Introduction

No-shows are a significant problem in health care that carry a substantial financial cost for providers and worsen outcomes for patients. We are working with health care providers to help combat the issue of doctor's appointment no-shows. The goal of the project is to analyze no-show data from medical appointments to identify patterns in the types of patients and appointments that end up being no-shows. By determining which appointments are at risk of turning into no-shows, we can assist clinic administrators in designing interventions to help prevent no-shows, which will decrease costs and improve patient outcomes.

Description of your dataset

Our initial dataset contains 110,527 observations and 14 variables of appointments at medical clinics in São Paulo, Brazil. The variables include:

Patient Id: 61,744 Unique observations

#Indicates identification of the patient

Appointment Id: 110,527 unique observations

# Indicates identification of the appointment

Gender: 71,840 Female observations, 38,687 Male observations

#Indicates sex

Scheduled day: 110,527 Unique observations

# Indicates date and time that the appointment was scheduled

Appointment day: 110,527 Unique observations

# Indicates date and time of the appointment

Age: 104 Unique observations

Min: -1

1st Quartile: 18

Median: 37

Mean: 37.09

3rd Quartile: 55

Max: 115

#Indicates age of the patient, negative values refer to unborn children

Neighborhood: 81 Unique observations

#Indicates the location of the appointment

Scholarship: 10,861 scholarship observations; 99,666 non-scholarship observations

#Indicates participation in Bolsa Família, a social welfare program in Brazil

Hypertension: 21,801 hypertension observations; 88,726 non-hypertension observations

#Indicates presence of hypertension

Diabetes: 7,943 diabetes observations; 102,584 non-diabetes observations

#Indicates presence of diabetes

Alcoholism: 3,360 alcoholism observations; 107,167 non-alcoholism observations

#Indicates presence of alcoholism

Handicap: 3 level four observations; 13 level three observations; 183 level two observations; 108,286 level zero observations

#Indicates level of disability from 0-4

SMS received: 35,482 observations of SMS received; 75,045 observations of non-SMS received

#Indicates whether one or more confirmation text messages were opened by the patient

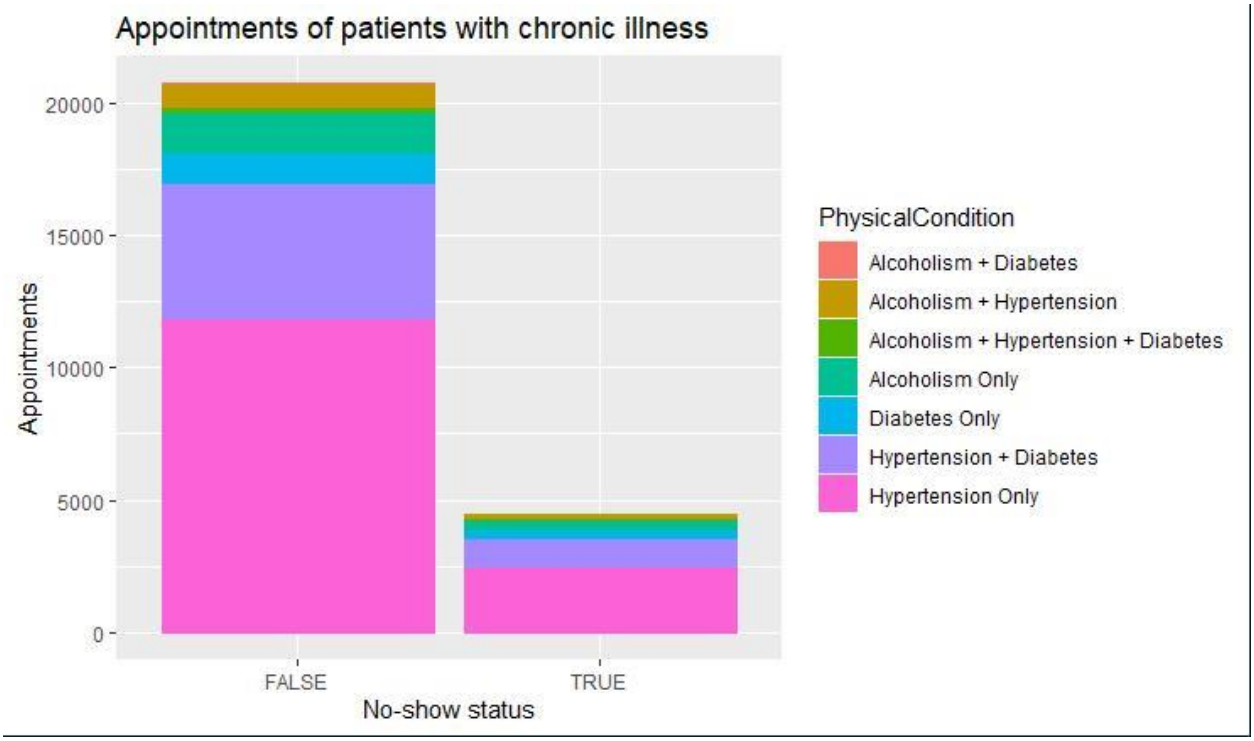
No-show: 22,319 observations of no-show; 88,208 observations of show

