A STUDY TO REDUCE THE NUMBER OF PREVENTABLE EMERGENCY VISITS AT COMMUNITY LEVEL

A Thesis

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Health Informatics

---

by

ANUSHKA KANOONGO

Dr. Patricia Alafaireet, Thesis Supervisor

MAY 2023

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled,

A STUDY TO REDUCE THE NUMBER OF PREVENTABLE EMERGENCY VISITS AT COMMUNITY LEVEL

presented by Anushka Kanoongo, a candidate for the degree of Master of Science [Health Informatics], and hereby certify that, in their opinion, it is worthy of acceptance.

____

Professor Patricia Alafaireet

____

Professor David Moxley

____

Professor Sue Boren

DEDICATION

I would like to take this time to thank my parents and family for their throughout support in my endeavors. I want to extend special thanks to my mom for being there for me even at the odd times of the day. I am grateful to my friends who constantly cheer me up and be there for me.

I am grateful for this opportunity and all the people who directly or indirectly helped me in completing this study successfully.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to Dr. Patricia Alafaireet, my thesis advisor, for her support and guidance throughout this research and graduate studies. I am grateful to her for the opportunity to do research and providing me with invaluable guidance. Her leadership has deeply inspired me. It is a great privilege and honor to learn from her.

I am thankful to all my faculties, especially Dr. Sue Boren, Dr. Peter Tonellato, Dr. Mihail Popescu, Dr. David Moxley, and Dr. Brian Hensel for teaching me valuable aspects of health informatics which helped in the completion of this study.

I would also like to thank my committee members for letting my defense be an enjoyable moment and for their brilliant comments and suggestions.

Last but not the least, the entire Health management and Informatics department for their efforts to make this an enjoyable experience.

Thank you!

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

3A’s	Accessibility, Affordability, Availability
AHRQ	Agency for Healthcare Research and Quality
CDC	Centers for Disease Control and Prevention
CMS	Center for Medicare & Medicaid Services
ED	Emergency Department
EHR	Electronic Health Record
ER	Emergency
HCUP	Healthcare Cost and Utilization Project
HHS	U.S. Department of Health and Human Services
NCHS	National Center for Health Statistics
NEDS	Nationwide Emergency Department Sample
RCCCP	Randolph County Caring Community Partnership

ACADEMIC ABSTRACT

Introduction: Emergency Department overcrowding is a worldwide issue. The impact of increased unnecessary and preventable emergency visits in hospitals leads to reduced quality of care, inability to care for critically ill patients, increased errors, and mortality rates. The study aims at identifying ways to predict avoidable ER visits through implementing machine learning algorithms and intervening them at a community level.

Materials and Method: The data was collected on a community level Electronic Health Record (CCMO) database and was provided for further investigation by the Randolph Caring Community Partnership (RCCCP). The exploratory data analysis was conducted in the RStudio and later the machine learning models were implemented in the jupyter notebook. 14 features were selected out of the 43 features through step wise logistic regression method to predict the ER visits.

Result: The data of 595 patients was collected during the period of 2018 to 2021. The AUC score of the Decision Tree and the logistic regression model was 0.54 and 0.61 respectively. The support vector machine had the accuracy of 0.58 in predicting ER visits.

Conclusion: The study depicted that it is possible to have prediction models at the community level to reduce the burden at emergency department. A more focused data collection specific for the prediction of unnecessary ER visit will be able to produce a more feasible model.

INTRODUCTION

Overcrowding is one of the issues faced by many emergency departments around the world (Morley et al., 2018). The Centers for Disease Control and Prevention (CDC) has reported a significant growth in the number of emergency visits per year from 2010 to 2016. The emergency visits have seen a cumulative growth rate of 7.51% and compound annual growth rate (CAGR) of 1.21%. The overall rate is higher than the population cumulative growth rate of 4.46% and CAGR of 0.73%. The average cost of treating 10 common primary care treatable conditions at a hospital ED is $2,032, more than $1,800 higher than in primary care settings (United Health Group, 2019). Emergency Department visit cost 12 times higher than at a physician office or 10 times higher than at an urgent care (United Health Group, 2019). An estimated 13% to 27% of ED visits in the United States could be managed in physician offices, clinics, and urgent care centers, saving $4.4 billion annually (Weinick, et al., 2010; Tapia et al., 2022).

Statistically, females have higher rate of utilizing emergency services (HHS, 2021). The rural population are among the highest number of treat-and-release visits to the emergency department (HHS, 2021). The Department of Health and Human services also concluded in their report that the age group 18-44 visit to an Emergency Department in the non-mental health/ substance use treat and release category and mental health/ substance use category. It is conceding with the fact that these users are more in the small metropolitan or micropolitan areas (HHS, 2021). The highest number of ER visits are found to be in the most vulnerable areas (HHS, 2021). The Department of HHS and CMS believes that targeting the population who need care for behavioral health problems and ensuring enough primary care providers will certainly aid in reducing the number of non-urgent emergency visits.

The HHS report also indicates that community services and community health centers can significantly reduce the number of preventable ER visits while focusing on the social determinants for better care delivery. It is shown that the rural counties with a community health center have less ED use among the ones that doesn't have a community center (HHS, 2021).

BACKGROUND

The Randolph County, Missouri is 53 of 115 county in the clinical care factors as per Missouri Health Statistics (Missouri Health, 2022). The Clinical Care Rank as described by MOHealth is "a weighted combination of Uninsured, Primary Care Physicians, Dentists, Mental Health Providers, Preventable Hospital Stays, and Mammography

Screening factors”. There is one primary care physician per 2751 persons (Missouri Health Atlas, 2021). Fifteen percent of the population under the age of 65 does not have a health insurance as per the report (Missouri Health, 2022). The county have 1 mental health provider for 1076 persons and 1 primary care provider per 1285 persons (MOHealth, 2022).

The county recorded 4796 Emergency visits per 100,000 Medicare enrollees in the year of 2022 (MOHealth, 2022). The median age of female and male is 40 and 38 respectively (US. Census Bureau, 2017-21). The Randolph county ranks 87 out of 115 for the Health Behaviours. MO Health describes Health Behaviours as “ a weighted combination of Adult Smoking, Adult Obesity, Food Environment Index, Physical Inactivity, Access to Exercise Opportunities, Excessive Drinking, Alcohol-Impared Driving Deaths, Sexually Transmitted Infections, and Teen Birth factors”.

LITERATURE REVIEW

Data Sources

We searched Medline (1946–2022) for eligible journal articles and randomized controlled trials using combinations of the following search terms: (1) Emergency Service (medical subject headings (MeSH)), unnecessary emergency visit, Telemedicine (MeSH), Preventable emergency visit (MeSH), or reducing unnecessary visit; (2) Decision Support System, Information Technology (MeSH), Internet (MeSH), Clinical Information System, Medical Informatics; (3) Models, Statistical, Algorithms, Machine Learning, Predictive Modeling Models, Theoretical. We also systematically searched the reference lists of included studies and relevant reviews.

Study Selection and Data Extraction

The investigator reviewed the titles and abstracts of the identified citations and identified eligible articles based on the following criteria. Inclusion criteria included any randomized controlled trial or journal article. The included studies were identifying ways to deal with the emergency department overcrowding. Studies published in English language were included. The investigators collected the following information from each article that was eligible: descriptions of the study design, country, sample size, duration of study, technology used, cohort identification process, success rate, limitations, application.

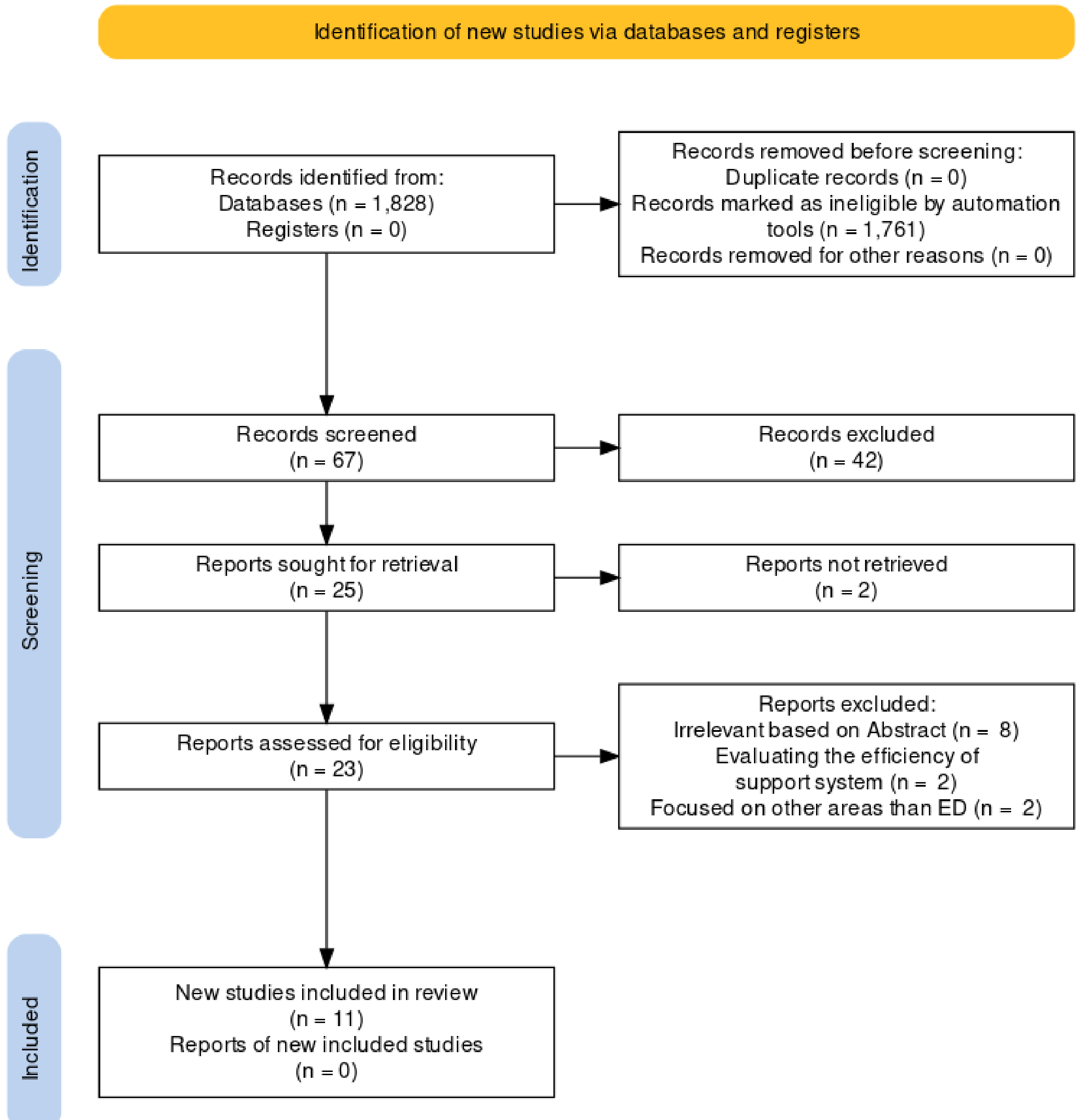


Figure 1. PRISMA Flow Chart - Study selection

A total of 1,828 research articles were identified through Medline search of the terms. Out of which 1,761 articles were deemed only slightly related to the topic of interest and hence, were removed from the selection. After careful review of title, abstract, and objective of the 67 studies, 11 articles were selected for this review. (Fig 1)

Study Characteristics

The majority of the selected studies were conducted in the United States (45%). The selected studies were equally distributed between retrospective study and controlled trials (27%). All the studies were published between the years of 2014 and 2022 showing that the ED overcrowding has been an issue for quite a while. The maximum proportion of these studies addressed the solutions of emergency overcrowding (90%) or consequences of ED crowding (54%) whereas only 27% of the articles were focused on establishing an association between the causes of ED overcrowding. Only 3 articles were focused on informatics or community services-based solutions for emergency overcrowding.

Impact of ED overcrowding on patients:

Tapia and Salway researched the consequences of ED overcrowding on patients at length. The research study suggests that overcrowding impacts the quality of care, increases the chances of medical errors, and increases mortality rates (Salway, 2017). It could also mean that some critically ill patients might not be able to have access to emergency care in time. One of the reasons behind overcrowding at the emergency department is the unavailability of inpatient beds. It leads to an increased length of stay in the emergency. Di Somma et al. (2014) showed evidence that patient satisfaction is positively associated with the length of stay in the department.

