STAT 405 GROUP 8 FINAL PROJECT

Weather-Based Clustering of Indian Cities

Flora He, Jiawei Zhang, Lexiang Liu, Yuchen Wang, Huiyu Li

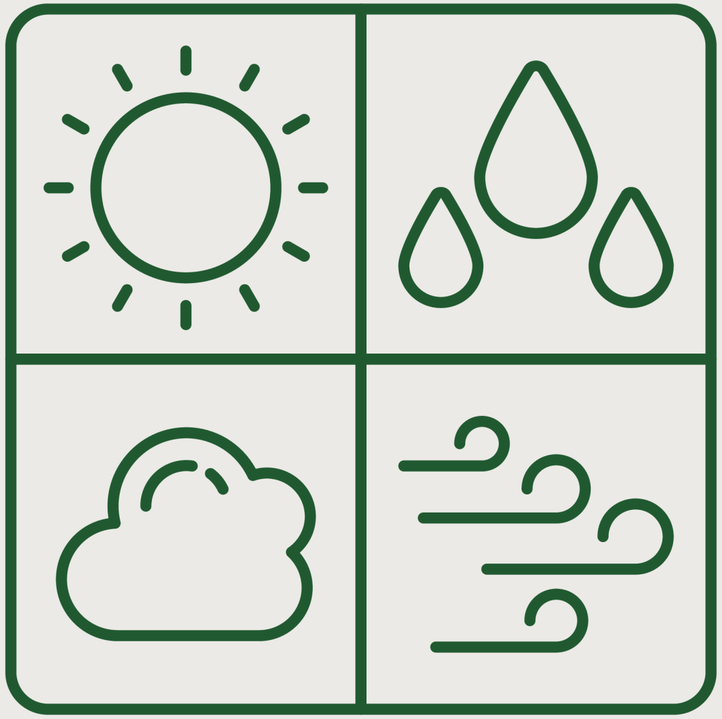
01
PROJECT BACKGROUND

DATASET

- Weather information (2010-2024) for 5585 cities across India.
- Key variables: temperature, humidity, precipitation, wind, air pressure, etc.

MOTIVATION

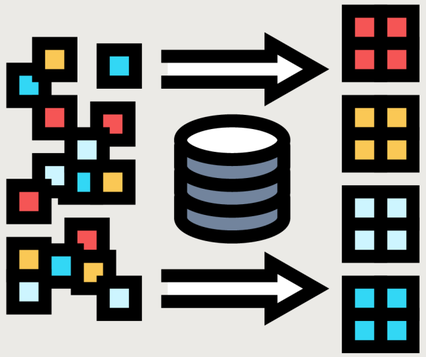
- Explore how Indian cities group together based on their weather patterns, providing insights into the natural distribution and similarities across different regions.



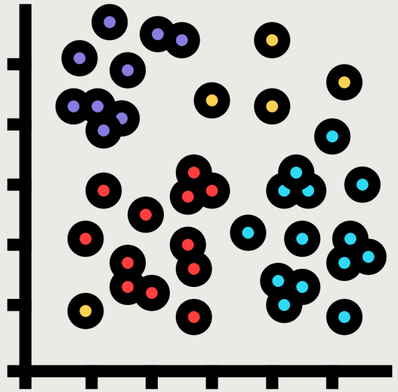
02
PARALLEL COMPUTING

GENERAL WORKFLOW

DATA PREPARING AND PROCESSING



CLUSTERING



GRAPHING

Download and extract data — One step, do once

↓

data/cityname.csv files — One step, do once

Process: Abbigeri.csv Abdullahnagar.csv ... 4685 files — 4685 parallel jobs, repeatable

merged.csv — 1 job, repeatable

↓

kmeans_results.csv — 1 job, repeatable

↓

Visualization — 1 job, repeatable

DATA PREPARING

Weather_Data_Scraping_and_Analysis/

weather-data-scraping-and-analysis.ipynb

weather.csv

w_d_1/

1762 cityname.csv files

w_d_2/

2087 cityname.csv files

w_d_3/

w_d_3/

836 cityname.csv files

→

Weather_Data_Scraping_and_Analysis/

weather-data-scraping-and-analysis.ipynb

weather.csv

data/

All 4685 cityname.csv files
(replaced spaces with underscores in file names)

FEATURE SELECTION

- 19 features in original dataset
- 5 features after selection
- 'temperature_2m', 'relative_humidity_2m', 'precipitation', 'pressure_msl', 'cloud_cover'

DATA PROCESSING

For each .csv file, do:

