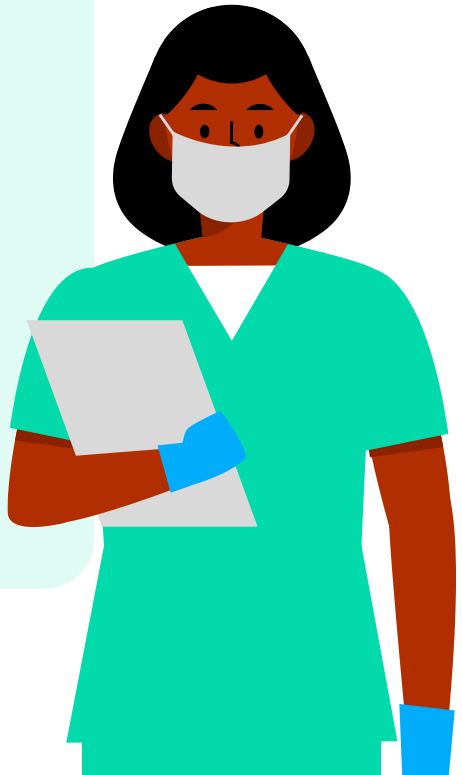
أكاديمية سدايا
SDAIA Academy

SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

Medical Appointments

Done by : Abdulrahman Aljubaylan

Business Understanding



Overview

The **2030 Vision** of Saudi Arabia seeks to Incease the Country Income that is not based on Oil that include The Health Sector where decresing the unnecessary appointments to save the lost Money and to use the lost money in other beneficial areas such as investments and Hospitals Development and Other.

Project Objective

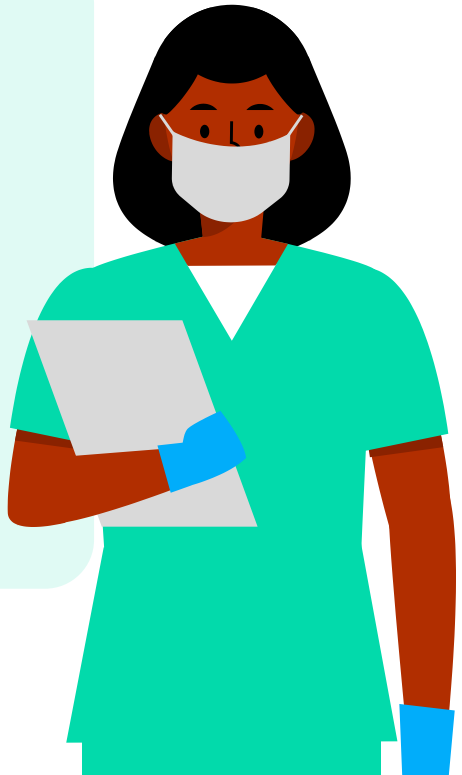
- measure the patients awareness about the health & appointments.
- find out the major factors for not Showing up for the appointment.
- give Solutions for the Main faced issue .
- measure the factors and decrease it .

Stakeholders

The Target Audience

- The Hospitals
- The Clinics
- The Insurance Companies
- Ministry of Health (MOH)
- Any related Business

Data Collection & Preparation



Dataset Overview

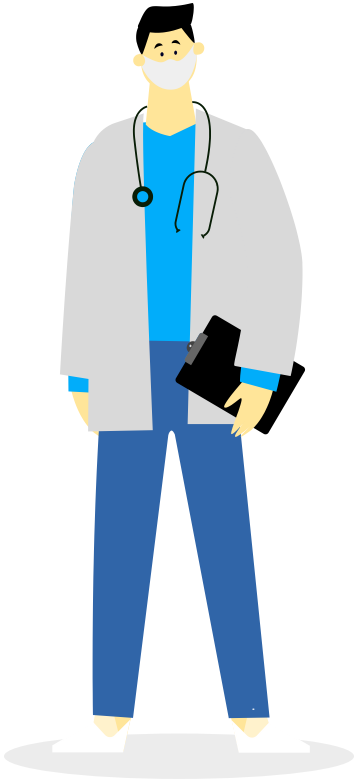
• it is a Dataset that contain the patients general Information. It consists of 14 Features , Which indicates the number of columns and 110,133 number of the observations which indicates the number of rows.

