# Healthcare Appointment No-Show Prediction

### Tools Used

- **Programming:** Python (Pandas, Scikit-learn)
- **Visualization & Reporting:** Power BI

---

# 1. Data Collection & Preprocessing

**Given dataset:**The dataset contains historical medical appointment records with fields including:

- Patient ID
- Appointment Date and Day
- Age
- Gender
- SMS Reminder (Yes/No)
- Scholarship (Medical aid)
- No-show status (Target variable)
- Hypertension, Diabetes, Alcoholism, Handicap
- Neighbourhood

### Cleaning Steps in Python

- Removed invalid or negative values for age
- Dropped duplicate records
- Converted categorical features (e.g., `No-show`, `Gender`, `SMS_received`) to numerical
- Converted date fields to datetime objects
- Engineered new features: `Appointment_Day`, `Wait_Days`, `Is_Weekend`

---

# 2. Exploratory Data Analysis (EDA)

### Key Findings:

- **SMS Reminders:** Patients who received SMS reminders were more likely to show up.
- **Age Factor:** Elderly patients (60+) had better attendance. Teenagers had higher no-show rates.
- **Days of the Week:** Friday appointments had higher no-show rates.
- **Lead Time (Wait Days):** Longer wait times were linked to higher no-shows.
- **Scholarship Influence:** Patients with medical aid (scholarship) had a slightly higher no-show rate.

---

# 3. Predictive Modeling

### Model Used: Decision Tree Classifier

**Feature Set:**

- Age
- SMS_received
- Day_of_week
- Wait_Days
- Scholarship
- Hypertension
- Diabetes
- Gender
- Handicap

**Train-Test Split:**

- **Train:** 80%
- **Test:** 20%

**Model Performance:**

- **Accuracy:** ~76%
- **Precision:** 72%
- **Recall:** 70%
- **F1 Score:** 71%

**Interpretation:** The model moderately predicts the likelihood of a no-show and can be used to flag high-risk appointments.

---

# 4. Power BI Dashboard

### Dashboards Developed:

- **Appointment Trends Dashboard:**
  - Shows total appointments, no-shows, show rates by day of the week.
- **Demographic Insights:**
  - Heatmaps of age vs. no-show rate.
- **Impact of SMS Reminders:**
  - Bar chart comparing show rates between SMS-received vs. not.
- **Neighborhood Analysis:**
  - Geo-distribution of high no-show regions.

---

# 5. Optimization Recommendations

### Based on Insights:

- **Automated SMS Reminders**
  Implement multiple reminders (1 day before and on the morning of the appointment).
- **Optimize Scheduling**
  Avoid high no-show days (e.g., Fridays) for first-time or high-risk patients.
- **Overbooking Strategy**
  Slightly overbook during known high no-show slots based on model probability.
- **Patient Follow-ups**
  For patients with high model-predicted no-show likelihood, follow up with calls or WhatsApp messages.
- **Lead Time Management**
  Keep waiting periods under 7 days wherever possible.

---

# 6. Deliverables Summary

| Deliverable | Description |
|---|---|
| **Prediction Model** | Python-based Decision Tree Classifier with 76% accuracy |
| **BI Dashboard** | Power BI dashboard covering show/no-show trends, SMS effectiveness, age demographics, and location analysis |
| **Recommendation Report** | Strategic suggestions for reducing no-shows through scheduling optimization and communication improvements |

---

Submitted by **Aishwarya P**