Programming Lab – Project Assignment

Development of an Interactive Data-Analytics Tool

Prof. Dr. Sebastian Bremm

Submitted by:

Aiman Ashfaq
Aya Zirari
Thi Diem Quynh Tran (1449369)
Mohamed Akhouaji

Summer semester 2024

Study program: Wirtschaftsinformatik (B.Sc.)

Outline

1. Introduction

1.1 Project Overview

The main objective of this project was to create an interactive web dashboard using the Python Django framework and ECharts, Chart.js libraries for front-end interactivity. The purpose of the web tool is to provide business managers with insightful data visualizations that support informed decision-making. This tool was developed as part of a programming lab project at university, with the target audience being professors simulating the role of business managers.

1.2 Scope

The project focuses on analyzing sales data, customer distribution, product performance, and store operations across multiple years. It aims to provide managers with a comprehensive overview of key metrics to inform strategic business decisions. The tool covers functionalities such as year-over-year sales comparisons, product popularity analysis, customer distribution mapping, and order distribution by distance. Technologies used include Python, Django, AJAX, and data visualization libraries like Echarts, Chart.js and Leaflet.

1.3 Unique Features

One of the standout features of this tool is its integration of multiple data visualizations and interactive elements, allowing users to explore data dynamically. The interactivity enables managers to drill down into specific datasets, adjust parameters, and gain deeper insights without navigating away from the dashboard. This dynamic data exploration is designed to make the tool not only informative but also engaging and user-friendly.

1.4 Main Functions

The dashboard is divided into four main pages, each focusing on a different aspect of the data:

- **Homepage:** Provides key metrics on overall sales for all years, offering a quick overview of the business's performance.
- **Products Page:** Focuses on product performance, including charts on product sales, size popularity, and ingredient usage.
- **Customers Page:** Analyzes customer distribution and behavior, providing insights into customer demographics and purchasing patterns.
- **Stores Page:** Evaluates store performance with metrics such as total revenue by store, order distribution by distance, and heat maps of customer locations.

1.5 User Experience

The dashboard was designed with simplicity and ease of use in mind, ensuring that users can navigate and understand the data with minimal effort. The layout is intuitive, with clearly labeled charts and interactive elements that provide immediate feedback. By prioritizing user experience, the tool ensures that managers can quickly access the information they need to make informed decisions without being overwhelmed by complex data representations.

2. Installation Guideline

2.1 Prerequisites

Before you begin, ensure you have met the following requirements:

- Python 3.12.3
- pip (Python package installer)
- Virtualenv (optional but recommended)
- Git (to clone the repository)
- MySQL server and client installed

2.2 Installation Steps

Step 1: Clone the Repository

First, clone the repository from GitHub to your local machine:

```
git clone https://github.com/thidiemquynhtran/ProgLab.git
cd ProgLab
```

Step 2: Set Up a Virtual Environment (Optional)

It's a good practice to use a virtual environment to manage dependencies. You can create one using `virtualenv`:

```
python -m venv venv
```

Activate the virtual environment:

- On Windows:

  ```
  venv\Scripts\activate
  ```

- On macOS/Linux:

  ```
  source venv/bin/activate
  ```

Step 3. Install Dependencies

With the virtual environment activated, install the project dependencies using pip:

```
pip install -r requirements.txt
```

Step 4: Download the SQL Dump File

Download the SQL dump file from the provided link: Download dbpizatest.sql

Or from this link:
https://drive.google.com/file/d/13qsSLJHpEWWur3RW0bRm522sH7GeiLSR/view?usp=sharing

Step 5: Configure and Load the Database

Ensure your MySQL server is running, and you have created a database for the project. You can create a database using the MySQL client:

```
mysql -u root -p
```

In the MySQL prompt, run:

```
CREATE DATABASE dbpizatest;
CREATE USER 'youruser'@'localhost' IDENTIFIED BY 'yourpassword';
GRANT ALL PRIVILEGES ON yourdatabase. TO 'youruser'@'localhost';
FLUSH PRIVILEGES;
EXIT;
```

If you have a database dump file (`dbpizatest.sql`), you need to load it into MySQL. Use the following command:

```
mysql -u root -p dbpizatest < dbpizatest.sql
```
Or:
```
mysql -u youruser -p dbpizatest < dbpizatest.sql
```

Step 6: Migrate the database

To initialize the database, run the following commands:

```
python manage.py migrate
```

Step 7: Run the Development Server

Now you can start the development server:

```
python manage.py runserver
```

Step 8: Access the Dashboard

Open your web browser and navigate to `http://127.0.0.1:8000/data_analysis/index.html` to see the project running.

With the project running, you can start analyzing your data. Refer to the user guide for detailed instructions on how to use the tool effectively.

3.Database & Backend

3.1 Database Schema:

The project uses a relational database to store and manage the data needed for the dashboard. The schema includes several key tables that organize data related to orders, products, customers, and stores.

3.2 Table Descriptions:

- **SalesSummary Pie:**
    - **Purpose:** Stores aggregated sales data by product category, year, and month for the pie chart visualizations.

    - **Columns:** Name (product category), Year, Month, Revenue.
- **TotalSalesByMonthBar:**
    - **Purpose:** Holds total sales revenue data aggregated by year and month for bar chart visualizations.
    - **Columns:** Year, Month, Revenue.
- **order_distances:**
    - **Purpose:** Contains data about the distance between customers and stores for analyzing order distributions by distance.
    - **Columns:** orderID, customerID, storeID, orderDate, total, customer_lat, customer_lon, store_lat, store_lon, distance.
- **order_distance_aggregates:**
    - **Purpose:** Stores aggregated data of orders by rounded distance for better visualization.
    - **Columns:** distance_rounded, order_count, avg_order_value.
- **store_revenue_items:**
    - **Purpose:** Aggregates total revenue and total items sold by store for scatter plot visualizations.
    - **Columns:** storeID, total_revenue, total_items.
- **product_size_popularity:**
    - **Purpose:** Tracks the popularity and revenue of different product sizes for analysis.
    - **Columns:** product_size, total_sales, total_revenue.

3.3 Database Management:

- **SQL-Scripts:**
    - SQL scripts are used to create and populate the tables, ensuring that the data is correctly structured for analysis and visualization.
    - Example: The script for creating and inserting data into SalesSummaryPie aggregates revenue by product category and date.

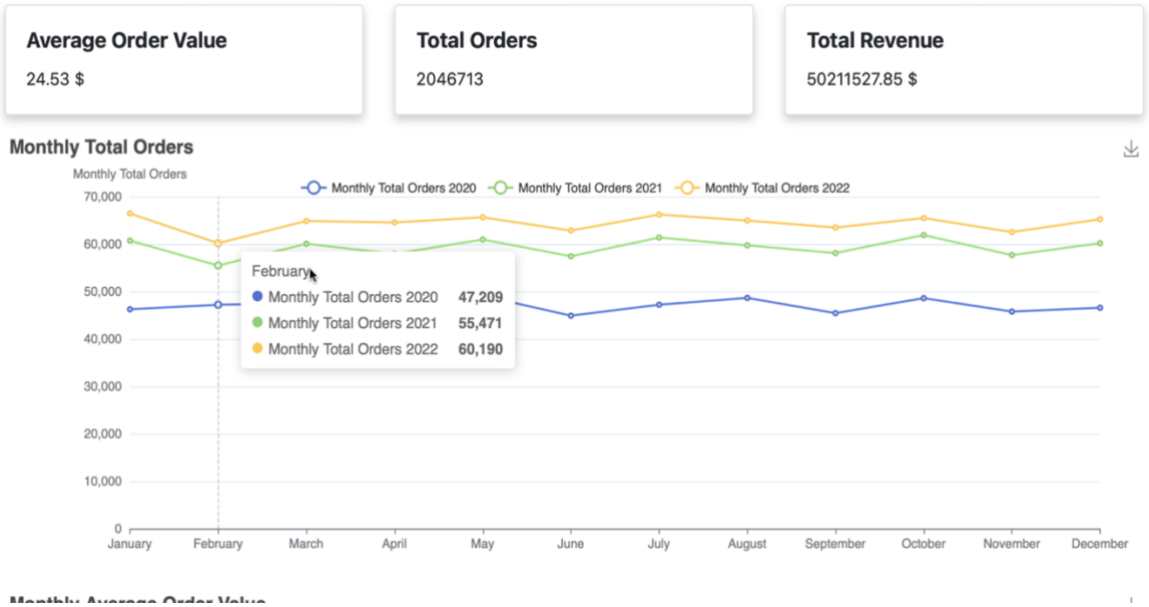
3.4 Backend Development:

- **Django Models:**
    - Django ORM is used to define models that map to the database tables, facilitating data retrieval and manipulation.
    - Example: The SalesSummaryPie model in Django corresponds to the SalesSummaryPie table in the database.
- **Data Processing:**
    - Data is processed and aggregated using SQL queries before being visualized in the dashboard. This ensures efficient handling of large datasets and quick response times for user interactions.
- **APIs:**
    - Django views and RESTful APIs are implemented to serve the processed data to the frontend, allowing for dynamic updates and interactions.

