

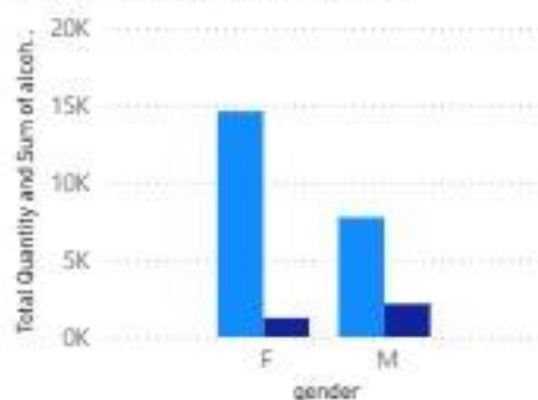
Patient Appointment Dataset Analysis Report

Project Title:

Exploratory Analysis and Visualization of Medical Appointment Data

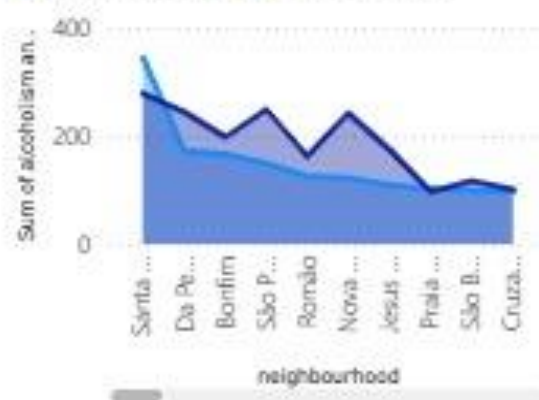
Total Quantity and Sum of alcoholism by gender

● Total Quantity ● Sum of alcoholism



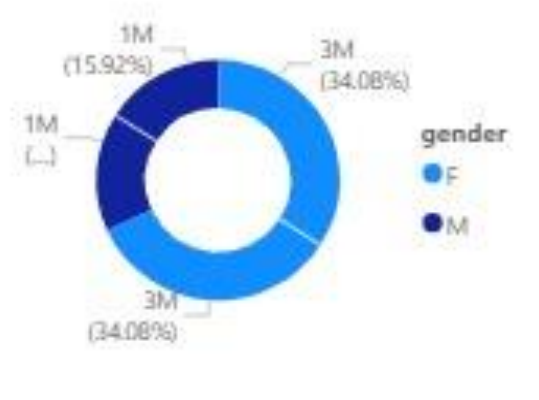
Sum of alcoholism and Sum of diabetes by neighbourhood

● Sum of alcoholism ● Sum of diabetes



Sum of age and Total Sales by gender

● Sum of age ● Total Sales



Key influencers

Top segments

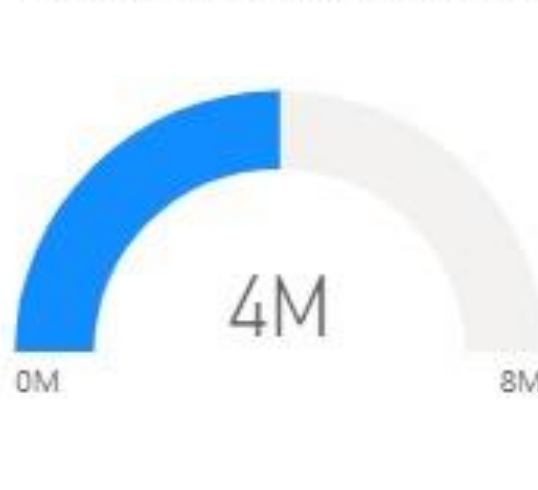
What influences sms_received to

Increase

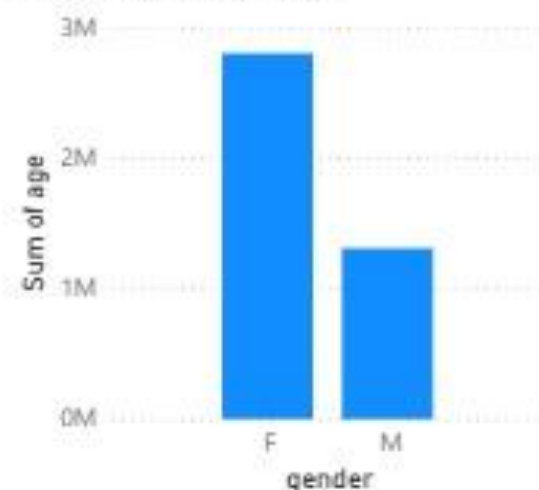
?



