Group Assignment – DTVC, Dr. Philippe Blaettchen

Objective

You will explore **how data collection can support decision-making** for a new or existing small business. This assignment will dive into web scraping and APIs to collect and analyze real-world data. The focus is on applying these skills to an industry of your choice.

Scenario

Imagine you are consulting for a small business or launching a new venture. Your goal is to **understand the market, competition, or customer preferences**. Data will be your primary tool for uncovering actionable insights.

Instructions

In your groups:

1. Define Your Business Context:
    - Select an industry or business idea (e.g., sustainable clothing, artisanal coffee shops, local tourism, or fitness services).
    - Clearly define the main question(s) your analysis will aim to answer (e.g., optimal pricing, competitive positioning, customer segmentation).
2. Research Planning:
    - Identify potential data sources (e.g., competitor websites, product review platforms, public directories, or relevant APIs).
    - Sketch a research plan detailing:
        - Key analyses you will perform.
        - Specific data points you aim to collect.
        - The relevance of these data points to your business question(s).
3. Data Collection Strategy:
    - Develop a strategy to gather data from identified sources using tools such as:
        - Web scraping with BeautifulSoup or Selenium, or more advanced scraping tools.
        - APIs for accessing structured data.
        - A mix of both, emphasizing scraping.
    - Prioritize sources based on relevance and ease of collection.
4. Pipeline Development:
    - Build a pipeline to:
        - Collect raw data.
        - Clean and preprocess the data.
        - Engineer relevant features.

    - Your pipeline should end with one or more clean DataFrames ready for analysis.
    - Generate a lexicon for the DataFrame(s), including one column each for feature names, descriptions, and units.
5. Analysis and Insights:
    - Use Python to perform:
        - Summary statistics.
        - Visualizations.
        - Preliminary analyses that address your business question(s).
    - Highlight key findings and potential implications for the business.
6. Presentation:
    - Record a 10–12-minute video presentation, summarizing:
        - Your business context, main question(s) you were aiming to answer, and research process.
        - The data and insights collected.
        - Recommendations for the business based on the findings. Clearly highlight how your data and analysis support answering your main question(s).

Deliverables

By the assignment deadline indicated on Moodle, you are required to complete the following:

1. Code Submission:
    - Submit all Python code used for scraping, cleaning, and analyzing data.
    - Include a README file if necessary for clarity.
    - Ensure your code is thoroughly commented. It should be executable and understandable by a non-expert with coding experience.
2. Video Presentation:
    - Upload a video on a cloud platform (e.g., private YouTube video) summarizing your process and insights.
    - Use slides or visual aids as necessary.
    - Make sure to provide a link to your video as part of your submission materials on Moodle.

Assessment Criteria

Your submission will be evaluated against five criteria:

1. Use of Data Collection Techniques (20%)
   Effective and Appropriate application of scraping and API skills.
2. Quality of Research Plan and Execution (20%)
   Relevance and thoroughness of the research plan and data collection strategy.
3. Data Pipeline Quality (20%)
   Clean, organized, and functional pipelines.

4. Insightfulness of Analysis (20%)
   Relevance and creativity of insights and recommendations.
5. Presentation Quality (20%)
   Clarity, professionalism, and engagement of the video presentation.