COVID-19 VACCINATION DATA VISUALIZATION BY PUBLIC HEALTH UNIT OF ONTARIO

|  |  |
| --- | --- |
| Group Number: 22 |  |
|  |  |
|  |  |
| Team Members: | Instructor Name: |
| Kartik Shah | Prof. Fadi Alzhouri |
| Drasti Shah (0693615) |  |

Introduction:

Thanks to the Covid-19 vaccination drive which has picked up pace across the world, that we are now in a position to beat, control and avoid the widespread coronavirus. However, what has been of extreme importance during the whole pandemic is our access to the huge amount of data and availability of tools to manipulate and visualize it.

The aim of our project was to use one such dataset on real- time Covid-19 vaccination statistics to develop a dynamic and synchronous visual dashboard which gives us insights about the current vaccination numbers, hospitalizations according to various filters like age groups and month of vaccinations. The data has been extracted from the Ontario government’s official website, transformed using Python and stored in a database in MySQL. This data on being loaded in Tableau has then been used to create a dynamic dashboard with useful insights.

The report discusses our objective and procedure to create the visualizations. It further goes on to describe data and the steps followed for data preprocessing. The references and resources are listed followed by the appendix containing the code performed to follow the Extract-Transform-Load (ETL) process.

Project Objectives:

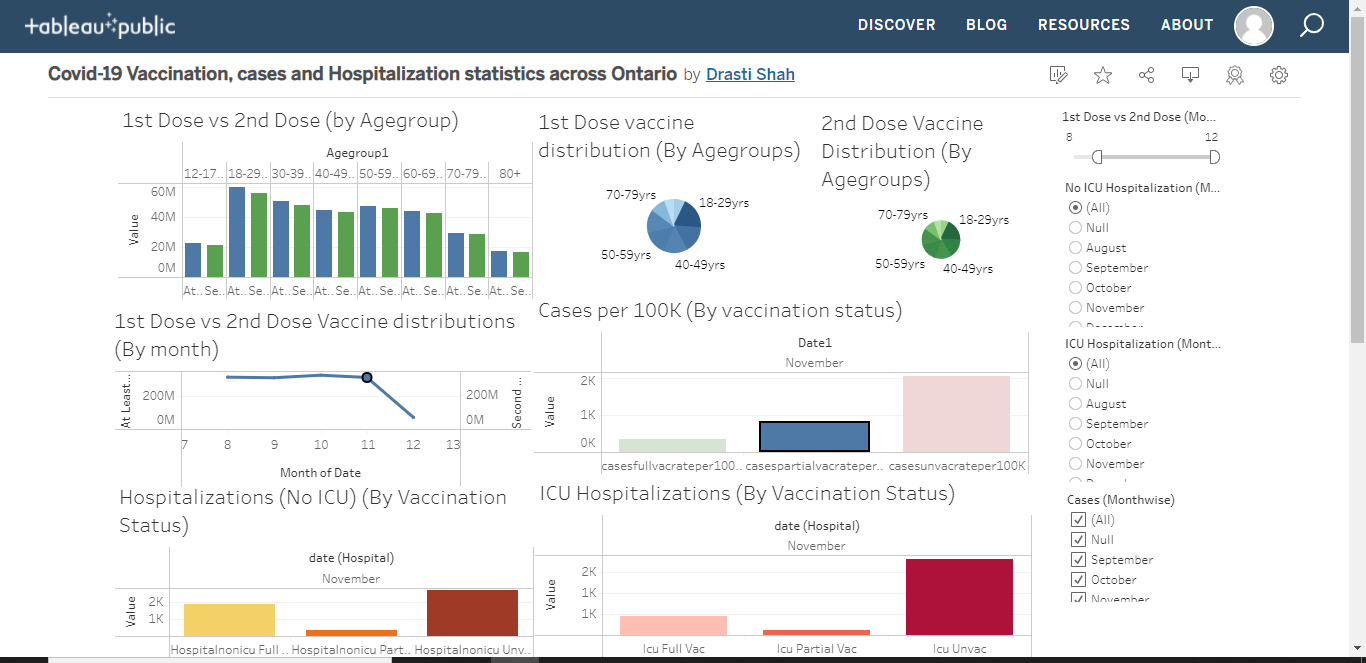
As students we are always taught that “A picture speaks a thousand words.” With this prime objective in our mind, we began exploring the Covid-19 statistics and the vaccination numbers as well. Our goal was to present as much information as we could in a concise format for the users to understand and analyse.

Moreover, Frank Lloyd Wright once said that "Form follows function—that has been misunderstood. Form and function should be one, joined in a spiritual union." Indeed, this is true and form and function should go hand-in-hand. With the intention to balance the design and functionalities of the Covid-19 vaccination dataset and others, and to develop a dashboard which updates itself with the real-time data, we have created multiple insightful visualizations using Tableau.

