Jan 29, 2025 | Weekly-Sprint

Things to do:
- Team members share the datasets they've found and come together to select the datasets they will use for the course project.
- Email or meet with TA or Dr. Mallavarapu **(Last date for approval: Feb 25, 2025)**
- Write a brief description about that dataset and provide a link to the files.
  Give credits: provide link of the source where was the data taken from and names of people or organization (if available) that collected the data
  Context: Provide context as to why was it collected as specified by the source
  Who: Generally, when data is collected a part of the context also states, who was it collected from (people) or about the process that generated the data. Most of the time, this will be about the people who contributed the data.
- *Wrangling*: Clearly summarize your current understanding of the data-wrangling, and clean the data. Describe steps.
- *Data Merging*: Summarize data-merging issues (steps and challenges), both specifically targeting your data sets.
- Make a submission. Just one student is enough. Make sure to list all team members' names along with their contributions to receive credit. These contributions should also be present in Github.

6% of the course grade. **Deadline: Due by 11:59 pm on March 3, 2025 (online).**

Things that were resolved in the meeting:
- Group members shared the datasets that they've found so far.

Things that are pending:
- Team must choose the final datasets to work on and receive approval before the approval deadline
- Write a brief description about the dataset. Provide context as to why the data was collected, who collected it and who it was collected from and the data collection process.
- Merge and clean datasets and summarize the process.
- Make a submission before part 1 deadline

Feb 5, 2025 | Weekly-Sprint

Agenda for group meeting
- Group members further presented and discussed possible datasets, including speculating on what the end group project should look like.
- Some members still need to present their group project

Things to do

- Nidhi and Khush will upload their proposed datasets to the GitHub README, and everyone will upload a document in the drive proposing a basic outline and workflow for the possible project that can be undertaken along with it.

Feb 10, 2025 | Weekly-Sprint

Agenda for the meeting
- Khush and Nidhi to present their datasets and project pitches.
- The team members will finalize the dataset to be presented to Dr. Mallavarapu later today

Update:
- The team voted on Khush's medicine oriented dataset to be pitched to the professor.
- Malik's dataset about county demographics and food access by county remain as backups at any point in the process
- We do plan to pitch both of the datasets to the professor though, to get a better insight into what direction we should be looking in

To-do
- Nidhi needs to present her dataset.

Feb 10, 2025 | Data Selection Review

- It was concluded that it is not possible to perform a meaningful merge between the biological datasets, due to a difference in the meaning of keys and columns. It was better suited for classification and predictive modelling.
- The county demographics and food access datasets were approved, but were very clean. Null values would have to be introduced by the professor to get an opportunity to clean the data.
- We are currently temporarily finalizing the county and food datasets but will be on the lookout for something better if we find it.

To-do
- Get the final confirmation for the dataset from Dr. Aditi by email, whether it is changed or not.
- If not, then write an email to get the dataset with the null values introduced into the dataset.

Feb 11, 2025 | Weekly-Sprint

- Wellness day !
- Not many updates. Still on the lookout for an alternative datasets

Feb 14, 2025 | Weekly-Update

We finalised the dataset proposed by Malik over text. Steve Uwa will be emailing Dr. Mallavarapu to introduce the null values.

Feb 15, 2025 | Weekly-Update

Dr. Mallavarapu responded with the updated dataset. We are ready to begin exploring the new dataset and make relevant data cleaning decisions.

Feb 15, 2025 | Weekly-Sprint

Agenda for team meeting:
- Discussed what was to be done with each column.
- Explored the different types of null values and the new dataset
- Concluded that the dataset had singular columns with many kinds of null values.
- Divided tasks for data cleaning

To-do
- Khush and Steve will work on the county dataset and Nidhi will work on the food dataset with help from Malik and Malik will work on finally merging the datasets together, and Nidhi will review.

Feb 20, 2025 | Weekly-Sprint

Agenda for the meeting:
- Review the data cleaning procedures after what has been learned in lecture
- Discussed best way to merge the two datasets

Conclusions and to-dos
- We decided on taking a "primary key" approach to merging the datasets. The key is a unique combination of the concatenated county name and state names abbreviated(in that order). The full names of states in the food dataset would need to be abbreviated using some external library (we plan on using 'us')
- Nidhi will create the primary keys
- Steve and Khush will clean all the numerical columns to ensure that all the data types such as float and int were consistent. (Check through df.info())

- Malik will carry out the final merging after verification of the process and will look over the final dataset

Feb 22, 2025 | Weekly-Sprint

Agenda for the meeting
- Nidhi is done with creating the merged primary key column.
- Steve is done with cleaning the county df
- Everyone provided brief explanations of why they made the choices they did. Especially the procedure for handling the numerical columns.

To-do
- Khush will work on cleaning the food df.

Updates:
- Khush finished cleaning the food dataframe but recognized some issues in the process. So he created a copy of the original food df and worked on that.
- Malik will apply the primary keys to the new modified food df and merge.

Feb 25, 2025 Feb 26, 2025 | Weekly-Update

- We realised that we did not check for duplicate values.
- Steve, Khush added the duplicate value check and Nidhi wrote the final dataset to a csv and uploaded it to GitHub.
- Everyone met and worked together on the documentation and submission requirements for HW 5b, including the data dictionaries, setting up the GitHub wiki pages, and the pdf submission.
- Discussed some future considerations for the project, including EDA / Data Viz tools and looking at final dataset columns.

Mar 3, 2025

- Got started on materials for midterm, including report and slides
- Created templates for the slides and the report
- Assigned tasks to everyone. Follow up on either Sunday or Monday

To do
- Complete the assigned tasks and attend + schedule a follow up meeting.

Mar 10, 2025

- Discuss more about EDA and what questions we'll likely ask
- More in-depth discussion about dataset

To do
- Get started on course project part 2 and attend + schedule a follow up meeting.

Mar 17, 2025

- Clarify what is expected for course project part 2 assignment
- Discuss what we want to get out of the dataset, ie., predicting food scarcity for a county
- Also a brief discussion on the subset of data we will use for EDA

To do
- Get started on course project part 2 and attend + schedule a follow up meeting.

Mar 24, 2025

- Discussed potential models that we like to use and are familiar with
- Looked over each others progress in EDA and discussed any improvements, suggestions and concerns

To do
- Finish course project part 2 and attend + schedule a follow up meeting.
- Complete analysis sketch
- Each team member complete their EDA and report results

Mar 29, 2025

- Discussed results from EDA and looked over changes from each team member
- Discussed any improvements, suggestions and concerns

To do
- Finish course project part 2 and attend + schedule a follow up meeting.
- Complete analysis sketch