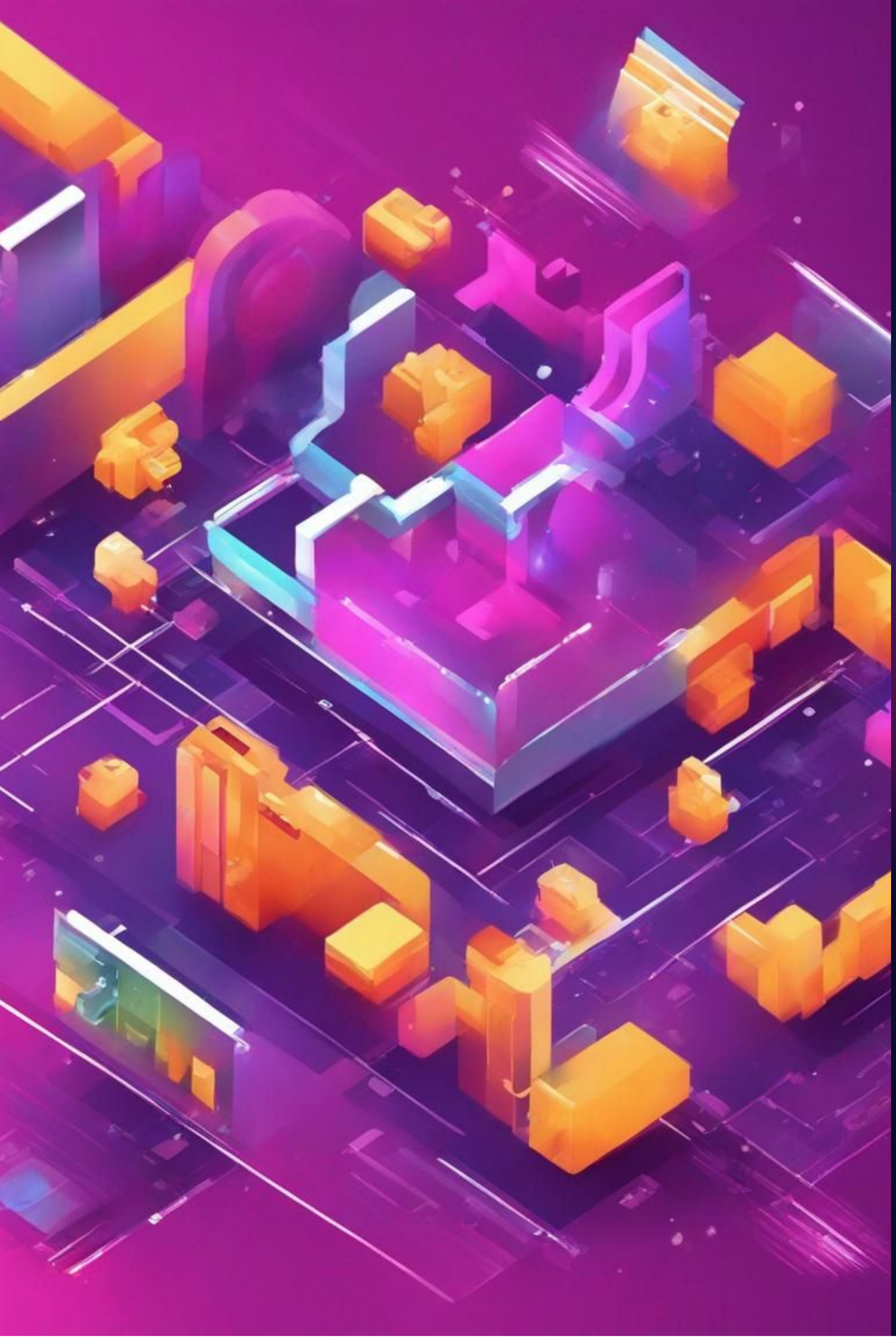


FMCG Supply Chain Optimization

The project aims to enhance the supply chain of a leading FMCG company in the instant noodles sector, focusing on improving efficiency and reducing operational costs.

