**Project Semester January–April 2025**

**DATA SCIENCE MINOR PROJECT REPORT**

**ON**

**Exploratory Data Analysis and Visualization on Climate Change**

**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**

**COURSE CODE: INT375**

1. **TECH COMPUTER SCIENCE AND ENGINEERING**

****

**LOVELY PROFESSIONAL UNIVERSITY**

**PHAGWARA, PUNJAB**

**PROJECT SUBMITTED BY:**

**Satendra Yadav (12304913)**

**Section: K23SG**

**Roll No.: 66**

**PROJECT SUBMITTED TO:**

**Dr. Manpreet Singh(32354)**

**DECLARATION**

I, **Satendra Yadav**, student of B.Tech – Computer Science and Engineering (Section K23RT) at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report titled:

**“Exploratory data analysis and visualisation on climate change”**

is based on my own intensive work and is genuine. The content of this report has not been submitted to any other university or institution for the award of any degree or diploma.

**Date:** 12-04-2025  
**Registration No.:** 12304913  
**Name:** Satendra Yadav

**CERTIFICATE**

This is to certify that **Ms. Satendra Yadav,** bearing Registration No. **12304913**, has successfully completed the **INT375** – Python Programming project titled:

**“Exploratory data analysis and visualisation on climate change”**

under my guidance and supervision. To the best of my knowledge, the present work is the result of her original development, effort, and study. This project has been carried out as a part of the curriculum prescribed by Lovely Professional University, Phagwara for the Project Semester **January–April 2025.**

**SignatureandNameoftheSupervisor  
Dr.Manpreet singh**

**ACKNOWLEDGEMENT**

I sincerely thank Dr.Manpreet Singh , for his guidance and support throughout this project. I also thank the faculty of the CSE Department at Lovely Professional University for providing the necessary resources and assistance.

**Satendra  
Reg. No.: 12304913**

| **S. No.** | **Topic** | **Page No.** |
| --- | --- | --- |
| **1** | **Introduction** | **4** |
| **2** | **Source Of Dataset** | **5** |
| **3** | **EDA Process** | **6** |
| **3.1** | **Data Loading and Inspection** | **6** |
| **3.2** | **Data Cleaning and Month Ordering** | **6-7** |
| **3.3** | **Basic Statistical Exploration** | **7-8** |
| **3.4** | **Grouping and Aggregation by Station** | **8** |
| **3.5** | **Visual Exploration Strategy** | **9** |
| **4** | **Data Visualization & Analysis** | **10** |
| **4.1** | **Temperature Trends by Station (Line Plot)** | **10-11** |
| **4.2** | **Rainfall Distribution Analysis (Bar Plot)** | **12-13** |
| **4.3** | **Heatmap of Max Temperatures** | **14-15** |
| **4.4** | **Boxplot of Max Temperatures by Station** | **16** |
| **4.5** | **Scatter Plot: Max vs Min Temperature** | **17** |
| **4.6** | **Rainfall Histogram (Distribution)** | **18** |
| **4.7** | **Violin Plot: Min Temperature by Month** | **19-20** |
| **4.8** | **Pairplot: Temperature & Rainfall** | **21-22** |
| **4.9** | **Rainfall Trends Comparison (Top 3 Stations)** | **23-24** |
| **4.10** | **Grouped Bar Chart: Avg Max & Min Temp (Top 5 Stations)** | **25-26** |
| **5** | **Conclusion** | **27** |
| **6** | **Future Scope** | **28-29** |
| **7** | **References** | **30** |

1. **INTRODUCTION:**

Climate change has emerged as one of the most pressing global challenges, with far-reaching impacts on ecosystems, economies, and public health. Understanding historical climate patterns is essential for predicting future trends and developing effective mitigation strategies. This project focuses on analyzing a cleaned dataset containing historical climate data up to the year 2000. The dataset includes key indicators such as temperature variations, CO₂ levels, and other environmental metrics across different regions and time periods. Using data science techniques such as exploratory data analysis (EDA), statistical testing, and visualization, the objective is to uncover trends, assess the extent of climate change over time, and identify potential correlations between various environmental factors. The goal of the project is not only to conduct thorough technical analysis but also to present the findings in a visually interpretable manner, enhancing public understanding and supporting data-driven environmental policy decisions.

1. **SOURCE OF DATASET:**

[**https://www.data.gov.in/**](%20%20%20%20%20%20https:/www.data.gov.in/)

**3. EDA PROCESS**

Exploratory Data Analysis (EDA) is a fundamental step in any data science project. In this project, EDA was performed to understand the structure, patterns, trends, and anomalies within the Climate Change dataset.