14 Feature

110,527 Record

Rows : 110,527 | Columns : 14

Dataset Features



01
PatientId
EX : 2.723732e+13

02
AppointmentID
EX : 5743405

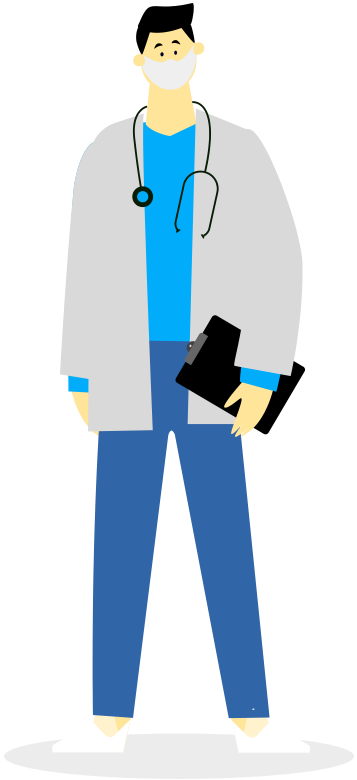
03
Gender
EX : F

04
ScheduledDay
EX : 2016-05-30T08:59:14Z

05
AppointmentDay
EX : 2016-06-01T00:00:00Z

06
Age
EX : 66

Dataset Features



07
Neighborhood
EX : ILHA DO PRINCIPE

08
Scholarship
EX : 0 | NO

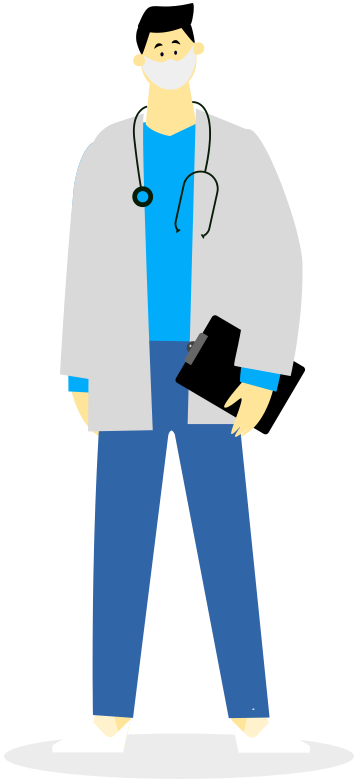
09
Hipertension
EX : 1 | YES

10
Diabetes
EX : 0 | NO

11
Alcoholism
EX : 0 | NO

12
Handcap
EX : 0 | NO

Dataset Features



13

SMS_received

EX : 0 | NO

14

No-show

EX : No

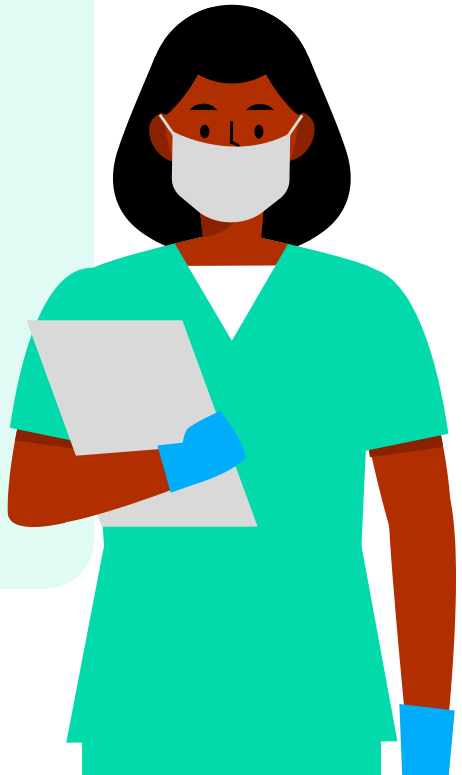
Data Preprocessing

The applied Approaches:

