Report On Mini Project

Name: SARANYA

Batch: ANB009

Contact Number: 8056431186

Mail ID: saransaran71218@gmail.com

Domine: Health care

Submission Date: 28-11-2025

Mentor Name: Kumaran M

Raw set Link:
https://drive.google.com/drive/folders/1-XOc_XNMO2Dxyrgmjh2cc7YXbbIrenmB?usp=sharing

Cleaned Data set Link:
https://drive.google.com/drive/folders/1_AY2ckv-LLhCCp3SMQsOV6tAiYJpBtfc?usp=sharing

Source link: https://www.kaggle.com/datasets/anshfactor/health-centre-data-raw

Analysis of Health Centre

Description: I have a dataset with 2489 records and 10 columns in an Excel file named Patient_Health_Centre_Dataset, to be messy! This "messiness" usually means the data isn't immediately ready for analysis and will require significant cleaning and shaping in the Power Query Editor. Each row represents a single record, likely corresponding to a patient visit or an individual patient profile.

Excel Cleaning:

Age Column: I used Replace value in **Age** column because mostly there is **47** that is i got average in age column so i filled **47 in blank cells**

Before

Age
8
91
3
82

After

Age
47
47
47
47
47
47
47
47
8
91
3
47
47
47
82

Gender Column:

I replace the values that **"O"** is **Others**, and i want to get a average for **gender column** to fill blank cell there is 783 others comparatively others gender so i filled **Other** in **blank cells**

Before

Gende
Other
O
Other
Other
Female
Other
Female
Female
Other
Female
Other
M
M
Male

After

Gende
Other
Other
Other
Other
Female
Other
Female
Female
Other
Other
Female
Other
Male
Male
Male

Cost Column:

I replace the values that **N/A** is **49325.86**, and i want to get a average for **Cost column** to fill blank cell there is **48** others comparatively others Cost values so i filled **49325.86** in **blank cells.**

Before

58309.86
43705.1
71412.86
44838.01
19146.02
17145.11
63856.26
87669.87
45591.85
N/A
32704.56
47411.73
20720.45
95306.19
95806.13

After

58309.86
43705.1
71412.86
44838.01
19146.02
17145.11
63856.26
87669.87
45591.85
49325.86
32704.56
47411.73
20720.45
95306.19
95806.13

Date Column:

Here the most common reason for this type of error is that the original data in the **Visit_Date column** has mixed date Type. I changed the data type in **date formats.**

Before

Visit_Dat
2022-07-29
2022-05-03
2022-03-09
2021-08-04
2023-01-18
2025-05-09
2023-09-17
2020-11-26
2024-07-04
08/09/2021
21/01/2021
2024-11-28
2024-02-27
2025-05-04
2022-01-16
2023-07-22
2022-06-29

After

Visit_Dat
29-07-2022
03-05-2022
09-03-2022
04-08-2021
18-01-2023
09-05-2025
17-09-2023
26-11-2020
04-07-2024
08-09-2021
21-01-2021
28-11-2024
27-02-2024
04-05-2025
16-01-2022
22-07-2023
29-06-2022

Name and Doctor Name Column:

There is so many blank cell in **Name(48) and Doctor Name(45) Column**
So i filled **Unknown** to all Blank cells of this both column

Before

Name	Doctor_Name

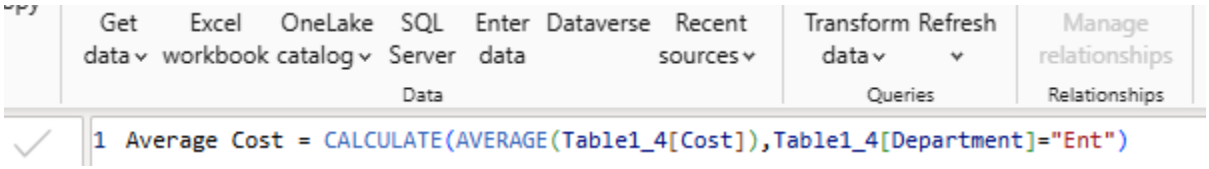
After

Name
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown

Doctor_Name
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown
Unknown

Power Bi visualisation

1. Card:



Formula:

Average Cost =

```
CALCULATE(AVERAGE(Table1_4[Cost]),Table1_4[Department]="Ent")
```

Observation for Average Cost (Department: Ent)

I created the measure that the average cost in the **ENT department is 47.50K**. This indicates that, on average, the ENT-related activities or treatments incur a cost of ₹47,500. This value helps in comparing ENT performance with other departments and identifying whether the cost level is high, moderate, or aligned with hospital standards.

