Data Analytics Academy

**Impact of Missed Appointments on Patient Health, Quality of Care,**

**and Hospital Finances**

Furhan Babar and Mary Grace Boyd

Data Analytics Academy – Final Capstone

Data Analytics Instructor and Coach: Niteen Kumar

# 

# 

# 

# 

**Table of Contents**

Introduction…………………………………………………………………………… 03

Background …………………………………………………………………………… 03

Data Source …………………………………………………………………………… 03

Hypothesis …………………………………………………………………………… 04

Approach and Methodologies …………………………………………………… 04

Conclusion …………………………………………………………………………… 10

References …………………………………………………………………………… 10

# Introduction

American Hospital Association currently replaced Patients’ Bill of Rights literature with a patient and family centered care brochure called “The Patient Care Partnership” that gives guidance to patients in understanding their expectations, rights, and responsibilities regarding their care (*The Patient Care Partnership | AHA*, 2018). In this brochure, patients, families, provider, care teams, and healthcare organizations focus on partnership and collaboration so that patients can make informed health care and treatment decisions by being engaged and accountable for their health.

Patients have rights and responsibilities when it comes to obtaining and receiving quality and safe care. One of the patient responsibilities is to be involved in their care by participating in medical appointments for preventative health, diagnostic, and treatment procedures. Missing these healthcare-related appointments “represent a significant risk marker for all-cause mortality” (McQueenie et al, 2019). In addition, the financial impact of missed appointments significantly affects an organizations’ ability to improve access and quality of care (Triemstra and Lowery, 2018).

# Background

Harlem Hospital Center’ fiscal year 2020 start on July 1, 2019 June 30, 2020. Throughout the fiscal year, “missed appointment” data has been recorded through EPIC documentation system. Analysis of the data and other variables recorded could potentially give us an insight as to how this information may impact financial sustainability and quality of care given to patients.

Understanding baseline information and data trend analysis is a vital component when undergoing quality improvement projects that may impact the overall vision, mission, and strategic pillars of the institution and the health system.

The specific strategic pillars involved are quality, safety, and financial sustainability that focuses. The primary goals under these strategies pillar are as follows:

* Improve quality of patient care and services
* Increase revenue streams

# Data Source

The Harlem Hospital Centers’ EPIC Reporting Work Bench: Missed Appointments for Fiscal Year 2019-2020 Data Set contains 21226 instances of data, with 50 attributes. The initial dataset represents one fiscal year worth of missed appointment data specific to HHC. The original data set data contains attributes such as patient’s name, medical record number, hospital account number, date of birth, gender, phone, state, zip code, language, religion, special needs, primary care provider, etc.

Environmental factors such as weather type, precipitation, snow depth, and other relevant attributes were added from National Oceanic and Atmospheric Administration’s (NOAA) National Centers for Environnemental Information (NCEI)

To ascertain impact of outside variables, certain attributes such as national and religious holidays and certain significant events were also obtained from websites such as Wikipedia via Google Search.

# Hypothesis

**Null Hypothesis (H0**): Patients demographic information (age and gender), environmental attributes, and appointment day and time have significant effect on the rate of missed appointments.

**Alternative Hypothesis (HA):** There is a significant financial loss during fiscal year 2020 at Harlem Hospital Center due to missed appointments.

# Approach and Methodologies

The below steps outline the approach for the research project:

1. Data Collection
2. Data Preparation
3. Data Wrangling
4. Exploratory Data Analysis
5. Data Visualization

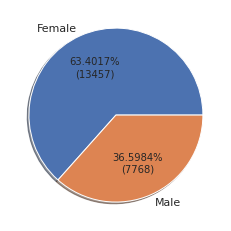
**Data Collection**: The initial raw data set was downloaded as an MS Excel spreadsheet file with leadership permission from the EPIC Reporting Work Bench. Environmental and miscellaneous attributes were added to the data set (as described under “Data Source”), ultimately creating the initial modified data set. The project dataset was initially hosted locally using data-encrypted work computers to ensure adherence to HIPAA policy and procedure. It was our intention to host the final dataset in Github after proper data preparation.

**Data Preparation**: The dataset characteristics has multiple variables and includes string and numerical values. Although initially tagged inappropriately as string data types, date and time attributes were also included in the data set. As the raw data initially included patient information that may violate their privacy rights, data preparation was executed locally in a secure, encrypted NYCHHC issued computer using Microsoft Excel program by deleting and/or replacing sensitive information with generic variables.