- Filling with zeros
- Filling with Values
- Removing Outliers
- Dropping The Null Values
- Filling with Forward Filling & Backward Filling

After conducting the different approaches, I continue with the best fit which is **Remove outlier values.**

Exploratory Data Analysis

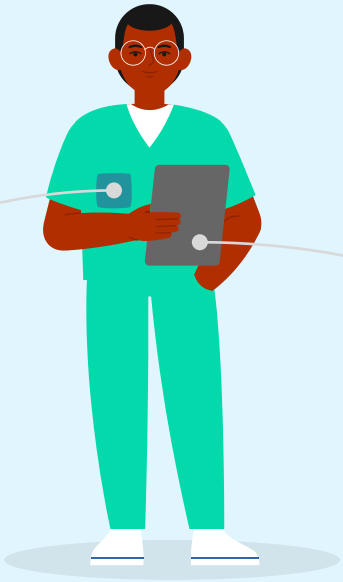


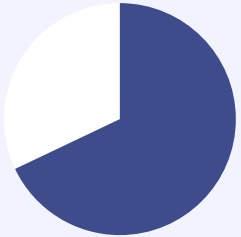
How Many are Showing up for their Appointment ?

20%

Not Showing up

22319



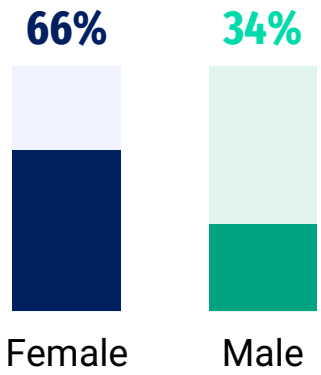


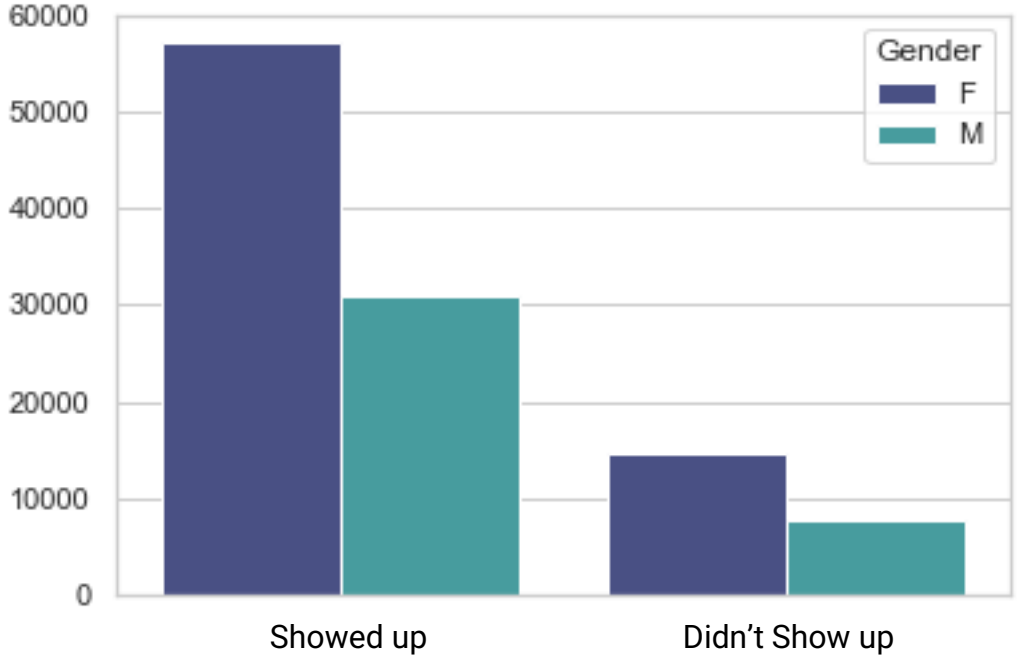
80%

Showing Up

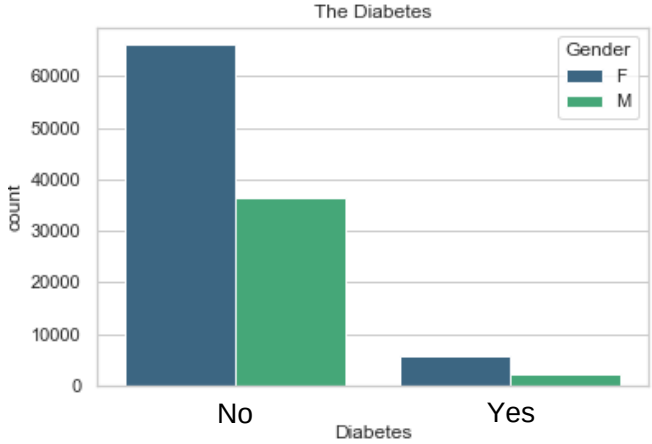
88208

The patients Gender & Appointments





The Patients Diseases



01 **Hipertension**

High blood pressure

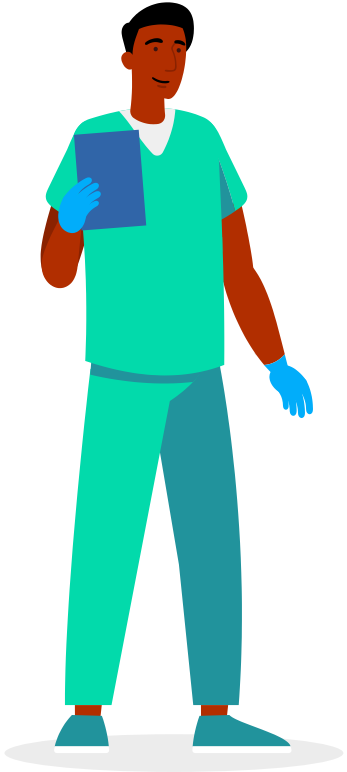
02 **Diabetes**

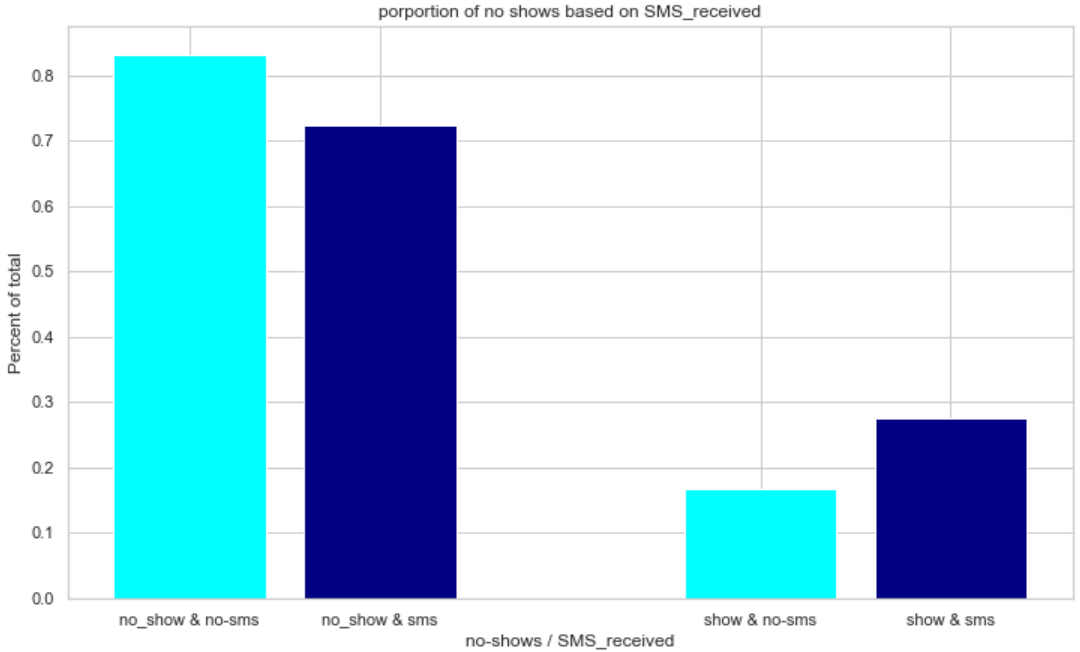
When your blood sugar goes up

03 **Handcap**

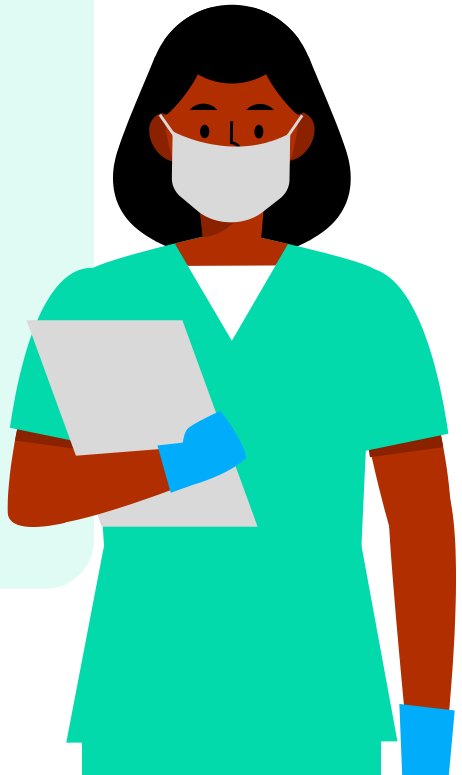
a physical or mental disability

The Relationship between the Showing up & SMS Messages





Modeling



