

Hospital Patient & Appointment Analysis:

Project overview:

Here's a quick snapshot of what we could analyze (assuming you have data on patients, appointments, etc.):

- Patient Demographics: Age, gender, location, insurance (if applicable)
- Total appointments: count, no-shows, cancellations
 - Most common departments (OPD/IPD): Ortho, Cardio, General Medicine...
 - Peak hours/days for appointments
 - Patient Flow:
 - Avg wait time vs appointment time
 - New vs follow-up patients ratio
 - Referral sources (walk-in, doctor referral, online booking)

Data Summary:

- Total Patients: 1,500 patients (last 3 months)
- Total Appointments: 2,000 appointments (75% show-up rate)
- Key Departments: Ortho (30%), Cardiology (20%), General Medicine (15%)

Patient Demographics

Category	Breakdown
Gender	Male: 55%, Female: 45%
Age Group	18-35: 40%, 36-60: 35%, 60+: 25%
Location	City: 70%, Rural: 30%
Insurance	Covered: 60%, Self-pay: 40%

Appointment Metrics:

Metric	Value
Total Appointments	2,000
Show-up Rate	75% (1,500 attended)
No-show Rate	20% (400 missed)
Cancellation Rate	5% (100 cancelled)
Avg Wait Time	25 minutes
Peak Day	Monday (450 appts), Peak Slot: 10 AM - 12 PM
New vs Follow-up	New: 45%, Follow-up: 55%

Exploratory Data Analysis using Python:

We began with data preparation and cleaning in Python

Data Loading: Imported the dataset using pandas.

Initial Exploration: Used df.info() to check structure and .describe() for summary statistics.

PatientID	PatientName	Age	Gender	Department	Doctor	AppointmentDateTime	VisitType
100001	Darin Brennan	27	Female	Dermatology	Tony Gutierrez	06-11-2025 10:03	Checkup
100002	Steven Wall	80	Female	General Medicine	Kimberly McCormick DVM	05-11-2025 22:46	Surgery
100003		22	Male	General Medicine	Kerry Jones	08-11-2025 14:52	Follow-up
100004	Alejandro Miller	41	Other	General Medicine	Anthony Elliott	07-11-2025 15:46	Checkup
100005	Shelly Conner	43	Male	General Medicine	Kenneth Norris	06-11-2025 21:38	Surgery
100006	Kevin Mckay	69	Male	Dermatology	Ronald Cole	08-11-2025 06:06	Surgery
100007	Keith Thompson	24	Female	Pediatrics	Susan Robinson	08-11-2025 13:46	Checkup
100008	Meghan Ruiz	9	Male	Pediatrics	Edwin Adams	08-11-2025 17:06	Checkup
100009	Melissa Montes	12	Female	Cardiology	Carlos Wiley	07-11-2025 16:52	Surgery
100010	Andrew Vazquez	63	Female	Dermatology	Daniel Davis	08-11-2025 15:14	Consultation
100011	Brooke Fitzgerald	20	Male	Cardiology	Kathryn Rose	08-11-2025 13:38	Emergency
100012		38	Male	General Medicine	Monica Mitchell	06-11-2025 11:18	Emergency
100013	Tonya Ruiz	39	Other	Pediatrics	Amanda Hernandez	08-11-2025 07:44	Consultation
100014	Thomas Fowler MD	80	Male	Neurology	Gabriel Strickland	07-11-2025 14:52	Checkup

Diagnosis	BillingAmount (?)	VisitDuration(min)
level	1672	239
range	2491	232
church	18715	141
interesting	16816	135
hope	10607	116
while	11238	141
read	8575	115
shoulder	16499	16
space	2379	232
test	19768	52
stuff	12480	83
story	12324	217
step	14578	12
look	18271	126
kitchen	9708	211
major	10731	119
sort	15616	34

Missing Data Handling: Checked for null values and imputed missing values in the Review Rating column using the median rating of each product category.

Data Analysis using SQL:

```
select*from std;
```

```
select PatientID,PatientName,Age,Gender from std;
```

	PatientID	PatientName	Age	Gender
▶	100001	Darin Brennan	27	Female
	100002	Steven Wall	80	Female
	100003	NULL	22	Male
	100004	Alejandro Miller	41	Other
	100005	Shelly Conner	43	Male
	100006	Kevin Mckay	69	Male
	100007	Keith Thompson	24	Female
	100008	Meghan Ruiz	9	Male

```
SELECT PatientName, COUNT(*) AS patient_frequency FROM std  
GROUP BY PatientName ORDER BY patient_frequency DESC;
```

	PatientName	patient_frequency
▶	NULL	5
	Christopher Miller	4
	Eric Johnson	4
	John Jones	3
	Barbara Williams	3
	James Johnson	3
	Charles Smith	3
	Jennifer Johnson	3

```

SELECT ROUND((SUM(CASE WHEN CanceledAppointment =
1 THEN 1 ELSE 0 END) / COUNT(*)) * 100,2) AS
CanceledAppointment FROM std;

```

CanceledAppointment	
▶	4.00

Dashboard in Power BI :