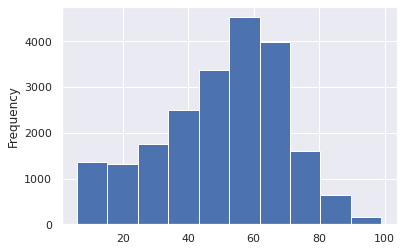
**Data Wrangling.** Manipulation of the data set via Google Colab using Python programming were used to pre-process the data. This step identifies the characteristics of the initial modified data set which is significant during data analysis and visualization process. The data set also had missing values and incorrect data types which was also corrected using data wrangling techniques. Certain irrelevant or repetitive attributes were also handled using Numpy package and Panda data frame manipulation tool.

**Exploratory Data Analysis (EDA-1)**: After the data was prepared and wrangled appropriately, the initial EDA was initiated using Python via Google Colab. We aimed to use different plots to explore the data and form baseline analysis to see which factors will be significantly appropriate for data visualization using dashboards.

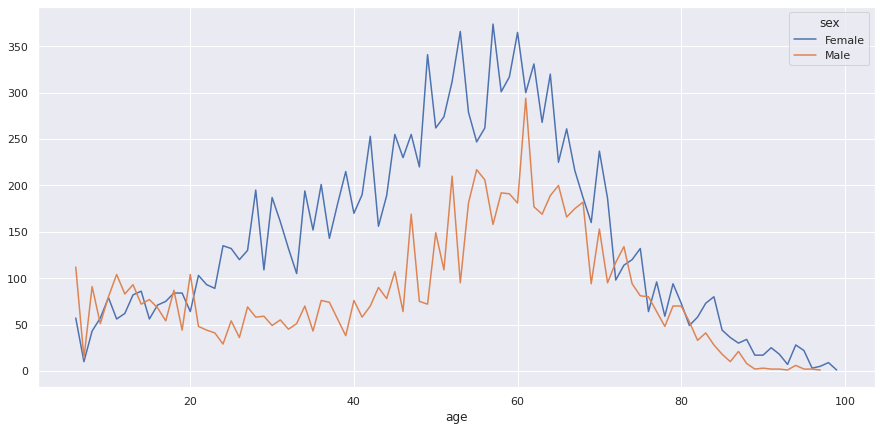
*Patient Gender:* The bar plot below shows female gender missed 13,457 or 63% of total missed appointments and male gender missed 7,768 or 37% of total missed appointments.



*Patient Age*. The age distribution was analyzed using the histogram plot which shows left or negative skewness. The data shows that the majority of missed appointments were from the older adult population. The data also shows that minimum age of patients that missed appointments is one and the maximum age is 99. The mean age is 50.12 and median age is 53.

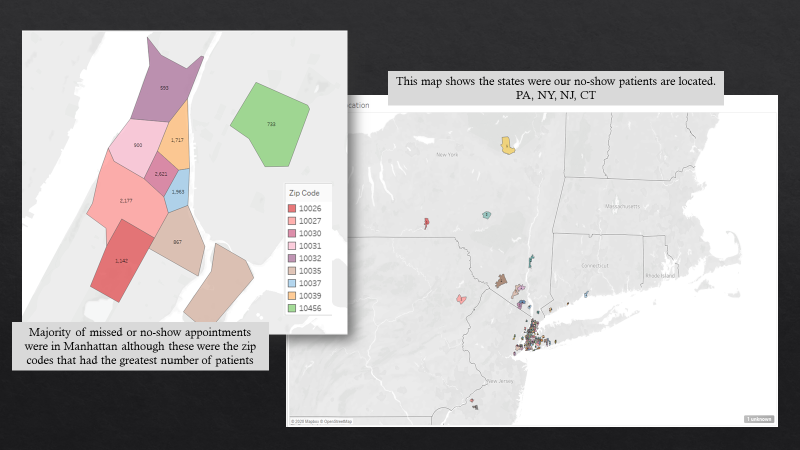


*Age and Gender Correlation.* When combined in a plot, correlation between two attributes maybe inferred. Looking at the line plot below, the correlation between the age and gender of patients that missed their appointments shows that female patients missed more appointments than their male counterparts across age groups with some exception on 0-20 and 70-80 age group.