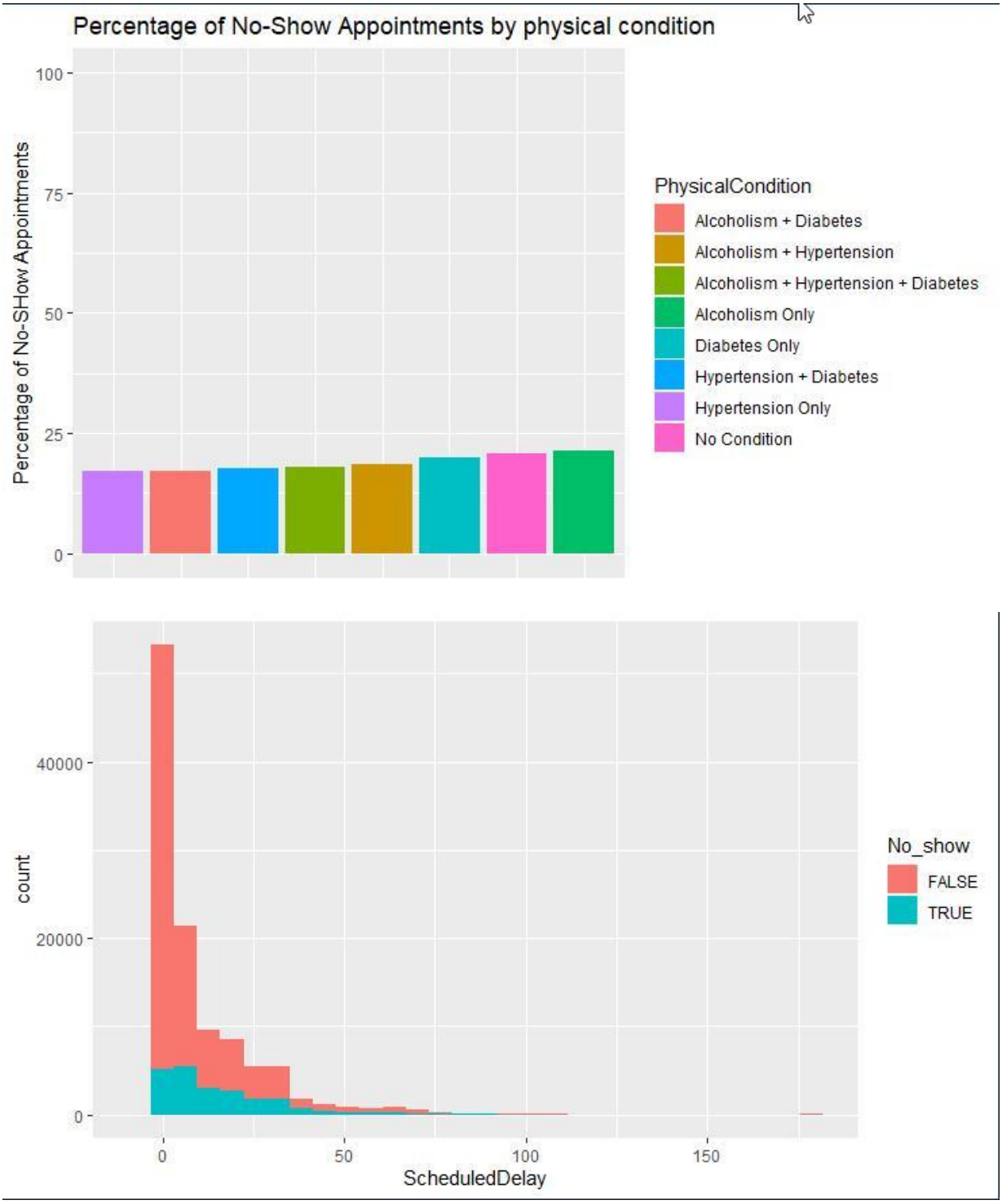
#Indicates if the patient was present for the appointment

