📊 Dataset Overview

Project Title:
"Data-Driven Environmental Analysis of Urban Heat Island Effects Across Cities"

🔍 Dataset Column Details

Column Name	Data Type	Description
City Name	object	Name of the city or urban area
Latitude	float64	Geographic latitude coordinate
Longitude	float64	Geographic longitude coordinate
Elevation (m)	float64	Elevation above sea level in meters
Temperature (°C)	float64	Average surface temperature in degrees Celsius
Land Cover	object	Primary land cover type (e.g., Water, Green Space, Built-Up)
Population Density (people/km²)	int64	Number of people per square kilometer
Energy Consumption (kWh)	float64	Annual per capita energy consumption in kilowatt-hours
Air Quality Index (AQI)	int64	AQI score, a measure of air pollution (lower is better)
Urban Greenness Ratio (%)	float64	Percentage of urban area covered by vegetation
Health Impact (Mortality Rate/100k)	float64	Mortality rate per 100,000 population due to heat or pollution impacts
Wind Speed (km/h)	float64	Average annual wind speed in kilometers per hour
Humidity (%)	float64	Average annual relative humidity percentage
Annual Rainfall (mm)	float64	Total annual rainfall in millimeters
GDP per Capita (USD)	float64	GDP per person in U.S. dollars, proxy for economic strength

---

🎯 Project Statement

Objective

This project aims to analyze environmental and socioeconomic data from 500 cities to study **urban heat island (UHI) effects**, their relationship with urban infrastructure, land cover, climate, and public

health. The goal is to uncover meaningful patterns that help policymakers and researchers identify risk zones and prioritize urban planning strategies.

🧭 Key Questions & Goals

🌐 Geographic & Climate Analysis

- Which cities are hottest or coolest on average?
- How does elevation affect urban temperature?
- What land cover types are associated with lower temperatures?

🏙️ Urban Structure & Impact

- How does population density correlate with temperature or AQI?
- Do greener cities have better air quality or lower mortality rates?
- Is energy consumption a driver of poor air quality?

📊 Statistical & Aggregation Tasks

- Group cities by land cover or climate zone and compare average values.
- Rank cities by mortality rate, AQI, or greenness.
- Find cities with extreme values (e.g., max temperature or lowest greenness).

🔗 Correlations & Trends

- What's the relationship between temperature and greenness?
- Do higher GDP cities have better environmental indicators?
- Explore humidity, rainfall, and wind speed as moderating factors for UHI.

🩺 Public Health Insights

- Identify regions with both high temperatures and high mortality.
- Which cities are most at risk based on combined environmental stressors?

🛠️ Tools & Technologies

- **Platform**: Jupyter, Python, or SQL-based tools
- **Libraries**: Pandas, Seaborn/Matplotlib, Scikit-learn (for clustering/correlation)
- **Optional**: PostgreSQL / SQLite for structured analysis

📈 Expected Outcomes

By the end of the project, you should be able to:

- Create a **profile for each city** based on heat risk and environmental resilience
- Visualize **relationships between urban structure and climate impact**
- Generate **data-driven insights** for environmental policy recommendations
- Develop **query templates** or Python notebooks for geospatial/environmental analysis.

Questions

📋 **Urban Heat Island Dataset – Analytical Questions**

🔍 **Basic Exploration**

1. **List all unique cities in the dataset.**
2. **Show the first 10 rows of city-level environmental data.**
3. **What are the different land cover types?**
4. **Which cities have the highest population density?**
5. **Display all cities with their average temperature.**

🎯 **Filtering with Conditions**

6. **Find cities with temperature above 35°C.**
7. **List cities where Urban Greenness Ratio is below 20%.**
8. **Show cities with AQI worse than 150 (poor air quality).**
9. **Which cities have elevation below 100 meters?**
10. **Filter cities with mortality rates above 30 per 100k.**

📊 **Sorting and Ranking**

11. **Rank cities by highest temperature.**
12. **Sort cities by best air quality (lowest AQI).**
13. **Show top 10 cities by GDP per capita.**
14. **Order cities by Urban Greenness Ratio (descending).**
15. **List cities with lowest energy consumption.**

📈 Aggregations and Counts

16. Count total number of cities in the dataset.
17. How many cities have “Green Space” as land cover?
18. What is the average temperature by land cover type?
19. Compute the average mortality rate across all cities.
20. Count cities in each AQI range (e.g., Good, Moderate, Poor).

🔗 Join-Based or Hypothetical Queries

21. (Hypothetical) Join city data with climate zones and compare temperatures.
22. (Hypothetical) Combine with population health data to correlate comorbidities.
23. (Hypothetical) Link to government spending on green infrastructure per city.
24. (Hypothetical) Join cities to their country or region and compare averages.
25. (Hypothetical) Match cities to pollution source types (industrial, traffic).

🧠 Subquery Logic

26. List cities hotter than the average temperature.
27. Find cities with AQI worse than the dataset’s median AQI.
28. Show cities with mortality rate in the top 10%.
29. Find cities with above-average greenness but below-average temperature.
30. Identify cities with GDP per capita above the 75th percentile.

🔤 String/Text Analysis

31. Find all cities with names starting with “San”.
32. Count how many cities have “Green” in the Land Cover field.
33. Show land cover types that include the word “Water”.
34. Replace “Green Space” with “Vegetation Zone” in output (display only).
35. List all cities whose names contain a number (if any).

🗓️ Temporal/Environmental Trend Simulation

36. Simulate seasons: find cities with high humidity and high rainfall.
37. Identify cities likely to suffer drought (low humidity, low rainfall).
38. List windy cities with low temperatures.
39. Rank cities by potential for solar energy (high temperature, low rainfall).

40. List coastal cities (e.g., based on Land Cover = “Water”).

🧮 Advanced Aggregations

41. Calculate average AQI per land cover type.
42. Find standard deviation of temperature across all cities.
43. Determine average GDP per region (if location added).
44. Count cities with mortality rate > 25 and greenness < 20%.
45. Find the correlation between GDP per capita and AQI.

🌎 UHI-Specific Insights

46. Which cities show all three: high temperature, high AQI, and low greenness?
47. What percentage of cities are classified as having “Green Space”?
48. Which cities have a balance of low temperature and high GDP?
49. What are the top 10 cities most vulnerable to urban heat islands?
50. Create an environmental score: score = (Temperature + AQI - Greenness) and rank cities.