No Show Appointments Dataset Report

Introduction

- **Problems Faced:** The Dataset we are going to analyze talks about the criterias for patients of several clinics and the end result of which if a patient has attend their appointment or not and what aspects affect the attendance for the appointments.
- **Data Gathering:** Here we used a CSV file that contains all the pertinent information that would help us decide what are the criterias that affect the patients decision to attend the appointments and make us able to predict future appointments attendance.
- **Solution Highlight:** First is we clean the data gathered, then we analyze these data and visualize our finding so we could answer all the questions we have about the data.

Data

Columns Definitions

- **PatientId**: Represent a unique number for each number to be identified with
- **AppointmentId**: Represent a unique number for each appointment to be identified with
- **Gender**: M for Males & F for Females
- **ScheduledDay**: Represent the day where the appointment was booked
- **AppointmentDay**: Represent the day of the appointment to be attended
- **Age**: Represent the age of the patient
- **Neighborhood**: Represent the location of each hospital
- **Scholarship**: indicates whether or not the patient is enrolled in Brasilian welfare program
- **Hipertension**: Represent if the patient has hypertension
- **Diabetes**: Represent if the patient has Diabetes
- **Alcoholism**: Represent if the patient is alcoholic
- **Handicap**: Represent if the patient is Handicapped
- **SMS_received**: 1 Represent that the patient received the SMS set to them and 0 represent the patient did receive SMS
- **No-show**: Yes if the patient didn't show up and No if they did show up

Steps Taken

- Change the name of No-show Column to Attended the I switched the values if Yes and No so now in the Attended column has two values, Yes represent that the patient attended there appointment and No represent that they didn't attend.
  The reason for the update is that it decreases confusion
- Change AppointmentId to AppointmentID for consistency
- Change the name of Handcap Column to Handicap
- Change the values (1,2&3) from the Handicap column into 1 since the column represent only True and False statement then any other values that would represent a True statement are redundant and should be converted to 1
- Replace F with Female & M with Male in Gender column for less confusion
- Change the datatype of the ScheduledDay and AppointmentDay to datetime
- Added AppMonthName that represent the name of the month of each appointment
- Added AppDayName that represent the name of the day of week of each appointment
- Added Health_Issues_Num that represent the count of health issues for each patient
- Added Health_Issues column which represent the names of the issues for each patient
- Added AgeGroups column that represent the 4 age groups for analysis (Under 21, from 21 to 35, from 35 to 50, Over 50)

Analysis

Questions and Conclusions

1. Does a certain gender has higher commitment rate to the appointments they book?

   → Males have approximately 80% attendance rate while Females have 79% so both males and females approximately have the same attendance rate there for gender is not a factor that we would take in consideration.

2. Which Appointment months does this data cover?

   → This data set only covers the appointments in April, June and May

3. Does a certain month has higher attendance rate?

   → April has 80% attendance rate , June has 82% attendance rate and May has 79% attendance rate .

   since all three have approximately the same attendance rate that means that the month does not affect the attendance.

4. What is the difference in the influx of patients in each month?

   → May has the most influx with 64073 appointments then comes June with 21569 appointments and finally comes April with 2602 appointments.

5. Does the clinic has any Holidays?

   → Yes Sunday is a Holiday

6. What is the difference in the influx of patients in each Day?

   → All week days has approximately near influx except for Saturday which only has 30 appointments.

   This means that the people don't book much appointments in Saturday since it's not a Holiday but still has much less appointments.

7. Is Age a factor on how much does a person goes to the clinic?

   → Yes, The analysis shows that people Under 21 and the people Over 50 are the most to visit the clinic while people Over 21 and Under 50 visit the clinic less.

8. Does the neighborhood of the clinic has an effect on the number of appointments?

   → Yes, as the top 10 neighborhoods has much higher appointments than the bottom 10.

9. Are there a lot of people who got the Bolsa Familia Scholarship?

   → Only 10,861 out of 110527 (Approximately 0.1% get the scholarship)

10. What is the difference in ratios of appointments attended in people who got the scholarship Vs those who didn't get it?

    → 76% of people who got the scholarship attended their appointments while 80% of people who didn't get the scholarship attended their appointments.