Sum of age and Sum of scholarship

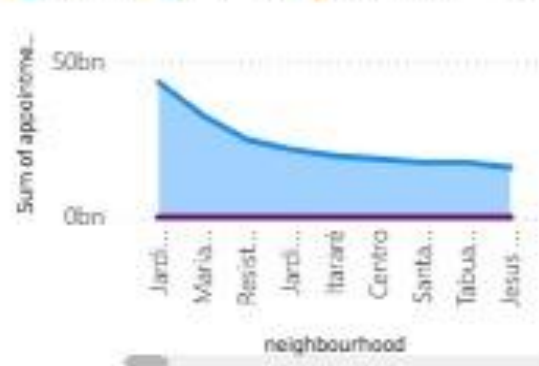


Sum of age by gender



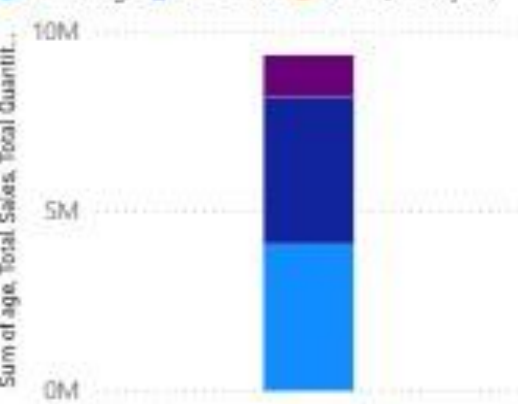
Sum of appointmentid, Sum of diabetes, Sum of handicap and Sum of hipertension by neighbourhood

● Sum of a... ● Sum of di... ● Sum of h...



Sum of age, Total Sales, Total Quantity and Sum of waiting_days

● Sum of age ● Total Sales ● Total Quantity



Total Quantity, Sum of scholarship and Sum of patientid



Sum of alcoholism, Sum of patientid, Sum of sms_received and Sum of scholarship



scheduledday	Sum of hipertension
10-11-2015 12:43:56	0
03-12-2015 13:47:28	1
07-12-2015 16:10:59	0
07-12-2015 16:12:42	1
Total	21801

Sum of age, Total Sales, Total Quantity and Sum of waiting_days

● Sum of age ● Total Sales ● Total Quantity



1. Objective

The primary objective of this project was to analyze and visualize patient appointment data to understand various factors affecting medical appointment attendance. The project aimed to:

- Identify trends in patient behavior.
- Examine demographic and medical variables like age, gender, scholarship, alcoholism, etc.
- Create an interactive Power BI dashboard for stakeholders to explore insights easily.

2. Tools and Technologies Used

- **Google Colab:** For data cleaning and preprocessing using Python (Pandas, NumPy).
- **Power BI:** For interactive dashboard creation and visualization.
- **CSV Dataset:** cleaned_data_final_ready.csv

3. Data Cleaning Process (Google Colab)

Using Python in Google Colab, the following data cleaning steps were applied:

- Removed duplicates and irrelevant columns.
- Handled missing/null values appropriately.
- Converted date columns to datetime format.
- Encoded categorical data where required (e.g., gender).
- Created derived columns for better analysis (e.g., waiting_days from appointment dates).
- Filtered out unrealistic or outlier values (e.g., negative ages, extreme waiting times).

4. Dashboard Overview (Power BI)

The Power BI dashboard consists of multiple visual tiles, each offering insights into different aspect of the dataset.

Key Visual Insights

- **Total Quantity and Alcoholism by Gender:**
Male patients showed higher counts in both total appointments and alcoholism.
- **Alcoholism and Diabetes by Neighborhood:**
Certain neighborhoods (e.g., Jd. Cambuci) have higher rates of alcoholism and diabetes.
- **Age and Total Sales by Gender (Donut Chart):**
Female patients accounted for a higher portion of the total.
- **Scholarship Distribution:**
A small percentage of patients received government scholarships, as represented in the half-doughnut gauge chart.
- **Patient Demographics by Age and Gender:**
Females outnumber males in most age brackets, with age distribution peaking in the middle-aged group.
- **Appointment Attendance and Medical Conditions by Neighborhood:**
Trends were shown in diabetes, hypertension, and disability distribution by neighborhood.
- **Waiting Days Distribution:**
High variability in the number of days between scheduling and actual appointments.
- **Influencer Analysis (Key Influencers Panel):**
While this visual was not fully populated, it was designed to identify factors influencing whether patients received SMS notifications.

5. Key Findings

- **Gender Differences:** Females have a higher appointment rate and are more likely to benefit from scholarships.
- **Medical Conditions:** Chronic conditions like hypertension and diabetes are common and vary by location.
- **Alcoholism:** Strongly correlated with certain neighborhoods and male gender.
- **Attendance Factors:** Waiting days and whether SMS was received may influence patient no-shows (requires further model-based analysis).
- **Scholarship Program:** Underutilized, though it plays a role in accessibility for some patients.

6. Recommendations

- **Targeted SMS Campaigns:** To reduce no-shows, especially in neighborhoods with higher waiting times.
- **Focus on High-Risk Groups:** Tailored healthcare interventions for neighborhoods with high alcoholism or chronic disease prevalence.
- **Improve Accessibility:** Expand the scholarship program to increase reach to underprivileged groups.

7. Conclusion

- This project successfully used Python and Power BI to transform raw healthcare appointment data into actionable insights. The dashboard enables stakeholders to explore patient demographics, medical conditions, and behavior patterns effectively. With further enhancements like predictive modeling or machine learning integration, this analysis can help optimize appointment scheduling and healthcare delivery.