Modeling Overview

The Dataset Name	Medical Appointment
The Target Label	Did you show in your appointment ?
The Machine Learning Type	Supervised Learning
The Modeling Category Type	Classification
The Programming Language	Python
The Main Library	Scikit-learn

Feature Engineering

The applied Mechanisms:

- Mapping

- Label encoding

- One Hot encoding

After conducting the different mechanisms and try them all, I continue with the best fit which is **Label encoding.**

Challenges & Solutions

The Challenges :

- **The label Imbalance.**
- **beating the baseline model.**

The Solution :

After Conducting many methods to fix the faced issues, I found out that the best solutions is to Go back in Data Cleaning and Preprocessing and Increase The Data Quality.

Models

The Models Results :

The Baseline	70.0%
Logistic Regression	0.79
Decision Tree Classifier	0.72
KNN	0.76
Random Forest	0.78

Conclusion

I proposed a Medical Dataset that:

1.Filled with Messy Data

2.Filled with biased Data

3.Lack of Specification

After I finished the project , I could :

1.Measure The Medical Appointments

2.Get The Main factors that Make Patients not showing up

3.Determine if a patient will show up or not

Future Work

In the near future :

- Add the project into Different blogs.

- Share The Results with the Local Community

- seek to continue to develop the project.

- Publish The work and add more Data.

- Deploy The model with Flask and make a complete application.

Acknowledgement

Mr.Modar & Mr.Moez for their time,effort and dedication for helping me. I learned a lot from them and enjoyed working with them.

SDAIA ACADEMY for supporting me and provide me with this wonderful Data Science Bootcamp that gave me the Strong Background I need so I can take my Career to the Next level.

Kaplan Team for supporting me and provide me with all the Resources I need to Increase my knowledge and work practically with different creative new projects.

Thank you