Final Project | Cohort 7 | Stream 3

Please go to the following link (https://github.com/ruhidm/Palette_Cohort_7) and access crop yield by Rural Municipality (RM) and RM shapefiles.
The goal of the final project is to make sure you are ready for a data scientist and analyst positions. To succeed in the interviews, you will be challenged with your communication and especially presentation skills besides your natural analytical and data-driven skills. You need to provide 2 things to pass the final project:

1. The project in **presentation slides (50%)** (PowerPoint or Google Slides). You will present your findings to the class.
   Here is a general outline for presenting a data science project in a slide deck:

   Introduction: Start with an attention-grabbing opening, such as a question or a compelling statistic, to introduce the problem being solved. Provide context for the project, and give a brief overview of the key findings.

   Problem Statement: Clearly define the problem and explain why it is important. Highlight the objective of the project and the questions being answered.

   Data Collection and Preprocessing: Discuss the sources of data (refer to Government of Saskatchewan) used in the project and any preprocessing steps taken to clean or transform the data. Provide examples of the data and visualizations to help the audience understand the data.

   Exploratory Data Analysis (EDA): Show the key findings from the EDA, including any patterns, relationships, or anomalies found in the data. Use visualizations and graphs to support the key findings.

   Methodology: Explain the methods and algorithms used in the project, including any feature engineering, splitting data for train, or model selection processes. Models are required to use:
   - ➔ Unsupervised Machine Learning:
     - K-Means Clustering
     - Spectral Clustering

Results: Present the results of the analysis, including any key insights, findings, or visualizations. Highlight the strengths and weaknesses of the results.

Conclusion: Summarize the key takeaways from the project, and discuss the implications and potential applications of the results. Provide recommendations for future work.

*** DO NOT SCREENSHOT YOUR CODE AND PASTE ON SLIDES**
*** You are more than welcome to use multiple crops.**

2. The project contains your **coding in a Jupyter Notebook (50%)** file.

Objective: A brief explanation of the problem being solved, the objective of the project, and the questions being answered.

Extract, Transform, Load (ETL): The code for collecting and preprocessing the data, including any data cleaning, transformation, or manipulation steps. This section should also include any relevant explanations or comments to help the reader understand the code.

Exploratory Data Analysis (EDA): The code and visualizations used to explore and understand the data, including any descriptive statistics, plots, and correlation analysis. This section should provide insights into the data and highlight any interesting patterns or relationships.

Methodology: The code for implementing the data science methodology, including any feature engineering, model selection, or hyperparameter tuning processes. This section should also include any relevant explanations or comments to help the reader understand the code.

Results: The code for generating the results of the analysis, including any key insights, findings, or visualizations. This section should provide a summary of the results, and highlight the strengths and limitations of the analysis.

Conclusion: A summary of the key takeaways from the project, and a discussion of the implications and potential applications of the results.

Future Work: An outline of any future work that could be done to extend the project, or address limitations and challenges encountered.