Assignment Task:

1. Data Generation:

```python
import csv
import random

cities = [
    "Ahmedabad", "Surat", "Gandhinagar", "Vadodara", "Rajkot", "Bhavnagar", "Anand", "Junagadh", "Nadiad", "Mehsana",
    "Navsari", "Vapi", "Amreli", "Palanpur", "Surendranagar"]
colleges = [
    "IIT Gandhinagar", "Nirma University", "Sardar Vallabhbhai Patel Institute of Technology",
    "MS University", "Dhirubhai Ambani Institute of Information and Communication Technology",
    "Gujarat University", "L.D. College of Engineering", "CEPT University", "Ahmedabad University", "Parul University"]
programs = ["Data Science", "Robotics", "AI", "Electric Vehicles"]
lead_sources = ["Instagram", "LinkedIn", "College Collaboration", "Google Form", "Mass-Mailing", "Whatsapp"]

def generate_row(lead_id):
    location = random.choice(cities)
    college = random.choice(colleges)
    year_of_study = random.choice(["1st", "2nd", "3rd", "4th"])
    program_interest = random.choice(programs)
    lead_source = random.choice(lead_sources)
    return [lead_id, location, college, year_of_study, program_interest, lead_source]

file_name = "Intern-assignment_data.csv"

with open(file_name, mode='w', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(["Lead ID", "Location", "College", "Year of Study", "Program Interest", "Lead Source"])
    num_rows = 13000
    for i in range(1, num_rows + 1):
        lead_id = f"LD{i}"
        row = generate_row(lead_id)
        writer.writerow(row)

print(f"CSV file '{file_name}' with {num_rows} rows has been generated.")
```

Output: 13,000 rows data generated and saved in csv file.

2. Demographic Analysis:

Question: Analyze the data to identify trends in lead sourcing based on **location**, **college**, and **year of study**. Highlight any significant patterns or anomalies.

Top 3 **location**:

1. Nadiad
2. Amreli
3. Surendranagar

Top 3 **college**:

1. Gujarat University
2. IIT Gandhinagar
3. CEPT University

Based on **above top 3 location and college** identify **top year of study** and **top college (From above top 3 college)** corresponding **lead source**:

Lead source	Top year of study	Top college
College Collaboration	4th	IIT Gandhinagar
Google Form	2nd	Gujarat University
Instagram	2nd	CEPT University
LinkedIn	4th	CEPT University
Mass-mailing	2nd	CEPT University
Whatsapp	1st	IIT Gandhinagar

3. Program Analysis:

Question: Compare the interest levels for various e-learning programs.
Provide a recommendation on which demographics should be targeted more aggressively for each Program.

Interest for various e-learning programs:

Program	Interest
Robotics	3379
Electric Vehicles	3260
Data Science	3244
AI	3117

Recommendation demographics for each program:

Program	College	Location	Year of study
Robotics	CEPT University IIT Gandhinagar L.D college	Mehsana Surendranagar Vapi	4th
Electric Vehicles	CEPT University Gujarat University MS University	Bhavnagar Navsari Rajkot	3rd
Data Science	DAIICT IIT Gandhinagar MS University	Gandhinagar Vadodara Palanpur	2nd
AI	Gujarat University Nirma University Parul University	Amreli Nadiad Surat	1st

4. Projections and Predictions:

Question: Based on the generated data, make a data-driven projection about lead conversion rates and suggest a budget allocation strategy for marketing and targeting.

Conversion rate is decided for lead source.

Return On Investment (ROI) = (Revenue from conversion – Marketing Cost) / Marketing Cost

Suggest a budget allocation strategy for marketing and targeting:

If ROI is high then select lead source which is low conversion rate.
If ROI is low then select lead source which is high conversion rate.

5. Data Preprocessing:

Question: Document the steps taken to clean and preprocess the data for analysis.

1. Load Dataset
2. Handle Missing Values
3. Handle Inconsistent Data
4. Remove Duplicates
5. Validate Datatype

6. Presentation of Finding:

Data Analysis performs in Excel:
Link: https://github.com/jay-bambhaniya/Tech-Analogy-Assignment/blob/main/Intern-assignment_data.xlsx