# Healthcare Appointment No-Show Prediction

## 1. Introduction

Missed healthcare appointments, also known as no-shows, cause inefficiencies in medical facilities by wasting time, disrupting workflows, and increasing operational costs. Predicting which patients are likely to miss appointments allows healthcare providers to take proactive measures, such as sending reminders or overbooking intelligently. This project uses historical appointment data to build a predictive model and dashboard that helps optimize appointment scheduling and improve clinic resource utilization.

## 2. Abstract

This project focuses on using real-world medical appointment data to develop a machine learning model that can predict appointment no-shows. After cleaning and analyzing the data, a Decision Tree Classifier was trained, and its performance was enhanced using hyperparameter tuning. Key trends affecting patient attendance—such as SMS reminders, appointment weekdays, and patient age—were identified through exploratory data analysis (EDA).

## 3. Tools Used

- **Python**: Data preprocessing, modeling (Pandas, Sklearn), and visualization.
- **Jupyter Notebook**: For interactive code execution.
- **Matplotlib / Seaborn**: For exploratory data visualization.

## 4. Steps Involved in Building the Project

1. **Data Import and Cleaning**
   - Loaded the dataset (appointments.csv), removed rows with negative age or inconsistent data.
   - Converted date fields to datetime format and standardized column names.
   - Encoded categorical variables such as gender, no-show, and SMS_received.
2. **Exploratory Data Analysis (EDA)**
   - Analyzed patterns in patient behavior, including the effect of SMS reminders, waiting time, and appointment day.
   - Created visualizations such as bar plots and pie charts to identify insights.
3. **Feature Engineering**
   - Calculated waiting_days from ScheduledDay and AppointmentDay.
   - Extracted appointment_weekday and binned ages into groups.
4. **Model Building**
   - Trained a Decision Tree Classifier using sklearn.
   - Split data into training and testing sets for evaluation.
5. **Hyperparameter Tuning**
   - Used GridSearchCV to find the optimal values for max_depth, min_samples_split, and min_samples_leaf.
   - Improved model performance and avoided overfitting.

6. **Prediction for New Entry**
    - Added functionality to predict a specific patient's attendance status using their attributes.

## 5. Conclusion

This project illustrates the value of applying data science techniques to real-world healthcare challenges. With a well-tuned decision tree model and an insightful Power BI dashboard, medical staff can better predict no-shows and optimize scheduling. The outcome is a more efficient clinic operation, improved resource allocation, and better patient outcomes. Future extensions may include testing ensemble models like Random Forest or integrating a real-time prediction system into hospital scheduling software.