Assignment 1
Exploratory Data Analysis

Goal

In this assignment, you will perform an exploratory analysis on "**Citywide Public Computer Centers**" dataset to better understand the shape & structure of the data, investigate initial questions, and develop preliminary insights & hypotheses. Your final submission will take the form of a report consisting of captioned visualizations that convey key insights gained during your analysis.

- Follow the link for more information about the dataset: https://data.cityofnewyork.us/Social-Services/Citywide-Public-Computer-Centers/sejx-2gn3
- **Due Date: 30/11/2023. Penalty applies for late submission. (5% per day which includes Blockades, Strikes, Govt. Holidays)**

Phase 1:

In the first phase, you should seek to *gain an overview* of the shape & structure of your dataset. What variables does the dataset contain? How are they distributed? Are there any notable data quality issues? Are there any surprising relationships among the variables? Be sure to also perform "sanity checks" for patterns you expect to see!

Phase 2:

In this phase, you should investigate your initial questions, as well as *any new questions* that arise during your exploration. For each question, start by creating a visualization that might provide a useful answer. Then refine the visualization (by adding additional variables, changing sorting or axis scales, filtering or sub-setting data, *etc.*) to develop better perspectives, explore unexpected observations, or sanity check your assumptions. You should repeat this process for each of the given questions. In this context, I have already provided you the questions for exploring the dataset.

You need to answer the following question and provide the necessary visualization to support the explanation:

1. What is the distribution of these computer centers across different boroughs and neighborhoods?
2. Which community districts and council districts have the highest and lowest concentrations of computer centers?
3. How many of the computer centers are wheelchair accessible?
4. How many computer centers offer technology-related courses?
5. Are there any trends or correlations between computer centers offering affordability connectivity programs and their location within certain boroughs or neighborhoods?
6. How does the distribution of wheelchair-accessible centers vary by borough?
7. Create a bar chart to visualize the number of computer centers that offer services in different languages.
8. Create a scatter plot to visualize the geographic distribution of computer centers with and without wheelchair accessibility.
9. Visualize the differences in affordability connectivity programs between neighborhoods in different boroughs.
10. Use correlation matrices or heatmaps to identify potential relationships between the presence of productivity tools and the availability of assistive technologies.

Final Deliverable

Your final submission should take the form of **a PDF report –** similar to a slide show or comic book – **that consists captioned visualizations detailing your discovered insights from the given questions**.

- You need to answer each question with providing the proper visualization of the data to support your explanation.

Each visualization image should be a screenshot exported from a visualization tool or you can import Notebook as PDF file following the similar convention, accompanied with a title and descriptive caption (1-4 sentences long) describing the insight(s) learned from that view. Provide sufficient detail for each caption such that anyone could read through your report and understand what you've learned. You are free, but not required, to annotate your images to draw attention to specific features of the data. You may perform highlighting within the visualization tool itself, or draw annotations on the exported image.

The end of your report should include a brief summary of main lessons learned from the dataset.