Impact of ED overcrowding on workforce:

Tapia et al. (2017) argued that ED overcrowding negatively impacts the staff as they are engaged for more extended periods than required when 60% of emergency visits are deemed unnecessary. The resources are limited and should use effectively, whereas the issue puts an extra burden on the physicians and nurses to provide better care in less time. It compromises the quality of care and leads to increased medical errors. It is also one of the leading causes of workforce burnout. (Di Somma, 2014)

Table 1

Literature Review

Author/Country/Year	Study Design	Aim	Sample	Method Summary	Outcome Measure	Findings Summary
Tapia et al. / USA / 2022	Retrospective Chart Review	We seek to determine the proportion of patients that visited the ED at BAMC that were primary care eligible and would presumably be readily manageable in an appointment-based setting.	84000 visits during a period of 1 year (2019-2020): pre-covid and post-covid	Standard Deviation, Mann Whitney U-test and Chi-Square test.	Whether each patient could be managed in a non-ED setting (primary care eligible)	1. 3 out of 5 patients were categorized as primary care eligible. 2. The unnecessary emergency visits are costly for the healthcare. Especially puts extra burden on the workforce. 3. There is a decrease in total number of patients post-covid. 4. Overcrowding affects the quality of services as the staff and equipment are engaged for longer period than required.

Salway et al. / USA / 2017	Descriptive Research Study	This paper will endeavor to answer some of the vital questions concerning ED overcrowding and propose some possible solutions to this critical issue.		Observational Method, Literature Review		1. Admitted patients are not timely transferred due to bed unavailability resulting in increased number of patients in Emergency department 2. Inability to care for the sick patient is a major consequence of ED overcrowding. Other consequences include increased LOS, increased walkouts, reduced quality of care, and increased medical errors.

						3. Solutions can include improved bed management, a provider conducting the triage, early discharge, smoothing of elective admissions.
McKenna et al. / USA / 2019	Descriptive Research Study	This paper discusses the causes of emergency department overcrowding, provides a brief overview of the drastic consequences, and discusses possible cures that have been successfully implemented.		Observational Method, Literature Review		1. Solutions targeted at the emergency department flow are not as effective because of the longer inpatient waiting period. Increasing the size of ED does not reduce the strain on the limited workforce. 2. Targeting the following areas – elective

						admissions smoothing, early and weekend discharges, capacity plans could decrease the load on the workforce and overcrowding of ED.
Morley et al. / Australia / 2018	Systematic Review	The aim of this systematic review was to critically analyze and summarize the findings of peer-reviewed research studies investigating the causes and consequences of, and solutions to, emergency department crowding.	102 research studies published between January 2000 and June 2018	Literature Review	Only articles investigating the causes, consequences, and solutions for ED overcrowding.	1. Consequences reported in the study are adverse outcomes for patients such as treatment delays, increased mortality, and poorer pain management. 2. Some of the solutions mentioned are early physician assessment in the ED, opening general practitioners’

						clinics, increase access to primary care, and more effective communication.
Burton et al. / UK / 2021	Retrospective Study	Hypothesized that a regional urgent and emergency care system would show typical features of a complex system: specifically, it would show a stable pattern of frequent attendance which followed a power law distribution.	Healthcare data collected from the ED in the region covering over 5.5 million residents from year 2014 to 2017.	Measure of Deprivation, power law scaling and Kolmogorov-Smirnoff test	P-value > 0.05: distribution indistinguishable from power law.	1. The frequent ED visitors represent one continuous and uninterrupted distribution of attendance and should be treated like one. 2. The overcrowding requires a system as well as individual solutions.
Hanlon et al. / USA / 2022	Retrospective Observational Study	The objective of this retrospective, observational study was to assess the mediating	EHR of 147,496 adults from year 2015 to 2017.	Descriptive statistics, Causal mediation analyses, social	Assessment of resource utilization outcomes	1. There is a positive correlation between social vulnerability and number of

		effect of medical complexity on the relationship between social vulnerability and four acute care resource use outcomes—number of hospitalizations, emergency department (ED) visits, observation stays, and total visits.		vulnerability index		hospitalizations.
Di Somma et al. / Italy / 2014	Manuscript	To provide valuable contributions in the understanding of ED crowding solutions.				1. Patient satisfaction is positively associated with the length of stay in ED 2. ED overcrowding leads to increased medical errors, mortality.

						3. Overcrowding is not only internal but also a public health crisis. 4. Rule of having 98% ED patients to be seen, admitted, or discharged within first 4 to 6 hours is effective. However, it compromises quality and burdens the workforce at times. 5. Development of Acute Medical Units (AMUs) is also a viable solution.
Yarmohammadian et al. / Australia / 2016	Systematic Review	The objective of this review article is to present strategies with an important role in the	30 articles were included.	Literature Review	Assessed articles with real patient flow in ED.	1. Involving physician in the triage team as a leader. 2. Introduce alternative

		improvement of patient flow, delay in services, and overcrowding of the ED.				short-stay units such as fast track, ambulatory area, minor injury units.
Belmin et al. / France / 2022	Pragmatic Randomized Clinical Trial	To implement a system that produces alerts when the machine learning algorithm identifies a short-term risk for an emergency department (ED) visit	206 patients (mean age 85) and 109 Home aides	The home care aide will fill the eHealth app for each patient during their visit. The app will estimate risk of admission in ED in next 14 days. The HA notifies the nurse and caretaker for closer supervision to avoid ED visit.	F – Score and P-Value < .05 was considered significant.	1. An alert triggering eHealth system might be effective in reducing the ED visit (P<.001, Sensitivity - 83%, Usability of the eHealth system - 90%)
Enard et al. / USA / 2014	Quasi experimental study	To implement and assess the impact of patient navigators for the Primary care related	13,642 patients between 18 years to 65 years old	Once the triage deduces the emergency level as low, primary care eligible;	Odds Ratio, continuous mean measure of visits and cost. (p<.05)	1. They observed significant decline in the mean visits and cost.

		ED visits to reduce preventable visits.	were included.	patient navigators communicate with patient to understand access barrier and solve any issue to the best of their capabilities.		
Pierssens et al. / Canada / 2022	Cohort Observational Study	To assess the safety of a redirection process of low-acuity ED patients to a nearby clinic using an electronic clinical support system that helps patient identification and appointment scheduling.	2140 low acuity patients were included in the study	Questionnaires were filled and analyzed through descriptive statistics.	Rate of redirected patient returning to any ED within 48 hours and within 7 days. Satisfaction rates.	1. Only 2.8% patients returned unexpectedly within first 48 hours. 2. 4.8% patients were returned within 7 days. 3. The satisfaction rate of redirected patients was high. 4. The redirection

							model is safe to implement in ED for better allocation of resources.

Suggested Solutions:

Two research studies argued involving a physician in the triage facilitates better bed and resource management and could be a viable solution for ED overcrowding. (Morley, 2018; Yarmohammadian, 2016). Most of the research articles reviewed arguably indicate that bed management and smoothing elective discharges can change Emergency management during overcrowding. (McKenna, 2019; Salway, 2017). The suggestion from Morley et al. (2018) of opening general practitioners' unit in the nearby areas showed significant improvement in reducing emergency visits. Di Somma et al. (2014) theoretically suggested that having acute medical units in place could be a practical solution for reducing overcrowding at the emergency centers. A systematic literature review showed evidential proof that alternative short stay units can be beneficial in managing primary care patients who visit the emergency department often (Yarmohhammadian, 2016).

One of the research studies showed the usage of health informatics and home aides to predict the need for an emergency visit by the patient in the next 14 days (Belmin, 2022). Belmin's innovative idea of using predictive modeling to understand whether the patient needs emergency care helped the healthcare workforce to provide treatment and care early on and prevent the visit. The alert-triggering eHealth system showed significant results in the field. Involving the community health workers as the patient navigators for primary care eligible patients in the emergency department pinpoint the core cause of unnecessary visits and community services help in solving the issue. (Enard, 2014). Pierssens et al. (2022) proved redirecting low-acuity patients to nearby clinics is a safe approach and could reduce overcrowding in the emergency department. Pierssens' approach helps critically ill patients in getting treatment and care faster.

Table 2

Suggested solutions for ED overcrowding

Including Physician in the Triage team
Smoothing elective discharge
Rule of making an admit/discharge decision within first 4 to 6 hours
Opening general practitioners' clinic whenever required
Development of Acute Medical Units
Introducing alternative short stay units like minor injury units or ambulatory areas

Involving home care aides in having eHealth system for future ED visit prediction
Involving Community Health workers as patient navigators
Redirecting low acuity patients to nearby clinics

Research Question

Can a machine learning model be implemented to predict ER visits and prevent them by intervention at the community level?

Aims of the study includes:

- To understand the role of 3A's (Accessibility, Affordability, and Availability) in the increasing ER visits.
- To explore whether having a Primary Care Physician impacts overall rate of ER visits.
- To implement multiple machine learning model to predict avoidable ER visits based on the data available.

METHODS

Institutional Review Board: The University of Missouri Institutional Review Board determined that the study does not constitute human subjects research as per the Department of Health and Human Service regulatory definitions.

Data Sources: The data used in this study is de-identified data provided by the Randolph Caring Community Partnership (RCCCP). The data gets collected on their community level EHR CCMO.

Data Characteristics: The dataset consists of 595 de-identified patients' eligible for the research. The data was collected from 01/01/2018 to 01/01/2022. The age of these patients ranged from 18 to 84 years. The male participants were accounted for the 32.3% of the data whereas there were 67.7% female participants. The data constitute of 43 variables and inclusive of following details:

- Client ID
- Age At Prog Entry
- County
- State
- Zip Code
- Race
- Ethnicity
- Gender
- Assessment Date
- Average hours of sleep (nightly)?
- Do you have a health condition?
- Do you have a Primary Care Physician?
- Do you have health insurance?

- Do you have issues and/or barriers accessing the COVID 19 vaccine?
- Do you use illegal drugs? Please rate on the following scale:
- For what condition do you take medication?
- Have you had a mammogram in the previous 12 months?
- Have you had a well-woman checkup in the previous 12 months?
- How far do you travel to access health care?
- How many Emergency Room visits in the past year?
- How many hospitals stays in the past year?
- How many different prescriptions do you take each day?
- How often does anyone, including family, insult or talk down to you?
- How often does anyone, including family, physically hurt you?
- How would you rate your overall physical health?
- If "Other" type of healthcare, please describe.
- If no insurance, why?
- If other barriers, what are those?
- In the past 12 months has the electric, gas, oil, or Water Company threatened to shut off services in your home?
- In the past 12 months, has lack of transportation kept you from medical appointments, meetings, work or from getting things needed for daily living?
- Other health insurance

- Please describe any additional health conditions not listed in the previous question.
- Please describe any additional housing needs.
- Think about the place you live. Do you have problems with any of the following?
- What are your barriers for accessing healthcare?
- What health conditions do you currently have?
- What is your housing situation today?
- Where do you currently go for healthcare?
- Which housing financial needs do you have?
- Within the past 12 months, the food you bought just didn’t last and you didn’t have money to get more.
- Within the past 12 months, you worried that your food would run out before you got money to buy more.
- Would you like to become pregnant in the next year?

(Appendix I)

<u>Data Cleaning and Pre-processing:</u> The data was extracted in the form of spreadsheet and the open-ended response variables like describe the barriers or prescriptions for what health condition and others were removed from the dataset. There were 97 duplicates found in the dataset. These duplicates were the participants with same client ID but

entered at different date in the system. As the goal of this study is to understand what leads to the ER admission and building a prediction model, the duplicates were kept in the dataset as they can provide useful insights.

