Project Title:

EV Charging Station Siting Analysis

Introduction/Business Problem:

Background:

This code was written as part of the requirements of the IBM/Coursera Data Science Professional Career Track Specialization. Students were asked to choose a problem and utilize skills acquired during this specialization to solve the selected problem.

Business Problem Outline:

Electric Vehicles (EVs) are gaining popularity due to absence of local emissions, and certain technical advantages that EVs have over conventional vehicles (faster acceleration, no need for periodic service requirements etc.). This is not to say that there are no barriers to the widespread acceptance of EVs. One of the primary drawbacks is the lack of access to charging equipment at public locations.

In this project, we look at the city of Raleigh in North Carolina, where I currently reside. The objective is to try to come up with a siting analysis for public EV charging station installations - preferably close to local shops and restaurants so that people can charge their cars while getting groceries or sharing a meal with their loved ones. One of the indirect benefits of EV charging that is often talked about is contribution to the local economy - i.e. people tend to spend money on nearby shops when waiting on charging their EVs.

To be specific, the area of interest in this case is within a 100 km radius of downtown Raleigh in North Carolina. This approximately covers the towns of Durham, Raleigh and Chapel Hill. As of 2020, the population of the larger Raleigh-Durham-Chapel Hill Combined Statistical Area (CSA) is estimated at 2.03 million. Hailed as a technological hub within the state, the mean age of residents range between 26 to 35 years. This area may be considered a suitable case study location for testing out algorithms ranking retail facilities that could benefit from EV charger installations.

Note that the approach that I am taking is pretty simplistic - I am aiming to identify retail locations (primarily grocery shops, restaurants and some service industries) that have relatively few EV chargers in the vicinity. A list of suitable retail locations that could benefit from an EV charging installation in its neighborhood is provided as input to a clustering algorithm. This algorithm determines locations that could improve access to EV charging in our region of interest.

Outputs include a map with clusters of retail locations that could benefit from EV charger installations and a list of suitable locations for EV charging installations.

Rubenka Bandyopadhyay, PhD. Date Submitted: 5/31/2020

Target Audience

This project is primarily aimed at any one interested in Data Science applications in the Energy sector, but I hope that other Data Science enthusiasts will also find ways to repurpose the code. Data Science techniques and/or Machine Learning Algorithms often continue to remain abstract ideas until we apply them to problems. I hope this project provides some insight into translating abstract ideas into specific outputs that could aid decision making in business ventures.

Data Sources:

API Foursquare - This database provides information on locations of popular restaurants. Link: https://foursquare.com/

AFDC Database - This database provides information on preexisting EV charging locations. Link: https://afdc.energy.gov/fuels/electricity_locations.html#/find/nearest?fuel=ELEC

Methodology: Use k-means clustering algorithm to identify suitable locations

Step 0:

Import relevant Python libraries

Step 1:

Read in the data from AFDC and Foursquare API

Next up - identification of trending spots in the Raleigh...
There are a few different ways of identifying trending spots. For purposes of this project we consider two specific Foursquare categories:
(a) 'Food', and (b) 'Shops & Services'.
Refer to API Foursquare List of Venue Categories for full list of categories

Step 2:

Now that we have the geographic coordinates of businesses in the area, let us determine the distance between each venue obtained from the foursquare API and each EVSE outlet location: Procedure to calculate the distance has been obtained from this article in medium: Finding the distance between two lists of geographic coordinates by Dana Lindquist. The distance computed here is a haversine distance. This assumes the earth is a true sphere which makes for a relatively fast computation. The sklearn computation assumes the radius of the sphere is 1, so to get the distance in miles we multiply the output of the sklearn computation by 3959 miles, the average radius of the earth. To get the distance in kilometers this number would be 6371 km.

Step 3:

Now let's count the number of EVSE Stations that are at distance which is less than or equal to 5 km from the prospective venues

Step 4:

Let us now **cluster** those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.

Finally, let's reverse geocode those k-means clustering centers to get the addresses which can be presented to stakeholders.

Discussion

While this project does a good job demonstrating how data science can be used to inform energy infrastructure siting decisions, the analysis could be improved in many different ways. A few are listed here but should not be considered an exhaustive list:

(a) Socio-economic as well as demographic factors need to be considered when prioritizing siting locations. Here is an excellent article summarizing policies and actions being taken to expand access to electric transporation among low income groups and apartment dwellers

(b) Siting analysis may need to take into considerations the practicality of EV charging installations - zoning permits, transmission capacity to support supply of electricity for EV charging etc. are some factors that could affect EV charging installations

(c) Demand for EV charging is another really important factor. If there are not enough EVs driving through the region that need to stop and charge - there may not be any incentive for installation of EV chargers. North Carolina DMV has recently started releasing EV registration data in NC counties . This could indicate how many EV owners currently reside in NC. Market research could also be performed to explore the out-of-state EV traffic passing through NC.

Conclusion

This project provides a simple example of utilization of k-means clustering technique in the clean energy sector. Data cleaning and manipulation, application of algorithms, and subsequent data visualization are the primary steps involved in any problem that needs to be solved using data analytics. I hope that this notebook provides a source of reference for those starting to explore data-driven solutions to their business problems.