SIADS 593: Milestone I

Team Project Proposal

version 2022.07.27.1.CT

**Instructions:** please make a ***copy*** of this template file (do not edit original).

## **Proposal Title:**

## Evaluating Hospital Readmission Rates and Timely Care Metrics

## 

## 1. Team members

Please list your team members (2-3 max).

* Iris Lin
* Kasra Afzali
* Michael Light

## 2. Project summary

Summarize your proposed project in a few sentences.

#### What is your proposed project and why are you proposing it?

#### What are the question(s) you want to answer, or goal you want to achieve?

| * Unplanned hospital visits and readmission rates are critical indicators of the quality of care and patient outcomes in healthcare facilities. * Timely and effective care, particularly in emergency departments, is essential to prevent unplanned readmissions and improve overall patient satisfaction. * This project aims to analyze the relationship between hospital readmission rates and timely care metrics by merging data from unplanned hospital visits and timely and effective care, focusing on various conditions such as heart failure, pneumonia, and emergency department (ED) visits. * Understanding these relationships can help identify areas for improvement in hospital care, reducing readmission rates and enhancing patient outcomes. |
| --- |

## 3. Datasets

#### Describe one primary dataset and at least one secondary dataset. If other secondary datasets will be used please describe them as well.

#### The proposed datasets should exhibit different features/columns and/or different access methods, e.g., \*.csv file, \*.json file, API retrieval, web scraping, etc. Different time periods, for example, with the same features/columns is not considered a different dataset. Remember, the focus of the project in this Milestone course is to give you the opportunity to practice your data manipulation skills, so feel free to challenge yourself.

#### If you're unsure if your data sets are "different enough" describe the datasets and request a review via the *#siads593\_[semester]\_001\_project* Slack channel.

#### **Please note:** all proposed datasets ***MUST*** be publicly available to all members of the class (students, instructors, course support personnel, etc.). Use of proprietary datasets for this project is ***not*** permitted.

## 3.1 Primary dataset description

Describe your primary dataset. How is the data collected and how will you access it? Please share what features in the dataset are relevant to your topic. At a minimum, include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

| **Datasets:**  **Dataset 1: Unplanned Hospital Visits - Hospital**   * Source: Provider-level data on hospital return days, unplanned readmissions, and unplanned hospital visits after procedures. * Size: Includes features for various conditions (e.g., heart failure, heart attack, pneumonia) and metrics related to unplanned visits after procedures (e.g., colonoscopy, outpatient chemotherapy). * Key Features: Hospital return days, readmission rates, ED visits for outpatient chemotherapy, unplanned hospital visits after outpatient procedures. * Quality: Detailed data on patient outcomes and procedures, though potential for missing values in certain readmission categories.   + Size: 67,074   + Format: csv   + Public   + Has some missing/supressed values   Dataset 1:  [Unplanned Hospital Visits - Hospital | Provider Data Catalog (cms.gov)](https://data.cms.gov/provider-data/dataset/632h-zaca#data-table) |
| --- |

## 3.2 Secondary dataset(s) description

Describe your secondary dataset(s). How is the data collected and how will you access it? Please share what features in the dataset(s) are relevant to your topic and describe the data types you’re expecting. At a minimum, for each secondary dataset include the following information:

#### Short description (i.e., 1-3 sentences) of its key features

#### Estimated size (in records and/or bytes)

#### Location (give the URL or other access method)

#### Format (CSV, JSON, etc.)

#### Access method (download, web scraping, API, etc.)

| [Please use this space for your response. You may expand or contract this box as needed.]  **Dataset 2: Timely and Effective Care - Hospital**   * Source: Provider-level data on timely care, including emergency department metrics, vaccination rates, and procedure follow-ups. * Size: Covers a wide range of emergency and procedural care metrics, such as ED visit times, sepsis care bundles, and vaccination rates for healthcare workers. * Key Features: Emergency department volume, admit decision to ED departure time, influenza and COVID-19 vaccination rates for healthcare workers, and appropriate care for severe sepsis and septic shock. * Quality: Rich in time-based metrics and procedural outcomes but may require filtering based on psychiatric and non-psychiatric cases.   + Size: 105,455 rows   + Format: csv   + Public   + Has some missing/supressed values   Dataset 2:  [Timely and Effective Care - Hospital | Provider Data Catalog (cms.gov)](https://data.cms.gov/provider-data/dataset/yv7e-xc69) |
| --- |

## 3.3 [ ] Affirm: datasets are public.

Please write YES in the above box to confirm that your primary and secondary datasets are accessible and available to your classmates and the instructional team.

