Data Understanding Competition

Rules:

- Teams of maximum 3 people [recommended]. Doesn't have to be your project team.
- You can discuss general ideas with other teams but not repeat their exact analytics process.
- One submission per team.
- You are allowed to use any external datasets you see relevant.
- Your submission MUST include:
  i. A 5-10 pages report stating: (1) Team members' names and student IDs, (2) the used software packages and datasets, (3) instructions on how to download and install them, (4) a detailed description of the analytics process used, (5) 10+ charts describing the data and showing interesting trends and patterns plus a paragraph describing what is interesting in each chart and (6) Your findings.
  ii. All code developed to produce the results.
- Compress all files into one archive that you upload to OnQ.
- Late submissions will be penalized 1 point for each late day.

This competition consists of one tasks.

Planning Bus Routes for the City of Kingston

The City of Kingston promotes and fosters open government principles of participation, innovation, transparency and accountability. It provides a data catalogue to support these principles and as a first step in making it easier to view, obtain and use the information the City has gathered. The datasets are available at:
https://www.cityofkingston.ca/explore/data-catalogue

Use the city's datasets and any other datasets you find relevant to answer questions like supported by visualizations:

1) Is the bus service accessible to everyone in Kingston (walk, bike, park & ride to a bus stop)?
2) Are there some redundant bus routes that can be cancelled?
3) Assume that each driveway has two cars and that all people work in downtown Kingston, how many new bus routes are needed to make people leave their cars at home? How much would that reduce the gas emissions (make your own reasonable assumptions)?
4) Is there anything else you find interesting in the data that can help planning the bus routes?

The Kingston data catalog includes the following datasets (You can use some or all of them):

- **Addresses**: Municipal addresses and address locations.
- **Buildings**: Dataset of the various classifications of building rooflines for structures greater than 5 sq. metres within the City of Kingston, Ontario.
- **Cycling Facilities:** Existing on and off-road cycling infrastructure within the City of Kingston, Ontario, with the type of cycling facility that is currently in place identified.
- **Driveways**: Dataset representing the driven portion of a private or public drive for the purposes of parking, or to access a parking area.
- **Neighbourhoods:** Dataset of the boundaries of neighbourhoods within the City of Kingston, Ontario, as established from the 2011 census dissemination areas. Includes neighbourhood names.
- **Parking Areas:** Dataset representing the driven portion of a private or public parking area (not to be confused with a parking lot, which is a separate dataset).
- **Points of Interest:** Dataset of the locations of specific point locations within the City of Kingston, Ontario, that are of common interest to all. Includes name of each place, address and more.
- **Roads (Centreline):** Dataset of the street network within the City of Kingston, Ontario. Includes addressing, road classes, status, street names and more.
- **Roads (Paved Surface):** Dataset of the representation of the travelled portion of streets and roadways within the City of Kingston, Ontario.
- **Transit Schedules, Routes & Stops (static):** Kingston Transit General Transit Feed Specification (GTFS) for static schedule, route, and stop information.

- **Transit Service Alerts (real-time):** Kingston Transit General Transit Feed Specification (GTFS) for real-time service alerts.
- **Transit Trip Updates (real-time):** Kingston Transit General Transit Feed Specification (GTFS) for real-time trip updates.
- **Transit Vehicle Positions (real-time):** Kingston Transit General Transit Feed Specification (GTFS) for real-time vehicle positions.

Resources (You don't have to use them):

- Sample programs analyzing the city data: https://mayors-ic.github.io/software
- Heat Mapper: http://www.heatmapper.ca/
- Google Charts: https://developers.google.com/chart/interactive/docs/gallery
- GGobi: http://ggobi.org/
- Free license to ESRI ArcGIS for map visualization: https://my.queensu.ca/software-centre
- Google Maps APIs for map visualization: https://developers.google.com/maps/documentation/
    - Javascript tutorial: https://developers.google.com/maps/documentation/javascript/tutorial
- Free Tableau license for different visualizations: https://www.tableau.com/academic/students
- Free 30-days trial Watson Analytics: https://www.ibm.com/ca-en/marketplace/watson-analytics
- Matplotlib & Pylab: Python graph plotting libraries.
- Excel Power View