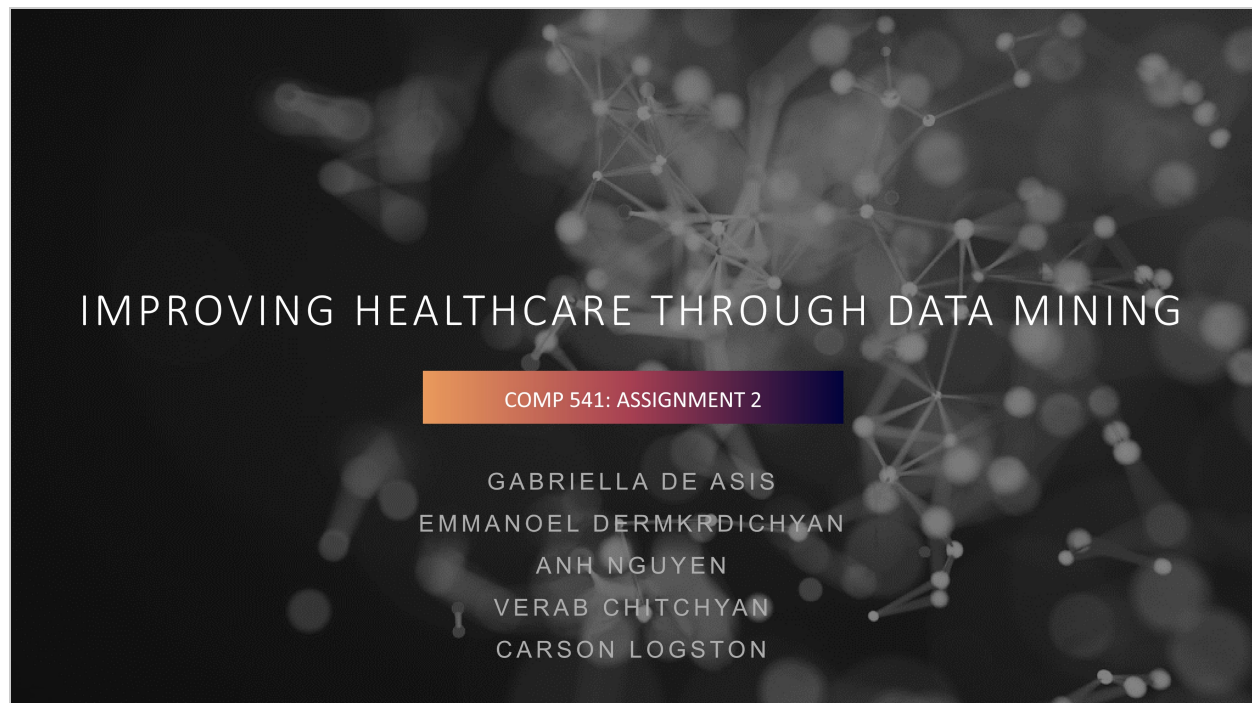


Improving Healthcare through Data Mining



- **Datasets acquired:** So far, we have found two datasets, Registered Nurse Shortage Areas and COVID-19 Reported Patient Impact and Hospital Capacity by State.
- **Locations:** We found both datasets from healthdata.gov
- **Methods of acquisition:** We found these datasets by researching COVID-19 cases and resulting nursing shortages.
- **Problems encountered:** With the massive size of the datasets, some online viewers did not work, and we needed to download the related files before being able to see the dataset. Other than the ability to preview the data before downloading it, we did not encounter any problems.

DESCRIPTION OF DATA



DATA TYPE



TYPE OF DATASETS



ATTRIBUTE TYPES

The data we acquired are listed as table in a CSV file. For the COVID-19 Reported Patient Impact data set, each file contains daily records from July 2020 until today. Each table contains columns of information on staff shortages, bed occupancy, patient population, ICU capacity, and so on. There are 53 rows in each table for the 50 states, DC, Puerto Rico, and the U.S. Virgin Islands. The Registered Nurse Shortage Areas dataset contains information on 72 areas within California. This table lists the area population, number of licensed nurses in the area, and the number of employed nurses. The nurse shortage for each area is classified as low, medium, high, or severe.

- **Type of data:** We are working with structured data since it can be described by a set of attributes. Attributes of our data set include state name, staff shortage, bed shortage, and other related information .
- **Type of datasets:** Our datasets can be categorized as tabular, sequential, and relational. It is tabular since that data is organized as a table with rows that represent a location. The columns represent attributes. The data is considered sequential since it is organized by date. This is important since we are trying to predict patterns of future patient influx and nurse shortages. It is a relational dataset since the tables have entries for rows and attributes of columns. This fits the description of a relational dataset. Furthermore, it has a primary key, which is the state name in one set and the name of an area within California in the other.
- **Attribute types:** Our data has nominal attributes such as state names. It also has ratio attributes like the number of hospitals that had a staffing shortage on a particular day in relation to the hospitals that did not have a shortage within the same state. Another ratio is the percentage of beds that are utilized and percentage of inpatients with COVID. Numeric attributes include the interval-scale attribute such as reporting cutoff dates and ratio-scale attributes. Discrete attributes include state names and counts such as critical staff shortage information and continuous attributes such as the percentages mentioned before.



PRESENTATION SUMMARY

DATASETS

DESCRIPTION OF DATA

TRAVNUR AGENCY