

## Documentation

1. The code starts by importing the necessary libraries for data analysis and modeling, including pandas, numpy, statsmodels, datetime, and matplotlib. It then reads a CSV file named 'data\_set\_.csv' into a pandas DataFrame.
2. The next step involves preprocessing the data. The code converts the date variables in the DataFrame to datetime format using the specified date format. It extracts the month and quarter information from the order and delivery dates, respectively, and adds these as new columns to the DataFrame.
3. The code selects a subset of variables of interest from the DataFrame and drops any rows with missing values. It calculates the number of days between the order date and the requested delivery date and adds this as a new column called "Delivery Time". It also removes rows with invalid item values.
4. To prepare the data for modeling, the code converts the "value" and "items" columns to integer type. It then calculates the total order value by multiplying the price and quantity, and adds this as a new column called "Total Value". Additionally, it calculates the choice probability based on the delivery time and total value using a logistic regression formula and adds this as a new column called "Choice Probability".
5. The code defines the independent variables (X) as the order and delivery months, and the dependent variable (y) as the choice probability. It fits a logistic regression model using the logit function from the statsmodels library. The results of the regression are printed using the summary2() method.
6. Next, the code calculates the "x\_values" based on the logistic regression coefficients and the order and delivery months. It then adds 11 to the "x\_values" to calculate the predicted number of items, and stores these predictions in a new column called "predicted\_items".
7. Finally, the code performs SARIMAX modeling using the predicted items as the target variable. It fits a SARIMAX model with specified orders and seasonal orders

using the SARIMAX class from the statsmodels library. The results of the SARIMAX model are printed using the summary() method.

8. In summary, this code reads and preprocesses a dataset, performs logistic regression analysis on choice probability, calculates predicted item quantities, and applies SARIMAX modeling to forecast future item quantities based on the provided dataset.