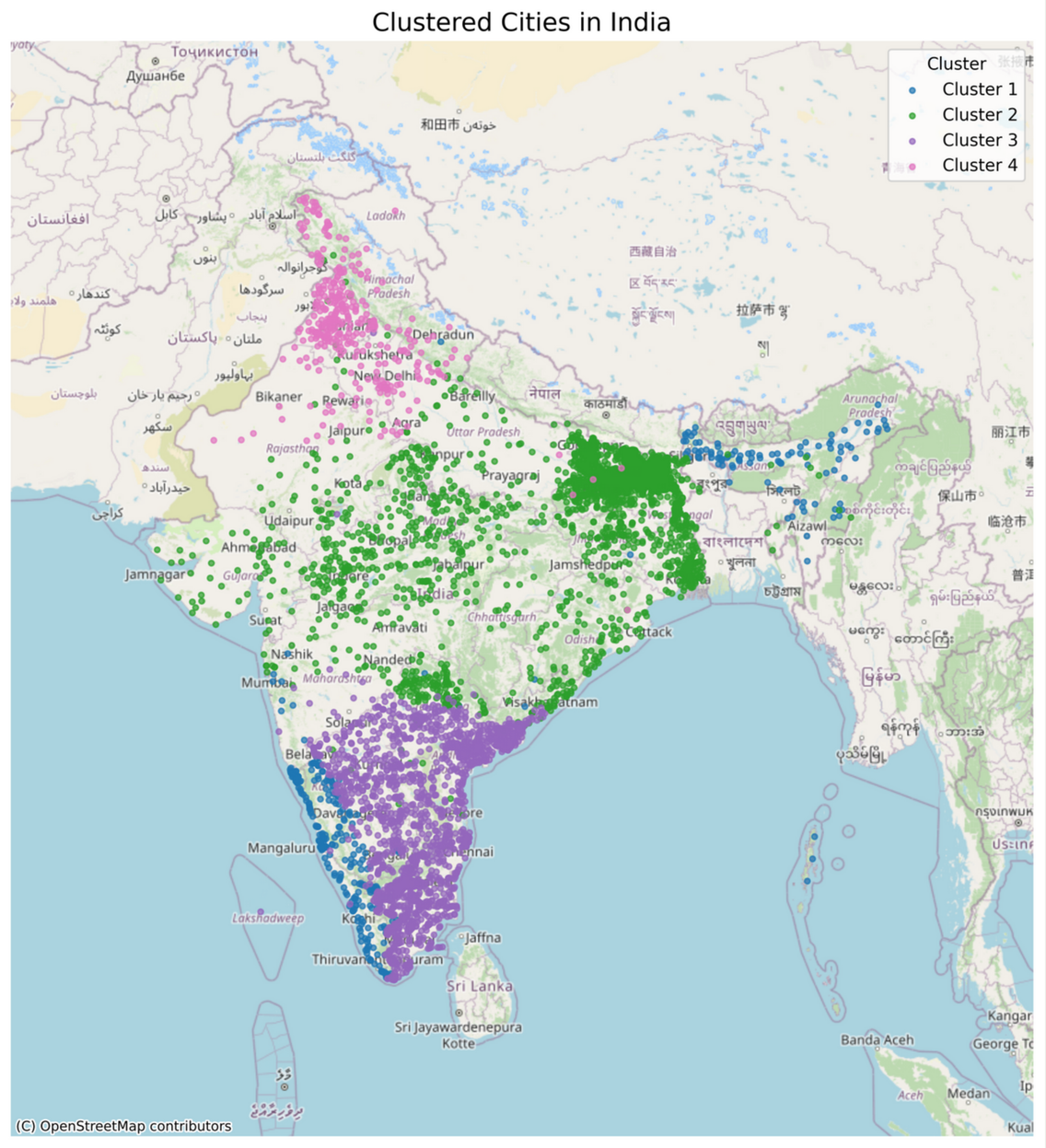
1. Select the recent 5-year rows
2. Extract mean values of the 5 variables over each month
3. Output a new cityname.csv file to another directory (1 row x 60 columns)
4. Merge all new cityname.csv files to merged.csv (added one column for city names)

