IBM Naan Mudhalvan - Artificial Intelligence

|  |  |
| --- | --- |
| **Date** | **28-10-2023** |
| **Team ID** | **1051** |
| **Project Name** | **Market Basket Analysis** |
| **Domain** | **Artificial Intelligence** |

**Problem Statement:**

Unveiling Customer Behaviour through Association Analysis: Utilize market basket analysis on the provided dataset to uncover hidden patterns and associations between products, aiming to understand customer purchasing behaviour and identify potential cross-selling opportunities for the retail business.

**Understanding of the problem statement:**

Scope & Purpose: The problem revolves around a retail business's desire to gain deeper insights into customer purchasing behaviour to boost sales and customer satisfaction. AI can be used to automate the process of analysing vast amounts of transactional data to uncover patterns and associations that might be difficult for humans to identify manually.

**Proposing a Solution:**

The proposed solution involves leveraging Artificial Intelligence techniques, particularly machine learning and data mining, to perform market basket analysis. Market basket analysis, when powered by AI, can handle large datasets, adapt to changing customer behaviour over time, and provide more accurate and real-time insights.

**Introduction:**

In today's highly competitive retail landscape, understanding and effectively responding to customer behaviour is paramount to the success of any business. One powerful tool at the disposal of retailers is Market Basket Analysis, which enables them to uncover hidden patterns and associations between products, shedding light on intricate aspects of customer purchasing behaviour. This analysis is not only about understanding what customers buy but also about discovering what they buy together. In this context, we are presented with a compelling problem statement: "Unveiling Customer Behaviour through Association Analysis."

By taking this path, we embark on a journey to unlock the secrets of customer behavior, and in doing so, we equip retailers with the tools they need to stay ahead in an ever-changing and highly competitive market. The potential for growth and customer satisfaction lies within the patterns and associations waiting to be uncovered in the provided dataset, and the marriage of AI and Market Basket Analysis promises to be the key that unlocks this potential.

**Literature Survey:**

1. **A Technical Analysis of Market Basket by using Association Rule Mining and Apriori Algorithm**

This research paper discusses the use of association rule mining and the Apriori algorithm for market basket analysis. Market basket analysis is a data mining technique that identifies associations between items purchased by customers. The paper highlights how this analysis can assist retailers in expanding their marketing strategies by gaining insight into frequently purchased items. It also presents a survey of existing data mining algorithms for market basket analysis. The paper concludes with a discussion of the limitations of the Apriori algorithm and proposes the use of fuzzy logic for better selection of association rules.

1. **A Survey on Association Rule Mining**

Data mining is a process that uses various data analysis tools to discover patterns and relationships in large amounts of data. Association rule mining is an important aspect of data mining and is used for tasks such as market basket analysis. Many algorithms have been developed for association rule mining, including Apriori, FP-tree, and Fuzzy FP-tree. This survey explores different approaches to association rule mining and compares their advantages and disadvantages. It also discusses the use of fuzzy logic in association rule mining and the challenges associated with mining frequent patterns from fuzzy data.

1. **Direct-Indirect Association Rule Mining for Online Shopping Customer Data using Natural Language Processing**

This provides insights on how to analyze customer feedback and predict customer intents using advanced techniques like Apriori algorithm and sentimental analysis. The paper is organized into different sections, including literature review, proposed system overview, process pipeline, implementation details, and experimental results using sentimental analysis. The study also includes a 3D plot of review rating frequency to determine both the explicit and implicit relationship from customer review analysis. The analysis assists in identifying the implicit product to improve sales by providing offers for respective implicit products. Overall, the file aims to help businesses make informed decisions and improve their business models by analyzing customer feedback**.**

1. **An Overview of Classification Rule and Association Rule Mining**

This paper is about Classification Rule and Association Rule Mining, which are data mining techniques used in Market Basket Analysis to identify purchasing patterns and trends of customers. The study explains the process of extracting useful information from a large dataset and how these algorithms can support retailers in making informed decisions to increase their sales. This also includes references to various research papers related to data mining and image processing.

