DOCUMENTATION

SALES ANALYSIS USING PYTHON

**Skills:**

With this project I will be developing the following skills-

1. Python Programming
2. Data cleaning and preprocessing
3. Data visualization
4. Statistical analysis
5. Time-series analysis
6. Machine learning
7. Business acumen
8. Communication
9. Version control
10. Project Management

**Solved Business Problems:**

* **Identifying Sales Trends**

Understanding the overall trend in sales helps businesses adapt to market changes and make informed decisions on inventory, pricing, and marketing strategies.

* **Product Performance Analysis**

Businesses can identify their best-performing and underperforming products. This information helps in optimizing the product portfolio, marketing efforts, and investment allocation.

* **Pricing Strategy Optimization**

By analyzing sales data, businesses can evaluate the impact of different pricing strategies on sales volume and revenue. This information is crucial for setting competitive prices and maximizing profits.

* **Forecasting and Demand Planning**

Sales analysis aids in accurate forecasting, allowing businesses to plan production, staffing, and other resources based on anticipated demand.

**Identifying Sales Trend:**

Steps overview-

1. Load and explore sales data with Python.
2. Visualize time-series data to identify trends.
3. Apply statistical methods to smooth data and highlight trends.

**Dataset**

The dataset that we will be using for doing sales analysis is from the retail industry. It is historical sales data for 45 stores located at different locations and each store has some departments inside it. The store also runs some promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of which are the Super Bowl, Labor Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks.

We have 3 csv files- Stores, Features and Sales.

*Stores*

Anonymized information about the 45 stores, indicating the type and size of store

*Features*

Contains additional data related to the store, department, and regional activity for the given dates.

* Store - the store number.
* Date - the week.
* Temperature - average temperature in the region.
* Fuel\_Price - cost of fuel in the region.
* MarkDown1-5 - anonymized data related to promotional markdowns. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with a NA.
* CPI - the consumer price index.
* Unemployment - the unemployment rate.
* IsHoliday - whether the week is a special holiday week.

*Sales*

Historical sales data, which covers 2010-02-05 to 2012-11-01. Within this tab you will find the following fields:

* Store - the store number.
* Dept - the department number.
* Date - the week.
* Weekly\_Sales - sales for the given department in the given store.
* IsHoliday - whether the week is a special holiday week.

**Exploratory data analysis**

*Data Types and Missing Values*

* Check the data types of each column.
* Identify and handle missing values, especially in markdown data.

*Time Series Analysis*

* Convert the 'Date' column to a datetime type.
* Explore time-related trends, seasonality, and cycles.

*Store and Department Distributions*

* Understand the distribution of stores and departments in the dataset.

*Holiday Analysis*

* Explore the frequency and impact of special holiday weeks on sales.

**Preprocessing**

*Handling Missing Values*

* Deal with missing values in the markdown columns. Consider imputation or dropping columns if necessary.

*Data Merging*

* Merge the Sales, Features, and Stores datasets based on common columns like 'Store' and 'Date'.

*Feature Engineering*

* Extract additional features from the date, such as day of the week, month, and year.
* Consider creating lag features to capture historical sales patterns.

**Visualizations**

*Time Series Plots*

* Plot weekly sales over time to identify trends and seasonality.

*Time Series Decomposition*

* Use methods like seasonal decomposition to break down the time series into trend, seasonal, and residual components.

*Statistical Analysis*

* Perform statistical tests to identify significant trends or changes in sales patterns.

*Correlation Analysis*

* Analyze the correlation between sales and other features like temperature, fuel price, CPI, and unemployment.

*Store-wise and Department-wise Analysis*

* Create visualizations to understand sales distribution across different stores and departments.

*Holiday Impact Visualization*

* Visualize the impact of special holiday weeks on sales using appropriate charts.

*Markdown Effect*

Investigate the impact of promotional markdowns on sales. Plot markdowns against sales to identify correlations.

*Holiday Sales Boost*

Assess the impact of holidays on sales. Compare holiday weeks to non-holiday weeks.

*Store Type and Size Analysis*

Explore how store types and sizes influence sales. Plotting store size against sales may reveal interesting patterns.

**Recommendations**

*Optimal Markdown Strategy*

* Identify the effectiveness of markdowns in driving sales and optimize the timing and magnitude of future promotions.

*Seasonal Inventory Management*

* Adjust inventory levels based on identified seasonal patterns to avoid overstocking or stockouts.

*Holiday Campaigns*

* Plan targeted marketing campaigns around holidays to leverage the observed sales boost during special weeks.

*Store-Specific Strategies*

* Tailor strategies for different store types and sizes based on sales trends.

*Weather Impact*

* Consider weather-related factors in sales planning, especially if there is a significant correlation with sales.

*Operational Efficiency*

* Identify areas for operational improvement based on sales trends and performance.

**DOCUMENTATION END**