4685 + 1 jobs in total

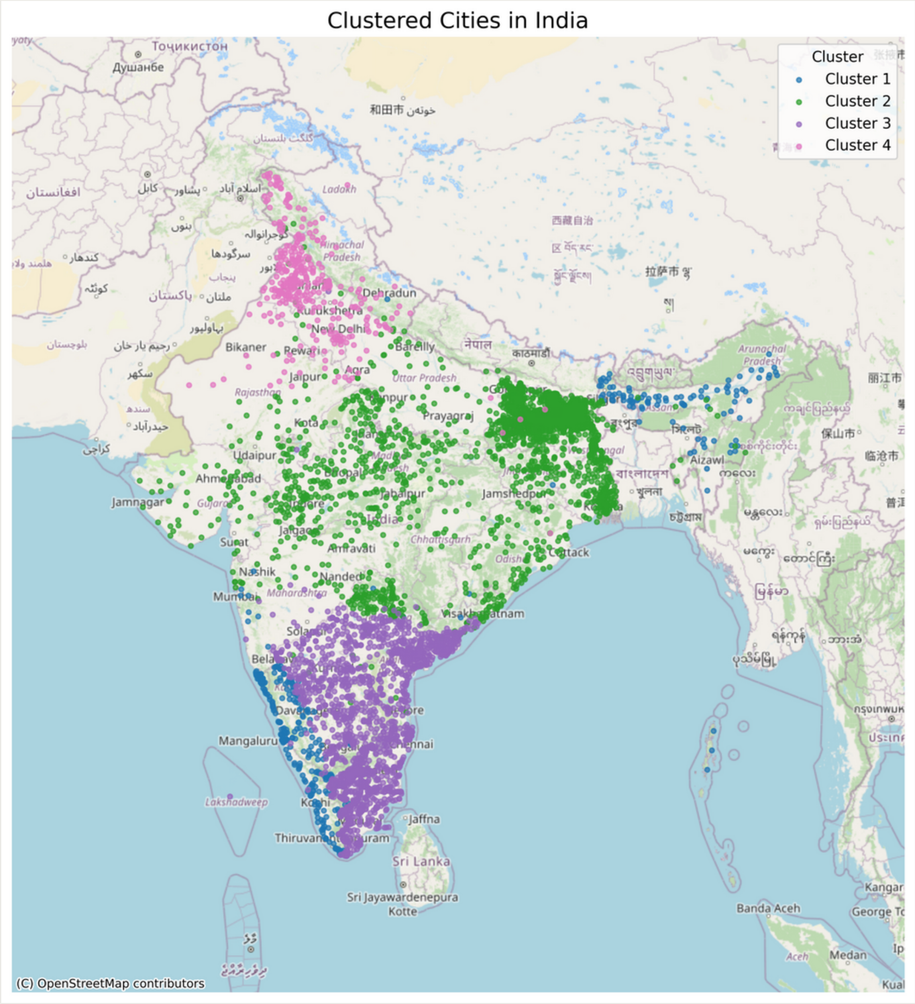
K-means Clustering

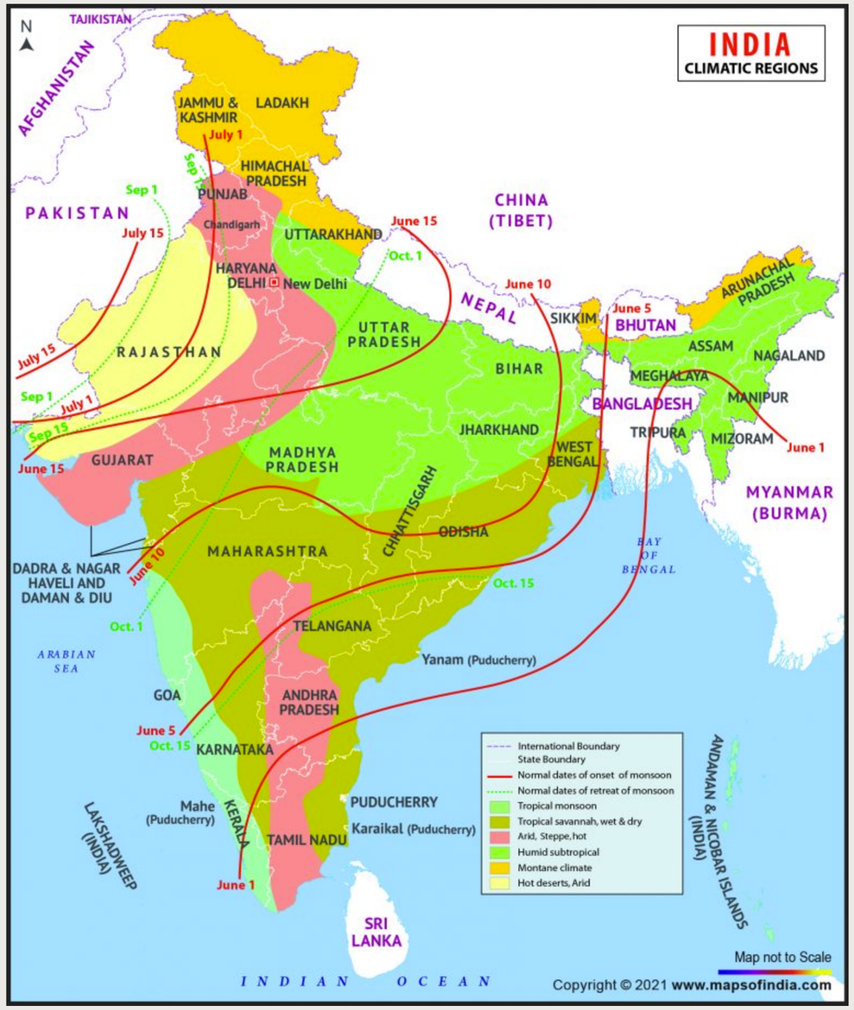
1. Read data from merged.csv
2. Scale data for better clustering
3. Apply K-means clustering
4. Merge data with geographic coordinates
5. Export result as kmeans_results.csv