2. Card:

Observation for Total Count of Patient ID (2488)

The total number of patients recorded in the **dataset is 2,488**. This indicates that **2,488 unique patient** entries have been captured, reflecting the overall patient volume handled during the selected period. This count

helps in understanding patient load, resource utilization, and departmental performance.

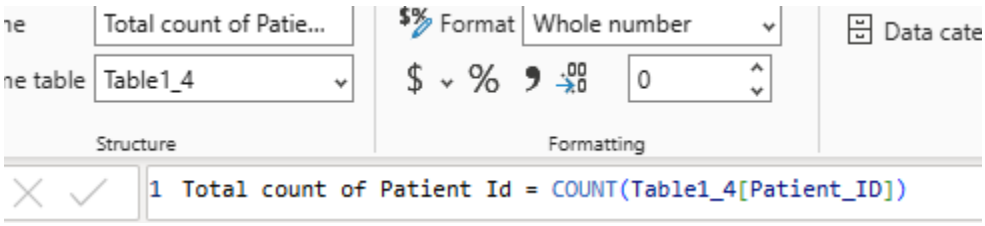
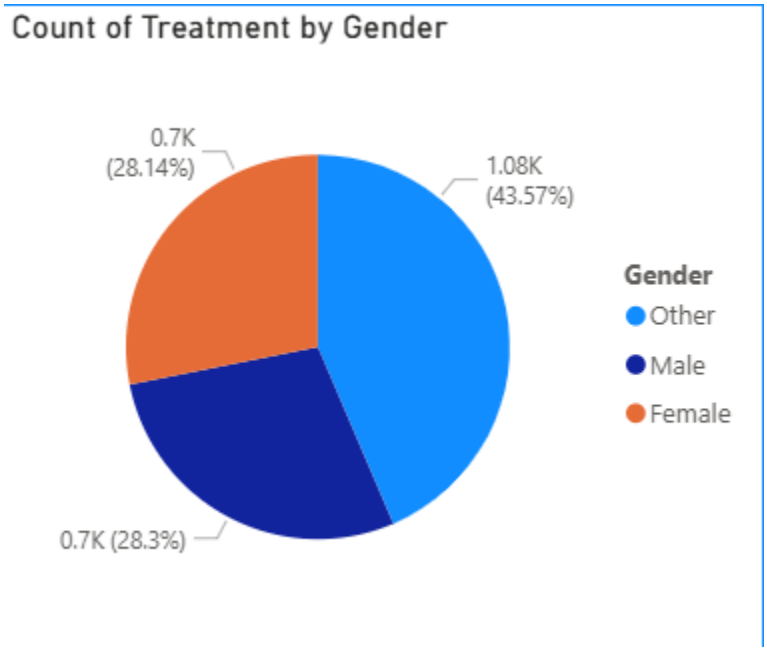


Chart visual:

1. Pie chart:

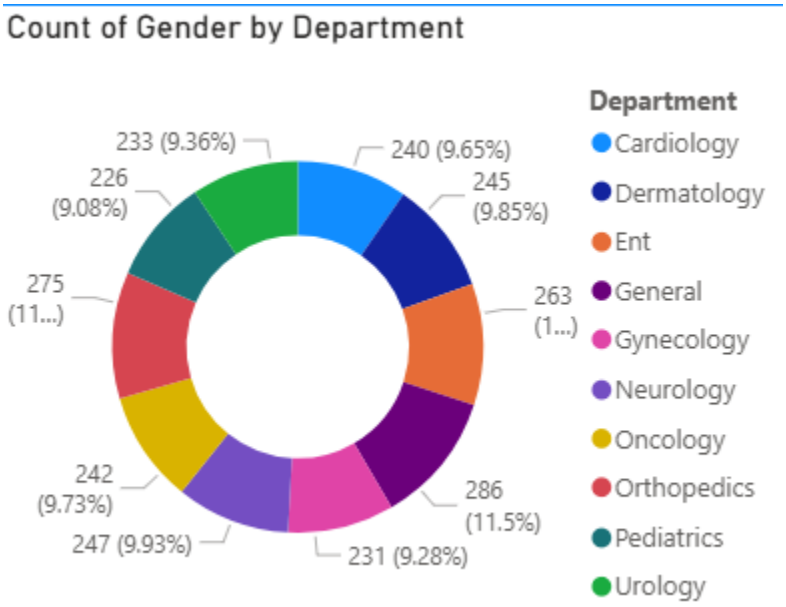
Observations: "Other" Gender Dominates Treatment Count: The largest proportion of treatments is attributed to the "Other" gender category, accounting for **43.57% of the total count**. This segment has a count of **1.08K.Male** and **Female** Counts are Nearly Identical: The counts for Male and Female treatments are almost the same.Female treatments account for **28.14% (Count: 0.7K)**.Male treatments account for **28.3% (Count: 0.7K)**.High Concentration in "Other" Group: The "Other" category accounts for significantly more treatments than either the Male or Female category individually, suggesting that the population receiving treatment is heavily skewed towards this group, or that this category is capturing a large, diverse population that requires treatment.Overall Treatment Counts: The total number of treatments visualized is approximately **2.48K** (calculated by summing the three counts: **1.08K + 0.7K + 0.7K)**



