ANALYZING THE FOOD AND AGRICULTURE ORGANIZATION CORPORATE DATASET USING CLUSTERING AND FITTING TECHNIQUES

NAME:ASHIKA MOHAN MUNGATH STUDENT ID:22063061

GitHub link: https://github.com/ASHIKAMOHAN/ADS-2.git

Kaggle Dataset link:https://www.kaggle.com/code/hungnd11/exploring-the-fao-dataset/input

ABSTRACT

- The aim of this study is to analyze the FAO dataset with a focus on geographical and temporal patterns in agricultural and food-related metrics.
- The study employs clustering techniques, including KMeans and hierarchical clustering, to identify spatial groupings in the dataset.
- Additionally, it explores temporal trends by fitting polynomial functions to time-series data for specific geographic areas.
- The overarching goal is to gain insights into the distribution of agricultural attributes and their evolution over time, providing a comprehensive understanding of regional and temporal dynamics within the FAO dataset.

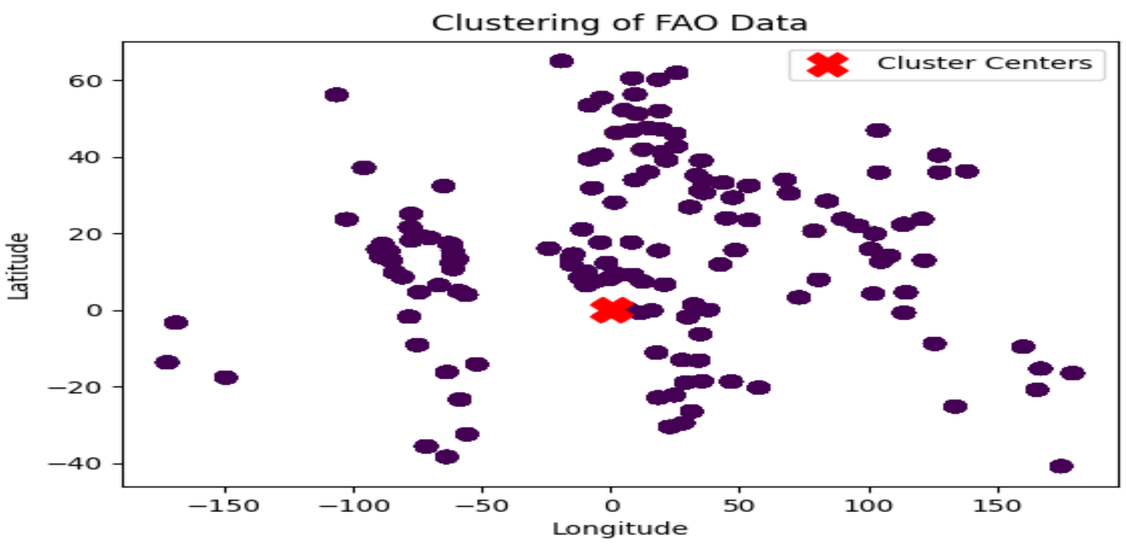
INTRODUCTION

- This study seeks to explore the intricacies of global agricultural trends using the extensive Food and Agriculture Organization (FAO) dataset.
- Our central objective is to employ clustering and polynomial fitting methods to unveil patterns in both space and time.
- As we navigate through the code and visualizations, our aim is to reveal compelling stories embedded within the data, establishing connections between each step to construct a comprehensive narrative of the global agricultural landscape.