4. Frontend Description

4.1 Homepage

Homepage Analysis



Relevance of Key Metrics

Average Order Value (AOV):

Importance: AOV indicates the average amount spent by customers per order. It's a crucial metric for understanding customer purchasing behavior and the effectiveness of pricing strategies.
Business Implications: A higher AOV can lead to increased profitability without needing to acquire more customers. Strategies such as upselling, cross-selling, and offering bundles can be utilized to increase the AOV.

Total Orders:

Importance: The total number of orders reflects the volume of transactions over a period. It provides insight into customer demand, sales performance, and market penetration.
Business Implications: Monitoring the number of orders helps in understanding sales trends and identifying peak periods. This information is vital for inventory management, staffing, and ensuring customer satisfaction during high-demand times.

Total Revenue:

Importance: Total revenue is the sum of all sales over a specific period and is a direct indicator of business performance and growth.

Business Implications: Revenue trends help in assessing the overall financial health of the business. It also aids in strategic planning, budget allocation, and investment decisions to drive further growth.

Monthly Total Orders Analysis

2020 Analysis:

- Orders are generally lower compared to following years.
- Notable dip in February with 47,209 orders.
- Relatively flat trend throughout the year, indicating minimal seasonal impact.

2021 Analysis:

- Increase in the number of orders compared to 2020.
- Higher number of orders in February with 55,471.
- Fairly consistent orders throughout the year, peaking around March and April.

2022 Analysis:

- Highest number of orders among the three years.
- February shows 60,190 orders, indicating growth compared to previous years.
- Steady and slightly increasing number of orders towards the end of the year.

Insights and Implications

Growth Trend:

- Clear upward trend in the number of orders from 2020 to 2022.
- Indicates successful customer retention and acquisition strategies.

Stable Revenue Generation:

- A steady average order value of $24.53 and increasing total orders ensure stable revenue growth.
- Total revenue reflects strong sales performance and effective pricing strategies.

Seasonal Stability:

- Monthly total orders show minimal seasonal fluctuations, indicating consistent demand throughout the year.
- Reduces dependency on seasonal peaks.

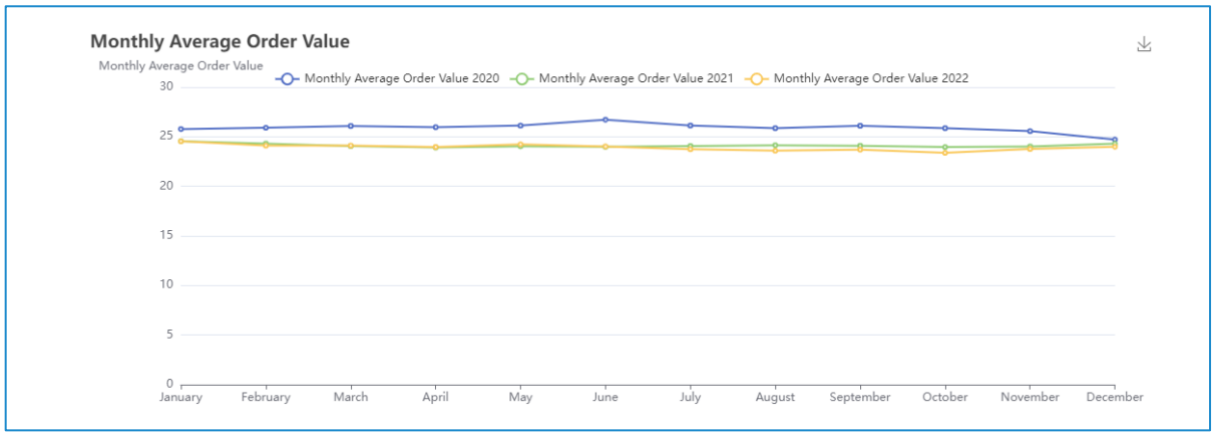
Predictive Insights

Future Order Growth:

- Likely continuation of the increasing number of orders based on current trends.
- Predictive models can forecast monthly orders, aiding in better resource and inventory planning.

Revenue Projections:

- Steady increase in orders and consistent average order value suggest proportionate revenue growth.
- Future revenue projections can be more accurately made using historical data



and trend analysis.

Monthly Average Order Value

2020 Analysis

- **Monthly Average Order Value**: Approximately $25.
- **Trend**: The AOV remains relatively stable throughout the year, with slight fluctuations.
- **Insights**:
  - Consistency in AOV indicates stable customer spending behavior.
  - Slight increases in certain months suggest opportunities for targeted promotions to boost sales further.

2021 Analysis

- **Monthly Average Order Value**: Consistently around $24.
- **Trend**: The AOV shows a consistent pattern similar to 2020 but is slightly lower.
- **Insights**:
  - Stability in AOV suggests steady customer spending habits.
  - The slight decrease compared to 2020 might indicate the need to review pricing strategies or market conditions.

2022 Analysis

- **Monthly Average Order Value**: Starts around $24 and shows a slight upward trend towards the end of the year.
- **Trend**: The AOV is consistent but shows a minor increase in the last few months.
- **Insights**:
  - The upward trend towards the end of the year indicates improving customer spending behavior.
  - The consistent AOV throughout the year suggests stable pricing and customer satisfaction.

Recommendations:

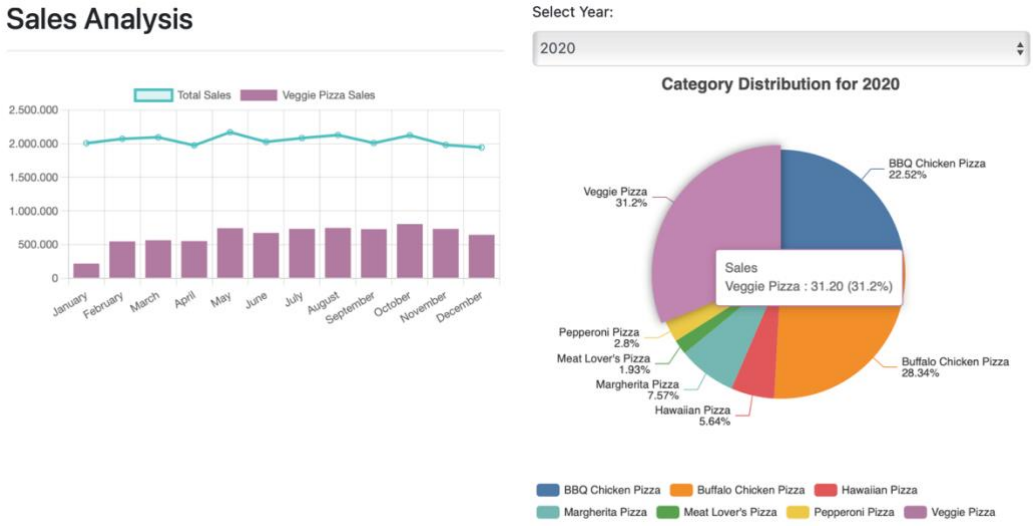
- Evaluate pricing strategies to identify potential areas for improvement.
- Introduce value-added services or bundle offers to encourage higher spending per order.
- Monitor market trends to understand external factors affecting AOV.

