

## Weekly Summary - Group 5

Emily Ye [zy367@drexel.edu](mailto:zy367@drexel.edu)

Junkai Ge [jg3944@drexel.edu](mailto:jg3944@drexel.edu)

Jerry Li [jl4533@drexel.edu](mailto:jl4533@drexel.edu)

Shengyang Dong [sd3666@drexel.edu](mailto:sd3666@drexel.edu)

Our team imported and initially analyzed the required datasets. We then performed a detailed exploratory analysis of the data, including adding new columns, examining store closures under different conditions, school and state holiday impacts, and handling missing values. We also merged the training data and store data to ensure data integrity and consistency. Based on this, the team performed correlation analysis to identify key factors affecting sales and displayed the relationships between the data through visualization.

## Member's contributions

Member/Task	Emily Ye	Junkai Ge	Jerry Li	Shengyang Dong
Project Planning	Define project objectives	Project Plan	Set Milestones	Assign Roles
Data Acquisition	N/A	Identify Data Sources	N/A	Data preprocessing
Initial Data Exploration	Data Overview	Data Quality Assessment	Data Overview	Correlation Analysis
Tool Setup	Environment Setup	N/A	Environment Setup	N/A

## Each group member's contribution

Emily Ye [zy367@drexel.edu](mailto:zy367@drexel.edu)

Define project objectives: Clearly outline the purpose and goals of the project.

Identify the key deliverables and desired outcomes. Specify the timeline and milestones.

Data Overview: Identify the sources of data and identify the structure of the data, acquire a basic understanding of the dataset

Environment Setup: List the software tools and applications needed (e.g. development environments, libraries, frameworks).

Ensure compatibility and installation of the required software.

Junkai Ge [jg3944@drexel.edu](mailto:jg3944@drexel.edu)

Project Plan: Develop a detailed project plan, including timelines, tasks, and deliverables.

Identify Data Sources: Determine where to obtain the necessary data for the project.

Data Quality Assessment: Evaluate the quality of the data, including checking for missing values, inconsistencies, and duplicates.

Jerry Li [jl4533@drexel.edu](mailto:jl4533@drexel.edu)

Set Milestones: Establish major milestones and deadlines for each phase of the project.

Data Overview: Conduct a high-level overview of the data to understand its structure and contents.

Environment Setup: Ensure all team members have the necessary tools and environments set up

Shengyang Dong [sd3666@drexel.edu](mailto:sd3666@drexel.edu)

Data Preprocessing: Fill missing values in store.csv and test.csv.

Filter out data where the store is open and has sales greater than zero.

Correlation Analysis: Carry out correlation analysis on store opening data.

## Future Work

In the coming week, we will learn and understand the time series nature of sales data, identifying seasonal patterns, trends, and anomalies that may affect our forecasting models.