TITLE

EMPLOYEE DATA ANALYSIS AND INSIGHTS

****

Prepared by: NIRMAL KUMAR

Date:11/03/2025

Institution/Organization: KIET GROUP OF INSTITUTIONS

Course/Department: CSEAI

University Roll No. : 242801100300162

Submitted to : Mr. Mayank Lakhotia

🔍 **Background & Purpose**
In modern organizations, **employee data analysis** is essential for making informed decisions regarding **salaries, experience, and department-wise performance**. Understanding salary trends, departmental variations, and experience-based compensation structures helps organizations

optimize workforce management, improve employee satisfaction, and ensure fair pay structures.
This project aims to **analyze and visualize employee data** to extract meaningful insights, such as:
✓ Identifying **salary trends** across departments
✓ Finding **correlation between salary & experience**
✓ Detecting **outliers in salaries**
✓ Understanding **departmental salary distribution**
✓ Exploring **employee demographics**
By leveraging **Python, Pandas, Seaborn, and Matplotlib**, this study provides a **data-driven approach** to employee analytics, helping HR professionals and management make strategic workforce decisions.

METHODOLOGY:

🛠 Tools & Technologies Used

- **Python** – For data manipulation and analysis
- **Pandas** – For handling structured data (CSV file)

- **Matplotlib & Seaborn** – For visualizing salary distribution, correlations, and trends
- **Statistical Analysis** – Using correlation, outlier detection, and department-wise insights

🗁 **Steps Followed**

Step 1: Data Collection & Preprocessing

- The employee dataset (CSV format) was loaded into Python using Pandas.
- Initial checks were performed to inspect missing values and data types.

Step 2: Exploratory Data Analysis (EDA)

- **Summary statistics** were computed, including min, max, mean, and standard deviation of salary, experience, and other numerical attributes.
- **Department-wise average salary & experience** were analyzed.
- **Correlation analysis** was performed between salary and experience.

Step 3: Advanced Analysis & Outlier Detection

- **Outliers in salary** were detected using the **Interquartile Range (IQR) method**.
- Employees with **highest & lowest salaries** were identified.

Step 4: Data Visualization

- **Salary Distribution** (Histogram with KDE)
- **Experience vs Salary Scatterplot** (Color-coded by department)
- **Department-wise Salary Boxplot** (To show salary spread)
- **Salary Growth Trendline** (Regression plot)
- **Employee Distribution Pie Chart**

Step 5: Interpretation & Insights

- The results were analyzed to identify key patterns, correlations, and department-wise salary trends.

CODE TYPED:

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

# Load the employee data CSV file
file_path = "employee_data.csv"  # Update if needed
df = pd.read_csv(file_path)

# Display basic dataset information
print("📌 Basic Information:\n", df.info())

# Display first few rows
print("\n📌 First Few Rows of the Dataset:\n",
df.head())

# -------------------- SUMMARY STATISTICS --------------------

print("\n📌 Summary Statistics:\n", df.describe())




# -------------------- DEPARTMENT-WISE ANALYSIS --------------------

# Average salary and experience by department
```

```
dept_analysis = df.groupby("Department")[["Salary", "Experience"]].mean()
print("\n📌 Department-wise Average Salary and Experience:\n", dept_analysis)

# Count of employees in each department
dept_count = df["Department"].value_counts()
print("\n📌 Employee Count per Department:\n", dept_count)

# -------------------- CORRELATION ANALYSIS ---------------
# Compute correlation between Experience & Salary
correlation = df[["Experience", "Salary"]].corr()
print("\n📌 Correlation Between Experience & Salary:\n", correlation)

# -------------------- HIGHEST & LOWEST SALARY EMPLOYEES --------------------

highest_salary = df.loc[df["Salary"].idxmax()]
lowest_salary = df.loc[df["Salary"].idxmin()]
print("\n📌 Employee with Highest Salary:\n", highest_salary)
print("\n📌 Employee with Lowest Salary:\n", lowest_salary)

# -------------------- OUTLIER DETECTION --------------------

# Identify outliers in salary using IQR method
Q1 = df["Salary"].quantile(0.25)
```

```
Q3 = df["Salary"].quantile(0.75)
IQR = Q3 - Q1
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

outliers = df[(df["Salary"] < lower_bound) |
(df["Salary"] > upper_bound)]
print("\n📌 Outliers in Salary:\n", outliers if not
outliers.empty else "No outliers detected.")

