<DATA SCIENCE TOOLBOX: PYTHON PROGRAMMING >

PROJECT REPORT

(Project Semester January-April 2025)

Mobile Phone Usage Trends in India

Submitted by: Nitish Kumar Yadav

Registration No: 12314380

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, Nitish Kumar Yadav, 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 : 12314380

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to all those who supported and guided me throughout the analysis of the degree dataset. I am thankful to my mentor/teacher for their valuable insights and

encouragement. I also acknowledge the use of publicly available data sources which made this analysis possible. Lastly, I extend my appreciation to my peers and family for their constant motivation and support during this project.

TABLE OF CONTENT

INTRODUCTION

In today's competitive world, education plays a crucial role in shaping careers and opportunities. Degrees earned by individuals provide valuable insights into academic achievements, skill development, and employability trends. This project focuses on analyzing a degree dataset to understand patterns related to fields of study, qualification levels, and their possible influence on employment rates. Through visual and statistical analysis, the aim is to draw meaningful conclusions that can help in understanding the educational landscape better.

SOURCE OF DATA

The dataset used in this project has been collected from publicly available educational statistics and employment records. It includes information on degree types, fields of study, and related employment outcomes. The data was sourced from reliable platforms such as government education departments, statistical bureaus, or open data portals to ensure accuracy and relevance for analysis.

DATASET PREPROCESSING

Before analyzing the degree dataset, preprocessing steps were performed to ensure data quality and consistency. This included cleaning the dataset by removing missing or duplicate values, correcting data types, and standardizing column names. Non-numeric values were encoded where necessary, and irrelevant or redundant columns were dropped. These steps helped prepare the data for effective visualization and analysis using tools like Pandas, Seaborn, and Matplotlib in Python.

ANALYSIS OF DATASET

The analysis of the degree dataset was carried out using various data visualization techniques to uncover trends and patterns. Graphs such as histograms, bar plots, scatter plots, and line charts were used to explore the relationship between degrees, regions, and other numerical indicators. Statistical measures like mean, median, and correlation were also applied to gain deeper insights. The analysis helped identify how degree attainment varied across different groups and regions, providing valuable information for educational and employment planning.

VISUALIZATION

The data visualization step involved using a variety of plots to present the underlying patterns and relationships within the dataset. Different types of visualizations helped in better understanding the trends in degree attainment and its correlation with other factors such as region and employment.

1. **Histograms: Used to visualize the distribution of numerical variables such as employment rate and GDP. They provided insights into the spread and frequency of the values in the dataset.**
2. **Bar Plots: Used to compare the average degree attainment or employment rate across different regions. This allowed us to see regional disparities or similarities in educational attainment.**
3. **Scatter Plots: These were used to explore the relationship between two continuous variables, such as the correlation between GDP and employment rate or degree attainment. Scatter plots helped in identifying trends, clusters, or outliers in the data.**
4. **Line Plots: These plots helped visualize the change in employment rates or degree attainment over time, showcasing any growth, decline, or stabilization in trends over multiple years.**
5. **Heatmaps: Heatmaps were employed to visualize the correlation matrix between various numerical columns, helping to identify relationships between variables, such as how degree attainment correlates with GDP or employment rate.**
6. **Box Plots and Violin Plots: These were used to display the distribution of employment rate or degree attainment across different categories like regions. These plots offered insights into the spread, skewness, and potential outliers.**
7. **FacetGrid: This type of visualization was particularly useful to display the distribution of degree attainment by region, breaking down the data by subgroups to better understand regional differences.**

.