There were no missing values except for the questions directed specifically towards women population (mammogram and well women check-ups) were blank for male population. Here is a brief overview of the data:

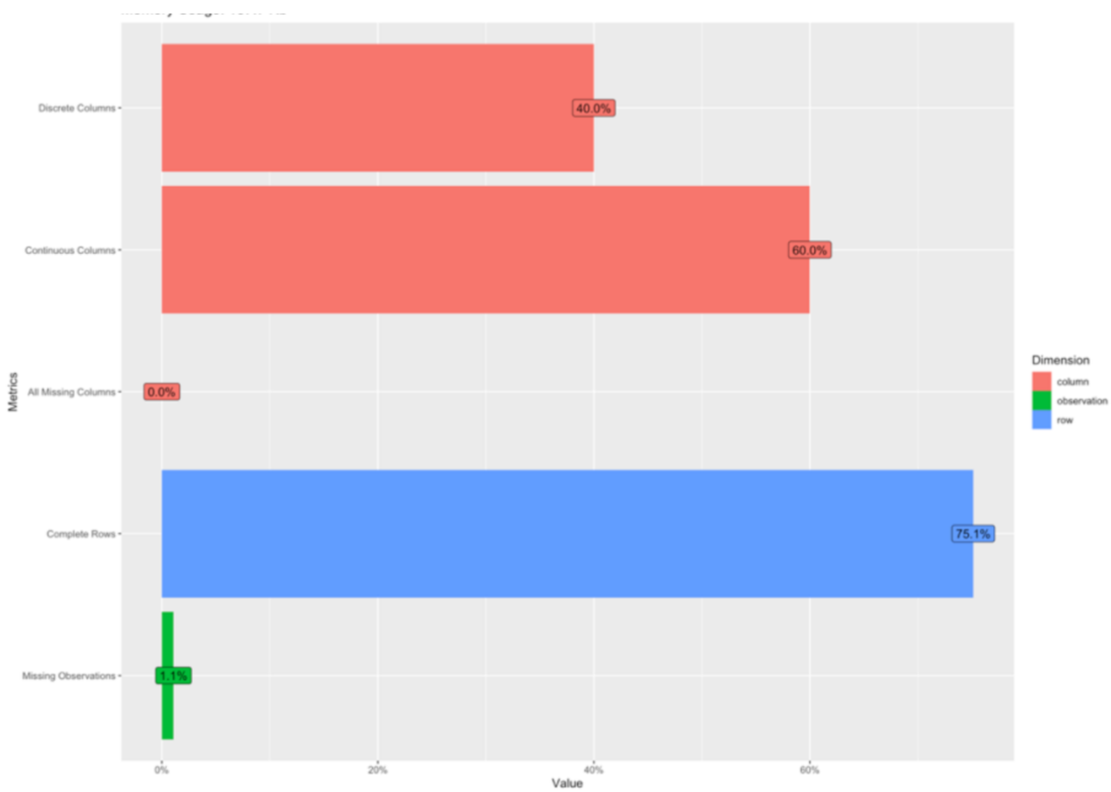


Figure 2. Raw Data Overview

Data was cleaned using the Microsoft Excel. For the major variables like Affordability, Accessibility, Availability, Health conditions, Insurance, and having a Primary Care Physician were all transformed into multiple columns with each having 0 = No and 1 = yes as substitute. The formula used to transform these columns is:

=IF(ISNUMBER(SEARCH("……", #CellNumber)), 1, 0)

The name of all these variables were manually changed. For the ER Visits, we used the same given formula and used it to transform a new column for more than 3 visits to the ER.

Data Exploration: The data exploration was conducted through RStudio using the DataExplorer package. We also added a column to the dataset using R for changing the Age_at_Entry to the ageGrp column. Here is the code:

```
install.packages("pacman")

library("pacman")

library(DataExplorer)

require(DataExplorer)

install.packages("tidyverse")

require("tidyverse")

df <-

read.csv("/Users/anushkakanoongo/Desktop/Thesis/Preprocessed_dat

aFile.csv", header = TRUE)

DataExplorer::create_report(df)

df <- df %>%
  mutate(AgeGrp =
```

```
case_when(Age_at_Entry >= 18 & Age_at_Entry <=24 ~ '18-24',
          Age_at_Entry > 24 & Age_at_Entry <= 30 ~ '25-30',
          Age_at_Entry > 30 & Age_at_Entry <= 36 ~ '31-36',
          Age_at_Entry > 36 & Age_at_Entry <= 42 ~ '37-42',
          Age_at_Entry > 42 & Age_at_Entry <= 48 ~ '43-48',
          Age_at_Entry > 48 & Age_at_Entry <= 54 ~ '49-54',
          Age_at_Entry > 54 & Age_at_Entry <= 60 ~ '54-60',
          Age_at_Entry > 60 & Age_at_Entry <= 66 ~ '61-66',
          Age_at_Entry > 66 & Age_at_Entry <= 72 ~ '67-72',
          Age_at_Entry > 72 & Age_at_Entry <= 78 ~ '73-78',
          Age_at_Entry > 78 ~ 'Above 78'))
write.csv(df,
"/Users/anushkakanoongo/Desktop/Thesis/Cleaned_dataFile.csv")
```

The rest of the data exploration was conducted using Python in the Jupyter Notebook.

Here is the python Code:

```
df =
pd.read_csv('/Users/anushkakanoongo/Desktop/Thesis/Cleaned_dataF
ile.csv')
fig = make_subplots(
   rows=8, cols=2, subplot_titles=('Heart_conditions', 'Diabetes',
                          'Kidney_disease','Breathing_issues',
                          'Depression', 'Opioid_use',
                          'Illegal_substanceUse', 'Affordability',
                          'Accessibility','Availability',
                          'No_Medical_Insurance', 'Primary_doctor',
                          'Other_HealthConditions', 'Gender', 'Race'),

   specs=[[{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}],
       [{"type": "domain"}, {"type": "domain"}]]
)

colours = ['#4285f4', '#ea4335', '#fbbc05', '#34a853']
```

```
fig.add_trace(go.Pie(labels=np.array(df['Heart_conditions'].value_coun
ts().index),
                 values=[x for x in
df['Heart_conditions'].value_counts()], hole=.35,
                 textinfo='label+percent', rotation=-45,
marker_colors=colours),
           row=1, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Diabetes'].value_counts().ind
ex),
                 values=[x for x in df['Diabetes'].value_counts()],
hole=.35,
                 textinfo='label+percent', marker_colors=colours),
           row=1, col=2)

fig.add_trace(go.Pie(labels=np.array(df['Kidney_disease'].value_count
s().index),
                 values=[x for x in df['Kidney_disease'].value_counts()],
hole=.35,
                 textinfo='label+percent', rotation=-45,
marker_colors=colours),
           row=2, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Breathing_issues'].value_cou
nts().index),
                 values=[x for x in
df['Breathing_issues'].value_counts()], hole=.35,
                 textinfo='label+percent', rotation=-45,
marker_colors=colours),
           row=2, col=2)

fig.add_trace(go.Pie(labels=np.array(df['Depression'].value_counts().i
ndex),
              values=[x for x in df['Depression'].value_counts()],
hole=.35,
                 textinfo='label+percent', marker_colors=colours),
           row=3, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Opioid_use'].value_counts().i
ndex),
                 values=[x for x in df['Opioid_use'].value_counts()],
hole=.35,
                 textinfo='label+percent', marker_colors=colours),
           row=3, col=2)
```

```
fig.add_trace(go.Pie(labels=np.array(df['Illegal_substanceUse'].value_
counts().index),
                values=[x for x in
df['Illegal_substanceUse'].value_counts()], hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
            row=4, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Affordability'].value_counts().i
ndex),
                values=[x for x in df['Affordability'].value_counts()],
hole=.35,
                textinfo='label+percent', marker_colors=colours),
            row=4, col=2)

fig.add_trace(go.Pie(labels=np.array(df['Accessibility'].value_counts().i
ndex),
                values=[x for x in df['Accessibility'].value_counts()],
hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
            row=5, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Availability'].value_counts().in
dex),
                values=[x for x in df['Availability'].value_counts()],
hole=.35,
                textinfo='label+percent', marker_colors=colours),
            row=5, col=2)

fig.add_trace(go.Pie(labels=np.array(df['No_Medical_Insurance'].value
_counts().index),
                values=[x for x in
df['No_Medical_Insurance'].value_counts()], hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
            row=6, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Primary_doctor'].value_count
s().index),
                values=[x for x in df['Primary_doctor'].value_counts()],
hole=.35,
```

```
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
          row=6, col=2)

fig.add_trace(go.Pie(labels=np.array(df['Other_HealthConditions'].valu
e_counts().index),
                values=[x for x in
df['Other_HealthConditions'].value_counts()], hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
          row=7, col=1)

fig.add_trace(go.Pie(labels=np.array(df['Gender'].value_counts().inde
x),
                values=[x for x in df['Gender'].value_counts()],
hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
          row=7, col=2)

fig.add_trace(go.Pie(labels=np.array(df['Race'].value_counts().index),
                values=[x for x in df['Race'].value_counts()], hole=.35,
                textinfo='label+percent', rotation=-45,
marker_colors=colours),
          row=8, col=1)


fig.update_layout(height=3200, font=dict(size=14),
showlegend=False)

fig.show()
## creating function to get model statistics
import numpy as np
import statsmodels.api as sm
def get_stats():
   x = data[x_columns]
   results = sm.OLS(y, x).fit()
   print(results.summary())
get_stats()
```

The demographic table was created using python package TableOne.

Here is the code used:

```
!pip install tableone
from tableone import TableOne, load_dataset

dataset = df
columns = [ 'Race',
      'Gender', 'Health_Condition', 'PCP',
      'X0_Medical_Insurance', 'Utility_shutdown',
      'Transportation_as_a_barrier', 'Primary_careProvider',
      'Urgent_care', 'Emergency_Room', 'Health_Department',
      'Rent_issues', 'Utilities_payIssues', 'Deposit_Issues',
      'Accessibility', 'Affordability',
      'Availability', 'Heart_conditions', 'Diabetes', 'Kidney_disease',
      'Breathing_issues', 'Depression', 'Opioid_Use',
      'Illegal_substanceUse', 'Other_HealthConditions']

categorical = [ 'Race',
      'Gender', 'Health_Condition', 'PCP',
      'X0_Medical_Insurance', 'Utility_shutdown',
      'Transportation_as_a_barrier', 'Primary_careProvider',
      'Urgent_care', 'Emergency_Room', 'Health_Department',
      'Rent_issues', 'Utilities_payIssues', 'Deposit_Issues',
      'Accessibility', 'Affordability',
      'Availability', 'Heart_conditions', 'Diabetes', 'Kidney_disease',
      'Breathing_issues', 'Depression', 'Opioid_Use',
      'Illegal_substanceUse', 'Other_HealthConditions']

mytable = TableOne(dataset, columns=columns,
categorical=categorical)
mytable.to_csv('/Users/anushkakanoongo/Desktop/Thesis/Demographi
cs.csv')
print(mytable.tabulate(tablefmt = "fancy_grid"))
```

Feature Elimination: The feature elimination process was conducted through the step wise logistic regression in the python.

Here is the code for transforming the columns in binary values for better implementation of the ML models:

```
for col in ['Gender']:
    if df[col].dtype == 'O':
        le = LabelEncoder()
        df[col] = le.fit_transform(df[col])

# One-hot encode columns with more than 2 unique values
```

```
df = pd.get_dummies(df, columns=['Race', 'Vaccine_Barriers',
'Avg_sleepHRS', 'Travel_for_healthcare'], prefix = ['Race', 'Vaccine_',
'Sleep_hours_', 'Travel_'])
```

Here is the code for the Step wise Logistic Regression Analysis (OLS = Ordinary Least Square) for feature elimination:

```
## creating function to get model statistics

import numpy as np

import statsmodels.api as sm

def get_stats():

    x = data[x_columns]

    results = sm.OLS(y, x).fit()

    print(results.summary())

get_stats()
```

