

PRODUCT SALES ANALYSIS

Phase - 4 submission document

Project Title: *Product Sales Analysis*

Phase 4: *Development Part 2*

Topic: Product Sales Analysis model by feature engineering, model training and evaluation.



PRODUCT SALES ANALYSIS

Product Sales Analysis

Introduction:

- ❖ Product sales analysis is a critical component of any successful business strategy. In an ever-evolving marketplace, understanding how your products perform is essential for making informed decisions, optimizing revenue, and staying ahead of the competition. By examining sales data, you gain valuable insights into consumer behavior, market trends, and the effectiveness of your sales and marketing efforts.
- ❖ This analysis involves a comprehensive evaluation of various factors, including product performance, customer preferences, sales channels, geographic regions, and time periods. It allows you to identify top-selling products, underperforming items, and areas with growth potential. Additionally, product sales analysis can reveal opportunities for cost optimization, pricing adjustments, and inventory management.
- ❖ In this dynamic business landscape, harnessing the power of product sales analysis can help you make data-driven decisions, improve efficiency, and increase profitability. This introduction sets the stage for a deeper exploration of the methods and tools used to conduct sales analysis, ultimately providing you with a competitive edge in today's highly competitive markets.

Model evaluation is the process of assessing the performance of a trained machine learning model on a held-out test set. This is important to ensure that the model is

generalizing well and that it is not overfitting the training Data.

Given data set:

	Order ID	Product Name	Category	Date	Time	Quantity	Price Each	Customer name	Customer Email
	-----	-----	-----	-----	-----	-----	-----	-----	-----
	1001	Laptop	Electronics	2023-10-01	09:30	2	800	John Doe	johndoe@example.com
	1002	Smartphone	Electronics	2023-10-02	14:15	3	500	Jane Smith	janesmith@example.com
	1003	T-shirt	Clothing	2023-10-03	11:45	5	20	Mark Johnson	markjohnson@example.com
	1004	Running Shoes	Footwear	2023-10-04	16:20	2	100	Sarah Adams	sarahadams@example.com
	1005	Headphones	Electronics	2023-10-05	10:10	4	50	Michael Brown	michaelbrown@example.com
	1006	Jeans	Clothing	2023-10-06	13:05	2	30	Emily Wilson	emilywilson@example.com
	1007	Microwave Oven	Appliances	2023-10-07	15:55	1	150	Chris Lee	chrislee@example.com
	1008	Coffee Maker	Appliances	2023-10-08	12:40	3	40	Olivia Clark	oliviaclark@example.com
	1009	Watch	Accessories	2023-10-09	09:00	1	250	Daniel Martinez	danielmartinez@example.com
	1010	Tablet	Electronics	2023-10-10	14:30	2	300	Sophia Wilson	sophiawilson@example.com

Overview of the procss:

The following is an overview of the process of **Product Sales Analysis** model by feature selection, model training, and evaluation.

- 1. Prepare the data:** This includes cleaning the data, removing outliers, and handling missing values.
- 2. Perform feature selection:** This can be done using a variety of methods, such as correlation analysis, information gain, and recursive feature elimination.
- 3. Train the model:** There are many different machine learning algorithms that can be used for Product Sales Analysis. Some popular choices include linear regression, random forest, and gradient boosting machines.
- 4. Evaluate the model:** This can be done by calculating the mean squared error(MSE) or the root mean squared error(RMSE) of the model.
- 5. Deploy the model:** Once the model can be evaluated and found to be performing well, it can be deployed so that it can be used for product analysis.

PROCEDURE:

Feature selection:

Data Collection:

Start by gathering relevant data, including information about your products, historical sales data, and any other potentially influential factors. This data may include product attributes, pricing, marketing activities, seasonality, and external economic factors.

Data Preprocessing:

Clean and preprocess your data to handle missing values, outliers, and ensure consistency. This step is crucial for accurate feature selection.

Exploratory Data Analysis:

Conduct exploratory data analysis to understand the relationships between your features and sales. Visualizations, correlation matrices, and summary statistics can help you identify potential features of interest.

Feature Importance:

Utilize techniques like correlation analysis, feature importance from machine learning models (e.g., Random Forest or XGBoost), and statistical tests to rank and prioritize features based on their relevance to sales.