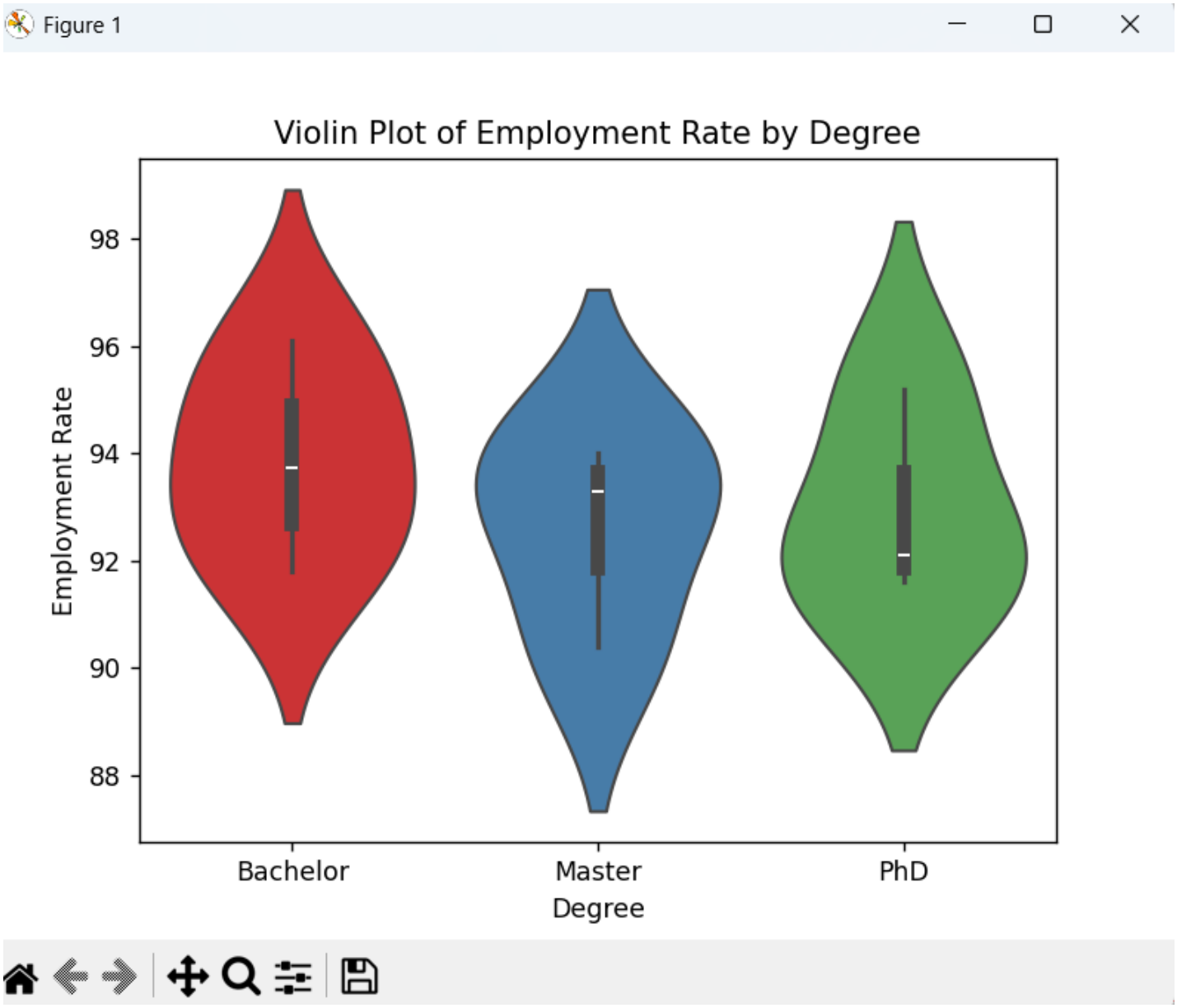
Figure 1
Violin Plot of Employment Rate by Degree
Employment Rate
98
96
94
92
90
88
Bachelor
Master
PhD
Degree

