**DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING**  
**PROJECT REPORT**  
(Project Semester January-April 2025)  
**District Level Data Analysis using Python**

Submitted by  
**Bhavesh Kanwar**  
Registration No: 12321008  
Programme and Section: **B.Tech CSE**

**Section :K23GD**  
Course Code: **INT375**

Under the Guidance of  
**Baljinder Kaur (UID: 28968)**

**Assistant Professor**  
Discipline of CSE  
**Lovely School of Computer Science and Engineering**  
Lovely Professional University, Phagwara

**Declaration**

I, Bhavesh Kanwar, student of B.Tech  under CSE Discipline at Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date:11-04-2025.

Registration No: 12321008  
Name: Bhavesh Kanwar

**Certificate**

This is to certify that **Bhavesh Kanwar**, bearing Registration No. 12321008 has completed **INT375** project titled **“District Level Data Analysis using Python”** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort, and study.

Name of the Supervisor: **Baljinder Kaur(UID:28968)**   
Designation: Assistant Professor  
School of Computer Science and Engineering  
Lovely Professional University, Phagwara, Punjab  
Date:11-04-2025

**Acknowledgement**

I would like to express my sincere gratitude to my project guide **Baljinder Kaur mam**, Assistant Professor, School of Computer Science and Engineering, for his valuable guidance and continuous support throughout the development of this project. I would also like to thank Lovely Professional University for providing me with the resources and environment to complete this work.

**Table of Contents**

1. Introduction
2. Source of Dataset
3. EDA Process
4. Analysis on Dataset
   * General Description
   * Specific Requirements, Functions and Formulas
   * Analysis Results
   * Visualization
5. Conclusion
6. Future Scope
7. References

**1. Introduction**

This project titled **"District Level Data Analysis using Python"** aims to explore and analyze agricultural fertilizer consumption data across Indian districts using Python. The goal is to understand trends in the use of Nitrogen, Phosphate, and Potash and visualize the spatial and statistical characteristics of this data.

**2. Source of Dataset**

The dataset used for this project is **"ICRISAT-District Level Data.csv"**, which provides district-wise consumption data of fertilizers in India. It includes:

* Nitrogen per hectare of gross cropped area
* Phosphate per hectare of gross cropped area
* Potash per hectare of gross cropped area
* Total consumption in tons

Source: International Crops Research Institute for the Semi-Arid Tropics (ICRISAT)

**3. EDA Process**

* **Data Importing**: Used pandas to load the dataset.
* **Data Cleaning**: Stripped column names and removed rows with missing fertilizer data.
* **Feature Engineering**: Created a new feature – Total NPK (kg/ha).
* **Summary Statistics**: Used .describe(), .info(), and .isnull().sum().

**4. Analysis on Dataset**

**a) General Description**

The analysis focuses on:

* Exploring NPK fertilizer distribution.
* Identifying outliers.
* Understanding correlation between nutrients.
* Analyzing state-level nitrogen usage.

**b) Specific Requirements, Functions and Formulas**

* Libraries: pandas, numpy, matplotlib.pyplot, seaborn
* Custom Feature:

python

CopyEdit

df['Total NPK (kg/ha)'] = df[fertilizers].sum(axis=1)