Prediction Models: Decision Tree, Logistic Regression, and Support Vector Machine Model was implemented for predicting the ER visits. The further validation of these models was done by accuracy levels, confusion matrix, and Receiver Operating Characteristic AUC scores. Here is the code:

```
#Setting the label and features for the model

features = data1[['Age_at_Entry',

       'Gender', 'PCP', 'Utility_shutdown',
```

```
    'Emergency_Room', 'Utilities_payIssues',
    'Heart_conditions', 'Diabetes',
    'Breathing_issues', 'Other_HealthConditions',
    'Race_Black or African American',
    'Race_White',
    'Travel__21+ miles']]

labels = data1['More_than_3ERvisits']
#Dividing the dataset into training and testing dataset (80% training,
20% testing)
train_df = data1[:int(len(df)*0.8)]
val_df = data1[int(len(df)*0.8):]
#The imbalance of the data was then balanced using oversampling
technique.
class_0 = train_df[train_df['More_than_3ERvisits'] == 0]
class_1 = train_df[train_df['More_than_3ERvisits'] == 1]

class_1 = class_1.sample(len(class_0),replace=True)
train_df = pd.concat([class_0, class_1], axis=0)
print('Data in Train:')
print(train_df['More_than_3ERvisits'].value_counts())
```

```
class_0 = val_df[val_df['More_than_3ERvisits'] == 0]
class_1 = val_df[val_df['More_than_3ERvisits'] == 1]

class_1 = class_1.sample(len(class_0),replace=True)
val_df = pd.concat([class_0, class_1], axis=0)
print('Data in Test:')
print(val_df['More_than_3ERvisits'].value_counts())

#Converting the dataframe to array for better functioning model.
x_train = np.array(train_df[['Age_at_Entry',
    'Gender', 'PCP', 'Utility_shutdown',
    'Emergency_Room', 'Utilities_payIssues',
    'Heart_conditions', 'Diabetes',
    'Breathing_issues', 'Other_HealthConditions',
    'Race_Black or African American',
    'Race_White',
    'Travel__21+ miles']])

y_train = np.array(train_df['More_than_3ERvisits'])

x_val = np.array(val_df[['Age_at_Entry',
    'Gender', 'PCP', 'Utility_shutdown',
```

```
        'Emergency_Room', 'Utilities_payIssues',
        'Heart_conditions', 'Diabetes',
        'Breathing_issues', 'Other_HealthConditions',
        'Race_Black or African American',
        'Race_White',
        'Travel__21+ miles']])

y_val = np.array(val_df['More_than_3ERvisits'])
#Implementing Decision Tree classifier with accuracy and confusion
matrix outputs:
clf = DecisionTreeClassifier()

# Train Decision Tree Classifer
clf = clf.fit(x_train,y_train)

#Predict the response for test dataset
y_pred = clf.predict(x_val)



score = accuracy_score(y_val, y_pred)
print("Accuracy of the decision tree model is", score)
```

```
cm = confusion_matrix(y_val, y_pred)

plt.figure(figsize=(7,5))

ax = sns.heatmap(cm/np.sum(cm),fmt='.2%', annot=True,
cmap='Greens')

ax.set_xlabel('\nPredicted Values')
ax.set_ylabel('Actual Values ');

ax.xaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])
ax.yaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])

plt.show()
#For logistic regression classifier
from sklearn import model_selection
from sklearn.model_selection import RepeatedKFold
from sklearn.model_selection import cross_val_score
logreg = LogisticRegression()
logreg.fit(x_train, y_train)

y_pred = logreg.predict(x_val)
```

```
accuracy = accuracy_score(y_val, y_pred)

print("Accuracy of the logistic regression model is", accuracy)


print(classification_report(y_val,y_pred))


cm = confusion_matrix(y_val, y_pred)

plt.figure(figsize=(7,5))


ax = sns.heatmap(cm/np.sum(cm),fmt='.2%', annot=True,

cmap='Greens')


ax.set_xlabel('\nPredicted Values')

ax.set_ylabel('Actual Values ');


ax.xaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])

ax.yaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])

plt.show()


#Support Vector Machine Model

svm = SVC(kernel= 'linear', random_state=1, C=0.1)

svm.fit(x_train, y_train)
```

```
# Model performance

y_pred = svm.predict(x_val)
print('Accuracy: %.3f' % accuracy_score(y_val, y_pred))

print(classification_report(y_val,y_pred))

cm = confusion_matrix(y_val, y_pred)
plt.figure(figsize=(7,5))

ax = sns.heatmap(cm/np.sum(cm),fmt='.2%', annot=True,
cmap='Greens')

ax.set_xlabel('\nPredicted Values')
ax.set_ylabel('Actual Values ');

ax.xaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])
ax.yaxis.set_ticklabels(['ERvisits < 3','ER Visits >= 3'])
plt.show()

#ROC AUC
```

```
from sklearn.metrics import roc_auc_score

from sklearn.metrics import roc_curve

logit_roc_auc = roc_auc_score(y_val, clf.predict(x_val))

fpr, tpr, thresholds = roc_curve(y_val, clf.predict_proba(x_val)[:,1])

plt.figure()

plt.plot(fpr, tpr, label='Decision Tree (area = %0.2f)' % logit_roc_auc)

plt.plot([0, 1], [0, 1],'r--')

plt.xlim([0.0, 1.0])

plt.ylim([0.0, 1.05])

plt.xlabel('False Positive Rate')

plt.ylabel('True Positive Rate')

plt.title('Receiver operating characteristic')

plt.legend(loc="lower right")

plt.savefig('DecisionTree')

plt.show()


logit_roc_auc = roc_auc_score(y_val, logreg.predict(x_val))

fpr, tpr, thresholds = roc_curve(y_val,

logreg.predict_proba(x_val)[:,1])

plt.figure()

plt.plot(fpr, tpr, label='LogisticcRegression (area = %0.2f)' %

logit_roc_auc)
```

```
plt.plot([0, 1], [0, 1],'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.savefig('Log_ROC')
plt.show()
```

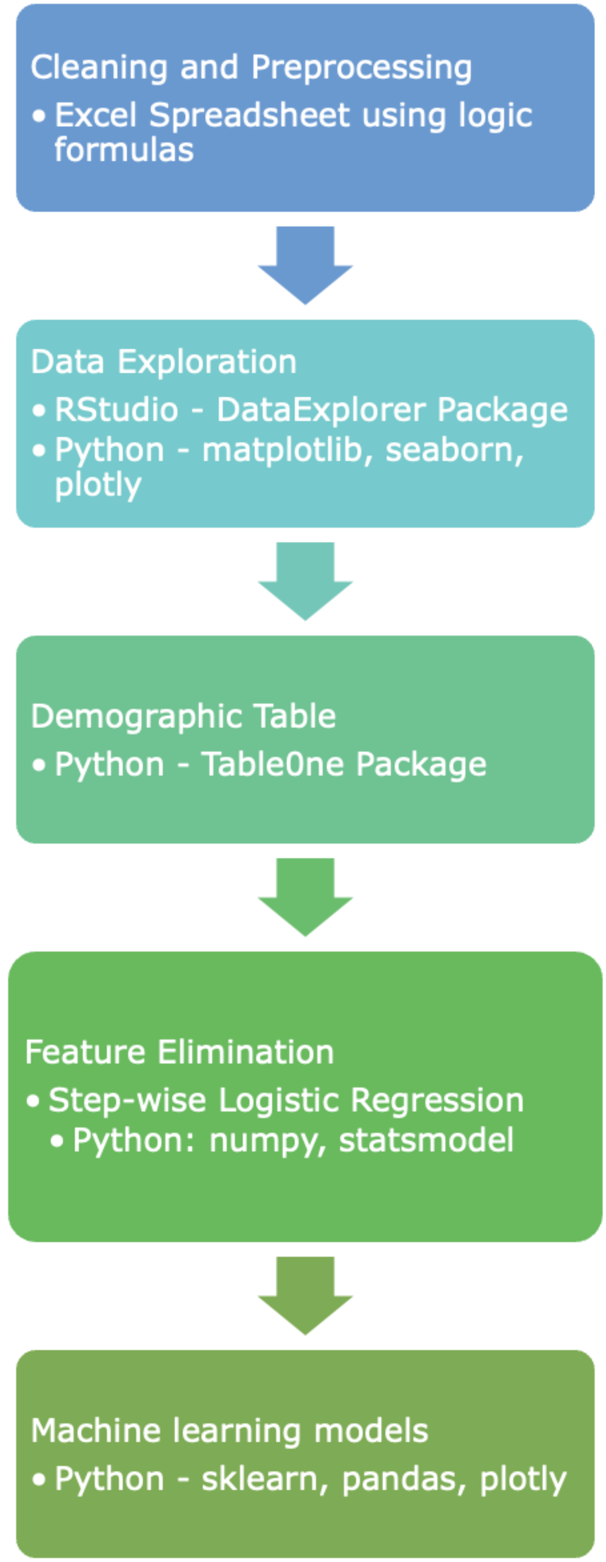


Figure 3. Methods

RESULTS

Exploratory Data Analysis:

75% of the participants are White, almost 14% are Black or African Americans.

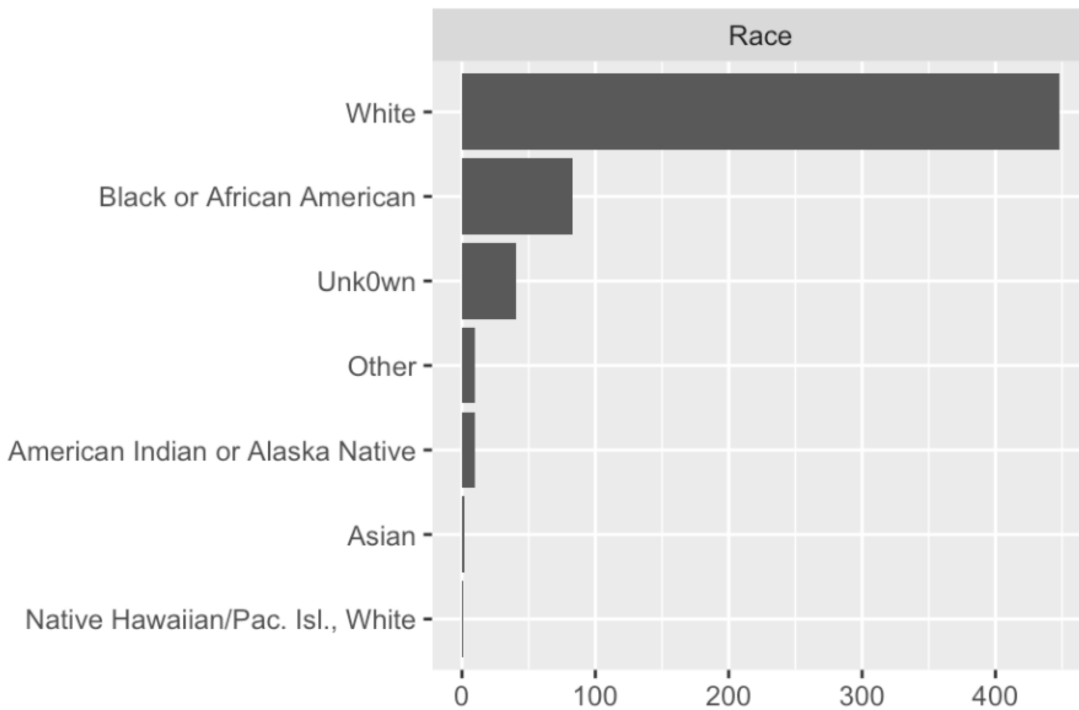


Figure 4. Race Stats

Around 68% of these participants have one or more health conditions. 40% of them don't have any medical insurance and 52% of the participants doesn't have a primary care physician. For 30% of the participants, emergency room is there first point for any healthcare needs. Affordability issues are faced by approximately 54%.

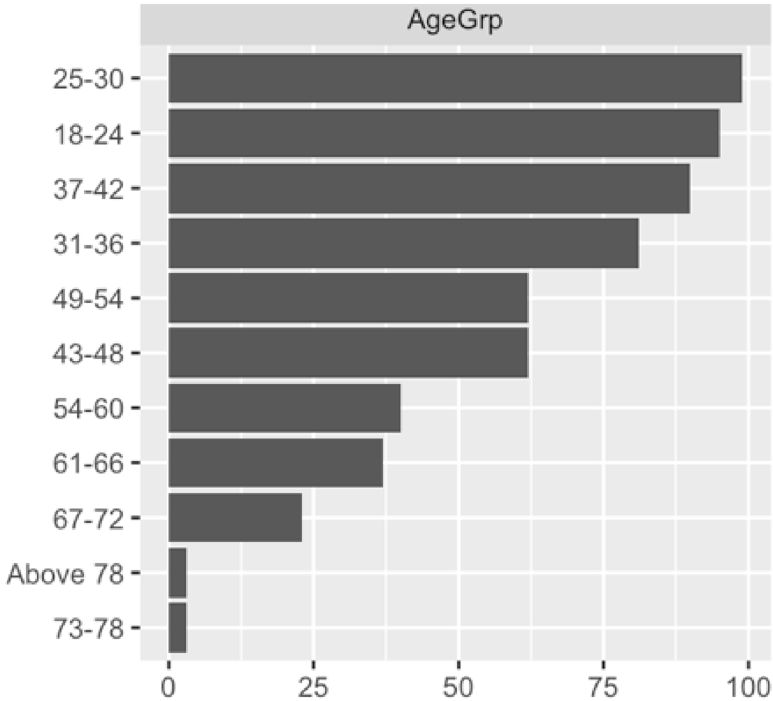


Figure 5. Age Group Distribution

Among the health conditions, depression seems to be the biggest issue as 61.2% of these participants associate themselves with it.

