

Data Assignment

Xiyu Zhang

2/17/2022

Part 1

Clearly and briefly identify at least one broad question of interest that can be explored through data science methods.

We could use data science to target their marketing to specific consumer demographics (segments).
Also, we could do “crisis” problem-solving using data science methods.
Third, we could investigate the infrastructures.

Part 2

Identify at least three data sets related to the question(s) from part 1. Briefly describe each data set.

You will benefit from data of diverse types. For example, datasets that contain geospatial, temporal, and hierarchical aspects, with varying units of analysis, can give you more to work with. Also, being able to merge the datasets could prove valuable.

I am trying to investigate the energy infrastructures.

1. The first data set I choose is called “Solar”.

It is from the link “<https://atlas.eia.gov/datasets/eia::solar-2/about>”. • How the data are created: It is created by U.S. Energy Information Administration.

- If there is a clear data dictionary: yes, as it shows in “<https://services7.arcgis.com/FGr1D95XCGALKXqM/arcgis/rest/services/ElectricPowerPlants/FeatureServer/9%22/>” • The unit of analysis of the data: In this data set, units of analysis are solar utilities.
- The dimensions of the data: solar power(MW), install(MW), Hydro(MW), and so on.
- Any missingness in the data:
- The limitations of the data: : It is only a feature layer of the nation-wide data set.

2. The second data set I choose is called “Pumped Storage”

It is from the link “<https://atlas.eia.gov/datasets/eia::pumped-storage/about>”. • How the data are created: It is created by U.S. Energy Information Administration.

- If there is a clear data dictionary: yes, as it shows in “<https://atlas.eia.gov/datasets/pumped-storage/explore?location=37.475521%2C-118.495999%2C3.59&showTable=true%22/>” • The unit of analysis of the data: In this data set, units of analysis are pumped storage utilities.
- The dimensions of the data: utility ID, Sector name, Install MW, Total_MW, and so on.
- Any missingness in the data: No.

- The limitations of the data: : It is only a feature layer of the nation-wide data set.
3. The third data set I choose is called “Petroleum”
- It is from the link “<https://atlas.eia.gov/datasets/eia::petroleum-2/about>”. • How the data are created: It is created by U.S. Energy Information Administration.
- If there is a clear data dictionary: yes, as it shows in “<https://atlas.eia.gov/datasets/petroleum-2/explore?location=30.778341%2C-118.495999%2C3.59&showTable=true%22/>” • The unit of analysis of the data: In this data set, units of analysis are petroleum utilities.
 - The dimensions of the data: utility ID, Sector name, Install MW, primary source, Total_MW, and so on.
 - Any missingness in the data: No.
 - The limitations of the data: : It is only a feature layer of the nation-wide data set.