Healthcare Data Analytics Project

27th June 2025- 4th July 2025

Content

Summary

This project is all about helping hospitals work better by using data. We looked at 10,000+ hospital appointment records from Jan–Feb 2025 to understand patient wait times — like which departments and doctors have longer waits, and when delays usually happen. After cleaning the data and doing some analysis, we found that most appointments happen on weekdays, especially midweek. Wait times are higher in the late morning and afternoon. The Neurology department and doctor DR102 had the longest wait times, while departments like Dermatology and Pediatrics handled many patients smoothly. We created interactive dashboards using Power BI to clearly show these patterns and help the hospital make smart decisions. We also suggested easy-to-follow solutions like better scheduling, using AI tools, real-time queue tracking, and learning from efficient departments. Even though the data had some gaps (like no info on missed appointments or patient feedback), this project shows how data can be used to improve patient care, reduce stress for staff, and make hospital visits smoother for everyone.

Project overview

In today's healthcare system, operational efficiency plays a critical role in ensuring timely delivery of high quality patient care. One of the most important challenges healthcare centers face is managing patient waiting time, which directly impacts satisfaction, revenue, reputation , clinical outcome and productivity of staff. This project focuses on analyzing hospital data for understanding patterns related to wait time related to doctor wise and department wise wait duration. With the help of healthcare analytics we can uncover patterns, develop interactive dashboards and generate actionable insights which help hospital administration in optimization in appointments and improve overall operation of hospitals. This project aligns with the growing need for data-driven decision-making in modern healthcare environments.

Problem Statement

In many hospitals, Long patient wait times are a common challenge. They not only affect patient satisfaction but also lead to missed appointments, stressed staff, and reduced overall efficiency. These delays often result from poor appointment scheduling, uneven distribution of doctor workload, and limited visibility into day-to-day operations. Even though hospitals collect large volumes of data, most lack the tools or processes to turn that data into useful insights. As a result, valuable opportunities to optimize patient flow and resource utilization are often missed.

This project aims to analyze hospital appointment and wait time data to identify underlying patterns, highlight inefficiencies, and offer data-driven solutions that support better scheduling and smarter hospital management.

Potential Impact

If these issues remain unaddressed for a long time, they can lead to several negative consequences on both patients and hospital operations.. Some key impacts are mentioned below:-

1. Lower patient satisfaction, due to long wait time and poor service.
2. Missed appointments.
3. Loss of revenue to hospitals.
4. Damage to Hospital reputation due to negative comments.
5. Lower productivity of staff especially doctors and nurses because of burnout.
6. Patients will opt for competitors.
7. It will worsen a patient's condition due to late diagnosis.
8. Resource misallocation.

Dataset Description

The dataset used in this project consists of hospital appointment records from a period of January 2025 to February 2025. It contains 10000 rows and 7 columns, capturing various aspects of patient visits and hospital operations.

- Columns
    - Patient_ID : Unique identifier of each patient
    - Patient_name : name of patient
    - Appointment_date : scheduled date of the appointment
    - Arrival_time : Patient arrived at clinic
    - Consultation_start_time : time when doctor consultation begins.
    - Department : department visited (e.g., Cardiology, Neurology)
    - Doctor_Id : attending doctor

Patientt_id	Patient_name	appointment_date	Arrival_time	Consultation_start_time	Department	Doctor_id
P100000	Anay Shankar	10-02-2025	11:57	12:16	cardiology	DR100
P100001	Nirvan	09-01-2025	09:08	09:40	cardiology	DR108
P100002	Sara beth	03-01- 2025	10:57	11:34	cardiology	DR101
P100003	Kiara kakar	08-02-2025	11:12	11:58	cardiology	DR108

Table: Dataset Sample

This data includes a mix of **categorical, datetime, and numerical variables**, making it suitable for detailed descriptive analysis and dashboard creation.

Objectives

The primary objectives of this project are:

- **Identify key issues:** Utilize the dataset to pinpoint areas of poor performance, such as excessive wait times or scheduling bottlenecks.
- **Analyze Root Causes**: Employ data analytics tools (SQL, Excel, Power BI) to uncover patterns, such as peak wait times or departments with consistent delays.
- **Propose Solutions**: Develop recommendations to optimize appointment scheduling, resource allocation, or process improvements.
- **Demonstrate Impact**: Create dashboards and KPI cards to compare “before” and “after” metrics, showcasing the effectiveness of proposed solutions.
- **Build a Portfolio-Worthy Project:** Develop a comprehensive case study that demonstrates problem-solving skills and real-world applicability, appealing to recruiters in the healthcare analytics field.

