The forecast algorithm used in this code is a Long Short-Term Memory (LSTM) network, which is a type of recurrent neural network (RNN) capable of learning long-term dependencies in time series data. LSTM networks are well-suited for time series forecasting tasks as they can learn patterns in the data over time and use this information to make predictions.

In this solution, we first convert the data from wide format to the long format using pandas melt function. In the wide format, each date is a separate column and the sales for each item code, category, and state are spread across multiple columns. In the long format, there is a single 'date' column and the sales for each item code, category, and state are stacked in a single 'sales' column. This allows for easier analysis and visualization of the time series data. The time series data is then preprocessed further by grouping it by date, item code, category, and state, and calculating the mean sales for each group, which can be considered as a form of aggregation as multiple entries are combined into a single entry according to the mean sales of each group. The data is then converted to a supervised learning format using a sliding window approach, where each sample consists of a sequence of past sales values of 60 days and the target is the sales value at a future time step which is 90 days in the case. The data is also filtered to ensure that the item code, category, and state are consistent within each sample.

The LSTM network is trained using the preprocessed data, with the mean squared error (MSE) loss function and the Adam optimizer. The model is trained for 10 epochs with a batch size of 256 and a learning rate of 0.0003. The best model is saved using a callback based on the validation loss, leading to a WMAPE validation forecast of around 92.5%.

Here are three key insights from the data that could help sales and demand planning teams:

- The data shows a clear seasonal pattern, with sales increasing in the months leading up to the holiday season and decreasing in the months afterwards. This information could be used to adjust production and inventory levels to meet expected demand.
- The data also shows a significant amount of variability in sales, with some items selling much better than others depending on the state location. This information could be used to identify popular items and focus marketing efforts on these products.
- The data shows that a significant proportion of sales (around 62%) are for items with a sales value of 0. This is an indication of intermittent demand in this time series. This could indicate that there are some items in the catalog that are not selling well and could be discontinued or marked down to clear inventory.

To improve forecasting accuracy, additional indicators that could be incorporated into the model include:

- Promotional activities: Information about promotional activities, such as discounts or advertising campaigns, could be used to predict spikes in demand for certain items.
- External factors: External factors such as economic indicators, government policies and regulations could also be incorporated into the model to improve forecasting accuracy.

- Competitor data: Information about competitor products and pricing could be used to predict changes in demand for certain items.