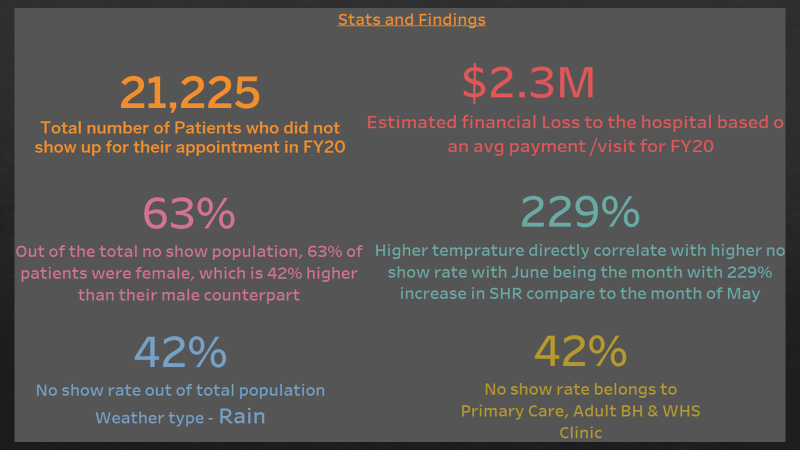


**Exploratory Data Analysis (EDA-2)**: As string data types are challenging to work with in Python for exploratory data analysis, we also used Tableau programming to work with our data set.

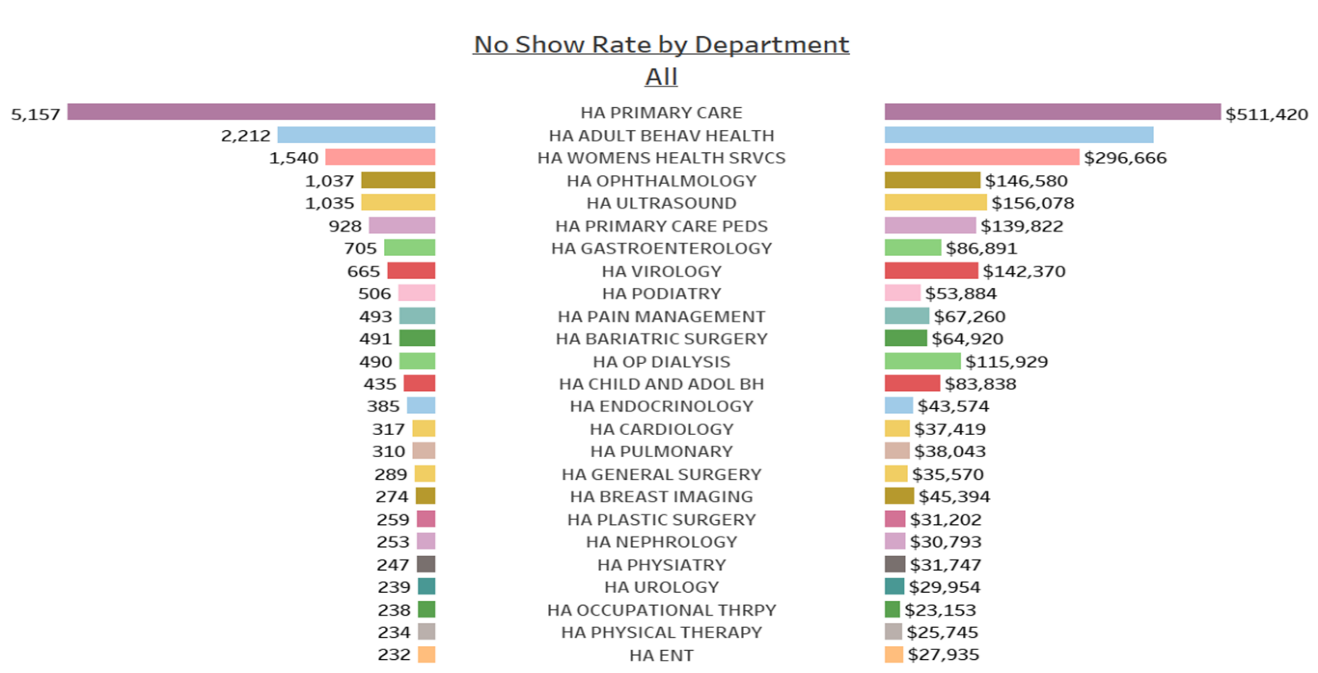
*Patient state and zip code*. As seen below, using Tableau allowed us to visually explore and analyze the data regarding patient location. In the left map visualization, we were able to ascertain the top 10 zip codes that has the greatest number of patients that missed their appointments. We also determined which states our patients lives in.



