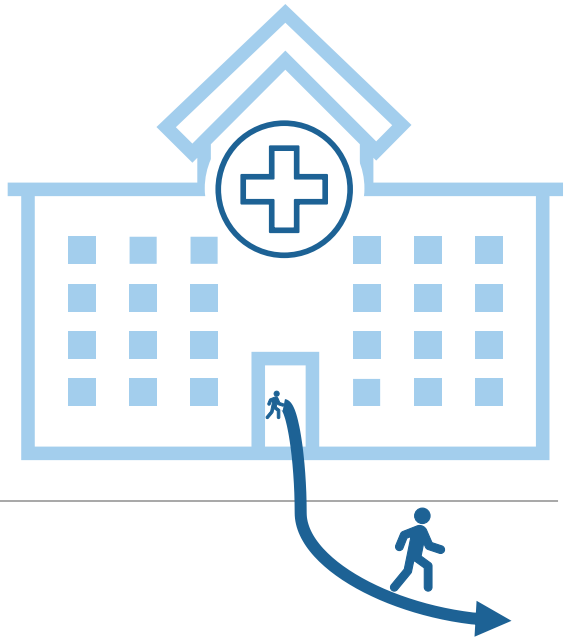
Patient Discharge Encounters



KENSUKE MORRIS

Overview

CalED
Emergency Department Design Company

Company

CalED is a fictitious company that develops user experience (UX) related solutions for healthcare facilities and patients.

CalED has created successful UX solutions relating to emergency visits and admissions.

Problem

To build upon their success, UX managers are considering designing UX solutions related to patient discharges but are unclear about which patient demographic they should target.

They are particularly interested in race groups and their association with discharge encounters in healthcare facilities in California.

Additionally, the managers are interested in finding new groupings for the healthcare facilities that can be useful for future UX research.

Objective

To conduct a preliminary analysis of discharge encounters related to patient race groups to find the main race groups and discharge profiles.

The analysis results should give UX managers an initial understanding of the main race groups at healthcare facilities, and a possible new grouping for facilities based on the main race group(s).

GitHub Resources: Pythons Scripts, Tableau Storyboard, Reflections

Project Resources

Skills

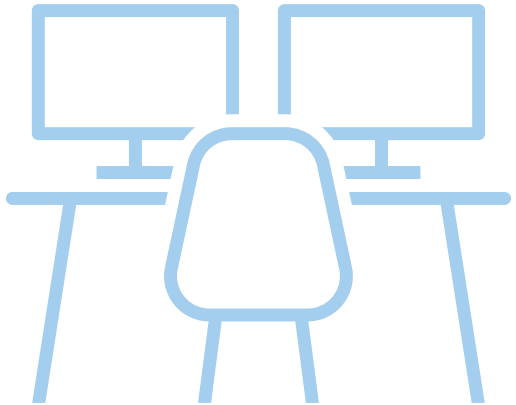
- Cleaning
- Wrangling
- Exploration
- Aggregations
- Regression
- Clustering
- Visualization
- Storytelling

Datasets

- Hospital ED – Characteristics by Facility (Main)
- Hospital ED – Encounters by Facility

Role

Data Analyst



Tools: **Python** from Preparation to Analytics, **Tableau** for Dashboard Reporting

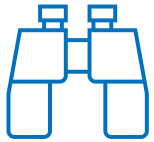
Questions

Preparation

1. Where can I locate ED data in about California?
2. Is there race- and discharge-related variables for my analysis?

Analysis

1. Which are the most frequent discharge profiles?
2. What are the main race groups?
3. Are there new groups for facilities based on the chosen race group?
4. Are there any additional points that can supplement the UX managers' questions?



