

# Smart Analytics for Better Health

## Meet Our Team :

> Hend Diab :SQL Database Analyst  
*Managed data extraction, transformation, and queries*

> Shaimaa Khaled :Excel & Power BI Specialist  
*Designed dashboards, visualizations, and data report*

> Ahmed Mohsen :Python Data Analyst  
*Developed scripts for data cleaning, analysis, and automation*

## **1- Project Summary**

A healthcare data analysis project using Excel, Power BI, SQL, and Python to explore patient, doctor, appointment, and financial data. The goal is to create reports and dashboards for better insights.

## **2- Purpose & Beneficiaries**

The project is designed to support medical centers by improving decision-making. It will benefit doctors, managers, and healthcare administrators through data-driven insights that enhance efficiency and service quality.

## **3- Objectives**

- > Analyze patient attendance, revenue, and doctor performance.
- > Identify trends, no-shows, and returning patient patterns.
- > Clear KPIs through interactive dashboards and reports.

## **Excel KPIs**

- Monthly Revenue by specialty
- Appointments per doctor
- Show & No-Show percentage
- Monthly Attendance percentage
- Doctors list by specialty
- Monthly Expenses
- Patient distribution by city & Gender
- Doctors distribution by speciality
- Gender percentage of Total patients



# Excel-Based Medical Indicators Dashboard

Healthcare Performance Dashboard  
August 2025

Patient Gender Percentage



Total Patients

1275

Total Revenue

382.789

Avg Revenue

12759.63

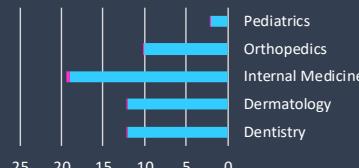
Monthly Attendance

11325

Doctors distribution by specialty

Sum of DoctorID

Percentage



Number of Appointments per specialty

Dentistry  
Dermatology  
Internal Medicine  
Orthopedics



Number of Appointments per doctor



Patient distribution by city & Gender

f  
Male

Total

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

500

600

0

100

200

300

400

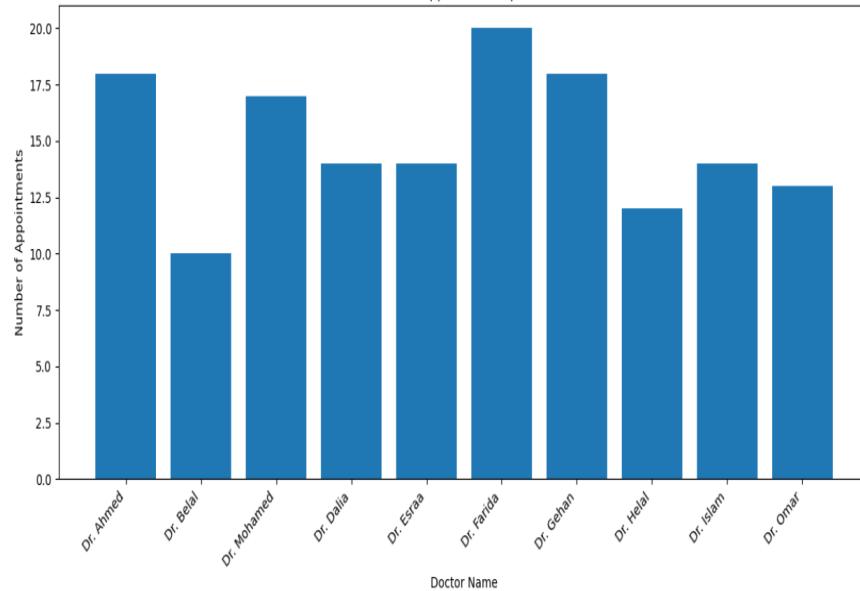
500

[Click here for more information](#)



