Problem Statement

Background:
In the realm of market analysis and forecasting, understanding the intricate patterns within time series data is paramount for informed decision-making. With the advent of machine learning techniques, it's now possible to delve deeper into historical market data to predict future trends accurately. In this context, we have at our disposal a dataset containing monthly market data spanning multiple years, encompassing various regions, commodities, and pricing information.

Objective:
The primary objective of this project is to develop a robust time series machine learning model capable of accurately forecasting market trends based on historical data. By leveraging advanced algorithms, we aim to predict the quantity and prices of commodities for future months, empowering stakeholders to make proactive decisions regarding production, procurement, pricing strategies, and resource allocation.

Data Description:
The dataset comprises the following columns:

- market: The market or commodity under consideration.
- month: The month for which the data is recorded.
- year: The year corresponding to the recorded data.
- quantity: The quantity of the commodity traded or available.
- priceMin: The minimum price of the commodity during the month.
- priceMax: The maximum price of the commodity during the month.
- priceMod: The mode or most frequently occurring price of the commodity during the month.
- state: The state or region where the market is located.
- city: The city where the market is situated.
- date: The specific date of the recorded data.

Task:
The task involves several key steps:

1. Data Preprocessing: Cleaning the dataset, handling missing values, and encoding categorical variables.
2. Exploratory Data Analysis (EDA): Analyzing the temporal patterns, identifying seasonality, trends, and anomalies within the data.
3. Feature Engineering: Creating relevant features such as lagged variables, rolling statistics, and seasonal indicators.
4. Model Selection and Training: Evaluating various time series forecasting models such as ARIMA, SARIMA, Prophet, and LSTM, selecting the most suitable one, and training it on the dataset.
5. Model Evaluation: Assessing the model's performance using appropriate metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).
6. Fine-tuning and Validation: Fine-tuning the model parameters, validating its performance on unseen data, and iterating if necessary.

Outcome:
The ultimate goal is to deploy a production-ready machine learning model capable of generating accurate forecasts for market quantity and prices for future months. The insights derived from this model will aid stakeholders in making data-driven decisions, optimizing inventory management, pricing strategies, and resource allocation, thereby enhancing overall efficiency and profitability in the market ecosystem.