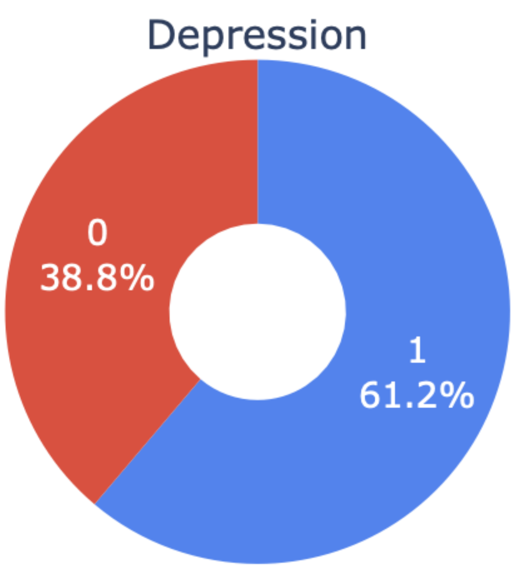


Figure 6. Depression statistics

Table 3

Demographic Characteristics

		Overall
n		595
Race, n (%)	American Indian or Alaska Native	10 (1.7)
	Asian	2 (0.3)
	Black or African American	83 (13.9)
	Native Hawaiian/Pac. Isl., White	1 (0.2)
	Other	10 (1.7)
	Unknown	41 (6.9)
	White	448 (75.3)
Gender, n (%)	Female	403 (67.7)
	Male	192 (32.3)
Health_Condition, n (%)	0	187 (31.4)
	1	408 (68.6)
PCP, n (%)	0	307 (51.6)
	1	288 (48.4)
X0_Medical_Insurance, n (%)	0	354 (59.5)
	1	241 (40.5)
Utility_shutdown, n (%)	0	412 (69.2)
	1	183 (30.8)
Transportation_as_a_barrier, n (%)	0	379 (63.7)
	1	216 (36.3)
Primary_careProvider, n (%)	0	421 (70.8)
	1	174 (29.2)
Urgent_care, n (%)	0	500 (84.0)
	1	95 (16.0)
Emergency_Room, n (%)	0	420 (70.6)
	1	175 (29.4)
Health_Department, n (%)	0	474 (79.7)
	1	121 (20.3)
Rent_issues, n (%)	0	559 (93.9)
	1	36 (6.1)
Utilities_payIssues, n (%)	0	559 (93.9)
	1	36 (6.1)
Deposit_Issues, n (%)	0	581 (97.6)

	1	14 (2.4)
Accessibility, n (%)	0	457 (76.8)
	1	138 (23.2)
Affordability, n (%)	0	275 (46.2)
	1	320 (53.8)
Availability, n (%)	0	505 (84.9)
	1	90 (15.1)
Heart_conditions, n (%)	0	475 (79.8)
	1	120 (20.2)
Diabetes, n (%)	0	497 (83.5)
	1	98 (16.5)
Kidney_disease, n (%)	0	588 (98.8)
	1	7 (1.2)
Breathing_issues, n (%)	0	408 (68.6)
	1	187 (31.4)
Depression, n (%)	0	231 (38.8)
	1	364 (61.2)
Opioid_Use, n (%)	0	573 (96.3)
	1	22 (3.7)
Illegal_substanceUse, n (%)	0	489 (82.2)
	1	106 (17.8)
Other_HealthConditions, n (%)	0	227 (38.2)
	1	368 (61.8)

0 = No, 1 = Yes; PCP = Primary Care Provider; X0_Medical_Insurance = No medical insurance; Primary_careProvider, Urgent_care, Emergency_Room, Health_Department = First point of contact for healthcare.

Almost 26% of the 595 participants have more than 3 ER visits in the past 12 months.

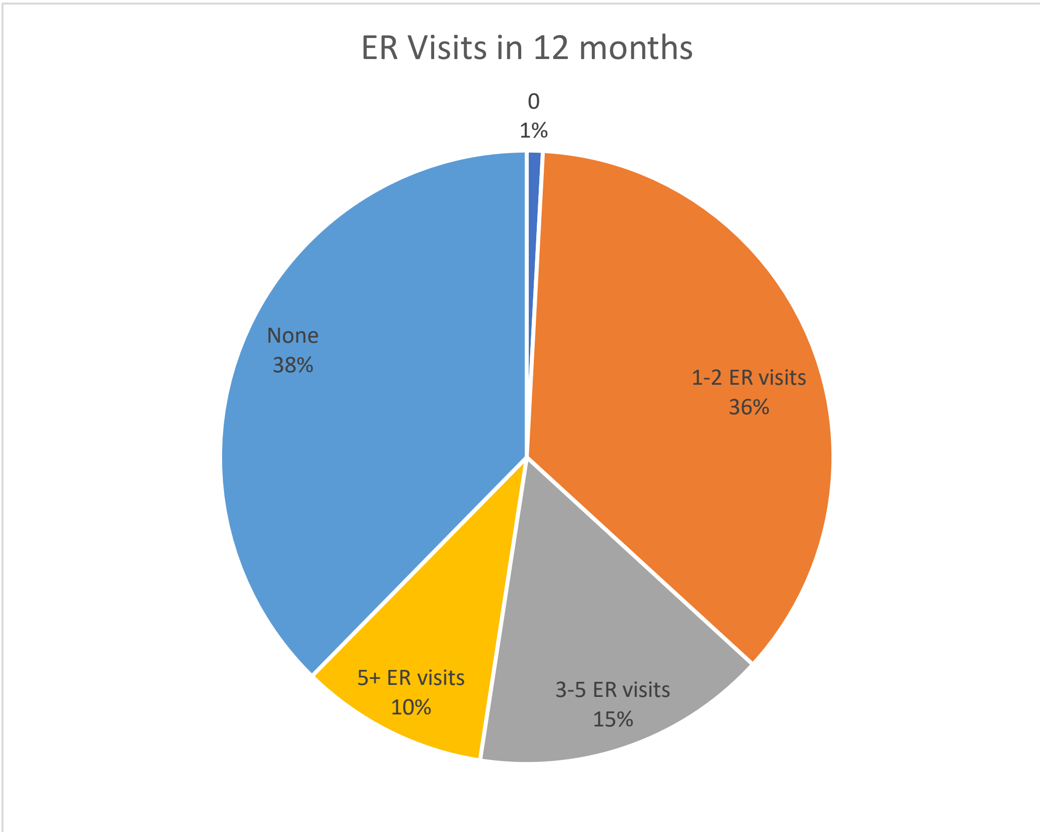


Figure 7. ER Visits in a year.

A brief overview of the data shows that the age group between 25-40 faces the challenge of accessibility, affordability, and availability the most in the area. They are also the age group having the most ER visits among the population.

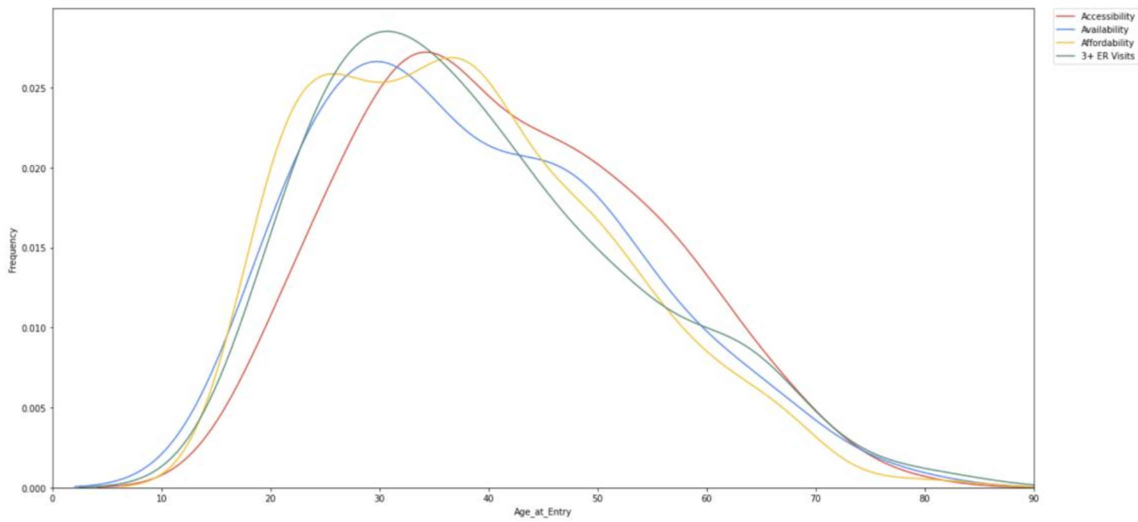


Figure 8. 3A's and 3+ ER visits

The Illegal Substance use is more common among the age of 30-40. However, they are also the ones having the most depression issues.

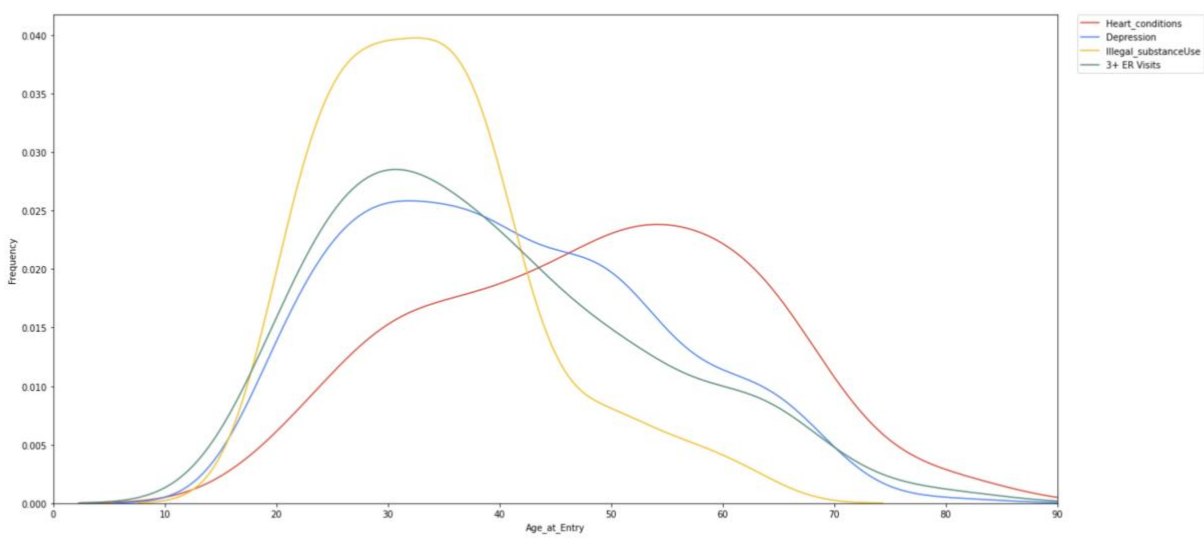


Figure 9. Health Conditions Overview

Majority of the participants must travel more than 21 miles for healthcare.

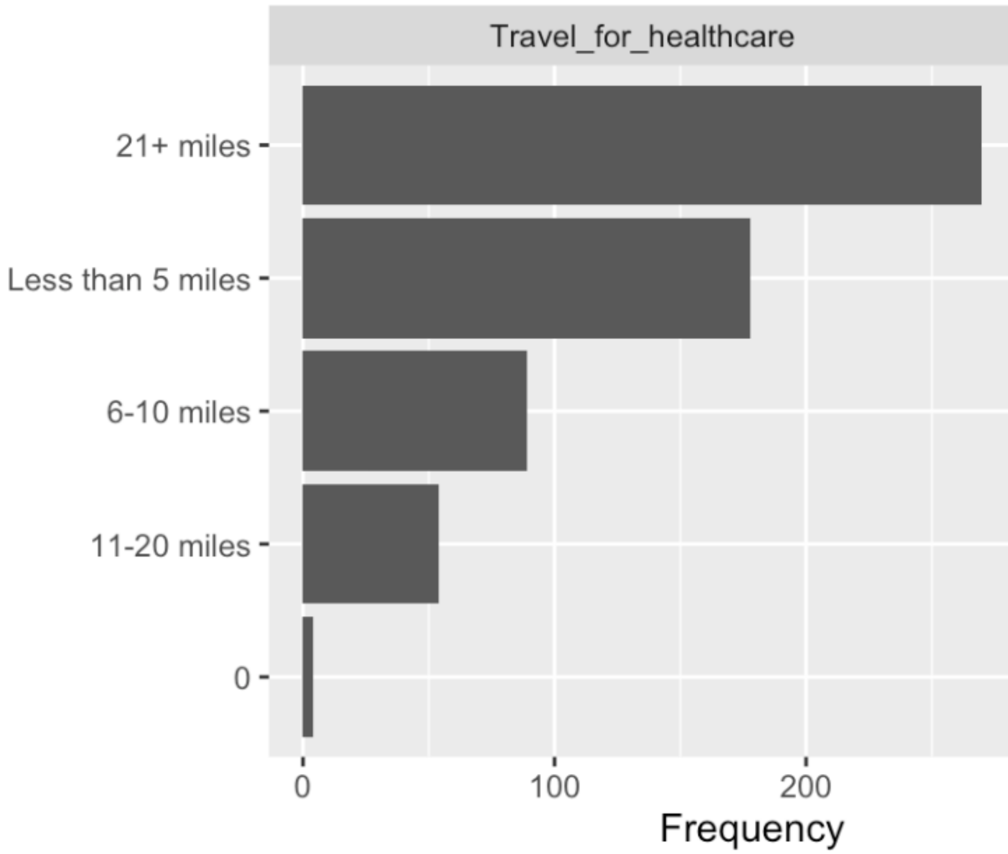


Figure 10. Travel for Healthcare

Feature Elimination: Out of 46 variables, 31 variables were eliminated in 4 iterations and 14 variables were selected for the prediction model. Following features were found significant enough to be included in the model. Here is the result of the logistic regression:

Table 4.