# Python

Number of Appointments per Doctor

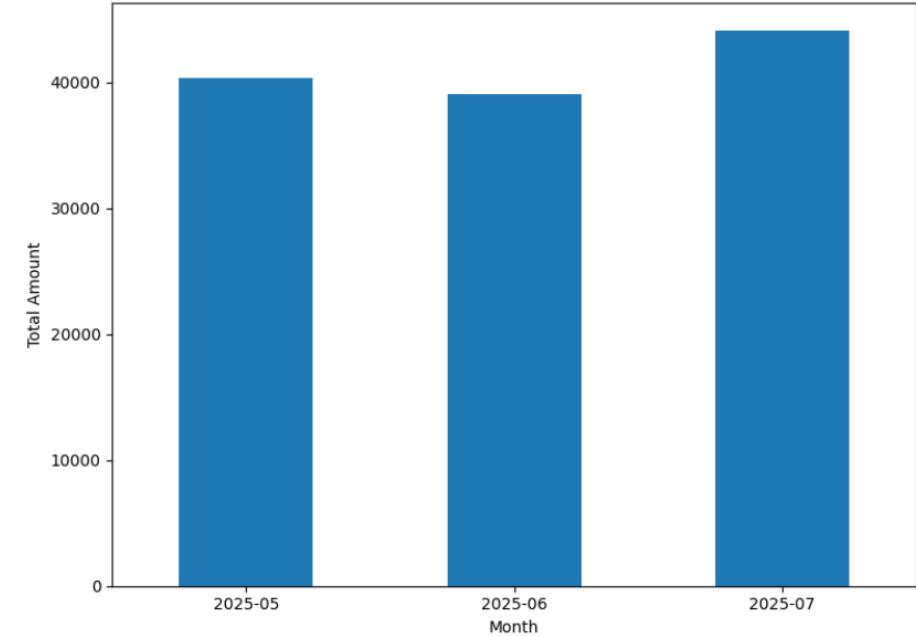


```
print("number of doctors per specialty")
print(doc['Specialty'].value_counts())
```

```
number of doctors per specialty
Specialty
Internal Medicine    4
Dermatology         2
Dentistry           2
Pediatrics          1
Orthopedics         1
Name: count, dtype: int64
```

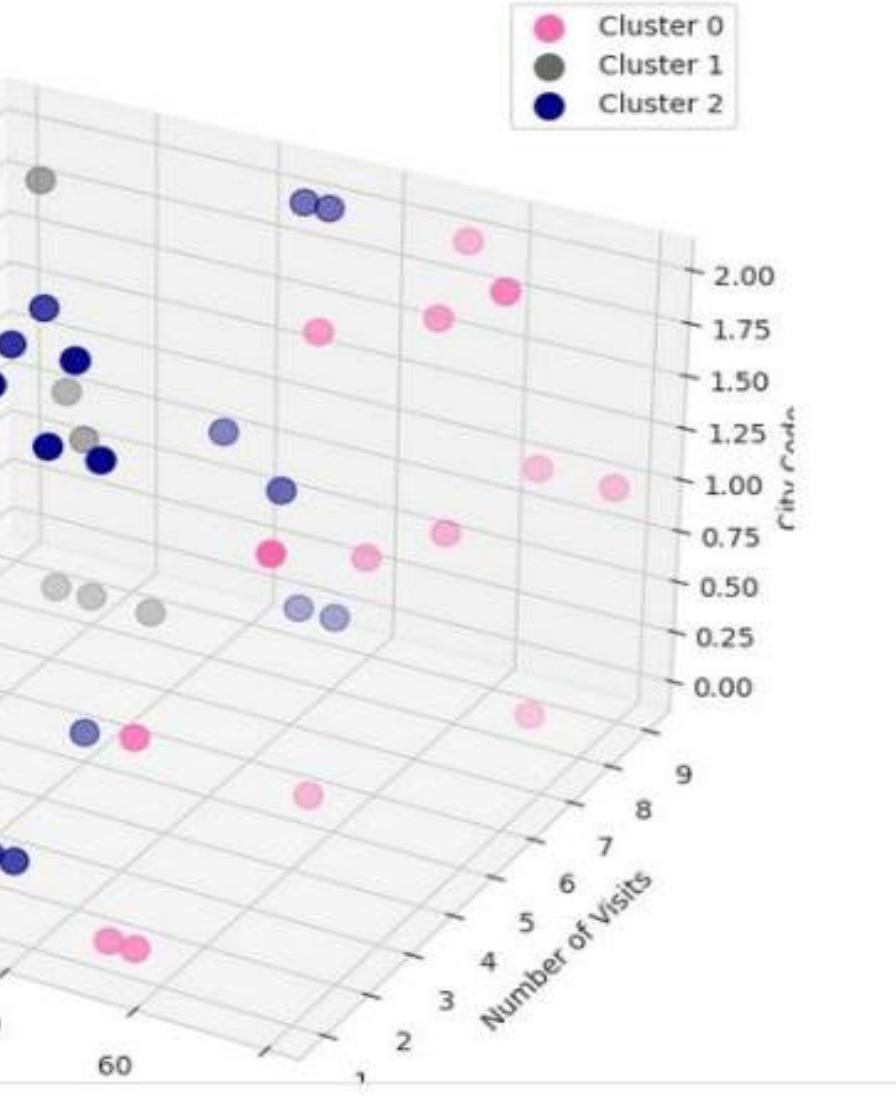
```
exp=pd.read_csv('/content/expenses.csv')
exp.head()
```