11. Does the number of health issues a patient have affect the number of appointments they have with the clinic and their attendance rate?

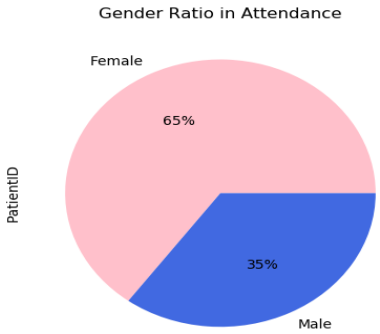
    → Not necessarily, as the patients with no health issues has attendance rate of 79.6% while patients with all 4 health issues has only 69.2% attendance rate.

12. Does it help when sending SMS to patients to remind them of their appointments?

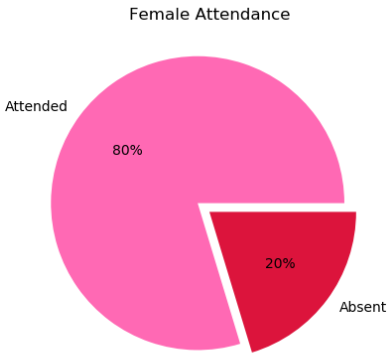
    → No, in fact when no SMS is sent the attendance rate is 83% while when there is a SMS sent the attendance rate decreases to 72%

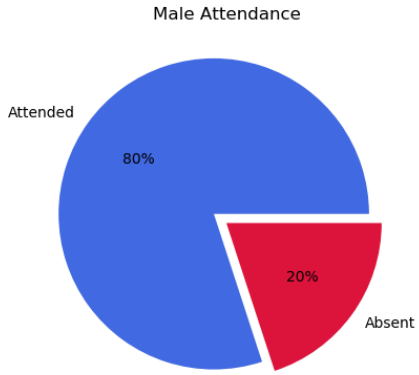
Outputs & Visualizations

1. Figure 1 is a Pie chart that represent the difference between number of Male and Female appointments.

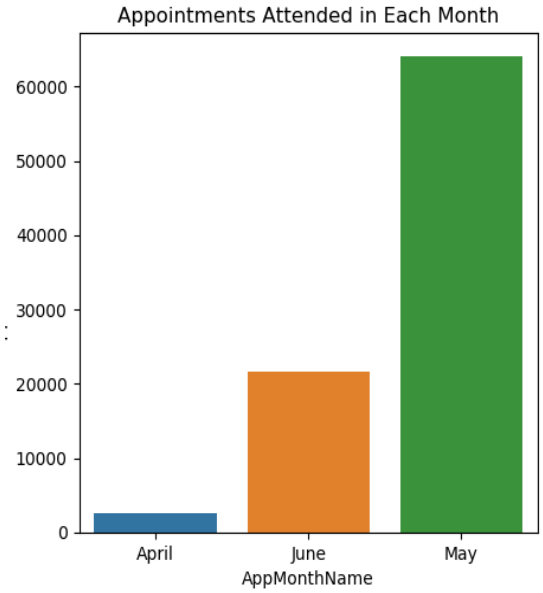


2. Figure 2 and 3 are a Pie charts that represent the ratio between the attendance and absence in Females and Males appointments.

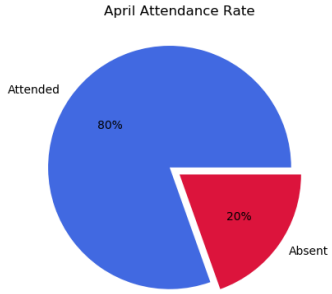


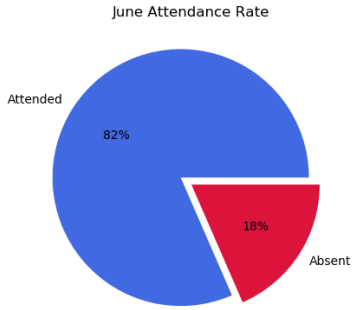


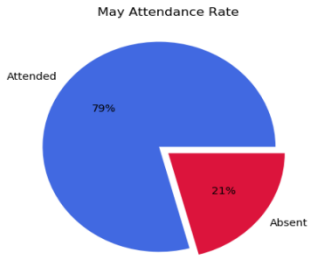
3. Figure 4 is a Bar chart the represent the difference in attendance in each of the months we have



