**Case Study: No-Show Medical Appointments Dataset**

**🎯 Objective**

To analyze patterns from patient appointment data and understand factors that may be associated with patients not showing up for their scheduled medical appointments.

**Understand Patient Behavior Around Missed Medical Appointments**

By analyzing the dataset, we aim to:

1. **Identify patterns or factors** that may contribute to patients missing their medical appointments (e.g., age, gender, chronic illness, reminders, location).
2. **Evaluate the effectiveness of interventions**, like SMS reminders, in reducing no-shows.
3. **Provide actionable insights** for healthcare providers to:
   * Improve appointment attendance.
   * Reduce wasted time and resources.
   * Improve patient care and clinic efficiency.

**🧾 Dataset Description**

The dataset includes information about medical appointments in Brazil and whether patients showed up or not. Each row represents a patient appointment with the following fields:

* **PatientId**: Unique identifier for a patient.
* **AppointmentID**: Unique identifier for the appointment.
* **Gender**: Patient’s gender.
* **ScheduledDay**: Timestamp when the appointment was scheduled.
* **AppointmentDay**: The day the appointment was supposed to take place.
* **Age**: Age of the patient.
* **Neighbourhood**: Location of the patient.
* **Scholarship**: Whether the patient is enrolled in a social welfare program.
* **Hipertension / Diabetes / Alcoholism / Handicaps**: Medical conditions flagged as binary (1 = Yes, 0 = No).
* **SMS\_received**: Whether an SMS reminder was sent.
* **No-show**: Indicates if the patient missed the appointment ("Yes") or not ("No").

**🔍 Data Exploration & Key Observations**

**1. No-show Rate**

* From the sample, 3 out of 18 patients **did not show up** for their appointments.
* **No-show Rate = (3 / 18) \* 100 ≈ 16.7%**

**2. Gender Distribution**

* Female patients: 14
* Male patients: 4
* Among the no-shows, all 3 are female.
* Observation: **Females are more frequent in the dataset**, but further investigation is needed to assess if they are more likely to miss appointments.

**3. Age Factor**

* Ages range from **8 to 76**.
* No-shows were observed in the age groups: **23, 29, 40**.
* No clear trend with age from this small sample, but mid-aged adults may be more likely to miss appointments.

**4. Chronic Conditions**

* Out of the 3 no-shows:
  + 1 had a handicap
  + None had hypertension or diabetes
  + All 3 had no alcoholism
* Observation: Chronic conditions were not clearly linked to no-shows in this subset.

**5. SMS Reminders**

* 2 out of 3 no-shows **received SMS reminders**, while 1 did not.
* This suggests **SMS reminders alone may not guarantee attendance**.

**6. Neighbourhood Analysis**

* The **Nova Palestine** neighborhood has the **highest number of appointments (7)**.
* Among the no-shows:
  + 1 was from **Nova Palestine**
  + 1 from **Goiabeiras**
  + 1 from **Conquista**
* Observation: **No-shows are spread across different neighborhoods**, so locality might not be a strong factor.

**📌 Summary of Insights**

| **Feature** | **No-show Trend** |
| --- | --- |
| Gender | All no-shows were **female** |
| Age | No-shows in **mid-age** (23–40) |
| SMS Received | 2 out of 3 no-shows got SMS |
| Handicap | 1 out of 3 no-shows was handicapped |
| Neighbourhood | No clear locality-based trend |
| Chronic Illness | Not strongly correlated in this data |

**✅ Recommendations**

1. **Investigate why mid-aged female patients tend to miss appointments**—it could be work/life balance or transport-related.
2. **Improve communication beyond SMS**, approach for whatsapp, such as phone calls or app-based reminders.
3. **Further analyze full 110K+ records** for statistical validation of these trends.
4. **Consider adding cancellation options** to distinguish between intentional no-shows and emergencies.

.

**🧠 Python Questions for No-Show Appointments Dataset**

**📁 Data Loading and Exploration**

1. How would you read this dataset into a pandas DataFrame?
2. How can you check for missing or null values in the dataset?
3. How do you view the first 10 rows and the structure (columns and data types) of the DataFrame?
4. How would you get the number of unique patients in the dataset?
5. How would you check for duplicate rows based on PatientId and AppointmentID?

**📊 Descriptive Statistics and Aggregation**

1. How do you calculate the percentage of patients who did not show up for their appointment?
2. How can you group the data by gender and count the no-shows in each group?
3. How would you find the average age of patients who showed up vs. those who didn’t?
4. How can you count how many patients received SMS reminders but still didn’t show up?
5. How would you determine the neighborhood with the highest no-show rate?

**📅 Date and Time Analysis**

1. How would you convert ScheduledDay and AppointmentDay to datetime objects?
2. How do you calculate the number of days between the scheduling date and appointment date?
3. How would you find the average scheduling lead time for no-shows vs. shows?
4. How can you filter appointments that were scheduled on the same day as the appointment?

**🧼 Data Cleaning and Preparation**

1. How do you handle incorrect or outlier values in the Age column (e.g., negative values)?
2. How would you rename columns to follow snake\_case naming conventions (e.g., No-show to no\_show)?
3. How would you convert the No-show column into a Boolean column?
4. How can you drop columns that are not relevant to the analysis (e.g., AppointmentID)?

**📈 Basic Visualization-Oriented Questions**

1. How would you prepare data to plot the no-show rate by age group?
2. How do you summarize the number of appointments over time (daily or weekly)?

Bottom of Form