Project Title

**Medical Appointment No Shows**

**Team Names:**

Mohamed Mohsen Fouad

Aya Mahmoud Mohamed

Amira Elsayed Mohamed

Aya Sherif Samir

Shrouk Abdelhamid Soliman

**Supervised By:**

Dr. Alaa Eldin Abdallah

2021/2022

Contents

[**Problem** 3](#_Toc90929824)

[**Data sources** 4](#_Toc90929825)

[**Steps to building data warehouse model** 5](#_Toc90929826)

[**DIMENSIONS and fact table** 8](#_Toc90929827)

[**Star schema** 9](#_Toc90929828)

[**Present the result** 11](#_Toc90929829)

**Problem**

Missed appointments, often referred to as no-shows,

are a problem for any clinic for a variety of reasons. If an

appointment is scheduled and not kept it could prevent

another patient who might have wanted the appointment

time from being scheduled. Missed appointments decrease

clinician productivity and may decrease income.

Example questions we would want to answer after DW from this are like below:

* What factors are important for us to know to predict if a patient will show up for their scheduled appointment?
* If patients received SMS, Are showing up?
* What is the number of patients who Show or are No-show in their appointment?

Background

Missed primary care appointments lead to poor disease control and later presentation to care. No-show rates are higher in clinics caring for underserved populations and may contribute to poorer health outcomes in this group. The objective of this project was to analyze the reasons for the patients not showing, for helping in decision making to reduce that.

**Data sources**

This dataset collects information from 100k medical appointments in Brazil and is focused on the question of whether or not patients show up for their appointment.

Several characteristics of the patient are included in each row.

Dataset Description:

|  |  |  |
| --- | --- | --- |
| Name | Value(s) | Description |
| PatienID | number | identification of a patient |
| AppointmentID | number | identification of each appointment |
| Gender | F or M | it says 'F' if female or 'M' if man |
| ScheduledDay | date | tells us on what day the patient set up their appointment |
| AppointmentDay | date | the day of the actual appointment, when they have to visit the doctor |
| Age | number | how old is the patient |
|  |  |  |
| Neighbourhood | string | indicates the location of the hospital |
| Scholarship | 0 or 1 | indicates whether or not the patient is enrolled in Brasilian welfare program Bolsa Família |
| Hipertension | 0 or 1 | indicates if the patient has hipertension |
| Diabetes | 0 or 1 | indicates if the patient has diabetes |
| Alcoholism | 0 or 1 | indicates if the patient is an alcoholic |
| Handcap | 0 or 1 | indicates if the patient is handicaped |
| SMS\_received | 0 or 1 | 1 or more messages sent to the patient |
| No-show | 1 or 0 | 0 if the patient showed up to their appointment, and 1 if they did not show up |

(Source on [Kaggle](https://www.kaggle.com/joniarroba/noshowappointments))

**steps to building data warehouse model**

**Data warehousing** is a business analyst's dream all the information about the organization's activities is gathered in one place; open to a single set of analytical tools you must plan your data warehouse system.

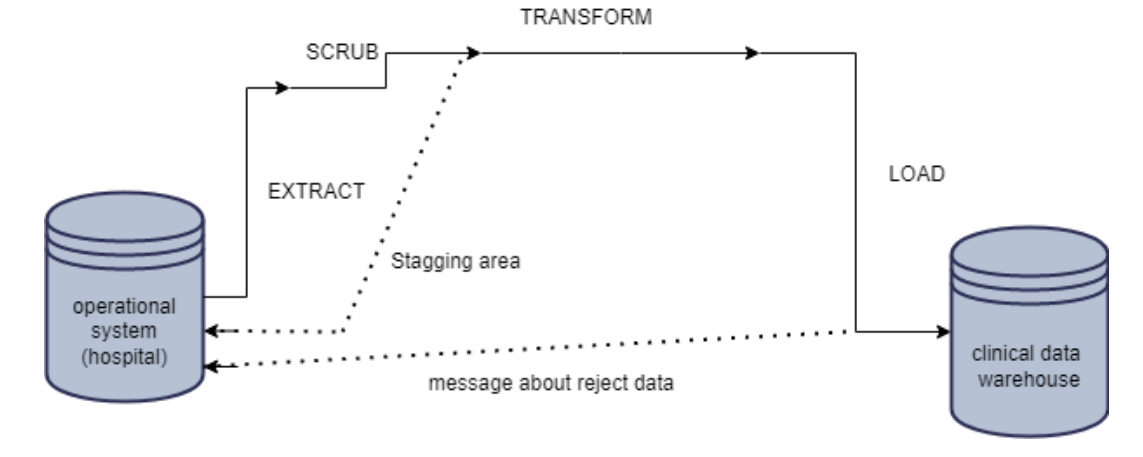
**Architectural Design of Clinical/Medical Data Warehouse:**