* Visualization: sns.histplot, sns.boxplot, sns.heatmap, sns.violinplot

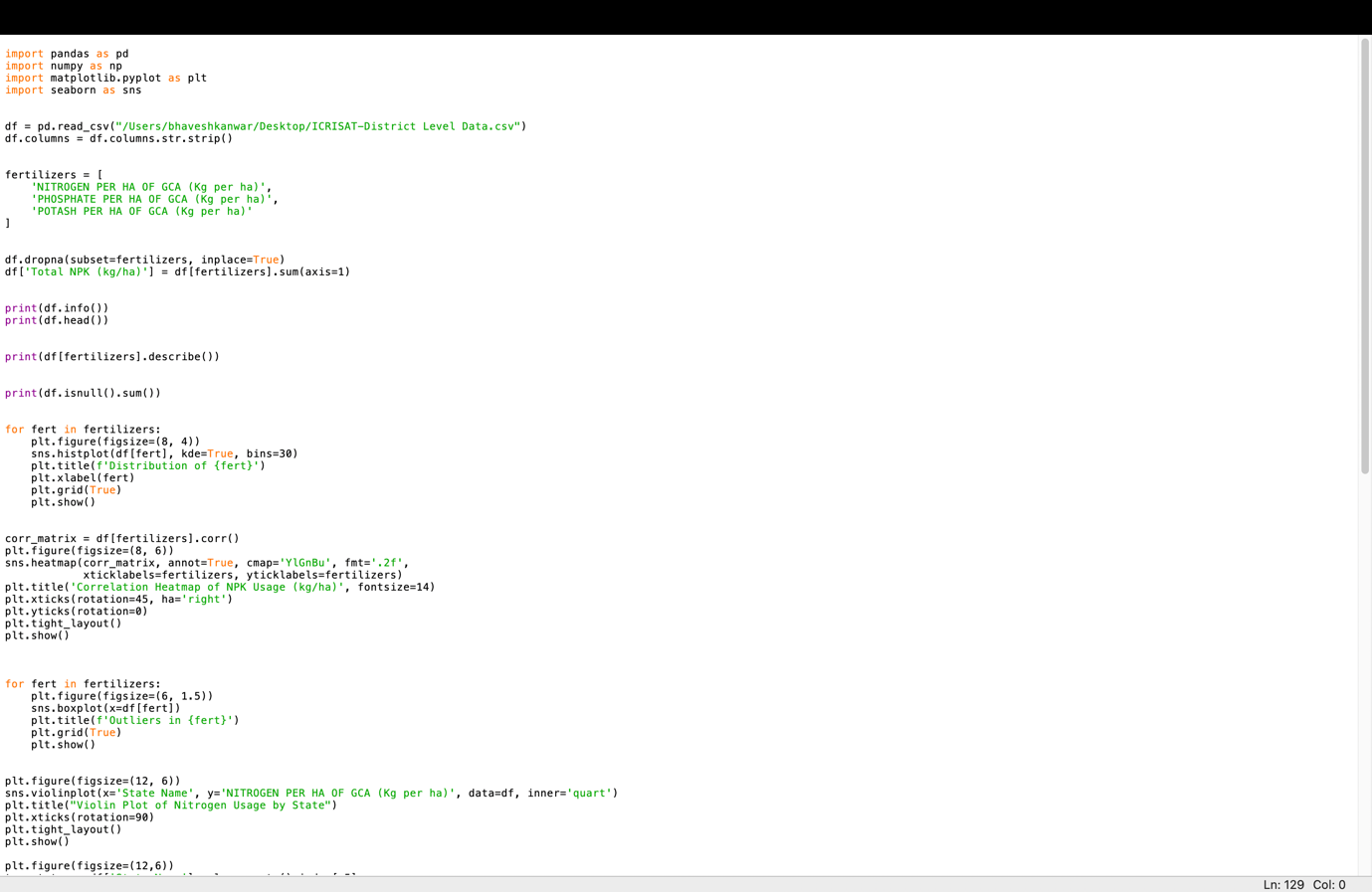
**c) Analysis Results**

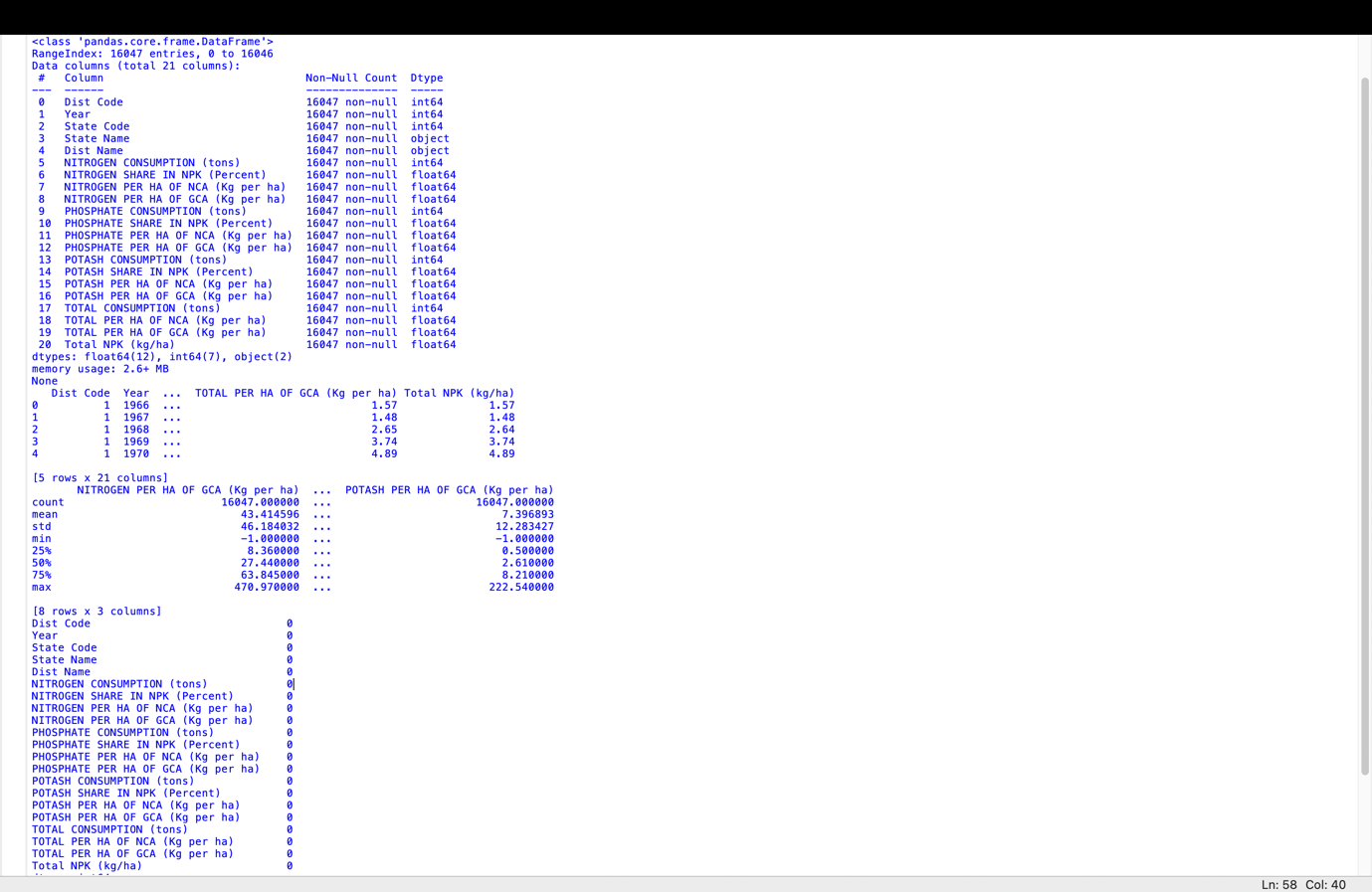
* Nitrogen, Phosphate, and Potash showed varied distribution across districts.
* Positive correlation found between N, P, and K usage.
* Outliers present in all three fertilizers.
* States show different nitrogen usage patterns.

**d) Visualization**

Visualizations were generated and saved as PNG images:

