Group: Seble Alemu, Shaveen Saadee, Atty Ehui, Catherine Choi, and Alexander Davis
Final Project Reflection Paper
Course: DS 2002
Prof. Williamson
TA: Albert Huang

Exploring the Relationship Between Food Insecurity and Climate Change

We had the opportunity to explore the relationship between food insecurity and climate change by analyzing the Global Hunger Index (GHI) and climatological disaster data. Our project aimed to uncover how the frequency and severity of natural disasters, such as droughts, correlate with levels of food insecurity at the national level. This experience provided us with valuable insights into data science workflows, from selecting datasets and implementing ETL pipelines to analyzing data and leveraging cloud storage. While we encountered several challenges, the project offered a rich learning experience that improved both our technical and teamwork skills.

Data Selection

Finding reliable datasets was a straightforward and collaborative process. We quickly identified the GHI dataset, which provided peer-reviewed and well-documented information on global hunger levels. Additionally, climatological disaster and water and waste statistics datasets were readily available on reputable platforms. The accessibility and relevance of these sources made data selection relatively smooth and allowed us to focus on integrating the datasets effectively. However, aligning the datasets was not without its difficulties. While each dataset offered valuable insights, they often differed in scope and structure. For example, the GHI dataset reported annual data, while some climatological datasets aggregated data over longer periods. This misalignment required us to standardize timelines and ensure compatibility between variables, which added complexity to the data selection phase.

Data Cleaning

Data cleaning turned out to be one of the most tedious and time-consuming parts of our project. Each dataset came with its own set of challenges that required meticulous attention to detail. Fixing column headers was a key task, as many were either vague or inconsistent, making it difficult to understand the data at a glance. Standardizing these headers to be more specific and intuitive took longer than expected but was essential for making the data easier to interpret and analyze. Another major challenge was dealing with empty rows and ensuring that only meaningful observations were retained. This step required careful filtering to avoid accidentally removing useful data. Converting observations from strings to numeric values was also a critical and time-intensive process. Ensuring that data types were consistently floats rather than mixed types added an extra layer of complexity, especially when working with large datasets. Additionally, aligning data across different timeframes required interpolation or aggregation to create consistency between datasets. While these tasks were repetitive and often frustrating, they were vital for creating a clean, reliable foundation for analysis. By the end, we came to

appreciate how important these steps were in ensuring that our results were accurate and meaningful.

ETL Setup and Implementation

The ETL (Extract, Transform, Load) process was a key learning opportunity for our group. Designing a conceptual pipeline was relatively easy, but implementing it in Python highlighted several technical challenges. The extraction phase required us to work with datasets in various formats, such as CSV and JSON, which demanded flexibility in our approach. The transformation phase was particularly demanding. Cleaning the data, normalizing variables, and merging datasets involved a great deal of trial and error. We used Python libraries like Pandas and NumPy to automate repetitive tasks and streamline the process, but the learning curve was steep for some team members. Loading the cleaned data into a MongoDB database added another layer of complexity, especially as we worked to ensure that the database was both efficient and scalable. Despite these challenges, building a functional ETL pipeline was a rewarding experience. By collaborating and troubleshooting together, we were able to create a system that reliably prepared the data for analysis while accommodating potential updates.

Data Analysis

The data analysis phase was both exciting and challenging. Using Python, we conducted exploratory data analysis (EDA) to uncover trends and relationships between food insecurity and climatological disasters. Creating visualizations such as time-series graphs, scatter plots, and heatmaps helped us identify patterns and communicate our findings effectively. One major challenge during this phase was accounting for confounding variables that could influence food insecurity, such as economic inequality or political instability. While we were able to identify correlations between natural disasters and hunger levels, isolating causal relationships proved more complex. This highlighted the need for advanced statistical techniques, which were beyond the scope of this project but remain an area for future growth.

Cloud Storage

Setting up cloud storage on Google Cloud was a challenging but insightful experience for our group. One of the biggest hurdles was establishing the proper authorization credentials to store the data. Initially, we struggled to configure these credentials correctly, which caused delays and added complexity to the process. Sharing and securely storing these credentials among team members was another challenge; there were instances where improper handling led to unauthorized access or delays in collaboration, which highlighted the need for better security practices. Additionally, configuring permissions to allow seamless collaboration without compromising data security required careful thought and trial and error. We had to balance enabling team members to access the data efficiently while minimizing the risk of potential data breaches. Although these challenges were frustrating at times, they taught us the importance of proper credential management and access control in real-world data projects. By the end of the

project, we were able to set up a secure and accessible system that supported reproducible analysis and team collaboration, though this process showed us how critical attention to security and organization is in cloud-based workflows.

Technical Challenges
One of the key lessons we learned was the importance of meticulous data cleaning and preparation. While these tasks were often tedious, they were essential for producing meaningful analysis. We also gained a better understanding of how to design and implement an ETL pipeline, ensuring that our workflows were both efficient and adaptable.

Team Coordination
This project reinforced the value of effective teamwork and communication. While we generally worked well together, there were moments of misalignment, particularly during the data cleaning phase, when task responsibilities weren't always clear. Regular check-ins and clear task delegation helped us overcome these challenges and move forward as a cohesive group. We also learned to appreciate the diverse strengths and perspectives each team member brought to the table.

Conclusion
Through this project, we developed a range of technical skills, including Python programming for data cleaning, visualization, and database management. We also gained experience with cloud storage and learned the importance of documenting workflows for collaboration and reproducibility. Looking ahead, we hope to build on these skills by learning more advanced statistical and machine learning techniques, which would enable us to conduct deeper and more nuanced analyses. Additionally, improving our proficiency with cloud computing and database design will help us tackle larger, more dynamic projects in the future.

This project was a valuable learning experience that allowed us to apply data science techniques to a pressing global issue. By working through challenges in data selection, ETL implementation, analysis, and cloud storage, we gained a deeper appreciation for the complexity and impact of data-driven decision-making. As a group of students, we are proud of what we accomplished and excited to carry the lessons and skills from this project into future endeavors. Overall, it's been such a meaningful class, and we're grateful for the opportunity to learn so much in such a practical, engaging way.

Thank You, Prof. Williamson,
Seble Alemu, Shaveen Saadee, Atty Ehui, Catherine Choi, and Alexander Davis