HealthCare Appointment – No Show Predictions

**Introduction**

Healthcare providers across the world face a persistent challenge of patients not attending scheduled medical appointments. These **“no-shows”** create significant operational inefficiencies, including wasted physician time, underutilized resources, longer waiting times for other patients, and financial losses for hospitals and clinics. Understanding the factors that influence patient attendance and predicting no-shows in advance can play a vital role in optimizing scheduling systems, improving patient care delivery, and reducing costs.

This project addresses the problem by analyzing healthcare appointment data, identifying trends behind patient behavior, and developing a predictive machine learning model to estimate the likelihood of no-shows. Furthermore, interactive dashboards in **Power BI** were built to provide healthcare administrators with actionable insights for better decision-making.

## ****Abstract****

The project analyzes healthcare appointment data to predict whether patients will miss their scheduled appointments. By applying a Decision Tree model in Python (Sklearn, Pandas) and exploring the dataset with Power BI, important behavioral patterns such as the effect of SMS reminders, patient age groups, and weekday scheduling were identified. A prediction model was built to assist healthcare providers in improving scheduling efficiency. Additionally, Power BI dashboards were developed to visualize patient demographics, no-show trends, and influencing factors, offering actionable recommendations.

## ****Tools Used****

* **Python**: Data cleaning, preprocessing, feature engineering, and predictive modeling.
* **Sklearn**: Building and evaluating the Decision Tree model.
* **Pandas & Matplotlib/Seaborn**: Exploratory data analysis (EDA), visualization, and statistical insights.
* **Power BI**: Dashboard creation for interactive visual analytics.
* **Dataset**: The “Medical Appointment No Shows” dataset containing 14 columns such as Age, Gender, Neighborhood, Appointment Date, Scheduled Date, SMS reminders, and the target variable No-show.

**Steps Involved in Building the Project**

1. **Data Import & Cleaning**
   * Loaded the dataset into Python using Pandas.
   * Standardized column names and removed extra spaces.
   * Converted ScheduledDay and AppointmentDay into proper datetime objects.
   * Checked for missing/invalid values and ensured correct data types.
   * Encoded No-show column: *No = 0 (attended)*, *Yes = 1 (missed)*.
2. **Exploratory Data Analysis (EDA)**
   * Visualized **appointment counts** by age, gender, and weekdays.
   * Analyzed the **effect of SMS reminders** on reducing no-shows.
   * Observed that younger patients (<30 years) and certain weekdays (Monday/Friday) had higher no-show rates.
   * Generated plots such as bar charts, scatterplots, and count plots in Python and refined insights using Power BI dashboards.
3. **Model Building & Evaluation**
   * Selected features: Age, Gender, Neighborhood, SMS reminders, Appointment weekday.
   * Trained a **Decision Tree Classifier** to predict no-shows.
4. **Power BI Insights**  
   Designed interactive dashboards with:
   * **Line Chart** showing trends of missed appointments over time.
   * **Bar Chart** comparing no-shows across regions and neighborhoods.
   * **Donut Chart** illustrating the contribution of each age group.
   * **Filters (Slicers)** to analyze by gender, region, and category.
5. **Optimization Recommendations**
   * **SMS reminders** are effective — ensure all patients receive timely reminders.
   * Implement **age-specific strategies** (extra follow-up for younger patients who tend to miss appointments more often).
   * Reschedule or overbook **high no-show days** (e.g., Mondays/Fridays) to minimize empty slots.
   * Consider **penalty or incentive systems** for repeat no-show patients.

**Conclusion**

The **Healthcare Appointment No-Show Prediction** project demonstrates how predictive analytics and visual dashboards can be integrated to solve real-world healthcare challenges. Using machine learning, we can forecast the likelihood of patients missing appointments, enabling proactive scheduling adjustments. At the same time, Power BI dashboards provide healthcare administrators with an easy-to-understand, interactive platform to explore the data and uncover behavioral trends.

By combining data science with business intelligence tools, this project not only highlights the **importance of predictive modeling in healthcare operations** but also emphasizes the role of **visual storytelling for decision-making**. Implementing these strategies can improve hospital efficiency, enhance patient care, and reduce the financial burden of no-shows.