1. **Market Basket Analysis for Mobile Showroom**

This paper provides information on Market Basket Analysis, a data mining technique used to examine customer buying patterns and increase sales. The document discusses the purpose and workings of Market Basket Analysis, as well as how businesses can use the findings to improve their sales strategies. It also covers different data mining algorithms used for Market Basket Analysis and their differences. The PDF includes examples and references to further reading on the topic.

**Challenges:**

**1.Data Quality and Preprocessing:**

* AI models heavily depend on the quality of data. Retail datasets may have missing values, inconsistencies, and noise, which need to be addressed before analysis.
* Cleaning and preprocessing large datasets can be time-consuming and resource intensive.

**2.Data Privacy and Ethics:**

* Handling customer data requires strict compliance with privacy regulations (e.g., GDPR, CCPA). Ensuring data anonymity while retaining its utility is a complex task.
* Ethical concerns regarding the use of customer data for marketing and analysis must be addressed transparently.

**3.Scalability:**

* Retail businesses often deal with massive amounts of data. Scaling AI algorithms to handle such large datasets efficiently can be challenging.

**4.Algorithm Selection and Tuning:**

* Choosing the right AI algorithms (e.g., Apriori, FP-growth, deep learning) and tuning their parameters for specific retail datasets can be complex.
* Inaccurate algorithm selection can lead to suboptimal results.

**5.Real-time Analysis:**

* Achieving real-time analysis and responsiveness to changing customer behaviour can be demanding. AI systems must process data quickly and provide timely recommendations.

**6.Interpretable Models:**

* Complex AI models may provide accurate predictions but lack interpretability. Retailers need to understand and trust the insights generated by AI systems.

**7.Infrastructure and Resources:**

* Building and maintaining the necessary AI infrastructure, including hardware and software, can be costly.
* Acquiring and retaining AI talent for development and maintenance can be challenging.

**8.Dynamic Customer Behaviour:**

* Customer behaviour is not static. It evolves over time, influenced by various factors (seasonal trends, external events, marketing campaigns). AI models must adapt to these changes.

**9.Overfitting and Generalization:**

* AI models may be overfit to historical data, capturing noise rather than meaningful patterns. Ensuring models generalize well to new data is crucial.

**10.Customer Trust:**

* Building and maintaining trust with customers is vital. AI-driven recommendations should enhance the customer experience without feeling invasive or overly personalized.

**Advantage:**

* Identify cross-selling and up-selling opportunities. Market basket analysis can help businesses identify which products are frequently purchased together. This information can then be used to develop marketing campaigns that promote complementary products.
* Improve store layout and product placement. Market basket analysis can help businesses understand how customers move through their stores and which products they are most likely to purchase. This information can then be used to improve the store layout and to place products in more prominent locations.
* Optimize inventory levels. Market basket analysis can help businesses identify which products are frequently purchased together and which products are seasonal. This information can then be used to optimize inventory levels and to avoid stockouts.
* Understand customer behaviour. Market basket analysis can help businesses understand how customers shop and what their needs and preferences are. This information can then be used to develop products and services that meet the needs of different customer segments.

**Disadvantage:**

* Requires large amounts of data. Market basket analysis requires a large amount of customer purchase data to be effective. This data can be difficult and expensive to collect.
* Can be complex to analyse. Market basket analysis can be a complex process to analyse, especially for large datasets. This requires specialized skills and knowledge.
* Does not provide insights into why customers purchase certain products together. Market basket analysis can only identify which products are frequently purchased together. It cannot determine the causal relationship between them.
* Privacy concerns. Market basket analysis can raise privacy concerns, as it involves collecting and analysing customer purchase data. Businesses need to take steps to protect customer privacy and to ensure that data is used responsibly.

**Additional insights for Market Basket Analysis:**

* AI-powered market basket analysis involves using machine learning algorithms to automatically discover frequent item sets and association rules in the transactional data.
* Deep learning models can be employed to identify more complex patterns and trends in customer behaviour.
* AI can enable the real-time analysis of customer interactions, allowing retailers to respond quickly to changing trends and customer preferences.
* Natural Language Processing (NLP) can be used to analyse customer reviews and feedback, providing additional insights into product associations and customer sentiment.
* AI-driven recommendation systems can suggest cross-selling opportunities to customers in real-time, improving the overall shopping experience and increasing sales.