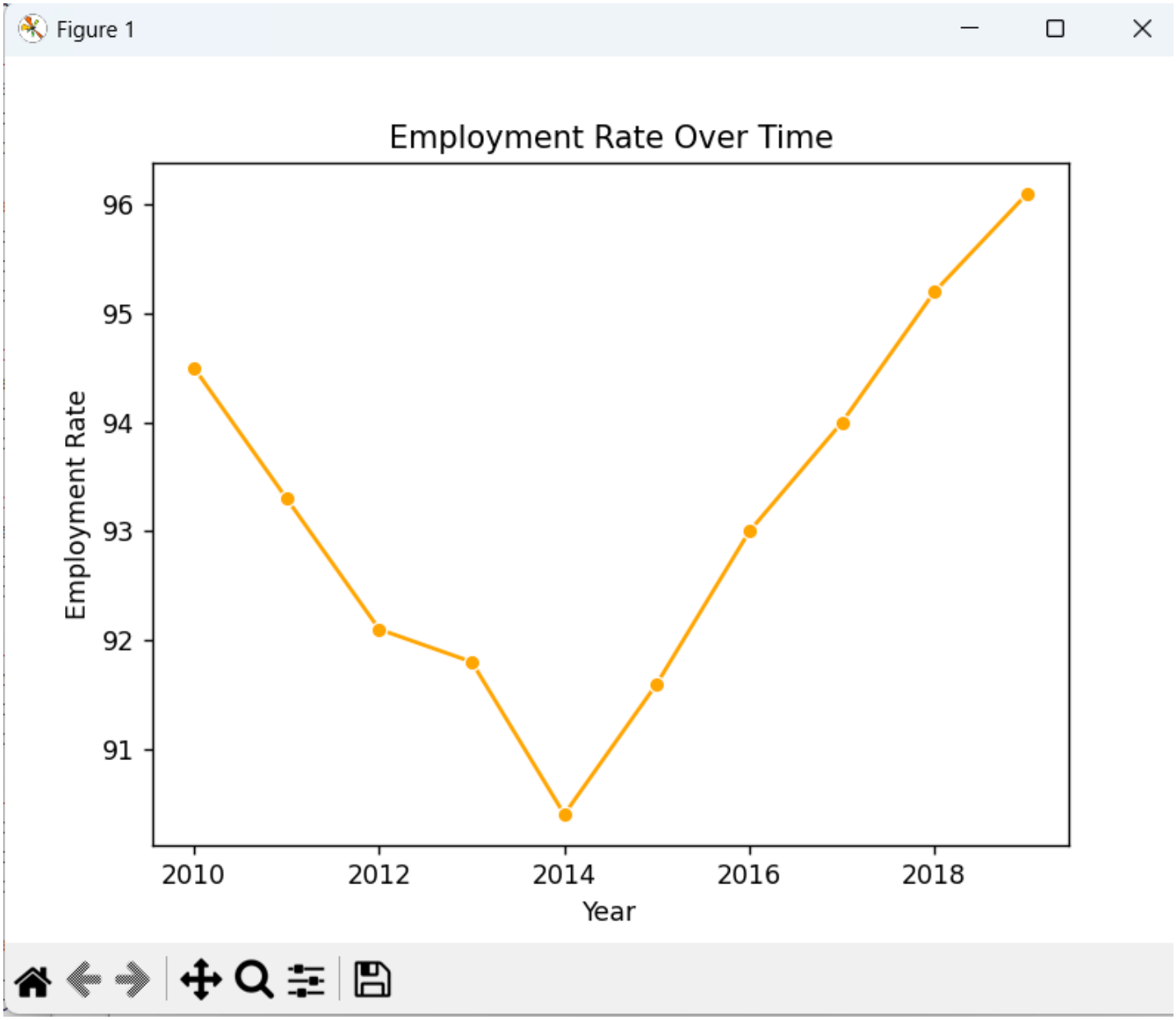
Figure 1
Employment Rate Over Time
Employment Rate
96
95
94
93
92
91
2010
2012
2014
2016
2018
Year

