**Future Sales Prediction Model**

## ***Team members:***

* *Saai krahaanth. S Ja (411621104045)*
* *Balaji .A (411621104005)*
* *Vigneshwaran.V (411621104055)*
* *Yuven .K (411621104056)*
* *Praveen .R(411621104041)*



# DEFENITION

The Future Sales Prediction Model is designed to assist businesses in forecasting their sales performance. This model leverages historical sales data, market trends, and advanced machine learning techniques to make accurate predictions. The key components of this model include data preprocessing, feature engineering, model selection, and evaluation.

# DESIGN THINKING

## 1.Data Source:

* Historical Sales Data:
* This is the most fundamental data source. It includes information about past sales, such as date, product, quantity sold, and revenue generated.
* This data allows you to identify trends and seasonality in sales.
* Product Data:
* Information about the products you sell, including attributes like product category, price, brand, and any product-specific features.
* Product data helps in understanding how different product attributes impact sales.

## 2. Data Preprocessing:

### Data Cleaning:

* + Handle missing data: Identify and handle missing values in your dataset through techniques like imputation (e.g., filling missing values with means or medians) or removing rows with missing data.
  + Outlier detection: Detect and deal with outliers that can skew predictions. You can remove outliers or transform them using methods like winsorization.
  + Data validation: Ensure the integrity of your data by validating that it conforms to expected formats and ranges.

### Data Transformation:

* + Feature scaling: Normalize or standardize numerical features to bring them to a common scale. This is particularly important for algorithms sensitive to the scale of input features, like gradient-based methods.
  + Encoding categorical variables: Convert categorical variables into numerical representations. Common techniques include one-hot encoding or label encoding.
  + Handling time series data: If working with time series data, make sure to sort the data by date and handle any irregular time intervals. You may also need to aggregate data over specific time windows.
  + Log transformations: Apply logarithmic transformations to data that exhibits exponential growth, which can make it more amenable to linear modeling.

3. Feature Engineering:

### Price-Related Features:

* + Incorporate features related to product prices, such as average price, price changes, or discounts. These can influence sales patterns.

### Customer Features:

* + If you have customer data, engineer features related to customer behavior, such as the average purchase frequency, customer lifetime value, or customer segments.

4. Model Selection:

### Time Series Models:

* + If working with time series data, consider models like ARIMA (AutoRegressive Integrated Moving Average), Exponential Smoothing, or Seasonal Decomposition of Time Series (STL) for capturing temporal patterns and seasonality.

### Ensemble Models:

* + Consider using ensemble techniques like Bagging (e.g., Random Forest) or Boosting (e.g., XGBoost) to combine the strengths of multiple models for improved accuracy.

5. Training and Validation:

### Data Preparation:

* + Ensure your dataset is properly prepared, cleaned, and preprocessed, as discussed in previous responses. This includes handling missing values, encoding categorical variables, and scaling features.

### Data Splitting:

* + Split your dataset into training, validation, and test sets. The training set is used to train the model, the validation set helps tune hyperparameters and monitor model performance, and the test set is reserved for final evaluation.

### Feature Selection:

* + Identify the most relevant features to include in your model based on domain knowledge and feature importance analysis. Remove or transform less important features to simplify the model.

6. Prediction and Evaluation:

### Mean Absolute Error (MAE):

* + MAE measures the average absolute difference between the predicted sales and the actual sales. It gives you a sense of the average magnitude of errors in your predictions.

### Mean Squared Error (MSE):

* + MSE measures the average squared difference between predicted and actual sales. It penalizes larger errors more heavily than MAE, making it sensitive to outliers.

### Root Mean Squared Error (RMSE):

* + RMSE is the square root of MSE. It provides a measure of the average error in the same units as the target variable (sales). RMSE is widely used and often preferred when larger errors should be penalized.

# CONCLUSION

The Future Sales Prediction Model serves as a powerful tool for businesses to anticipate future sales trends, allocate resources efficiently, and stay competitive in dynamic markets. This abstract model provides a high-level overview of the data science process involved in creating a future sales prediction model. Actual implementations may vary in complexity and require expertise in data analysis, machine learning, and domain knowledge specific to the industry and business.