IV. Data:

The first step in gathering data is to properly identify what data is required in relation to the purpose of the study. The identified data requirements are:

- **LRT Stations:** an updated list of LRT stations that traverse the Metro Manila area. Each LRT station is identified through a unique name, its governing district and the municipality.

  URL:
  https://en.wikipedia.org/wiki/List_of_rail_transit_stations_in_the_Greater_Manila_Area

Station	District/Barangay	City/Municipality	Connections	Opened	Nearby landmarks
Green Line (Line 1, Light Rail Manila Corporation)					
Roosevelt	Bago Bantay	Quezon City		22 October 2010	Walter Mart Muñoz, INC Muñoz Church, Muñoz Market, Quezon City General Hospital
Balintawak	Balintawak			22 March 2010	Balintawak Interchange, North Luzon Expressway, Puregold Balintawak, Ayala Malls Cloverleaf
Yamaha Monumento	Grace Park	Caloocan		12 May 1985	Bonifacio Monument, Caloocan Cathedral, Manila Central University, University of the East Caloocan,
5th Avenue					La Loma Cemetery, Caloocan City Hall
R. Papa	Tondo	City of Manila			
Abad Santos					Chinese Cemetery
Blumentritt	Santa Cruz		Blumentritt		Chinese General Hospital, North Cemetery
Tayuman					Department of Health, SM City San Lazaro, University of Santo Tomas
Bambang					Metropolitan Medical Center
Doroteo Jose			Recto		Dr. Jose Fabella Memorial Hospital, Far Eastern University, Isetann Cinerama Recto, Saint Stephen's Parish
Carriedo					Binondo Church, Escolta Street, Golden Mosque, Plaza Lacson, Plaza Miranda, Quiapo Church, Santa Cruz Church
Central Terminal	Ermita		Lawton (Pasig River Ferry)		Central Post Office, City Hall, Intramuros, Manila Cathedral, Metropolitan Theater, National Museum of Fine Arts, SM City Manila
United Nations					National Museum of Anthropology, National Museum of Natural History, National Bureau of Investigation, National Library, Rizal Park, Supreme Court, World Health Organization
Pedro Gil					Philippine General Hospital, Robinsons Place Manila, University of the Philippines Manila

- **Station Attributes:** an updated list of features surrounding the identified LRT stations. Feature refers to an infrastructure that has prominent effect on the classification of the LRT station. One example is the number of businesses surrounding the station, number of transportation terminals among others.

  Source: https://foursquare.com/

Station	Nearby landmarks
Roosevelt	Walter Mart Muñoz, INC Muñoz Church, Muñoz Market, Quezon City General Hospital
Balintawak	Balintawak Interchange, North Luzon Expressway, Puregold Balintawak, Ayala Malls Cloverleaf
Yamaha Monumento	Bonifacio Monument, Caloocan Cathedral, Manila Central University, University of the East Caloocan,
5th Avenue	La Loma Cemetery, Caloocan City Hall
R. Papa	
Abad Santos	Chinese Cemetery
Blumentritt	Chinese General Hospital, North Cemetery
Tayuman	Department of Health, SM City San Lazaro, University of Santo Tomas
Bambang	Metropolitan Medical Center
Doroteo Jose	Dr. Jose Fabella Memorial Hospital, Far Eastern University, Isetann Cinerama Recto, Saint Stephen's Parish
Carriedo	Binondo Church, Escolta Street, Golden Mosque, Plaza Lacson, Plaza Miranda, Quiapo Church, Santa Cruz Church
Central Terminal	Central Post Office, City Hall, Intramuros, Manila Cathedral, Metropolitan Theater, National Museum of Fine Arts, SM City Manila
United Nations	National Museum of Anthropology, National Museum of Natural History, National Bureau of Investigation, National Library, Rizal Park, Supreme Court, World Health Organization
Pedro Gil	Philippine General Hospital, Robinsons Place Manila, University of the Philippines Manila

The implemented methodology for data collection is:

1. The list of LRT station is critical since this will act as an input to the Geopy python package which will provide the geospatial coordinates of the location. Each station will have a corresponding latitude and longitude values.

   The LRT station will be scraped from the Wikipedia page using the Beautiful Soup library in python. Beautiful Soup is a python package which allows parsing of HTML to deconstruct it into logical components which allows for easier extraction of data.

   Beautiful Soup together with Pandas library will accomplish scraping of unstructured data from the Wikipedia page to a data frame that is suited for further processing and analysis.

2. The geographical location of each LRT station will then be used as input in the Geopy package which will provide the geospatial coordinates of the each LRT station.

   The Geopy package allows for the creation of an instance that allows querying of the location's longitude and latitude values. The extracted coordinates will act as input to the Foursquare API to allow access to the different attributes surrounding the location (LRT Station).
   Through pandas, the coordinates will be pre-processed and transformed into a data frame that is concatenated with the list of LRT stations.

3. The geospatial coordinates will be integrated into the Foursquare API syntax which will provide access to the database containing the available attributes of each LRT station.

   The extraction of available attributes will be limited to 1,000 meters which is a reasonable estimation of acceptable walking distance and clustering will be contingent on the top 10 attributes that will be extracted from the API call.

   The attributes determine the criticality of the LRT station in relation to the attributes it is serving nearby. To put it into perspective, an LRT station whose top attributes revolve around business establishments imply the peak volume of passengers being serviced by the system.

   This will be essential information in determining how to decongest that particular station or what additional improvements are required to increase the level of service.

The pre-processed data will then be visualized to assess its quality and administer corrective measures as deemed required by the model.

The model will the cluster the stations based on their attributes which will define the proper actions required to achieve the objective of the study.