Data are extracted from several sources and transformed into a data staging area which is then stored in the data warehouse for further analysis. There are different stages of building informational databases (Data warehouse). They are namely known as extracting data, scrubbing, and cleaning data, transforming it into a new structure, and loading it into the medical information system’s repositories.

Diagram

Description automatically generated

**Figure 1. Data Warehouse Architecture for Medical Information System**

**Process of Clinical/Medical Data Warehouse:**

**Figure 2.Data Warehouse Building Process**

* **Data Acquisition**

Data acquisition area is the portion of a data warehouse where data are coming from various resources, move all the data into the staging area then load into the warehouse repositories. The source data are categorized into different categories and are in a different format so that they can be converted into suitable formats. It has different services and functions

**Data Extraction**

Data is taken from the database and applied the filtration technique to determine the clinical data warehouse.

**Data Transformation**

Extracted data will map to the data warehouse repository. The data mining process is applied to the database for extracting the data

**Data Loading**

The data loading function is applied when the design had been completed and a large amount of data is loaded into the warehouse. Then the data loading function will work after the data warehouse starts functioning.

* **Data Storage**

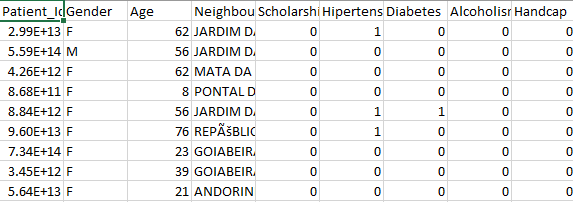
This area is used for loading the data into the warehouse daily. In the data warehouse, data are normalized and structured. In this stage, transformation and integration are completed between different data files and databases.

* **Medical Information System**

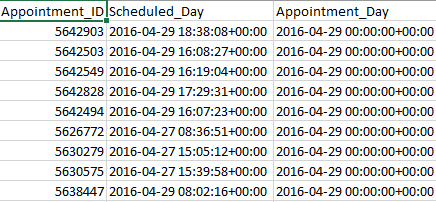
At this stage, the health information is collected from the clinical data warehouse. It is easy for doctors in the decision-making process. After locating the source of data, the data warehouse model proceeds to design the clinical data warehouse; it will adopt the star model with multiple dimensions and fact tables. Then create a valuable report for the doctors which is used for the decision process and analyze the medical information through clinical data cube.

**DIMENSIONS and fact table**

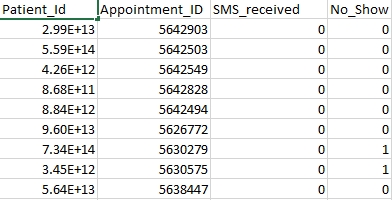
* **Dimensions Tables**
  + Patient\_Dimension