2. Donut chart:

Observations:

1. Near-Equal Distribution: The most striking feature is that the number of individuals (gender count) for each of the ten departments is almost the same, falling within a very narrow range.The counts range from a **minimum of 226 (Pediatrics)** to a **maximum of 286 (General)**.
2. The total percentage for each department hovers closely **around 10%.**
3. **Largest Department**: The General department has the highest count of **individuals at 286**, which represents **11.5% of the total count**.
4. **Smallest Department:** The Pediatrics department has the lowest count of **individuals at 226**, representing **9.08% of the total coun**t.
5. **Highly Uniformity (Low Variance)**: There is no single department that heavily dominates the total count.
6. The difference between the **largest (General, 286)** and **smallest (Pediatrics, 226)** department is only 60 individuals, which, in the context of the total count (**approximately 2,500**), indicates high uniformity.

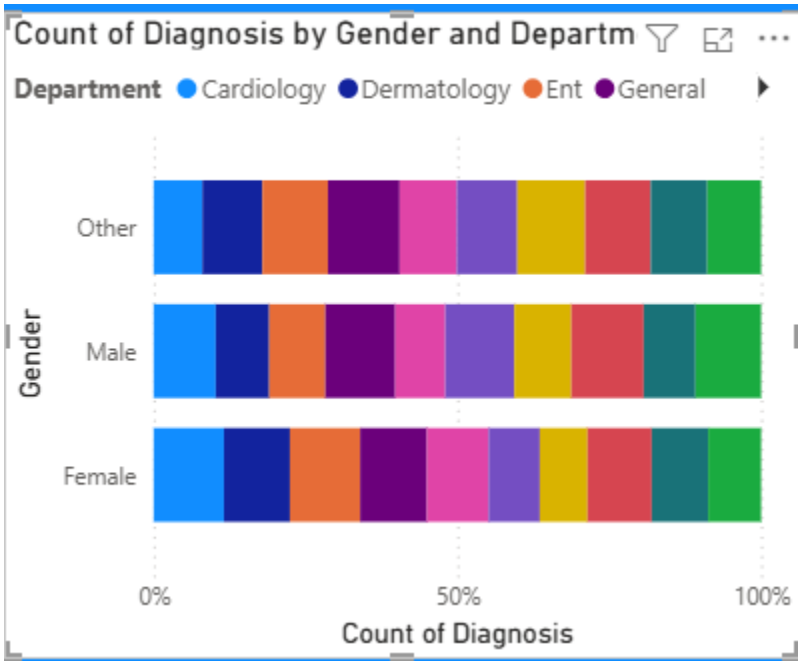


3. 100% stacked bar chart:

Observations:

Uniform Department Distribution Across Genders: The most significant observation is the nearly identical percentage distribution of diagnoses across the departments for all three gender groups (**Female, Male, and Other**).Even Spread: In all three horizontal bars (representing the three genders), each colored segment (representing a Department)