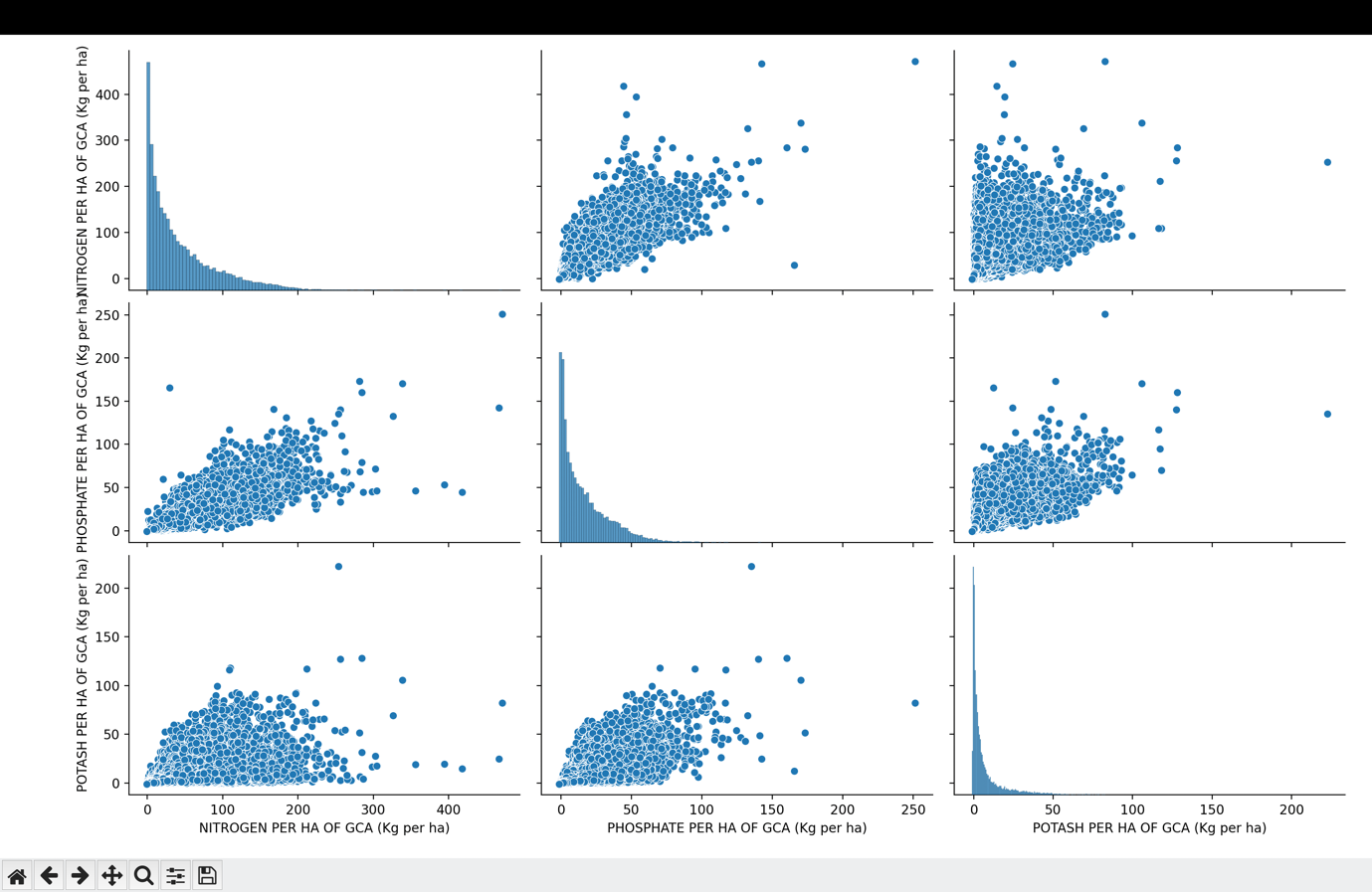
* Histogram of each nutrient (N, P, K)
* Boxplot for outliers
* Heatmap of correlations
* Violin plot for state-wise nitrogen usage

