PATIENT APPOINTMENT NO-SHOW ANALYSIS USING PYTHON EDA FOR A MULTI-CLINIC U.S. HEALTHCARE SYSTEM

1. Background

A regional U.S. healthcare network operating 18 outpatient clinics across 3 states struggled with high patient no-show rates, leading to underutilized staff time, delayed care, and revenue loss. Despite having access to EMR (Electronic Medical Records) and appointment booking logs, their internal team lacked a consolidated analytical view.

We were engaged to conduct a Python-based exploratory data analysis (EDA) on appointment data to help administrators understand the **who**, **when**, and **why** behind no-shows—and use those insights to revise scheduling policies and improve engagement.

2. Objective

- To analyze patterns in appointment attendance and identify factors correlated with no-shows
- To compare attendance rates across clinics, time slots, demographics, and communication channels
- To generate insights that inform operational changes like reminder policies and time slot distribution
- To produce visual, shareable outputs that clinical and administrative teams can interpret easily

3. Data Used

The client provided **9 months of appointment data** across 18 clinics, totaling **162,000 records** with the following key fields:

- Appointment_ID
- Clinic_Location
- Appointment_Date, Time_Slot (Morning / Afternoon / Evening)
- No_Show (1 = Missed, 0 = Attended)
- Patient_Age

- Gender
- Insurance_Type (Medicaid, Medicare, Private, Self-pay)
- Appointment_Booked_By (Phone, Online, Walk-in)
- Reminder_Sent (Yes/No)
- Days_Between_Booking_And_Appointment

Additional metadata included clinic-level staffing schedules and historical reschedule frequency.

4. Methodology

4.1 Data Preprocessing

- Cleaned and standardized Time_Slot and Insurance_Type fields
- Derived fields such as:
  - Lead_Time = days between booking and appointment
  - Age_Group (0–18, 19–35, 36–60, 60+)
  - Is_Reminded (flag based on Reminder_Sent)
- Removed incomplete entries and no-show duplicates due to same-day rebooking

4.2 Exploratory Analysis

- Univariate analysis of no-show rate by:
  - Age group
  - Time slot
  - Appointment lead time
- Bivariate and multivariate analysis:
  - No_Show rate by Insurance_Type × Clinic_Location
  - Effect of reminders vs no reminders
  - Visualization of no-show heatmaps by weekday and time

4.3 Tools Used

- pandas, numpy for transformation
- seaborn, matplotlib, plotly for visualizations

- statsmodels for basic statistical testing
- Report prepared in Harvard format with inline Python visuals

5. Key Results

- **Overall no-show rate**: 21.3%
- **Highest no-show time slots**:
  - Evening (28.4%) > Afternoon (22.6%) > Morning (15.9%)
- **Lead Time Effect**:
  - Appointments booked **more than 7 days in advance** had a **32% higher** no-show rate
- **Reminder Effect**:
  - **No reminder** = 29.1% no-show
  - **Reminder sent** = 14.8% no-show
- **Medicaid patients** had the highest no-show rate at **31%**, compared to **12%** for privately insured
- Some clinics had 2× the network average—flagged for further operational review

6. Report Output

- **PDF Report (16 pages)**:
  - Summary charts showing no-show trends by time, location, age, and payer
  - Interpretation notes for non-technical managers and clinical staff
  - Clustered bar graphs for reminder vs non-reminder impact
  - Tables summarizing top ZIP codes and demographic trends associated with missed visits
- **Python Notebook**:
  - Fully annotated with markdown cells explaining each analysis step
  - Exportable visual plots and KPI summary table functions
- **Excel Workbook**:
  - Clinic-level no-show comparison dashboard

- Suggested reschedule buffer time logic based on predicted no-show probability

7. Business Impact

- Clinics adjusted evening appointment quotas, moving more follow-up visits to morning slots
- Admin team adopted **SMS + email reminders** for appointments booked >5 days in advance
- Medicaid-heavy clinics received operational support to implement pre-call reminders
- Quarterly no-show rate dropped from **21.3% to 15.7%** after implementation
- Insights integrated into staff utilization tracking system, improving resource planning

8. Future Scope

- Build a no-show **prediction model using logistic regression**
- Integrate appointment feedback data (e.g., rescheduling reasons)
- Develop a real-time dashboard showing daily no-show risk by clinic
- Automate reminder logic using Python + Twilio API based on risk tier
- Extend to **walk-in flow analysis** to optimize overbooking strategy