Summary for the Manager

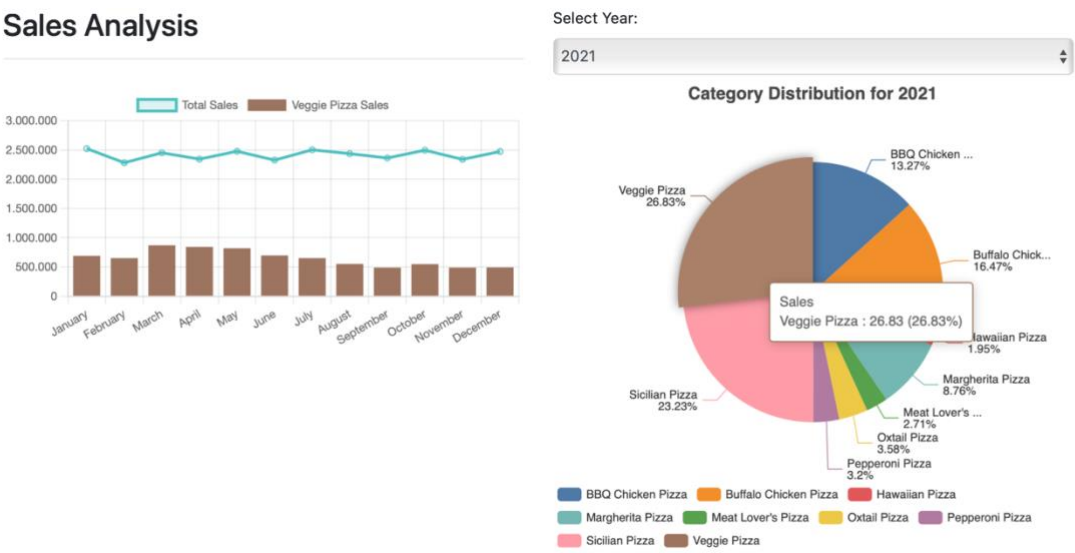
The analysis of the Monthly Average Order Value over the three years provides valuable insights into customer spending behavior. By maintaining a stable pricing strategy, exploring opportunities for value-added services, and leveraging periods of higher AOV for targeted marketing, the manager can optimize revenue and ensure sustained business growth.

4.2 Products

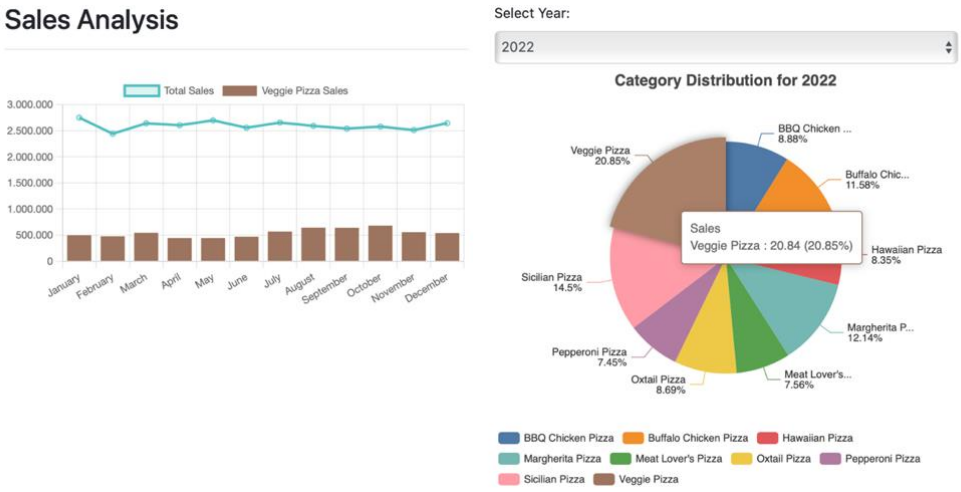
Data Analysis and Insights for 2020-2022



- **Total Sales:**
  - Sales remained consistent around 2,000,000 to 2,500,000, with a slight dip in February and a peak in April.
- **Veggie Pizza Sales:**
  - **2020:** Veggie Pizza sales were steady across the months, indicating consistent popularity.
- **Category Distribution:**
  - **2020:**
    - Veggie Pizza: 31.2%
    - Buffalo Chicken Pizza: 28.34%
    - Sicilian Pizza: 22.52%
    - Other varieties like Margherita, Hawaiian, and BBQ Chicken Pizzas held moderate shares.



- **Total Sales:**
  - Sales maintained a stable range like 2020, indicating a strong customer base.
- **Veggie Pizza Sales:**
  - **2021:** Veggie Pizza continued to show strong sales but was joined by significant sales of Sicilian and Margherita pizzas.
- **Category Distribution 2021**
  - **2021:**
    - Veggie Pizza: 26.83%
    - Sicilian Pizza: 23.23%
    - Buffalo Chicken Pizza: 16.47%



- **Total Sales:**
  - Total sales showed consistent performance, reflecting stable demand and strong brand loyalty.
- **Veggie Pizza Sales**
  - **2022:** Veggie Pizza remained prominent, but the sales distribution among different pizzas became more balanced.

- **Category Distribution 2022**
  - Veggie Pizza: 20.85%
  - Sicilian Pizza: 14.5%

- Margherita Pizza: 12.14%

Comparative Insights

1. **Sales Stability and Consistency:**
    - Over the three years, total sales have remained remarkably consistent. This stability indicates a loyal customer base and a steady demand for pizza products, unaffected significantly by external factors.
    - Minor fluctuations in total sales suggest that the business is not significantly affected by seasonal trends or external disruptions.
2. **Shifts in Pizza Preferences:**
    - **Veggie Pizza:** Although Veggie Pizza remains the most popular choice, its market share has been gradually decreasing from 31.2% in 2020 to 20.85% in 2022. This shift indicates that while Veggie Pizza is still favored, customers are increasingly trying other varieties.
    - **Sicilian Pizza:** The share of Sicilian Pizza increased from 22.52% in 2020 to 23.23% in 2021 but saw a slight decrease to 14.5% in 2022. This trend suggests initial growth in popularity, followed by stabilization.
    - **Buffalo Chicken Pizza:** There was a notable decrease in its share from 28.34% in 2020 to 16.47% in 2021, and further to 11.58% in 2022. This decline suggests changing customer preferences away from this variety.
    - The emergence of varieties like Margherita and Oxtail Pizzas shows potential growth areas and opportunities for targeted promotions.
3. **Diverse Product Portfolio:**
    - The distribution of sales among different pizza types has become more balanced over the years. This diversification indicates a healthy mix of product offerings that caters to varying customer tastes.
    - By diversifying the pizza offerings, the business can cater to a wider range of preferences, enhancing customer satisfaction and loyalty.
4. **Market Strategy Implications:**
    - **Promotions and Marketing:** To leverage the consistent popularity of Veggie Pizza, targeted marketing campaigns can be designed to further solidify its position as a signature product. Additionally, promoting emerging favorites like Margherita and Sicilian pizzas could attract new customers.
    - **Product Development:** The gradual shift in preferences suggests a need for continual product innovation. Introducing new flavors or limited-time offerings could maintain customer interest and drive sales.
    - **Inventory Management:** Understanding the stable yet shifting demand allows for better inventory planning. Ensuring adequate stock of popular varieties while also accommodating new and emerging favorites can optimize inventory turnover and reduce waste.
5. **Strategic Planning:**
    - **Long-Term Trends:** The ability to filter and analyze data by year provides valuable insights into long-term trends. These insights can inform strategic decisions regarding product development, market expansion, and promotional activities.
    - **Customer Engagement:** Maintaining high levels of customer engagement through consistent quality, new product offerings, and regular promotions can sustain and grow the customer base.

