import pandas as pd

import os

from openpyxl import load\_workbook

# File path and name

file\_path = r"C:\Users\USER\Desktop\OFFICE WORK"

file\_name = "TIER TWO 2024.xlsx"

full\_path = os.path.join(file\_path, file\_name)

# Output file name

output\_file = "TIER TWO 2024 - CONSOLIDATED.xlsx"

output\_path = os.path.join(file\_path, output\_file)

try:

# Load the workbook to get all sheet names

workbook = load\_workbook(full\_path)

sheet\_names = workbook.sheetnames

workbook.close()

print(f"Found {len(sheet\_names)} sheets in the workbook:")

for i, sheet in enumerate(sheet\_names, 1):

print(f"{i}. {sheet}")

# List to store all dataframes

all\_dataframes = []

# Read each sheet and combine them

for sheet\_name in sheet\_names:

try:

# Read the sheet with headers in row 2 (index 1)

df = pd.read\_excel(full\_path, sheet\_name=sheet\_name, header=1)

# Add a column to identify which sheet the data came from

df['Source\_Sheet'] = sheet\_name

# Only add non-empty dataframes

if not df.empty and len(df.columns) > 1: # More than just the Source\_Sheet column

all\_dataframes.append(df)

print(f"✓ Successfully read sheet '{sheet\_name}' - {len(df)} rows")

else:

print(f"⚠ Skipped empty sheet '{sheet\_name}'")

except Exception as e:

print(f"✗ Error reading sheet '{sheet\_name}': {str(e)}")

if all\_dataframes:

# Combine all dataframes

consolidated\_df = pd.concat(all\_dataframes, ignore\_index=True, sort=False)

print(f"\n📊 Consolidation Summary:")

print(f"Total sheets processed: {len(all\_dataframes)}")

print(f"Total rows in consolidated data: {len(consolidated\_df)}")

print(f"Total columns: {len(consolidated\_df.columns)}")

# Create a new workbook with the consolidated data

with pd.ExcelWriter(output\_path, engine='openpyxl') as writer:

# Write consolidated data to 'Consolidated' sheet

consolidated\_df.to\_excel(writer, sheet\_name='Consolidated', index=False)

# Optional: Create a summary sheet

summary\_data = []

for sheet\_name in sheet\_names:

if sheet\_name in consolidated\_df['Source\_Sheet'].values:

row\_count = len(consolidated\_df[consolidated\_df['Source\_Sheet'] == sheet\_name])

summary\_data.append({'Sheet\_Name': sheet\_name, 'Row\_Count': row\_count})

summary\_df = pd.DataFrame(summary\_data)

summary\_df.to\_excel(writer, sheet\_name='Summary', index=False)

print(f"\n✅ Successfully created consolidated workbook: '{output\_file}'")

print(f"📁 Location: {output\_path}")

print(f"\nThe new workbook contains:")

print(f"- 'Consolidated' sheet with all data")

print(f"- 'Summary' sheet showing row counts per original sheet")

print(f"- 'Source\_Sheet' column to identify data origin")

else:

print("❌ No data found to consolidate. Please check your file and sheet contents.")

except FileNotFoundError:

print(f"❌ File not found: {full\_path}")

print("Please check:")

print("1. The file path is correct")

print("2. The file name is correct (including .xlsx extension)")

print("3. The file exists in the specified location")

except Exception as e:

print(f"❌ An error occurred: {str(e)}")

print("\n" + "="\*50)

print("Script execution completed!")