

## Iteration 3: Sales Forecasting Project

Ayman Mushtaq Ahmad, Amisha Tiwari, Ronhit Neema

November 10, 2024

## Contents

<b>1</b>	<b>Project Update</b>	<b>3</b>
<b>2</b>	<b>Task Breakdown</b>	<b>3</b>
2.1	Ayman Mushtaq Ahmad . . . . .	3
2.2	Ronhit Neema . . . . .	3
2.3	Amisha Tiwari . . . . .	3
<b>3</b>	<b>Skills and Tools Used</b>	<b>3</b>
<b>4</b>	<b>Progress Review</b>	<b>4</b>
<b>5</b>	<b>Submission for This Iteration</b>	<b>4</b>
<b>6</b>	<b>Our GitHub Repository link</b>	<b>4</b>

# 1 Project Update

The primary focus for this iteration was to complete data preprocessing, including thorough data cleaning, exploratory data analysis (EDA), and initial visualization of insights. Each team member contributed to this phase, ensuring a streamlined and accurate dataset for model implementation in subsequent iterations.

## 2 Task Breakdown

### 2.1 Ayman Mushtaq Ahmad

- Task: Data Cleaning
- Description: Ayman led the data cleaning process, ensuring data quality and consistency. Key steps included handling missing values, removing outliers, and standardizing data formats as per project requirements.

### 2.2 Ronhit Neema

- Task: Exploratory Data Analysis (EDA)
- Description: Ronhit performed EDA to uncover patterns, correlations, and data distributions, providing insights that will guide feature selection and model training in future iterations.

### 2.3 Amisha Tiwari

- Task: Data Visualization and Feature Engineering
- Description: Amisha focused on creating visual representations of the cleaned data and added a new feature column to enhance model interpretability. This step provides an initial framework for understanding data trends and distributions.

## 3 Skills and Tools Used

- **Pandas and NumPy:** Used for data cleaning and processing.
- **Matplotlib:** Utilized for visualizing data distributions, correlations, and trends during the EDA and visualization stages.
- **Jupyter Notebooks:** Employed as the primary environment for EDA and data visualization.

## 4 Progress Review

With data preprocessing complete, we are on schedule to begin model selection and training in the next iteration. The cleaned and preprocessed dataset has prepared us for effective model experimentation and implementation.

## 5 Submission for This Iteration

- All tasks outlined for this iteration are complete, with a structured and clean dataset ready for modeling.
- A comprehensive report of our data preprocessing and EDA progress is submitted, including documented insights and visualizations.
- This report reflects our completion of preprocessing and readiness for modeling in the upcoming week.

## 6 Our GitHub Repository link

- We have started using GitHub, and each member is contributing to the project. We are following industry-standard practices in Git, where issues are created and assigned, and each project member works on these in respective branches. PRs are then raised and approved by other members before merging into the main branch.
- Our Project Repo on GitHub is available at: [https://github.com/ahmadaym/IDMP\\_SalesForecasting\\_Project](https://github.com/ahmadaym/IDMP_SalesForecasting_Project)