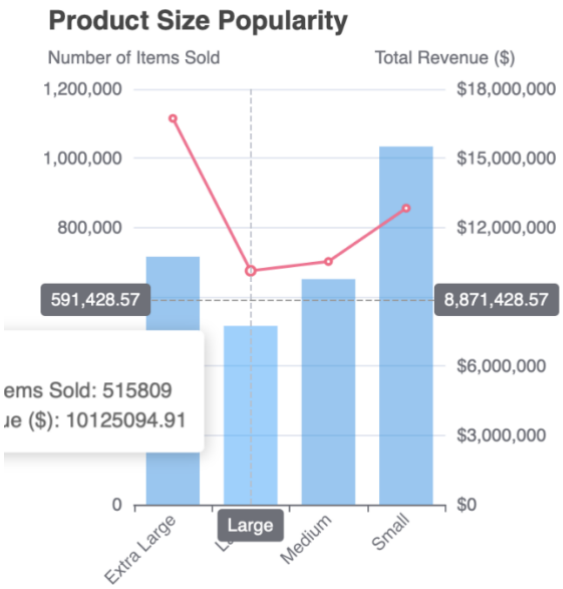
Future Predictions

1. **Continued Popularity of Veggie Pizza:**
    - Veggie Pizza is expected to remain a top seller, though its market share may continue to slightly decrease as customers explore other varieties.
2. **Growth in Diverse Pizza Varieties:**
    - Varieties like Margherita and Oxtail Pizzas may see increased sales as their market shares have shown growth potential.
    - The balanced distribution among different pizza types suggests that introducing new flavors could be well-received by customers.
3. **Stable Sales Trend:**
    - Given the consistent sales performance over the past three years, it is likely that the business will continue to experience stable demand.
    - Regular promotions and new product introductions will be key to maintaining and potentially increasing this stable trend.
4. **Enhanced Customer Engagement:**
    - Engaging with customers through targeted marketing, promotions, and high-quality offerings will be crucial for sustaining loyalty and attracting new customers.
    - Data-driven decisions will help in timely adjustments to inventory and marketing strategies, ensuring that the business remains aligned with customer preferences and market trends.

Summary

The analysis of sales data from 2020 to 2022 reveals a stable market with a loyal customer base, alongside evolving preferences towards a more diverse range of pizza offerings. By leveraging these insights, businesses can enhance their strategic planning, optimize inventory management, and drive marketing efforts to align with customer trends, ensuring sustained growth and customer satisfaction.

Product Size Popularity Analysis



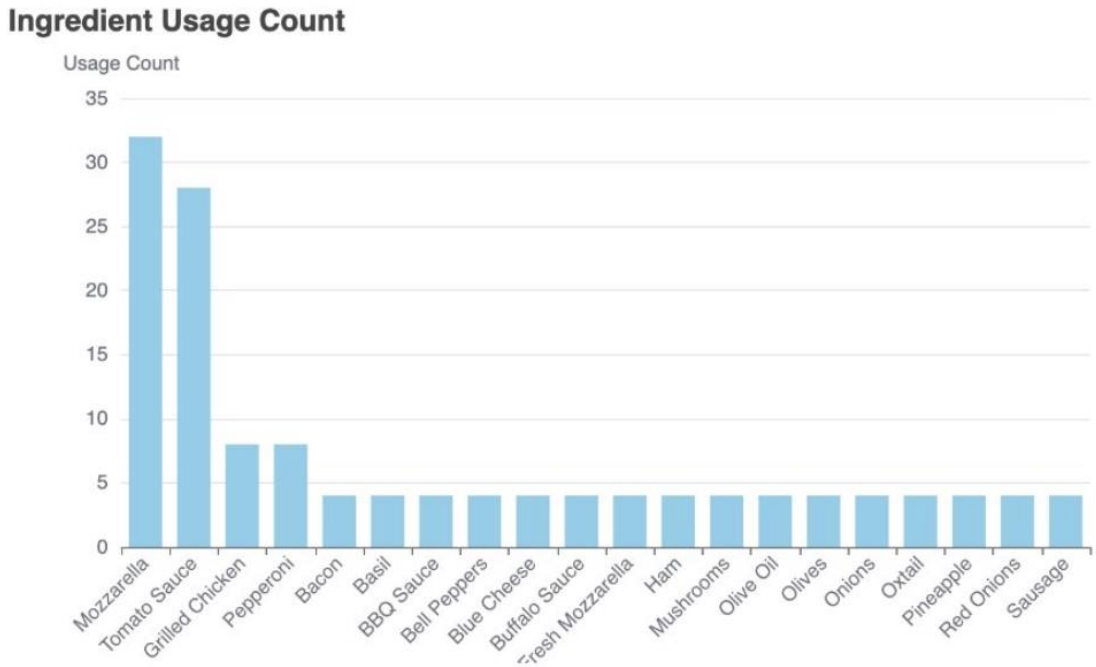
Items Sold and Revenue:

- **Small Size:** Highest number of items sold, around 1,041,428.57 units, generating the highest revenue of approximately $15,621,428.57.
- **Extra Large Size:** Lowest number of items sold but generates significant revenue, indicating a higher price point.
- **Medium and Large Sizes:** Moderate number of items sold with substantial revenue, showing balanced popularity.

Insights:

- Small-sized pizzas are the most popular, indicating customer preference for this size, potentially due to price or portion suitability.
- Extra-large pizzas, while not sold in high numbers, contribute significantly to revenue, suggesting a premium pricing strategy.
- Balanced sales of medium and large sizes indicate diverse customer preferences, providing opportunities for targeted promotions based on size preferences.

Ingredient Usage Analysis



Ingredient Usage Count:

1. **Mozzarella:** Highest usage count, indicating it is a key ingredient in most pizzas.
2. **Tomato Sauce and Grilled Chicken:** Also have high usage, essential in various pizza varieties.
3. **Pepperoni and Bacon:** Moderate usage, popular toppings.
4. **Other Ingredients:** Basil, BBQ Sauce, Bell Peppers, Blue Cheese, Buffalo Sauce, etc., have lower but consistent usage, showing diverse ingredient choices across different pizzas.

Insights:

- Mozzarella and Tomato Sauce are core ingredients, critical for inventory management.
- High usage of Grilled Chicken and Pepperoni indicates their popularity as toppings.
- Consistent usage of a variety of ingredients highlights the diversity in pizza offerings and the need to maintain a well-stocked inventory of all ingredients.

Summary and Strategic Insights

The consistent sales figures over three years show a stable market with loyal customers. Shifts in pizza preferences, with Veggie Pizza gradually losing share while other varieties gain, suggest evolving customer tastes. Balanced sales distribution among pizza sizes and varied ingredient usage indicates diverse customer preferences, offering opportunities for targeted marketing and product development. By leveraging these insights, the business can enhance strategic planning, optimize operations, and drive growth through informed decision-making.

