Train Sales Forecasting Project Overview

Purpose

The Train Sales Forecasting Project is designed to predict future train ticket sales based on historical data. Utilizing data-driven insights, the project aims to optimize ticket allocation, pricing strategies, and improve overall operational efficiency. This approach is crucial for enhancing customer satisfaction and maximizing revenue.

Methodology

The project employs advanced data analytics and machine learning techniques to analyze past sales trends. Key factors considered include days to departure, train types, seasonal variations, and historical booking patterns. By simulating different booking scenarios, the model provides robust predictions of future sales.

Python Code

The accompanying Python script serves as the backbone of the analysis. It performs the following key functions:

1. Data Preparation: The script loads and preprocesses the train sales data, setting the stage for analysis. This includes handling missing values, encoding categorical variables, and feature engineering.

2. Forecasting Model: A Random Forest Regressor model is implemented for forecasting, chosen for its ability to handle non-linear data and provide accurate predictions.

3. Model Validation: The script includes a mechanism for cross-validation to assess the model's performance, ensuring its reliability and accuracy.

I have successfully computed a MASE score of 64.40 in the Additive model, specifically for ‘days prior’ combined with ‘day of week’, which signifies a notable improvement in forecasting accuracy when integrating the 'day of week' variable.

In a prior academic endeavor, I engaged in a comparable project focused on forecasting airline data, achieving a commendable score in the Mean Absolute Scaled Error (MASE) metric. Regrettably, the specifics of this project are confidential and cannot be shared publicly due to a Non-Disclosure Agreement (NDA).