The key steps in the EDA process were as follows:

* 1. **Data Loading and Inspection**

The first step in any data analysis project is loading and inspecting the dataset. Using Python's **pandas** library, we load the CSV file and examine its structure.

A screen shot of a computer

AI-generated content may be incorrect.

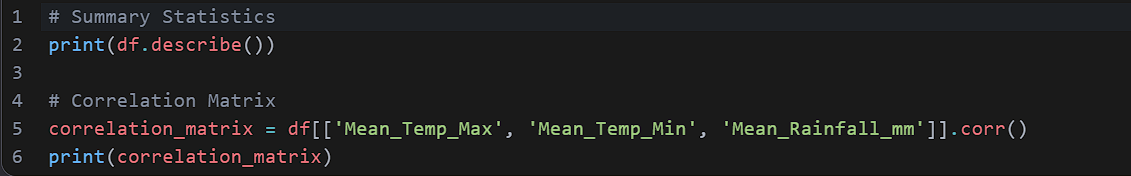
**3.2. Data Cleaning and Month Ordering**

To ensure consistency, we renamed the columns for easier access and converted the "Month" column into a categorical variable with a predefined order (January to December).



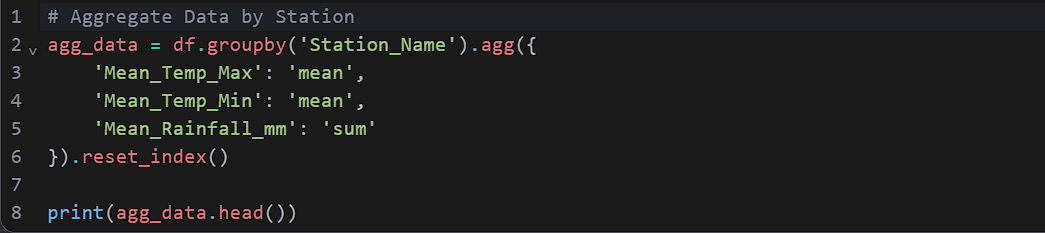
**3.3. Basic Statistical Exploration**

We performed basic statistical analysis to understand the distribution of temperature and rainfall data.



**3.4. Grouping and Aggregation by Station**

Data was grouped by a station to calculate aggregate statistics such as average maximum and minimum temperatures and total rainfall.



**3.5. Visual Exploration Strategy**

Visualization plays a crucial role in understanding trends and patterns. We employed various plots such as line charts, bar charts, heatmaps, boxplots, and scatter plots to explore relationships between variables.

**4.1 Temperature Trends by Station (Line Plot)**

A line plot was created to visualize monthly temperature trends for a specific station (e.g., Abu).

A screenshot of a computer screen

AI-generated content may be incorrect.

