

Healthcare Appointment No-Show Prediction

A Data-Driven Analysis to Improve Patient Attendance

- PRIYADARSHINI GUPTA

Abstract

Missed medical appointments reduce healthcare efficiency and increase operational costs. This project analyzes healthcare appointment data to identify patterns associated with patient no-shows and to support improvements in scheduling and patient communication. Using exploratory data analysis, a decision tree model, and an interactive Power BI dashboard, the project highlights key factors influencing appointment attendance and proposes data-driven optimization strategies.

1. Introduction

Patient no-shows are a common challenge in healthcare systems, leading to wasted resources and longer waiting times. Identifying factors associated with missed appointments can help providers improve operational efficiency and patient engagement. This project analyzes historical appointment data to examine no-show behavior and to translate analytical findings into practical, business-focused insights. The emphasis is on interpretability and actionable outcomes rather than complex modeling.

2. Tools and Technologies Used

- Tools Used: Python (pandas, matplotlib, scikit-learn), Power BI
- Model Used: Decision Tree Classifier

The dataset includes patient demographics, appointment details, waiting time between booking and appointment, SMS reminder status, and attendance outcome (show/no-show).

3. Methodology and Project Steps

3.1 Data Preparation

The dataset was cleaned to remove inconsistencies and irrelevant fields. A key engineered feature was waiting days, representing the delay between scheduling and appointment date. Patient ages were grouped into categories (children, adults, middle-aged, seniors) to improve interpretability.

3.2 Exploratory Analysis

Exploratory data analysis examined no-show rates across age groups, waiting periods, and SMS reminder status. Percentage-based comparisons were used to ensure fair interpretation across groups of different sizes.

3.3 Modeling

A Decision Tree Classifier was trained to identify patterns associated with appointment attendance. Feature importance was extracted to understand which variables contributed most to the model's decision-making process.

3.4 Visualization

Insights were presented using an interactive Power BI dashboard designed for non-technical stakeholders. Tooltips were added to provide contextual explanations and improve interpretability.

4. Key Findings

- Longer waiting periods are associated with higher no-show rates.
- No-show behavior varies across age groups.
- Patients who received SMS reminders generally showed better attendance.
- Feature importance analysis identified waiting days and age as the most influential factors in predicting appointment outcomes.

Feature importance reflects the variables that most influenced the model's predictions and should not be interpreted as direct causation.

5. Optimization Recommendations

- Based on the analysis, the following actions are recommended:
- Reduce appointment waiting times where possible.
- Prioritize reminder strategies for high-risk age groups.
- Expand the use of SMS reminders, particularly for appointments scheduled far in advance.
- Continuously monitor no-show trends using dashboard-based reporting.

6. Conclusion

This project demonstrates how data analysis can support operational improvements in healthcare scheduling. By identifying key factors associated with missed appointments and presenting insights through an interactive dashboard, the analysis enables informed decision-making. Implementing targeted, data-driven interventions can help reduce no-show rates, improve efficiency, and enhance patient care outcomes.