patients

* **Income-Based Discounts: Offer discounts based on patients' income levels. Patients who fall below a certain income threshold could be eligible for reduced fees or sliding-scale payments**
* **Senior Citizen Discounts: Provide special discounts for senior citizens to accommodate their healthcare needs, considering their fixed incomes and potentially higher healthcare expenses**
* **Flexible Payment Plans: Offer flexible payment plans or financing options to patients who may struggle to pay for healthcare services upfront. This can include zero-interest payment plans, extended payment periods, or low down payment requirements**

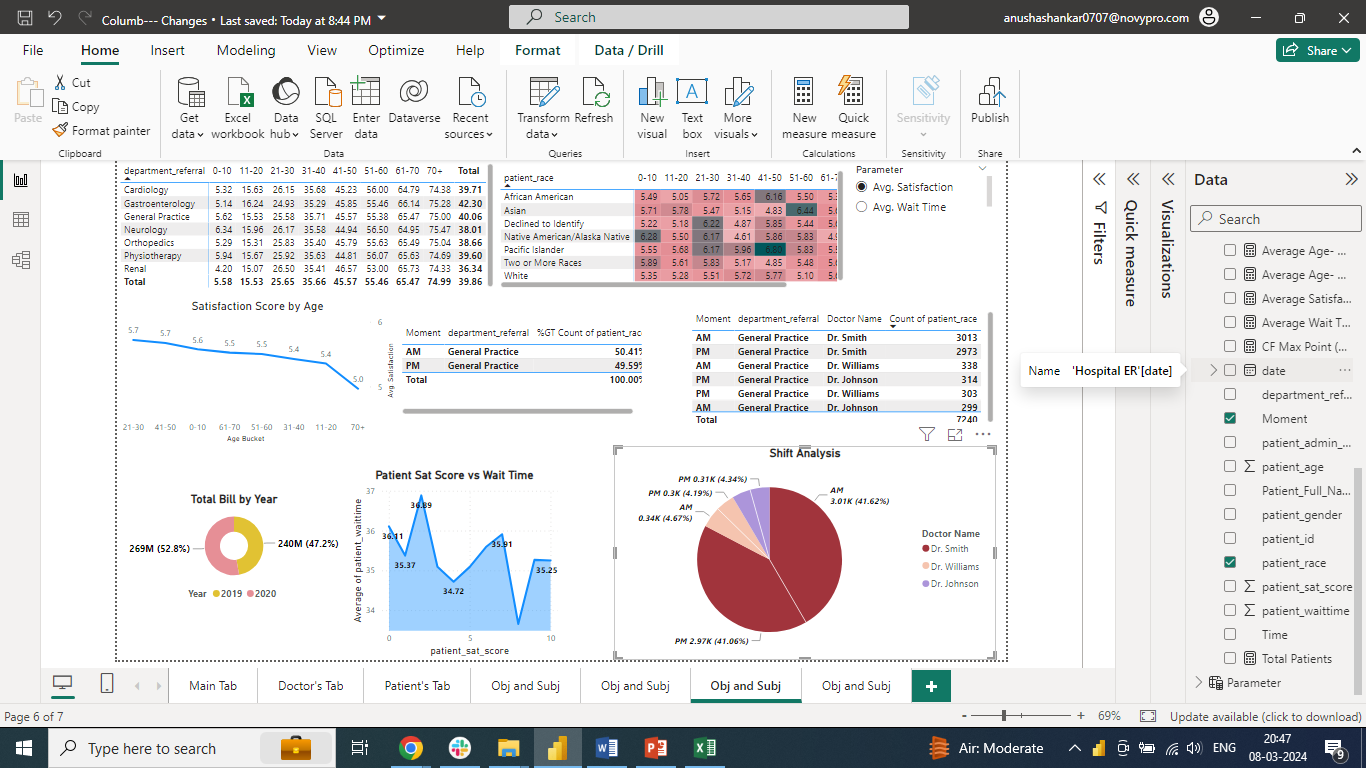
1. Say you need to align the doctors of the “General Practice” department to work in one of the two shifts, how will you identify what will these two shifts' timings be, and how will you divide the doctors in these two shifts? And also will this 2 shift policy be helpful for the hospital? –

* **By viewing the data via Matrix, we observe that all the 3 doctors are working in both the shift timings**
* **However, Dr. Smith consults more number of patients (5986) compared to Dr.Johnson (613) and Dr.Williams(641). Since all the 3 doctors belongs to “General Practice” department. We can make Dr.Johnson and Dr.Williams to work in PM shift as we can see Dr.Smith consulting about 2973 patients in the second shift**
* **Yes, the 2 shift policy is helpful because by implementing this policy we can align Dr.Smith to AM shift and other 2 doctors to PM shift equalising the workload of all the doctors**





**Dr.Smith’s patients vists in AM/PM as shown below:**



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

**Power BI Gateway is a tool provided by Microsoft as part of its Power BI suite of business analytics tools. It enables secure data transfer between on-premises data sources and the Power BI service in the cloud. The Power BI Gateway acts as a bridge between your organization's data sources, such as databases, files, and other on-premises data repositories, and the Power BI service, allowing users to access up-to-date data for reporting, analysis, and visualization purposes.**