4.3 Stores

Monthly Total Sales of All Stores

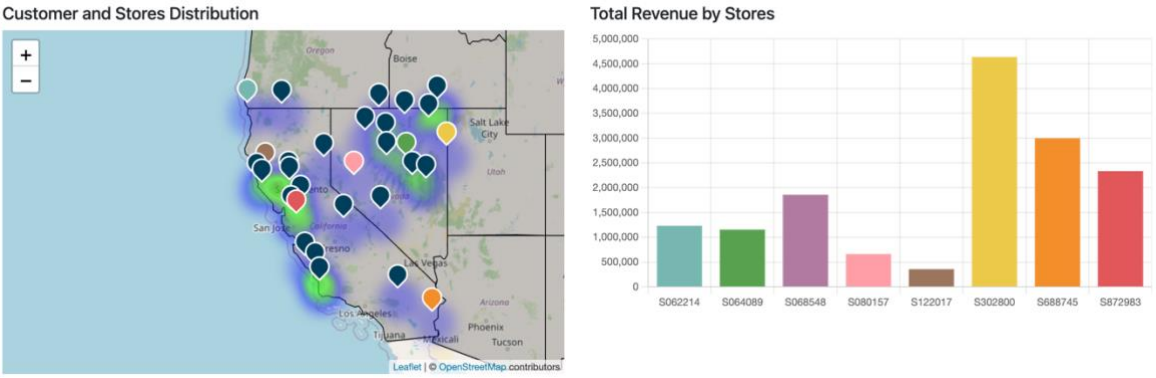
1. **Trend Analysis:**

    - **2020:** Total sales are the lowest among the three years, with a noticeable dip in April.
    - **2021:** Sales in 2021 are higher than in 2020 with a steady trend and slight fluctuations, indicating a recovery
    - **2022:** This year shows the highest sales, consistently outperforming the previous years.
    - December Peak: All three years show a peak in sales during December, likely due to the holiday season, which typically drives higher consumer spending.
    - Mid-Year Dips: Slight dips in sales occur around April and July, likely due to seasonal slowdowns, or less promotional activity during these months.
2. **Insights and Recommendations**

    - Implement targeted marketing campaigns and promotions during the holiday season to maximize sales.

- Develop strategies to boost sales during the mid-year dip periods, such as special promotions or new product launches.

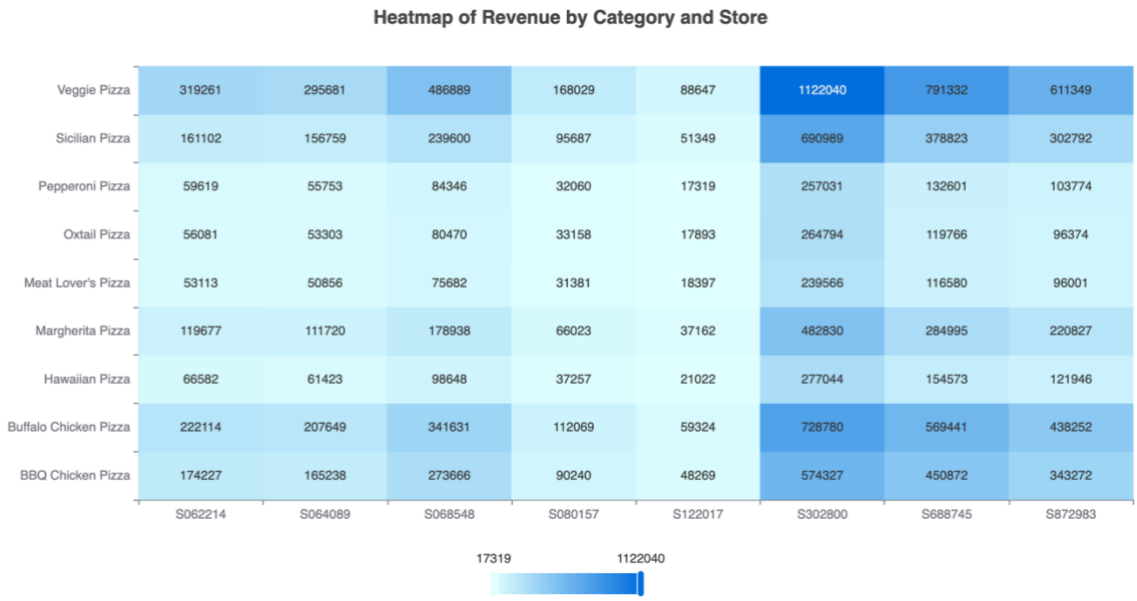
Customer and Stores Distribution: This map shows the geographical distribution of customers and stores, with a heatmap overlay indicating the density of customer locations. The interactive feature allows users to click on location markers on the map to view the corresponding store's revenue in the bar chart. This helps in quickly identifying which stores are performing well and which are not, and where they are located based on their geographical heatmap on the left.



1. **Customer Density:**
    - High-density areas are shown in purple, indicating regions with the most customers.
    - California has the most customers, particularly in urban areas like San Francisco, Los Angeles, and San Diego.
2. **Store Locations:**
    - Stores are mostly situated in high-density customer areas, showing effective market coverage.
    - Areas like Nevada and parts of Utah have fewer stores, each state has only one store.
3. **Total Revenue by Stores**
    - Stores with yellow and orange colors generate double sales more than other stores in California and Nevada. So, we can consider expanding more stores into Utah and Arizona as these areas have lower store density but high customer potential.
    - Top-performing stores, they have higher revenue likely due to high customer density, effective management, or favorable location.
    - Low-performing stores may need targeted strategies to boost performance, such as local marketing or operational improvements.

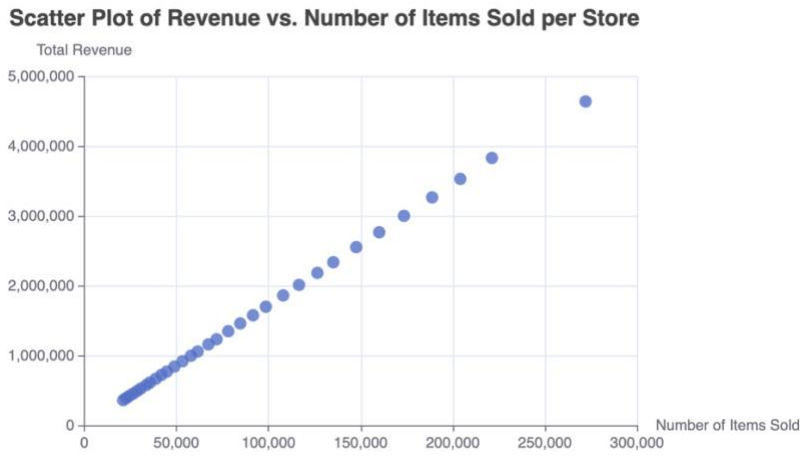
Heatmap of Revenue by Category and Store

This heatmap displays the revenue generated by various pizza categories across different stores. The stores in this heatmap correspond to the ones in the previously discussed bar chart. So, we can go into detail to analyze how each category performs in each store.



1. **General Analysis:**
    - Veggie Pizza and Buffalo Chicken Pizza generate the most revenue across most stores.
    - Sicilian, BBQ and Margherita Pizza performs well across several stores,
    - Pepperoni, Oxtail and Meat Lover's Pizza show lower sales compared to Veggie and Sicilian Pizza.
    - Some stores like S080157 and S112017 generate low revenue across all categories, indicating potential issues that need addressing.
    - Store S302800 dominates in almost all categories, especially Veggie Pizza, Sicilian Pizza, and Buffalo Chicken Pizza, suggesting a strong overall performance.
2. **Insights and Recommendations:**
    - It is recommended to promote popular categories like Veggie Pizza and Buffalo Chicken Pizza by introducing new variants or combo deals.
    - We can boot sales for lower-performing categories like Pepperoni and Oxtail, Pizzas with special promotions or limited-time offers.

Scatter Plot of Revenue vs. Number of Items Sold per Store



This scatter plot represents the relationship between the total revenue and the number of items sold per store. Each point on the plot corresponds to a store, with the x-axis showing the number of items sold and the y-axis representing the total revenue.

1. **General Analysis:**

    - There is a clear positive correlation between the number of items sold and the total revenue. As the number of items sold increases, the total revenue also increases. This indicates that higher sales volumes directly contribute to higher revenues.