# -------------------- DATA VISUALIZATION --------------------

# Plot Salary Distribution
plt.figure(figsize=(8, 5))
sns.histplot(df["Salary"], bins=10, kde=True,
color="blue")
plt.title("Salary Distribution")
plt.xlabel("Salary")
plt.ylabel("Frequency")
plt.show()

# Plot Experience vs Salary
plt.figure(figsize=(8, 5))
sns.scatterplot(x=df["Experience"], y=df["Salary"],
hue=df["Department"], palette="coolwarm")
plt.title("Experience vs Salary")
plt.xlabel("Years of Experience")
plt.ylabel("Salary")
plt.show()
```

```
# Boxplot of Salary by Department
plt.figure(figsize=(10, 6))
sns.boxplot(x="Department", y="Salary", data=df,
palette="Set2")
plt.title("Salary Distribution by Department")
plt.xlabel("Department")
plt.ylabel("Salary")
plt.xticks(rotation=45)
plt.show()

# Salary Growth with Experience (Trendline)
plt.figure(figsize=(8, 5))
sns.regplot(x=df["Experience"], y=df["Salary"],
scatter_kws={"color": "red"}, line_kws={"color":
"black"})
plt.title("Salary Growth with Experience")
plt.xlabel("Experience (Years)")
plt.ylabel("Salary")
plt.show()

# Department-wise Employee Distribution (Pie Chart)
plt.figure(figsize=(6, 6))
plt.pie(dept_count, labels=dept_count.index,
autopct="%1.1f%%",
colors=sns.color_palette("pastel"))
plt.title("Employee Distribution by Department")
plt.show()
```

SCREENSHOTS:

NIRMAL_202401100300162.ipynb

colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo

File Edit View Insert Runtime Tools Help

Commands + Code + Text

```
plt.show()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20 entries, 0 to 19
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   EmployeeID  20 non-null     int64
 1   Age         20 non-null     int64
 2   Department  20 non-null     object
 3   Experience  20 non-null     int64
 4   Salary      20 non-null     int64
dtypes: int64(4), object(1)
memory usage: 932.0+ bytes
📌 Basic Information:
 None

📌 First Few Rows of the Dataset:
   EmployeeID  Age Department  Experience  Salary
0           1   23    Finance           8   93563
1           2   28    Finance           2   41742
2           3   37         HR           8   56905
3           4   23         HR          23  138397
4           5   55         IT          29   96879

