

E-commerce-Data-Analysis for Sales Strategy

The following data analysis techniques will guide Kmart in refining its sales strategy for 2020, leveraging insights from 2019 data. By employing a structured 10-step approach—from data retrieval and preparation to exploration, modeling, and application development—Kmart will uncover essential insights, such as the best months for sales, top-performing cities, ideal times for advertising, and the most popular products. These insights will be key in enabling Kmart to optimize its sales performance and enhance strategic decision-making for the upcoming year.

1. Research Data

Before beginning any analysis, you need to gain a clear understanding of the data and the context in which it's being used. This step involves:

- **Understanding the Business Context:** Kmart wants to analyze its 2019 e-commerce sales data to determine the most effective sales strategy for 2020. The objective is to extract actionable insights that could drive better sales performance.
 - **Understanding the Data:** The dataset includes sales transactions, product details, order times, and city information. Analyzing these will help identify patterns, correlations, and trends.
 - **Defining Key Metrics:** Key performance indicators (KPIs) such as total sales, total quantity sold, best-selling products, and sales by city or time of day will be the focus.
-

2. Data Retrieval

In this step, you acquire and load the data into your working environment:

- **Access the Data:** Obtain the dataset, usually provided in CSV, Excel, or database format.
 - **Initial Inspection:** Understand the structure and content of the dataset by reviewing the first few rows and columns, and confirm that all necessary data is included.
-

3. Data Preparation

Data preparation is the crucial step where you clean, preprocess, and structure the data to make it suitable for analysis:

- **Handling Missing Data:** Identify and handle any missing or null values in the dataset (e.g., filling in missing values or removing rows with missing data).
 - **Data Type Conversion:** Convert columns like date or time into the appropriate data types for analysis, ensuring they are recognized as dates or times.
 - **Feature Engineering:** Extract additional useful features, such as month, day of the week, or hour, from the datetime column to help analyze sales trends by time.
 - **Filtering Data:** Filter out any irrelevant or incorrect data points, such as transactions with zero sales or invalid product codes.
-

4. Data Exploration

Exploratory Data Analysis (EDA) helps you understand the data distribution, relationships between variables, and potential trends or anomalies:

- **Summary Statistics:** Calculate basic statistics like mean, median, standard deviation, and range to understand the general characteristics of the data.

- **Visualizing Trends:** Plot graphs to visualize important patterns, such as sales by month, sales by city, and sales distribution by product.
 - **Correlation Analysis:** Look for correlations between variables, such as sales and product price, or time of day and purchase frequency.
-

5. Data Modeling

Data modeling involves selecting and building appropriate models to predict future outcomes or understand relationships in the data:

- **Choosing a Model:** Depending on the business question, models such as linear regression, time series forecasting, or machine learning algorithms like decision trees and random forests could be used.
 - **Feature Selection:** Identify which features (e.g., month, time of day, city, product price) are most predictive of sales and should be included in the model.
 - **Train-Test Split:** Split the data into training and testing sets to ensure that the model can generalize to new, unseen data.
-

6. Linear Regression

In this step, you build a linear regression model to understand the relationship between sales and other factors:

- **Regression Analysis:** Perform linear regression to predict sales based on features like the month, price, and city. This will help assess which factors have the most impact on sales.
- **Model Evaluation:** Evaluate the regression model using metrics such as Mean Squared Error (MSE) to determine its predictive accuracy.
- **Insights:** From the regression model, you can gain insights into how different variables influence sales. For example, you may find that time of day or price significantly impacts sales.

7. Cleansing and Transforming Data

Data cleansing and transformation ensure that the dataset is in the best shape for modeling:

- **Outlier Detection:** Identify and handle outliers or extreme values that could skew the results.
 - **Normalizing Data:** Standardize numerical features (such as sales, price, and quantities) to a similar scale, especially if using machine learning algorithms that are sensitive to feature scales.
 - **Encoding Categorical Variables:** If necessary, encode categorical variables (such as city or product category) into numerical values to be used in modeling.
-

8. Exploratory Data Analysis (EDA)

Revisit the data with a more in-depth analysis, digging deeper into specific trends:

- **Sales by Time:** Identify the best times for sales (e.g., time of day, month, day of the week) to optimize ad placements and marketing strategies.
 - **Sales by Product:** Analyze which products sell the most and if there are any patterns in terms of product category, price, or other features.
 - **Sales by City:** Examine how sales vary by city, allowing you to tailor marketing strategies to high-performing areas.
 - **Heatmaps/Visualizations:** Use heatmaps or other visual tools to identify strong correlations between sales and other variables.
-

9. Building Models

Once the data is clean and prepared, you can build more advanced predictive models for forecasting and decision-making:

- **Predictive Models:** Implement more complex models like Random Forests, Gradient Boosting, or even neural networks, depending on the problem's complexity.
 - **Time Series Forecasting:** If predicting future sales, use time series models to account for seasonal trends and temporal dependencies.
 - **Model Tuning:** Fine-tune the model by optimizing hyperparameters to improve accuracy and generalization.
 - **Model Evaluation:** Test the model on a separate testing dataset and evaluate its performance using metrics such as accuracy, precision, recall, and F1 score for classification or RMSE for regression.
-

10. Presenting and Building Application

The final step is to present the findings and create a usable application for stakeholders:

- **Visualization and Reporting:** Create clear, visually engaging reports or dashboards to present the key findings, such as the best month for sales, top-selling products, and recommendations for sales strategy.
- **Decision Support:** Provide actionable insights for Kmart's sales team, such as the optimal times to run ads, which cities to focus on for promotions, and product bundling strategies.
- **Building Interactive Dashboards:** Use tools like **Streamlit**, **Dash**, or **Power BI** to create an interactive dashboard where Kmart's team can explore the data and adjust parameters (e.g., changing the time of day, selecting a specific product, etc.) to see how different factors impact sales.
- **Deployment:** If the models are to be used for ongoing decision-making, deploy them in a real-time environment where they can be regularly updated with new sales data.

In conclusion, by applying the 10-step data analysis process, Kmart can unlock valuable insights from its 2019 sales data, providing a solid foundation for refining its sales strategy in 2020. This approach will allow Kmart to identify critical patterns such as peak sales periods, high-performing cities, optimal advertising times, and the best-selling products. With these actionable insights, Kmart will be better equipped to make data-driven decisions, enhance customer engagement, and ultimately improve sales performance in the coming year.