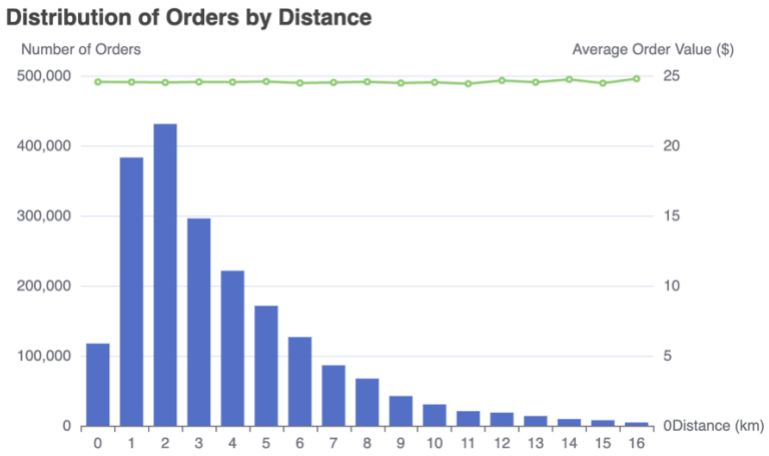
    - The stores with the highest revenue, nearing 5 million $, are also those that sell the most items, which is the one in the state of Utah. This store is benefiting from high traffic, effective marketing, and possibly a diverse product range that attracts more customers.

    - The middle segment of the scatter plot shows stores with moderate performance, both in terms of revenue and the number of items sold. And at the lower left segment are low-performing stores.
2. **Insights**

    - Implement **strategies to increase the number of items sold** in mid-performing stores such as special promotions, loyalty programs, and bundling of popular items to encourage higher purchase volumes. They could help move them towards the top-performing segment.

Distribution of Orders by Distance

This bar chart represents the distribution of orders based on the distance from the store, while the line chart shows the average order value ($) across these distances.



1. **Order Distribution by Distance:**
    - The number of orders is highest, within the 0-2 km range from the store. Orders peak at 2 km with over 400,000 orders.
    - As the distance increases, the number of orders decreases significantly. Orders drop sharply beyond 5 km.

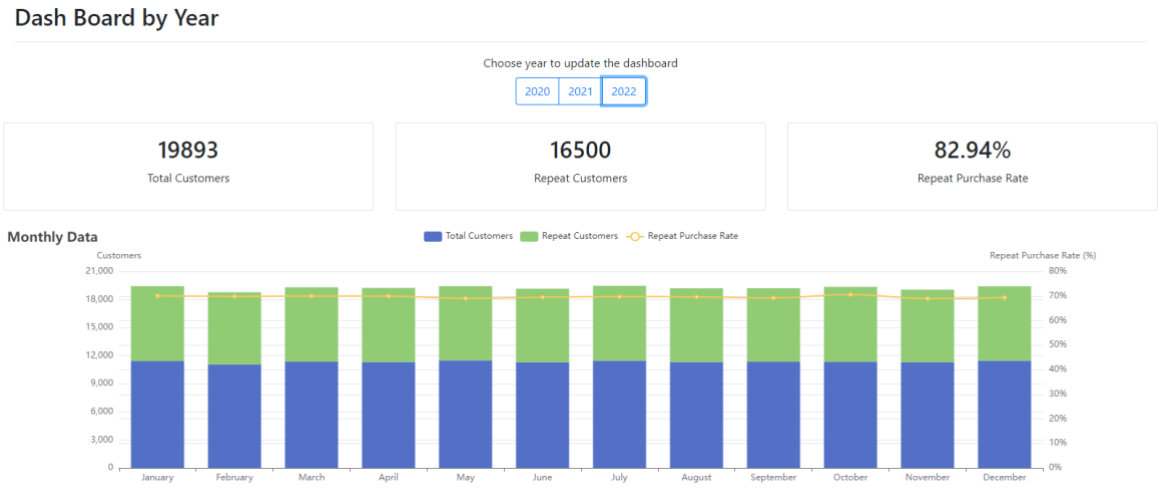
- There are very few orders beyond 10 km, indicating a strong preference for customers to order from nearby stores.

2. **Average Order Value:**
   - The average order value remains relatively constant across all distances, hovering around $25.
   - There is no significant variation in the average order value based on the distance, suggesting that AOV does not depend heavily on how far the customer is from the store.

Detailed Insights

1. **Customer Proximity:**
   - Most customers prefer to order from stores within a 5 km radius, with the highest concentration within 2 km. This highlights the importance of store location in attracting nearby customers.
   - We should focus on optimizing delivery routes within the 5 km radius for faster and efficient service. It can enhance customer satisfaction.
   - The decline in orders beyond 5 km may be due to longer delivery times, higher costs, or decreased service efficiency.
2. **Optimize Store Locations:**
   - Place new stores within 2-5 km of dense customer areas to maximize orders.
   - From the analysis aboved, we find out that Utah and Arizona are areas with fewer stores but high customer potential. So, we can consider opening new stores in dense customer areas within a 2-5 km radius.

4.4 Customer



The Customer Dashboard provides a comprehensive view of the customer dynamics over the years 2020, 2021, and 2022. This dashboard includes key metrics such as Total Customers, Repeat Customers, and Repeat Purchase Rate, along with monthly data visualizations and customer segments by sales. This analysis will help the manager understand customer behavior and identify opportunities for business improvement.

Relevance to the Manager

This dashboard is crucial for the manager as it:

- Helps monitor the growth and retention of customers over time.
- Identifies patterns in customer purchases and repeat behavior.
- Highlights segments of customers that are most valuable to the business.
- Provides insights into the effectiveness of customer retention strategies.

2020 Analysis

- **Total Customers**: 18,354
- **Repeat Customers**: 14,994
- **Repeat Purchase Rate**: 81.69%

Monthly Data:

- The chart shows a stable number of total and repeat customers across all months, indicating consistent business performance.
- The repeat purchase rate is displayed below 70% on the chart, which is inconsistent with the key metric of 81.69%. This discrepancy suggests a need for reviewing the calculation method or data representation.

2021 Analysis

- **Total Customers**: 19,458
- **Repeat Customers**: 16,058
- **Repeat Purchase Rate**: 82.53%

Monthly Data:

- There is a slight increase in total and repeat customers compared to 2020, indicating business growth.
- The repeat purchase rate remains around 70% on the chart, though the key metric shows 82.53%, suggesting a need for data consistency.

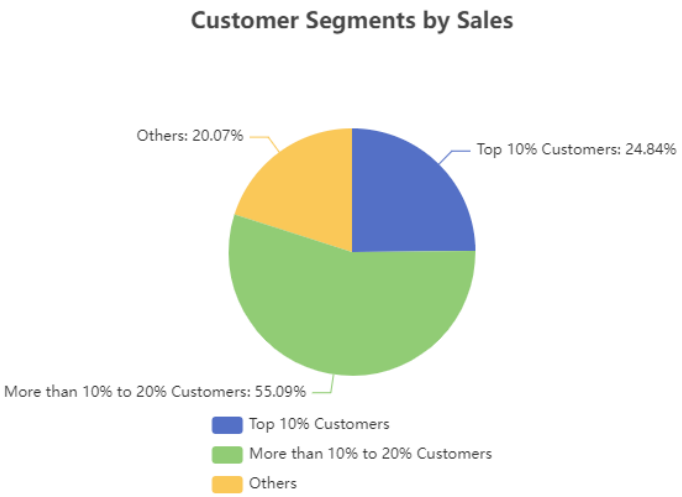
2022 Analysis

- **Total Customers**: 19,893
- **Repeat Customers**: 16,500
- **Repeat Purchase Rate**: 82.94%

Monthly Data:

- Further growth in total and repeat customers compared to 2021, showing continued business expansion.
- The repeat purchase rate remains relatively stable, close to 70% on the chart, with the key metric indicating 82.94%.

Customer Segments by Sales



Customer Segments by Sales 2020

- **Top 10% Customers**: 24.84%
- **More than 10% to 20% Customers**: 55.09%
- **Others**: 20.07%

Insights for 2020:

1. **Customer Stability**: The stable customer numbers indicate that the business maintained a consistent customer base throughout the year.
2. **High-Value Segments**: The top 10% and 10-20% customer segments contribute significantly to sales, emphasizing the importance of these customer groups.

