COVID Relief Project Scope

Proposed hypothesis 1: As the age of a person increases the number of hospitalizations and deaths in the 70+ age group will be the highest fatality(deaths/total cases).
Proposed hypothesis 2: Also,there is a direct correlation between the age of a person and the density of the boroughs. A 70+ person living in a denser borough will have a higher chance of exposure and have a higher fatality than someone living in a less dense borough.

Research Statement

We are tasked with looking at available information to identify those in the city with the biggest risk of covid. It is documented that elderly are the ones who are most at risk for the covid virus. By analyzing the trends of Covid-19 as it applied to how age affects mortality in NYC we will be able to reach a conclusion and prove/disprove our hypothesis.

Target Audience

- Policy makers: local, state and federal government
- Hospital administrators
- Social entrepreneurs : Ex TKH
- Non-profit organizations: Ex NY Cares

Research Beneficiaries

- Healthcare workers and essential staff.
- Local small businesses owners and employees.
- Non profit Organizations

Study Methodology

Proposition of data factors to be assessed

- Age
- Neighborhood Density NYC 5 boroughs
- Infection Rate
- Fatality Rate

Parameters being used for research

- Large enough sample to represent the actual population.
- Median
- Mode
- Standard Deviation
- Standard Error

Analysis

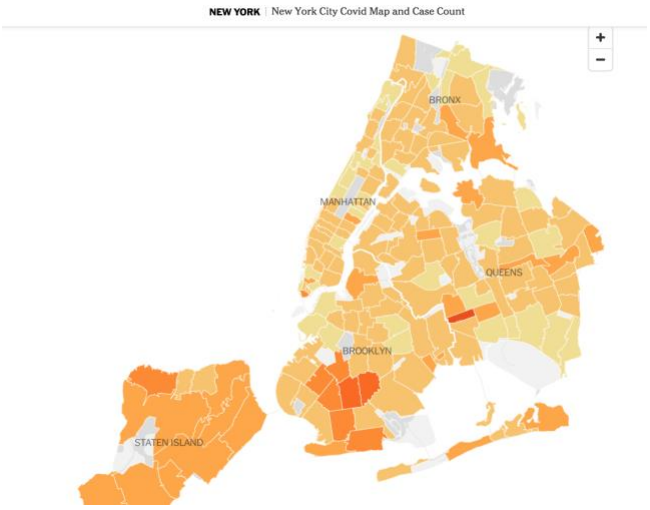
We will use quantitative and statistical analysis by using a combination of Bokeh, Power BI and Tableau to leverage data via CSV files and create a dashboard containing all the visualizations using python plots from Seaborn and matplotlib. Some of the variables we will use are age, ethnicity, location and occupation to gauge what groups suffered the most COVID-19 fatalities.

Visualizations proposals:

Heat Maps

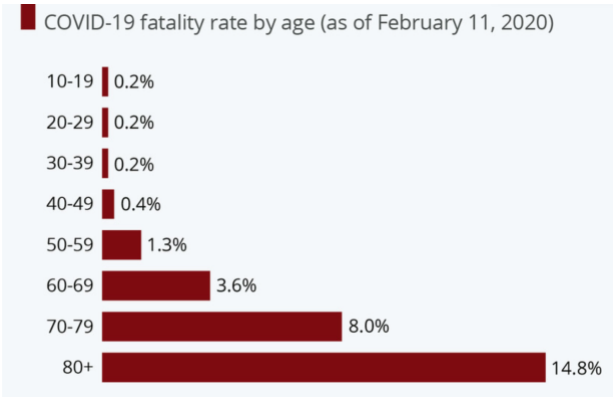
We can create three separate heat maps comparing:

- Location and rates of infection
- Correlation between neighborhood density and infection rate

