**✅ Paso 1: Librerías necesarias**

Para trabajar con archivos .xls (formato antiguo de Excel), necesitas instalar estas librerías:

bash

CopiarEditar

pip install pandas xlrd matplotlib

**¿Qué hace cada una?**

* pandas: para leer, procesar y analizar los datos.
* xlrd: para leer archivos .xls.
* matplotlib: para graficar.

## ✅ Paso 2: Código base para leer .xls y graficar gastos vs abonos

python

CopiarEditar

import pandas as pd

import matplotlib.pyplot as plt

# Cargar archivo .xls

ruta\_archivo = "reportCollection.xls" # Cambia por tu ruta local

df = pd.read\_excel(ruta\_archivo)

# Verifica qué columnas tiene tu archivo

print("Columnas:", df.columns)

print(df.head())

# Supongamos que las columnas se llaman: 'Fecha', 'Descripción', 'Monto', 'Tipo' (cargo/abono)

# Si los nombres son distintos, ajústalos aquí:

col\_fecha = 'Fecha'

col\_monto = 'Monto'

col\_tipo = 'Tipo'

# Filtramos gastos y abonos

df[col\_tipo] = df[col\_tipo].str.lower() # para evitar problemas de mayúsculas

gastos = df[df[col\_tipo].str.contains("cargo|gasto|debito", na=False)]

abonos = df[df[col\_tipo].str.contains("abono|deposito|credito", na=False)]

# Suma total

total\_gastos = gastos[col\_monto].sum()

total\_abonos = abonos[col\_monto].sum()

# Gráfico de torta

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

plt.pie([total\_gastos, total\_abonos],

labels=["Gastos", "Abonos"],

autopct="%1.1f%%",

colors=["red", "green"])

plt.title("Distribución de Gastos vs Abonos")

plt.show()

**✅ Paso 3: Siguientes pasos posibles**

Una vez esto funcione:

* Podemos agregar **clasificación de categorías** (salud, arriendo, etc).
* Crear un **dashboard interactivo** si más adelante lo haces con PyQt o Tkinter.
* O guardar los reportes automáticamente como imágenes o PDF.

### 🧪 Código con clasificación automática

python

CopiarEditar

import pandas as pd

import matplotlib.pyplot as plt

def limpiar\_monto(valor):

try:

return float(str(valor).replace('$', '').replace('.', '').replace(',', '.').strip())

except:

return 0.0

# Cargar archivo .xls

archivo = "reportCollection.xls"

df = pd.read\_excel(archivo)

# Normalizar nombres de columnas si es necesario

df.columns = df.columns.str.strip().str.capitalize()

# Limpiar montos

df["Cargo"] = df["Cargo"].apply(limpiar\_monto)

df["Abono"] = df["Abono"].apply(limpiar\_monto)

# Crear columna "Monto" y "Tipo"

df["Monto"] = df["Cargo"] - df["Abono"]

df["Tipo"] = df["Monto"].apply(lambda x: "Gasto" if x > 0 else "Abono")

# Clasificación automática

def clasificar(descripcion):

desc = str(descripcion).lower()

if any(palabra in desc for palabra in ["aguas", "enel", "luz", "gas"]):

return "Servicios Básicos"

elif "arriendo" in desc:

return "Arriendo"

elif "farmacia" in desc or "salcobrand" in desc or "cruz verde" in desc:

return "Salud"

elif "supermercado" in desc or "jumbo" in desc or "lider" in desc:

return "Compras"

elif "transf" in desc or "depósito" in desc or "abono" in desc:

return "Ingreso"

else:

return "Otros"

df["Categoría"] = df["Descripción"].apply(clasificar)

# Filtrar solo gastos para el gráfico

df\_gastos = df[df["Tipo"] == "Gasto"]

# Agrupar y graficar

categorias = df\_gastos.groupby("Categoría")["Monto"].sum()

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

categorias.plot(kind='bar', color='orange')

plt.title("Gastos por Categoría")

plt.ylabel("Monto ($)")

plt.xticks(rotation=45)

plt.grid(True)

plt.tight\_layout()

plt.show()