Questions to be Answered

1. What is the average, median, and maximum wait time across all appointments?
2. Which departments have the highest and lowest average wait times?
3. Which doctor has the highest and lowest average wait times?
4. Do certain doctors consistently have longer wait times compared to others?
5. At what times of the day or days of the week or day of month are wait times generally higher?
6. Are some departments overbooked while others are underutilized.
7. Do doctors with a higher volume of appointments also have a higher wait time?

Methodology

1. **Problem Identification, potential Impact identification and dataset observation**:
    a. In this project, we aimed to identify **patterns in patient wait-time variation** with respect to different **departments and doctors** using appointment data. Our goal was to understand how these inconsistencies in wait times could lead to **negative consequences for the hospital.**
    b. Understanding the possible negative consequences of wait-time variations, It helped shape the direction of our analysis. It led us to ask key questions such as average wait time, mean wait time etc.

c. The dataset observation phase helped us **understand the overall structure and quality of the data**, and informed our strategy for cleaning and analysis. Specifically, it allowed us to:
    i. Identify the **key columns of interest.**
    ii. Understand the **data types**, formats, and relationships between columns.
    iii. Detects issues such as **missing values**, **inconsistent formats**, or **duplicate entries.**
    iv. Clarify whether any **assumptions** were needed, such as standard appointment durations.
    v. Determine what kind of **data wrangling** is necessary.

2. **Descriptive analysis stage**:
    a. This stage helped me to uncover **baseline trends and distributions** within the appointment dataset. We can identify KPIs such as Average wait time, department wise Average wait time, highest and lowest wait time department wise , doctor wise, appointments trends monthly, weekly and hourly basis.
    b. These descriptive insights provided a **clear understanding of operational bottlenecks** and formed the foundation for deeper, targeted analysis in the later stages.
3. **Dashboard Generation stage:**
    a. After data Wrangling and Descriptive analysis, an interactive dashboard was developed using Power BI tool to visualize key metrics and patterns effectively.
    b. Key components of the dashboard included:
        i. Appointment trends
        ii. Average wait time
        iii. Doctor wise and department wise appointment distribution
        iv. Time series analysis of appointments.
    c. The dashboard enabled stakeholders (e.g., hospital management, scheduling staff) to quickly identify **inefficiencies**, **overburdened departments**, and **opportunities for scheduling improvements.**

Insights

1. **Dashboard Page 1: Overview**

a. There are **5 departments** and **10 doctors** in the hospital.
b. The average waiting time is **~ 31.46 mins**.
c. Overall appointments are **10k, out of which around 76% are scheduled for weekdays and 24 % scheduled for weekends**, which indicates high operational workload for doctors and other hospital staff on weekdays. Hospital administration must ensure adequate staffing, time slot distribution and other resources during this peak time.
d. During weekdays, **Wednesday(1595) , Thursday(1565) and Friday(1553)** have high appointments as compared to other weekdays. While during the weekend **Saturday(1284)** witnessed lowest appointments. Low weekend appointment may be due to doctor availability or patient preference
e. Due to high appointments in mid week, hospital need to do better resource management(slot booking or staff rotation.
f. Appointments distribution among the department is relatively balanced:
    i. **Neurology(20.31%),**
    ii. **Dermatology (20.25%),**
    iii. **Orthopedics (19.42%),**
    iv. **Cardiology (19.83%),**
    v. **Pediatrics (20.19%)**
g. Average daily count of appointments revolves around **330~340** . It indicates the normal flow of patients in the hospital indicating **normal operational behavior** without major fluctuations.

2. **Dashboard page 2: Departmental view insights**

a. Neurology department **Faces High appointment Load(2031) with Longer Wait Times(31 min).** There is a higher probability of potential **overload of patients**, complex medical conditions or **insufficient consultation capacity**, which may lead to patient dissatisfaction if not addressed.

