# *Healthcare Appointment No-Show Prediction Analysis*

-Rohan Prabhakar

In this project, the main aim was to predict whether a patient would show up for a scheduled medical appointment using historical appointment data. I began by importing essential Python libraries such as pandas and numpy for data manipulation, matplotlib for visualization, and several modules from scikit-learn for preprocessing and modeling. The dataset was loaded using pandas.read_csv(), and the dataset included features such as Gender, Age, Scholarship, Hipertension, Diabetes, Alcoholism, SMS_received, and the target variable No-show.

I first proceeded to clean the data by removing unnecessary columns and checking for missing values, ensuring that the dataset was fit for this model. For preprocessing, categorical features such as Gender and No-show were converted into numerical formats using Label Encoding, making them suitable for model input. I then split the dataset into training and testing sets using train_test_split and trained a Decision Tree Classifier. After training the model, I evaluated its performance and achieved an accuracy of approximately 80.02% which indicates a reasonably good predictive capability .

I also analyzed feature importances from the decision tree, revealing that a few features likely SMS_received or Age were more influential than others in determining whether a patient would attend their appointment.I also generated several plots for visualizations of feature importance, class distributions, and possibly age-related patterns.

## From Dashboard

The Healthcare Appointment Analysis dashboard provides a overview of patient attendance behavior across various factors. The most effective insight is the high no-show rate, with approximately 88,200 appointments missed out of a total of

110,000, indicating that nearly 80% of scheduled patients did not attend. This is a significant concern for healthcare service efficiency and resource planning.

Demographic analysis reveals that females represent a larger portion of the patient population which is about 65%, but both genders follow similar attendance trends. Age appears to play a critical role in attendance, as younger patients—especially children and adolescents—tend to have higher no-show rates, while attendance improves steadily with age.

From a geographic perspective, certain neighbourhoods like Jardim Camburi and Maria Ortiz report the highest number of missed appointments, suggesting possible accessibility or socioeconomic challenges in these areas. Medical factors such as alcoholism, diabetes, and hypertension are also represented, with over 3,300 cases of alcoholism, 7,900 patients with diabetes, and significant overlap between no-shows and those under scholarship programs or with hypertension. This may imply that patients with chronic conditions or financial aid are still facing barriers to care.

The influence of communication is evident: out of 35,000 patients who received SMS reminders, attendance did improve, but not substantially, suggesting that while reminders help, they are not sufficient on their own. Lastly, the attendance trend by date shows a noticeable drop around the middle of May, which could indicate a system-level or external event affecting patient turnout during that period.

Overall, the dashboard highlights the importance of targeted interventions in specific demographics, neighbourhoods, and patient groups to reduce no-show rates and improve healthcare delivery efficiency.