📄 Healthcare Appointment No-Show Prediction Report

📜 1. Project Overview

Objective:
Predict whether a patient will miss their medical appointment using data analysis and machine learning, and visualize insights through an interactive Power BI dashboard to support healthcare scheduling decisions.

Dataset Source:
KaggleV2-May-2016.csv (from Kaggle)

📊 2. Dataset Description

Original Dataset Columns:

Column	Description
PatientId	Unique patient identifier
AppointmentID	Unique appointment identifier
Gender	Gender of the patient
ScheduledDay	Date when appointment was scheduled
AppointmentDay	Date of the actual appointment
Age	Age of the patient
Neighbourhood	Location of the clinic
Scholarship	Whether on welfare (Bolsa Família)
Hipertension	1 = Yes, 0 = No
Diabetes	1 = Yes, 0 = No
Alcoholism	1 = Yes, 0 = No
SMS_received	1 = Received SMS reminder
No-show	'Yes' = Did not show, 'No' = Showed up

🔧 3. Data Preprocessing (Python)

- Dropped unneeded columns (PatientId, ScheduledDay, etc.)
- Converted No-show from text to binary (1 = No-Show, 0 = Show)
- Added derived columns:

    - weekday: Weekday name of appointment
    - NoShowLabel: Readable version of no-show status
- Cleaned file saved as: cleaned_appointments_for_powerbi.csv

---

🤖 4. Machine Learning Model

- **Model**: Decision Tree Classifier
- **Libraries**: pandas, sklearn
- **Features Used**: age, gender, sms_received, weekday
- **Target Variable**: no_show
- **Accuracy**: ~76%

Code in: healthcare_no_show_prediction.ipynb

📈 5. Power BI Dashboard Overview

Dashboard File: no_show_dashboard.pbix

Main Visuals:

- **KPI Cards**:
    - Total Appointments
    - Total No-Shows
    - No-Show Rate
    - SMS Sent Count
- **Charts**:
    - Stacked Column: Age vs No-Show
    - Bar Chart: Weekday vs No-Show
    - Pie Chart: Show vs No-Show Distribution
    - Bar Chart: Appointments by Neighbourhood
- **Slicers**: Gender, Age, Weekday, SMS Received, NoShowLabel
- **Patient Table**: Includes gender, age, weekday, SMS, and label

Preview:

📌 6. Insights & Observations

- **SMS Reminders Help**: Patients who received SMS were less likely to miss appointments.
- **Age Impact**: No-shows were more common in younger age groups.
- **Weekday Trend**: Mondays and Tuesdays had higher no-show rates.
- **Neighborhood Differences**: Some neighborhoods had significantly higher no-show counts.

💡 7. Recommendations

- Send SMS reminders to high-risk patient groups.
- Avoid scheduling high-risk patients on Mondays.
- Analyze neighborhood-wise issues (e.g., transport or awareness).
- Use the model to flag likely no-shows and confirm with follow-up.

📦 8. Project Files

File Name	Purpose
KaggleV2-May-2016.csv	Original dataset
cleaned_appointments_for_powerbi.csv	Cleaned dataset for dashboard
healthcare_no_show_prediction.ipynb	Data cleaning and modeling
no_show_dashboard.pbix	Power BI dashboard file
dashboard_preview.png	Dashboard image
report.pdf	This report

🙋‍♂️ 9. Author

- **Name**: **Saurabh Walde**
- **Email**: Saurabhwalde722@gmail.com
- **GitHub**: https://github.com/SaurabhWalde