Recommendations:

- Investigate the discrepancy in repeat purchase rates to ensure data accuracy.
- Focus on retaining high-value customers through loyalty programs and personalized offers.
- Develop strategies to convert mid-tier customers into high-value customers.

Customer Segments by Sales 2021

- **Top 10% Customers**: 25.09%
- **More than 10% to 20% Customers**: 54.42%
- **Others**: 20.49%

Insights for 2021:

1. **Growth in Customers**: The increase in total and repeat customers suggests effective customer acquisition and retention strategies.
2. **Consistent High-Value Segments**: The top customer segments continue to contribute significantly to sales.

Recommendations:

- Ensure data consistency between key metrics and chart visualizations.
- Continue enhancing customer retention programs to maintain high repeat purchase rates.
- Leverage data analytics to identify and target high-potential mid-tier customers.

Customer Segments by Sales 2022

- **Top 10% Customers**: 25.25%
- **More than 10% to 20% Customers**: 54.24%
- **Others**: 20.51%

Insights for 2022:

1. **Sustained Growth**: The continued increase in customer numbers indicates successful business strategies and market appeal.
2. **Valuable Customer Segments**: The consistent contribution from top customer segments highlights their importance to revenue.

Recommendations:

- Focus on retaining and engaging the growing customer base with tailored marketing efforts.
- Strengthen loyalty programs to ensure high repeat purchase rates.
- Use insights from high-value customers to refine customer acquisition strategies, targeting similar profiles.

Analysis and Insights from the Scatter Plot of CLV vs. Number of Orders

Overview

CLV Definition is a metric that represents the total net profit a company can expect to generate from a customer throughout their entire relationship

CLV= (Avg. Purchase value * Avg. Frequency rate) * (Avg. of years a customer stays alive/ total of customers)

This scatter plot displays the relationship between Customer Lifetime Value (CLV) and the number of orders per customer. Each point on the plot represents a customer, with the x-axis showing the number of orders and the y-axis representing the CLV.

Key Observations

1. **Positive Correlation:**
    - There is a strong positive correlation between the number of orders and the CLV. As the number of orders increases, the CLV also increases.
    - This suggests that customers who place more orders tend to have a higher lifetime value, which is expected as they generate more revenue over time.
2. **Linear Relationship:**
    - The data points form a linear pattern, indicating a direct relationship between the number of orders and the CLV.
    - The linearity suggests that each additional order consistently adds a similar amount to the customer's lifetime value.

Detailed Insights

1. **High-Value Customer Segment:**
    - The top-right cluster of the scatter plot represents high-value customers who have placed many orders. These customers are likely highly loyal and regular buyers.
    - Retaining these customers is crucial for sustained revenue growth. Strategies to enhance their experience and maintain their loyalty should be prioritized.
2. **Low- to Mid-Value Customer Segment:**
    - The dense cluster towards the bottom-left indicates many customers with lower order volumes and consequently lower CLV.
    - Converting these customers into higher-value customers by encouraging more frequent purchases could significantly boost overall revenue.
3. **Potential Growth Opportunities:**
    - The clear linear relationship suggests potential for growth by increasing the number of orders from existing customers.
    - Marketing strategies such as personalized offers, loyalty programs, and targeted promotions can help in achieving this.

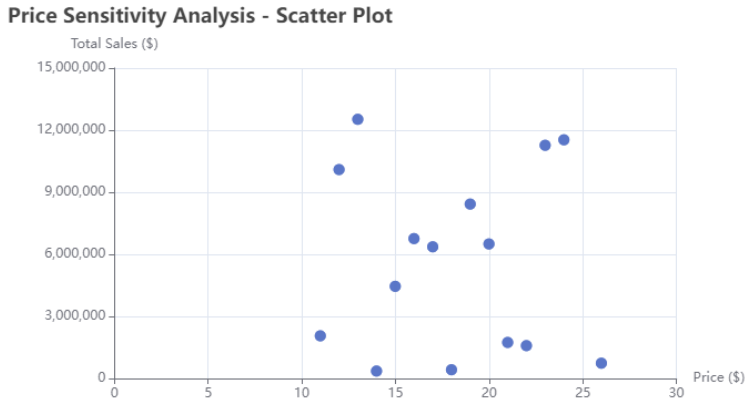
Recommendations

1. **Focus on Retaining High-Value Customers:**
    - Implement loyalty programs, exclusive offers, and personalized services to retain and reward high-value customers.

    - Engage with these customers through personalized communication and feedback mechanisms to ensure high satisfaction and continued loyalty.
2. **Increase Order Frequency for Low-Value Customers:**
    - Encourage low- to mid-value customers to increase their purchase frequency through targeted promotions, discounts on bulk purchases, and subscription services.
    - Use data analytics to identify the purchasing patterns of these customers and tailor marketing efforts to fit their preferences.
3. **Optimize Customer Acquisition Strategies:**
    - Analyze the characteristics of high-value customers and use these insights to refine customer acquisition strategies.
    - Focus on acquiring new customers who resemble the high-value segment in terms of demographics, preferences, and purchasing behavior.

By implementing these recommendations, PizzaToGo can enhance customer retention, increase order frequency, and ultimately boost the overall customer lifetime value, leading to sustained revenue growth.

Scatter Plot Price Sensitivity



Overview

This scatter plot represents the relationship between price ($) and total sales ($). Each point on the plot represents a data point of total sales at a particular price level.

We have 15 prices

Key Observations

1. **Price and Sales Relationship:**
    - The scatter plot shows that total sales vary across different price points, indicating how sensitive sales are to changes in price.

    - There is no single clear pattern of how price affects total sales, suggesting a complex relationship influenced by various factors such as market conditions, customer preferences, and competitive pricing.
2. **High Sales Clusters:**
    - The highest total sales, approaching $15,000,000, are observed at price points around $10 and $15. This indicates that lower to mid-range prices might be optimal for maximizing total sales.
    - There is another cluster of high sales around the $20-$25 price range, but the total sales here are slightly lower than the highest cluster.
3. **Low Sales Points:**
    - Several points with total sales below $5,000,000 are scattered across various price points, indicating that not all price increases or decreases lead to higher sales.
    - Particularly, prices above $20 show varied total sales, with some points significantly lower than others, suggesting that higher prices might not consistently lead to high sales volumes.

Detailed Insights

1. **Optimal Price Points:**
    - Price points around $10 to $15 appear to generate the highest total sales, suggesting these might be the optimal price ranges for maximizing revenue.
    - Adjusting pricing strategies to focus on these ranges could potentially increase total sales.
2. **Price Elasticity:**
    - The spread of points across different price levels indicates varying degrees of price elasticity. Some customers might be more sensitive to price changes, while others may prioritize product value or quality.
    - The mixed results at higher price points suggest that while some customers are willing to pay more, a significant portion might be deterred by higher prices, leading to lower sales.
3. **Strategic Pricing:**
    - To optimize sales, it might be beneficial to employ dynamic pricing strategies, adjusting prices based on demand, competitive pricing, and customer feedback.
    - Offering discounts, promotions, or bundle deals around the optimal price points ($10-$15) could further boost sales and attract price-sensitive customers.

Recommendations

1. **Focus on Optimal Price Ranges:**
    - Concentrate pricing strategies around the $10-$15 range to maximize total sales. Regularly review and adjust prices within this range based on market conditions and sales performance.
2. **Dynamic Pricing Strategy:**
    - Implement a dynamic pricing model to adjust prices in real-time based on demand, competition, and other market factors. This can help capture more sales at optimal price points.

    - Use data analytics to monitor customer response to price changes and refine the pricing model accordingly.
3. **Promotions and Discounts:**
    - Introduce promotions, discounts, and special offers around the $10-$15 price range to attract more customers and drive higher sales volumes.
    - Consider limited time offers and loyalty programs to incentivize repeat purchases and enhance customer loyalty.
4. **Customer Segmentation:**
    - Segment customers based on their price sensitivity and purchasing behavior. Tailor marketing efforts and pricing strategies to different segments to maximize overall sales.
    - For price-sensitive customers, focus on value deals and promotions. For less price-sensitive customers, emphasize product quality and unique features.
5. **Competitive Analysis:**
    - Regularly analyze competitor pricing to ensure competitive pricing strategies. Adjust prices to remain competitive while maintaining profitability.
    - Monitor market trends and adjust prices dynamically to capitalize on changing market conditions.

