

TEAM 8 TECHNOLOGIES

Regional Energy Challenges: A Comprehensive Analysis of City and County Energy Profiles

- Arpan Pradhan
- Jahanvi Reddy
- Mingchao Yao



CONTENTS

- 1. Executive summary**
- 2. Business use case**
- 3. Data collection**
- 4. Data profile**
- 5. Data implementation tools**

Executive Summary

Amid global concerns over pollution and the ongoing climate crisis, our research focuses on the United States. We conduct a thorough analysis of city and county energy profiles, using data on states, demographics, energy use, and greenhouse gas emissions. Our main goal is to understand energy consumption patterns, efficiency factors, and their environmental impact at the regional level.

We will use data analysis and visual aids to evaluate the environmental effects of energy use and transportation nationwide. Our primary focus is identifying regions with high carbon emissions and proposing solutions.

Business Use Case

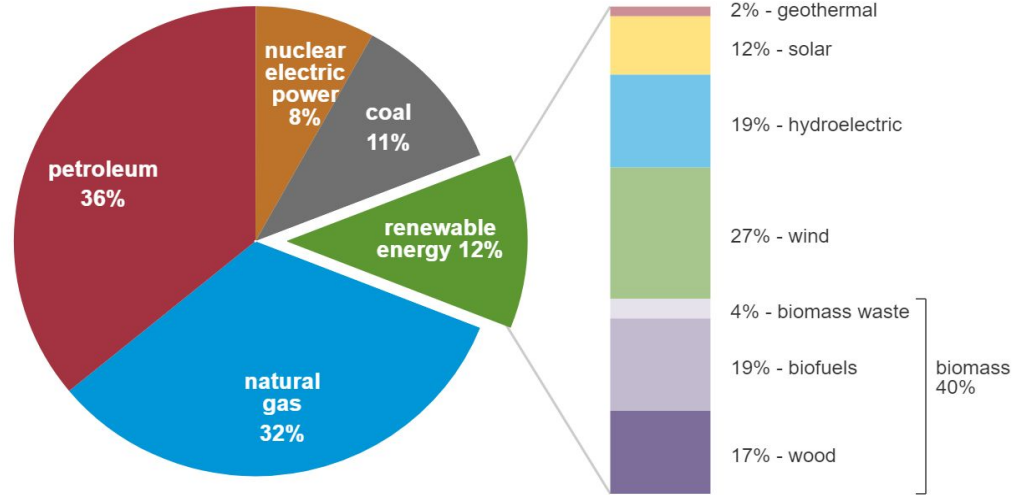
1. The transportation miles and fuel consumption are important indicators of the greenhouse gas emissions, through listing key related variables and analyzing these variables, the further recommendation based on data is practically useful to reduce the gas emissions for the regional areas.
2. Electricity generating and natural gas consumption are the main process of producing dioxide carbon and other greenhouse gases, we would focus on this perspective from different kinds of consumers to analyze the data.
3. By the analysis of different kinds of powers' consumption, we would classify the regional or city profiles so that we can give the related recommendations on improving the local environments reducing the greenhouse gas emissions.

Early adoption of sustainable energy practices not only confers a competitive advantage by showcasing a commitment to environmental responsibility but also attracts eco-conscious consumers and partners. This demonstration of commitment enhances the company's reputation and garners favorable views from customers, investors, and employees.

U.S. primary energy consumption by energy source, 2021

total = 97.33 quadrillion
British thermal units (Btu)

total = 12.16 quadrillion Btu



Data source: U.S. Energy Information Administration, *Monthly Energy Review*, Table 1.3 and 10.1, April 2022, preliminary data
Note: Sum of components may not equal 100% because of independent rounding.

Data Collection

Data Source: City and County Energy Profiles

The dataset used for our analysis is sourced from Data.gov, a reliable and comprehensive platform that provides access to a wide range of public datasets. The "City and County Energy Profiles" dataset offers valuable information on energy consumption, demographics, greenhouse gas emissions, and more, focusing on city and county-level data across the United States.

Data Profile

Our analysis encompasses two primary datasets.

- The first dataset comprises comprehensive energy profiles for major cities across the United States.

Data Size: This dataset contains data on approximately 23,437 cities.

