

Unified Mentor Supply Chain Demand Forecasting Project - Adimalla Nithin Siddhartha

This document summarizes the data preparation and preprocessing phases of the Supply Chain Demand Forecasting project, aimed at preparing the dataset for a neural network model.

1. Data Acquisition and Initial Inspection

The process began with loading the supply chain data from the provided CSV file. An initial inspection was performed to confirm the data was loaded correctly and to understand the basic structure and columns of the dataset.

2. Data Quality Assessment

A crucial step involved assessing the quality of the data, specifically by checking for missing values across all relevant columns.

- **Observation:** It was observed that the dataset did not contain any missing values, which simplified the preprocessing required.

3. Data Transformation

To make the data suitable for a neural network model, several transformations were applied.

- **Data Type Review:** The data types of each column were reviewed to ensure they were appropriate for numerical processing. While some columns were already in a suitable format (e.g., numerical prices), others required attention.
- **Feature Encoding:** Categorical features, such as product types, customer demographics, and location, were identified. These features were converted into a numerical format using one-hot encoding. This process creates new binary columns for each category, allowing the model to interpret these non-numerical inputs.
- **Feature Selection:** Based on the project's goal of demand forecasting, relevant numerical features and the newly encoded categorical features were selected to form the final dataset used for modeling.

4. Data Partitioning

The prepared dataset was divided into two distinct sets: a training set and a testing set.

- **Process:** The data was split with the 'Number of products sold' designated as the target variable (what we aim to forecast) and the remaining selected features as the input for the model. An 80/20

ratio was used, allocating the majority of the data for training the model and reserving a portion for evaluating its performance on unseen data.

- **Observation:** The split resulted in appropriately sized training and testing datasets for both the features and the target variable, ready for the modeling phase.

5. Target Variable Analysis

Before proceeding to model building, the distribution of the target variable ('Number of products sold') was analyzed.

- **Process:** A visualization (histogram with KDE) was created to show the frequency of different values for the number of products sold.
- **Observation:** The visualization provided insights into the spread and pattern of demand, which is valuable for understanding the forecasting problem and interpreting model results.