

TECH ALPHARETTA

2972 Webb Bridge Rd, Alpharetta, GA 30009

**Missing Doctor Appointments**

Doctor’s appointment no-shows are a serious issue in the public health care field. Missed appointments are associated with poorer patient outcomes and cost the health care system in the US nearly $200 each. Therefore, it comes as no small surprise that reducing the rate of no-shows has become a priority in the United States and around the world. Numerous studies have been undertaken in order to determine the most effective means of reducing rates of absenteeism at with varying degrees of success.

# Introduction to Dataset

Dataset is drawn from 300,000 primary physician visits in Brazil across 2014 and 2015. The information about the appointment was automatically coded when the patient scheduled the appointment and then the patient was marked as having either attended or not. The information about the appointment included demographic data, time data, and conditions concerning the reason for the visit.

A total of 14 variables are included in the Data set. The variables and the description of the values are as follows

1. **Age**: integer age of patient
2. **Gender**: M or F
3. **AppointmentReservationDate**: date and time appointment was made
4. **AppointmentDate**: date of appointment without time
5. **DayOfTheWeek**: day of the week of appointment
6. **Status**: Show-up or No-Show
7. **Diabetes**: 0 or 1 for condition (1 means patient was scheduled to treat condition)
8. **Alcoholism**: 0 or 1 for condition
9. **Hypertension**: 0 or 1 for condition
10. **Smoker**: 0 or 1 for smoker / non-smoker
11. **Scholarship**: 0 or 1 indicating whether the family of the patient takes part in the Bolsa Familia Program, an initiative that provides families with small cash transfers in exchange for keeping children in school and completing health care visits
12. **Tuberculosis**: 1 or 0 for condition
13. **SMSReminder**: 0 ,1 ,2 for number of text message reminders sent to patient about appointment
14. **WaitingTime**: integer number of days between when the appointment was made and when the appointment took place.

**Objective**

The first step to solving the problem of missed appointments is identifying why a patient skips a scheduled visit in the first place. What trends are there among patients with higher absence rates? Are there demographic indicators or perhaps time-variant relationships hiding in the data?

Your responsibility as Data Scientist/Analysis is to find reasons behind missed appointments through data provided, examine the data to identify any trends present.