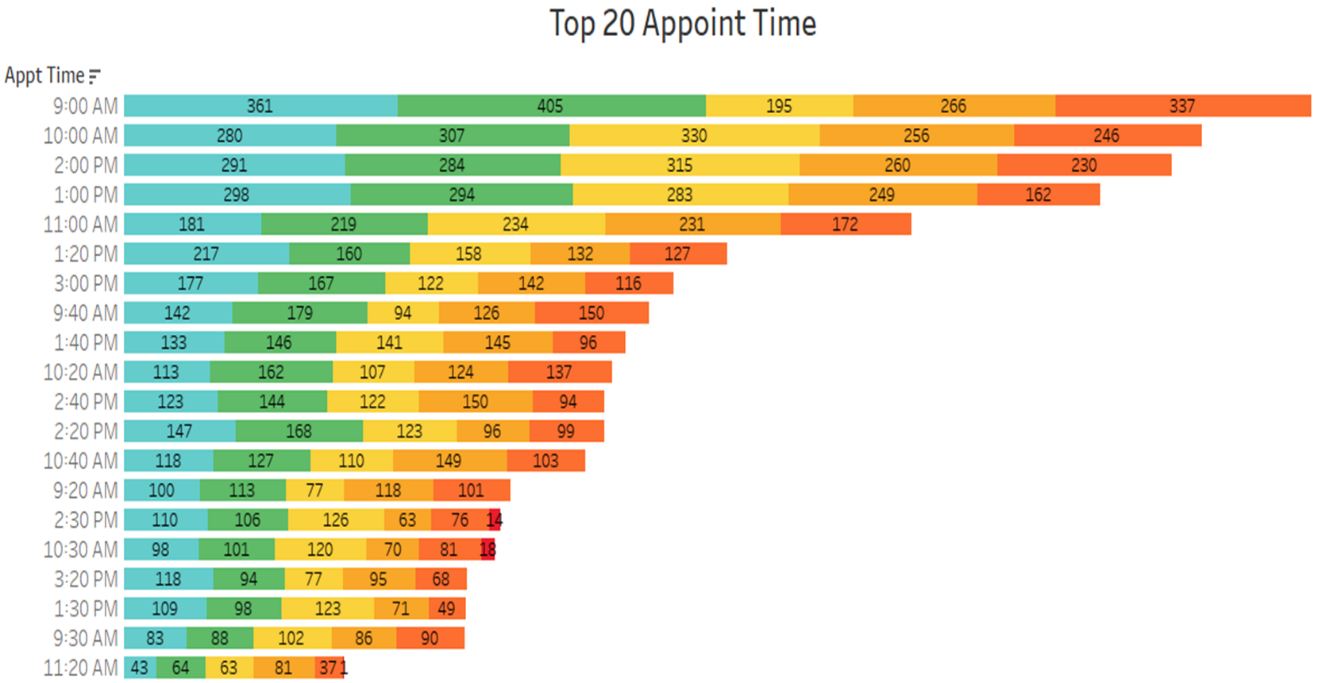
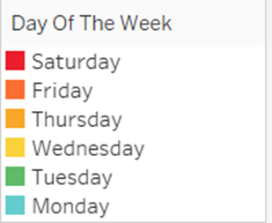
*Statistical findings*. The information below along with other worksheets shows some of the statistical findings obtained from using Tableau programming.



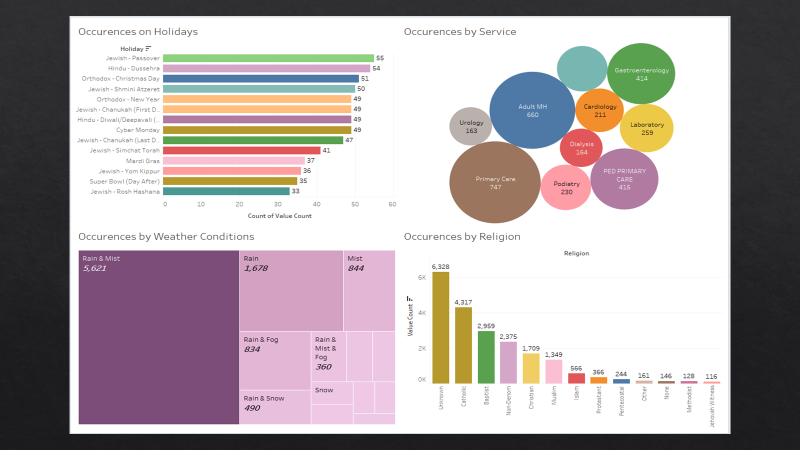
*Financial Impact by Department*. Primary Care, Adult Behavioral Health (BH), Womens Health Services (WHS) and Ophthalmology department accounted for more then 47% of the total missed appointments (or No Show) with the total lost revenue of $1.3 million.