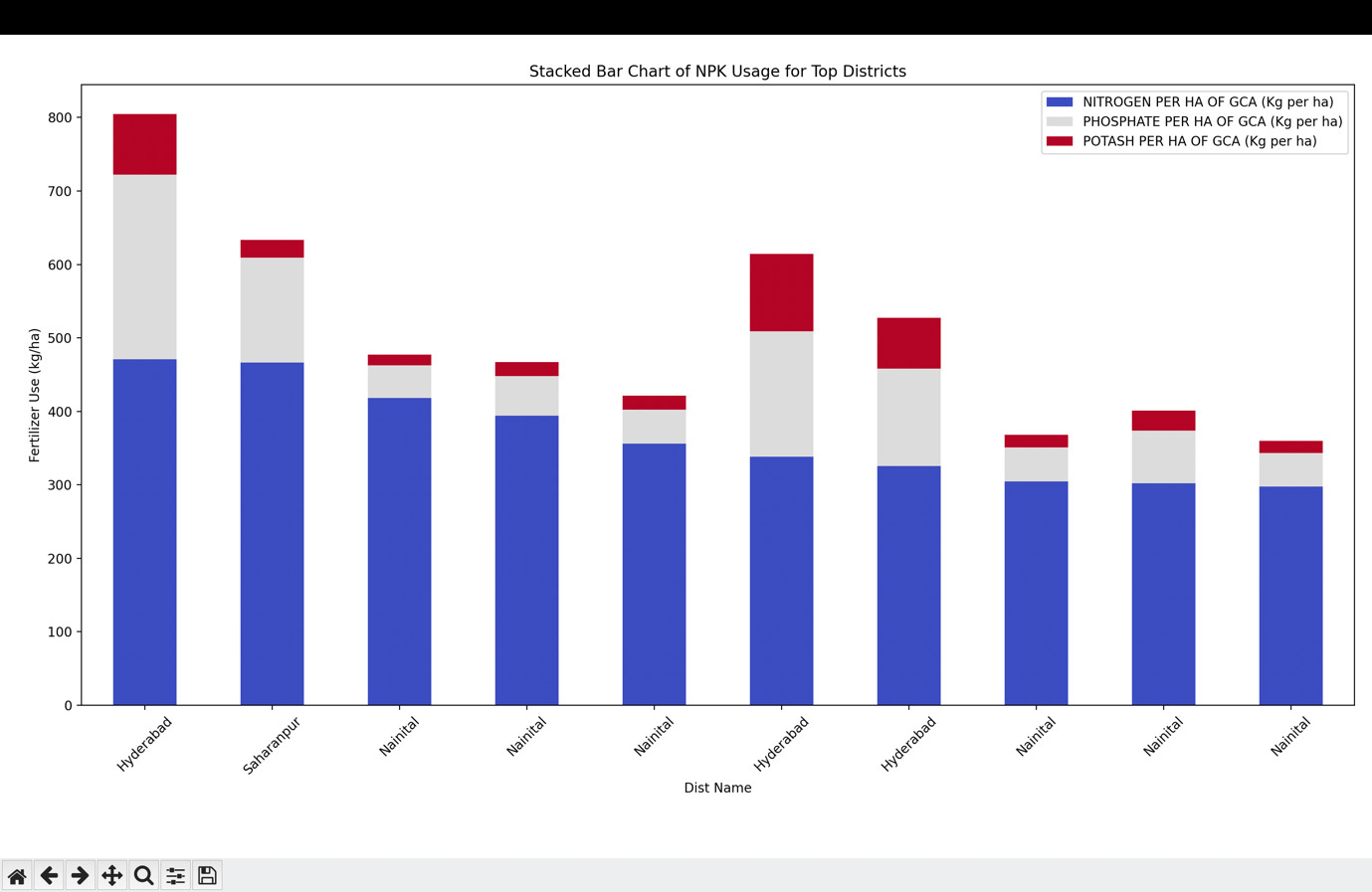


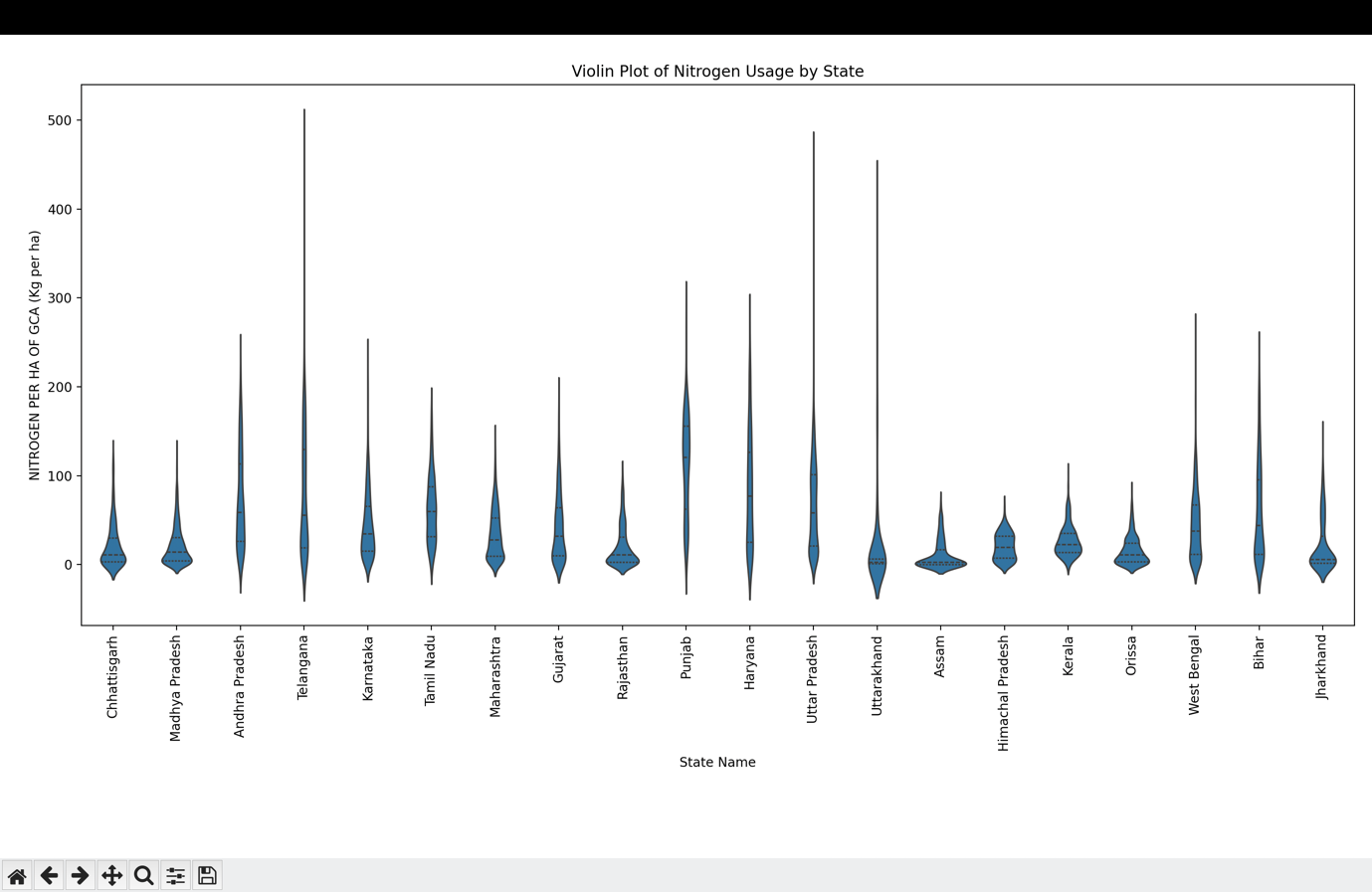


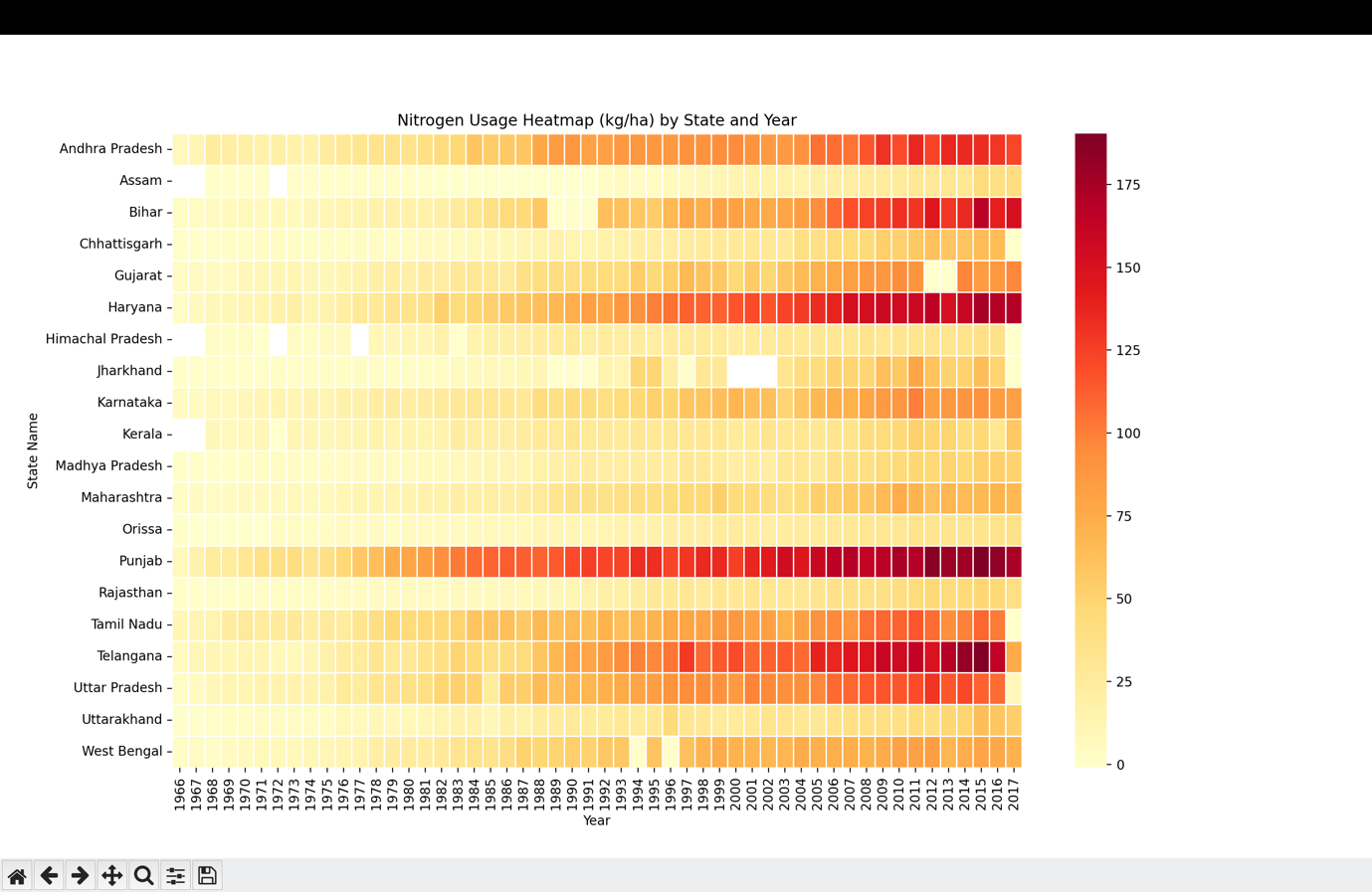
A screenshot of a pie chart

AI-generated content may be incorrect.









A graph with a bar and a line

AI-generated content may be incorrect.

[1] Linkedin like - <https://www.linkedin.com/feed/update/urn:li:activity:7316781744500682752/>

[2] Github - https://github.com/Bhaveshkanwar

**5. Conclusion**

The project titled *"District Level Data Analysis using Python"* successfully achieved its goal of analyzing and visualizing fertilizer consumption patterns across various districts in India. Through systematic exploratory data analysis (EDA), we uncovered key trends, such as the uneven distribution of Nitrogen, Phosphate, and Potash usage, the presence of notable outliers, and significant correlations between the nutrients.

The visualizations provided intuitive insights, making the findings accessible for both technical and non-technical audiences. This analysis not only sheds light on current agricultural fertilizer practices but also builds a foundation for deeper data-driven decision-making in the agricultural sector.  