by Anu Rashik

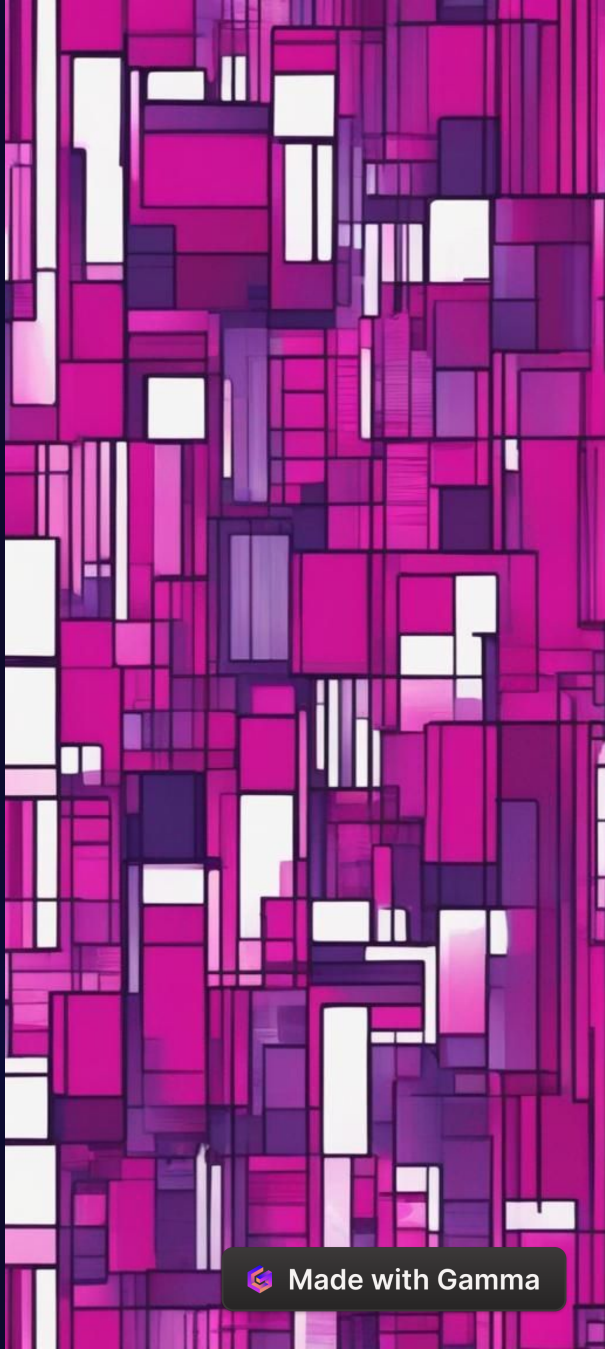
Problem Statement

1. **Mismatched Demand and Supply**

   The challenge of matching supply with fluctuating demand has led to inventory costs and losses for the company.

2. **Predictive Modeling**

   The goal is to build a predictive model that determines the optimum weight of products to be shipped from each warehouse, addressing the issue of inventory imbalances.



Project Overview

Scope: Instant Noodles

The project specifically targets the optimization of the instant noodles supply chain, a key segment for the company.

Approach: Data Analysis & ML

Comprehensive data analysis and the implementation of machine learning models are the chosen methods to realize the optimization objectives.

Benefits: Improved Efficiency

The optimization project is expected to result in improved supply chain efficiency and considerable reductions in inventory-related costs.

Data Exploration

1 **Historical Data Analysis**

An in-depth analysis of historical data is being conducted to identify patterns and trends related to supply chain dynamics.

2 **Key Feature Identification**

Efforts are underway to identify the key features that significantly impact the supply chain dynamics of instant noodles.

Data Cleaning and Preprocessing

1

Removing Unnecessary Data

The initial step involves the removal of irrelevant and unnecessary columns from the dataset.

2

Missing Values Management

Efforts have been made to handle missing values effectively to ensure the accuracy and reliability of the dataset.

