Logo

Description automatically generated

Week 2 Assignment – MASSACHUSETTS COVID-19 EXPLORATORY DATA ANALYSIS  
  
Group 1

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

Sneha Sharma Satheesha - [satheesha.s@northeastern.edu](mailto:satheesha.s@northeastern.edu)

Pranay Bindela – [bindela.p@northeastern.edu](mailto:bindela.p@northeastern.edu)

Northeastern University

November 2022 – ALY 6150

Week 2 Assignment – MASSACHUSETTS COVID-19 EXPLORATORY DATA ANALYSIS  
  
Group 1

For our group assignment we are using statewide data from the covid tracking project: <https://covidtracking.com/data/state/massachusetts>.The dataset has daily statewide metrics across a wide variety of areas.

|  |  |
| --- | --- |
| **Field Name** | **Description** |
| **date** | Date of the measurement |
| **state** | **ma** for every row in this dataset (column to be removed) |
| **death** | Cumulative count of death confirmed + death probably |
| **deathConfirmed** | Cumulative count of death confirmed |
| **deathIncrease** | Daily increment of death count |
| **deathProbable** | Cumulative count of probably death attributable to COVID-19 |
| **hospitalized** | Cumulative count of hospitalized (weekly) |
| **hospitalizedCumulative** | Cumulative count of hospitalized (weekly) |
| **hospitalizedCurrently** | Count currently hospitalized |
| **hospitalizedIncrease** | Daily increment of hospitalized |
| **inIcuCumulative** | Cumulative count of ICU visits attributable to COVID-19 (Blank – column to be removed) |
| **inIcuCurrently** | Current count of ICU visits attributable to COVID-19 |
| **negative** | Cumulative count of negative tests |
| **negativeIncrease** | Daily increment of negative tests |
| **negativeTestsAntibody** | (Blank – column to be removed) |
| **negativeTestsPeopleAntibody** | (Blank – column to be removed) |
| **negativeTestViral** | (Blank – column to be removed) |
| **onVentilatorCumulative** | (Blank – column to be removed) |
| **onVentilatorCurrently** | Count on ventilator currently |
| **positive** | Cumulative count of positive tests (Antibody + Antigen + Viral) |
| **positiveCasesViral** | Cumulative count of positive viral tests |
| **positiveIncrease** | Daily increment of positive increase |
| **positiveScore** | (Blank – column to be removed) |
| **positiveTestsAntibody** | (Blank – column to be removed) |
| **positiveTestsAntigen** | (Blank – column to be removed) |
| **positiveTestsPeopleAntibody** | Cumulative count of tests positive for antibodies |
| **positiveTestsPeopleAntigen** | (Blank – column to be removed) |
| **positiveTestsViral** | Cumulative count of positive tests |
| **recovered** | Cumulative count of recovered (weekly) |
| **totalTestEncountersViral** | (Blank – column to be removed) |
| **totalTestEncountersViralIncrease** | (Blank – column to be removed) |
| **totalTestResults** | Cumulative count of tests |
| **totalTestResultsIncrease** | Daily increment of tests |
| **totalTestsAntibody** | (Blank – column to be removed) |
| **totalTestsAntigen** | (Blank – column to be removed) |
| **totalTestsPeopleAntibody** | Cumulative count of antibody tests |
| **totalTestsPeopleAntigen** | Cumulative count of antigen tests |
| **totalTestsPeopleViral** | Cumulative count of viral tests |
| **totalTestsPeopleViralIncrease** | Daily increment of viral tests |
| **totalTestsViral** | How is this different to totalTestResults? |
| **totalTestsViralIncrease** | How is this different to totalTestResultsIncrease? |

Samples of rows within the dataset after removing duplicate columns, blank columns and columns that do not provide useful information:

Table

Description automatically generated

Descriptive Statistics of the dataset:

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:

Chart, histogram

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**

<Conclusion here>

**References**

1. The COVID Tracking Project – Massachusetts <https://covidtracking.com/data/state/massachusetts>