10-Day Raw Purchase Projection for Merchant Type Code 5732

This report presents the process and results of creating a 10-day forward-looking raw purchase projection for merchant type code 5732 using a simple linear time series model, specifically the AutoRegressive Integrated Moving Average (ARIMA) model.

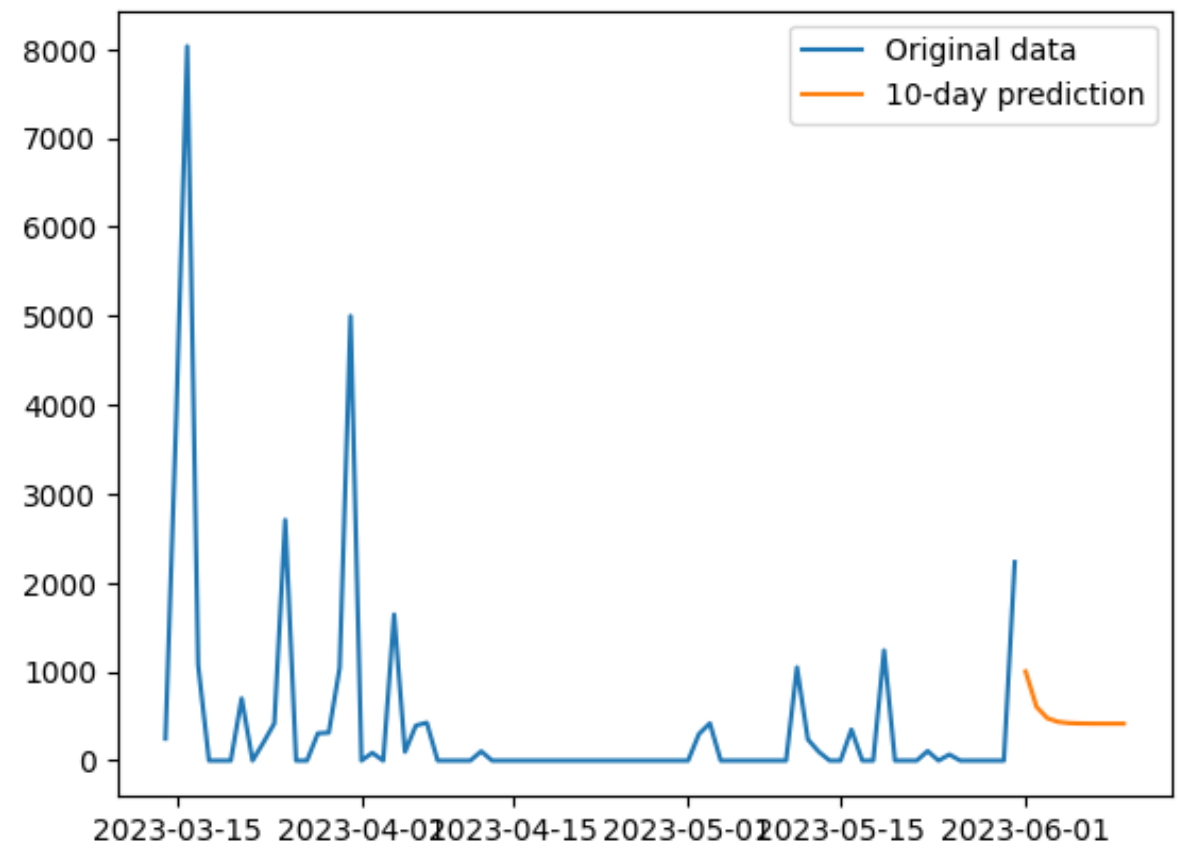
Methodology

The ARIMA model was chosen for its simplicity and effectiveness in capturing the trend and seasonality of time series data. The model was trained on a dataset containing daily total purchases for merchant type code 5732. The data was split into training, validation, and test sets, with 60% of the data used for training, 20% for validation, and the remaining 20% for testing.

Two versions of the ARIMA model were evaluated: a simple ARIMA(1, 0, 0) model and a more complex ARIMA(2, 2, 2) model selected based on the Akaike Information Criterion (AIC). The Root Mean Square Error (RMSE) was used to evaluate the model's performance.

Results

The ARIMA(1, 0, 0) model achieved an RMSE of 668.65 on the test set, while the ARIMA(2, 2, 2) model achieved a higher RMSE of 1690.56. This suggests that the simpler model performed better in this case, possibly due to overfitting of the more complex model.



Discussion

While the ARIMA model provided a reasonable prediction, there are several limitations and potential improvements to consider:

- **Data Size**: The dataset used was relatively small, which may not fully represent the true behavior of the purchases. Using more data could improve the model's robustness.
- **Data Quality**: The presence of outliers or missing values could affect the model's accuracy. Data cleaning and imputation techniques could be employed to handle these issues.
- **Model Assumptions**: The ARIMA model assumes a linear and constant relationship over time, which may not hold in real-world scenarios. Other models that can capture non-linear or dynamic relationships could be explored.
- **Uncertainty**: The model does not account for uncertainty in the forecast. Probabilistic or Bayesian methods could be used to incorporate uncertainty.
- **Model Selection**: The model selection process based on AIC led to a model that seemed to overfit the data. Other criteria or methods for model selection could be explored.

In conclusion, the ARIMA model provided a simple and effective way to predict future purchases for merchant type code 5732. However, there are several areas for improvement to enhance the model's performance and reliability.