Statistical Data Analyst

QUESTION 1:

The director of York Region's Infectious Disease Control Division wants to better understand the workload associated with our COVID case counts to optimize staffing assignments.

You have been asked to develop a dashboard to support these staffing decisions, with a focus on three case investigation teams:

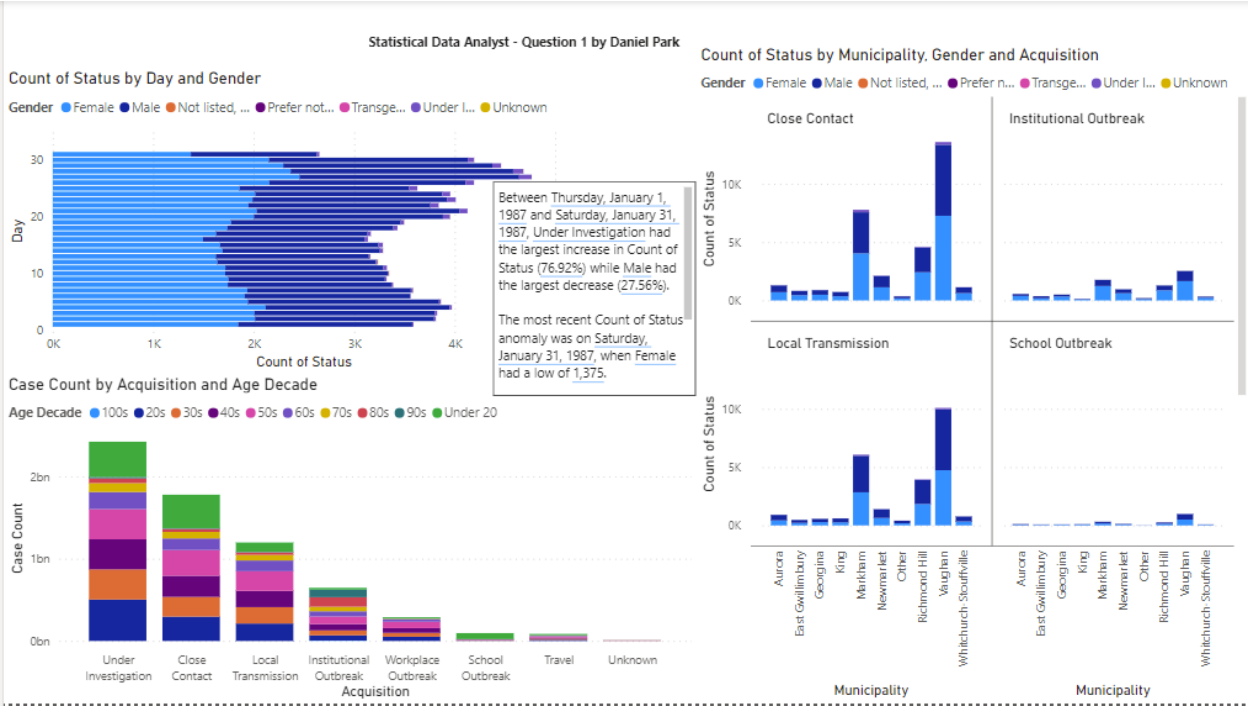
- The Institutional Outbreak team
- Travel Case Management team
- Community Case Management team (who follow up on all cases that are not associated with outbreaks or travel).

The data is available on http://york.ca/covid19data, downloaded from the bottom left corner of the dashboard.

Please prepare a feature visual that would be used to support the above business problem:

- You can use any method you like to prepare the visualization.
  - PowerBI is preferred as it is our standard internal visualization software.
  - Tableau or other visualization software is also acceptable.
- Please include a paragraph explaining your design choices in preparing this visual.

Answer to Question 1:



For the visualization, I decided to use Power BI and created 3 visuals and a text summarizing some of the information. I made sure that the visuals weren't redundant, and that each visual

conveyed new information to the audience. I hoped to consider the most relevant and meaningful criteria to create my dashboard.

