**🔹 1.Paste the Code**

Copy and paste this code into your cardio\_model.py file:

# Step 1: Setup and Import Libraries

# Data manipulation and analysis

import pandas as pd

import numpy as np

# Visualization

import matplotlib.pyplot as plt

import seaborn as sns

# Machine Learning tools

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

from sklearn.preprocessing import StandardScaler

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import classification\_report, confusion\_matrix, roc\_auc\_score

**🔹 2. Save the File**

* Press Ctrl + S or go to **File → Save**.

**🔹 3. Run the File**

* Open the **terminal** in VS Code:  
  Press Ctrl + ~ or go to **View → Terminal**
* Type:
* python cardio\_model.py

**🔹 If You See No Errors:**

✅ Success!

**⚠️ If You See an Error Like ModuleNotFoundError: No module named 'pandas':**

It means you haven’t installed the required libraries. Install them one by one:

pip install pandas numpy matplotlib seaborn scikit-learn

Then run your script again:

python cardio\_model.py

Perfect! Now that your imports are set and working, you're ready to start **Step 1: Exploratory Data Analysis (EDA)**.

**✅ Step 1: Exploratory Data Analysis (EDA)**

**🔹 What we’ll do:**

1. **Load the dataset**
2. **View basic info**
3. **Check for missing values**
4. **Understand distributions and relationships**
5. **Visualize key insights**

**🧠 Paste the following into your cardio\_model.py below your import lines:**

# Step 1: Exploratory Data Analysis (EDA)

# Load the dataset

df = pd.read\_csv("cardio\_train.csv", sep=';') # change path if needed

# Basic structure

print("First 5 rows:")

print(df.head())

print("\nInfo:")

print(df.info())

print("\nSummary statistics:")

print(df.describe())

# Check for missing values

print("\nMissing values:")

print(df.isnull().sum())

# Distribution of target variable

sns.countplot(data=df, x='cardio')

plt.title("Distribution of Cardiovascular Disease")

plt.xlabel("Cardiovascular Disease (0 = No, 1 = Yes)")

plt.ylabel("Count")

plt.show()

# Correlation heatmap

plt.figure(figsize=(12, 8))

sns.heatmap(df.corr(), annot=True, fmt='.2f', cmap='coolwarm')

plt.title("Feature Correlation Heatmap")

plt.show()

**🔔 Note:**

* Make sure your **cardio\_train.csv** file is in the **same folder** as your script. If not, update the path.
* If you haven't downloaded the dataset yet, [get it from Kaggle](https://www.kaggle.com/datasets/mahmudulhaqueshawon/cardiovascular-disease) and extract the CSV.

Let me know once you've run this — I’ll help you analyze the output or move on to Step 2 (Feature Engineering).