Regression Results

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:     More_than_3ERvisits   R-squared (uncentered):                   0.342
Model:                             OLS   Adj. R-squared (uncentered):              0.326
Method:                  Least Squares   F-statistic:                              21.60
Date:                 Wed, 26 Apr 2023   Prob (F-statistic):                    2.05e-44
Time:                         18:57:11   Log-Likelihood:                         -313.62
No. Observations:                  595   AIC:                                      655.2
Df Residuals:                      581   BIC:                                      716.7
Df Model:                           14
Covariance Type:             nonrobust
=================================================================================================
                                   coef    std err          t      P>|t|      [0.025      0.975]
-------------------------------------------------------------------------------------------------
Age_at_Entry                    -0.0034      0.001     -2.791      0.005      -0.006      -0.001
Gender                          -0.0868      0.038     -2.261      0.024      -0.162      -0.011
Health_Condition                -0.0036      0.041     -0.087      0.931      -0.084       0.077
PCP                              0.0733      0.038      1.925      0.055      -0.001       0.148
Utility_shutdown                 0.0909      0.038      2.370      0.018       0.016       0.166
Emergency_Room                   0.1573      0.039      4.050      0.000       0.081       0.234
Utilities_payIssues              0.1124      0.074      1.510      0.132      -0.034       0.259
Heart_conditions                 0.0682      0.047      1.465      0.144      -0.023       0.160
Diabetes                         0.0771      0.049      1.575      0.116      -0.019       0.173
Breathing_issues                 0.1161      0.040      2.893      0.004       0.037       0.195
Other_HealthConditions           0.1415      0.039      3.649      0.000       0.065       0.218
Race_Black or African American   0.1273      0.064      1.973      0.049       0.001       0.254
Race_White                       0.1284      0.046      2.780      0.006       0.038       0.219
Travel__21+ miles                0.0715      0.035      2.049      0.041       0.003       0.140
==============================================================================
Omnibus:                       77.526   Durbin-Watson:                   1.923
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               88.617
Skew:                           0.901   Prob(JB):                     5.71e-20
Kurtosis:                       2.427   Cond. No.                         188.
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

The age and gender are significantly critical (p value = 0.005, 0.024) to identify the ER visits. The travel for more than 21 miles (p value = 0.04) also seemed to be an important determinator of the increased ER visits in the area. The utility shutdowns are also seem to be one of the feature leading to the ER visits with p value of 0.01.

Machine Learning Models:

Decision Tree Classifier was found to be 0.53 accurate in predicting the ER visits. The model was significantly low in predicting more than 3 ER visits as it was correct only 13.10%.

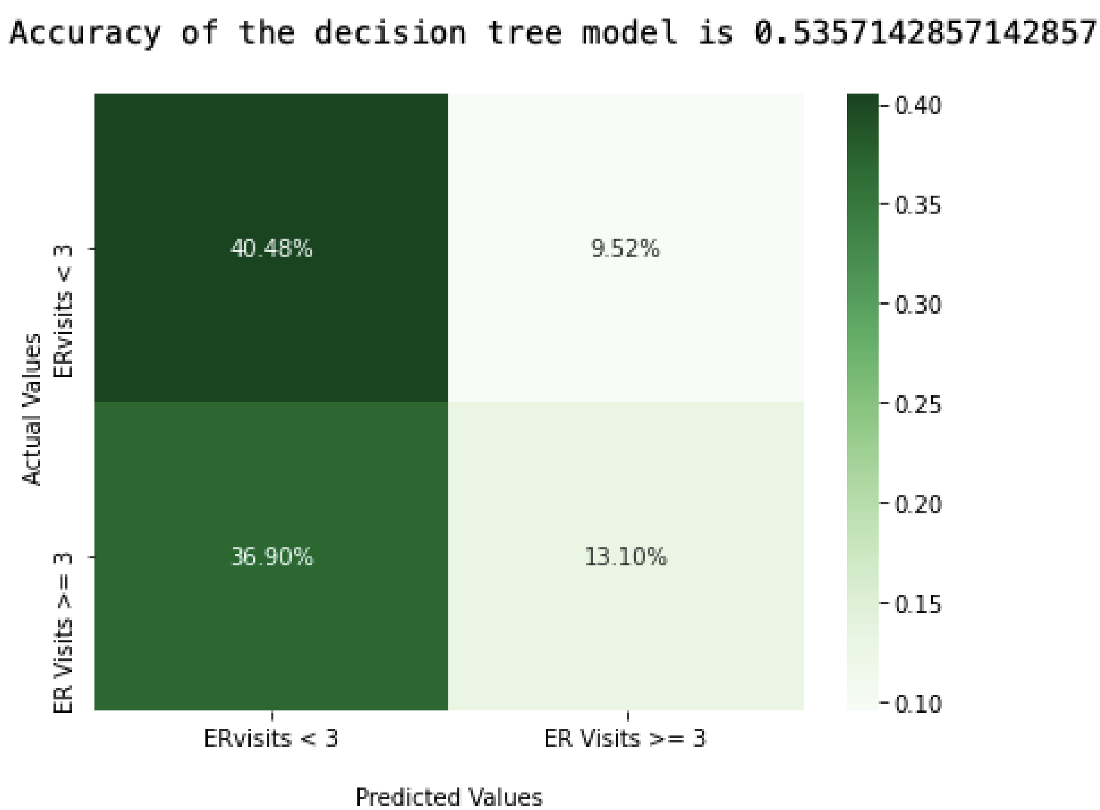


Figure 11. Decision Tree Confusion Matrix

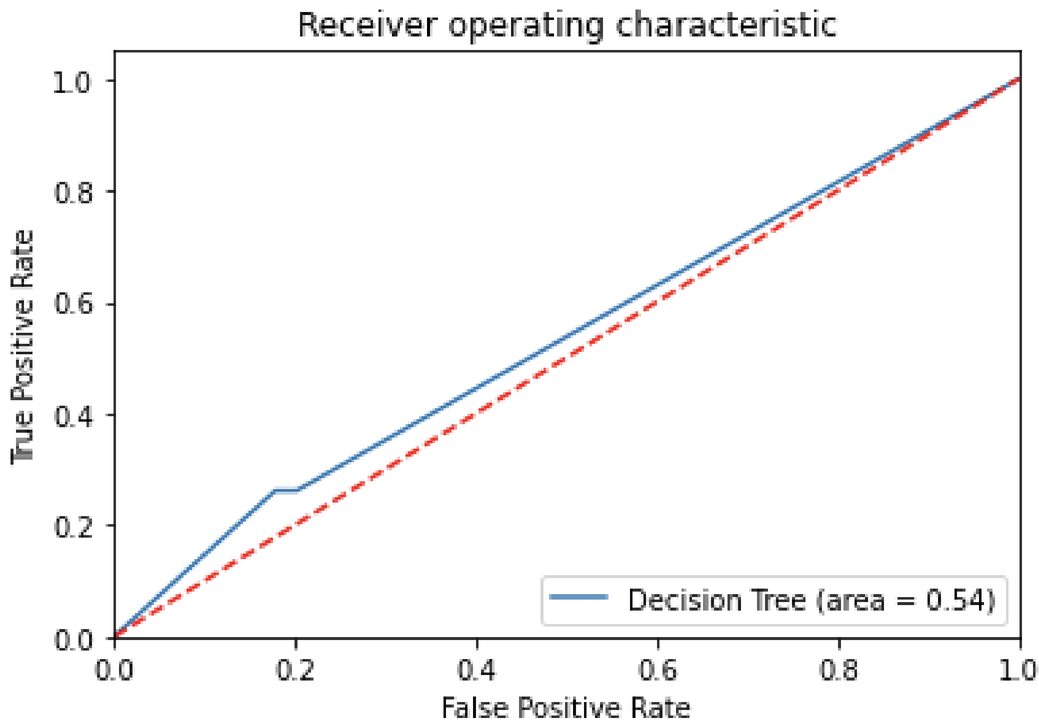


Figure 12. Decision Tree ROC

The support vector machine model had an accuracy of 0.589. It shown a significant better predictability for the more than 3 ER visits.

```
Accuracy: 0.589
              precision    recall  f1-score   support

           0       0.58      0.63      0.61        84
           1       0.60      0.55      0.57        84

    accuracy                           0.59       168
   macro avg       0.59      0.59      0.59       168
weighted avg       0.59      0.59      0.59       168
```

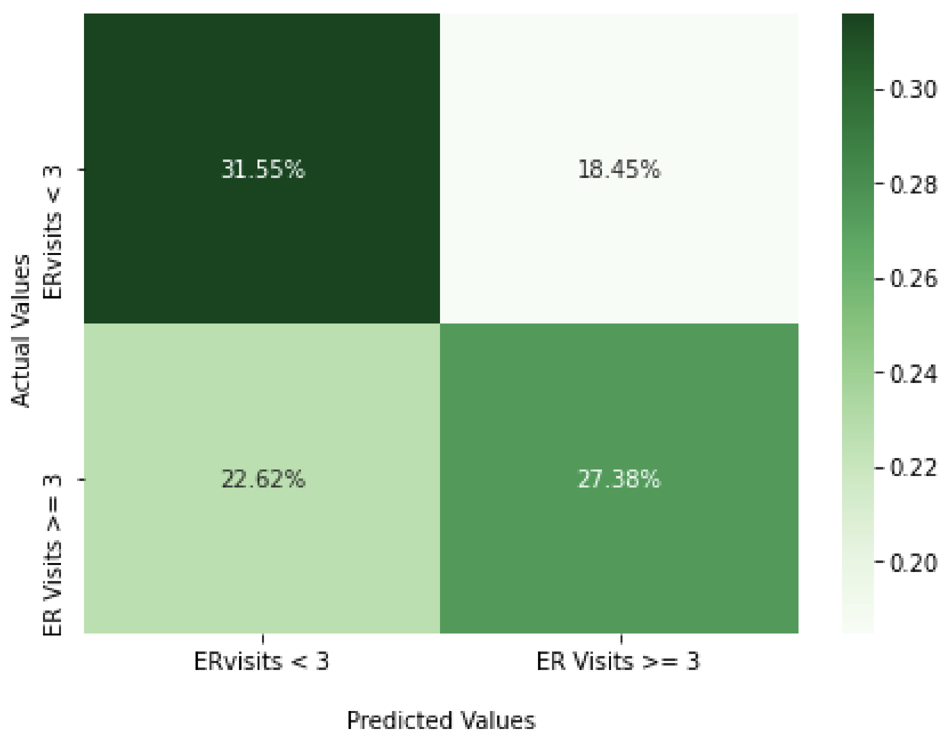


Figure 13. SVM Confusion Matrix

The Logistic Regression Model reached the accuracy of 0.60.