**Here's a breakdown of its key components and use cases:**

* **Data Connectivity: Power BI Gateway provides a secure and reliable connection between on-premises data sources and the Power BI service. It supports various data sources, including SQL Server, Oracle, MySQL, PostgreSQL, SharePoint, files (Excel, CSV), and more.**
* **Data Refresh: One of the primary use cases of Power BI Gateway is to enable scheduled data refreshes for Power BI datasets stored in the cloud. It allows Power BI reports and dashboards to stay up-to-date with the latest data from on-premises sources without manual intervention.**
* **Direct Query: Power BI Gateway supports Direct Query mode, allowing users to access and analyze large volumes of data stored in on-premises databases in real-time. This enables interactive and dynamic reporting without the need to import data into Power BI datasets.**
* **Personal and Enterprise Gateways: Power BI Gateway comes in two versions: Personal Gateway and Enterprise Gateway. The Personal Gateway is designed for individual users or small businesses, while the Enterprise Gateway is suitable for larger organizations with multiple users and data sources. The Enterprise Gateway offers centralized management, monitoring, and administration capabilities.**
* **Security: Power BI Gateway ensures secure data transfer between on-premises data sources and the Power BI service by using encryption and authentication protocols. It supports Windows authentication, Azure Active Directory (AAD) authentication, and OAuth authentication methods to ensure data security and compliance with organizational policies.**
* **Hybrid Deployments: Power BI Gateway enables hybrid deployments, allowing organizations to leverage both cloud-based and on-premises data sources in their analytics solutions. This flexibility is particularly beneficial for organizations with regulatory requirements or data privacy concerns that mandate keeping certain data on-premises.**

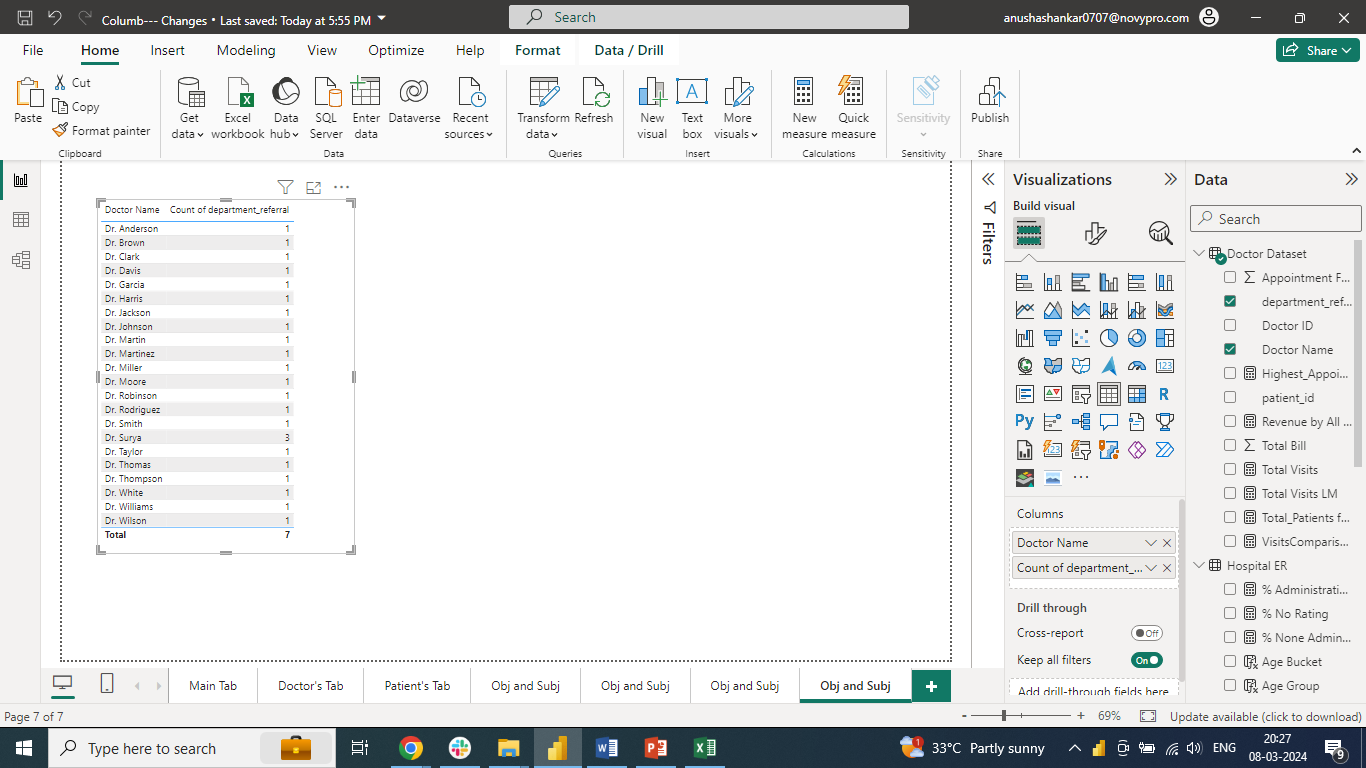
1. How would you approach this problem, if the objective and subjective questions weren't given?