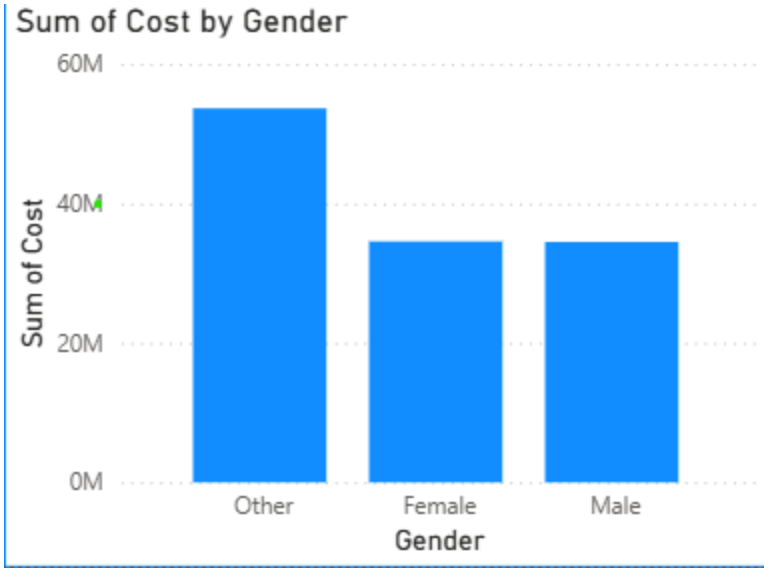
appears to take up a very similar, **small proportion of the total 100%**. This indicates that:
There is no significant department preference for any gender group based on the count of diagnoses.For example, Cardiology (the light blue segment) makes up roughly the same percentage of total diagnoses for Females, Males, and the Other group.



4. Clustered Column chart:

Observations:

- "Other" Gender Incurs the Highest Cost: The **"Other"** gender category has the highest total cost, significantly exceeding the costs for Female and Male categories.The sum of cost for the **"Other" group is approximately 55M** (55 million).
- Female and Male Costs are Nearly Identical: The total costs for the Female and Male categories are almost the same.Both **Female and Male costs are approximately 35M** (35 million).
- Large Cost Disparity: **The "Other" group's cost (approx. 55M)** is approximately 1.57 times the cost of the **Female or Male groups (approx. 35M each).**



5. Table chart:

Observations:

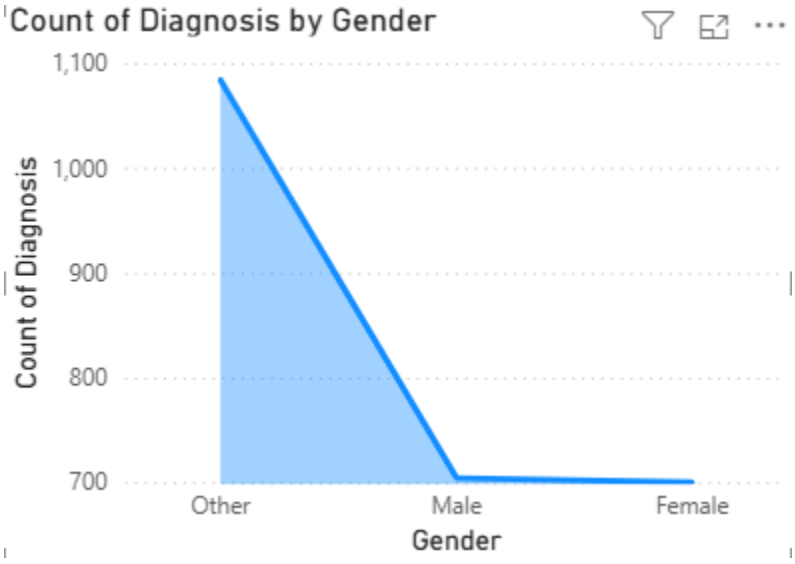
- **Count of Treatments/Visits (Volume)"Other" Gender has Highest Volume:** The "Other" gender group accounts for the largest number of treatments visits with **1,084.Male** and Female Volume is Similar: The volume for **Male (704) and Female (700)** is almost the same, indicating balanced usage between these two groups.Total Volume: The dataset represents a total of 2,488 treatments visits. Sum of Cost (Total Expenditure)
- **"Other" Gender has Highest Total Cost:** This group incurs the highest total expenditure at 5,36,58,979.84.Male and Female Total Costs are Similar: The total costs for **Male (3,44,97,049.16)** and **Female (3,45,66,716.42)** are almost identical, consistent with the treatment volume.Total Expenditure: **The overall expenditure is 12,27,22,745.42.**