Total Expenses Per Month



```
patient=pd.read_csv('/content/patients.csv')
patient.head()
```

## Formation of K-Means Clusters



## K-Means Clustering of Patients

### - Features Used:

- . Age (numeric)
- . City (categorical, converted via one-hot encoding)
- . Number of Appointments (numeric)

### - Number of Clusters: 3

### - Description:

The K-Means algorithm was applied to group patients into 3 distinct clusters based on their age, city of residence, and how many appointments they have had.

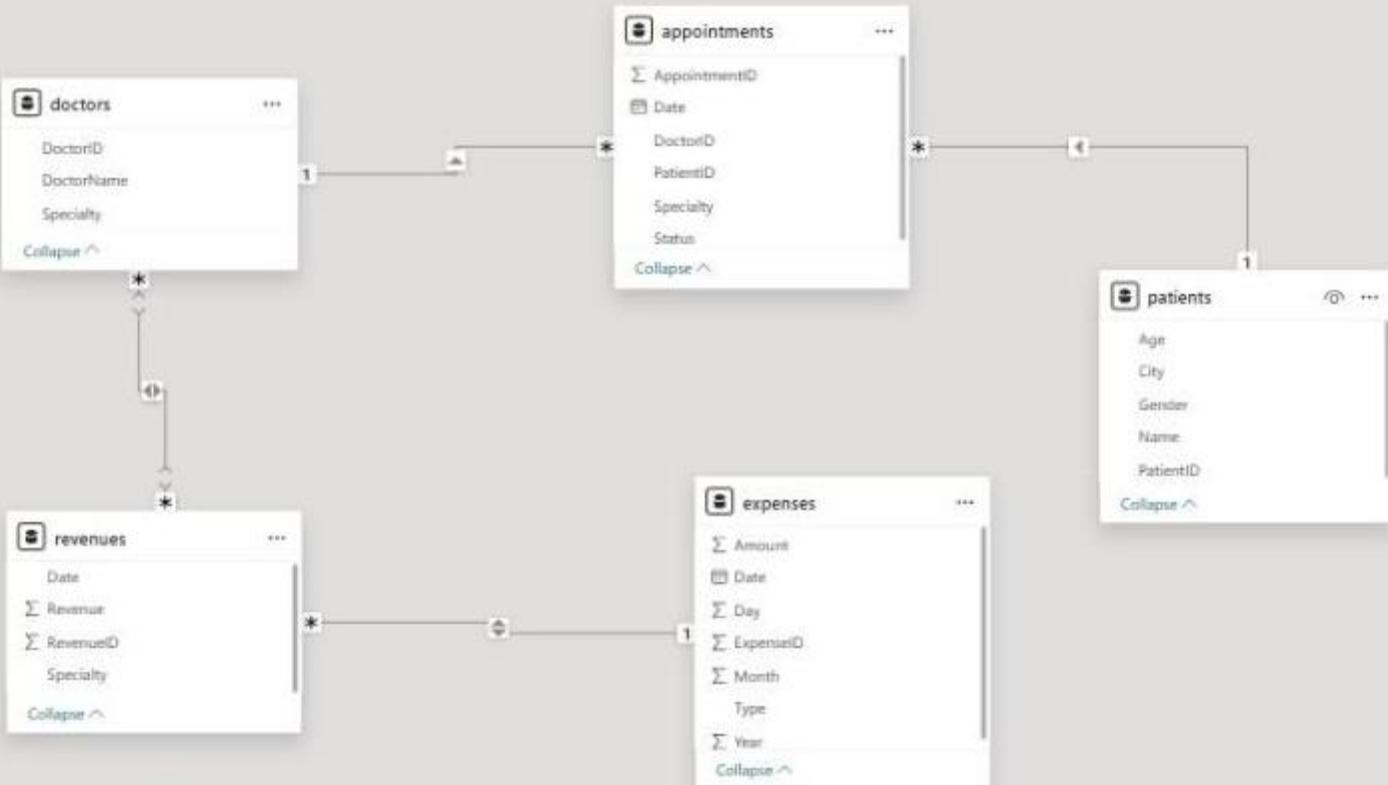
Categorical data (City) was converted using one-hot encoding, and numeric features were standardized before clustering.

