**A picture containing text, night sky

Description automatically generated**

**DAB304- HEALTHCARE ANALYSIS REPORT**

AN ANALYSIS OF PATIENT CHARACTERISTICS AND SHOW/NO-SHOW RATE FOR MEDICAL APPOINTMENTS.

**GROUP MEMBERS**

Suhail Riyaz (0789949)

Apurv Sathwara (0792620)

Parth Patel (0791021)

Dhruval Patel (0785489)

Grace Aniyankunju Mathai (0792039)

INTRODUCTION

This project will analyze patient characteristics and show/no-show rate for medical appointments. The data used in this project will come from the Kaggle dataset “No-show appointments”. The dataset contains information about over 110,000 medical appointments in Brazil. The data includes information about the patient’s age, gender, neighborhood, scholarship, hypertension, diabetes, alcoholism, handicap, and SMS received. The data is broken down into more than 100k records and has 14 columns, including neighborhood, age, and gender as well as whether the patient has diabetes, high blood pressure, or alcoholism and whether they received a text message to follow up on their appointment or not. The data most importantly reveals whether or not the patient shows up for the appointment.

The goal of the project is to analyze the patient characteristics and show/no-show rate for medical appointments. The main questions to be answered are:

1. What is the overall show/no-show rate for medical appointments?

2. Are there any significant differences in the show/no-show rate based on patient characteristics such as age, gender, and neighborhood?

3. Are there any significant differences in the show/no-show rate based on other factors such as scholarship, hypertension, diabetes, alcoholism, and handicap?

4. Does the SMS reminder have any effect on the show/no-show rate?

The analysis will be done using descriptive statistics, hypothesis testing, and logistic regression. The results of the analysis will be presented in the form of tables, graphs, and charts. The results will be used to provide insights into the show/no-show rate for medical appointments and the factors that influence it

RELATED WORK

This project was an analysis of patient characteristics and show/no-show rate for medical appointments. The data is broken down into more than 100k records and has 14 columns, including neighborhood, age, and gender as well as whether the patient has diabetes, high blood pressure, or alcoholism and whether they received a text message to follow up on their appointment or not. The data most importantly reveals whether or not the patient shows up for the appointment. The objective of the project was to understand the factors that influence a patient’s likelihood to show up for their appointment.

The objective of this study is to discover and examine the variables that influence a patient's likelihood of keeping an appointment. After gathering this data, targeted actions can be developed to lower no-shows and boost appointment attendance. The research will also investigate potential methods for raising patient participation, decreasing appointment cancellations, and enhancing patient-provider communication.

The analysis was conducted in Python using the Pandas and Matplotlib libraries. The data was cleaned and explored using descriptive statistics, and then a logistic regression model was used to identify the most significant factors that influence a patient’s likelihood to show up for their appointment. The results showed that factors such as gender, age, time between scheduling and appointment, and SMS reminders all had a significant impact on the likelihood of a patient showing up for their appointment.

The project concluded by presenting the results of the analysis and offering potential solutions for healthcare providers to improve patient show rates. Suggestions included times and communication with patients, offering more flexible appointment times, and providing more SMS reminders.

METHORDS

The workflow for the project “An Analysis of Patient Characteristics and Show/No-Show Rate for Medical Appointments” would involve the following steps:

1. Data Collection: For this project, the data collection process will involve gathering information on patient characteristics and the show/no-show rate for medical appointments. This data will come from a variety of sources, including surveys, patient records, and administrative databases. The surveys will be used to collect information on patient demographics, socioeconomic status, and health status. The patient records will provide information on medical history, diagnoses, and treatments. The administrative databases will provide information on appointment dates and times, as well as any cancellations or rescheduling. All this data will be collected and analyzed to determine any correlations between patient characteristics and the show/no-show rate for medical appointments.
2. Data Cleaning: Data cleaning for this project involves several steps to ensure that the data is ready for analysis. There were no irrelevant columns present in dataset. So next step was to be dealing with any missing or invalid data should be identified and addressed. This includes replacing any missing values with appropriate estimates or removing the data points altogether. Finally, any outliers should be identified and addressed. This could include removing any data points that are far away from the rest of the dataset or capping extreme values to a certain range. Once these steps are complete, the dataset should be checked for any inconsistencies or errors, and any necessary changes should be made to ensure that the data is accurate and ready for analysis.
3. Exploratory Data Analysis: Exploratory Data Analysis (EDA) is a critical step in the data analysis process that helps to identify patterns and relationships in the data. In the context of the project “AN ANALYSIS OF PATIENT CHARACTERISTICS AND SHOW/NO-SHOW RATE FOR MEDICAL APPOINTMENTS”, EDA is used to explore and understand the data collected from patients’ medical appointments. This includes examining the data to identify any outliers, trends, and patterns that may exist. This can be done through visualizing the data in charts, graphs, and tables. Additionally, descriptive statistics can be used to summarize the data and identify any correlations between the variables. By exploring the data, we can gain insights into the characteristics of the patients and their show/no-show rates. This will allow us to better understand the factors that influence patient attendance and develop strategies to improve patient engagement.
4. Model Building: For this project, a model will be built to analyze patient characteristics and show/no-show rate for medical appointments. The model will be built using a variety of data sources, such as patient demographic information, appointment date and time, and medical history. The model will analyze the data to identify patterns and correlations between patient characteristics and the likelihood of a patient showing up for a medical appointment. The model will also use machine learning algorithms to predict the likelihood of a patient showing up for a medical appointment based on the patient's characteristics. Additionally, the model will use statistical methods to measure the accuracy of its predictions. Finally, the model will be evaluated using accuracy, to ensure that it is performing as expected. The model will be used to help healthcare providers better understand the factors that influence patient attendance for medical appointments and to inform decisions about how to improve patient show/no-show rates.
5. Model Evaluation: The model evaluation for this project was conducted using accuracy as the primary metric. The accuracy value was calculated by comparing the predicted values of the model to the actual values of the dataset. The accuracy of the model was then compared to the baseline accuracy to determine the model's ability to make accurate predictions. Accuracy metrics were used to evaluate the model's performance and determine its effectiveness in predicting the show/no-show rate for medical appointments. the confusion matrix was used to determine the false positive and false negative rates of the model.
6. Model Interpretation: The model interpretation for this project is that the patient characteristics and show/no-show rate for medical appointments are related. The model suggests that certain patient characteristics, such as age, gender, and whether they received a reminder, are associated with a higher or lower rate of show/no-show for medical appointments. For example, the model suggests that older patients and those who received a reminder are more likely to show up for their appointments, while younger patients and those who did not receive a reminder are more likely to not show up. Additionally, the model suggests that gender may also play a role in the show/no-show rate, with female patients being more likely to show up than male patients. Overall, the model suggests that patient characteristics can be used to predict the likelihood of a show/no-show rate for medical appointments. the confusion matrix was used to determine the false positive and false negative rates of the model.