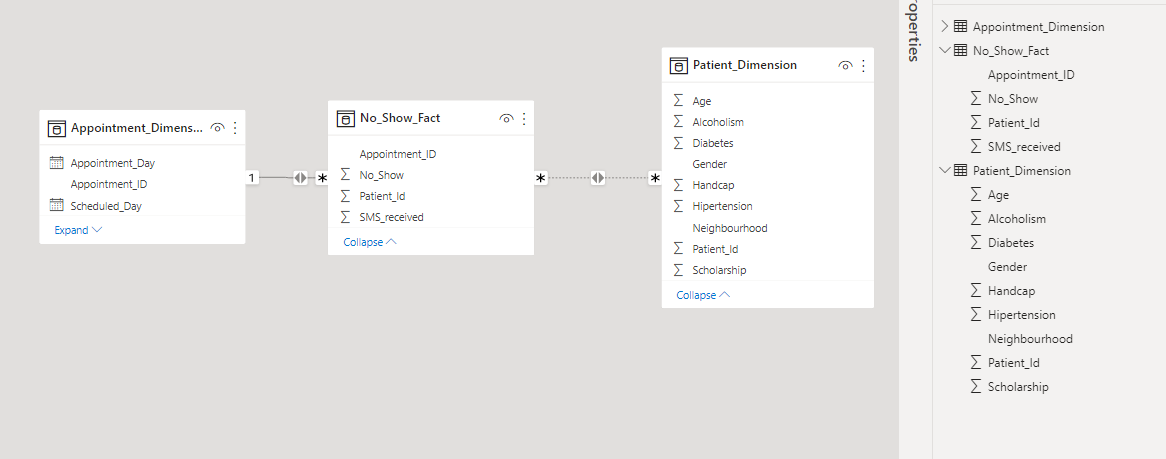
* + Appointment\_Dimension



* **Fact Table**
* No\_Show \_Fact



**star schema**



|  |
| --- |
| Patient\_Dim |
| Patient\_ID (PK) |
| Gender |
| Age |
| Neighborhood |
| ScholarShip |
| Hipertension |
| Diabetes |
| Alcoholism |
| Handcap |

|  |
| --- |
| Appointment\_Dim |
| Appointment\_ID (PK) |
| Schedule\_Day |
| Apointment\_Day |

**Star Schema Model**

1

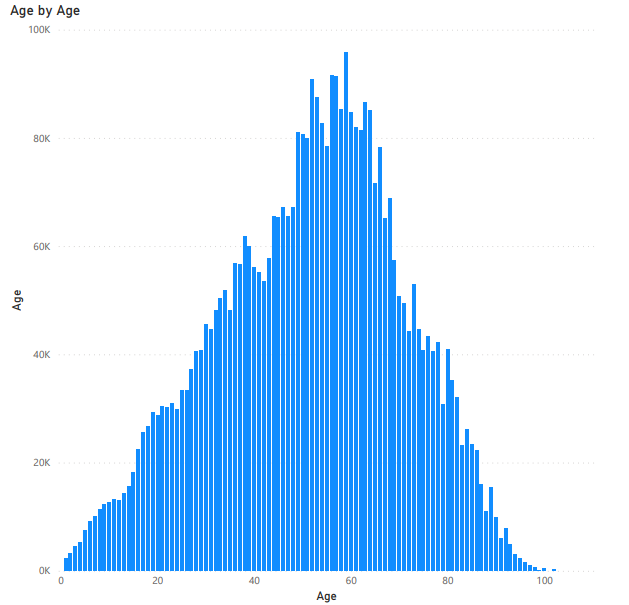
M

M

1

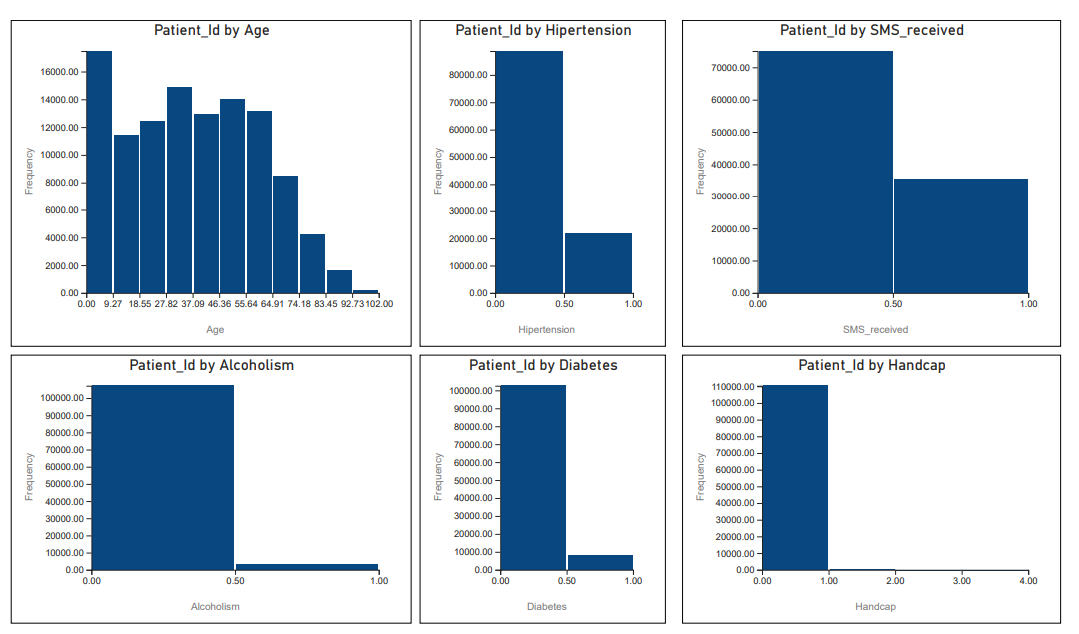
|  |
| --- |
| No\_Show\_Fact |
| Appointment\_ID (FK) |
| Patient\_ID (FK) |
| No\_Show |
| SMS\_received |