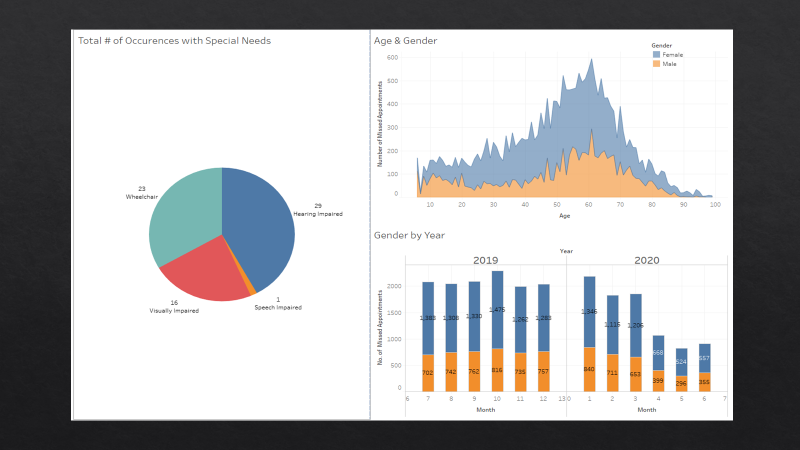


*Appointment Time and Day of the Week*. Below graph shows that 9:00 am followed by 10:00 am appointment times are the moset missed by patients. In addition, Tuesday had the highest rate across all appointment times.



**Data Visualization**. Visualization of data starts with EDA using Python and Tableau programming as seen in the previous pages. The use of dashboard is an invaluable tool when presenting complicated data findings and analysis. We used Tableau to initialize multiple worksheets of data analysis and created a dashboard that will showcase the ideas compiled in an organized and meaningful sequence. See example of our dashboard below.





**Conclusion**

In conclusion, multiple attributes such as gender, age, location, weather type, and appointment time and day of the week contributed significantly on the total cases of missed appointments at Harlem Hospital Center. These findings should be utilized as baseline information when thinking of quality improvement projects that may decrease missed appointment visits. In addition, a significant negative financial impact to Harlem Hospital Center was determined during analysis of the data set. An estimated total loss of $2.3 million dollars during the 2020 fiscal year may negatively impact patient care therefore missing opportunities to improve our patients’ quality of life

# References

*Holidays and observances in United States in 2020. (n.d.). TimeandDate.Com. Retrieved September 1, 2020, from https://www.timeanddate.com/holidays/us/*

*Local Climatological Data (LCD) Dataset Documentation*. (2020, September 2). National Climatic Data Centers. https://www1.ncdc.noaa.gov/pub/data/cdo/documentation/LCD\_documentation.pdf

McQueenie, R., Ellis, D. A., McConnachie, A., Wilson, P., & Williamson, A. E. (2019, January 11). *Morbidity, mortality and missed appointments in healthcare: a national retrospective data linkage study*. BMC Medicine. https://bmcmedicine.biomedcentral.com/articles/10.1186/s12916-018-1234-0#citeas

National Centers for Environmental Information (NCEI). (n.d.). *Datasets | Climate Data Online (CDO) | National Climatic Data Center (NCDC).* Climate Data Online (CDO). Retrieved September 2, 2020, from https://www.ncdc.noaa.gov/cdo-web/datasets#LCD

National Weather Service Corporate Image Web Team. (2020*). National Weather Service Climate*. National Weather Service Forecast Office | New York, NY. https://w2.weather.gov/climate/xmacis.php?wfo=okx

*The Patient Care Partnership | AHA.* (2018, January 18). American Hospital Association. https://www.aha.org/other-resources/patient-care-partnership

Triemstra, J. D., & Lowery, L. (2018, November 19*). Prevalence, Predictors, and the Financial Impact of Missed Appointments in an Academic Adolescent Clinic*. NCBI | PMC | US National Library of Medicine | National Institute of Health. <https://www.cureus.com/articles/14071-prevalence-predictors-and-the-financial-impact-ofmissed-appointments-in-an-academic-adolescent-clinic>

Wikipedia contributors. (2003, December 13). *Public holidays in the United States*. Wikipedia. https://en.wikipedia.org/wiki/Public\_holidays\_in\_the\_United\_States