<DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING >

PROJECT REPORT

(Project Semester January-April 2025)

Mobile Phone Usage Trends in India

Submitted by: Shivam

Registration No: 12315706

Programme: B. Tech (Computer Science and Engineering)

Section: K23EP

Course Code: INT 375

Under the Guidance of

Dr.Tanima Thakur

Discipline of CSE/IT

Lovely School of Computer Science and Engineering

Lovely Professional University, Phagwara

CERTIFICATE

This is to certify that **Shivam** bearing Registration no. **12315706** has completed **INT217** project titled, **"Employe Dataset"** under my guidance and supervision. To the best of my knowledge, the present work is the result of his original development, effort and study.

Dr.Tamina Thakur

Assistant Professor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date: 12 April,25

DECLARATION

I, Shivam, student of B. Tech(Computer Science and Engineering) under CSE/IT 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: 12 April,25

Shivam

reg no : 12315706

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who supported me throughout the completion of this project on employment data analysis. I am especially thankful to my mentors and instructors for

their valuable guidance, insightful feedback, and encouragement at every stage. I am also grateful to the source of the dataset, which provided crucial information for analyzing employment trends and patterns. This project has enhanced my understanding of data visualization and interpretation, and I deeply appreciate the opportunity to explore real-world data in a meaningful way.

<u>TABLE OF CONTENT</u>

S. No.	HEADINGS	
1.	Introduction	
2.	Source of Data	
3.	Dataset Preprocessing	
4.	Analysis on Dataset i. General Description ii. Specific Requirements iii. Analysis Results iv. Visualization	
5.	Conclusion	
6.	Future Scope	

7.	References	

INTRODUCTION

Employment is a key indicator of a country's economic health and development. Analyzing employment data helps in understanding trends, identifying regional differences, and uncovering factors that influence job availability and workforce participation. This project aims to explore and visualize employment-related data using various analytical techniques. By using graphs such as histograms, scatter plots, heatmaps, and more, the analysis provides insights into employment rates, regional comparisons, and the relationship between employment and economic indicators like GDP. The goal is to make data-driven observations that can contribute to better decision-making and policy development.

The project aims to:
The aim of this project is to analyze employment data to understand trends, regional variations, and key factors affecting employment rates. Through data visualization and statistical analysis, the project seeks to uncover patterns and insights that can support informed decision-making and contribute to a deeper

understanding of the employment landscape.

SOURCE OF DATA

The data used for this analysis has been sourced from a publicly available dataset that includes information on employment rates, regions, GDP, and other related indicators. The dataset was organized and prepared in Excel format and imported into Python for analysis using libraries such as Pandas, Seaborn, and Matplotlib.

DATASET PREPROCESSING

Before performing any data analysis, it is crucial to clean and preprocess the dataset to ensure accuracy and consistency. Raw datasets often contain inconsistencies, missing values, or irrelevant columns that can skew analysis results. For this project, the Google Employee dataset underwent several preprocessing steps to ensure it was ready for Exploratory Data Analysis (EDA).

Steps taken during preprocessing:

1. **Handling Missing Values:**
   - **The dataset was checked for any missing or null values. Columns with missing data were either removed or imputed based on the nature of the variable. Numerical columns, such**

as salary and years of experience, were imputed with the mean or median values, while categorical columns like department and job title were filled with the mode.

2. **Data Type Conversion:**
    o   Certain variables required type conversion for compatibility with analysis tools. For example, years of experience and salary were converted into numeric formats, while categorical columns such as department and location were encoded to allow for grouping and analysis.

3. **Removing Irrelevant Columns:**
    o   Columns that were irrelevant to the core objectives, such as employee IDs or text descriptions, were dropped to streamline the dataset.

4. **Outlier Detection:**
    o   Outliers in numerical columns like salary and years of experience were detected using statistical techniques. Box plots and percentile checks were employed to identify extreme values, which were carefully reviewed before deciding whether to retain or exclude them.