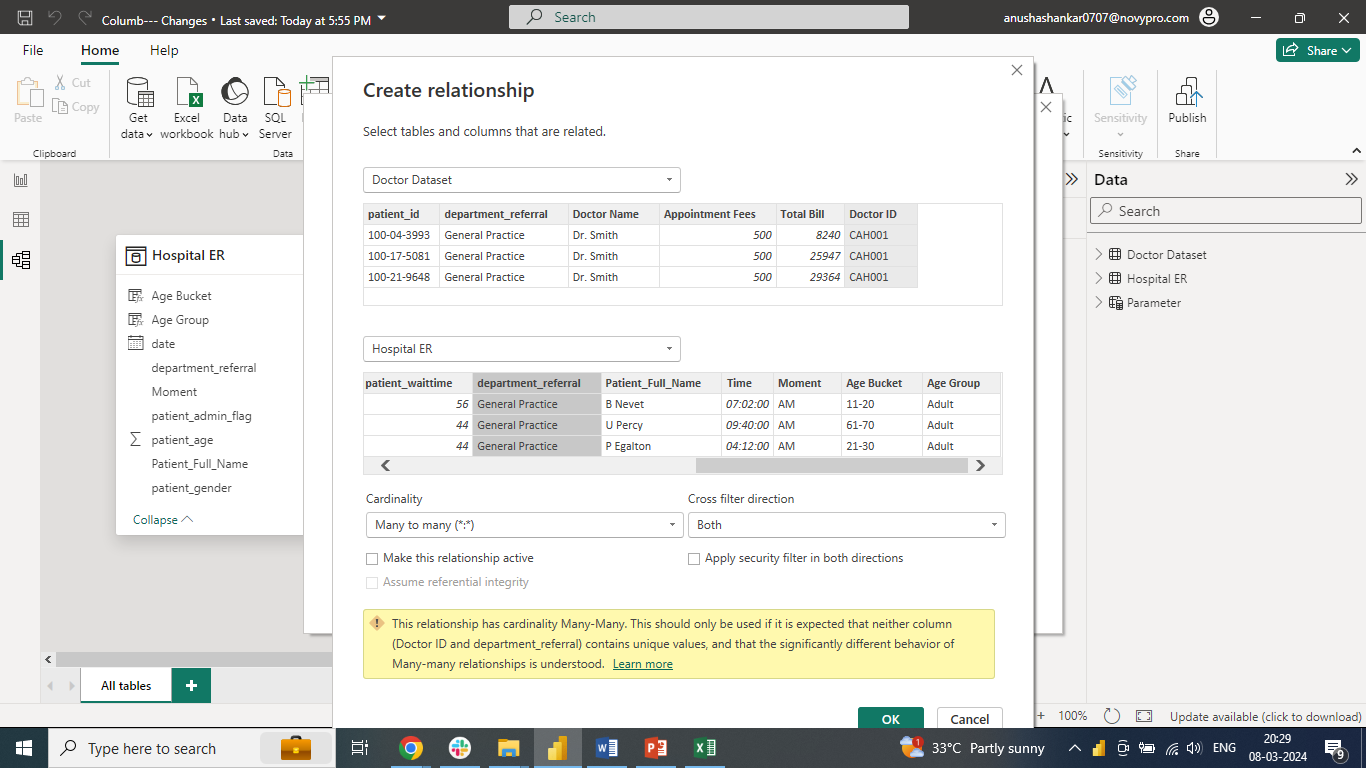
* **Define the objectives and scope of the "Columbia Asia" project, outlining the key questions the Power BI solution should address.**
* **Perform data cleaning, transformation, and modelling to prepare the data for analysis.**
* **Create a data model in Power BI that accurately represents the relationships between different data entities.**
* **Design and develop interactive reports and dashboards in Power BI to visualize the data and answer the key questions identified during requirement gathering.**
* **Use a variety of Power BI visualizations, such as charts, graphs and tables, to present insights effectively.**
* **Incorporate interactive features like slicers, filters, and drill-downs to allow users to explore the data dynamically.**

**By following this approach, we can develop solution for the "Columbasia" project, enabling stakeholders to gain valuable insights from their data and make informed decisions**

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

* **Doctor ID and Department has “Many to Many” relationship**
* **As per the table below, we can see that Dr.Surya is attending 3 different departments**





**Note: In Power BI, a many-to-many relationship refers to a scenario where one record in one table can relate to multiple records in another table, and vice versa**

**Please access the Power BI report using the Link**

[**https://app.powerbi.com/groups/me/reports/7ff7d267-a193-4328-8844-430cea2360c0/ReportSection08b0cc655ef670295db0?experience=power-bi**](https://app.powerbi.com/groups/me/reports/7ff7d267-a193-4328-8844-430cea2360c0/ReportSection08b0cc655ef670295db0?experience=power-bi)

