Dashboard Insights and Findings

Assignment Task

1. Data Generation

I had created a dataset with 15,000 rows, including the following details:

- Columns: Locations, Colleges, Year of Study, Program Interests, and Lead Sources.

- Distribution: Varied distribution of values for realistic insights.

---

2. Demographic Analysis

Trends in Lead Sourcing

- **Locations**: Ahmedabad contributed the highest number of leads (3,354), followed by Mumbai (2,125) and Delhi (1,789). Jaipur had the least leads (619).

- **Colleges**: The top contributors were:

  o Indian Institute of Management Ahmedabad (IIM-A): 1,142 leads.

  o Nirma University: 1,111 leads.

- **Years of Study**:

  o 2nd-year students contributed the most leads (39.73%).

  o 1st-year students contributed the least (9.69%).

Patterns and Anomalies

- Ahmedabad's dominance suggests it as a key market for targeting leads.

- The low contribution from 1st-year students might indicate a lack of interest or reach within this group.

---

3. Program Analysis

Interest Levels

- **Most Popular Programs**:

  o Environmental Science: 3,128 leads.

  o Data Science: 2,198 leads.

- **Least Popular Program**:

  o Electric Vehicles: 1,210 leads.

Dashboard Insights and Findings

Recommendations

- Focus marketing efforts on Environmental Science and Data Science, especially in high-performing locations like Ahmedabad.

- Develop strategies to promote Electric Vehicles in regions with fewer leads but potential growth, such as Indore or Jaipur.

---

4. Projections and Predictions

Lead Conversion Rates

Assuming a conversion rate of 20%:

- Environmental Science: Expected conversions = 3,128 × 20% = 626 leads.

- Data Science: Expected conversions = 2,198 × 20% = 440 leads.

Budget Allocation Strategy

Allocate the marketing budget based on:

1. **Top Locations**: Ahmedabad, Mumbai, and Delhi.

2. **Top Programs**: Environmental Science and Data Science.

3. **Effective Channels**: LinkedIn and College Collaborations.

---

5. Data Preprocessing

Programming Language: Python.

Libraries Used: pandas, numpy, and random.

Steps Taken:

1. **Dataset Generation**:

   o Used Python libraries to generate a dataset of 15,000 rows.

   o The dataset includes columns for Lead ID, Location, College/University, Year of Study, Program Interest, and Lead Source.

2. **No Missing Data**:

   o The dataset was generated programmatically, ensuring no missing values.

3. **Standardized Data**:

   o The script defines consistent column names and formats, avoiding issues with inconsistent data.

4. **Distributions**:

   o **Locations**: Defined probabilities for each city, e.g., Ahmedabad (22%), Mumbai (14%), and so on.

   o **Years of Study**: Applied probabilities like 10% for 1st year, 40% for 2nd year, etc.

   o **Programs and Lead Sources**: Used predefined distributions for realistic and varied data.

5. **Realistic Data**:

   o Used numpy's random.choice with defined probabilities for controlled randomness.

   o Varied colleges based on the location to reflect realistic data patterns.

6. **File Export**:

   o Saved the generated dataset as lead_dataset.csv for use in Power BI.

Outcome:

- The dataset was clean and ready for analysis without requiring further preprocessing.

---

6. Presentation of Findings

Your Power BI dashboard effectively:

- Displays KPIs such as total leads, colleges, locations, lead sources and program interests.

- Provides dynamic filtering through slicers for college, lead source, and program interest.

- Includes a refresh button, implemented using bookmarks, to reset all visuals and revert back to the original dashboard without any filters, providing a clean view of the data.

- Highlights trends with interactive visualizations:

   o Bar charts for lead distribution by location and program interest.

   o Pie charts for year of study and lead sources.

   o Area chart for Top 5 Colleges by leads.