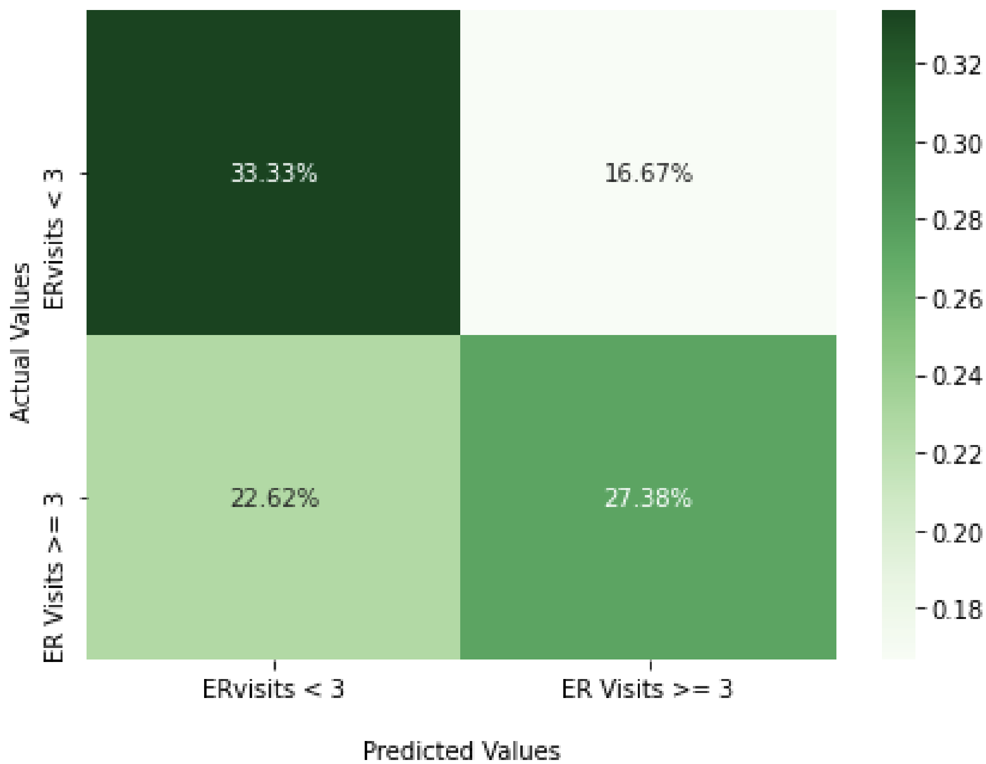


Figure 14. LogReg Confusion Matrix

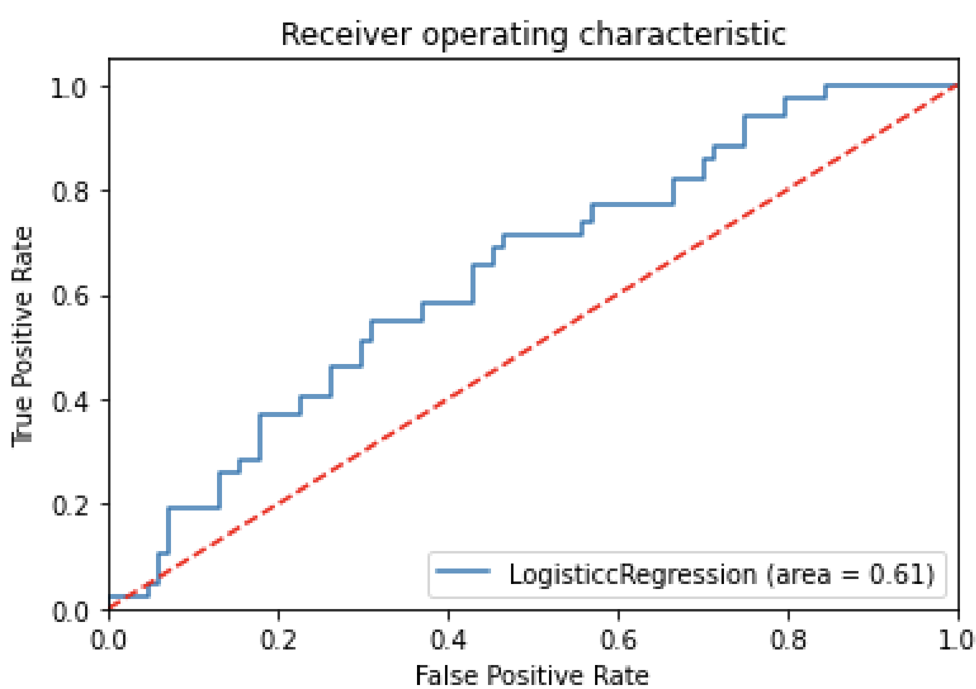


Figure 15. LogReg ROC curve

On comparison, the logistic regression model performed the best among the three prediction models with 0.6 accuracy.

DISCUSSION

The results show a moderate ability in predicting the avoidable ER visits to be able to intervene at the community level. However, more data needs to be procured to have a feasible model in place.

Surprisingly, the 3A's of Healthcare (Accessibility, Availability, and Affordability) were not deemed significant in predicting the increase of ER visits. Although, some variables like traveling more than 21 miles for health care services, and utility shut down were significant in developing the model.

The usage of machine learning models and artificial intelligence could change the approach towards the healthcare. It is vital that we include community level organizations in this process for an effective model. Based on the research conducted for this study, this is a second attempt to implement such models. A more holistic approach in this direction will provide more useful insights to reduce the number of preventable ER visits.

LIMITATIONS

- The data was collected at a community level and no inputs were taken from any other sources like hospital.
- Due to time constraints, further investigation of the models weren't possible.
- Insufficient sample size to produce desirable outcome.
- More data features are required to have a reliable and ubiquitous model.

CONCLUSION

The study shows that it is possible to have prediction models at the community level to reduce the burden at emergency department. Intervention at the community level would result in impacting at the root of the increasing number of preventable emergency visits. This approach would lead to reduced burden at the Emergency department, cost reduction, and better healthcare delivery.

Future researchers can focus on collecting data from multiple sources to build a dashboard / alert system for the community services for better care delivery on the ground. They can also develop a tool for the community health workers for indicating if someone at risk of going to the ER visits for prevention of it as early as possible.

REFERENCES

- Belmin, J., Villani, P., Gay, M., Fabries, S., Havreng-Théry, C., Malvoisin, S., Denis, F., & Veyron, J.-H. (2022). Real-world Implementation of an eHealth System Based on Artificial Intelligence Designed to Predict and Reduce Emergency Department Visits by Older Adults: Pragmatic Trial. Journal of Medical Internet Research, 24(9), e40387. https://doi.org/10.2196/40387
- Burton, C., Stone, T., Oliver, P., Dickson, J. M., Lewis, J., & Mason, S. M. (2022a). Frequent attendance at the emergency department shows typical features of complex systems: Analysis of multicentre linked data. Emergency Medicine Journal, 39(1), 3–9. https://doi.org/10.1136/emermed-2020-210772
- Chan, S. S. W., Cheung, N. K., Graham, C. A., & Rainer, T. H. (2015). Strategies and solutions to alleviate access block and overcrowding in emergency departments. Hong Kong Medical Journal = Xianggang Yi Xue Za Zhi, 21(4), 345–352. https://doi.org/10.12809/hkmj144399
- Di Somma, S., Paladino, L., Vaughan, L., Lalle, I., Magrini, L., & Magnanti, M. (2015). Overcrowding in emergency department: An international issue. Internal and Emergency Medicine, 10(2), 171–175. https://doi.org/10.1007/s11739-014-1154-8

- Enard, K. R., & Ganelin, D. M. (2013). Reducing preventable emergency department utilization and costs by using community health workers as patient navigators. Journal of Healthcare Management / American College of Healthcare Executives, 58(6), 412–427; discussion 428.
- Feral-Pierssens, A.-L., Morris, J., Marquis, M., Daoust, R., Cournoyer, A., Lessard, J., Berthelot, S., & Messier, A. (2022). Safety assessment of a redirection program using an electronic application for low-acuity patients visiting an emergency department. BMC Emergency Medicine, 22(1), 71. https://doi.org/10.1186/s12873-022-00626-4
- Geographic Variance in Maryland's Potentially Preventable Emergency Visits: Comparison of Explanatory Models. (n.d.). https://doi.org/10.1177/0193945919867938
- Jarvis, S., Flemming, K., Richardson, G., & Fraser, L. (2022). Adult healthcare is associated with more emergency healthcare for young people with life-limiting conditions. Pediatric Research, 92(5), 1458–1469. https://doi.org/10.1038/s41390-022-01975-3
- Locklear, M. (2022, September 30). Emergency department crowding hits crisis levels, risking patient safety. YaleNews.

https://news.yale.edu/2022/09/30/emergency-department-crowding-hits-crisis-levels-risking-patient-safety

- McKenna, P., Heslin, S. M., Viccellio, P., Mallon, W. K., Hernandez, C., & Morley, E. J. (2019). Emergency department and hospital crowding: Causes, consequences, and cures. Clinical and Experimental Emergency Medicine, 6(3), 189–195. https://doi.org/10.15441/ceem.18.022
- Hanlon, Alexandra L. PhD; Pauly, Mark V. PhD; Huang, Liming PhD; Lozano, Alicia J.; Hirschman, Karen B. PhD; McCauley, Kathleen PhD; Press, Matthew MD; Naylor, Mary D. PhD. Medical Complexity Mediates Healthcare Resource Use in the Socially Vulnerable. Journal of Healthcare Management: May-June 2022 - Volume 67 - Issue 3 - p 173-191 doi: 10.1097/JHM-D-21-00072
- Morley, C., Unwin, M., Peterson, G. M., Stankovich, J., & Kinsman, L. (2018). Emergency department crowding: A systematic review of causes, consequences, and solutions. PLoS ONE, 13(8), e0203316. https://doi.org/10.1371/journal.pone.0203316
- Salway, R. J., Valenzuela, R., Shoenberger, J. M., Mallon, W. K., & Viccellio, A. (2017). EMERGENCY DEPARTMENT (ED) OVERCROWDING: EVIDENCE-BASED ANSWERS TO FREQUENTLY

ASKED QUESTIONS. Revista Médica Clínica Las Condes, 28(2), 213–219. https://doi.org/10.1016/j.rmclc.2017.04.008

- Tapia, A. D., Howard, J. T., Bebo, N. L., Pfaff, J. A., Chin, E. J., Trueblood, W. A., April, M. D., Long, B. J., Long, A. N., Fernandez, W. G., & Schauer, S. G. (2022b). A Retrospective Review of Emergency Department Visits That May Be Appropriate for Management in Non-Emergency Settings. Military Medicine, 187(9–10), e1153–e1159. https://doi.org/10.1093/milmed/usab553
- Yarmohammadian, M. H., Rezaei, F., Haghshenas, A., & Tavakoli, N. (2017). Overcrowding in emergency departments: A review of strategies to decrease future challenges. Journal of Research in Medical Sciences: The Official Journal of Isfahan University of Medical Sciences, 22, 23. https://doi.org/10.4103/1735-1995.200277
- Lane, B. H., Mallow, P. J., Hooker, M. B., & Hooker, E. (2020). Trends in United States emergency department visits and associated charges from 2010 to 2016. The American Journal of Emergency Medicine, 38(8), 1576–1581. https://doi.org/10.1016/j.ajem.2019.158423
- Ramdhani, Abdullah & Ramdhani, Muhammad & Amin, Abdusy. (2014). Writing a Literature Review Research Paper: A step-by-

step approach. International Journal of Basic and Applied Science. 3. 47-56.

- Bennett H. Lane, Peter J. Mallow, Maria B. Hooker, Edmond Hooker, Trends in United States emergency department visits and associated charges from 2010 to 2016, The American Journal of Emergency Medicine, Volume 38, Issue 8, 2020, Pages 1576-1581, ISSN 0735-6757, https://doi.org/10.1016/j.ajem.2019.158423.
- Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health & Human Services. Trends in the Utilization of Emergency Department Services, 2009-2018. 2021. https://aspe.hhs.gov/pdf-report/utilization-emergency-department-services
- HealthGroup, U. (2019). *Avoidable ed visits - unitedhealth group*. 18 Million Avoidable Hospital Emergency Department Visits Add $32 Billion in Costs to the Health Care System Each Year. Retrieved April 12, 2023, from https://www.unitedhealthgroup.com/content/dam/UHG/PDF/2019/UHG-Avoidable-ED-Visits.pdf
- Yang, Y., Yu, J., Liu, S., Wang, H., Dresden, S., & Luo, Y. (2022). Predicting Avoidable Emergency Department Visits Using the

NHAMCS Dataset. *AMIA ... Annual Symposium proceedings. AMIA Symposium*, *2022*, 514–523.

