Initially, I loaded all the provided datasets into Google Colab and developed three separate scripts to analyze them: one using pure Python, another using Polars, and the third using Pandas. For each script, I performed descriptive statistical analysis on the respective dataset to ensure consistency across tools. Once I was satisfied with the outputs, I converted each notebook into a standalone `.py` script, making them runnable from the command line as per the project requirements.

During this process, I encountered several errors and challenges, which I carefully documented below to help others who might face similar issues in the future. Additionally, I utilized AI resources such as ChatGPT and Google Gemini to troubleshoot and refine my code, ensuring the scripts were accurate, efficient, and aligned with best practices.

Error Log & Troubleshooting Documentation

General Setup & File Access Issues

1. **Google Colab Upload Failure**
   - **Error**: `RangeError: Maximum call stack size exceeded`
   - **Cause**: `files.upload()` failed in Google Colab for a ~5MB file due to internal Colab limitations.
   - **Fix**: Mounted Google Drive instead and accessed the file directly from there.
2. **Missing Column During Grouping**
   - **Error**: `KeyError: 'Page Admin Top Country'`
   - **Cause**: Attempted grouping on a non-existent column.
   - **Fix**: Used `df.columns` to inspect column names and updated the grouping logic.

Pure Python Script Issues (`pure_python_stats.py`)

3. **Unterminated String**
   - **Error**: `SyntaxError: unterminated string literal`

     Cause: Line:

     ```
     with open(filename, newline=', encoding='utf-8') as csvfile:
     ```

     was incorrectly written (`newline=','` instead of `newline=''`).

   - **Fix**: Corrected to:

     ```
     with open(filename, newline='', encoding='utf-8') as csvfile:
     ```

4. **NameError: filename is not defined**
   - **Cause**: Variable `filename` was used without being defined.
   - **Fix**: Replaced with `DATA_PATH` or defined the variable before use.

5. **Missing Column in GROUP_KEYS**
    - **Error**: Script fails silently or misbehaves if `GROUP_KEYS = ['Page Id', 'Ad Id']` don't match CSV column names.
    - **Fix**: Ensured exact column names using `reader.fieldnames` or `df.columns`.

Errors encountered with Pandas Script (`pandas_stats.py`)

6. **No major Pandas syntax errors** were encountered in this script, but user needed:
    - Clarification on how to:
        - Use `describe()`, `value_counts()`, `nunique()`
        - Handle grouping and nested aggregation (optional)
    - Request to convert `.ipynb` notebook to a `.py` CLI-compatible file
7. **CLI Conversion Steps**
    - **Need**: Add block to handle `sys.argv` input file and optional grouping

    Change:

    ```
    import sys
    if len(sys.argv) > 1:
        filename = sys.argv[1]
    ```

Errors encountered with Polars Script (`polars_stats.py`)

8. **Wrong Use of `.sort()` with Extra Positional Arg**
    - **Error**: `Expr.sort() takes 1 positional argument but 2 positional arguments (and 1 keyword-only argument) were given`

      Cause: Used `sort(col, descending=True)` instead of `sort(by=col, descending=True)`

      Fix: Corrected to:

      ```
      .sort(by="counts", descending=True)
      ```

9. **Missing "counts" Column After value_counts()**
    - **Error**: `ColumnNotFoundError: "counts" not found`

      Cause: Polars sometimes auto-generates count column with non-standard name.

      Fix:

        - Dynamically infer the count column:

          ```
          count_col = [c for c in vc.columns if c != col][0]
          ```

10. **Empty Value Counts Resulting in IndexError**
    - **Error**: `IndexError: list index out of range`

      Cause: Attempted to access `[0]` from an empty list after value_counts.

      Fix: Added safeguard:

      ```
      if vc.shape[0] == 0: continue
      ```

11. **Column Inference for Top Categorical Values**
    - **Fix**: Added dynamic column selection to extract top most frequent values:

      ```
      count_col = [c for c in vc.columns if c != col][0]
      ```

12. **Dataset Reload Failure**
    - **Error**: `FileNotFoundError` when reloading after session reset
    - **Fix**: User reuploaded datasets for continued analysis

Notebook to CLI Conversion Issues

13. **`.ipynb` to `.py` Conversion**
    - **Need**: Convert interactive notebooks to runnable scripts
    - **Fix**:
      - Strip all notebook-specific cells and I/O
      - Add `if __name__ == "__main__"` block with `sys.argv` input
      - Save as `.py` and test from terminal