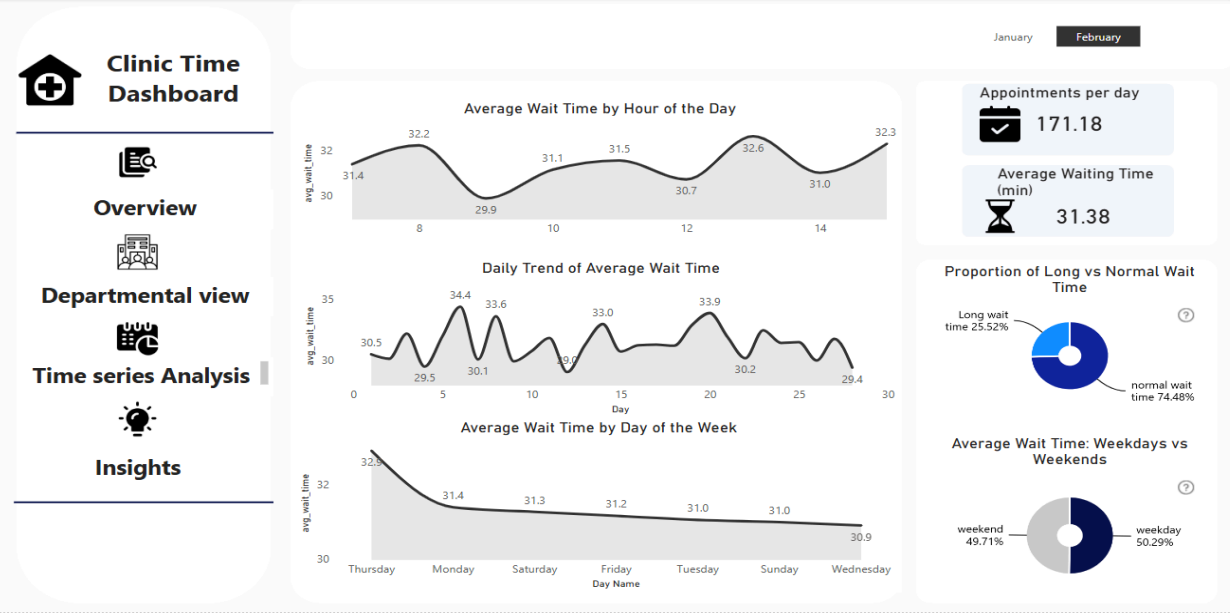
b. **Dermatology(31.20 min, 2025) & Pediatrics(31.31 min ,2019)** has Low average Wait Time with High Patient Volume. It shows both departments handled over 2000 appointments with low overall average wait. It may be due to **efficient scheduling**, possibly shorter consultation durations, or well-managed workflows.

c. **Cardiology (31.56 min, 1983) and orthopedics(31.47 min , 1942)** shows overall balanced performance. These departments average wait time is little over average wait time(31.46) which indicates they need minor optimization in operations.

d. Doctor **DR102** has the **highest average wait time (~32.8 min)** and relatively **high appointment volume (~1030+)**. While doctors like DR106, DR107,DR108 handled a good number of patients with lower average wait time. It indicates workflow efficiency or less severe cases or better staff coordination.

e. Wait time variation among doctors helps hospitals to redistribute appointments to doctors who have lower Average wait time and handle a low number of patients. It will improve patient flow and reduce unnecessary delay.

3. **Dashboard Page 3: Time Series Analysis**



a. **Early morning and afternoon witnessed peak average wait time**. High average wait time recorded at **7 AM (31.8 min) , 1 PM (31.8 min) and 3 PM (32.9 min)**. **Minimum average wait time is around 9 AM and 2 PM** which indicates smoother operations or low number of appointments.

b. Possible causes for late afternoon spike in average wait time are:
   i. Overlapping appointments.
   ii. Delayed start cascading into midday.
   iii. Break time.

c. There is a **high fluctuation in average wait time during a period of 30 days**. This might indicate there are some kind of operational issues such as variable appointment load, staffing issue or doctor availability. **Day 6 (33.38 min), day 8 (32.90 min), day 14(32.80 min), day 20 (33.40 min) experienced high average time and day 4 (30.7 min) , day7 (30.7 min) and day 9 (29.2 min) experienced low average wait time.**
   i. It may indicate there is lack of uniform control scheduling and execution across the month.
   ii. The hospital witnessed an average of 169 appointments per day, which indicated stable traffic. So we need to optimize the process.

d. If we talk about waiting time in terms of proportion between weekdays and weekends. **Weekday waiting time** is 50.29% and weekend waiting time is 49.71 % which shows Wait time is **almost equal on both** despite lower appointments on weekends. There may be possibility that Weekend staffing may be low or **resource allocation is inefficient or weekends** appointments were allocated to critical cases.

e. If we compare the normal waiting period (75 %) and the higher waiting period (25 %). **1 in every 4** patients is **waiting more than the expected limit**. Hospital administration needed to focus on this issue as well.