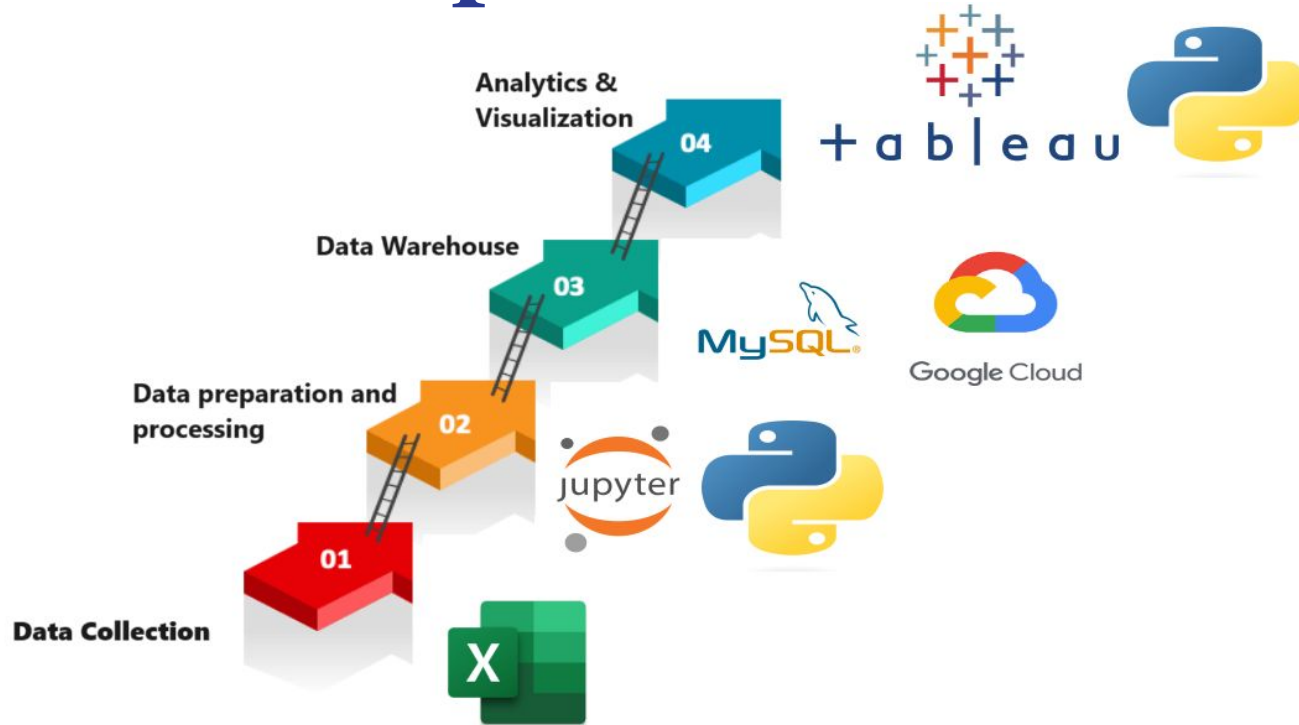
- The second dataset focuses on county-level data spanning the entire United States.

Data Size: It includes information on around 3,147 counties, providing a broad geographical context for our analysis.

The database uses various data types such as INT, VARCHAR, DATE, and more to store different kinds of information related to the energy profiles. The database is in good shape overall, but we need to address missing data before we can use it reliably for analysis.

The datasets are presented in a structured .csv format, and they collectively possess a size of 36,702 kilobytes.

Data Implementation Tools



Additionally we will also be using MongoDB, QGIS and Alteryx

Data preparation steps

	STEP 1	STEP 2	STEP 3	STEP 4	STEP 5	STEP 6	STEP 7
TYPE	INITIAL DATA ASSESSMENT	DATA CLEANING	DATA TRANSFORMATION	DATA INTEGRATION	DATA REDUCTION	HANDLING OUTLIERS	FINAL DATA REVIEW
OBJECTIVE	<ul style="list-style-type: none">• DATA COMPLETENESS;• CONSISTENCY;• FORMAT	<ul style="list-style-type: none">• MISSING DATA HANDLING;• REMOVING DEPLICATES	<ul style="list-style-type: none">• NORMALIZATION• CATEGORICAL DATA ENCODING	<ul style="list-style-type: none">• COMBINE CITY AND COUNTRY DATASETS	<ul style="list-style-type: none">• REDUCE DATASET FOR ANALYSIS EFFECNECY	<ul style="list-style-type: none">• IDENTIFY AND MANAGE OUTLIERS	<ul style="list-style-type: none">• ENSURE DATA READY
PROCEDURE	Load data file into a analysis platform to perform the assessment	Use functions to identify missing data or remove them, and find and remove duplicate records	Min-Max scaling or Z-score normalization can be used. Also use one-hot encoding or label encoding	Merge datasets on common identifiers	Apply feature selection techniques to remove irrelevant or redundant data columns	Use IQR, Z-score to detect outliers	Perform a final review of the dataset

PLATFORM CONSIDERATIONS FOR DATA PROCESSING

STEP	TOOLS/TECHNIQUES/PLATFORMS
DATA COLLECTION	<ul style="list-style-type: none">✓ Public data source: credible datasets from Data.gov✓ API and PYTHON Beautiful Soup: complement the datasets✓ Other ways: To do data consistency checks
DATA TRANSFORMATION	<ul style="list-style-type: none">✓ Python with Pandas and SciPy Libraries Considerations:<ul style="list-style-type: none">(1) handling large datasets;(2) functionality: data manipulation for data cleaning and transformation;(3) for statistical analysis and machine learning

EER Diagram