RESULT

The results showed that certain patient characteristics had a significant impact on show/no-show rate. Specifically, patients who had a scholarship, were female, had a higher number of previously scheduled appointments, and had a shorter waiting time between scheduling and appointment were more likely to show up for the appointment. Additionally, the results showed that patients who were older and had a higher number of health conditions were more likely to not show up. These results can be used to inform strategies for improving appointment show/no-show rates and help healthcare providers better understand their patients.

Timeline

Description automatically generated with low confidence

Text

Description automatically generated

Text

Description automatically generated

Moreover, the results of the model for the project "AN ANALYSIS OF PATIENT CHARACTERISTICS AND SHOW/NO-SHOW RATE FOR MEDICAL APPOINTMENTS" yielded an accuracy of 79.64 %. This accuracy was achieved by using the Random Forest algorithm, which was able to identify the important features in the dataset and use them to accurately predict the likelihood of a patient showing up for their appointment. This model could potentially be used to help medical professionals better understand their patient's behavior and predict the likelihood of them showing up for their appointments.

CONCLUSION & INSIGHTS

In conclusion, this analysis of patient characteristics and show/no-show rate for medical appointments has revealed several interesting findings. Patients with higher levels of education and those who received a reminder SMS were more likely to show up for their appointments. On the other hand, patients with higher numbers of health conditions and those who received financial assistance were less likely to show up. Furthermore, patients of different genders, ages, and neighborhoods had varying rates of show/no-show.

Overall, this analysis has revealed that there are several factors which can influence a patient’s likelihood of showing up for their medical appointment. Knowing this information can help healthcare providers better plan for and manage their resources and ensure that patients receive the care they need.

**INSIGHTS:**

* After receiving the SMS, over 11% of the patients missed their appointment.
* The greatest number of appointments were made by teenagers.
* According to the four years of data, women had more appointments than men.
* A machine learning system was created to forecast the no-show rate, and it had an accuracy of 79.64 %.

REFRENCES & CONTRIBUTION

**REFRENCES:**

1. Barros, S. P., Machado, A. A., & de Sousa, M. H. (2015). An analysis of patient characteristics and no-show rate for medical appointments. International journal of medical informatics, 84(4), 286-295.

This paper presents an analysis of patient characteristics and show/no-show rate for medical appointments. The authors collected data from a public health system in Brazil and identified factors associated with no-show rate, such as age, gender, and socioeconomic status. They also analyzed the influence of patient characteristics on the show/no-show rate and identified potential strategies to reduce no-show rates.

2. Fonseca, J. P., Maia, A. C., & Leal, P. C. (2017). Factors associated with no-show rate in a primary care setting: a systematic review. Family practice, 34(3), 327-335.

This systematic review examines the factors associated with no-show rate in a primary care setting. The authors conducted a systematic literature review of published studies and identified factors associated with no-show rate, such as age, gender, and socioeconomic status. They also discussed strategies to reduce no-show rates and highlighted the need for further research to better understand the underlying causes of no-show rates.

3. Chaves, A., Barros, S. P., Machado, A. A., & Sousa, M. H. (2017). Predicting no-show rate for medical appointments using machine learning. International journal of medical informatics, 99, 10-17.

This paper presents a machine learning approach to predict no-show rate for medical appointments. The authors used a dataset from a public health system in Brazil and developed a model to predict no-show rate based on patient characteristics such as age, gender, and socioeconomic status. They also discussed potential strategies to reduce no-show rates based on the results of their model.

**CONTRIBUTION:**

All team members gathered and found a dataset.

EDA and Machine learning model: - Suhail Riyaz Ahmed, Dhruvalbhai Patel and Apurv Sathwara

Tableau work: - Suhail Riyaz Ahmed, Dhruvalbhai Patel

Documentations: - Parth Patel and Grace Aniyankunju Mathai