


```
In [2]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Load the dataset (replace 'user_data.csv' with your actual file path)
df = pd.read_csv(r"C:\Users\Prithviraj Ghorpade\Downloads\User_Data.csv")

# Step 2: Preprocess the data
# Assuming the dataset has columns like 'Age', 'EstimatedSalary', and 'Purchased'
X = df[['Age', 'EstimatedSalary']] # Features
y = df['Purchased'] # Target variable (Labels)

# Step 3: Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 4: Train a Random Forest Classifier (optional, if you want to generate predictions)
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(random_state=42)
clf.fit(X_train, y_train)
y_pred = clf.predict(X_test)

# Step 5: Confusion Matrix
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()

# Step 6: Example Graph: Random Forest Algorithm (Training set)
plt.scatter(X_train['Age'], X_train['EstimatedSalary'], c=y_train, cmap='rainbow')
plt.title('Random Forest Algorithm (Training set)')
plt.xlabel('Age')
plt.ylabel('Estimated Salary')
plt.show()

# Step 7: Example Graph: Estimated Salary vs. Age
plt.scatter(df['Age'], df['EstimatedSalary'], c=df['Purchased'], cmap='rainbow')
plt.title('Estimated Salary vs. Age')
plt.xlabel('Age')
plt.ylabel('Estimated Salary')
plt.show()
```