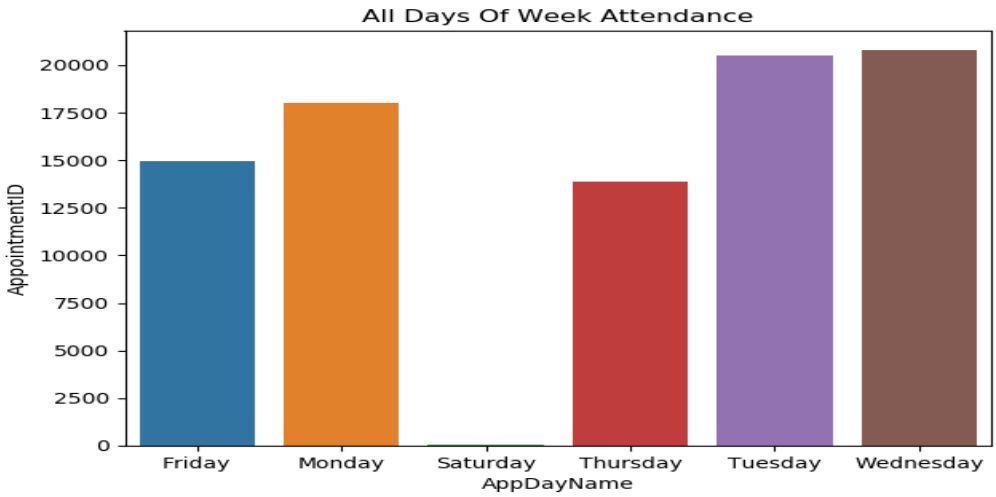
4. Figures 5,6,7 are Pie charts that represent the attendance rate for April, June, May consecutively.



