

Python Project

Download the dataset Project_dataset.csv from canvas. The dataset has following columns.

GEOID10	Census Tract ID (10 digit)
County Code	County that the tract belong to
Latitude	Latitude of the tract
Logitude	Longitude of the tract
Total Pop	Total Population of the tract
Mean Comm Time	Mean commute time for workers over 16 years old in the tract
Frac_College	Number of people aged 25 or older who have a bachelor's degree, master's degree, professional school degree, or doctorate degree, divided by the total number of people aged 25 or older in a tract.
Med HHincome	Median household income
Pop_Density	Number of residents per square mile, calculated by dividing the total tract population with tract land area given in square miles.
Frac_Living	Fractions of population living in poverty in the tract
Jobs Total	Total Number of jobs in the tract.
Frac Nonwhite	Fraction of non-white population
Underground	Number of Underground Tanks in the tract.
Fire Station	Number of fire station in the tract
Hazardous	Number of hazardous sites in the tract.
Gas Station	Number of Gas stations in the tract
Economic Mob	Economic Mobility Index of the tract

Create a jupyter notebook to perform the below tasks.

Data Exploration and Cleaning

1. How many columns are in the dataset? How many tracts' records are there?
2. Is there any missing data? Which Column has the missing values?
3. If there is missing data in a column, replace the missing values with the median of that column

Querying a Dataset

1. What is the maximum and minimum Mean Commute Time?
2. What is the maximum value for fraction of residents living in poverty in a census tract? Which census tract has the maximum fraction of residents living in poverty?
3. How many census tracts have more than 10 hazardous waste sites?
4. How many census tracts do not have any fire stations?
5. How many census tracts have more than the average number of underground tanks?
6. What is the mean population density of the census tracts with 10 or more Gas Stations vs. without any Gas Stations?
7. What is the average of median house hold income for the tracts where 80% or more people are non-white vs. 20% or less people are non-white
8. Which census tract has the highest and lowest economic mobility? What are the fractions of college educated population there?
9. What are the averages of median household income for each county? Sort them in ascending order
10. Group the data by number of Hazardous sites. What are the average population densities?

Finding Correlations

1. What is the correlation between median household income and fraction of residents with a college degree?
2. Plot the correlation between median household income and fraction of residents with a college degree.
3. Plot the correlation between median household income and fraction of residents living with poverty.
4. What is the correlation between Economic Mobility and fraction of residents living in poverty?
5. Plot the correlation between Economic Mobility and fraction of residents with a college degree.
6. Plot the correlation between Economic Mobility and fraction of residents living in poverty.

Data Visualization

1. Show the histogram for each columns
2. Use the color map to visualize population density and median household income for NC geo tracts.
4. Use the color map to visualize number of gas stations and median household income for NC geo tracts.
5. Use the color map to visualize number of total jobs and the value of Economic mibility for NC geo tracts.
6. Use the color map to visualize the fraction of people living with poverty and the value of Economic mobility for NC geo tracts.