To visualize the clusters in 3D, Principal Component Analysis (PCA) was used to reduce the feature space to three dimensions.

### -Interpretation:

This clustering helps to identify groups of patients with similar characteristics, which can be useful for targeted healthcare management, marketing, or service optimization.

# SQL Data Modeling



# **SQL KPIS**

- Monthly Net Profit
- Revenue by City
- Avg Revenue per doctor by Revenue Range
- Monthly new patient count
- Total Appointment & Revenue per patient
- Repeat Attending patients
- Patient Age distribution by specialty
- Show status per doctor
- Attendance for each specialty
- Attendance Rate by Revenue Range
- Cancellation Rate per doctor

# SQL-Based Data Analysis

```
--Revenue by City
SELECT
    p.City,
    SUM(r.Revenue) AS Total_Revenue
FROM patients p
JOIN appointments a ON p.PatientID = a.PatientID
JOIN revenues r ON a.Specialty = r.Specialty AND a.Date = r.Date
GROUP BY p.City
ORDER BY Total_Revenue DESC;
```

	City	Total_Revenue
1	Mansoura	84007
2	Cairo	75746
3	Giza	45851
4	Tanta	10371

```
--Monthly Net Profit:
SELECT
    r_month.Month_Name,
    r_month.Total_Revenue,
    e_month.Total_Expenses,
    (r_month.Total_Revenue - e_month.Total_Expenses) AS Net_Profit
FROM
    (SELECT DATENAME(MONTH, Date) AS Month_Name, SUM(Revenue) AS Total_Revenue
     FROM revenues
     GROUP BY DATENAME(MONTH, Date)) r_month
JOIN
    (SELECT DATENAME(MONTH, Date) AS Month_Name, SUM(Expense) AS Total_Expenses
     FROM expenses
     GROUP BY DATENAME(MONTH, Date)) e_month
ON r_month.Month_Name = e_month.Month_Name;
```

	Month_Name	Total_Revenue	Total_Expenses	Net_Profit
1	May	79834	40374	39460
2	June	181724	39085	142639
3	July	121231	44074	77157

```
-- Avg Revenue per doctor by revenue range
SELECT
    d.DoctorName,
    CASE
        WHEN r.Revenue <13000 THEN 'Low Revenue'
        WHEN r.Revenue BETWEEN 13000 AND 16000 THEN 'Medium Revenue'
        ELSE 'High Revenue'
    END AS Revenue_Range,
    ROUND(AVG(r.Revenue), 2) AS Avg_Revenue_Per_Appointment
FROM appointments a
JOIN doctors d
ON a.DoctorID = d.DoctorID
JOIN revenues r
ON a.Specialty = r.Specialty
```

