# SSNIT Audit Tool - Final Robust Version

import os

import pandas as pd

from IPython.display import display

# 1. SETUP - UPDATE THESE!

main\_folder = r"C:\Users\USER\Desktop\STALLION TIGER COMPANY LTD"

output\_file = r"C:\Users\USER\Desktop\SSNIT\_Audit\_Report.xlsx"

# 2. CONFIGURATION - Update these to match your column names

COLUMN\_NAMES = {

'ssnit\_col': 'SSNIT', # Change to your SSNIT column name

'name\_col': 'Name', # Change if different

'salary\_col': 'Salary' # Change if different

}

# 3. FUNCTION TO LOAD DATA

def load\_all\_ssnit\_data(main\_folder):

all\_data = []

for year in range(2014, 2025):

year\_folder = os.path.join(main\_folder, str(year))

for month in ['jan', 'feb', 'mar', 'apr', 'may', 'jun',

'jul', 'aug', 'sep', 'oct', 'nov', 'dec']:

file\_path = os.path.join(year\_folder, f"{month}{year}.xlsx")

if os.path.exists(file\_path):

try:

# Read Excel forcing SSNIT as string

df = pd.read\_excel(file\_path, header=0, dtype={COLUMN\_NAMES['ssnit\_col']: str})

# Standardize column names

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

col\_map = {

'ssnit': COLUMN\_NAMES['ssnit\_col'].lower(),

'name': COLUMN\_NAMES['name\_col'].lower(),

'salary': COLUMN\_NAMES['salary\_col'].lower()

}

# Rename columns

df.rename(columns={

col\_map['ssnit']: 'SSNIT',

col\_map['name']: 'Name',

col\_map['salary']: 'Salary'

}, inplace=True)

# Clean SSNIT data (remove spaces, force string)

df['SSNIT'] = df['SSNIT'].astype(str).str.strip()

df['Source'] = f"{month.capitalize()} {year}"

all\_data.append(df)

except Exception as e:

print(f"Error reading {file\_path}: {str(e)}")

return pd.concat(all\_data, ignore\_index=True) if all\_data else pd.DataFrame()

# 4. PROCESS DATA

def process\_ssnit\_data(df):

if df.empty:

return pd.DataFrame()

# Ensure required columns exist

required\_cols = ['SSNIT', 'Name', 'Salary', 'Source']

missing\_cols = [col for col in required\_cols if col not in df.columns]

if missing\_cols:

raise KeyError(f"Missing columns: {missing\_cols}. Please check COLUMN\_NAMES configuration.")

# Final type conversion for safety

df['SSNIT'] = df['SSNIT'].astype(str)

# Sort and group

df.sort\_values(by='SSNIT', inplace=True)

blank\_row = pd.DataFrame({col: '' for col in df.columns}, index=[0])

grouped\_df = pd.DataFrame()

for ssnit, group in df.groupby('SSNIT'):

grouped\_df = pd.concat([grouped\_df, group, blank\_row], ignore\_index=True)

return grouped\_df

# 5. RUN THE AUDIT

print("Starting SSNIT audit...")

combined\_data = load\_all\_ssnit\_data(main\_folder)

if combined\_data.empty:

print("No data found! Please check:")

print(f"1. Folder path: {main\_folder}")

print(f"2. Column names configuration: {COLUMN\_NAMES}")

else:

try:

final\_report = process\_ssnit\_data(combined\_data)

final\_report.to\_excel(output\_file, index=False)

print(f"Success! Report saved to: {output\_file}")

print("\nPreview:")

display(final\_report.head(15))

# Additional analysis

print("\nSummary Statistics:")

print(f"Total SSNITs found: {final\_report['SSNIT'].nunique()}")

print(f"Total entries processed: {len(final\_report)}")

except Exception as e:

print(f"Error: {str(e)}")

print("Process completed.")