

TravNur was a project created for COMP 541: Data Mining

For the purpose of this project, TravNur is an agency that assigns traveling nurses to the hospital. The goal of this project is to anticipate the hospital's need for additional staffing with a 70% accuracy 1 week before the shortage occurs.

Datasets used:

- COVID-19 Reported Patient Impact and Hospital Capacity by State
- Source: [healthdata.gov](https://healthdata.gov)

Description of Dataset

File Type: CSV

Time period: January 2020 – April 2021

Information: staff shortages, bed occupancy, patient population, ICU capacity, suspected COVID cases, actual COVID cases, and more (63 columns in total) There are 53 rows per date (one for each state plus DC, Puerto Rico, and the U.S. Virgin Islands)

Cleaning and Analysis

- Prior to July 15, 2020, many of the entries were missing information. No reason was given on the website as to why this might be. Our best guess is that states were not required to provide that information prior to July 15, 2020. Because of the amount of incomplete information, we removed those entries from our study.
- New time period: July 2020 – April 2021
- Changed data types to string and int from float64 and object data types
- Using the correlation function on the columns of our dataset, we found that there was a high correlation (81%) between staffing shortages and total number of patients with covid and suspected to have covid. Patients were divided into total adult patients and total pediatric patients. There was also a high correlation between anticipated staffing shortage and patients with covid and suspected to have covid.
- `project4_group1_removal.py`
  - Takes dataset: COVID-19\_Reported\_Patient\_Impact\_and\_Hospital\_Capacity\_by\_State\_Timeseries\_aftrJuly15.csv
  - Removes rows with null values
  - Generates new dataset: AftrJuly15\_RowsRemoved\_1.csv
- `project4_group1_outliers.py`
  - Removes outliers more than 3 points from the standard deviation

- projec4\_group1\_normalization.py
  - Displays normalized graphs and relevant data

#### Project 5: Tested different models

- project5\_group1\_model.py
- project5\_group1\_model2.py
- project5\_group1\_model3.py
- project5\_group1\_model4.py
- project5\_group1\_model5.py

#### Project 6: Tried a linear regression model vs. a random forest classifier

- The linear regression model came out with more accuracy. See report for details.