**Capstone Project Proposal (Data Analysis Pathway)**

**Project Title:** *Lifestyle Factors in Relation to Home Ownership*

**Proposed By:***Lauren Sapp*

**GitHub Repository:**[*https://github.com/laurenmsapp1/lifestyle\_home\_ownership*](https://github.com/laurenmsapp1/lifestyle_home_ownership)

**Project Overview**

Briefly describe the overall purpose of your project. What are you trying to analyze, explain, or uncover? Why does this matter? Write 2–3 sentences that summarize the big picture.

**Data Sources**

Dataset #1:

* **Dataset Name: Regression Dataset for Household Income Analysis**
* **Source:** Kaggle
* **Relevant fields**: Age, Occupation
* **Link:** <https://www.kaggle.com/datasets/stealthtechnologies/regression-dataset-for-household-income-analysis>

Dataset #2:

* **Dataset Name: Lifestyle and Sleep Patterns**
* **Source**: Kaggle
* **Relevant fields**: Age, Occupation
* **Link**: <https://www.kaggle.com/datasets/minahilfatima12328/lifestyle-and-sleep-patterns>

*(You must include at least two datasets from different sources. Additional optional datasets may be listed here as well.)*

**Research Objectives**

State your main questions and objectives. You should include:

* **Primary Question(s):** the key relationships or patterns you want to test.
* **Secondary/Exploratory Questions:** additional comparisons or trends you want to explore.

Explain briefly why these objectives are valuable or interesting.

**Data Preparation Approach**

Describe how you will prepare the data for analysis. Be specific:

* How you will align or merge datasets (keys, years, geography).
* How you will address missing values (replacement, dropping rows, etc.).
* How you will address outliers or extreme values.
* What new variables or indices you may create.
* How you will structure your relational database (at minimum two tables from different sources, joined through a common key).

**Current Status**

Summarize what you have already completed. Examples:

* Data acquired and inspected.
* Early cleaning or standardization steps.
* Prototype functions or exploratory plots.
* Evidence that the datasets can be successfully merged.

**Deliverables (Remaining Work)**

List the specific things you still need to complete for your project. Break them into **required** tasks (must be done to meet project requirements) and **stretch** tasks (optional, if time allows).

* **Required Tasks (must be completed):***(Example: finish cleaning literacy dataset; implement 3 Python functions; build SQLite schema and load tables; produce 3 required visualizations; write README and data dictionary.)*
* **Stretch Goals (optional, if time allows):***(Example: add ACS demographics to analysis; build interactive dashboard; run regression model.)*

**Project Timeline**

Outline the sequence of major steps or milestones leading up to the deadline. You may structure this as weeks, phases, or checkpoints. Examples:

* Phase 1: Acquire and clean data, set up database.
* Phase 2: Exploratory analysis, build functions, first draft visuals.
* Phase 3: Deeper analysis, refine visuals, draft report.
* Phase 4: Finalize deliverables, polish repo, record presentation.

Be as specific as possible about what you plan to complete at each step.

**Additional Considerations**

Note any assumptions, limitations, or risks you anticipate. Examples:

* Geographic regions or variables you will exclude (and why).
* Additional datasets you may add if time allows.
* Possible challenges (e.g., very large files, inconsistent variable names, data sparsity).