ADVANCED PROGRAMMING FINAL PROJECT - PROJECT GROUP 6

Team

Adwoa Nhyira Attafuah-Wadee

Shreeya Umesh Chitnis

Lina Botero

Problem

Helping customers of retail brands prepare towards Black Friday purchases, with an understanding of likely deal prices, and the sentiment towards the different products on sale.

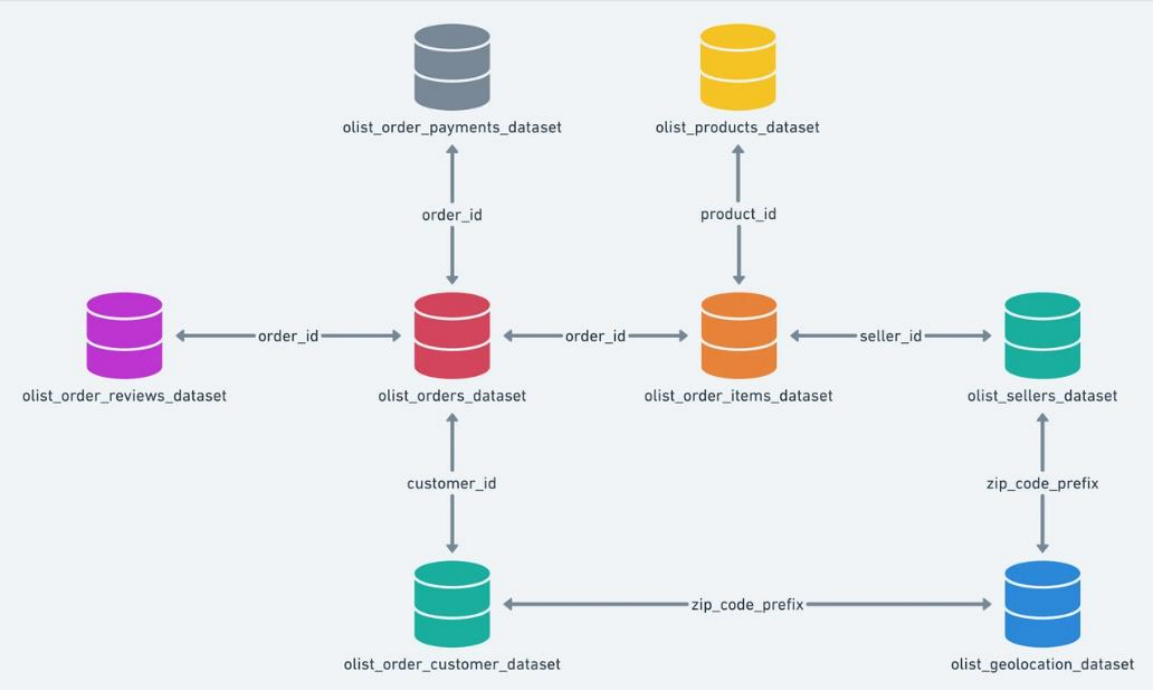
Solution: Black Friday Dashboard

The Black Friday Analysis Dashboard is a web-based application developed on Streamlit, designed to provide users with comprehensive insights into historical Black Friday data, and predictions of what the prices will look like for Black Friday in the following year. The dashboard can be manipulated to display different types of data from various product categories to provide a complete user experience.

Dataset

The dataset used is a Brazilian ecommerce public dataset of orders made at Olist Store. Olist is an ecommerce marketplace platform that connects small and medium-sized businesses to customers. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. The features of the dataset allow viewing orders from multiple dimensions, specifically, order status, price, payment, freight performance, customer and seller location, product attributes and customer reviews. The product categories and reviews are in Portuguese.

Data Schema



Data Preparation

Black Friday occurs in the month of November, specifically on the Friday after Thanksgiving (in the United States). It is heavily practiced in Brazil, with some businesses offering Black Friday deals for two weeks in advance of the actual day or the whole month of November.

To focus on Black Friday sales only, the dataset was filtered to contain sales between two weeks to Black Friday and the day itself.

For sentiment analysis, some of the datasets were merged into one data frame to begin analysis. These datasets are the olist_order_dataset, olist_order_items_dataset, olist_products_dataset,olist_orders_customers_dataset,and olist_orders_reviews_dataset.

For the predictive model, these datasets were merged: olist_order_dataset, olist_order_items_dataset, olist_products_dataset, olist_orders_customers_dataset, olist_orders_reviews_dataset, and olist_sellers_dataset.

Features and Functionalities of Black Friday Dashboard

Below are the key functionalities and features of the dashboard:

Descriptive Statistics: The dashboard offers descriptive statistics of historical Black Friday data, allowing users to explore various metrics such as sales volume, revenue, and discount percentages.

Visualizations: Various visualizations, including bar charts, word clouds, images, are available to present data in an intuitive and visually appealing format. Users can customize visualizations based on their preferences and requirements.

Sentiment Analysis: The dashboard implements sentiment analysis on product reviews for all the product categories. It evaluates the overall sentiment and provides insights into public perception of a product category, specifically whether sentiment is positive, negative, or neutral, and the sentiment compound score per product category. It includes an analysis of ratings (1 to 5 stars), and overall sentiment with that product category after purchase.

Product Price Prediction: A predictive model is integrated into the dashboard to forecast future prices of products based on historical data. This feature helps users anticipate market trends and make informed purchase plans.

User-friendly Interface: The dashboard features a user-friendly interface with interactive elements and customizable settings. There are clear filters to be able to navigate and manipulate the data according to user preferences. Users can easily navigate through different sections, explore data, and generate insights with minimal effort.

Data Integration and Translation: The dashboard integrates data from various sources, including Brazilian Black Friday data from Kaggle. The data was translated from Portuguese to English, enabling broader accessibility and analysis.

Customization and Personalization: Users can customize and personalize their dashboard experience by selecting specific metrics, time periods, and visualization options. This flexibility allows users to tailor the dashboard to their specific needs and preferences.

Export and Sharing: The dashboard can be easily shared with a link, so users can not only enjoy it themselves but also share insights with friends, family, colleagues, or stakeholders.

Tools, APIs and Libraries

Deal Price Predictive Model

For the predictive model, these libraries were used:

- Pandas – to manipulate, load and pre-process data
- LabelEncoder – to encode categorical target labels with integer values.
- train_test_split – to split the dataset into training and testing sets.
- RandomForestRegressor – to build a random forest regression model to predict target variables.

Sentiment Analysis

For the sentiment analysis, these libraries were used:

- NLTK – to tokenize text and analyse sentiment in text
- DeepL – to translate Portuguese to English
- Pandas – to manipulate, load and pre-process data
- Re – to remove special characters in review comments
- Matplotlib – to visualize average review scores against product categories

Black Friday Dashboard

For the Black Friday dashboard, these libraries were used:

- Streamlit – to create the visualization dashboard
- Pandas – to manipulate, load and pre-process data
- Matplotlib – to create the visualizations
- WordCloud – to create the Word Clouds
- BeautifulSoup – to display the images from Google in the dashboard

Overall, the Black Friday Analysis Dashboard provides a comprehensive platform for analysing historical Black Friday data, conducting sentiment analysis, and gaining valuable insights into consumer behaviour and market trends. It empowers users to make informed decisions and optimize their strategies for future Black Friday events.