Solutions

1. **Redistribute Load (Doctor & Department Level)**:
    a. The neurology department and doctor DR102 overloaded with high average wait times.
    b. Here are some solutions which resolve this issue:
        i. **Hospital can reassign patients** to underutilized doctors.
        ii. Hospital can create a rotational **system** to share high-demand patients among multiple qualified doctors.
        iii. Administration can create **buffer slots** between appointments for DR102 and Neurology to absorb overruns.
2. **Time Slot Optimization (Hourly Analysis)**
    a. Early morning and afternoon sessions witnessed high average wait time. This tissue can be resolved by several ways:
        i. **Limit new bookings** during peak slots unless urgent.
        ii. Book appointments based on severity of patients. Prioritize severe cases in the morning and general OPD in later hours.
        iii. Allow **auto-suggestions** during booking to guide patients to off-peak slots.
3. **Weekday Load Balancing**
    a. Midweek (Wednesday–Friday) handles the highest patient load.
    b. Here are some solutions for this:
        i. **Distribute appointments** more evenly across the week through auto-suggestions.
        ii. Open **extra slots on lighter days (e.g Friday and Saturday).**
        iii. Assign backup doctors or tele-consultation options during high-load days
4. **Real-time Queue Management**
    a. High wait time affects 25 % of patients. It can be resolved through:
        i. Implement **real-time patient queue dashboards** visible to reception/admin staff.
        ii. Introduce **check-in kiosks or mobile check-ins** to reduce manual delays.
        iii. Notify patients about **expected wait time** on their phones via SMS or in-app alerts.
        iv. Automatically **flag critical delays** to floor managers.
5. **Smart Scheduling & AI Use**

    a. AI based **scheduling systems help** hospitals to operate smoothly by predicting delays, Recommend slot length based on consultation history and Adjust daily doctor distribution dynamically.
    b. Even AI can track patterns like **repeat no-shows**, **chronic delays**, and flag them.
6. **Departmental Efficiency Audits.**
    a. Some departments are more efficient (e.g., Dermatology), while others have high delays (e.g., Neurology). Here are some solutions which can bring efficiency in an over-booked department.
        i. Conduct **quarterly department-level audits**. Administration can compare patient flow time, average wait, and consultation duration, and interview staff to identify pain points.
        ii. Apply **best practices** from efficient departments (e.g., staggered shifts, digital forms).
7. **Optimize Weekend Usage**
    a. Only 24% of appointments are on weekends, yet wait times remain similar to weekdays. It shows there is some kind of problem in operations and patient handling. We can mitigate this problem by various means:
        i. Improve **staffing and consultation spread** on weekends to reduce inefficiencies.
        ii. Hospital can **Promote weekend booking** by offering:
            1. Shorter wait times
            2. Discounts for follow-ups
            3. Special clinics (e.g., child vaccination on Sunday).

Project Limitation

1. Data is limited to only two months, due to this we cannot capture long term trend or seasonal variation. For example : Festivals, holidays, or flu seasons can affect appointments, but may not be visible here.
2. Absence of Patient Demographics from the data which make this analysis less broad.
3. This dataset also lacks critical information such as no-shows, appointment cancellations, and consultation durations. The absence of these variables may introduce bias and limit the accuracy of the analysis.
4. Manual delays not tracked such as number of breaks taken by staff, system login , Administrative queue lags etc.
5. The analysis assumes similar consultation durations for all patients, while in real scenarios situations are different for example : Neurology or Orthopedics might take longer and follow-ups may be faster.

6. No Patient Feedback data can be found. The analysis focuses only on **quantitative metrics** like wait time, not qualitative experience.
7. There is no information about staffing pattern, condition of physical infrastructure.