Preparation

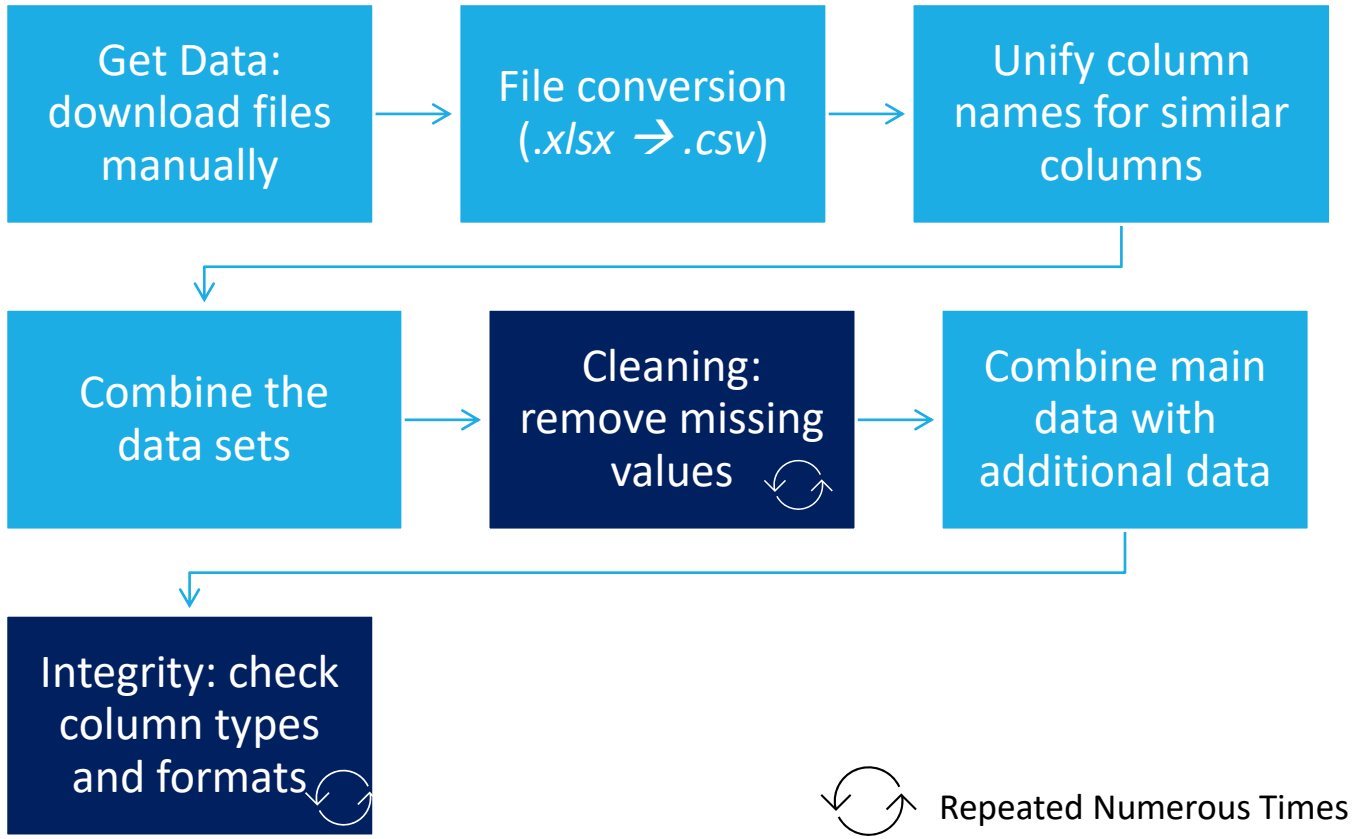
Answers to questions

1. The data was from an authoritative source, the Office of Statewide Health Planning & Development (OSPHD)

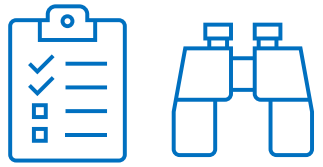
2. The data contained discharge encounters that were separated based on type. For example, routine discharge, discharge against medical advice etc.

I excluded 2005 - 2011 from the main dataset because the additional data was from 2012

My steps during the preparation stage



Challenges



Challenge: The sourced datasets contained inconsistent column names and variable types specifically for categorical variables.

Solution: Before concatenation by year, I reviewed each dataset manually to compare column names and matching variable types. Since the naming conventions were not consistent, this process took more time than anticipated. I excluded irrelevant columns during cleaning; this saved more time during data exploration.

Challenge: During data exploration, race group encounters in the dataset were in a "wide" structure, hindering the exploration of race group encounters by year.

