Task no: 2

Results...!!!

1.Customer Churn Dataset Preview:

```
Preview of Raw Dataset (first 5 rows):
   customerID  gender  SeniorCitizen Partner Dependents  tenure PhoneService  \
0  7590-VHVEG  Female              0     Yes         No       1           No
1  5575-GNVDE    Male              0      No         No      34          Yes
2  3668-QPYBK    Male              0      No         No       2          Yes
3  7795-CFOCW    Male              0      No         No      45           No
4  9237-HQITU  Female              0      No         No       2          Yes

      MultipleLines InternetService OnlineSecurity  ... DeviceProtection  \
0  No phone service             DSL             No  ...               No
1                No             DSL            Yes  ...              Yes
2                No             DSL            Yes  ...               No
3  No phone service             DSL            Yes  ...              Yes
4                No     Fiber optic             No  ...               No

  TechSupport StreamingTV StreamingMovies        Contract PaperlessBilling  \
0          No          No              No  Month-to-month              Yes
1          No          No              No        One year               No
2          No          No              No  Month-to-month              Yes
3         Yes          No              No        One year               No
4          No          No              No  Month-to-month              Yes

              PaymentMethod MonthlyCharges  TotalCharges Churn
0          Electronic check          29.85         29.85    No
1              Mailed check          56.95        1889.5    No
2              Mailed check          53.85        108.15   Yes
3  Bank transfer (automatic)          42.30       1840.75    No
4          Electronic check          70.70        151.65   Yes

[5 rows x 21 columns]
```

1.Dataset: **Telco Customer Chur**

2.Each row = **1 customer's record**

3.Columns = **21 features (customer info + services + billing)**

4.Includes **demographics** (gender, senior citizen, dependents, partner)

5.Includes **services** (phone, internet, streaming, security, support)

6.Includes **account info** (tenure, contract, payment method, charges)

7.Target column = **Churn** (Yes/No → whether customer left or stayed)

2.Customer Churn Dataset Preview:

```
Cleaning data...
Data cleaned! New shape: (7032, 20)

Training set: (5625, 19), Testing set: (1407, 19)

Feature Types:
Numeric: ['tenure', 'MonthlyCharges', 'TotalCharges']
Categorical: ['gender', 'SeniorCitizen', 'Partner', 'Dependents', 'PhoneService', 'MultipleLines', 'InternetService',
```

Total Records after Cleaning: **7032 rows × 20 columns**

Train/Test Split:

- Training set → 5625 rows × 19 features
- Testing set → 1407 rows × 19 features

Feature Types:

- Numeric: tenure, MonthlyCharges, TotalCharges
- Categorical: 16 columns (like gender, Partner, Contract, PaymentMethod, etc.)

👉 In short: You now have a clean dataset with 19 input features (3 numeric + 16 categorical) ready for model training and evaluation.

3.Hyperparameter Tuning Complete:

```
Starting Hyperparameter Tuning...
Fitting 3 folds for each of 12 candidates, totalling 36 fits

Best Parameters: {'classifier__max_depth': 10, 'classifier__min_samples_split': 2, 'classifier__n_estimators': 200}
Best CV Accuracy: 0.8018

Tuned Random Forest Results:
Final Accuracy: 0.7903
              precision    recall  f1-score   support

           0       0.83      0.89      0.86      1033
           1       0.63      0.51      0.57       374

    accuracy                           0.79      1407
   macro avg       0.73      0.70      0.71      1407
weighted avg       0.78      0.79      0.78      1407


Pipeline saved as 'telco_churn_pipeline.joblib'
```

Tried 36 model settings with cross-validation.
Best Random Forest parameters: max_depth=10, min_samples_split=2, n_estimators=200.Accuracy: **80% (CV)**, **79% (test set)**.
Predicts non-churn well (precision 0.83, recall 0.89).

Struggles more with churn cases (precision 0.63, recall 0.51).
Final model saved as **telco_churn_pipeline.joblib**.

4. Model Performance Results:

```
Training Logistic Regression...
Logistic Regression Results:
Accuracy: 0.8045
              precision    recall  f1-score   support

           0       0.85      0.89      0.87      1033
           1       0.65      0.57      0.61       374

    accuracy                           0.80      1407
   macro avg       0.75      0.73      0.74      1407
weighted avg       0.80      0.80      0.80      1407


Training Random Forest...
Random Forest Results:
Accuracy: 0.7861
              precision    recall  f1-score   support

           0       0.83      0.89      0.86      1033
           1       0.62      0.51      0.56       374

    accuracy                           0.79      1407
   macro avg       0.73      0.70      0.71      1407
weighted avg       0.78      0.79      0.78      1407
```

