

Tools

- Please use VS code for the best experience

How to use the interactive dashboards:

- Make sure your port 8050 is free to use.
- Run the script.
- In order to run another dashboard, make sure to reset the kernel in the original script.

0_data_cleaning.ipynb

This file preprocesses and cleans the two provided datasets. The results will be stored in `cleaned_data` folder.

0_merging.ipynb

Please run this file after `0_data_cleaning.ipynb`. In this file, the two datasets merge and create `cleaned_data/merged.csv`

1_Dataset1_processing.ipynb

Dataset 1 specific analysis.

1_process_one_hot_encoding.ipynb

In this file, I implemented a one-hot encoding method to analyze the impact of different categories on each other including themselves by calculating their correlations.

This file includes an interactive dashboard to filter data fields.

1_process_plots.ipynb

This file generates different visualizations which will be stored in `report/dataset2` folder. It also includes an interactive dashboard for trends and distributions.

1_processing.ipynb

This file includes severity analysis and interactive dashboard.

2_histogram.ipynb

Specific for generating histogram results.

2_scatter_plot.ipynb

This file includes an interactive dashboard to analyze the effect of one data field on another one.

2_scatter_plot_merged.ipynb

This file generates figures for analyzing the effect of one data field on another one using `merged.csv` file generated in `0_meerging.ipynb` step.

3_prediction_dashboard.ipynb

In this file a predicting interactive dashboard is implemented to forecast the future injuries based on different filters.

Extra Resources

- I used Ontario and Canada's websites to find National Occupational Classification (NOC) and primary occupation code = North American Industry Classification System codes (NAICS) datasets to confirm they match with the "Occupation_[Category]_Code" in Dataset2 and "PRIMARY OCCUPATION CODE" column in Dataset1, respectively.
- I also used this website <https://worldpopulationreview.com/cities/province/ontario> to get a free available dataset for Ontario's city populations and used it in the number of injuries vs. city graphs.