Midterm Project – Medical Appointment No Shows

Courtney Drysdale

Regis University

MSDS 670 Data Visualization

Professor John Keonig

July 27, 2025

**Midterm Project – Medical Appointment No Shows**

For this project, I selected a dataset about medical appointment no shows. This dataset is made up of 110,527 medical appointments in a city in Brazil over a few months in 2016. It has 14 variables: Patient ID, Appointment ID, Gender, Scheduled Date, Appointment Date, Age, Neighborhood, Scholarship, Hypertension, Diabetes, Alcoholism, Handicap, SMS received, and (most importantly) No show.

I cleaned the data and created new variables based on the provided variables for the day of the week the appointment was scheduled and how many days there were between the day the appointment was scheduled and when the appointment actually happened. I also converted the dates to a better format so that I could parse out the days of the week, and I changed the categories that were 0 and 1 representing true or false to Yes and No. If I were doing machine learning with this dataset, I would like the numerical data, but changing it to categorical made it easier to use pivot tables and create charts with the data.

**Research Question**

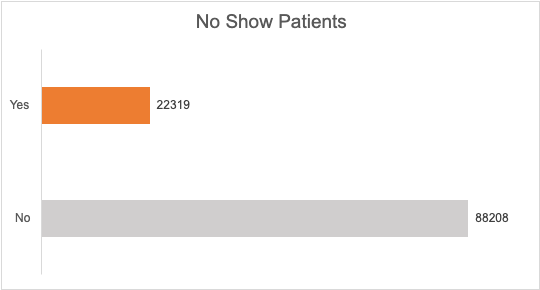
Since I am looking at visualizations for this dataset, I chose to look at which different groups “no showed” to their appointments more often. My research question was “Which groups of people are more likely to no show a medical appointment?” Looking for answers to this question would allow the medical practice to target the groups most likely to no show in the future to prevent future no shows.

**Methodology**

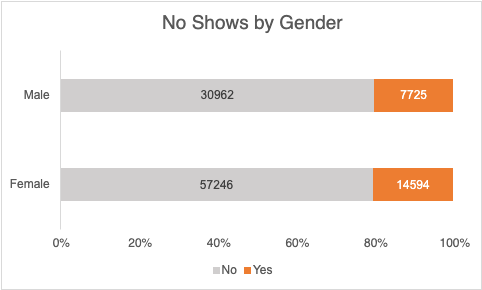
For this research question, I will start by looking broadly at the demographics of the entire group. I’ll look at the differences in certain variables between the “no show” and “showed up” groups within the dataset. I will identify groups of people at risk of no showing and summarize the dataset for the medical group.

**Results**

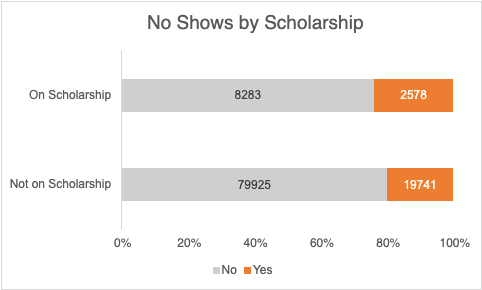
First, I looked at the big picture – how many patients no showed? Out of 110,527 appointments, 22,319 patients no showed, while 88, 208 attended their appointment. I used pivot tables on my dataset to get this information, then made a bar chart showing the information.



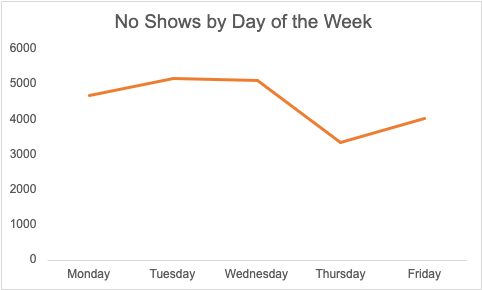
Next, I was curious if there might be a difference between male and female patients. Again, I used a pivot table to get this information in a usable way to create a chart. When looking at the numbers, it appears that there are twice as many female patients who no showed as male patients. However, there are also about twice as many female patients in the dataset, so when we look at it proportionally, it is about the same amount of male and female patients who no show.



