Preprocessing and loading of dataset

to

produce demand prediction model

Introduction:

Data preprocessing transforms the data into a format that is more easily and effectively processed in data mining, machine learning and other data science tasks.

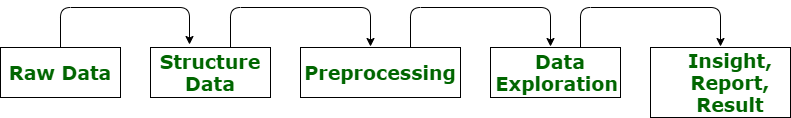
Data Loading is defined as copying data from one electronic file or database into another. Data loading implies converting from one format into another

Data loading:

data loading" refers to the process of acquiring, importing, and preparing data for analysis. It's one of the initial and critical steps in any data science or machine learning project. Data loading involves collecting data from various sources, such as databases, files, APIs, or web scraping, and then organizing and formatting it for further analysis.

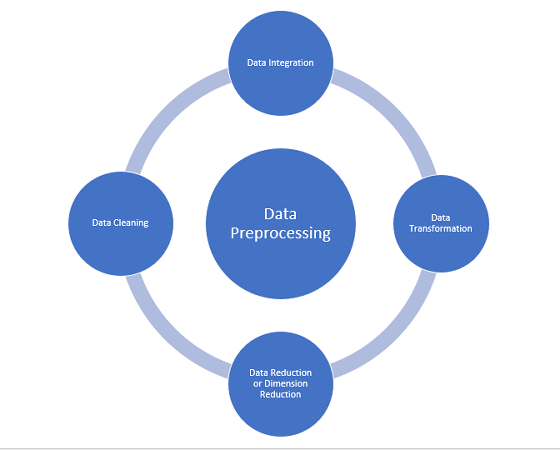
Data preprocessing:

Preprocessing simply refers to perform series of operations to transform or change data. It is transformation applied to our data before feeding it to algorithm. Data processing refers to perform operations on data to retrieve, transform, or change data, especially by computer.



Data preprocessing has four steps.

* Data quality assessment.
* Data cleaning.
* Data transformation.
* Data reduction



Benefits of data processing

* It’s a way to reduce the complexity of dataset
* Reduce the missing data
* Splitting data for training the dataset efficiency
* Transform data into usable form

Data Integration:

Combine historical sales data with external factors using a common identifier, such as date or product ID.

Train-Test Split:

Split the data into training and testing sets to evaluate the model's performance. Common splits are 70-30 or 80-20, but this can vary based on your dataset size and specific requirements.

Time Series Analysis:

If your data includes time series components, perform time series analysis to identify trends, seasonality, and other time-related patterns. This will help in selecting appropriate time series forecasting models.

Model Building:

Choose a suitable machine learning or time series forecasting model for demand prediction. Some popular options include linear regression, random forests, ARIMA, or deep learning models like recurrent neural networks (RNNs) or Long Short-Term Memory networks (LSTMs).

Monitoring and Maintenance:

Continuously monitor the model's performance and update it as needed. Demand patterns can change over time, so regular maintenance and retraining are essential.

Conclusion:

In conclusion, data preprocessing and loading are foundational steps in the journey to create a demand prediction model. Data loading involves the acquisition, importation, and organization of data from diverse sources, setting the stage for subsequent analysis. Data preprocessing, on the other hand, ensures that the data is in a suitable format for machine learning and data science tasks by addressing issues such as data quality, cleaning, transformation, and reduction.

By merging historical sales data with external factors, you create a comprehensive dataset, laying the groundwork for accurate demand predictions. The division of this data into training and testing sets is crucial for model evaluation, with time series analysis aiding in the recognition of temporal patterns and trends.

Lastly, continuous monitoring and maintenance of the model are essential to ensure it remains relevant and effective as demand patterns evolve over time.

In summary, data preprocessing and loading are the pillars upon which the demand prediction model is built, and these crucial steps lay the foundation for the accurate forecasting of product demand.