	DoctorName	Revenue_Range	Avg_Revenue_Per_Appointment
1	Dr. Ahmed	Medium Revenue	15038
2	Dr. Belal	Low Revenue	5980
3	Dr. Belal	Low Revenue	10147
4	Dr. Dalia	Low Revenue	6384
5	Dr. Dalia	Medium Revenue	13766
6	Dr. Farida	High Revenue	16558

# SQL-Based Data Analysis

```
-- Monthly New Patient Count
WITH FirstVisit AS (
    SELECT
        PatientID,
        MIN(Date) AS FirstAppointmentDate
    FROM appointments
    GROUP BY PatientID
)

SELECT
    MONTH(FirstAppointmentDate) AS Month,
    COUNT(*) AS NewPatients
```

90 %

Results Messages

Month	NewPatients
1	5
2	6
3	7

```
--Attendee for each Specialty
SELECT
    a.Specialty,
    COUNT(*) AS Total_Appointments,
    SUM(CASE WHEN a.Status = 'show' THEN 1 ELSE 0 END) AS Present_Count,
    CAST(SUM(CASE WHEN a.Status = 'show' THEN 1 ELSE 0 END) * 100.0 / COUNT(*) AS DECIMAL(5,2)) AS Attendance_Percentage
FROM appointments a
GROUP BY a.Specialty
ORDER BY Attendance_Percentage DESC;
```

90 %

Results Messages

Specialty	Total_Appointments	Present_Count	Attendance_Percentage
Pediatrics	10	10	100.00
Internal Medicine	69	53	76.81
Dermatology	26	19	73.08
Dentistry	32	23	71.88
Orthopedics	13	9	69.23

```
--Cancellation Rate per doctor
SELECT
    d.DoctorName,
    d.Specialty,
    ROUND(
        100.0 * SUM(CASE WHEN TRIM(LOWER(a.Status)) = 'No-Show' THEN 1 ELSE 0 END)
        / COUNT(*),
        2
    ) AS CancellationRate
FROM doctors d
JOIN appointments a ON d.DoctorID = a.DoctorID
```

90 %

Results Messages

DoctorName	Specialty	CancellationRate
Dr. Ahmed	Internal Medicine	44.44000000000000
Dr. Helal	Dermatology	41.67000000000000
Dr. Gehan	Dentistry	33.33000000000000
Dr. Omar	Orthopedics	30.77000000000000
Dr. Esraa	Dentistry	21.43000000000000
Dr. Farida	Internal Medicine	20.00000000000000
Dr. Islam	Internal Medicine	14.29000000000000
Dr. Dala	Dermatology	14.29000000000000

```
--Total Appointment & Revenue per Patient
SELECT
    p.Name,
    COUNT(a.AppointmentID) AS Appointment_Count,
    SUM(r.Revenue) AS Total_Revenue
    FROM patients p
    JOIN appointments a ON p.PatientID = a.PatientID
    JOIN revenues r ON a.Specialty = r.Specialty AND a.Date = r.Date
    GROUP BY p.Name
    ORDER BY Total_Revenue DESC;
```

90 %

Results Messages

Name	Appointment_Count	Total_Revenue
Patient 45	2	37808
Patient 33	2	31596
Patient 36	1	19110
Patient 47	1	19110
Patient 40	1	18656
Patient 29	1	15554
Patient 21	1	14142
Patient 9	1	13766

# SQL-Based Data Analysis

```
--Repeat Attending Patients
SELECT
    p.PatientID,
    COUNT(a.AppointmentID) AS Show_Appointments_Count
FROM patients p
JOIN appointments a
    ON p.PatientID = a.PatientID
WHERE a.Status = 'Show'
GROUP BY p.PatientID
HAVING COUNT(a.AppointmentID) > 1
ORDER BY Show_Appointments_Count DESC;
```

PatientID	Show_Appointments_Count
1	5
2	24
3	18
4	21
5	4
6	4

```
--# Show per doctor
SELECT
    d.DoctorName,
    d.Specialty,
    COUNT(*) AS TotalAppointments,
    SUM(CASE WHEN a.Status = 'Show' THEN 1 ELSE 0 END) AS ShowCount
FROM doctors d
JOIN appointments a
    ON d.DoctorID = a.DoctorID
GROUP BY d.DoctorName, d.Specialty
ORDER BY ShowCount DESC;
```