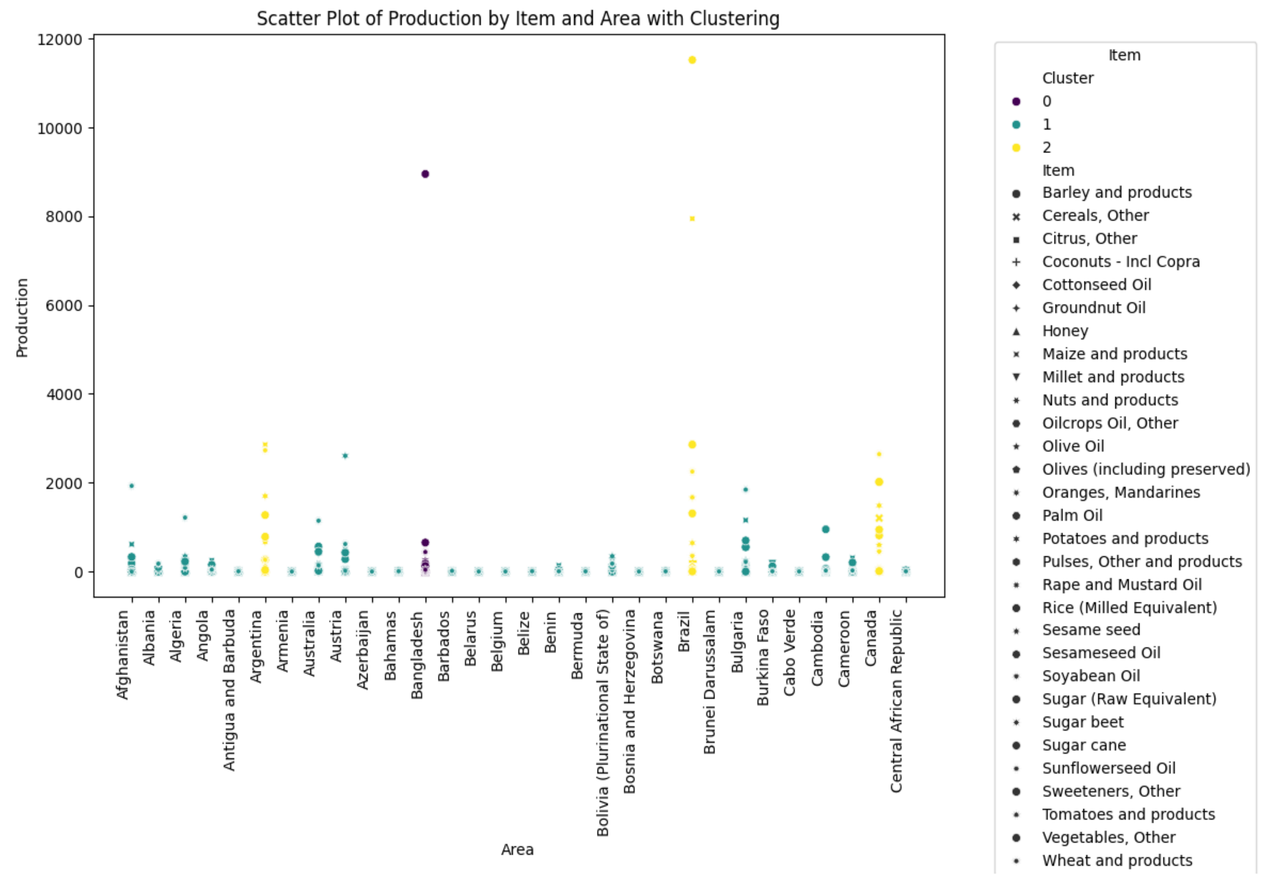
RESULT ANALYSIS

Goal 1: Clustering of a chosen dataset

- The initial step involves clustering geographical data using K-Means.
- The scatter plot showcasing these clusters not only provides a visual representation but also lays the foundation for understanding regional similarities.
- The red 'X' markers pinpoint the centroid locations, offering a snapshot of pivotal agricultural regions.
- The spatial and fleeting examples inside the FAO dataset give a more extensive hold of geological elements as well as those connected to rural, and food capacities.

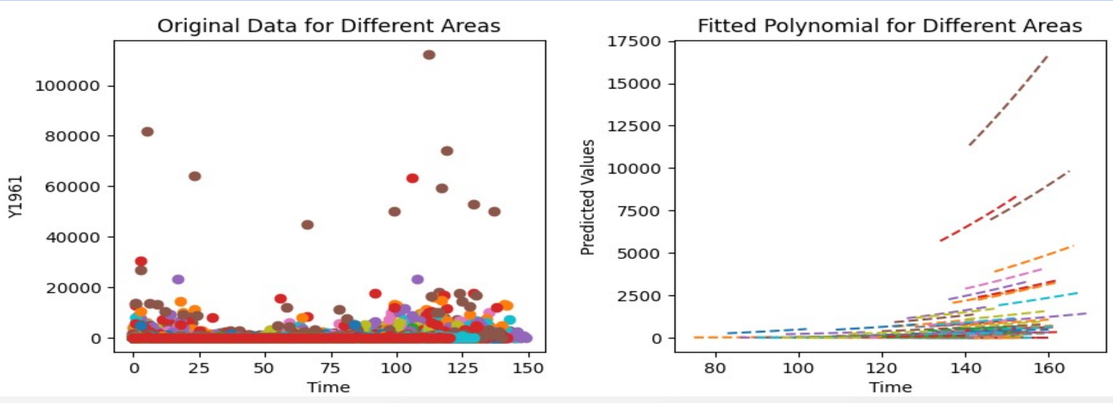


- The plot below shows production groups by item and area, with particular markers for each item and color-coded clusters, giving experiences into production designs in the chosen top 10 item and area.
- From this graph an areas overall production can be properly measured.