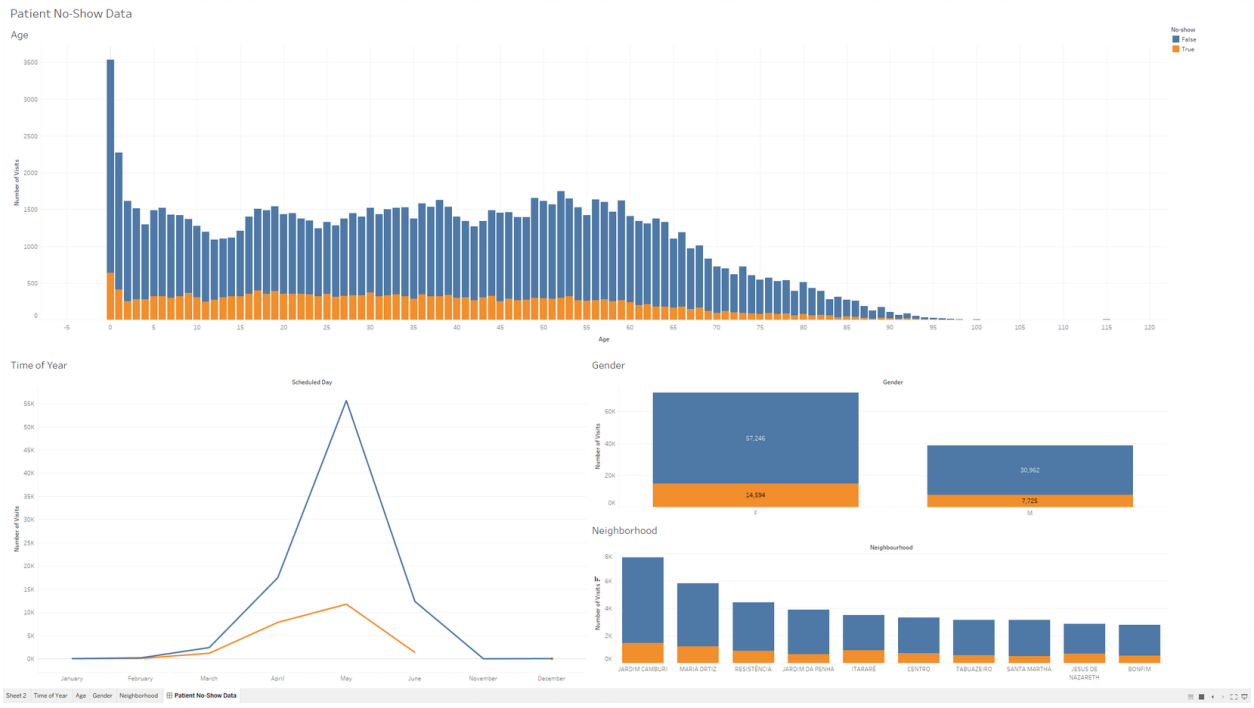
Table of the number of No-shows in each group of conditions:

Condition	No-show	% of No-shows	Show	% of Shows
Scholarship	2578	11.6	8283	9.4
Hypertension	3772	16.9	18,029	20.4
Diabetes	1430	6.4	6513	7.4
Alcoholism	677	3.0	2683	3.0
SMS received	9784	43.8	25,698	29.1

Exploratory visualizations:




Percentage of No-Show Appointments by physical condition
Percentage of No-SHow Appointments
100
75
50
25
0
PhysicalCondition
Alcoholism + Diabetes
Alcoholism + Hypertension
Alcoholism + Hypertension + Diabetes
Alcoholism Only
Diabetes Only
Hypertension + Diabetes
Hypertension Only
No Condition
count
40000
20000
0
0
50
100
150
ScheduledDelay
No_show
FALSE
TRUE



Completed steps:

- Understand and describe the problem
- Obtain data
- Load data into R to clean and preprocess
- Convert data types to more useable formats
- Reduce data
- Explore data in R and Tableau
- Identify preliminary patterns and connections

Future plan & Conclusion

Our preliminary plan is to characterize which patient attributes contribute the most to whether an appointment is a no-show. Next, we will attempt to develop a model to estimate the probability that a new appointment will be a no-show. If possible, we will develop a strategy to schedule appointments to minimize the risk of no-show appointments. We will achieve these goals by working through the 10 step process of a business intelligence solution and ultimately produce an actionable solution for our clients.