5. Major Challenges

Frontend:

- **Finding the Right Chart**: Selecting the appropriate type of chart to analyze data and derive meaningful insights can be challenging.
- **Interactive Features**: Implementing interactive makes frontend logic more complex and requires additional time to ensure smooth user interactions.
- **Visualization Optimization**: Ensuring that visual elements in the charts, such as colors, labels, and legends, are both meaningful and accurate is crucial for effective data representation.
- Libraries **like Chart.js and Leaflet**, while powerful, can sometimes fall short when it comes to handling complex data visualizations or specific customization requirements.
  For instance, by default, Leaflet markers do not offer extensive customization options for changing their appearance, such as color. => So, I had to create a custom icon as location marker in Leaflet involves using HTML and CSS to achieve the desired appearance and functionality.

Backend:

- **Creating New Tables in SQL**: Due to the large volume of data, displaying analytics directly on the page resulted in significant delays. To optimize performance, it was necessary to create new SQL tables for pre-aggregating and storing processed data. However, during the creation and migration of these SQL tables, we encountered numerous errors. The migration process frequently failed due to schema issues, data type mismatches, and constraints. It took several days to debug and correct these

issues. By refining our migration scripts, thoroughly testing them in a staging environment, and implementing proper error handling, we were able to successfully migrate and optimize the database.

- **Ensuring Accurate JSON Responses**: There were instances where the backend code, although seemingly correct, did not produce the expected results on the frontend. This was primarily due to discrepancies in the data structure and inconsistencies in the JSON responses. Continuous integration practices and automated tests were implemented to ensure that each change in the backend produced the correct JSON responses, ultimately achieving the desired frontend functionality.

- **Correcting URL Configurations**: Incorrect URL configurations led to numerous issues, including broken links and failed API calls. This hindered the ability to retrieve and display data accurately on the frontend. By standardizing URL configurations and implementing comprehensive routing tests, we resolved the discrepancies and ensured seamless navigation and data retrieval across the application.

- ***Migration Issues:*** While performing migrations, the Django models automatically created their own IDs as primary keys, which led to conflicts and errors.
  Solution: To resolve this the IDs in the database schema, even though they were not needed in the code, were created manually. This involved modifying the migration files to ensure that the primary key conflicts were resolved.

- **Data Extraction and Performance:** Initially, the data extraction methods within the code caused the webpage to load the charts very slowly. This was due to the heavy computations being done in real-time. To address this specific database tables to store precomputed data were created. The code was then modified to simply retrieve the data from these tables, significantly improving the performance and speed of loading the charts.
- **SQL Joins and Data Aggregation:** The project encountered difficulties with SQL joins, particularly because the orders table did not have a primary key, which complicated the join operations and led to incorrect results. To overcome this, the data by categories using SQL CASE statements was grouped and then performed the necessary calculations. This approach allowed for accurate data aggregation and correct chart distribution.

- **Pie Chart Distribution:** Creating the pie chart distribution was challenging due to the absence of a primary key in the orders table, which made join operations problematic. Grouping the data with SQL CASE statements and then calculating the category distributions made it possible to generate the correct data needed for the pie charts. This method ensured accurate representation of the data in the visualizations.

6. Frameworks and Tools Feedback

Technology	Advantages	Disadvantages
Django	Rapid development, scalability, security	Learning Curve, Monolithic, Opinionated

MariaDB	Open Source, Compatibility	Less Corporate Support, Lagging Features, Complexity
jQuery	Makes AJAX handling easier, suitable for a simple data analysis tool	Slow for modern web app compared to Angular or React
Chart.js	Easy to use, suitable for simple charts	Limited interactivity and advanced features
Echarts	A wide variety of chart types, extensive interactivity features	More complex, steeper learning curve
Leaflet	Simple, fast to run	Basic default styling, additional plugins needed

7. Distribution of Work

Task/ Feature	Stack	Assigned To	Status
Config database and Load data into the database	Database	Thi Diem Quynh Tran	Completed
Basic wireframe for web tool	Frontend	Mohamed Akhouaji	Completed
Three Key Metrics in Home Page	Frontend	Thi Diem Quynh Tran	Completed
Line chart for Monthly Total Orders	Frontend	Mohamed Akhouaji	Completed
Line chart for Monthly Average Oder Value	Frontend	Mohamed Akhouaji	Completed
Bar chart of Total Sales, Category Sales by year, month on products site	Frontend	Thi Diem Quynh Tran	Completed
Pie chart of Category Distribution by year, month	Frontend	Thi Diem Quynh Tran	Completed
Create bar chart of ingredient usage	Frontend	Thi Diem Quynh Tran	Completed
Mixed chart for Product Size Popularity	Frontend	Thi Diem Quynh Tran	Completed
Line chart for monthly total sales of all stores	Frontend	Thi Diem Quynh Tran	Completed
Heatmap for Store and Customer Distribution	Frontend	Thi Diem Quynh Tran	Completed
Bar chart of Total Sales by Stores	Frontend	Thi Diem Quynh Tran	Completed
Heatmap of Revenue by Store and Category	Frontend	Thi Diem Quynh Tran	Completed
Scatter plot of Revenue vs. Number of Items Sold	Frontend	Thi Diem Quynh Tran	Completed
Mixed chart for Distribution of Orders by Distance	Frontend	Thi Diem Quynh Tran	Completed
Three Key Metrics in Customers Site	Frontend	Thi Diem Quynh Tran	Completed
Mixed Chart for Total Customers, Repeat Customer, Repeat Purchase Rate	Frontend	Thi Diem Quynh Tran	Completed
Pie Chart for Customer Segments by Sales	Frontend	Thi Diem Quynh Tran	Completed
Scatter plot of CLV vs. Number of Orders	Frontend	Thi Diem Quynh Tran	Completed
Scatter plot of Price Sensitivity	Frontend	Thi Diem Quynh Tran	Completed

Bar Chart and Pie Chart: Category sales and overall month sales	Backend	Aiman Ashfaq	Completed
Line/Bar: Order distance	Backend	Aiman Ashfaq	Completed
Key metrics: Total revenue, Average Order Value, Repeat Purchase Rate (all years), Repeat Customer	Backend	Aiman Ashfaq	Completed
Product Size popularity	Backend	Aiman Ashfaq	Completed
Heatmap: Customer and store distribution	Backend	Aiman Ashfaq	Completed
Bar Chart: Repeat Customer	Backend	Aiman Ashfaq	Completed
Scatter plot: Revenue vs. Number of items sold per store	Backend	Aiman Ashfaq	Completed
Homepage/Line Chart: Total orders	Backend	Aiman Ashfaq	Completed
Homepage/Line Chart: Average Order Value	Backend	Aiman Ashfaq	Completed
Scatter Plot of CLV vs. Number of orders	Backend	Aya Zirari	Completed
Scatter Plot of Price Sensitivity Analysis	Backend	Aya Zirari	Completed
Pie Chart Customer Segments by Sales	Backend	Aya Zirari	Completed
Monthly Data Line Chart Repeat Purchase Rate	Backend	Aya Zirari	Completed
Monthly Data Bar Chart Total Customers	Backend	Aya Zirari	Completed
Bar Chart Ingredient Usage Count	Backend	Aya Zirari	Completed
Line Chart monthly total sales in products page	Backend	Aya Zirari	Completed
Bar Chart total customers	Backend	Aya Zirari	Completed
Bar Chart total revenues of stores by states	Backend	Aya Zirari	Completed
Heatmap of Revenues by Category and Store	Backend	Aya Zirari	Completed
Key metrics: Total customers and Total number of orders	Backend	Aya Zirari	Completed