**HEALTHCARE ANYALYTICS PROJECT (GROUP 2)**

Logo, company name

Description automatically generated

**PROJECT PROPOSAL**

Appointment Show/No-show

Suhail Riyaz Ahmed

Dhruvalbhai Sunilkumar Patel

Grace Aniyankunju Mathai

Parth Patel

Apurv Hiteshkumar Sathwara

Index

[1. Introduction](#_Toc52873031) 3

[2. Motivations](#_Toc52873033) 4

[3. Evaluation](#_Toc52873034) 5

[4. Resources](#_Toc52873041) 6

[5. REFERENCES](#_Toc52873041) 6

1. Introduction

A predictive analysis would be done on the data by the team to show visual representations of how various factors can affect appointments. In the first step a data clean up will be done using SSMS and Python .Tableau will be used to do a visualization to analyze and provide valuable insights and what can be done to help the hospitals to prepare for all situations.

This data used in this is the ‘Appointmentnoshowdata’ from Kaggle. It contains the data for medical appointments made by patients and other information and characteristics about the appointment, this data contains 13 columns. A portrayal of the data is given below in the table.

|  |  |
| --- | --- |
| **Column** | **Description** |
| PatientId | Patient id |
| AppointmentID | Unique ID for each appointment |
| Gender | Patients gender |
| Age | Patients age |
| AppointmentDate | The Date of appointment |
| Scholarship | Indicates if the patient is enrolled in a welfare program |
| Hypertension | Tells if the patients have hypertension |
| Alcoholism | Tells if the patients are suffering from alcoholism |
| Diabetes | Tells if the patients have diabetes |
| Handicap | Tells if the patients are handicap |
| SMS-received | Tells if the patients received SMS. |
| Neighborhood | The neighborhood in which the hospital is located. |
| No-show | Indicates NO if the patient doesn’t show up to their appointment |

1. Motivation

The motivation for this project is to determine how often or what kind of patients do not show up for the schedule appointments. We as a team, selected this data to examine what are the factors or reasons that patients do not show up for their appointments.

This project can help the hospital to prepare which patients are likely to cancel their appointments, which months to expect the appointment number to rise so that the hospital can be prepared. The motivation of this project was the rise of no-show in the appointment, after coming across a lot of questions like

* Which months are the most patients coming ?
* Which season has the highest impact ?
* Does any of the illness play a role in cancellation ?

We took this to explore how a hospital prepares for situations as such how they manage their resources and help the patients to receive better care.

1. Evaluation

We would evaluate the project on the basic of Show/No-show of appointments .

A tableau dashboard will be prepared to have a visual through which a predictive analysis would be done to see the reason which affect the cancellation what are the reason for appointments to rise.

A logistic regression model will help us to measure the success rate to identify the reasons which tells us why the patients don’t show up for the appointment.

Below are some Dummy visualization we would try to achieve and enhance it which can help us with the results and insights we could provide.

A picture containing chart

Description automatically generated

A picture containing chart

Description automatically generated

1. Resources

* Excel
* Tableau
* Python
* Kaggle Dataset Medical Appointments No-Show (<https://www.kaggle.com/datasets/joniarroba/noshowappointments>)
* Tableau prep builder
* SQL Server Management (SSMS)

1. References

* Data set directory, Kaggle,Data.world,Data.medicare and Government of Canada Free Dataset was explored for multiple data sets and to gain insights about different dataset: (<https://www.kaggle.com/datasets>)
* Generic Proposal Template for headers and content: [Generic Project Proposal Template - Casual.pm](file:///C:\Users\Grace%20Mathai\Downloads\Generic%20Project%20Proposal%20Template%20-%20Casual.pm)
* Source control was practiced in the team for the document sharing and updates – helped a lot in tracking changes and sharing insights to all team members: <https://github.com/>
* Spreadsheets were extensively used for EDA analysis – for Data derivation, Outliers, Graphs, and initial visualizations: Microsoft Excel