1

print("Name: Prachi Karande")

print("Roll no.: TACO22134")

2

*import* pandas *as* pd

*from* sklearn.model\_selection *import* train\_test\_split

*from* sklearn.preprocessing *import* LabelEncoder, StandardScaler

*from* sklearn.linear\_model *import* LogisticRegression

*from* sklearn.metrics *import* confusion\_matrix

*import* matplotlib.pyplot *as* plt

*import* seaborn *as* sns

3

df = pd.read\_csv(r"C:\Users\Admin\Desktop\Extended\_Employee\_Performance\_and\_Productivity\_Data.csv")

4

df.head()

5

df.info()

6

df.columns

7

df["Education\_Level"].unique()

8

categorical\_columns = ["Department", "Gender", "Job\_Title", "Education\_Level"]

*#Step 2: Apply Label Encoding to each categorical column*

le = LabelEncoder()

*for* col *in* categorical\_columns:

    df[col] = le.fit\_transform(df[col].astype(str))  *# Convert to string just in case there are NaNs*

df.dropna(*inplace*=True)

df.head()

9

df['High\_Performer'] = (df['Performance\_Score'] >= 4).astype(int)

df

10

plt.figure(*figsize*=(8, 6))

sns.boxplot(*x*='High\_Performer', *y*='Monthly\_Salary', *data*=df,

*palette*='Set2')

plt.title("Monthly Salary vs. High Performers")

plt.xlabel("High Performer (1 = Yes, 0 = No)")

plt.ylabel("Monthly Salary")

plt.show()

11

X = df[['Performance\_Score','Education\_Level', 'Employee\_Satisfaction\_Score', 'Promotions']]

Y = df['High\_Performer']

X\_train, X\_test, Y\_train, Y\_test =train\_test\_split(X, Y,*test\_size*= 0.2, *random\_state*= 42)

X\_test

12

*from* sklearn.linear\_model *import* LogisticRegression

*from* sklearn.metrics *import* confusion\_matrix

lr = LogisticRegression()

lr.fit(X\_train, Y\_train)

y\_pred = lr.predict(X\_test)

print("Confusion Matrix:\n", confusion\_matrix(Y\_test, y\_pred))

13

*from* sklearn.metrics *import* accuracy\_score

accuracy = accuracy\_score(Y\_test, y\_pred)

print("Accuracy: ",accuracy)

Accuracy:  1.0

prediction = lr.predict([[5, 1, 1.72, 2]])

print(prediction)