Healthcare Appointment No-Show Prediction

**Project Report**  
  
Tools Used: Python (Sklearn, Pandas), Power BI

# 1. Introduction

This project focuses on predicting if a patient is likely to miss their medical appointment. By working with a real-world dataset from Kaggle, the goal is to analyze key factors influencing patient attendance and develop strategies, including a predictive model, to help minimize appointment no-shows.

# 2. Dataset Overview

The dataset contains 110,527 medical appointments and information on 14 associated variables (characteristics) such as gender, scheduled day, appointment day, SMS reminders, and whether the patient showed up.  
Source: <https://www.kaggle.com/datasets/joniarroba/noshowappointments>

# 3. Methodology

## 3.1 Data Cleaning and Preparation

- Removed irrelevant columns like 'PatientId' and 'AppointmentID'  
- Renamed columns for readability  
- Converted date columns to datetime format  
- Created new features such as Day of Week, Waiting Days  
- Encoded categorical variables (e.g., Gender, No-show)

## 3.2 Exploratory Data Analysis (EDA)

Explored patterns using age, gender and SMS reminders. Identified the highest no-show rates by day of week and by other features.

## 3.3 Model Building

Used Decision Tree Classifier to predict the probability of a no-show.  
- Training/test split: 80/20  
- Evaluated using accuracy, precision, recall, and confusion matrix

# 4. Results & Insights

- Overall no-show rate: approximately 20%  
- SMS reminders were effective in reducing no-shows  
- Young adults and weekday appointments had higher no-show rates  
- Model accuracy: approximately 80%

# 5. Power BI Insights

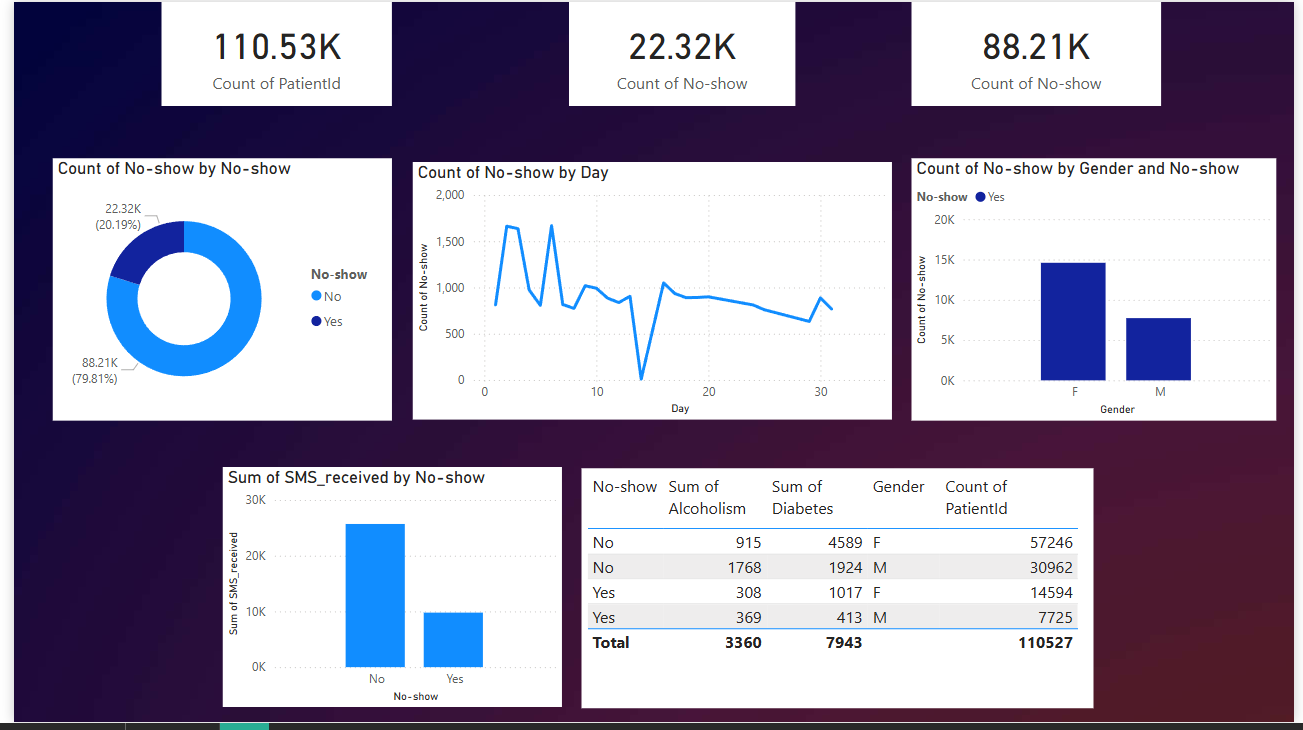
- Dashboard includes total appointments, show vs no-show ratio, no-shows by age, weekday, and SMS  
- Interactive filters allow slicing by gender, chronic conditions, and reminders  
- Insight pages highlight trends and high-risk groups visually

**6. Attachments**

* Goggle Colab Notebook having code for No-Show prediction

<https://colab.research.google.com/drive/1Ys3EN1hSdVKNEQYJKzbbvVKaPPH9Bqwu?usp=sharing>

* Screenshots of Power BI dashboard



# 7. Recommendations

- Send SMS reminders to all patients, especially young adults  
- Avoid scheduling too many appointments on days with historically high no-show rates, like Mondays.  
- Schedule follow-up calls for chronic patients with prior no-shows.

# 8. Conclusion

The project highlighted important factors contributing to missed appointments. Both the predictive model and the dashboard offer practical insights that can help streamline scheduling and enhance the overall efficiency of healthcare services.