Hugh Plunkett 23312173

DATE	TOPIC	DETAILS	Time Spent	Challenges faced	Solution
11/12/2024	Group meeting	Discussed about the topics, datasets and how we working style.	2 hour	-	-
17/12/2024	Group meeting	Define the problem statement regarding weather and grain	2 hour	-	-
18/12/2024	Data prepared	Found a valid dataset for weather in the USA, started to download daily data for 2022, 2023	1 hour	Finding a feature-rich dataset, relevant to our project goals	We found VisualCrossing - A free weather database aimed at data scientists
19/12/2024	Weather data downloaded	Finished downloading my weather data from visualcrossing.com, took 2 days due to limits on the free account	2 hour	-	-
20/12/2024	File combined	Combined the seperate weather files, one for each year 2020-2023, into one main weatherdata.csv file using python and pandas	1 hour	Using python to manipulate files by importing, merging and then saving them as one file	Easy using the tool Pandas
22/12/2024	Aquired full crop data	UN cereal production data from 2011 to 2022, need to change weather data to fit	4 hour	Finding a feature-rich dataset, relevant to my focus	The UN have plenty of crop production databases that are free to use
23/12/2024	Edit weather data	Removed year 2023 and worked on aquiring future years, 2018 and 2019	2 hour	Interacting with visualcrossings web api to download code	Done in browser
25/12/2024	Aquiring new weather data	Aquired years 2016 and 2017	1 hour	Interacting with visualcrossings web api to download code	Done in browser
26/12/2024	Cereal production data	Proprocessed our UN cereal data and selected United states data only, focusing on the years 2016 to 2022, got weekly crop numbers and prepared to merge with our weather data	4 hour	-	-
27/12/2024	Created weather week data	Created new weekly weather df using our weather data for 2016-2022. This will be merged with our new weekly crop numbers	3 hour	-	-
28/12/2024	Uploaded data to DB	Uploaded our 3 files to mongoDB. prepared to merge our two weekly files into one master weekly file for future processing	3 hour	Uploading to mongo db collection	Used MongoDBs Compass tool to import data

29/12/2024	Merged data in Mongo DB	Merged our weekly crop column into our weekly weather data based on year using mongo	2 hour	Using query to combine datasets so that weekly crop production is seen on weather data	Used pipeline tool within mongoDB
30/12/2024	Data Visualisations, preprocessing	Imported data into python. Graphing. Cleaning Data	4 hour	Graphing stastical details such as distributions	Used Seaborn plotting tool for python for nice graphs
31/12/2024	ML Algorithms	Implemented random forest regression and Support Vector Regression, Started evaluations	3 hour	Model performed poorly, showing signs of being overtrained	Changed the testing ratio from 0.2 to 0.3 to reduce overfitting
1/1/2025	ML Algorithms + Evals	Implemented linear regression model and furtehr evaluations on our 3 models	2 hour	-	-
2/1/2025	Project Report	Started work on the project report, finding revelant papers. 449 words in the related work and refrences section about RF regression	4 hour	Starting report in a manner which all could view and edit	Shared word document in shared folder created by Natalia
3/1/2025	Methodolgy section	Did work on methodology section in project report. 410 Words. Added refrence. CODE: Added Decision tree regression and Gradient boosting regression to python code for sake of comparisons with other group mebers.	3 hour	Added more models to project required extensive knowledge of skslearn library	Documentation is available on their website
4/1/2025	Traveling + Report work	Traveled Home, Worked on report and project journal	2 hour	-	-
5/1/2025	Final group meeting, Recording, Submi	Had final group meeting to discuss project, recorded my 5 minute discussion on my parts of the project	4 hour	-	-