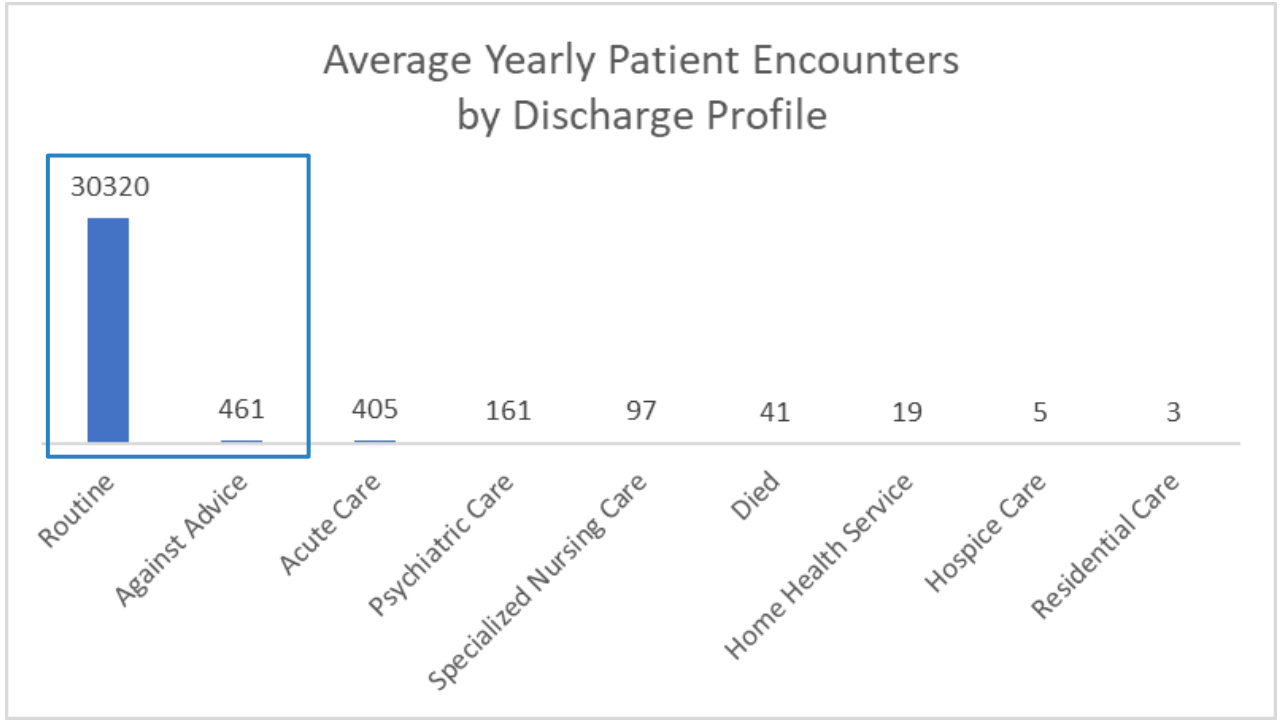
Solution: Prior to analysis, the race group variables were pivoted into one categorical variable with their corresponding values in an adjacent variable. I was then able to visualize race group encounters by year on a chronological graph.

Findings

1. Which are the most frequent discharge profiles?

On average, ED discharge profiles relating to **routine** and **against medical** advice were the most frequent discharge profiles.

I did not include **acute care** because it is assumed that the patient was transferred another ward and did not leave the healthcare facility.