📌 Summary Statistics:
        EmployeeID        Age  Experience         Salary
count     20.00000  20.000000   20.000000      20.000000
mean      10.50000  42.450000   16.750000  102503.150000
std        5.91608  13.092686   10.447236   32459.740566
min        1.00000  23.000000    2.000000   41742.000000
25%        5.75000  29.000000    8.000000   82244.750000
50%       10.50000  46.000000   15.500000  101315.500000
75%       15.25000  53.500000   24.250000  132247.500000
max       20.00000  58.000000   34.000000  144637.000000
```

✓ 3s completed at 10:30 AM

NIRMAL_202401100300162.ipynb

colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo

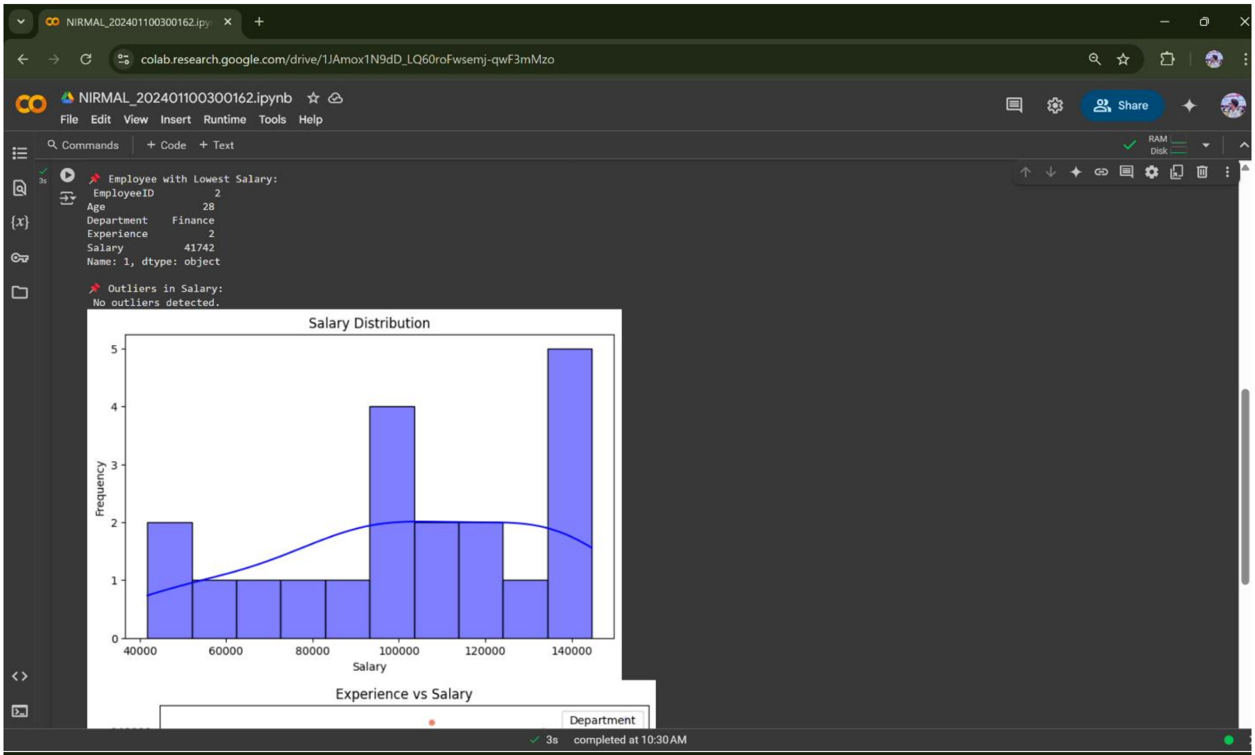
File Edit View Insert Runtime Tools Help

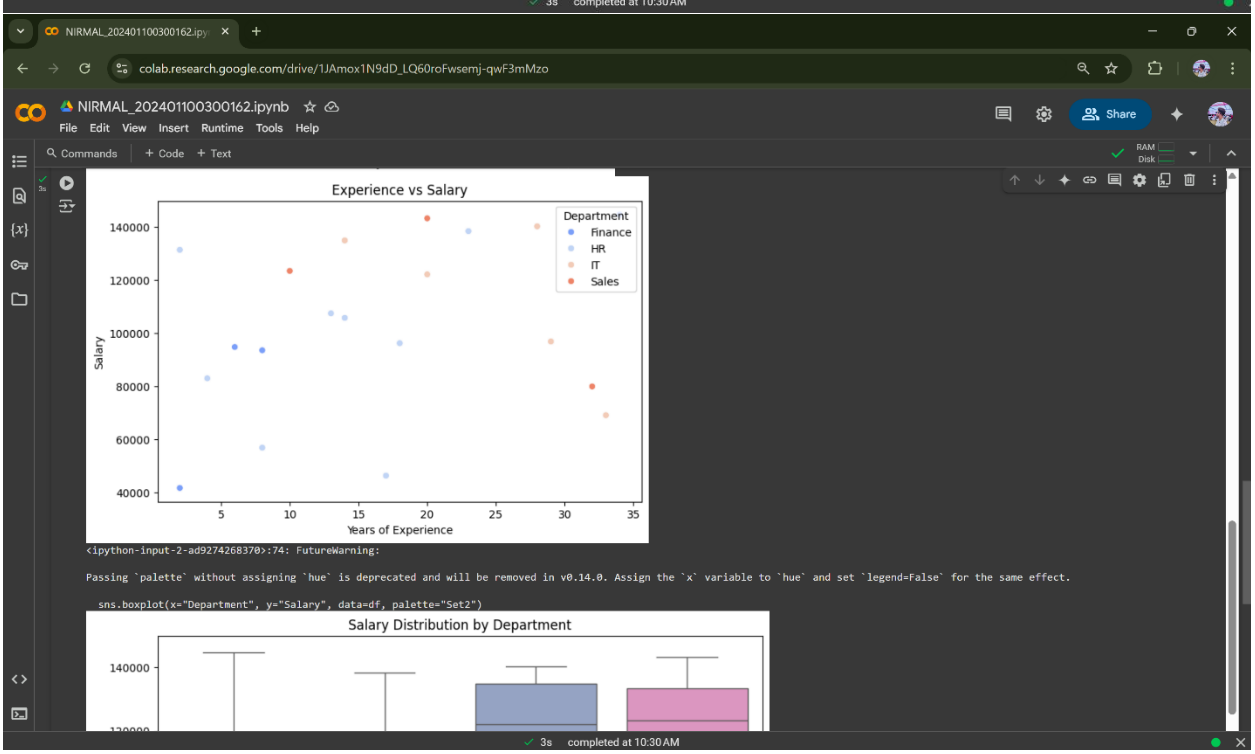
Commands + Code + Text

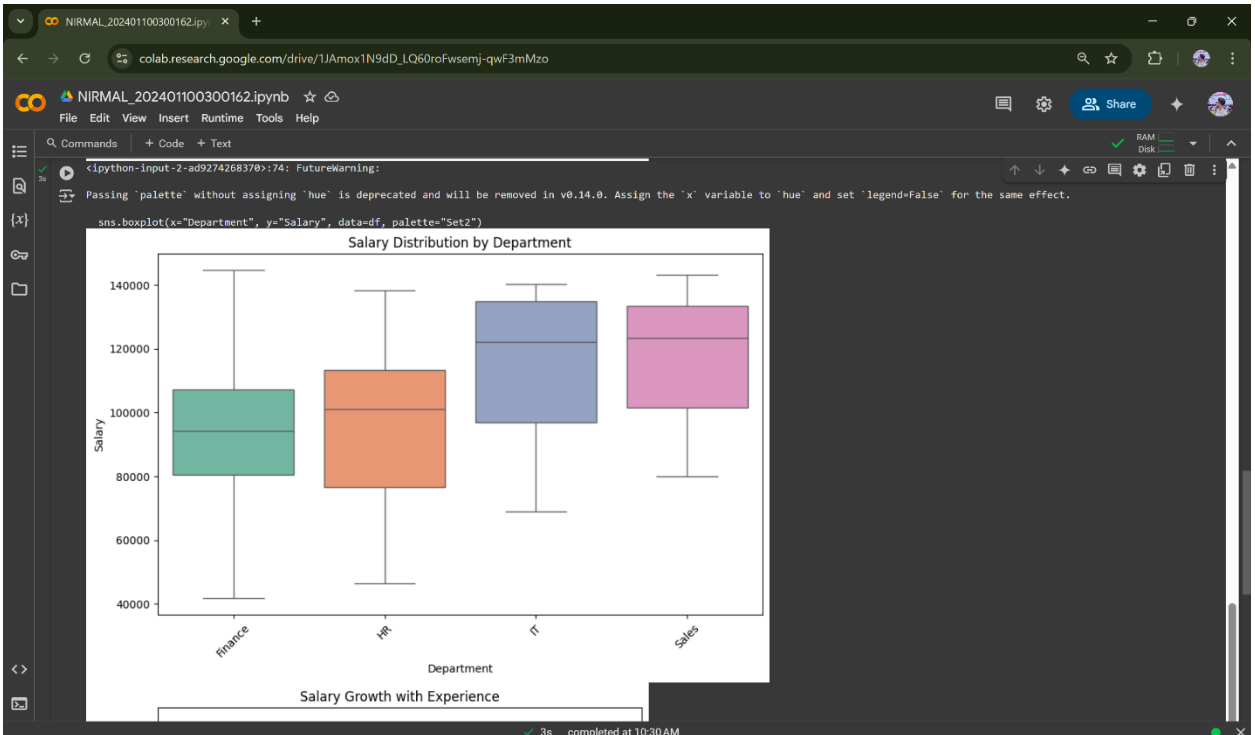
```
plt.show()
```

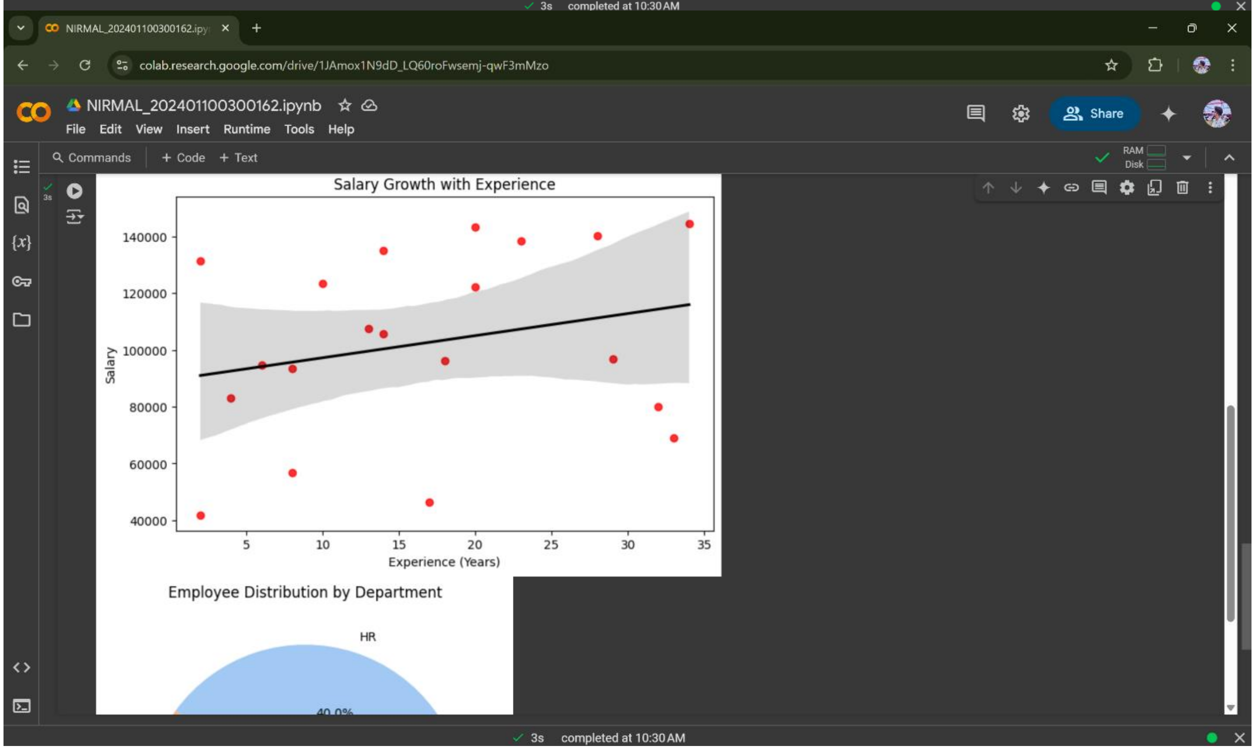