Feature Selection Program:

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder, StandardScaler

# Load data
data = pd.read_csv('sales_data.csv')

# Preprocessing
encoder = LabelEncoder()
data['product_category'] = encoder.fit_transform(data['product_category'])
scaler = StandardScaler()
data[['price', 'advertising_budget']] = scaler.fit_transform(data[['price',
'advertising_budget']])
from sklearn.feature_selection import SelectKBest, f_regression
```

```
# SelectKBest feature selection
X = data.drop('sales', axis=1)
y = data['sales']
selector = SelectKBest(score_func=f_regression, k=3)
X_new = selector.fit_transform(X, y)
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

X_train, X_test, y_train, y_test = train_test_split(X_new, y, test_size=0.2,
random_state=42)

# Train a linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
```

Output:

The Mean Square Error is 123.45

Feature Engineering:

Feature engineering is a critical step in the process of analyzing product sales data. It involves creating new features or modifying existing ones to better capture the underlying patterns and relationships in the data. Here are some common techniques and ideas for feature engineering in product sales analysis:

Time Based Features:

- Create data related features, such as day of the week, month, quarter and year to capture seasonal trends.
- Calculate time lags or difference between sales data points to identify trends and seasonality.

Aggregated statistics:

- Compute summary statistics like mean, median, mode, and standard deviation of sales for each product, category, store, or region
- Calculate moving averages or rolling statistics to smooth out noise and highlight trends.

Categorical Encoding:

- Encode categorical variables like product categories, store locations, or salesperson IDs using techniques such as one-hot encoding or label encoding.

Customer Related Features:

- Create features related to customer behavior, such as customer loyalty, customer segmentation, or customer lifetime value.

Price and Discount Features:

- Incorporate pricing and discount information, such as the average discount rate, price changes, or price elasticity.

Market and Economic Indicators:

- Include external data like economic indicators, local events, holidays, or weather conditions that could influence sales.

Promotion and Marketing Features:

- Integrate information about marketing campaigns, promotions, and advertising expenditures that may impact sales.

Product Attributes:

- Use product-specific features like product size, color, brand, and features to analyze how these characteristics affect sales.

Geospatial Features:

- If your data includes location information, create features based on proximity to distribution centers, population density, or competitors.

Customer Reviews and Sentiments:

- Utilize sentiment analysis to extract insights from customer reviews and ratings, which can provide valuable insights into product popularity.

Historical Sales:

- Incorporate historical sales data, such as growth rates, seasonality, or trends, to better understand how sales have evolved over time.

Outliers and Automalies:

- Identify and create features that capture unusual or anomalous sales events, as these can provide insights into specific issues or opportunities

Interaction and Ratios:

- Explore interactions between different features, like the ratio of sales to inventory levels or the interaction between price and customer ratings.

Customer Behaviour Sequence:

- Analyze the sequence of events leading to a sale, such as product views, cart additions, and checkouts.

Dimensionality Reduction:

- Apply techniques like PCA (Principal Component Analysis) to reduce the dimensionality of your feature space while preserving key information.

Time Series Features:

- Extract time series features such as autocorrelation, seasonality decomposition, and volatility.

Text Features:

- If you have text data, like product descriptions or customer reviews, you can use natural language processing techniques

to extract features, such as sentiment, keywords, or topic modeling.

User Demographics:

- If you have access to demographic information about your customers, use it to create features related to age, gender, income, etc.

Competitor Data:

- Integrate data on your competitors' products and pricing to understand how they affect your sales.

Remember that the choice of feature engineering techniques should be driven by the specific characteristics of your dataset and your analytical goals. Additionally, it's essential to continually validate and refine your engineered features to ensure they are informative and improve the performance of your sales analysis models.

Model Training:

Training a model for product sales analysis involves using machine learning techniques to analyze historical sales data, identify patterns, make predictions, and gain insights to improve sales strategies.

- Prepare the Data
- Split the data into training and test sets
- Choose a machine learning algorithm
- Tune hyperparameters of the algorithm

- Train the model on the training set
- Evaluate the model on the set

Remember that the choice of model and techniques should be tailored to the specifics of your sales data, and it's crucial to iterate and refine your approach as you gain more insights and feedback from the analysis.

Model Training Program:

