Semester Project – Python Data Analysis

Weather Trend system
and Statistical engine

Tools & Libraries: Python, Pandas, NumPy, Matplotlib
Course: Programming for Artificial Intelligence

PRESENTED BY : MUHAMMAD HAMAD KHAN
ZUHAD AHMAD
HASNAIN SHAH

PROBLEM STATEMENT

With increasing climate change concerns, raw weather datasets are:

- Extremely large and complex
- Difficult to interpret without statistical processing
- Not directly useful for decision-making

Problem Statement:
How can large-scale global temperature data be transformed into clear trends, statistics, and visual insights using Python?

DATASET DESCRIPTION

- Dataset Name: GlobalLandTemperaturesByCity
- Source: Kaggle (Berkeley Earth)
- Size: ~8.5 million rows, 7 columns
-

Key Attributes:

- dt — Date
- AverageTemperature
- AverageTemperatureUncertainty
- Country, City, Latitude, Longitude

Known Limitation: Missing values (NaN), especially in early records

DATA CLEANING & PREPARATION

Performed using Pandas:

- Removed missing values using dropna()
- Removed duplicate records
- Converted date column to datetime format
- Extracted new features: Year, Month, Year-Month
-

Purpose: Enable accurate time-series and statistical analysis

EXPLORATORY DATA ANALYSIS

Statistical analysis performed at country level:

- Mean (Average Temperature)
- Standard Deviation (Temperature variability)
- Variance

Insights:

- Identified hottest and coldest countries
- Measured climate stability vs volatility