**Design Thinking:**

Applying the Design Thinking model with AI:

* Empathize: AI can help gather and analyse customer data at scale to truly understand their

preferences, behaviours, and pain points.

* Define: Use AI to define the problem by identifying specific patterns or associations the retail

business aims to uncover.

* Ideate: AI-driven algorithms can brainstorm potential associations between products and

recommend various strategies for cross-selling.

* Prototype: Develop AI models that implement market basket analysis algorithms (e.g., Apriori,

FP-growth) and customize them to the retail business's needs.

* Test: Deploy AI models on the dataset to analyse customer purchase behaviour and validate the

associations found.

* Implement: Implement AI-driven recommendations and strategies based on insights gained, such

as optimizing product placements or creating personalized product recommendations.

* Iterate: Continuously refine AI models and strategies based on real-time customer data and

feedback.

**Phases of Development:**

The phases of development are,

1. Data collection
2. Data Preprocessing
3. Exploratory Data Analysis (EDA)
4. Market Basket Insights
5. Model Building & Evaluation
6. Visualization
7. Implementation

**Design and Innovation Strategies:**

**Data Collection and Feature Engineering**

Data collection: Collect data on customer transactions, such as the items purchased in each transaction, the time and date of the transaction, and any other relevant information.

One-hot encoding: One-hot encode categorical features, such as product categories or customer demographics. This will convert each categorical feature into a set of binary features, where each feature represents a possible value of the categorical feature.

Feature aggregation: Aggregate features to create new features that capture more information about the data. For example, you could create a feature that represents the total number of items purchased in each transaction.

Feature selection: Select the features that are most relevant to the market basket analysis task. This can be done using a variety of techniques, such as correlation analysis or mutual information.

**Data Pre-processing**

Clean and preprocess the data, removing any irrelevant information, handling missing values, and converting the data into a suitable format for analysis. This may involve converting the data to a binary format, where each item in a transaction is represented as a 1 or 0.

**Model Selection and Training**

Choose a model. There are a variety of machine learning models that can be used for market basket analysis, such as the Apriori algorithm, FP-growth algorithm, and association rule mining. The best model to use will depend on the specific dataset and the desired results.

Prepare the data. The data should be cleaned and pre-processed to ensure that it is in a format suitable for analysis. This may involve removing duplicate transactions, correcting errors in the data, and converting the data to a binary format.

Train the model. Once the data is prepared, the model can be trained on the dataset. This involves feeding the model the data and allowing it to learn the patterns and associations between products.

Evaluate the model. Once the model is trained, it should be evaluated on a holdout dataset to assess its performance. This can be done by calculating the accuracy, precision, recall, and F1 score of the model.

Deploy the model. Once the model is evaluated and found to be performing well, it can be deployed to production. This may involve integrating the model into a retail business's existing systems or developing a new application.

**Geographic Analysis**

Segment the customer base by geographical location. This can be done using customer addresses, postal codes, or other geospatial data.

Perform market basket analysis on each customer segment. This will identify the most frequently purchased items and item sets in each geographical location.

Compare the market basket analysis results across geographical locations. This will identify any regional differences in customer purchasing behaviour.

Use the market basket analysis results to develop targeted marketing campaigns and promotions for each geographical location.

For example, a retailer may find that customers in a particular region are more likely to purchase milk and bread together than customers in other regions. The retailer could then target this region with a promotion on milk and bread.

**Market Sentiment Analysis**

Market sentiment analysis can be used to understand customer purchasing behaviour and identify potential cross-selling opportunities by uncovering hidden patterns and associations between products. This can be done by using market basket analysis to identify the most frequently purchased items and item sets, as well as the rules that govern how items are purchased together. Once the association rules have been identified, they can be used to identify potential cross-selling opportunities.

**Explainable AI (XAI)**

Visualizations: XAI can be used to create visualizations that show how products are purchased together. This can help to identify patterns and relationships that would be difficult to see in the raw data.

Rule explanations: XAI can be used to generate explanations for the association rules that are identified. This can help to understand why certain products are purchased together and how this information can be used to improve the business.

Counterfactual reasoning: XAI can be used to perform counterfactual reasoning to explore how the association rules would change if certain conditions were different. This can help to identify the key factors that drive the associations.