## 4. Cleaning and manipulation

Describe how you will need to manipulate your datasets: how will you handle missing or anomalous data? How will you join your primary and secondary datasets? What cleaning and manipulation challenges, if any, do you anticipate?

| **Data Collection:**   * + Use the facility ID to merge the datasets for a more comprehensive analysis across hospitals. * **Cleaning Steps:**   + Handle missing values in key features such as hospital return days, ED visit times, and readmission rates.   + Standardize time-based features like admit decision to ED departure times to ensure consistency across datasets.   + Remove duplicates and any outliers in procedure follow-up times and emergency department metrics. * **Preparation Techniques:**   + Create new features, such as hospital efficiency scores, by combining readmission rates with emergency department performance metrics.   + Normalize metrics like vaccination rates and ED visit times for easier comparison across facilities.   + Merge datasets on facility ID to link readmission rates with timely and effective care measures.   + Filter data for specific patient groups (e.g., Heart Failure, Pneumonia, etc) for in-depth analysis of care disparities. |
| --- |

## 5. Analysis

Describe any analyses you plan to undertake. For each, please give the technique or approach and briefly explain what you expect to learn from it.

| [Please use this space for your response. You may expand or contract this box as needed.]  **Questions & Objectives:**   * How do timely and effective care measures, such as ED visit times and sepsis care, impact hospital readmission rates for specific conditions (e.g., heart failure, pneumonia)? * What is the relationship between vaccination rates of healthcare workers and hospital-wide readmission rates? * Can patterns in emergency department performance metrics predict the likelihood of unplanned readmissions or hospital return days?   **Formulations:**   * Use regression models to explore the relationship between readmission rates and emergency department performance metrics (e.g., admit decision to ED departure time). * Time-series analysis of hospital performance metrics over time to predict hospital return days for specific conditions. |
| --- |

## 6. Visualizations

Describe in 1-3 sentences at least **two** data visualizations that you plan to create. Include the chart type (e.g. bar chart, scatterplot, SPLOM, etc.) as well as the variables (features) you intend to plot.

| [Please use this space for your response. You may expand or contract this box as needed.]  **Graphs & Visualizations:**   * Scatter plots showing the relationship between admit decision to ED departure time and readmission rates for heart failure and pneumonia patients. * Heatmaps displaying the variation in vaccination rates and readmission rates across different facilities. * Time-series graphs tracking changes in emergency department efficiency and readmission rates over time for specific hospitals. |
| --- |

## 

## 7. Ethical considerations

Does your choice of data raise any ethical issues? If so, briefly describe the concern and how you plan to mitigate it.

| * **Patient Privacy:**   + Although the data is aggregated at the facility level, care must be taken to ensure no indirect re-identification of patients occurs when analyzing small hospitals or specific geographic regions.   + Plan to mitigate: Adhere strictly to CMS and provider-level data privacy standards, and avoid detailed analysis at an individual patient level to minimize the risk of re-identification. * **Geographic and Demographic Bias:**   + Hospitals in specific geographic regions or serving certain demographic groups may have performance variations due to external factors (e.g., socioeconomic disparities), leading to biased conclusions.   + Plan to mitigate: Control for regional and demographic factors during analysis to ensure that variations in performance metrics are not solely attributed to geographic or socioeconomic disparities. * **Healthcare Equity:**   + Disparities in timely care metrics and readmission rates may disproportionately affect vulnerable patient populations, such as those requiring psychiatric care or living in rural areas.   + Plan to mitigate: Ensure that disparities in readmission rates and emergency care for different patient groups (e.g., psychiatric vs. non-psychiatric) are analyzed and reported to promote healthcare equity. |
| --- |

## 8. Contributions

Indicate the contribution that each team member will make to the project.

| [Please use this space for your response. You may expand or contract this box as needed.] |
| --- |

### 

## Changelog

(2022.07.27.1.CT) Update for 593

(2021.07.24.1.AW) Adjust title, number sections, simplify section headings, edit text