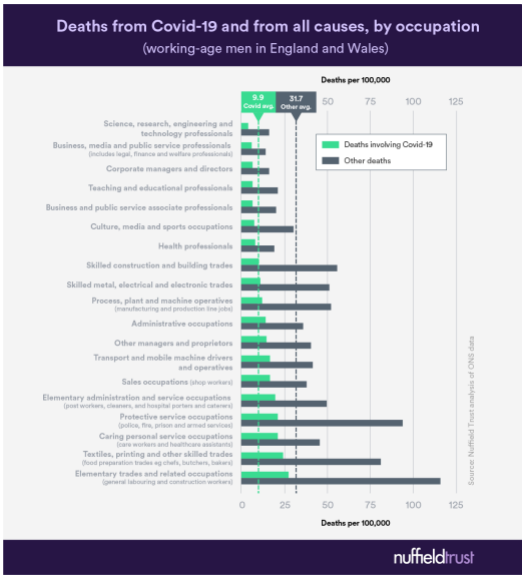


Bar Charts

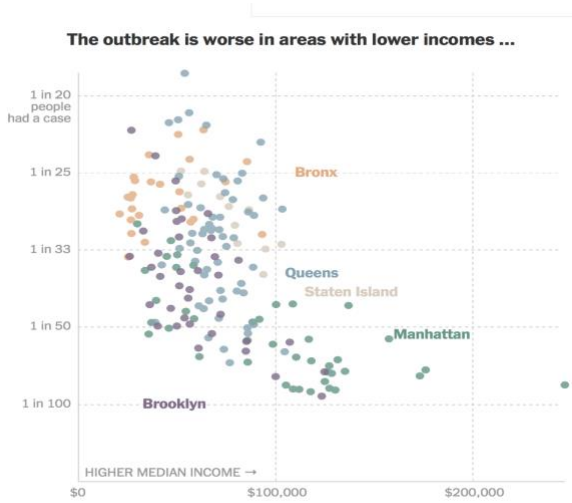
A bar plot will help us represent fatality rate percentages against the age variable.



Variables: Age, Fatality Rate



Scatter plot sample



We will use scatter plots to analyze the pattern between location density and infection rate.

Variables: Location density and rates of infection

Dataset/CSV Files:
https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm

COVID Relief Project Estimate

Timeline

This project will follow a 9 Month timeline starting at the approval of the project.

Milestones

Month 1-3 Emphasize, Define, Ideate

I. Research Proposal (Weeks 1-4):

Write an overview of our covid relief project. The overall value to the community of the research we will be doing. Our research assumptions and a description of those who would contribute to and benefit from the research and findings (user mapping). Finally, providing potential data sources.

II. Project Scope (Weeks 4-6)

We proposed a hypothesis for our covid relief research and the data factors to be assessed. We included the types of parameters we will use for the research and the types of analysis and visualizations our final dashboard will contain.

III. Project Estimate (Weeks 6-8)

We amended our research proposal and scope document with a timeline and phased hour estimate. Including our project breakdown with included and not included costs.

IV. Project Presentation (Week 8-12)

We created a slide desk to compliment our study proposal and support our study proposal pitch. The presentation will bring up things that we might want to add on or clarify before proceeding with prototyping and testing.

Month 4-6 Prototyping and Testing

I. Web Crawling, Data scraping and Cleaning (Weeks 12-16):

Scraping websites for data that will be beneficial to our project and cleaning data to identify what we will need to answer our research hypothesis. The cleaning and munging of data will be done in Python. Data can be prepared and ready for visualizations using matplotlib plots. Time for cleaning depends on how varied / distributed / noisy. Cleaning data will take roughly 4-5 hours per week.

II. Visualizations (Weeks 16-20)

Using Bokeh and Tableau to create visualizations with the data that was collected during the scrapping period. Also it is suggested to break down visualizations by categories, example: state wide (with state-level heatmaps). A drop down in the dashboard will be there to select the borough and view respective graphs.

III. Quality Assurance Validation (Weeks 20-24)

Go over the visualizations to make sure the hypothesis and research points are being answered to cover how covid-19 has impacted our main users.

Month 6-9 Dashboard Builtout

I. Dashboard and integration of Visualizations (Weeks 25-29)

Adding dashboards to Tableau from a cleaned dataset may take fixed time per day per week and would be an evolutionary process. To have an estimate, a sample storyboard takes 1 to two weeks to make. A summarized version of the visualizations will be integrated into a one page interactive dashboards.

II. Work on data Storytelling and Analyzation (Weeks 29-31)

Creating documentation or white paper to summarize and explain the findings presented on the dashboard can take about two weeks.

III. Dashboard Web Hosting(Week 31-33)

Hosting a web version of the dashboard including articles explaining each visualization can be done using a hosting provider that would allow uploading a content management system to include the database with all the research contents. 15 days is a very good estimate to present a working version that we can test for quality assurance.

IV. Testing and Maintenance (Week 33-36)

Test for quality assurance: can be implemented doing some analysis and regression testing to make sure the web page is working properly to display the visualizations and dashboards. In terms of the maintenance we can debrief how often a maintenance check will be run on the web hosting to verify it has the latest update for security and up-time for web availability purposes. It is recommended that for the lifetime of the research weekly wellness check be scheduled to make sure the website and the content is up and running.

Project Cost Breakdown:

- Hosting + domain $200 for 3 years.
- $7200 ($1,800 for each Data Scientist * 4)
- Miscellaneous: $1,600 (Licenses)
- Maintenance: $25/hr- 7 hours per week during the first 3 months after the end of the project.

Final Project Cost : $11,100.00

Additional Project Tasks Not Included:

- Security Maintenance
- Copywriting

Additional Project Costs Not Covered:

- Additional hosting after the 3 years - $200
- Additional research to update data(including additional visualizations)- $40 per overtime hours for up to 12 months.
- Additional Maintenance: $30 per hour for any additional maintenance following the 3 months included maintenance. Valid for up to 12 months after the completion of the project.