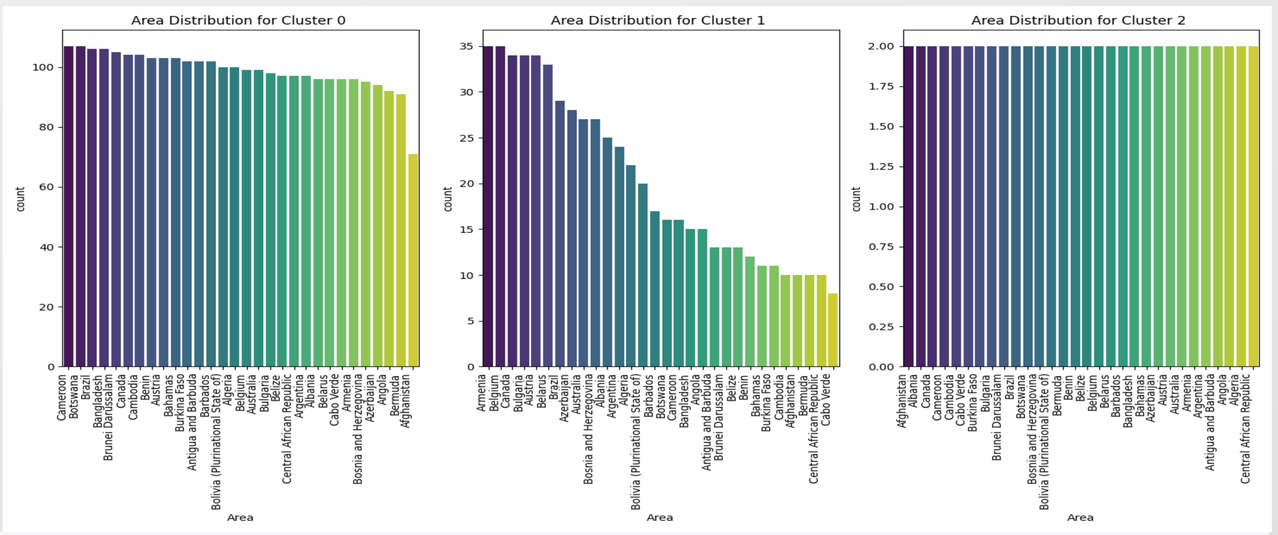
Goal 2: Fitting of a model

- One subplot displays the raw data values for each region while another features their corresponding fitted polynomial curves.
- Shifting from spatial to temporal analysis, we explore time-series data, revealing agricultural trends.
- Fitted polynomial curves elegantly depict metric evolution, each line telling a unique regional story.
- Dashed lines hint at future trajectories, aiding trend anticipation.

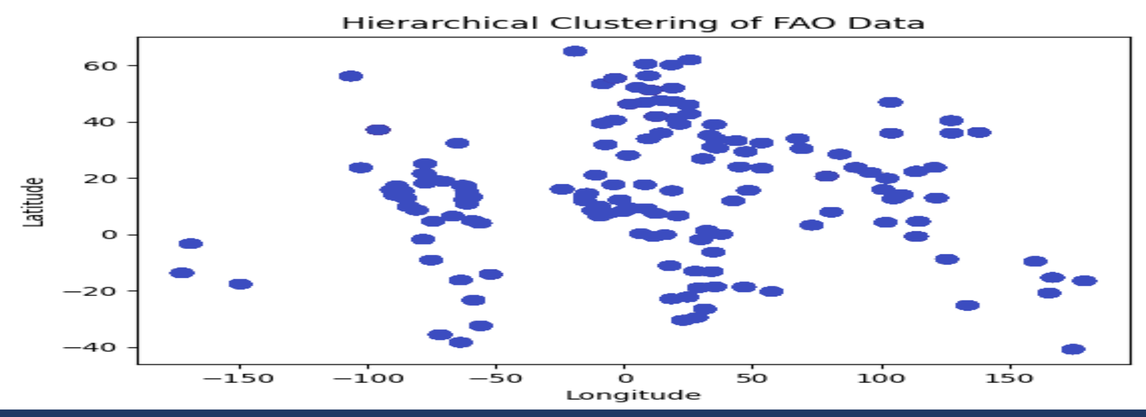


Goal 3: Clustering of Area

- Our narrative shifts to a comprehensive approach involving feature pre-processing and clustering through a pipeline.
- The count plots unravel the distribution of areas within each cluster, offering nuanced insights into how countries group together based on diverse agricultural attributes.
- This step establishes connections between spatial and temporal aspects, highlighting the interplay between geographic patterns and temporal trends.



- The final segment of our exploration introduces hierarchical clustering.
- The scatter plot resulting from Agglomerative Clustering provides an alternative perspective on relationships between countries.
- This insight bridges the spatial and temporal narratives, presenting a holistic understanding of global agricultural trends.



CONCLUSION

•In this respect, clustering and fitting analyses reveal a detailed picture of the complex mechanisms linking geography, economy, and food production.

•The resulting clusters show patterns in countries' economic evolution and food production, providing useful information for policymakers, researchers, and stakeholders.

• By including latitude and longitude, one is incorporating the spatial dimension into analysis emphasizing geographic aspects of agriculture and economic variables.

•This research adds to a better understanding of global forces, expectedly nudging initiatives towards sustainable agriculture and equitable economic growth.

•This diversity makes this dataset highly versatile and supports cross-disciplinary use, underscoring the necessity to stay mindful of diverse aspects in global analyses.