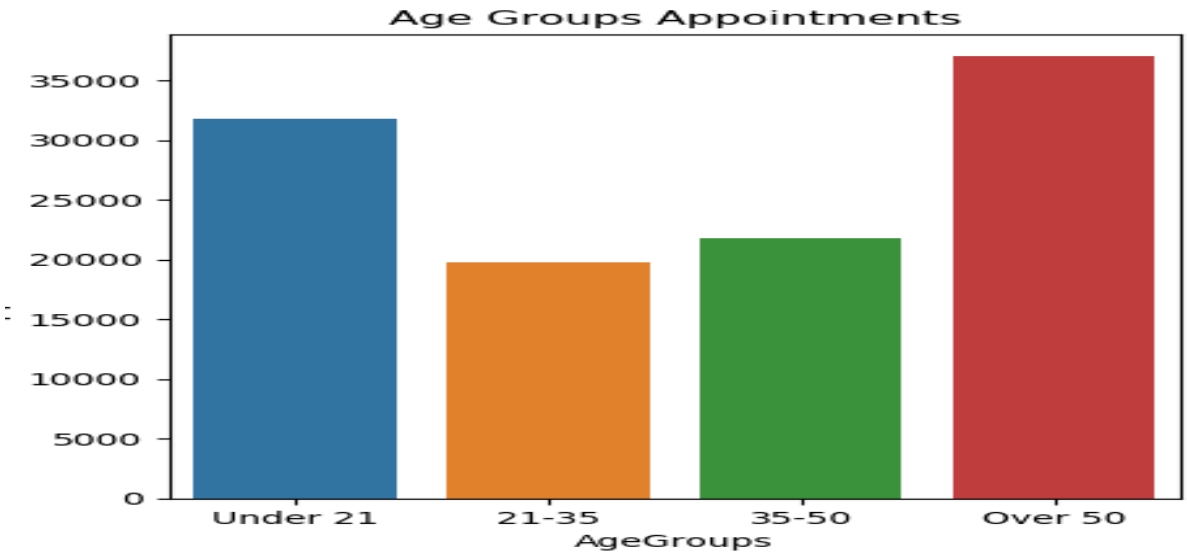




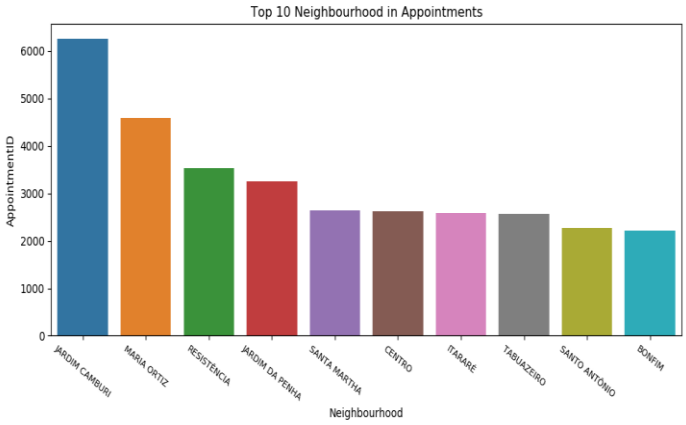
5. Figure 8 is a Bar chart that represent the attendance for each day of the week and conclude the holidays in the week.

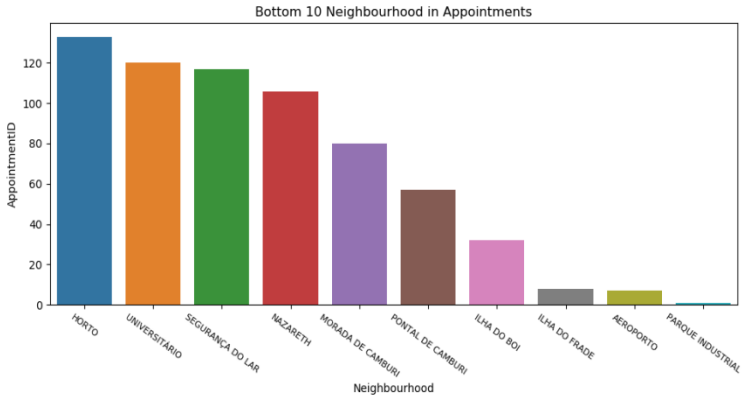


6. Figure 9 is a Bar chart the represent the attendance for each of the age groups we have (Under 21, 21-35, 35-50, Over 50).

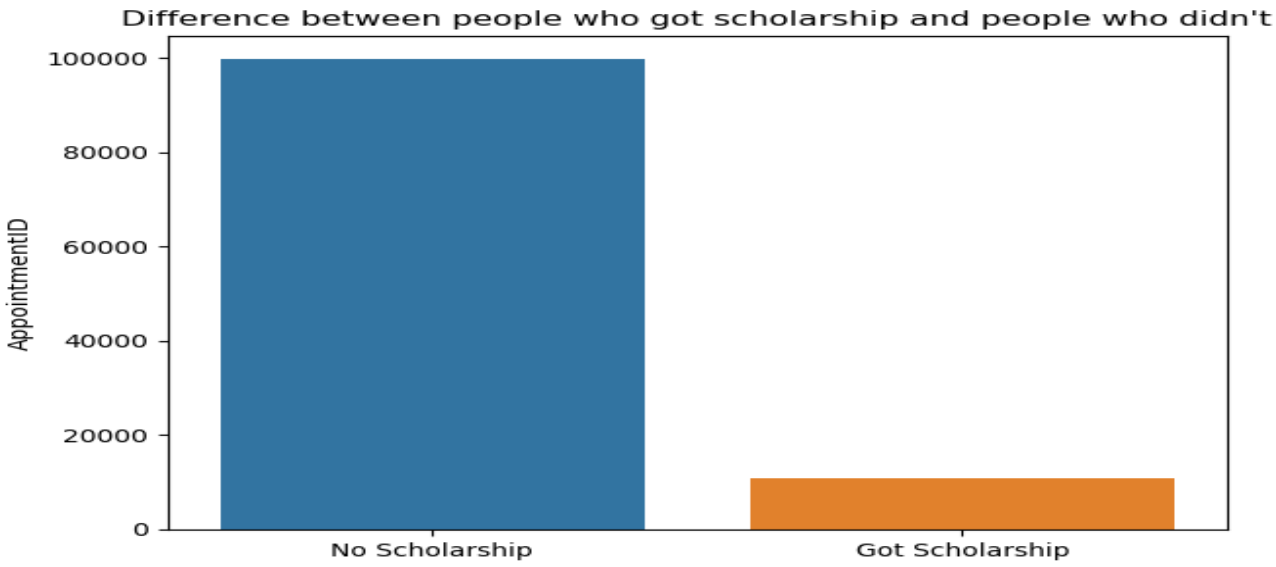


7. Figures 10 and 11 are Bar charts that represent the number of appointments in the top 10 neighborhoods a bottom 10 neighborhoods consecutively.