**Continuous Learning**

Collect and analyse new data on a regular basis to identify emerging patterns and associations between products. Use machine learning algorithms to develop predictive models that can forecast customer demand and identify potential cross-selling opportunities.

A screenshot of a diagram

Description automatically generated

Data collection

Feature Engineering

Geographical Analysis

Model Training

Data Pre-processing

Continuous Learning

Improvement & Case Study

Explainable AI

Market Sentimental Analysis

**Dataset Used:**

<https://www.kaggle.com/datasets/aslanahmedov/market-basket-analysis>

**Dataset Description:**

A "Market Basket Analysis" dataset typically consists of transaction records from a retail store or e-commerce platform. These records capture information about the items purchased by customers during their shopping trips. The dataset may include the following information:

1.Bill ID: A unique identifier for each shopping transaction.

2.Product Name: Identifiers or names of the products that were part of the transaction.

3.Quantity: The quantity of each product purchased in each transaction.

4.Date: The date and time when the transaction occurred.

5.Customer ID: An identifier for individual customers, if available.

6.Store Location: The location of the store or branch where the transaction took place.

**Data Preprocessing Steps:**

Before loading the data, the following preprocessing steps can be used to obtain error free data. The steps are,

1. Data cleaning
2. Data Transformation
3. Data Reduction
4. Data Formatting
5. Handling Imbalanced Data
6. Data Partitioning

**Data cleaning:**

Identify and handle missing values in the dataset. You can either remove records with missing values or impute missing values using techniques like mean, median, or mode imputation.

**Data Transformation:**

Convert categorical variables, like product names or categories, into binary (0/1) columns. This allows the machine learning algorithms to work with categorical data.

**Data cleaning:**

Reduce the size of the dataset if it's large. You may sample a subset for initial analysis or aggregate data to a higher level if fine-grained details are not necessary**.**

**Data Formatting:**

Ensure that the dataset is in the correct format for market basket analysis. This includes making sure that each record represents a transaction with associated items.

**Handling Imbalanced Data:**

Check for imbalances in the dataset, where some items or combinations of items may be rare. You might need to oversample or under sample to balance the dataset.

**Data Partitioning:**

To evaluate model, consider splitting the data into training and testing sets. This is particularly important when developing predictive models for market basket insights**.**

**Choice of Machine Learning Algorithm:**

In this Market Basket Insights, Logistic Regression is used as a Machine Learning Algorithm to predict binary outcomes, such as whether a specific product or combination of products is purchased in a transaction.

**Logistic regression:**

Logistic regression is a fundamental machine learning algorithm employed in Market Basket Insights to predict binary outcomes, particularly in the context of whether a specific product or a combination of products is purchased in a transaction.

**Key Characteristics:**

1. Binary Classification: Logistic regression is well-suited for binary classification tasks, where the goal is to predict a yes/no or true/false outcome. In Market Basket Insights, this can be applied to answer questions like whether a customer will purchase a specific product or not in each transaction.
2. Probabilistic Model: Unlike linear regression, which predicts continuous values, logistic regression models the probability of an event occurring. It calculates the likelihood of a binary outcome, typically expressed as a probability between 0 and 1.
3. Sigmoid Function: Logistic regression uses the sigmoid function to map the output of the algorithm to a probability value. This function ensures that the predicted probabilities fall within the appropriate range.

**Applications in Market Basket Insights:**

In Market Basket Analysis, logistic regression can be used in various ways:

1. Association Rule Mining: Logistic regression can help identify associations between products in a customer's transaction history. It can quantify the likelihood of one product being purchased given the presence or absence of another product.
2. Cross-Selling Opportunities: By analysing historical data, logistic regression can identify potential cross-selling opportunities. For example, it can determine the probability of a customer buying product A if they've already purchased product B.
3. Customer Segmentation: Logistic regression can be applied to segment customers based on their purchase behaviour. It can help classify customers into categories, such as high-probability buyers and low-probability buyers, allowing businesses to tailor marketing strategies accordingly.
4. Customer Churn Prediction: It can be used to predict the likelihood of a customer churning (stopping their purchases) based on their recent buying behaviour, enabling businesses to take proactive measures to retain customers.