✅ **Logistic Regression:**

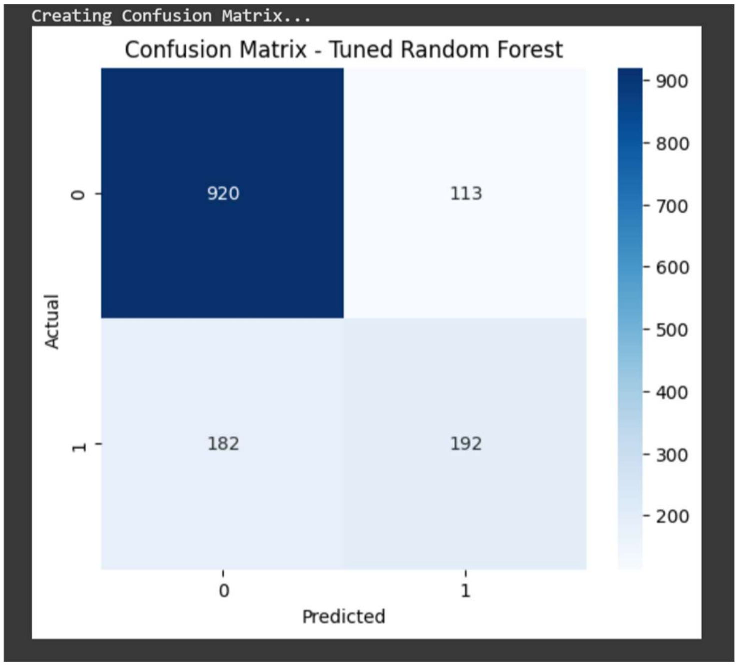
- **Accuracy:** 80.45%
- **Class 0 (Stayed):** Better predicted (85% precision)
- **Class 1 (Churned):** Harder to predict (65% precision)

🌳 **Random Forest:**

- **Accuracy:** 78.61%
- **Slightly worse** than Logistic Regression
- **Struggles more** with churned customers (62% precision)

📌 **Key Insight:**
Logistic Regression performed better for this dataset. Both models are better at predicting who will **stay** than who will **leave**.

5. Confusion Matrix:



✅ **Correct (Diagonal):**

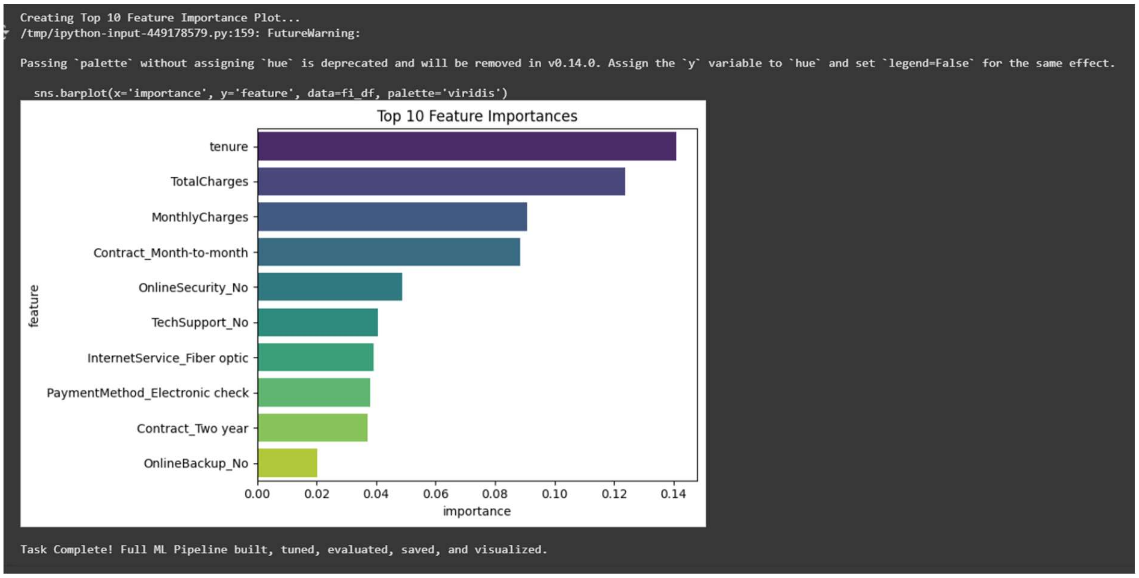
- **920**: Actually stayed → Predicted stayed
- **182**: Actually left → Predicted left

❌ **Errors (Off-Diagonal):**

- **113**: Actually stayed → Predicted left (False alarm)
- **192**: Actually left → Predicted stayed (Missed churn)

Key Takeaway:
Model is **good at identifying who stays** (920 correct), but **misses almost 200 customers** who actually churn. Needs improvement in detecting churn signals.

6.Top 10 Features Affecting Churn:



Most Important Features:

1. **Tenure** ← How long customer stayed
2. **TotalCharges** ← Total amount paid
3. **MonthlyCharges** ← Monthly payment amount

Service & Contract Features:

4. Month-to-month contract ← More likely to churn
5. No Online Security ← Higher churn risk
6. No Tech Support ← Higher churn risk
7. Fiber Internet ← Affects churn
8. Electronic Check payment ← Impacts retention
9. Two-year contract ← Less likely to churn
10. No Online Backup ← Increases churn chance

Insight:

Customer duration (tenure) and contract type are biggest churn predictors. Service features (security, support) also strongly influence retention.