MSA Practicum - Summary of Workload Distribution

Team 2
Tai Jin Yao (jtai30@gatech.edu); Koo Shi Ting Brenda (skoo37@gatech.edu);
Teo Rui Yang (rteo7@gatech.edu)

Task	Description	Team Member Contribution
Exploratory Data Analysis	We analyzed tweet records to assess relevance and content quality for weather-related topics, ensuring only pertinent data was retained. Processing the Twitter dataset, analyzing tweet frequencies over time, mapping contributions by country, and visualizing sentiment distributions to uncover trends and insights related to weather-related discussions	Jin Yao: Analyze foreign language tweets to determine their prevalence and the number of hashtags and weblink. Brenda: Analyze the distribution of tweets Rui Yang: Processing the Twitter dataset, analyzing tweet frequencies over time and visualizing sentiment distributions to uncover trends and insights related to weather-related discussions.
Data Cleaning / Transformation	Adding features (weather topic and country/region extraction). Tweets were filtered by relevance, removing non-weather tweets, foreign language entries, URLs, mentions, hashtags, and special characters. The text was tokenized and pre-processed for clarity.	Brenda & Jin Yao: Adding features and pre-processed the tweets Rui Yang - Tweets were filtered by relevance, removing non-weather tweets, foreign language entries, URLs
Methodology	Employed LDA for topic modeling, spaCy for geographic region extraction, and tested VADER, BERT/ClimateBERT for sentiment analysis. Decided to fine-tune ClimateBert for the project's objective Conducted sentiment analysis on individual tweets and created an aggregated view of the sentiments by grouping tweets based on weather topics and countries.	Rui Yang: Employed LDA for topic modeling, spaCy for geographic region extraction, and VADER, BERT/ClimateBERT for sentiment analysis. Jin Yao: Exploring the use of xlm-roberta-base-sentiment as an alternative to the VADER and TextBlob models. Performed LDA topic modeling. Brenda: Tested VADER and Textblob for sentiment analysis. Fine-tuned the Climatebert model to fit the project's objective.
Analysis, Results and Evaluation	Sentiments (positive, negative, neutral) were analyzed, with region and weather topics aggregation Used the fine-tune model to conduct sentiment analysis on individual tweets and created an aggregated view of the sentiments by grouping tweets based on weather topics and countries (final results). Evaluated the accuracy of the sentiment labels	Rui Yang: Sentiments (positive, negative, neutral) were analyzed Jin Yao: Sentiment distribution across different countries. Labeled tweets based on mean temperature and evaluated the fine-tuned model's accuracy Brenda: Developed the function to aggregate the sentiment analysis and provide the final results.
Code Compilation	Compilation of the code and functions to allow user to run it by following instructions in the README file and creating an environment with the requirements.txt file	Brenda

Midterm Report	Midpoint review of the project's progress, summarizing data filtering, model development, and preliminary findings related to market impact based on weather sentiment trends.	Each team member contributed equally to the Midterm Report.
Final Report		Each team member contributed equally to the Final Report.