Upload the Dataset

```
from google.colab import files
uploaded = files.upload()
```

```
Choose Files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to
```

Load the Dataset

```
import pandas as pd

# Replace with your actual filename
df = pd.read_csv('/content/churn_prediction (1).csv')
df.head()
```

Data Exploration

```
# Basic info
df.info()

# Descriptive statistics
df.describe()

# Preview column names
print("Columns:", df.columns.tolist())
```

Check for Missing Values and Duplicates python Copy Edit

```
# Check for missing values
print(df.isnull().sum())

# Check for duplicates
print("Duplicate Rows:", df.duplicated().sum())
```

Visualize a Few Features

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

# Set Seaborn style for better visuals
sns.set(style="whitegrid")

# Check if 'Gender' and 'Age' columns exist
if 'Gender' in df.columns:
    plt.figure(figsize=(6, 4))
    sns.countplot(data=df, x='Gender', palette='Set2')
    plt.title('Gender Distribution')
    plt.xlabel('Gender')
    plt.ylabel('Count')
    plt.show()
else:
    print("Column 'Gender' not found in DataFrame.")

if 'Age' in df.columns:
    plt.figure(figsize=(6, 4))
    sns.histplot(df['Age'], kde=True, color='skyblue', bins=30)
    plt.title('Age Distribution')
    plt.xlabel('Age')
    plt.ylabel('Frequency')
    plt.show()
else:
    print("Column 'Age' not found in DataFrame.")
```

Identify Target and Features

```
print(df.columns.tolist())

target_column = 'churn'
X = df.drop(target_column, axis=1)
y = df[target_column]
df.head()
```

Convert Categorical Columns to Numerical

```
# Identify categorical columns
cat_cols = X.select_dtypes(include='object').columns
print("Categorical Columns:", cat_cols.tolist())

# Apply label encoding temporarily (can be replaced with OneHot later)
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

for col in cat_cols:
    X[col] = le.fit_transform(X[col])
```

Convert Categorical Columns to Numerical

```
# Identify categorical columns
cat_cols = X.select_dtypes(include='object').columns
print("Categorical Columns:", cat_cols.tolist())

# Apply label encoding temporarily (can be replaced with OneHot later)
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()

for col in cat_cols:
    X[col] = le.fit_transform(X[col])
```

One-Hot Encoding

```
X = pd.get_dummies(X, drop_first=True)
```

Feature Scaling

```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

Train-Test Split

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42
)
```

Model Building

```
from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier(random_state=42)
model.fit(X_train, y_train)
```

Evaluation

```
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

y_pred = model.predict(X_test)
print("Accuracy:", accuracy_score(y_test, y_pred))
print(classification_report(y_test, y_pred))
sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d')
plt.show()
```

Make Predictions from New Input python Copy Edit

```
print(X.columns.tolist())

import pandas as pd

# Create a dictionary that includes all columns from training
new_input_dict = {
    'Age': 35,
    'Gender_Male': 1,
    'Gender_Female': 0,
    'Plan_Basic': 0,
    'Plan_Premium': 1,
    'MonthlyCharges': 5000,
    # ... include all other one-hot encoded or numeric features, set missing to 0
}

# Convert to DataFrame
new_input_df = pd.DataFrame([new_input_dict])

# Reindex to match training column order
new_input_df = new_input_df.reindex(columns=X.columns, fill_value=0)

# Scale
new_input_scaled = scaler.transform(new_input_df)

# Predict
prediction = model.predict(new_input_scaled)
print("Prediction:", "Churn" if prediction[0] == 1 else "Not Churn")
```

Convert to DataFrame and Encode (for prediction input)

```
input_dict = {
    'Age': [35],
    'Gender': ['Male'],
    'Plan': ['Basic'],
    # Add more fields as per your original dataset
}

input_df = pd.DataFrame(input_dict)

# Convert categorical variables
for col in input_df.select_dtypes(include='object'):
    input_df[col] = le.fit_transform(input_df[col])