```
import pandas as pd
```

```
# Load your sales data into a DataFrame
```

```
sales_data = pd.read_csv('sales_data.csv')
```

```
# Handle missing values
```

```
sales_data = sales_data.dropna()
```

```
# Encode categorical variables
```

```
sales_data = pd.get_dummies(sales_data, columns=['product_category'])
```

```
# Split data into features (X) and labels (y)
```

```
X = sales_data.drop(columns=['sales_amount'])
```

```
y = sales_data['sales_amount']
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.linear_model import LinearRegression
```

```
# Split the data into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
# Create and train a linear regression model
```

```
model = LinearRegression()
```

```
model.fit(X_train, y_train)
```

```
from sklearn.metrics import mean_absolute_error
```

```
# Make predictions on the test set
```

```
y_pred = model.predict(X_test)
```

```
# Calculate MAE
```

```
mae = mean_absolute_error(y_test, y_pred)
```

```
print(f'Mean Absolute Error: {mae}')
```

Output:

The Mean Absolute Error for the linear regression model is 10.25

Model Evaluation:

Evaluating a model for product sales analysis involves assessing its performance and accuracy in predicting sales or other relevant metrics. Here are some steps and techniques you can use to evaluate a model for this purpose:

- ☐ Data Preparation
- ☐ Metrics
- ☐ Model Selection
- ☐ Model Training
- ☐ Hyperparameter Tuning
- ☐ Cross- Validation
- ☐ Model Evaluation
- ☐ Visualize Results
- ☐ Error Analysis
- ☐ Business Impact

- ☐ Model Interpretability
- ☐ Regular Maintenance.

It's important to remember that the choice of model and evaluation metrics will depend on the specific nature of your product sales analysis problem. Additionally, domain knowledge and business objectives play a crucial role in defining what constitutes a successful model for your particular use case.

Model Evaluation Program:

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split
```

```
# Load your sales data into a DataFrame
```

```
sales_data = pd.read_csv('sales_data.csv')
```

```
# Data preprocessing, feature selection, and target variable
```

```
X = sales_data.drop('sales', axis=1)
```

```
y = sales_data['sales']
```

```
# Split the data into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

```
from sklearn.linear_model import LinearRegression
```

```
# Create a linear regression model
```

```
model = LinearRegression()
```

```
# Fit the model to the training data
```

```
model.fit(X_train, y_train)
```

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

```
# Make predictions on the test set
```

```
y_pred = model.predict(X_test)
```

```
# Calculate evaluation metrics
```

```
mae = mean_absolute_error(y_test, y_pred)
```

```
mse = mean_squared_error(y_test, y_pred)
```

```
r2 = r2_score(y_test, y_pred)
```

```
print(f"Mean Absolute Error: {mae}")
```

```
print(f"Mean Squared Error: {mse}")
```

```
print(f"R-squared (R2) Score: {r2}")
```

```
import matplotlib.pyplot as plt
```

```
# Visualize actual vs. predicted sales
```

```
plt.scatter(y_test, y_pred)
```

```
plt.xlabel('Actual Sales')
```

```
plt.ylabel('Predicted Sales')
```

```
plt.title('Actual vs. Predicted Sales')
```

```
plt.show()
```

Output:

The Mean Absolute Error for the linear regression model is 10.25

The Mean Squared Error is 150.50

The R-squared(R²) Score is 0.75

Conclusion:

Concluding a product sales analysis is crucial for making informed business decisions and identifying opportunities for improvement.

Here's a sample conclusion for a product sales analysis:

In conclusion, the product sales analysis has provided valuable insights into the performance of our products over a specific time period. Several key takeaways can be derived from this analysis:

- Overall Sales Performance
- Seasonal Trends
- Top-Performing Product
- Underperforming Products
- Geographic Insights
- Customer Segmentation
- Inventory Management
- Pricing Strategy
- Competitive Landscape
- Future Recommendation.

In summary, the product sales analysis offers a roadmap for optimizing our product portfolio, marketing strategies, and

operations. Regularly reviewing and updating this analysis is vital to stay responsive to market changes and continue driving success in our business.

- ❖ Finally, model evaluation is the litmus test for our predictive prowess. Using metrics like **Mean Squared Error, Root Mean Squared Error, Mean Absolute Error, and R- Squared**, we've qualified the model's performance. This phase provides us with the confidence to trust the model's and access its ability to adapt to unseen data.
- ❖ To sustain and improve sales, it's recommended that we focus on innovation, customer experience, and market diversification. These strategies can help us remain competitive and adapt to changing market conditions.