Overall, the project highlights the potential of Python-based data science techniques in understanding large-scale agricultural datasets and deriving meaningful conclusions that could assist policymakers, researchers, and farmers.

**6. Future Scope**

1. **Integration of Additional Data Features**:  
   Future work can incorporate more variables such as rainfall, soil quality, crop type, and overall agricultural productivity to understand how these factors influence fertilizer requirements.
2. **Time-Series Analysis**:  
   By including data from multiple years, it would be possible to perform time-series forecasting and observe how fertilizer usage trends change over time.
3. **Geospatial Visualization**:  
   Using libraries like Folium or GeoPandas, fertilizer consumption can be mapped geographically to create interactive district-wise and state-wise maps, offering a better spatial understanding.
4. **Predictive Modeling with Machine Learning**:  
   Machine learning algorithms such as linear regression, random forests, or gradient boosting could be applied to predict fertilizer needs based on environmental and agricultural inputs.
5. **Cluster Analysis**:  
   Districts can be grouped based on similar fertilizer consumption patterns using clustering algorithms like K-Means or Hierarchical Clustering, providing insights into regional similarities and differences.
6. **Policy Recommendations**:  
   Based on the analysis, specific recommendations could be made for optimizing fertilizer usage to balance agricultural productivity with environmental sustainability.
7. **Real-Time Dashboard Creation**:  
   Develop an interactive web-based dashboard using tools like Dash or Tableau to allow dynamic exploration of fertilizer data for different districts and states.

**7. References**

[1] ICISAT Dataset – http://data.icrisat.org/dld/src/crops.html  
[2] Ws McKinney, “Python for Data Analysis,” O’Reilly Media, 2018.  
[3] Seaborn & Matplotlib Documentation – https://seaborn.pydata.org  
**[4]** McKinney, Wes. *"Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython"*, 2nd Edition, O'Reilly Media, 2018.

**[5]** Hunter, J.D. *"Matplotlib: A 2D Graphics Environment."* Computing in Science & Engineering 9.3 (2007): 90-95.

**[6]** Waskom, Michael. *"Seaborn: Statistical Data Visualization."* Journal of Open Source Software, 6(60), 2021.

**[7]** Pedregosa, F., et al. *"Scikit-learn: Machine Learning in Python."* Journal of Machine Learning Research, 12 (2011), 2825–2830.

**[8]** Kaggle Tutorials - *https://www.kaggle.com/learn/data-visualization*

**[9]** Official Pandas Documentation - *https://pandas.pydata.org/docs/*