Gender	Sum of Cost	Count of Department
Other	5,36,58,979.84	1084
Female	3,45,66,716.42	700
Male	3,44,97,049.16	704
Total	**12,27,22,745.42**	**2488**

6. Table chart:

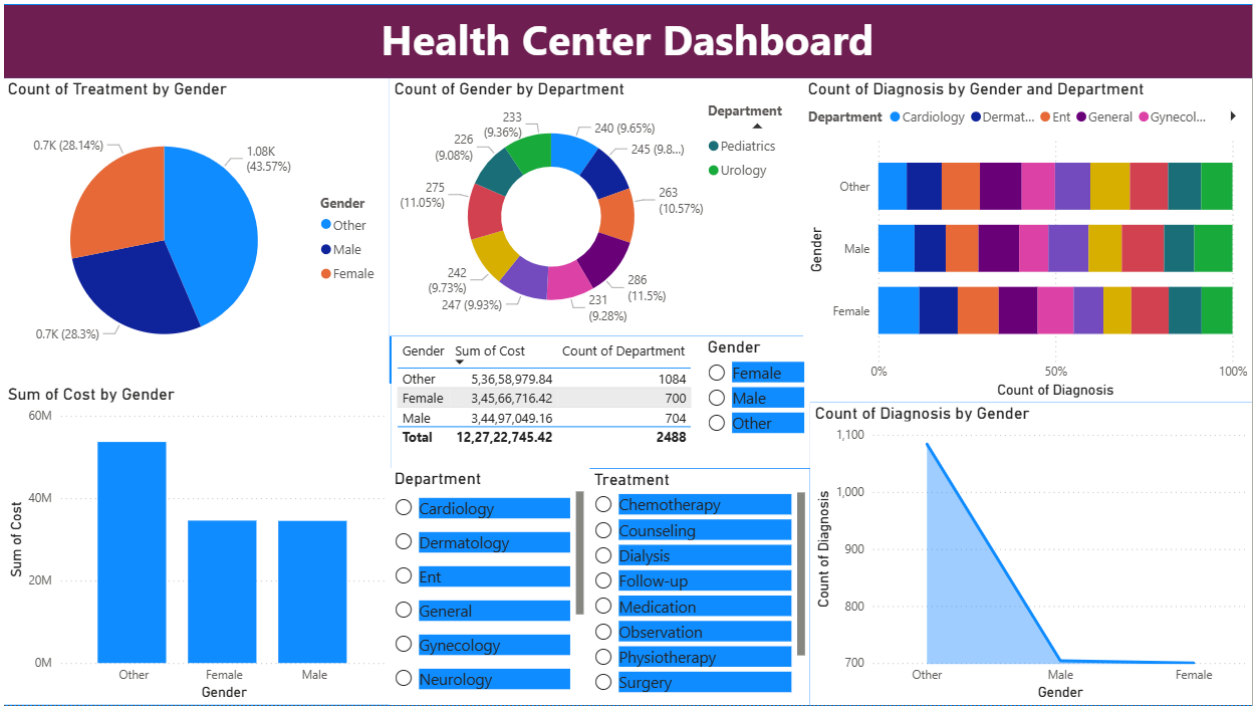
Observations:

- **"Other" Gender Dominates Diagnosis Count:** The "Other" gender category has a significantly higher count of diagnoses, peaking at approximately 1,080 to 1,090.
- **Male and Female Counts are Very Low and Similar:** The counts for Male and Female are dramatically lower than the "Other" category and are almost identical to each other, both hovering exactly at the 700 mark.
- **Large Disparity**: The number of diagnoses for the "Other" group is approximately **1.5 times (or 1084 / 700 = 1.55)** the count of diagnoses for either the Male or Female groups.
- **Visual Representation**: The chart visually emphasizes this disparity, showing a sharp drop-off in diagnosis count from the "Other" group to the Male and Female groups, which appear as a flat, low baseline.



Overall Observation:

The analysis reveals two major, contrasting themes: a significant disparity in volume and cost related to the "Other" gender group and a high degree of uniformity in departmental distribution and average treatment cost across all genders.



Conclusion:

In summary, the data suggests that while the volume of service is heavily concentrated in the "Other" gender category, the nature and cost of the services delivered are standardized across all three groups, as demonstrated by the consistent average cost per treatment and the even spread of diagnoses across departments.