**present the result**



Observations:

The patients are 37 years on average. 25% of patients are below 18 and most of them are above 55.



Observations:

Age: There are many very young people in the dataset but in general the patient’s age is distributed evenly and the number of patients goes drastically down for patients older than 60 years.

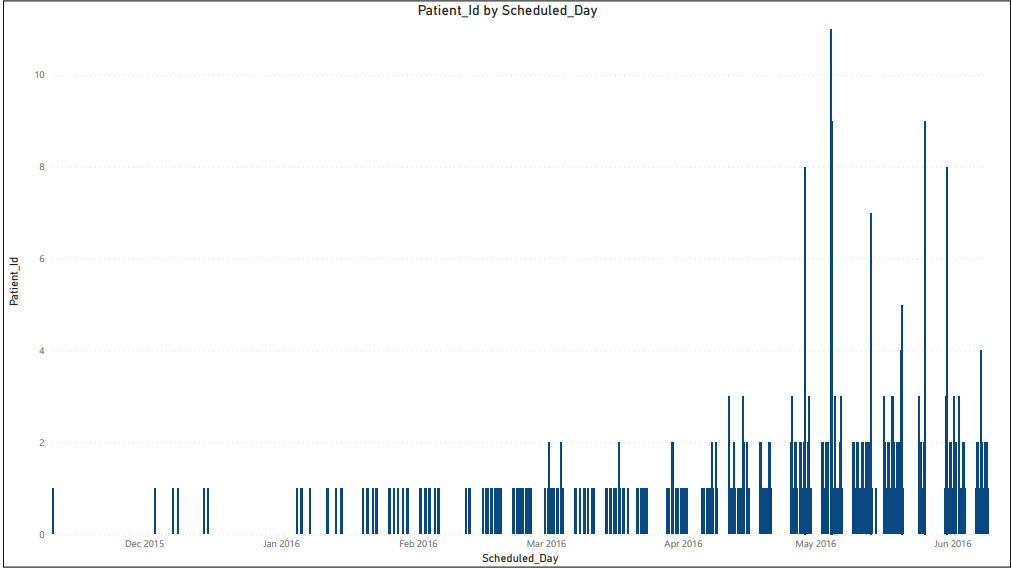
Hypertension: Most patients do not have hypertension diagnosed.

SMS\_Received: 70% of patients do not receive SMS regarding an appointment.

Alcoholism: Most of the patients are not alcoholics.

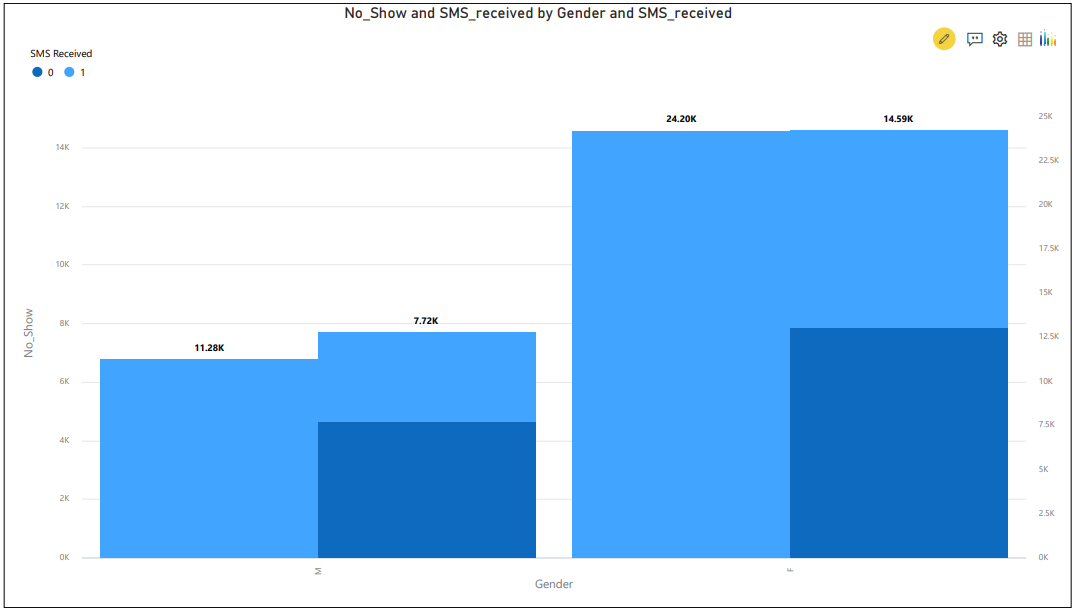
Diabetes: Most of the patients are not diabetes.