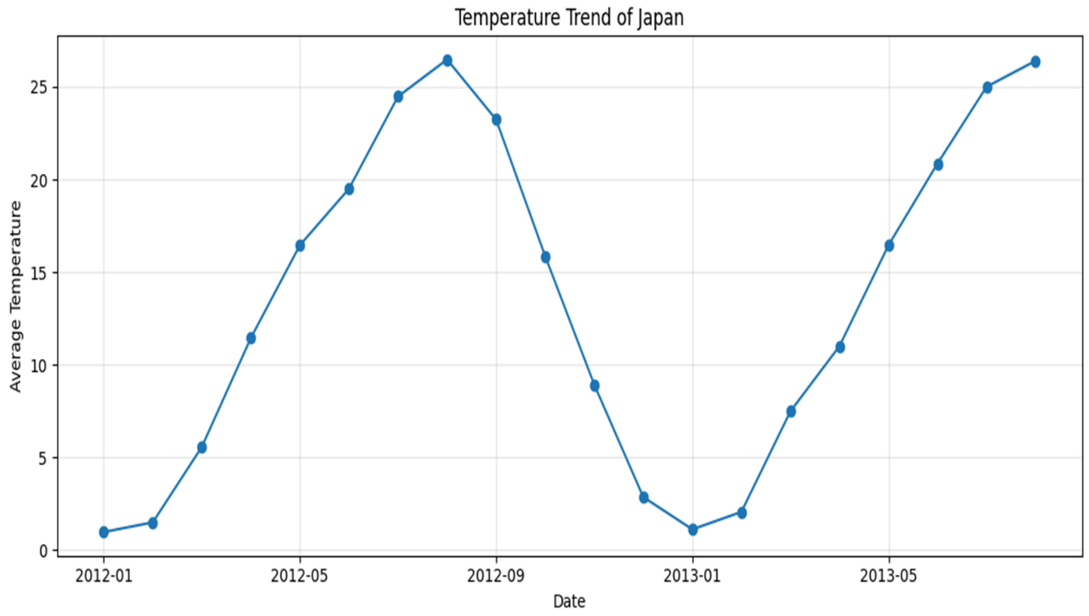
Semester Project – Python Data Analysis

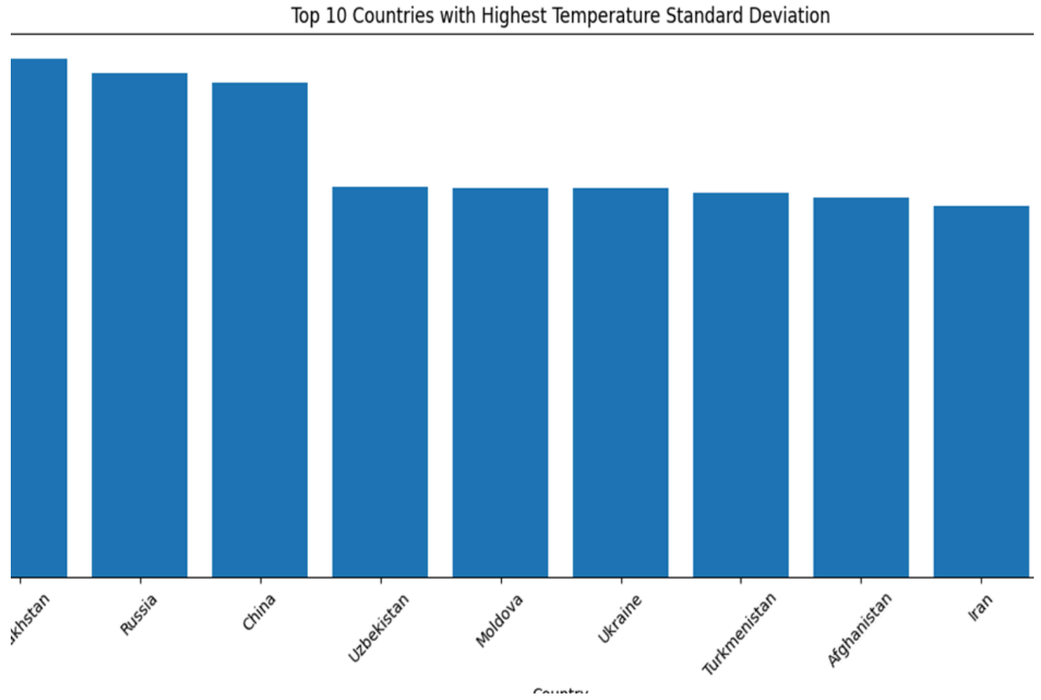
VISUALIZATIONS & INSIGHTS

Visual tools used:

- Bar charts (Top 5 hottest & coldest countries)
- Standard deviation bar chart (climate volatility)
- Global scatter plot (Latitude vs Longitude, colored by temperature)

Outcome: Geographic hotspots and global temperature patterns clearly visible





SYSTEM LOGIC

Object-Oriented Design implemented:

- load_data_set — Data loading, cleaning, preprocessing
- Statistical_Engine — Mean, SD, variance calculations
- Weather_Visualizer — All graph generation
- Weather_Report — Textual and predictive summaries

Menu-driven flow using loops and conditional logic

Semester Project – Python Data Analysis

FINAL RESULTS & INTERPRETATIONS

- Long-term global warming trend observed
- Equatorial regions show higher average temperatures
- High standard deviation indicates climate volatility

Interpretation: Statistical results support real-world climate change patterns

CHALLENGES & LIMITATIONS

- Large dataset required heavy cleaning
- Dropping NaN values reduced early historical data
- Prediction model uses simple linear regression

Limitation: Does not capture seasonality or non-linear climate effects

Semester Project – Python
Data Analysis

Conclusion & Future Work

Conclusion:

- Successfully built a modular Python-based statistical system
- Converted raw data into meaningful climate insights
-

Future Enhancements:

- Advanced time-series models (ARIMA)
- Interactive dashboards
- GUI-based system

GitHub & Credits

Credits:

- Kaggle & Berkeley Earth Dataset
- Pandas & Matplotlib Development Teams

Thank You

GITHUB REPOSITORY: HTTPS://GITHUB.COM/MHK61366-SVG/WEATHER-TREND-SYSTEM-AND-STATISTICAL-ENGINE.GIT