import pandas as pd

import random

from faker import Faker

from datetime import datetime, timedelta

# Initialize Faker

fake = Faker()

random.seed(42)

Faker.seed(42)

# Number of employees

num\_employees = 400

# Sample data pools

genders = ['Male', 'Female', 'Other']

positions = ['Software Engineer', 'Data Scientist', 'System Administrator', 'Product Manager', 'UX Designer', 'Business Analyst']

levels = ['Junior', 'Mid', 'Senior', 'Lead']

workloads = ['40h/week', '20h/week', '30h/week']

# ISCO-08 detailed unit groups relevant to IT

isco\_detailed = {

"1330": "ICT Services Managers",

"2152": "Electronics Engineers",

"2511": "Systems Analysts",

"2512": "Software Developers",

"2513": "Web and Multimedia Developers",

"2514": "Applications Programmers",

"2519": "Software & App Developers Not Elsewhere Classified",

"2521": "Database Designers and Administrators",

"2522": "Systems Administrators",

"2523": "Computer Network Professionals",

"2529": "ICT Professionals Not Elsewhere Classified",

"3511": "ICT Operations Technicians",

"3512": "ICT User Support Technicians"

}

# Function to generate contract period and renewal status

def generate\_contract\_period\_and\_status():

start\_date = fake.date\_between(start\_date='-3y', end\_date='today')

months\_duration = random.randint(6, 36)

end\_date = start\_date + timedelta(days=30 \* months\_duration)

start\_str = start\_date.strftime('%d-%m-%Y')

end\_str = end\_date.strftime('%d-%m-%Y')

period\_str = f"{start\_str} to {end\_str}"

today = datetime.today().date()

if end\_date.date() > today:

renewal\_status = random.choice(['Active', 'Pending', 'Renewed'])

else:

renewal\_status = random.choice(['Expired', 'Renewed', 'Not Renewed'])

return period\_str, renewal\_status

# Generate data

data = []

for i in range(1, num\_employees + 1):

employee\_id = f"E{i:04d}"

contract\_id = f"C{i:04d}"

gender = random.choice(genders)

full\_name = fake.name\_male() if gender == 'Male' else fake.name\_female() if gender == 'Female' else fake.name()

isco\_code, isco\_title = random.choice(list(isco\_detailed.items()))

position = random.choice(positions)

level = random.choice(levels)

workload = random.choice(workloads)

period\_date, renewal\_status = generate\_contract\_period\_and\_status()

data.append([

employee\_id,

contract\_id,

gender,

full\_name,

f"{isco\_code} - {isco\_title}",

position,

level,

workload,

period\_date,

renewal\_status

])

# Create DataFrame and save to CSV

columns = ['Employee ID', 'Contract ID', 'Gender', 'Full name', 'ISCO Group',

'Position', 'Level', 'Workload', 'Period Date', 'Renewal Status']

contract\_table = pd.DataFrame(data, columns=columns)

contract\_table.to\_csv("Contract\_Table.csv", index=False)

print("✅ Contract\_Table.csv has been created successfully.")