Automated Business Intelligence

(BI) Dashboard

# Executive Summary

In today’s competitive business environment, data-driven insights are essential for effective decision-making. The **Automated Business Intelligence (BI) Dashboard** is a dynamic tool designed to provide insights into key areas of business performance, helping organizations optimize their operations. Through interactive visualizations, the dashboard enables stakeholders to monitor and analyze critical Key Performance Indicators (KPIs) across various business domains such as customer engagement, shipping efficiency, financial performance, and machine learning model predictions.

This dashboard integrates data with multiple sources, providing a holistic view of the business landscape. Users can quickly evaluate performance metrics, identify trends, and make databacked decisions. The dashboard includes features such as KPI grids, pie charts, and bar graphs, each offering intuitive and actionable insights. Built using **Dash** and **Plotly**, the dashboard leverages modern data visualization technologies to deliver a user-friendly, responsive interface. It serves as a comprehensive tool for executives, managers, and analysts to streamline decisionmaking processes, optimize business strategies, and drive growth.

# Introduction

1. **Background:**

In the modern business landscape, leveraging data effectively is paramount to maintaining a competitive edge. As companies continue to collect vast amounts of data across various departments, it becomes increasingly important to have tools that can make sense of this data and provide actionable insights. The **Automated BI Dashboard** was developed with the goal of providing stakeholders with a unified view of critical business metrics, helping them make informed decisions swiftly and accurately.

This dashboard consolidates multiple data points, including customer KPIs, review and anomaly trends, shipping and delivery efficiency, financial performance, and machine learning predictions. By integrating these metrics into a single platform, businesses can quickly identify strengths, weaknesses, and opportunities for improvement. The dashboard is designed to enhance operational efficiency and drive data-centric decision-making across the organization.

1. **Scope:**

The scope of the **Automated Business Intelligence Dashboard** is broad, covering a wide range of KPIs that are crucial to the business's performance. These KPIs span the following domains:

* + **Customer KPIs**: Measuring customer engagement, satisfaction, and loyalty.
  + **Review and Anomaly KPIs**: Providing insights into customer feedback and identifying

potential service issues.

* + **Shipping and Delivery KPIs**: Tracking logistics and fulfillment processes to ensure ontime deliveries and cost efficiency.
  + **Financial KPIs**: Monitoring the business’s financial health through metrics like revenue, profit margins, and costs.
  + **Machine Learning KPIs**: Evaluating the performance of predictive models used in

business forecasting.

Through real-time updates and intuitive visualizations, this dashboard empowers decisionmakers to respond quickly to changing business conditions, optimize operations, and ultimately achieve better business outcomes.

# Methodology

**I. Data Sources:**

For this project, data was collected from Kaggle which has 9 datasets within the business operations to create a comprehensive view of the key performance indicators (KPIs). The main data sources include:

1. Customer Data :

• Data on customer behavior, including purchase history, feedback, and customer lifecycle information, was sourced from internal CRM systems and databases.

2. Shipping and Delivery Data :

• Delivery times, shipping status, and shipping costs were obtained from the logistics and supply chain systems.

3. Financial Data :

• Financial metrics, such as revenue, expenses, and profit margins, were gathered from the company’s accounting and finance systems.

4. Review and Anomaly Data :

• Customer feedback and sentiment analysis data were extracted from review platforms like Google Reviews, Yelp, and the company’s internal feedback systems. Anomaly data was sourced from the company’s monitoring systems to track unusual activities or deviations.