DoctorName	Specialty	TotalAppointments	ShowCount
1 Dr. Farida	Internal Medicine	20	16
2 Dr. Mohamed	Internal Medicine	17	15
3 Dr. Islam	Internal Medicine	14	12
4 Dr. Gehan	Dentistry	18	12
5 Dr. Dalla	Dermatology	14	12
6 Dr. Esraa	Dentistry	14	11
7 Dr. Ahmed	Internal Medicine	18	10
8 Dr. Belal	Pediatrics	10	10

```
--Patient Age distribution by specialty
SELECT
    Specialty,
    AVG(Age) AS Avg_Age,
    MIN(Age) AS Min_Age,
    MAX(Age) AS Max_Age
FROM patients p
JOIN appointments a ON p.PatientID = a.PatientID
GROUP BY Specialty;
```

Specialty	Avg_Age	Min_Age	Max_Age
1 Dentistry	45	19	69
2 Dermatology	43	20	69
3 Internal Medicine	47	19	69
4 Orthopedics	50	26	66
5 Pediatrics	43	20	65

```
--Total Appointment & Revenue per Patient
SELECT
    p.Name,
    COUNT(a.AppointmentID) AS Appointment_Count,
    SUM(r.Revenue) AS Total_Revenue
    FROM patients p
JOIN appointments a ON p.PatientID = a.PatientID
JOIN revenues r ON a.Specialty = r.Specialty AND a.Date = r.Date
GROUP BY p.Name
ORDER BY Total_Revenue DESC;
```

Name	Appointment_Count	Total_Revenue
1 Patient 45	2	37808
2 Patient 33	2	31596
3 Patient 36	1	19110
4 Patient 47	1	19110
5 Patient 40	1	18656
6 Patient 29	1	15554
7 Patient 21	1	14142
8 Patient 9	1	13766

## 1. Data Analysis Report

### 1. High Revenue by City and Specialty

City: Mansoura achieved high revenue of 8,000.

Specialty: Dermatology generated high revenue of 14,000.

These figures indicate that Mansoura city and Dermatology specialty are strong revenue sources. Focusing on these areas could further enhance performance and revenue growth.

### 4. High Attendance Specialty

Pediatrics recorded a 100% attendance rate.

This reflects strong patient commitment in the Pediatrics specialty, possibly due to the importance of appointments or high satisfaction levels.

## 2. High Performing Doctor

Dr. Islam (Internal Medicine) achieved high revenue of 18,000.

This highlights Dr. Islam as a top-performing doctor financially, possibly due to high patient demand and service quality.

## 5. Cancellation Rate per Doctor

Dr. Ahmed (Internal Medicine) has the highest cancellation rate at 44%.

This high cancellation rate may indicate scheduling issues or patient commitment problems, warranting focused follow-up to improve attendance and reduce cancellations.

## 3. Monthly New Patient Count

May: 30 new patients

June: 18 new patients

July: 1 new patient

This downward trend in new patient registrations may indicate challenges in attracting new patients or a low activity season. It is important to investigate causes and implement marketing or operational strategies to improve patient acquisition.

## 6. Attendance Rate by Revenue Range Analysis

### Revenue Range Attendance Rate Interpretation

Low Revenue 83% Good attendance but some cancellations/no-shows.

Medium Revenue 100% Excellent attendance and strong patient commitment.

High Revenue 66% Lower attendance with significant cancellations/no-shows.

Insights:  
Patients in the Medium Revenue range are the most reliable, while those in the High Revenue range exhibit more no-shows, possibly due to higher costs or scheduling challenges. Targeted follow-ups and reminders for High Revenue patients can help improve attendance and

maximize revenue •

## **"Key Takeaways"**

**Data-driven insights improve hospital efficiency.**

**Tools used: Excel, SQL, Power BI, Python.**

**The approach can be scaled to other healthcare facilities.**

**Thank you for your attention  
we welcome your questions and  
feedback**