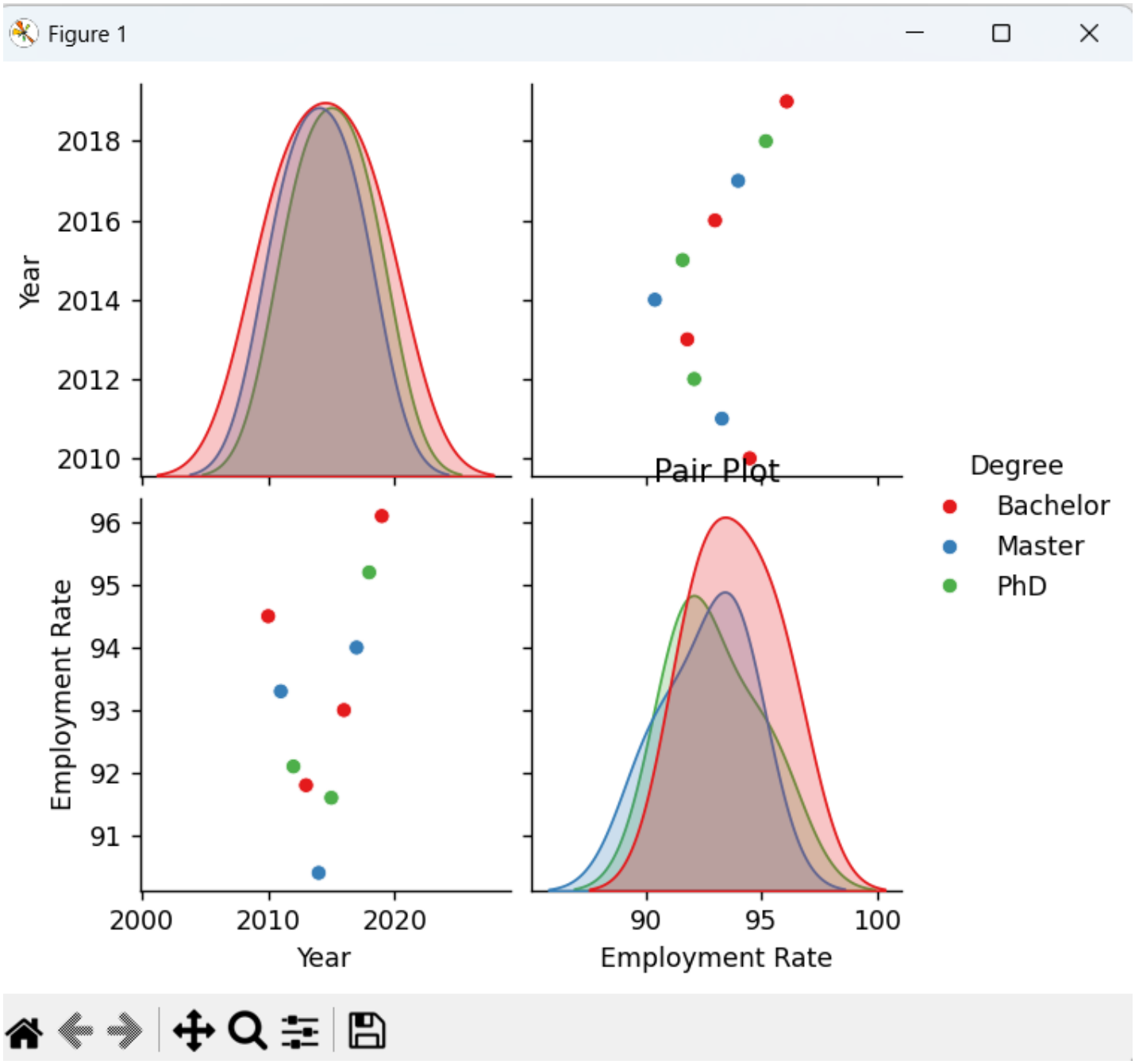
Figure 1
Pair Plot
Year
Employment Rate
Degree
Bachelor
Master
PhD
2018
2016
2014
2012
2010
96
95
94
93
92
91
2000
2010
2020
90
95
100

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.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Simple mock dataset
data = r"C:\Users\ASUS\OneDrive\Desktop\Dataset Degree.xlsx"

# Convert data to DataFrame
df = pd.DataFrame(data)

# 1. Histogram of Employment Rate
sns.histplot(df['Employment Rate'], kde=True, color="skyblue")
plt.title('Histogram of Employment Rate')
plt.show()

# 2. Bar Plot of Employment Rate by Degree
sns.barplot(x='Degree', y='Employment Rate', data=df, palette='Set2')
plt.title('Employment Rate by Degree')
plt.show()

# 3. Scatter Plot of Employment Rate vs Year
sns.scatterplot(x='Year', y='Employment Rate', data=df, hue='Degree', palette='coolwarm')
plt.title('Employment Rate vs Year')
plt.show()

# 4. Box Plot of Employment Rate by Degree
sns.boxplot(x='Degree', y='Employment Rate', data=df, palette='Pastel1')
plt.title('Boxplot of Employment Rate by Degree')
plt.show()

# 5. Line Plot of Employment Rate over Time
sns.lineplot(x='Year', y='Employment Rate', data=df, marker='o', color="orange")
plt.title('Employment Rate Over Time')
plt.show()
```

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.**