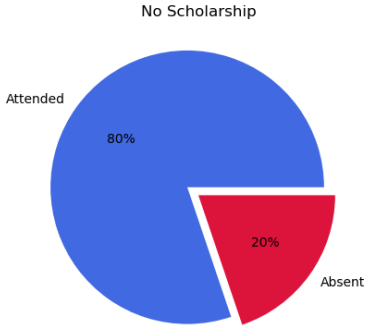


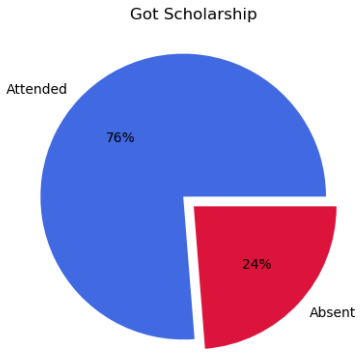


8. Figure 12 is a Bar chart that represent the difference in number of people who got the Bolsa Familia scholarship and people who didn't.



9. Figures 13 and 14 are Pie charts that represent the ration in attendance for people who didn't get the scholarship and people who did get it.



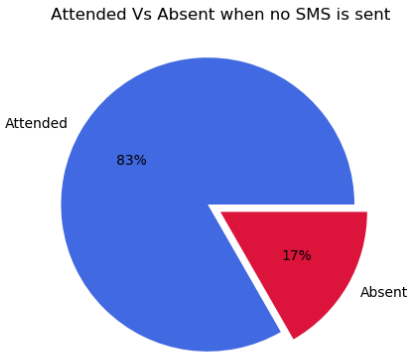


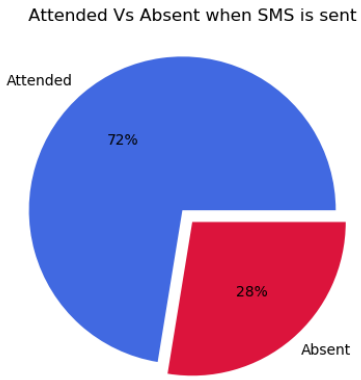
10. Figure 15 is a Table that compare between the health issues a patient have and their attendance rate.

Health Issues and thier Appointments

Health_Issues	Attended	Absent	Attendance Rate
Hipertension, Alcoholism, Handicap	28	1	96.6%
Diabetes, Handicap	37	3	92.5%
Alcoholism, Handicap	31	4	88.6%
Hipertension, Diabetes, Handicap	294	51	85.2%
Diabetes, Alcoholism	63	12	84.0%
Hipertension	11338	2325	83.0%
Hipertension, Diabetes, Alcoholism	201	42	82.7%
Hipertension, Diabetes	4841	1044	82.3%
Hipertension, Handicap	473	108	81.4%
Hipertension, Alcoholism	845	197	81.1%
Handicap	962	235	80.4%
Diabetes	1068	273	79.6%
No Health Issues	66512	17603	79.1%
Alcoholism	1506	416	78.4%
Hipertension, Diabetes, Alcoholism, Handicap	9	4	69.2%
Diabetes, Alcoholism, Handicap	0	1	0.0%

11. Figures 16 and 17 are Pie charts that represent the ration in attendance between people who received SMS reminder and people who didn't.





Results and Conclusions

Summary

In the analysis part of the report we concluded and visualized the factors that affect the attendance of the appointments and how much they affect it.

We use visualization to understand the data and the analysis better

Recommendations For Actions

- Since Saturday has way less appointments than the other day we could take it as a Holiday with Sunday
- Since sending SMS messages to the clients doesn't really affect the attendance we could stop it
- Make adds that would insinuate people with more health issues to attend since the people with no health issues has higher attendance rate than the people with all 4 health issues

Resources Used For the Project

N/A