Handicap: There are four handicap categories with most of the people not being handicapped.



Observations:

Most of the schedules were between April and June of 2016.

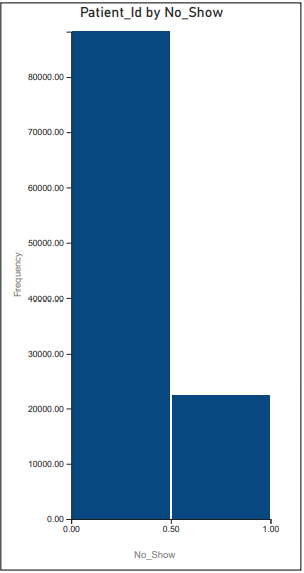


Observations:

Although the largest number of females received SMS, they missed their appointment.

I will focus on putting more light on answers to the following questions:

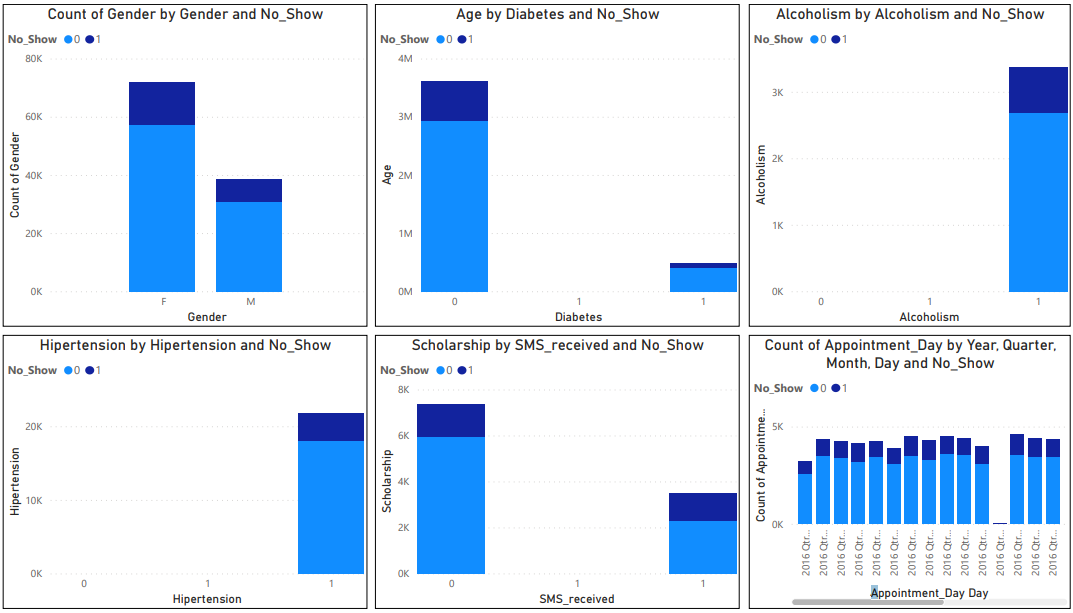
* What is the percent of patients missed their scheduled appointment?



Observations:

20% of appointments were missed.

* What factors are important to know in order to predict if a patient will show up for their scheduled appointment?



Observations:

There is no clear indication of any of these variables having bigger than others do affect show/no-show characteristics.