Clustering of Indian cities based on weather patterns



Comparison


Clustered Cities in India
Cluster
Cluster 1
Cluster 2
Cluster 3
Cluster 4
(C) OpenStreetMap contributors


INDIA
CLIMATIC REGIONS
N
TAJIKISTAN
AFGHANISTAN
PAKISTAN
CHINA (TIBET)
NEPAL
BHUTAN
BANGLADESH
MYANMAR (BURMA)
SRI LANKA
JAMMU & KASHMIR
LADAKH
HIMACHAL PRADESH
PUNJAB
Chandigarh
UTTARAKHAND
HARYANA
DELHI
New Delhi
RAJASTHAN
UTTAR PRADESH
BIHAR
SIKKIM
ARUNACHAL PRADESH
ASSAM
NAGALAND
MEGHALAYA
MANIPUR
TRIPURA
MIZORAM
JHARKHAND
WEST BENGAL
GUJARAT
MADHYA PRADESH
CHHATTISGARH
ODISHA
MAHARASHTRA
DADRA & NAGAR HAVELI AND DAMAN & DIU
TELANGANA
Yanam (Puducherry)
GOA
ANDHRA PRADESH
KARNATAKA
Mahe (Puducherry)
KERALA
PUDUCHERRY
Karaikal (Puducherry)
TAMIL NADU
LAKSHADWEEP (INDIA)
ANDAMAN & NICOBAR ISLANDS (INDIA)
ARABIAN SEA
BAY OF BENGAL
INDIAN OCEAN
July 1
Sep 1
July 15
June 15
Oct. 1
June 10
June 5
Oct. 15
June 1
Sep 15
International Boundary
State Boundary
Normal dates of onset of monsoon
Normal dates of retreat of monsoon
Tropical monsoon
Tropical savannah, wet & dry
Arid, Steppe, hot
Humid subtropical
Montane climate
Hot deserts, Arid
Map not to Scale
Copyright © 2021 www.mapsofindia.com

CONCLUSION

- From the comparison images, it can be observed that the classification map generated based on weather data shows a similar trend to the climatic map. This may be due to the close relationship between geological features and weather variation patterns.

- After this, in order to further improve the images, we hope to incorporate more weather data into the analysis and perform more detailed classification.


PAKISTAN
Baluchistan
Multan
Delhi
Jaipur
Agra
Jodhpur
Hyderabad
Karachi
Gulf of Oman
Muscat (Masqat)
U.A.E.
Ahmadabad
Surat
Nashik
Mumbai
Pune
Indore
Gwalior
Kanpur
Lucknow
Allahabad
Varanasi
Patna
Jabalpur
Nagpur
Raipur
Ranchi
Jamshedpur
Kolkata
Dhaka
Khulna
Chittagong
Guwahati
BHUTAN
BANGLA-DESH
MYANMAR (BURMA)
Rangoon (Yangôn)
Bassein
INDIA
Arabian Sea
Bay of Bengal
Bhubaneshwar
Vishakhapatnam
Hyderabad
Vijayawada
Kurnool
Bangalore
Mangalore
Mysore
Chennai
Calicut
Tiruchchirappalli
Cochin
Madurai
Trivandrum
Jaffna
Trincomalee
Laccadive Is (India)
Andaman Islands (India)
Socotra (Suqotrá)
Muzhur al Halaniyat
I. Masirah
Myeik
Kyunzu

Thank You