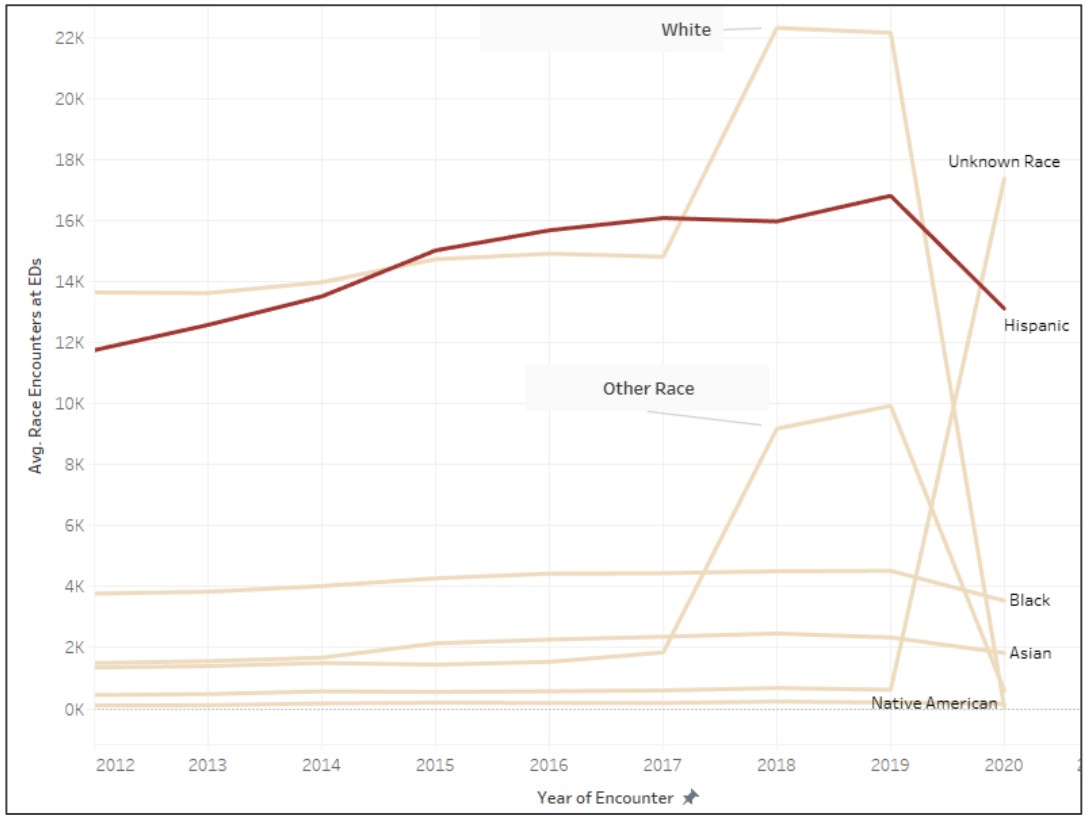
Findings

2. What are the main race groups?

Although **White** patient encounters were higher than **Hispanic** patients in recent years, **Hispanic** patient encounters were chosen because:

i. They had the strongest association with the two chosen discharge profiles (routine, against medical advice)

ii. Their encounter were available for 2020 in the dataset.



Findings

3. Are there new groups for facilities based on the chosen race group?

Low Encounters

Medium Encounters

High Encounters

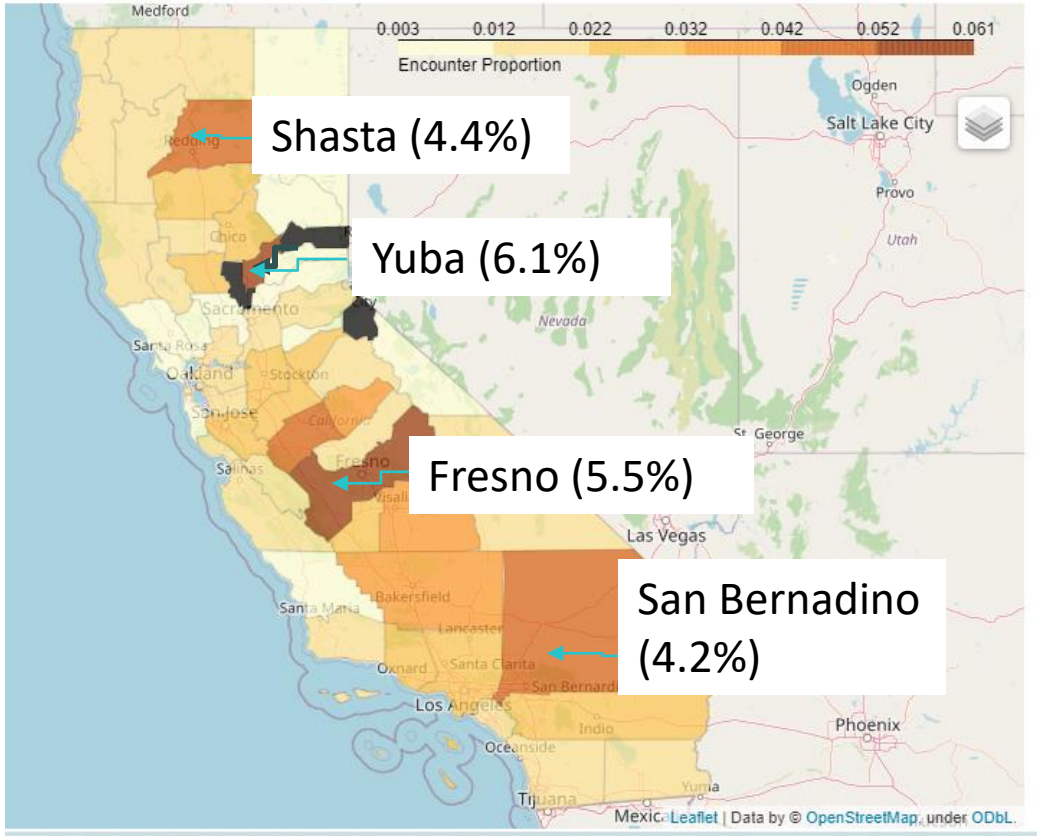
At a facility level, **three groups** based on **Hispanic** patient encounters were found.