In Market Basket Insights, logistic regression plays a pivotal role in making data-driven decisions by quantifying the likelihood of certain purchase behaviours. It allows businesses to not only understand customer preferences and associations but also to take targeted actions to enhance sales and customer satisfaction.

**Platform: Google Collab**

**Step by step Implementation & Screenshot:**

**Step 1: Import Packages & Load Dataset**

A screenshot of a computer

Description automatically generated

**Step 2: Data Preparation**

A screenshot of a computer

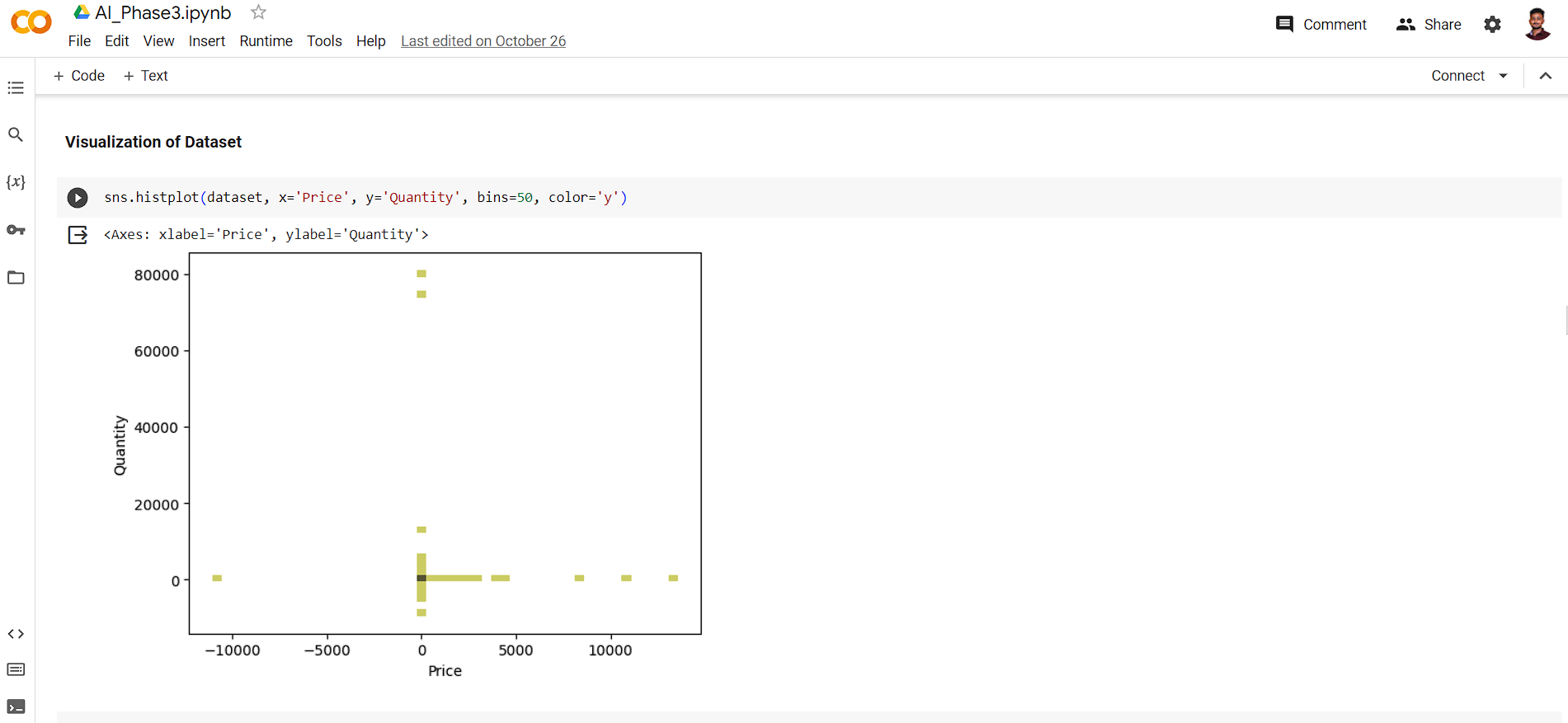
Description automatically generated

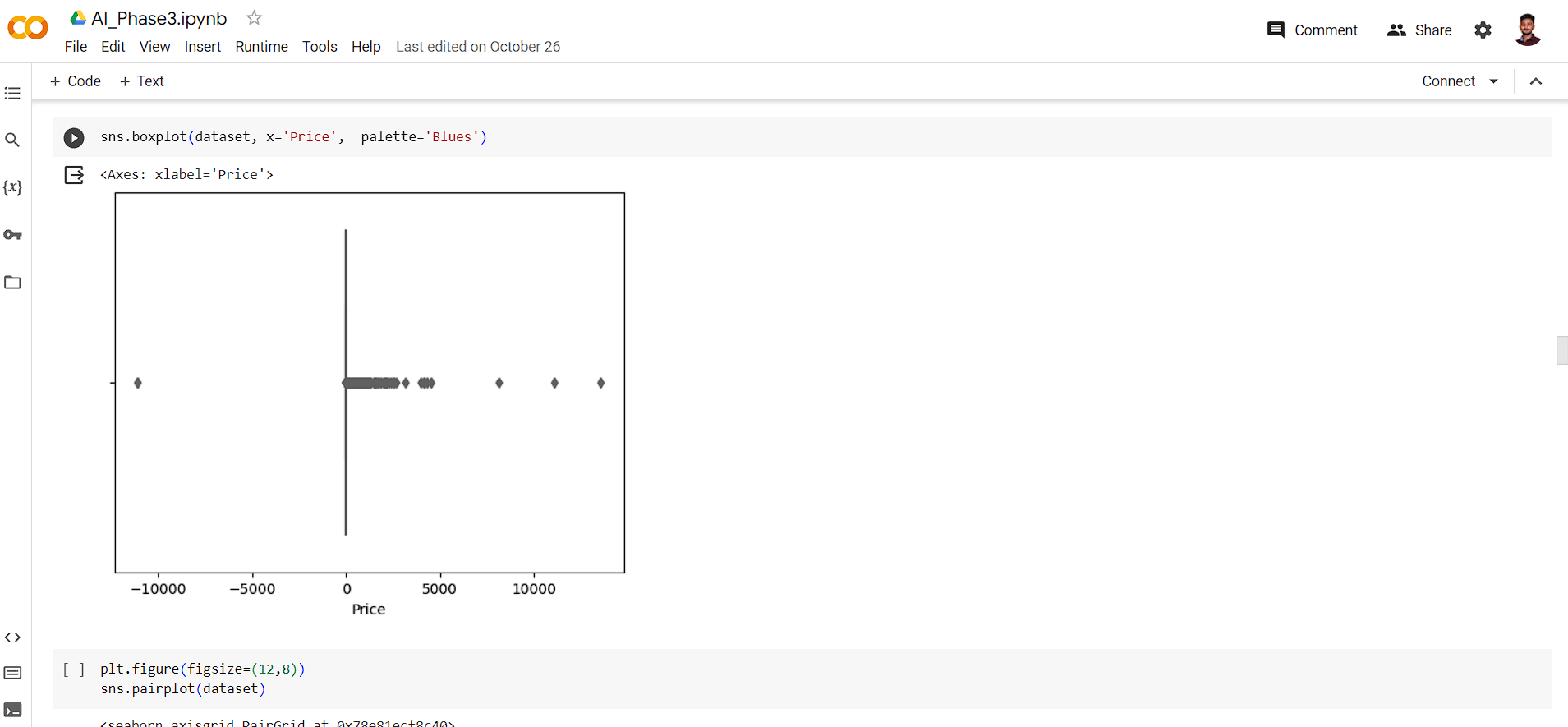
**Step 3: Training Model**

A screenshot of a computer

Description automatically generated

**Step 4: Visualization of Dataset**





A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Step 5: Machine Learning Algorithm (Logistic Regression)**

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

**Conclusion:**

In conclusion, the successful creation and deployment of Market Basket Insights using Artificial Intelligence has been completed successfully providing various insights on the consumer behaviour & analysis of the market. In conclusion, the project on Market Basket Analysis using Artificial Intelligence with a Kaggle dataset represents a significant contribution to understanding customer behaviour and optimizing retail strategies. This analysis leverages the power of AI and machine learning to extract valuable insights from transaction data, enabling businesses to make informed decisions. sss