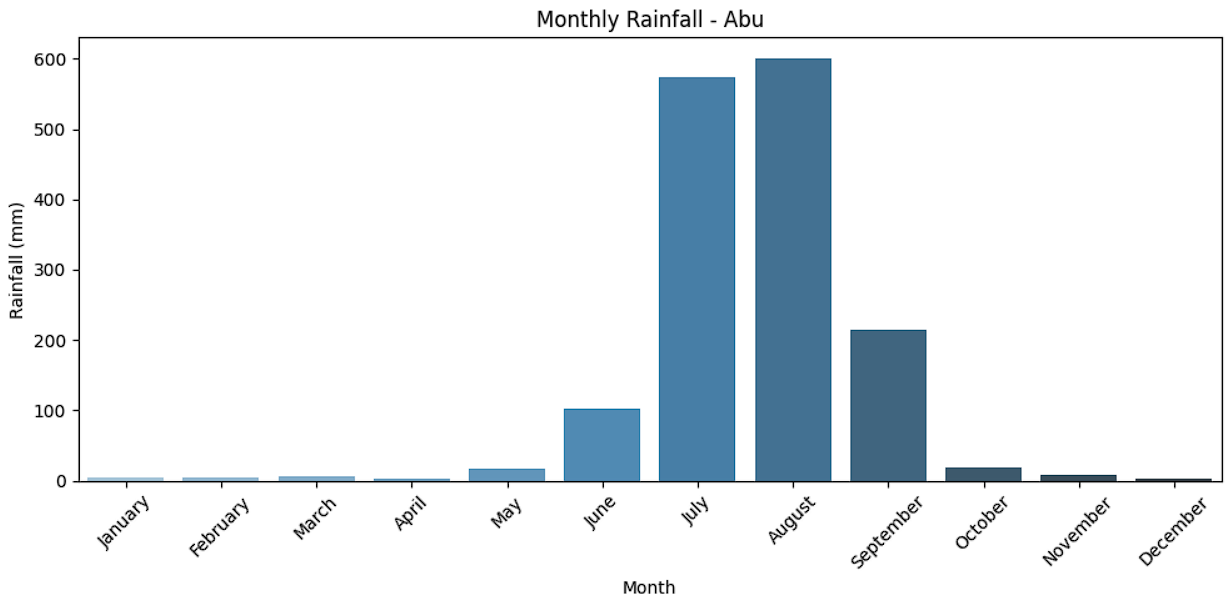
A graph of the temperature trends

AI-generated content may be incorrect.

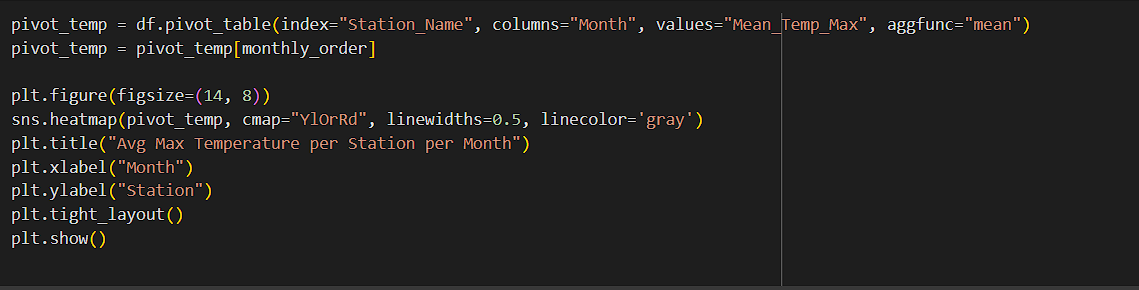
**4.2 Bar plot of mean rainfall for a station**

A computer screen with text

AI-generated content may be incorrect.



**4.3 Heatmap of average max temp per station per month**



A red and orange grid

AI-generated content may be incorrect.

1. Boxplot of max temperatures across all stations

A black screen with white text

AI-generated content may be incorrect.

A graph of different sizes and colors

AI-generated content may be incorrect.

1. **Scatter plot between max and min temp**

A black screen with text

AI-generated content may be incorrect.

A diagram of a rocket

AI-generated content may be incorrect.

Histogram of rainfall values

A computer screen with white text

AI-generated content may be incorrect.

A graph of a person

AI-generated content may be incorrect.

Violin plot of min temps by month

A black screen with white text

AI-generated content may be incorrect.

A chart of different colored shapes

AI-generated content may be incorrect.

**Pairplot of temperature and rainfall**

**A computer screen shot of a black background

AI-generated content may be incorrect.**

A group of blue and white graphs

AI-generated content may be incorrect.

Line plot comparing rainfall across 3 stations

A computer screen shot of text

AI-generated content may be incorrect.

A graph of a number of people

AI-generated content may be incorrect.

Grouped bar chart of avg max & min temp for top 5 stations

A screen shot of a computer code

AI-generated content may be incorrect.

A graph of different colored bars

AI-generated content may be incorrect.

**Conclusion**

Through this analysis, we observed significant variations in temperature and rainfall patterns across different stations. Key insights include seasonal trends, extreme weather events, and regional disparities. These findings can inform policymakers and researchers working on climate resilience strategies.

**Future Scope**

Future work could involve incorporating additional datasets, such as greenhouse gas emissions or land-use changes, to better understand drivers of climate variability. Advanced machine learning models could also predict future climate scenarios

**7. References**

* **Source Dataset**: **climate\_change\_upto\_2000\_1.csv**
* **Python Documentation - Pandas Library. Retrieved from:** [**https://pandas.pydata.org**](https://pandas.pydata.org)
* **Seaborn Visualization Library, Official Documentation. Accessed from:** [**https://seaborn.pydata.org**](https://seaborn.pydata.org)
* **Course Lecture Notes** – Data Science & Python, INT375,LPU.

8.**Linkedin:**

<https://www.linkedin.com/posts/satendrayadav21_climatechange-datascience-python-activity-7316865231769554944-JHC1?utm_source=share&utm_medium=member_android&rcm=ACoAAEdKiY8BOUzXQn1sTPSjTXFIRo0g3Hz5aAc>

**9.Github:**

[**https://github.com/satendra219/python-Project**](https://github.com/satendra219/python-Project)