```
📌 Department-wise Average Salary and Experience:
                Salary  Experience
Department
Finance      93680.750   12.500000
HR           95687.125   12.375000
IT          112647.600   24.800000
Sales       115535.000   20.666667

📌 Employee Count per Department:
 Department
HR         8
IT         5
Finance    4
Sales      3
Name: count, dtype: int64

📌 Correlation Between Experience & Salary:
            Experience    Salary
Experience    1.000000  0.250745
Salary        0.250745  1.000000

📌 Employee with Highest Salary:
 EmployeeID          8
Age                46
Department    Finance
Experience         34
Salary         144637
Name: 7, dtype: object

📌 Employee with Lowest Salary:
 EmployeeID          2
Age                28
Department    Finance
```

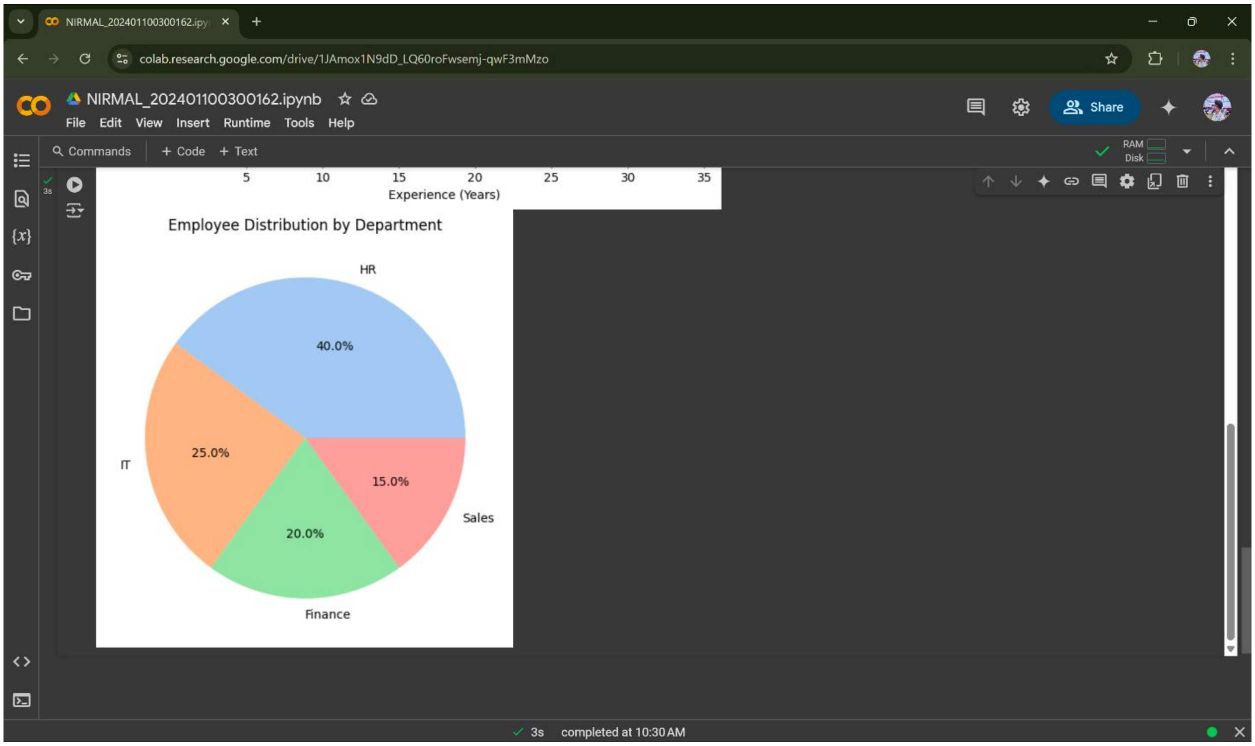
✓ 3s completed at 10:30 AM


NIRMAL_202401100300162.ipynb
colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo
File Edit View Insert Runtime Tools Help
Commands + Code + Text
Share
📌 Employee with Lowest Salary:
EmployeeID 2
Age 28
Department Finance
Experience 2
Salary 41742
Name: 1, dtype: object
📌 Outliers in Salary:
No outliers detected.
Salary Distribution
Frequency
Salary
Experience vs Salary
Department
3s completed at 10:30 AM


NIRMAL_202401100300162.ipynb
colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo
File Edit View Insert Runtime Tools Help
Commands + Code + Text
Share
Experience vs Salary
Department
Finance
HR
IT
Sales
Salary
Years of Experience
<ipython-input-2-ad9274268370>:74: FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.
sns.boxplot(x="Department", y="Salary", data=df, palette="Set2")
Salary Distribution by Department
3s completed at 10:30 AM


NIRMAL_202401100300162.ipynb
colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo
File Edit View Insert Runtime Tools Help
Commands + Code + Text
Share
<ipython-input-2-ad9274268370>:74: FutureWarning:
Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same effect.
sns.boxplot(x="Department", y="Salary", data=df, palette="Set2")
Salary Distribution by Department
Salary
140000
120000
100000
80000
60000
40000
Finance
HR
IT
Sales
Department
Salary Growth with Experience
3s completed at 10:30 AM


NIRMAL_202401100300162.ipynb
colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo
File Edit View Insert Runtime Tools Help
Commands + Code + Text
Share
Salary Growth with Experience
Salary
140000
120000
100000
80000
60000
40000
5
10
15
20
25
30
35
Experience (Years)
Employee Distribution by Department
HR
40.0%
3s completed at 10:30 AM


NIRMAL_202401100300162.ipynb
colab.research.google.com/drive/1JAmox1N9dD_LQ60roFwsemj-qwF3mMzo
File Edit View Insert Runtime Tools Help
Commands
+ Code
+ Text
Share
RAM
Disk
5 10 15 20 25 30 35
Experience (Years)
Employee Distribution by Department
HR
40.0%
IT
25.0%
15.0%
Sales
20.0%
Finance
3s
completed at 10:30 AM