Logo

Description automatically generated

Week 1 Assignment – LOCAL COVID-19 EXPLORATORY DATA ANALYSIS

Craig Perkins – [perkins.cr@northeastern.edu](mailto:perkins.cr@northeastern.edu)

Northeastern University

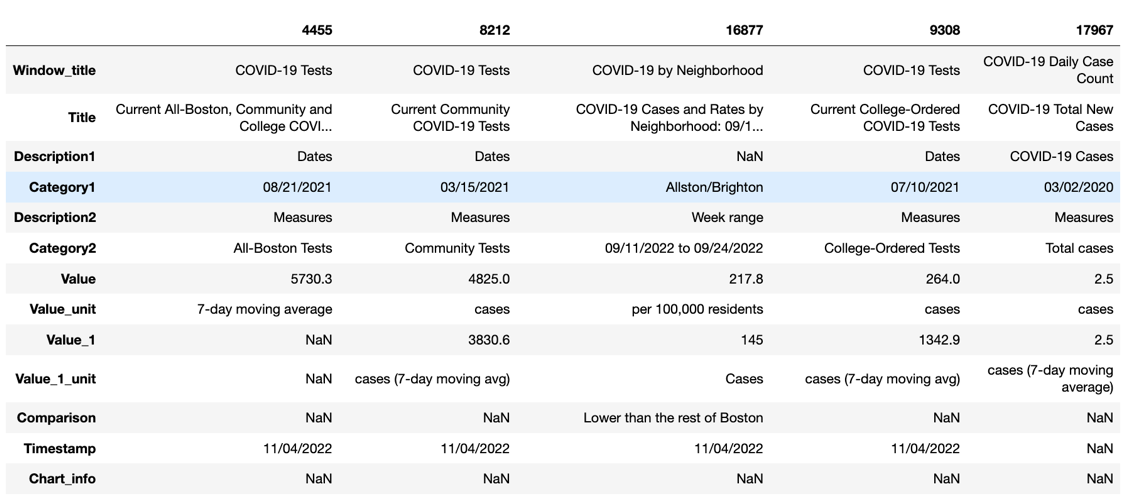
November 2022 – ALY 6140

Week 1 Assignment – LOCAL COVID-19 EXPLORATORY DATA ANALYSIS

For this week’s assignment I chose to use data from the Boston Public Health Commision1 with data on various covid-19 metrics like case count, deaths, hospitalizations and more for the Boston metropolitan area. For this analysis, I used the ` Boston\_COVID\_database\_download\_2022\_11\_04.csv` dataset which contains 3.4 MB (~20.5k rows by 13 columns) of data on COVID-19 dating back to January of 2020 up to November 4, 2022. This document will describe the data contained within the dataset and perform some initial data exploration to assess historical and recent trends in the number of cases, number of tests and the test positivity rate.

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| **Window\_title** | * **COVID-19 Positive Tests** * **COVID-19 Test Positivity** * **COVID-19 Tests** * **COVID-19 Emergency Department (ED) Visits** * **Confirmed Adult COVID-19 Hospitalizations** * **Occupied Adult Non-Surge ICU Beds at Boston Hospitals, All Patients** |
| **Title** | Name of the individual chart on the COVID-19 Metrics section of the BHCP website |
| **Description1** | Description of the first category (i.e. Dates, Covid-19 Cases) |
| **Category1** | The value of the grouping – usually a Date, Neighborhood, Race or Ethnicity |
| **Description2** | The description of the second category (If applicable) |
| **Category2** | The value of the second grouping (If applicable) |
| **Value** | The main metric being tracked |
| **Value\_unit** | Unit of the main metric being tracked |
| **Value\_1** | Secondary metric being tracked (often a rolling average, typically a 7-day avg.) |
| **Value\_1\_unit** | Unit of the secondary metric |
| **Comparison** | Use for neighborhood analysis to compare how the neighborhood compares to the city as a whole |
| **Timestamp** | Not positive on this, date of the data dump? |
| **Chart\_info** | This is not used for analysis, used to provide metadata about the dataset within the csv file |

Samples of rows within the dataset:



Descriptive Statistics of the dataset after dropping Chart\_info:

Table

Description automatically generated

The first analysis I performed was to visualize the number of positive tests by day using a line chart in pandas. Dates with N/A for the number of positive tests are treated as 0. Below is a chart of number of positive tests by day:

Graphical user interface, text, application

Description automatically generated

See the supplied jupyter notebook for more analysis, I only included the first chart of the analysis for this write-up.

**Conclusion**

In this assignment I performed an initial data analysis using data that will be used for the rest of the semester to get familiar with the dataset. Thanks to the efforts of the BPHC and other local institutions, we have rich datasets to analyze to obtain insight, create visualizations and to tell the story of the pandemic. Rich datasets allow us to learn more about complex systems by deploying analytical tools and can be used by people in authority to allocate resources efficiently. I hope to learn more in this class about the intersection of Healthcare and Analytics so that I can be prepared for careers in the healthcare industry as the industry transforms more into a data-driven era. There are many exciting initiatives in the healthcare industry unlocked by the power of cheaper and more ubiquitous computing power and it makes me exciting to see the future to come. Novel therapies, clinical decision support tools, precision medicine and recent innovations like mRNA vaccines show much promise and development of these technologies is accelerated by powerful computing power.

**References**

1. Boston Public Health Commission – Covid-19 Metrics <https://bphc-dashboard.shinyapps.io/BPHC-dashboard/>