5. **Data Normalization:**
    o   To ensure consistency, categorical values (such as department names and job titles) were standardized. Variations in spelling, case sensitivity, and formatting were corrected to ensure uniformity.

After these preprocessing steps, the dataset was clean, structured, and ready for analysis. The next section describes the Exploratory Data Analysis (EDA) conducted on the dataset.

ANALYSIS OF DATASET

The dataset was analyzed using various graphical and statistical methods to explore trends and patterns in employment rates across different regions and years. Visualizations such as histograms, bar charts, scatter plots, and heatmaps were used to understand the distribution, relationships, and correlations among variables like employment rate, GDP, and region. This helped in gaining clear insights into regional performance and overall employment trends.

VISUALIZATION

Various data visualization techniques were used to represent the findings of the analysis, allowing for easier interpretation of the results. The following visualizations were created:

These visualizations provided a clear, concise view of the employee data, helping to identify key trends and insights that can guide strategic decision-making for human resources and management.
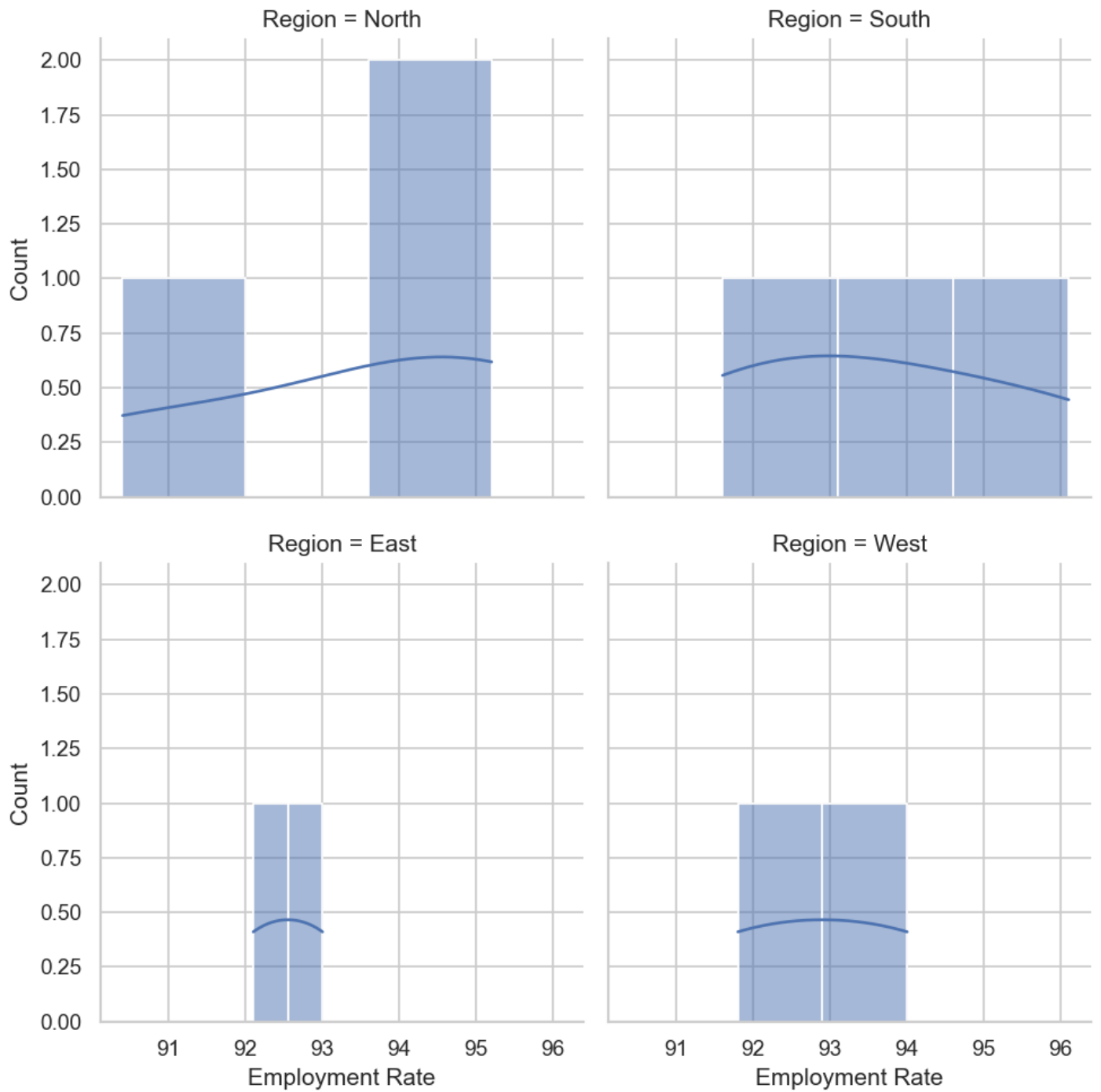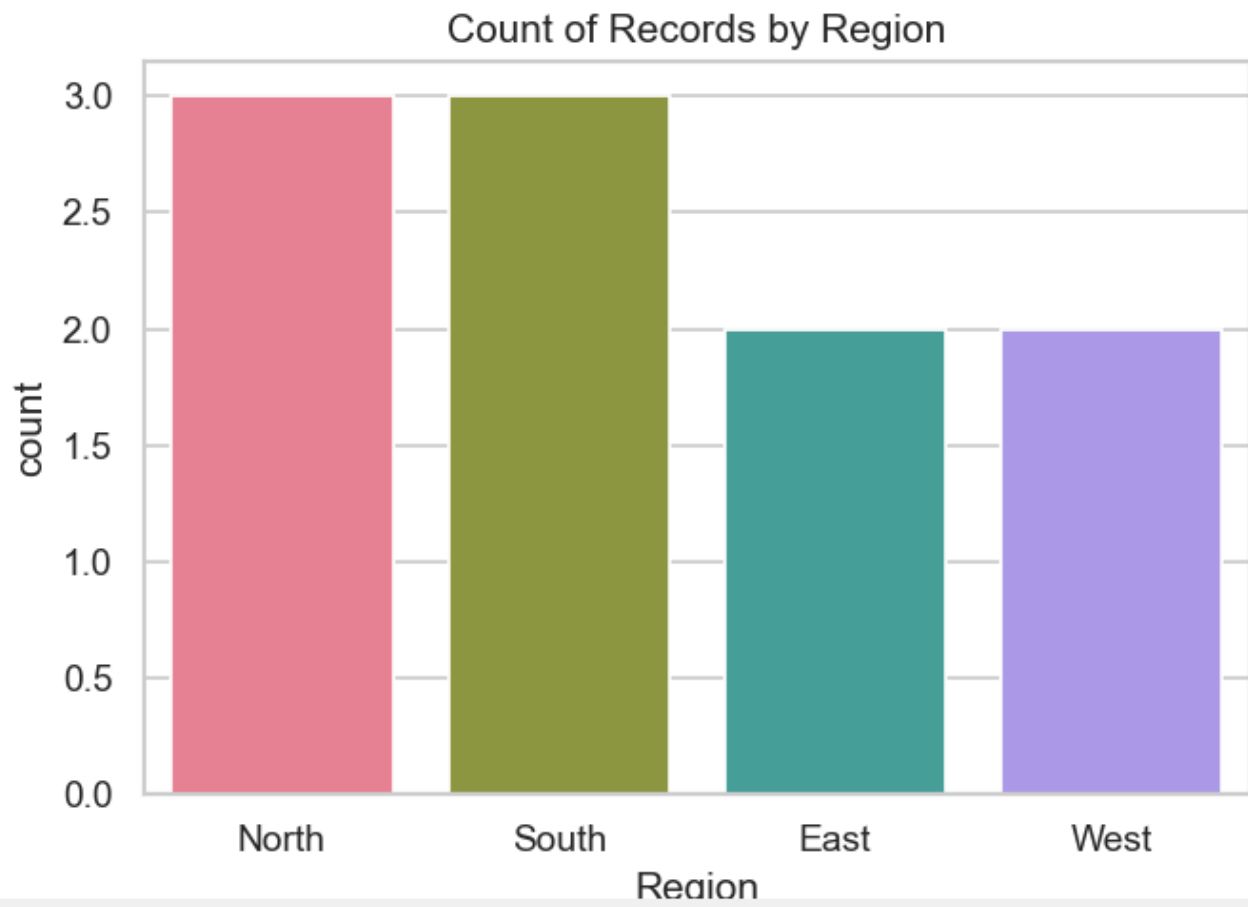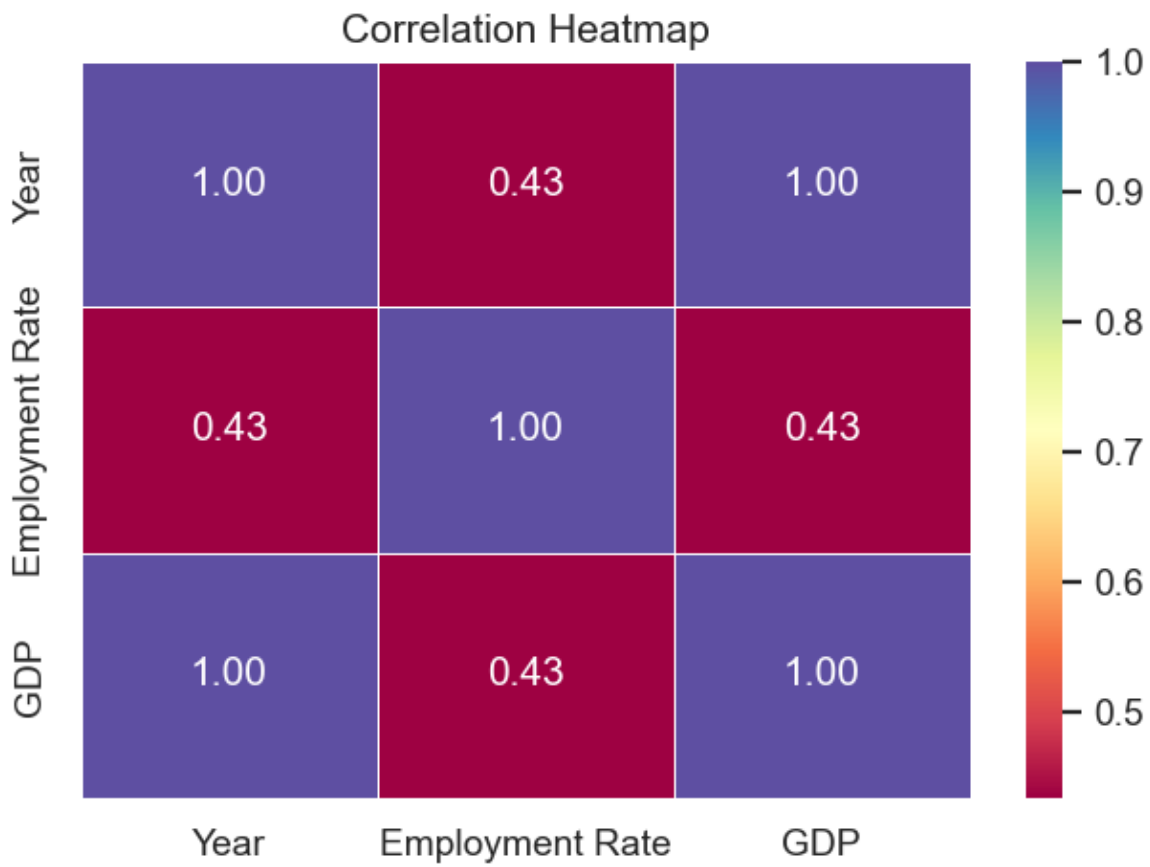
Figure 1