# Align columns
input_df = pd.get_dummies(input_df)
input_df = input_df.reindex(columns=X.columns, fill_value=0)

input_scaled = scaler.transform(input_df)
```

Predict the Final Grade

```
final_prediction = model.predict(input_scaled)
print("Final Prediction:", final_prediction)
```

Deployment - Building an Interactive App

```
# Simulate form input in Colab
user_input = pd.DataFrame({
    'Age': [30],
    'Gender': ['Male'],
    'Plan': ['Premium'],
    # Add other features...
})

# Encode, align, scale
for col in user_input.select_dtypes(include='object'):
    user_input[col] = le.fit_transform(user_input[col])\


user_input = pd.get_dummies(user_input)
user_input = user_input.reindex(columns=X.columns, fill_value=0)

user_input_scaled = scaler.transform(user_input)
prediction = model.predict(user_input_scaled)

print("Prediction:", "Churn" if prediction[0] == 1 else "Not Churn")
```

Create a Prediction Function

```
def preprocess_input(input_data, scaler, encoder, base_columns):
    """
    Preprocess input data: encode, one-hot, scale, and align columns.

    Args:
    - input_data (pd.DataFrame): Raw input data.
    - scaler (StandardScaler): Fitted scaler.
    - encoder (LabelEncoder): Fitted label encoder for categorical vars.
    - base_columns (list): List of original X.columns after one-hot.

    Returns:
    - np.array: Scaled and aligned feature vector.
    """
    data = input_data.copy()

    for col in data.select_dtypes(include='object').columns:
        data[col] = encoder.fit_transform(data[col])

    data = pd.get_dummies(data)
    data = data.reindex(columns=base_columns, fill_value=0)
    data_scaled = scaler.transform(data)

    return data_scaled

def predict_churn(input_dict, model, scaler, encoder, base_columns):
    """
    Make churn prediction from raw input dictionary.

    Args:
    - input_dict (dict): User inputs as key-value pairs.
    - model (trained model): Trained classifier.
    - scaler (StandardScaler): Trained scaler.
    - encoder (LabelEncoder): Trained label encoder.
    - base_columns (list): Reference for column alignment.

    Returns:
    - str: Prediction result.
    """
    input_df = pd.DataFrame([input_dict])
    processed = preprocess_input(input_df, scaler, encoder, base_columns)
    prediction = model.predict(processed)[0]
    return "Churn" if prediction == 1 else "Not Churn"
```

create the gradio interface

```
!pip install -q gradio
import gradio as gr

def predict_churn(age, gender, plan, monthly_charges):
    # Create input DataFrame
    input_dict = {
        'Age': [age],
        'Gender': [gender],
        'Plan': [plan],
        'MonthlyCharges': [monthly_charges]
    }
    input_df = pd.DataFrame(input_dict)

    # Encode
    for col in input_df.select_dtypes(include='object'):
        input_df[col] = le.fit_transform(input_df[col])

    # One-hot encoding (if needed)
    input_df = pd.get_dummies(input_df)
    input_df = input_df.reindex(columns=column_names, fill_value=0)

    # Scale
    input_scaled = scaler.transform(input_df)

    # Predict
    prediction = model.predict(input_scaled)[0]
    return "Churn" if prediction == 1 else "Not Churn"



iface = gr.Interface(
    fn=predict_churn,
    inputs=[
        gr.Number(label="Age"),
        gr.Dropdown(choices=["Male", "Female"], label="Gender"),
        gr.Dropdown(choices=["Basic", "Premium", "Gold"], label="Plan"),
        gr.Number(label="Monthly Charges")
    ],
    outputs="text",
    title="Churn Prediction App",
    description="Enter details to predict if a customer will churn."
)

iface.launch()
```

```
It looks like you are running Gradio on a hosted a Jupyter notebook. For the Gradio app to work, sharing must be enabled. Automatically

Colab notebook detected. To show errors in colab notebook, set debug=True in launch()
* Running on public URL: https://d54080e709ec2e3aa2.gradio.live

This share link expires in 1 week. For free permanent hosting and GPU upgrades, run `gradio deploy` from the terminal in the working dir
```