|  |  |
| --- | --- |
| Course | Data Analytics Boot Camp |
| Assignment | Project #1: EDA and Git Collaboration |
| Team | Group 3: Gurpal, Stephen, Carson, Clarajean |
| Due Date | Thursday July 27, 2023 |

**Overview**

Prepare a formal 10-minute presentation that covers the following topics:

· Questions that you found interesting and what motivated you to answer them.

· Where and how you found the data that you used to answer these questions.

· Data exploration and cleanup process (Jupyter Notebook)

· The analysis process (Jupyter Notebook)

· Conclusions: Numerical summary & Visualizations of the summary

· Implications of your findings: what do your findings mean?

**Assignment Requirements**

· Use Pandas to clean and format your dataset(s).

· Create a Jupyter Notebook describing the data exploration and cleanup process.

· Create a Jupyter Notebook illustrating the final data analysis.

· Use Matplotlib to create 8 visualizations of your data (2 per question/factor).

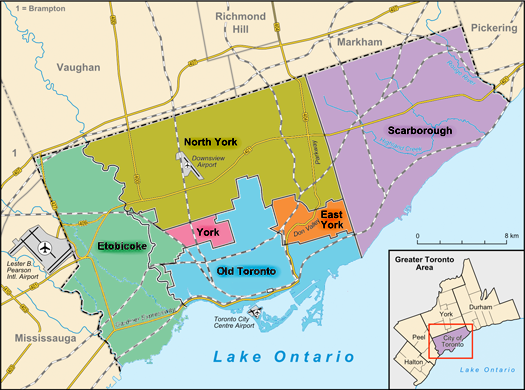
· Save PNG images of visualizations to distribute to class, professors and for the slide deck.

· Create a write-up summarizing your major findings. This should include a heading for each question/factor that you asked your data, as well as a short description of your findings and relevant plots.

· Bonus: Use at least one API – if you can find one with data pertinent to your primary research questions.

**Research Questions to Answer**

Districts of Toronto - Old Toronto, East York, North York, Scarborough, York, Etobicoke



1. Crime - To analyze the trend of total crime count over time per district. - STEPHEN
2. Pricing- To analyze the average real estate pricing per district and year over year growth/decline pricing trend over time per district? - GURPAL
3. School - Which district has the least to highest rated schools? - CARSON
4. Job Opportunities/Employers - Which district has a larger or lesser number of job opportunities/employers? - CLARAJEAN

**Datasets to Be Used**

Stephen - <https://open.toronto.ca/dataset/neighbourhood-crime-rates/>

Gurpal - <https://www.kaggle.com/datasets/alankmwong/toronto-home-price-index?resource=download>

Carson - search Geo API

Clarajean - <https://www.kaggle.com/datasets/youssef19/toronto-neighborhoods-inforamtion>

<https://open.toronto.ca/dataset/school-locations-all-types/>

**Project Description/Outline**

The purpose of the project is to determine which district in Toronto is the best choice to buy a home. Each member will get a dataset and filter the data to make it easier to understand. Each group member will utilise visual aids to analyse different factors and utilise different data sets in order to come up with a conclusion. Rank all districts from 1 to 6 in each factor and conclude which is the best district based on overall ranking across all factors.

The factors that we will use to analyse and come up with a conclusion to our question are crime rate, house pricing, school rating and amenities. Each group member will present their conclusions to the class.

**Rough Breakdown of Tasks**

Each team member completes all tasks under the Assignment Requirements for their own research question/factors below:

Crime - STEPHEN

Pricing- GURPAL

School - CARSON

Job Opportunities - CLARAJEAN