- *Home*. exploreMOhealth. (2023, April 17). Retrieved April 28, 2023, from https://exploremohealth.org/
- Udalova, V. (2023, April 19). *Most vulnerable more likely to depend on emergency rooms for preventable care*. Census.gov. Retrieved April 28, 2023, from https://www.census.gov/library/stories/2022/01/who-makes-more-preventable-visits-to-emergency-rooms.html
- Centers for Disease Control and Prevention. (2023, February 23). *FastStats - emergency department visits*. Centers for Disease Control and Prevention. Retrieved April 28, 2023, from https://www.cdc.gov/nchs/fastats/emergency-department.htm
- MOPHIMS - preventable hospitalization mica. (n.d.). Retrieved April 28, 2023, from https://healthapps.dhss.mo.gov/MoPhims/QueryBuilder
- *Scikit learn tutorial. Tutorials Point. (n.d.). Retrieved April 28, 2023, from https://www.tutorialspoint.com/scikit_learn/index.htm*

APPENDIX I. Data Description

Race	**Ethnicity**	**Gender**	**Average hours of sleep (nightly)?**	**Do you have a health condition?**	**Do you have a Primary Care Physician?**
Black or African American	Central American	Male	3-4	Yes	Yes
White	Spaniard	Female	5-7	No	No
American Indian or Alaska Native	Puerto Rican		Less than 3	None	
Other	Paraguayan		8+ hours		
Unknown	Mexican				
American Indian or Alaska Native, White	Bolivian				
Asian	Mexican American				
Black or African American, White					
Black or African American, Other					
Other, White					
Native Hawaiian/Pac. Isl., White					

Do you have health insurance?	Do you have issues and/or barriers accessing the COVID 19 vaccine?	Do you use illegal drugs? Please rate on the following scale:	For what condition do you take medication?	Have you had a mammogram in the previous 12 months?
[No Medical Insurance]	Sometimes	Non Use [1]		Not Applicable
[Medicaid]	No	Recreational/ Social Use [3]		No
[Medicaid] [Medicare]	Yes	Experimental Use [2]		Yes
[Private Insurance]		Regular use [4]		
[Medicare]		Dependence/ Compulsive Use [5]		
[Medicaid] [No Medical Insurance]				
[Other]				

How far do you travel to access health care?	How many Emergency Room visits in the past year?	How many hospital stays in the past year?	How many different prescriptions do you take each day?	How often does anyone, including family, insult or talk down to you?	How often does anyone, including family, physically hurt you?	How would your rate your overall physical health?
21+ miles	1-2 ER visits	3-5 hospital stays	4-6	0	0	Fair
Less than 5 miles	None	1-2 hospital stays	None	Never [1]	Never [1]	Good
11-20 miles	5+ ER visits	None	1-3	Sometimes [3]	Sometimes [3]	Poor
6-10 miles	3-5 ER visits	5+ hospital stays	11 or more	Rarely [2]	Fairly often [4]	Excellent
0	0	0	7-10	Fairly often [4]	Rarely [2]	0
			0	Frequently [5]	Frequently [5]	

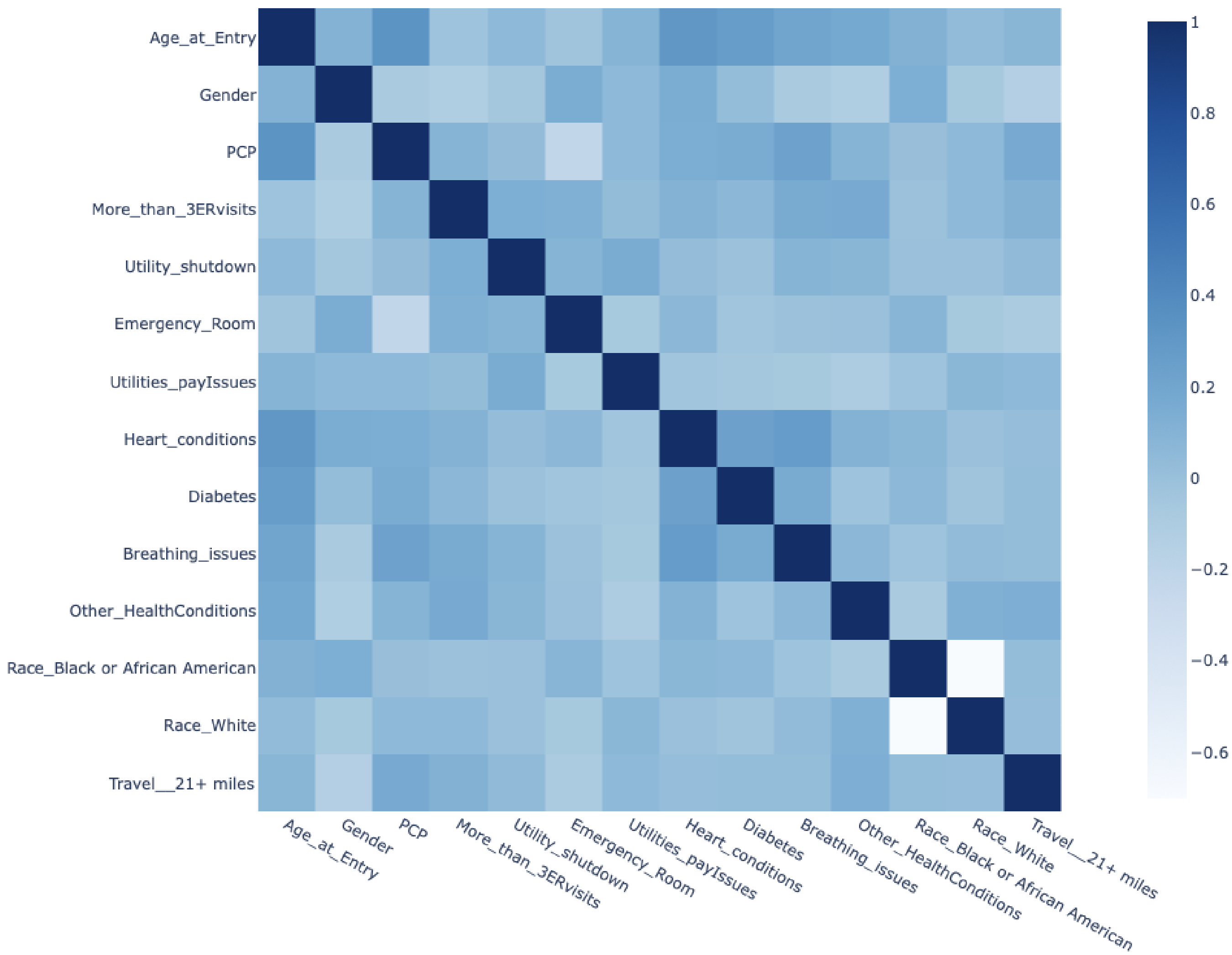
In the past 12 months has the electric, gas, oil, or Water Company threatened to shut off services in your home?	In the past 12 months, has lack of transportation kept you from medical appointments, meetings, work or from getting things needed for daily living? (Check all that apply)	Think about the place you live. Do you have problems with any of the following? (Check all that apply)	What are your barriers for accessing healthcare?	What health conditions do you currently have?
No	Yes, it has kept me from non-medical meetings, appointments, work, or getting things that I need	[None of the above]	[Accessibility]	[Heart Disease] [Diabetes or other blood sugar problems] [Other conditions]
Yes	No	[Inadequate heat]	[Affordability]	[Depression] [Other conditions]
Already shut off	Yes, it has kept me from medical appointments or getting medications	[Bug infestation]	[Accessibility]	[None]
		No or not working smoke detectors		[Heart Failure or an Enlarged Heart] [Breathing problems caused by emphysema or asthma] [Diabetes or other blood sugar problems] [Depression] [Other conditions]
		[Mold]		[Breathing problems caused by emphysema or asthma] [Other conditions]
		[Proximity to drug usage or sales]		[Diabetes or other blood sugar problems] [Depression]

		[Lead paint or pipes] [Oven or stove not working]		[Breathing problems caused by emphysema or asthma] [Depression]
		[Water leaks]		[Depression]
		[None of the above]		[Opioid Use/ Abuse]
				[Illegal Substance Use/ Abuse] [Other conditions]

What is your housing situation today?	**Where do you currently go for healthcare?**	**Which housing financial needs do you have? (Check all that apply)**	**Within the past 12 months, the food you bought just didn't last and you didn't have money to get more.**	**Within the past 12 months, you worried that your food would run out before you got money to buy more.**	**Would you like to become pregnant in the next year?**
I have housing.	[Health Department] [Primary Care Provider]	[Rent or Mortgage]	Sometimes true	Never true	Not Applicable
I have housing today, but I am worried about losing housing in the future.	[Primary Care Provider]	[Rent or Mortgage] [Deposit]	Often true	Sometimes true	No
I do not have housing (I am staying with others, hotel, shelter, outside or car)	[Urgent Care]	[Utilities]	Never true	Often true	Yes
	[None]				Unsure

	[Other]				Refused
	[Emergency Department]				ok either way

APPENDIX II. Correlation Analysis



APPENDIX III. Initial Regression Results

```
OLS Regression Results
==============================================================================
Dep. Variable:     More_than_3ERvisits   R-squared:                       0.148
Model:                             OLS   Adj. R-squared:                  0.085
Method:                  Least Squares   F-statistic:                     2.341
Date:                 Wed, 26 Apr 2023   Prob (F-statistic):           9.97e-06
Time:                         18:57:11   Log-Likelihood:                -302.91
No. Observations:                  595   AIC:                             689.8
Df Residuals:                      553   BIC:                             874.1
Df Model:                           41
Covariance Type:             nonrobust
=======================================================================================================
===
                                            coef    std err          t      P>|t|      [0.025      0.9
75]
-------------------------------------------------------------------------------------------------------
---
Age_at_Entry                             -0.0047      0.001     -3.232      0.001      -0.008      -0.
002
Gender                                   -0.0969      0.042     -2.334      0.020      -0.178      -0.
015
Health_Condition                         -0.0016      0.048     -0.034      0.973      -0.096       0.
092
PCP                                       0.0761      0.050      1.510      0.132      -0.023       0.
175
X0_Medical_Insurance                     -0.0696      0.049     -1.429      0.154      -0.165       0.
026
Utility_shutdown                          0.0826      0.041      2.024      0.043       0.002       0.
163
Transportation_as_a_barrier               0.0339      0.042      0.804      0.422      -0.049       0.
117
Primary_careProvider                     -0.0379      0.053     -0.713      0.476      -0.142       0.
066
```

Urgent_care	-0.0481	0.050	-0.970	0.332	-0.146	0.049
Emergency_Room	0.1442	0.041	3.523	0.000	0.064	0.225
Health_Department	-0.0417	0.048	-0.871	0.384	-0.136	0.052
Rent_issues	-0.0822	0.081	-1.014	0.311	-0.241	0.077
Utilities_payIssues	0.1367	0.082	1.658	0.098	-0.025	0.299
Deposit_Issues	0.0118	0.118	0.100	0.920	-0.219	0.243
Accessibility	0.0517	0.046	1.118	0.264	-0.039	0.143
Affordability	0.0354	0.045	0.789	0.431	-0.053	0.124
Availability	0.0036	0.051	0.070	0.944	-0.096	0.103
Heart_conditions	0.0725	0.049	1.491	0.136	-0.023	0.168
Diabetes	0.0896	0.051	1.762	0.079	-0.010	0.189
Kidney_disease	0.0659	0.168	0.392	0.695	-0.264	0.396
Breathing_issues	0.1172	0.042	2.814	0.005	0.035	0.199
Depression	-0.0174	0.042	-0.413	0.680	-0.100	0.066
Opioid_Use	0.0212	0.097	0.218	0.828	-0.170	0.212
Illegal_substanceUse	0.0281	0.053	0.530	0.596	-0.076	0.132
Other_HealthConditions	0.1292	0.042	3.101	0.002	0.047	0.211
Race_American Indian or Alaska Native	0.1273	0.140	0.911	0.363	-0.147	0.402

```
Race_Asian                                  -0.0239      0.272     -0.088      0.930     -0.559      0.
511
Race_Black or African American               0.1264      0.082      1.543      0.123     -0.034      0.
287
Race_Native Hawaiian/Pac. Isl., White       -0.3548      0.377     -0.941      0.347     -1.095      0.
385
Race_Other                                   0.0169      0.138      0.122      0.903     -0.254      0.
288
Race_Unk0wn                                  0.0388      0.090      0.432      0.666     -0.137      0.
215
Race_White                                   0.1228      0.071      1.737      0.083     -0.016      0.
262
Vaccine__0                                   0.0453      0.049      0.921      0.358     -0.051      0.
142
Vaccine__Sometimes                          -0.0266      0.064     -0.412      0.680     -0.153      0.
100
Vaccine__Yes                                 0.0346      0.063      0.548      0.584     -0.089      0.
159
Sleep_hours__0                              -0.3040      0.212     -1.433      0.152     -0.721      0.
113
Sleep_hours__3-4                             0.0935      0.061      1.534      0.125     -0.026      0.
213
Sleep_hours__5-7                             0.0320      0.053      0.605      0.546     -0.072      0.
136
Sleep_hours__8+ hours                        0.1192      0.060      2.000      0.046      0.002      0.
236
Sleep_hours__Less than 3                     0.1127      0.079      1.426      0.154     -0.043      0.
268
Travel__0                                    0.0469      0.183      0.256      0.798     -0.313      0.
407
Travel__11-20 miles                         -0.0520      0.067     -0.781      0.435     -0.183      0.
079
Travel__21+ miles                            0.0553      0.050      1.097      0.273     -0.044      0.
154
Travel__6-10 miles                          -0.0080      0.062     -0.129      0.897     -0.130      0.
114
```

```
Travel__Less than 5 miles                  0.0111      0.051      0.216      0.829      -0.089       0.
112
==============================================================================
Omnibus:                       73.790   Durbin-Watson:                   1.960
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               83.575
Skew:                           0.874   Prob(JB):                     7.11e-19
Kurtosis:                       2.438   Cond. No.                     1.00e+16
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.06e-26. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```