Capstone Project Proposal (Data Analysis Pathway)[The-Productivity-of-Remote-vs-in-office-Workers]

Project Title: *Which Vegetables Grow Best in Louisville, Kentucky*

Proposed By: *Tiffanie White*

GitHub Repository:
https://github.com/TiffanieWhite/Which-Vegetables-Grow-Best-In-Louisville-Kentucky-.git

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.

I decided to research and see what vegetables grow best in my growing zone in Kentucky. I tried to grow a few vegetables last year via container gardening and I only had very little success with some cherry tomatoes. I am going to analyze whether I possibly planted too soon or too late or whether maybe I didn't provide enough shade etc. It really matters to me because food is very expensive and I have noticed that certain foods just don't taste the same. It's like they're not homegrown, but engineered. I will be using all the information I find for my household to grow flowers and vegetables for our home and maybe to be able to sell a few.

Data Sources

List the datasets you will use. For each, include:

DATA SET# 1

Dataset name: Vegetables Dataset

Source: Kaggle

Relevant fields:

Link: https://www.kaggle.com/datasets/rudraprasadbhuyan/vegetables-dataset

DATA SET #2

Dataset name: Fruits and Vegetables Prices Datasets

Source: Kaggle

Relevant fields :

Link: https://www.kaggle.com/datasets/everydaycodings/produce-prices-dataset

(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):** Which vegetables grow the best in Kentucky?
>
> **Secondary/Exploratory Questions:** Growing/Production vs Grocery store production/price

Explain briefly why these objectives are valuable or interesting.

I decided to change my project to researching which vegetables grow best in Kentucky. My reasoning is I come from a crop farming background (my father's side and when I was much younger) and last year I tried my hand at patio/container gardening, but I was very much unsuccessful. I had a few tomato plants that produced probably about 20 cherry tomatoes and my watermelons stopped growing and producing which in turn made them ant food. This year I am researching and retrying in remembrance of my uncle who recently passed (to carry on the tradition) and because the food prices for my family of 4, of which 2 are special needs who only eat certain vegetables, has become astronomically high (especially since I have become unemployed). Not only is my goal to produce food for my home, but I am also hoping to have a little fruit/vegetable stand to get a little passive income.

Data Preparation Approach

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

> How you will align or merge datasets (keys, years, geography). I will merge my datasets regarding the type of vegetable and homegrown pricing (seed, soil, containers)

How you will address missing values (replacement, dropping rows, etc.).

I will address the missing values by dropping rows or researching for additional information.

How you will address outliers or extreme values.

I will address outliers or extreme values by restricting to a set loc.

What new variables or indices you may create.

I can create a new table with combinations of side by side price valuations.

How you will structure your relational database (at minimum two tables from different sources, joined through a common key).

I will structure my relational database joined by the name of the vegetable.

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).