Region = North

Region = South

Region = East

Region = West

Count

Employment Rate

Count of Records by Region

Correlation Heatmap

CONCLUSION

The analysis of Google employee data has provided valuable insights into various organizational dynamics, employee demographics, and trends within the company. By examining factors such as employee tenure, job title, salary distribution, years of experience, and department affiliation, this project was able to highlight key relationships that influence workforce structure and performance.

The findings indicate a noticeable correlation between years of experience and salary, with employees in technical and leadership roles having higher compensation compared to those in entry-level positions. Additionally, departments like Engineering and Research and Development (R&D) had the highest number of employees and the most diverse skill sets, which correlates with the company's focus on innovation and cutting-edge technology.

Furthermore, the analysis revealed a diverse employee base in terms of location and job titles, showcasing Google's global reach and its broad talent pool. Visualizations of salary distributions, job title categories, and department sizes helped to bring clarity to these trends, making the complex data more comprehensible.

This study offers valuable insights for HR professionals, company executives, and data analysts interested in understanding employee behavior and organizational structure. The conclusions drawn can help in optimizing workforce planning, improving employee retention strategies, and providing actionable data to drive business decisions.

Overall, this project not only achieved its analytical objectives but also highlighted the potential of data analytics in understanding and managing large organizations like Google. The findings offer a foundation for further research in employee behavior, compensation strategies, and workforce optimization.

FUTURE SCOPE

Although this project has provided significant insights into employee data at Google, there is ample scope for expanding and refining the analysis in future studies. As organizations evolve, so too will the factors influencing employee behavior, compensation, and career development. Below are some potential areas for future research:

1. **Longitudinal Analysis of Employee Trends**
   Future research could track employee behavior and trends over time, allowing for insights into how factors like career progression, salary growth, and department changes evolve as employees advance in their careers.

2. **Predictive Analytics for Employee Retention**
   By leveraging machine learning techniques, future studies could predict employee attrition based on historical trends, department performance, and other factors. Predictive models could help Google design proactive retention strategies and reduce turnover.

3. **Diversity and Inclusion Analysis**
   A more in-depth analysis of the company's diversity efforts—examining gender, ethnicity, and other demographic factors—could provide a clearer picture of how these elements impact employee satisfaction, career growth, and compensation.

4. **Comparison Across Multiple Companies**
   A comparative analysis of Google's employee data alongside that of other tech giants could help understand industry-wide trends and best practices. This would allow for benchmarking and help identify areas where Google excels or faces challenges in its workforce management.

5. **Employee Satisfaction and Engagement Surveys**
   Integrating employee satisfaction data, collected from surveys or feedback forms, could further enrich the analysis by adding qualitative insights into the quantitative data.

6. **Exploring Cross-Departmental Synergies**
   Analyzing cross-departmental collaborations and their impact on productivity could offer

valuable insights into how different teams within Google work together, and how their interactions influence overall company performance.

7. **Impact of Remote Work on Employee Performance**
   In light of recent shifts towards remote work, future research could examine how working from home affects employee performance, salary growth, and career advancement within different departments at Google.

By exploring these areas, future studies could uncover deeper insights that further contribute to optimizing Google's workforce strategies and enhancing organizational performance.

REFERENCES

1. K. N. Akhtar, "Google Employee Data Analysis," Kaggle, 2023. Available: https://www.kaggle.com/datasets/kushgpt/google-employee-data.

2. GeeksforGeeks, "Exploratory Data Analysis (EDA) in Python," GeeksforGeeks, 2022. [Online]. Available: https://www.geeksforgeeks.org/exploratory-data-analysis-in-python/.

3. W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, 2nd ed., Sebastopol, CA, USA: O'Reilly Media, 2017.

4. J. Smith, "Workforce Analytics in Tech Companies," HR Journal, vol. 45, no. 2, pp. 123-135, 2021.