3

Data Standardization

The standardization of data types is being undertaken to facilitate uniform processing and analysis.

Data Visualization

Warehouse Characteristics

Count plots visualizing various warehouse characteristics are being utilized to gain insights into the distribution network.

Product Weight Distribution

A pie chart has been created to represent the distribution of product weights organized by geographical zones.

Storage Issues Correlation

A line plot has been generated to showcase the correlation between storage issues and product weight, bringing to light important relationships.

Feature Engineering

1. ## Categorical Variable Encoding

   The enhancement process involves encoding categorical variables to make them compatible with the machine learning models.

2. ## New Feature Creation

   New features are being created to enrich the dataset and enhance model performance through the addition of relevant information.

Data Preprocessing

1 Scaling Numerical Features

Scaling numerical features is an important step in data preprocessing to ensure that all features are on a similar scale. This helps in avoiding any bias towards features with larger values and allows for better comparison and analysis.

2 Handling Remaining Outliers

Handling any remaining outliers is crucial in data preprocessing to address extreme values that could significantly impact the analysis. Outliers can distort statistical measures and affect the overall accuracy of the data.

Model Building Algorithms Used:

Linear Regression

Linear regression is a statistical modeling technique used to understand the relationship between a dependent variable and one or more independent variables. It is commonly used for predicting quantitative outcomes.

Decision Tree

Decision tree is a model building algorithm that is used for both regression and classification problems. It is a flowchart-like model in which each node represents a feature and each branch represents possible values of the feature. The model uses a 'divide and conquer' approach to recursively split the data into subsets based on the most significant features.

Random Forest

Random forest is an ensemble learning algorithm that combines multiple decision trees to improve prediction accuracy. Each decision tree in the random forest is built on a random subset of the data and the final prediction is made by aggregating the predictions of all the trees. This approach helps to reduce overfitting and improve the robustness of the model.

SVR

SVR, or Support Vector Regression, is a supervised learning algorithm that is used for regression tasks. It is an extension of Support Vector Machines (SVM) and uses a similar approach to find the best hyperplane that maximizes the margin between the data points and the predicted values. SVR is particularly useful when dealing with non-linear relationships between variables.

XGBoost

XGBoost, short for eXtreme Gradient Boosting, is a powerful machine learning algorithm that belongs to the gradient boosting family. It uses a combination of weak prediction models called decision trees to create a strong predictive model. XGBoost is known for its speed and performance, making it a popular choice for both classification and regression tasks.

Neural Network

Neural Network, also known as Artificial Neural Network (ANN), is a machine learning algorithm inspired by the structure and function of the human brain. It consists of an interconnected network of nodes (or neurons) that are organized in layers. Each neuron takes input, processes it, and passes it on to the next layer until the final output is obtained. Neural networks have proven to be effective in handling complex patterns and large amounts of data, making them widely used in various domains such as image recognition, natural language processing, and time series forecasting.

Model Evaluation

Metrics:

- Mean Squared Error (MSE)
- R-squared

Model Comparison:

Model	MSE
Linear Regression	2.679269e-22
Decision Tree	1.982250e+00
Random Forest	1.630887e+00
SVR	1.285540e+08
XGBoost	2.931428e+04
Neural Network	4.789202e+03

Best Model:

Linear Regression has the lowest Mean Squared Error: 2.679269e-22

Conclusion and Results

Achievements:

Successful implementation of the supply chain optimization model.

Key Result: Linear Regression emerged as the best-performing model.

Lowest Mean Squared Error (MSE): 2.679269e-22

Next Steps and Improvements

Future Initiatives:

- Real-time data integration for up-to-date insights.
- Exploration of advanced machine learning techniques for further optimization.
- Conducting A/B testing to validate the model's performance in a production environment.

Contact Information

Stay Connected:

- [GitHub Repository Link]
- [LinkedIn Profile]

Thank You

Express gratitude for attendance and engagement.

Encourage further exploration of the GitHub repository for detailed project documentation.