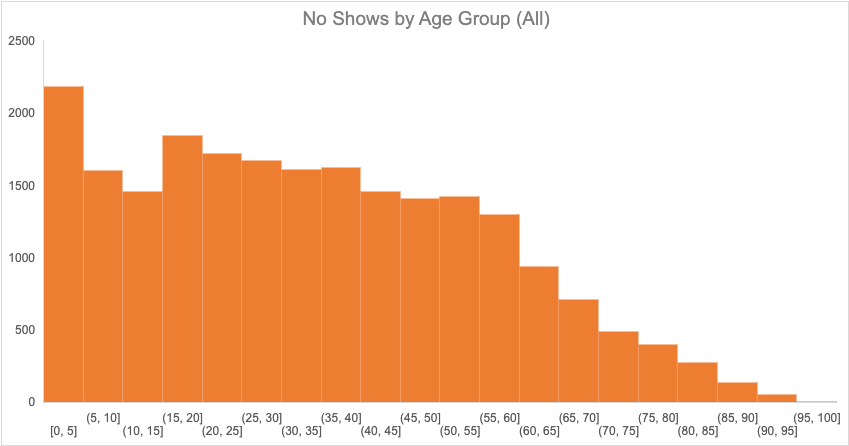
Being "on scholarship" in Brazil is the way they refer to government assistance for healthcare. While there are much fewer people on scholarship in the dataset, there are slightly more proportionally in this group to no show an appointment.

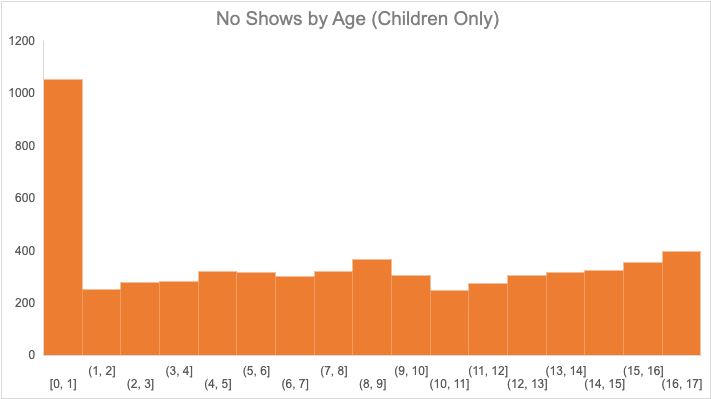


Next I looked at if the day of the week made a difference for no shows. No Shows happen every day of the week but happen more often earlier in the week. This dataset also had data for a few Saturday appointments (39) but there were so few no shows (9) that I left it off the chart.

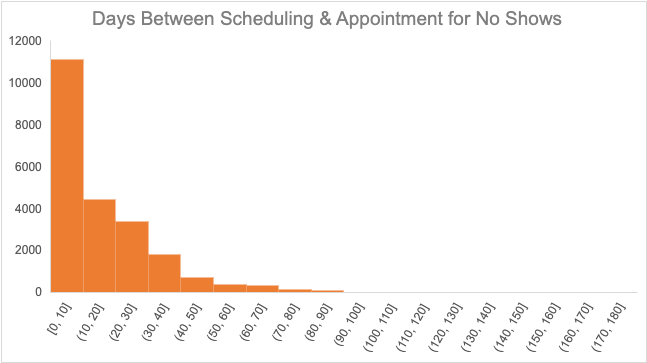


Next I looked at age groups to see if any age group had more no shows than the others. First I looked at the no shows on a histogram for all ages in 5 year bins.

 I noticed that there were more no shows toward the younger ages, so next I looked at no shows just among children. There were overwhelmingly more no shows for children under age 1 than any other group.



Finally, I looked at the number of days between scheduling the appointment and the actual appointment. Nearly half of the no shows were appointments that had been scheduled within the next 10 days.



**Conclusion**

By analyzing this data with visualizations, I came to the following conclusions about which groups no show medical appointments.

* The age group at most risk of no showing are children under age 1.
* Male and female patients no show at about the same rate.
* Patients on scholarship no showed at a slightly higher rate than those not on scholarship.
* More patients no showed earlier in the week.
* More patients no showed to appointments they had scheduled within 10 days.

Medical practices can use this information to determine which groups are most likely to no show and thus may need more outreach or reminders to show up for appointments. This exercise could be repeated in the future after implementing new outreach to determine if patients are showing up for more appointments rather than no showing.

**References**

Hoppen, J. (2016). *Medical Appointment No Shows.* <https://www.kaggle.com/datasets/joniarroba/noshowappointments/data>