QUESTION 2:
You have been assigned to work with an epidemiologist to understand COVID trends over time in York Region. The epidemiologist will help to explain the content and the clinical value of the data you are working with. As the statistical data analyst, your role is to develop the dashboard visualization.

The project deliverable is a dashboard that the epidemiologist can use for weekly presentations to a group of COVID case investigators and management. This presentation will highlight any changes in the recent trends as well as any subpopulations with higher risk (subpopulations can be defined by age group, municipality or vaccination status or other relevant characteristics).

The epidemiologist wants the dashboard to show rates over time, with the following formula:

$$\frac{\text{Total COVID Cases/Hospitalizations/Deaths per week}}{\text{Total Population}}$$

Since severe outcomes (hospitalizations and deaths) are less common than cases, the epidemiologist is expecting the trend to be very noisy and thinks this might be confusing to the audience of case investigators.

Our COVID-19 database includes a table of all confirmed COVID-19 cases from the beginning of the pandemic to present, including:

- demographic characteristic (age, gender, residential location);
- severity information (hospitalization and/or fatality);
- COVID-19 disease details (variant of concern or mutation of interest);
- vaccination status (doses and brands);
- testing type and history (PCR or Rapid Antigen Test); and
- involvement in an long-term care home or retirement home outbreak (if relevant).

Please prepare an analysis plan for how you would develop the requested dashboard, including the following:

- An outline of the analysis steps you would take to prepare this dashboard.
- Bullet points are acceptable.
- The outline may include: data preparation steps, analysis methods, software choices, plan for interpretation and analysis, any secondary analyses of interest, and other considerations for the analysis.
- Please include a brief explanation of the decisions that went into preparing this analysis plan, including your understanding of the audience's needs, and your rationale for the visualization and analysis choices to meet that need.
- **You do not need to do the analysis, instead outline the analysis plan**

Answer to Question 2:

Analysis Plan:

- Visualizations will be used and a dashboard will be created to provide insight into our dataset and identify trends in COVID-19 cases from the beginning of the pandemic to the present.
    - Tasks such as loading data, cleaning data, correcting errors, and choosing and configuring visualizations will be conducted.
    - Combining the data from multiple sources into coherent time-series information for the COVID-19 cases by loading the data
    - Upon loading the data, we need to view errors, as data is never in perfect shape for analysis when we get it - we need to clean it first.
        - Some errors include: changing number types and replacing values
    - Cleaning and transforming our data to get it ready for analysis and visualization.
        - Some cleaning and transformation tasks are: naming columns properly, giving columns meaningful names, and removing unnecessary columns or rows
    - After cleaning and transforming the data, our dataset is ready for analysis and visualization.
        - Select a plot template/type from the visualization pane
        - We can plot graphs and colour the data values of the chart with the regions to identify the trends of the COVID-19 cases
        - Add the appropriate legend and reorganize the field values if necessary
    - Examine the visual produced to see what kind of trends and relationships are present and create other visuals if necessary based on:
        - Examining recent trends, as well as other subpopulations characterized as higher risk(based on age, municipality, vaccination status, or other relevant characteristics), demographic characteristics, severity information, disease details, vaccination status, testing type and history, and involvement in a long-term care home or retirement home outbreak(if relevant)
        - If necessary, further evaluate trends for noise and explanations
    - Once we have created some visualizations, we can now create a page and put it all together in a dashboard.
        - Start a new page, and copy the visuals you created to the page
        - Align the visuals properly
        - Edit the interactions
        - Explain and clarify, and add notes to convey results or if there is anything that may be confusing to the audience of case investigators
    - Share the information for analysis in our visualization
        - Save our work to show it or to share it with others
        - Publish it on the Power BI service to make it available to the public, which can help showcase your skill