Fundamentals of Data Science 1

1. Dataset analyzed

The chosen dataset was **No-show appointments**

2. Question(s)

The main question is whether there are characteristics that indicate that a patient will not attend

To answer that question, It was analyzed 7 intermediate questions that investigate some features individually against no-show.

3. Discussion

Firstly, it was investigate the amount of data and type of each column. In addition, it was check if there were missing and duplicates. Some cleaning was made to eliminate patients with age less than 0, if exists and patients with appointment dates before scheduled dates.

After the cleaning process, some new columns were created in order to improve some following analysis.

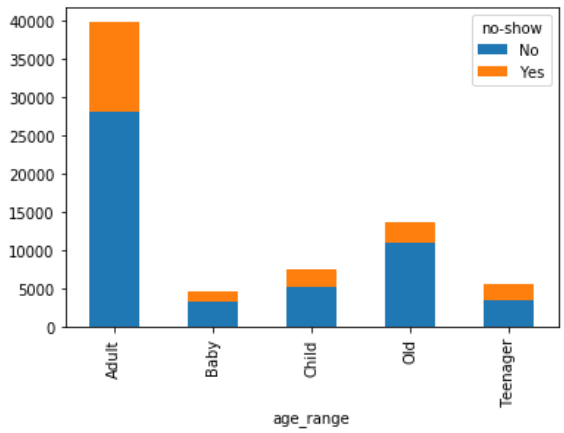
- **age_range** to represents babies, children, Adults and Agededs.
- **scheduled_weekday** and **appointment_weekday** to store the weekday of schedule and appointment
- **schedule_period_of_month** and **appointment_period_of_month** to store the period (beginning, middle or end) of the schedule and appointment
- **days_between_sched_appoint** the amount of days between scheduled date and appointment date

Finally, the analysis were made in order to answer the intermediate questions and main question.

4. Summary statistics

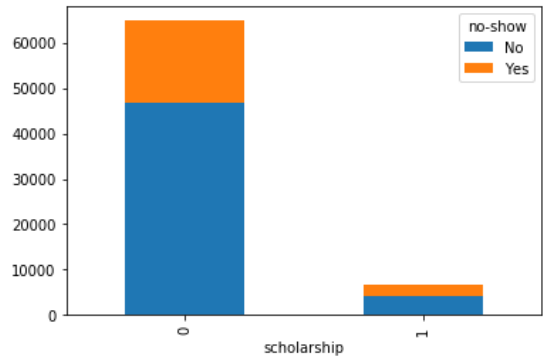
The main features involved in the question answer were ***age range*** of patients, ***receipt of a scholarship*** and ***interval between schedule and appointment***.

Age range



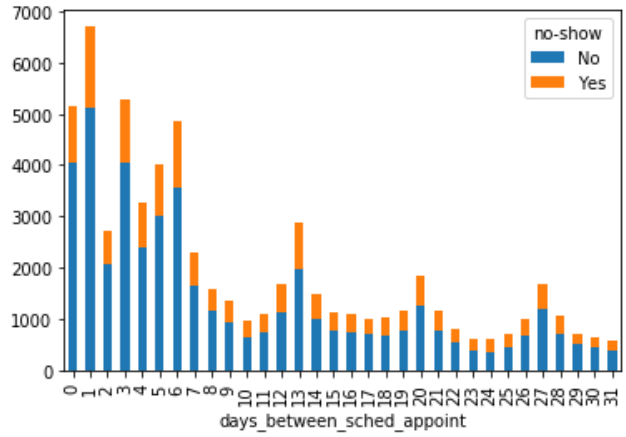
According to the plot, the age range Teenager has about 37% of no-show frequency. Baby and Child have about 30%. Those three types could be considered more likely not to show up.

Scholarship



According to the plot the amount of patients that receives scholarship is less than the amount of patients who don't. Despite of it, the proportion of those don't show up is greather in patient group who receives scholarship, about 35% against 28% who don't receive.

Interval between schedule and appointment



When the range of days between scheduling and appointment is less than 10 days the no-show rate is around 25%. From ten days onwards this rate is around 30% and its highest point (40%) when the interval is 24 days. We can conclude that when the consultation date is more than 10 days from the date the appointment was made, the number of no-shows tends to be higher.

By the answer of the previous intermediate questions, we got some information that could give some clue that a patient will not show up.

1. When the appointments are for Babies, Children or Teenagers, the frequency of no-show is greather than it is for Adults or Agededs.
2. Patients who receives scholarship have a slightly bigger no show frequency than those dont receive
3. It was detected that when the interval of days between schedule and appoint is greather than 10 days, the number of no-shows tends to increase.