Description of the Data and How It Will Be Used to Solve the Problem

The data used in the analysis will be derived from multiple sources.

First, we will need the list of FSAs (Forward Sortation Areas) assigned in Toronto. The list is available on the Wikipedia and must be scraped, processed and formatted as a dataframe to make it possible to merge it with other datasets.
[List of postal codes of Canada: M - Wikipedia](#)

A Canadian postal code is a six-character string that forms part of a postal address in Canada. Canada's postal codes are alphanumeric. They are in the format A1A 1A1, where A is a letter and 1 is a digit, with a space separating the third and fourth characters. A forward sortation area (FSA) is a geographical region in which all postal codes start with the same three characters. The dataset will consist of FSA codes, borough names and the list of neighbourhoods in each borough. This will be used as the basis for further data analysis, as outlined in the next paragraphs.

Second, we will need geospatial data to pair the FSAs with their geographical latitude and longitude. This will allow us to create a map and plot the FSAs as part of the data visualization component. This dataset can be obtained through python libraries where FSA information is passed as parameter. Several Python packages have been developed to make working with geospatial data in Python easier. The list of FSA codes from the first dataset will be used as input to retrieve latitude and longitude of each FSA.

Finally, we will need the veterinary clinic information extracted from Foursquare. Foursquare City Guide is a local search engine and discovery mobile app that helps users discover new places from a community of peers. It provides personalized recommendations of places to go near a user's current location based on the user's previous visits, likes and check-in history. The data will be retrieved via API calls, processed, cleaned and formatted to make it possible to merge it with other datasets. The API calls have a defined format with several pieces of information are concatenated and sent to Foursquare.

[https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={}&radius={}&limit={}&query={}](#)

    client_id, client_secret = tokens assigned to each registered user

    v = version, i.e. date of the search result dataset update
    radius = Limit results to venues within this many meters of the specified location. Defaults to a city-wide area. Only valid for requests that use categoryId or query. The maximum supported radius is currently 100,000 meters.

limit = Number of results to return

query = A search term to be applied against venue names

In return, Foursquare generates a dataset in .json format, which will be extracted, flattened and formatted as the dataframe compatible with the other datasets as mentioned above. The dataset will have venue names, categories (in this case the key word used for data retrieval will be 'veterinary'), venue latitude/longitude and the FSA the venue is located in (the first three characters of the postal code).

The data processing will include scraping, wrangling, analysis, merging/concatenation, dynamically building and execution of API queries, clustering and plotting a map with colour coded markers for easy identification of the areas with different venue counts. Markers will be superimposed on the Toronto area map with labels showing venue counts and location marker pop-ups showing FSA codes.