Findings

4. Are there any additional points that can supplement the managers' questions?

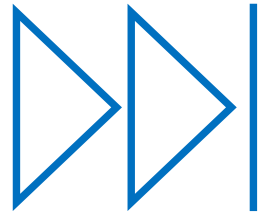
Since 2012, Yuba and Fresno had the highest proportions of discharges **against medical advice**. I assumed that larger counties would have higher proportions of discharge **against medical advice** but, the opposite was true. I can assume that county size does not dictate the proportion of discharge encounters **against medical advice**.

Routine discharge proportions were similar across counties.



Recommendations for Next Steps

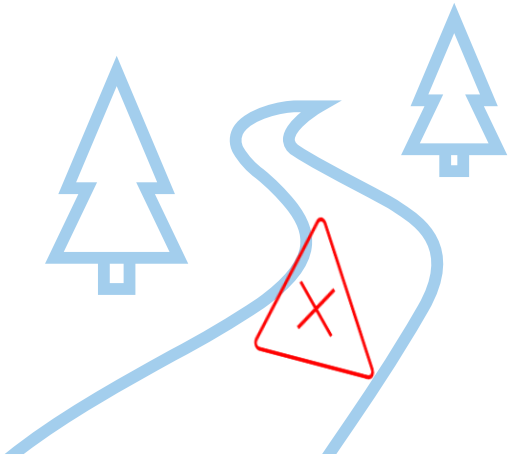
- Conduct a qualitative investigation into patients' reasons for discharge **against medical advice**. Use **Hispanic** patients as the first target group. Initially target **Yuba** county due to high proportions of discharge against advice and its small size.
- Segment healthcare facilities into **Low, Medium and High** encounters when choosing the samples for patient interviews.
- Survey patient preferences and concerns during discharge in California based on the results on of the qualitative inquiry.
- Since this was a preliminary analysis, collect additional data on a facility level related to race groups. On a patient level, anonymized data relating to patient discharges are also required, if possible.



GitHub Resources: Pythons Scripts, Tableau Storyboard, Reflections

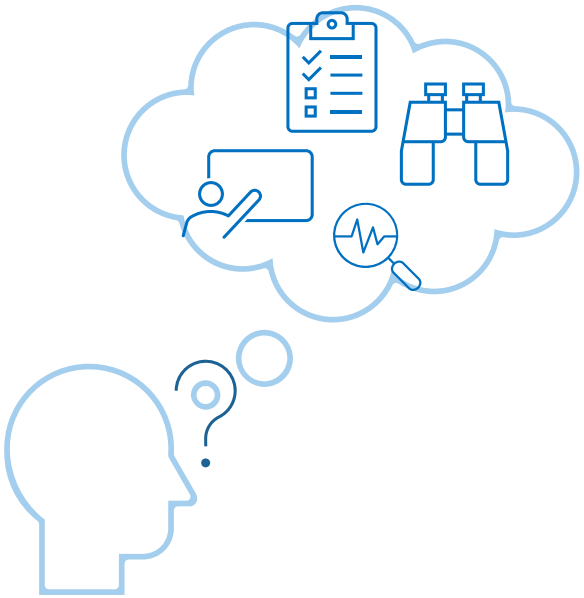
Data Limitations

- Aggregated data on a facility level was used for this preliminary analysis, thus, further analysis is required into clarifying the relationship between race groups and ED discharges were restricted.
- Repeated patient encounters could not be separated.
- Detailed discharge profiles for each race group were not available
- Despite the limitations, I think that the results of this preliminary investigation can assist the UX managers at CalED with their contextual understanding of discharge encounters in each county.



Retrospective

- I later realized that the main dataset chosen was adequate for a limited number of the initial questions. This was because the columns had aggregated values for discharge encounters of patients for each facility. Next time, I'll source "long data" as the main dataset to reduce limitations for further statistical analysis. The main dataset was structured as "wide" aggregated data.
- Next time, I'll subset discharge profiles, admission and visit encounters before cluster analysis. Additionally, before clustering, I will explore using a dimensionality reduction technique such as Principal Component Analysis.
- The many mistakes made during data wrangling and cleaning helped improve my skills. These mistakes increased my understanding of the reality of data cleaning and its importance, although time-consuming. If I had sourced a precleaned dataset, I might have finished this project faster. But the experience of data cleaning would have been one of my greatest losses.



GitHub Resources: Pythons Scripts, Tableau Storyboard, Reflections