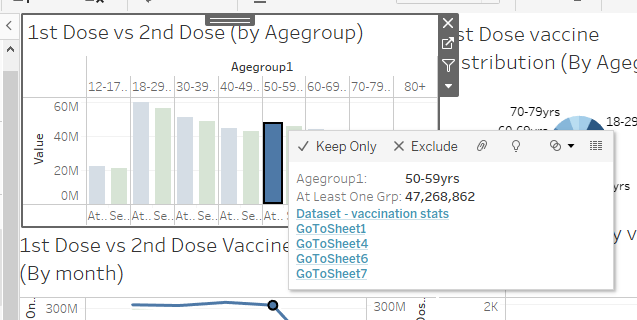
Secondly, we wanted to merge and put together the useful data from various different government websites to be able to connect and gain additional information from the same.

Description of visualization:

We have created a tableau [dashboard](https://public.tableau.com/app/profile/drasti.shah/viz/Covid-19VaccinationcasesandHospitalizationstatisticsacrossOntario/Dashboard1?publish=yes) which has all the information visualized and ready for the users to view. When viewed it looks like this:



Towards the extreme right, it has filters with the month of the data. As you apply the filters the relevant data visualizations show up. We have also given the links to the dataset and individual data visualizations for the users to view individually and analyse. When you click on the visualizations the following links open up (Note: The links vary for each visualization.):



On clicking the dataset link, you can view the dataset used to create the visualization and by clicking the Go to sheet links you can view the zoomed in individual visualizations along with their legends. This makes it easier for the users to analyse the data of interest. We have used bar charts, pie charts and line chart depending on the data in use. We have maintained consistency in our colours and fonts, provided headings as well as labeled our data. We have tried and added maximum features like filters, links to datasets and individual visualizations, to create a story for our users to view and make sense of easily.

Data description and Data Preprocessing:

Our [dataset](https://data.ontario.ca/en/dataset/covid-19-vaccine-data-in-ontario/resource/2a362139-b782-43b1-b3cb-078a2ef19524) has 11 attributes, divided and put into 11 unique columns namely –

* \_id: This is the id column with integer datatype. This column is set to primary key since it is unique.
* Date: This is the date column of the vaccination dose statistics across different Public Health Units. The datatype in Date in the “YYYY-MM-DD” format.
* PHU\_ID: The id of the Public Health Units. The datatype for this attribute is integer.
* PHU\_name: The name of the Public Health Unit. The datatype is String.
* Agegroup: The dataset has vaccination numbers per age group per PHU. The different age groups are 12- 17 yrs, 18- 29 yrs, 30-39 yrs, 40- 49 yrs,50- 59 yrs, 60- 69 yrs, 70- 79 yrs and 80+ yrs. The datatype is string.
* atLeastOneGrp: The cumulative statistics of the number of people who received first dose. The datatype is integer.
* secondDoseCumulative: The cumulative statistics of the number of people who received the second dose. The datatype is integer.
* totalPopulation: This is the total population of a Public Health unit. The datatype is integer.
* percentAtLeastOneDose: The percentage of people who received first dose. The datatype is float.
* percentFullyVaccinated: The percentage of people who received both doses. The datatype is float.

For Data preprocessing, we have followed the following steps:

* We first decided upon and acquired the dataset using Data API. This allows us to view real-time and dynamic visualizations as the data changes.
* We then looked for and imported all the libraries which are crucial to our project. These include pandas, SQLConnection, sqlalchemy and pygsheets among others.
* We then extracted our dataset by calling the API.
* Upon extraction of data, we first looked for missing values. There was a column ‘fully\_vaccinated\_cumulative’ which was entirely empty. So, we removed it. Another column ‘rankDate’ was of no particular use so we dropped it as well. The remaining columns were put to use in the visualization. There were some values labeled as undisclosed, missing and so on, which we deleted as well during the visualization process. We renamed certain columns to fit into the format.

Resources and References:

* Art of Visualization (2017, October 3). *Tableau How to create a Dashboard*.<https://www.youtube.com/watch?v=Nr31rv9tsJ8>
* Statistic Canada (2021, December 1). *Covid-19 Vaccine Data by Public Health Unit*.<https://data.ontario.ca/en/dataset/covid-19-vaccine-data-in-ontario/resource/2a362139-b782-43b1-b3cb-078a2ef19524>
* S. Eklavya (2020, April 10). *Python Pandas Dataframe to Google Sheets for Tableau Public LIVE*. towards data science. <https://towardsdatascience.com/python-pandas-dataframe-to-google-sheets-for-tableau-desktop-live-cc1f86982bca>
* *Pygsheets Reference*. <https://pygsheets.readthedocs.io/en/stable/reference.html#pygsheets.client.Client.open>
* Learn Google Spreadsheets (2020, October 6). *Google Sheets – Python API, Read and Write data*. <https://www.youtube.com/watch?v=4ssigWmExak>
* Google Sheets for Developers > Sheets API. *Method: spreadsheets.values.update.* <https://developers.google.com/sheets/api/reference/rest/v4/spreadsheets.values/update>
* Google Sheets for Developers > Sheets API. *Python Quickstart*. <https://developers.google.com/sheets/api/quickstart/python>
* *Tableau Desktop and Web Authoring Help.* <https://help.tableau.com/current/pro/desktop/en-us/actions_filter.htm>

Appendix: