Senior Data Analyst Recruitment - Take Home Assignment

The recruitment for the Senior Data Analyst position will consist of two parts, take-home assignment and an in-person interview. The following is information pertaining to the take-home assignment.

1.0 - Overview of the data analysis assignment

The goal of this part of the assessment is to evaluate the candidate's technical capabilities in the area of data analytics. The assessment will be completed as a take-home, the candidate will have four (4) days (including a weekend) to complete the assessment. The assessment will require the candidate to:

- access datasets
- perform analysis
- create visualizations
- save the work as an open source data analytics notebook (Python or R)
- share the notebook via code collaboration site (i.e. GitHub)

The assessment goals are:

1. Data analytics capabilities in either Python or R
2. Ability to perform data analytics within a "notebook" environment
3. Ability to work with large datasets
4. Coding abilities around automation (i.e. loops, functions, etc.)
5. Choosing and perform data analysis
6. Selection and presentation of visualizations
7. Capability to perform geospatial analysis
8. Ability to communicate data analytics effectively (story telling)
9. Understanding and capabilities using GitHub as a platform to share
10. Ability to clean and wrangle data

2.0 - Datasets for the assignment

The analysis will utilize a number of different datasets that are available on the City of Toronto's Open Data portal: https://open.toronto.ca/. Other datasets may be required based on the approach taken to complete the assignment.

The following datasets, as a minimum, are required for the analysis:

Dataset Name	*Open Data Portal URL*
Parking Tickets	https://open.toronto.ca/dataset/parking-tickets/
Green P Parking	https://open.toronto.ca/dataset/green-p-parking/
TTC Routes and Schedules	https://open.toronto.ca/dataset/ttc-routes-and-schedules/
City Wards	https://open.toronto.ca/dataset/city-wards/

3.0 The data analysis assignment

The following are the areas of evaluation during the assignment:

3.1 Analysis

1. What are the top 20 ticket infractions (frequency)
2. What are the top 20 ticket infractions (revenue)
3. Total revenue generated from all tickets
4. Are there alternative mobility options available for each of the top 20 infractions
    a. How far (as the crow flies) is the closest parking lot (Green P) to each of the top 20 infraction locations
    b. How far (as the crow flies) is the closest TTC stop to the top 20 infraction locations
5. Analyze the impact of the following on all infractions:
    a. Day of week
    b. Month
    c. Season (spring, summer, fall, winter)

3.2 Data Wrangling

1. Data cleaning
    a. Handling dirty and missing data
2. Data transformation
    a. Data converting and shaping
3. Feature engineering
    a. Creation of new attributes

3.3 Visualization

1. Visualizations must include:
    a. Distribution of infractions by:
        i. Year
        ii. Month
    b. Distribution of top 20 infractions by total ticket fines
2. Geographic distribution of top 20 infractions
    a. Location
    b. Count by ward
    c. Sum fines by ward

4.0 Delivery Checklist

The following is a list of deliverables and the method that is required for the assignment.

Section	Delivery	Method
3.1.1	Top 20 ticket infractions (frequency)	Value within notebook
3.1.2	Top 20 ticket infractions (revenue)	Value within notebook
3.1.3	Total revenue for all tickets	Value within notebook
3.1.4a	Distance to closest parking lots for top 20 infractions	Table within notebook
3.1.4b	Distance to closest TTC stop for top 20 infractions	Table within notebook
3.1.5a	Impact of day of week in all infractions	Table within notebook
3.1.5b	Impact of month of week in all infractions	Table within notebook
3.1.5c	Impact of season of week in all infractions	Table within notebook
3.3.1.a.i	Distribution of infractions by year	Chart within notebook
3.3.1.a.ii	Distribution of infractions by month	Chart within notebook
3.3.1.b	Distribution of top 20 infractions by fines	Chart within notebook
3.3.2.a	Geographic distribution (location) of top 20 infractions (count)	Map within notebook
3.3.2.b	Geographic distribution by ward for top 20 infractions (count)	Map within notebook
3.3.2.c	Geographic distribution by ward for top 20 infractions (revenue)	Map within notebook
N/A	Notebook (Python or R) containing analysis, visualizations, description of process, code comments, etc.	Code collaboration site

5.0 Conduct

The take-home assignment is to assess technical and analytical capabilities of the applicant. This is an individual assignment and is to be completed solely by the applicant, failure to do so could result in ineligibility for the position. The assignment is not to be shared, posted, or presented, in part or in its entirety, to individuals, organizations, forms, or media (social or traditional). The applicant is able to utilize their existing resources, print, digital, and online media to complete the assignment.