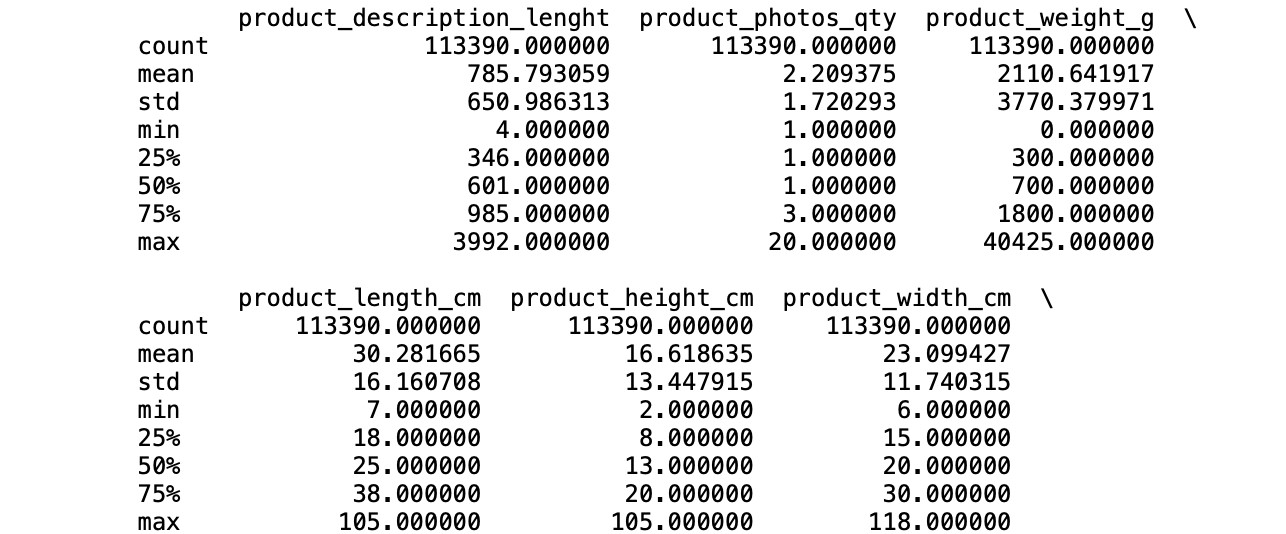
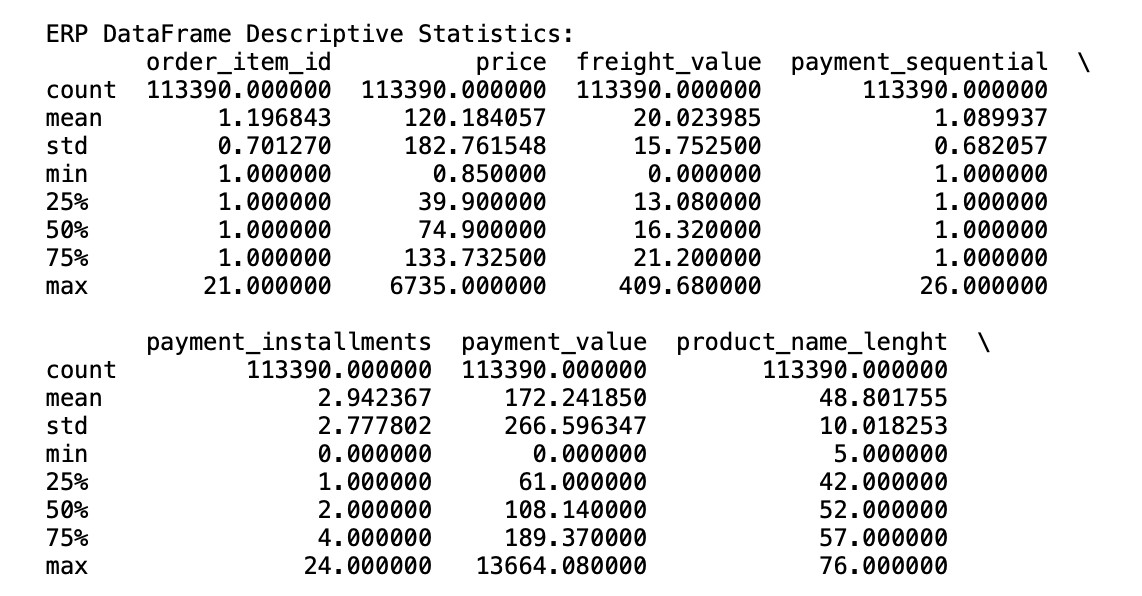
5. Machine Learning KPIs :

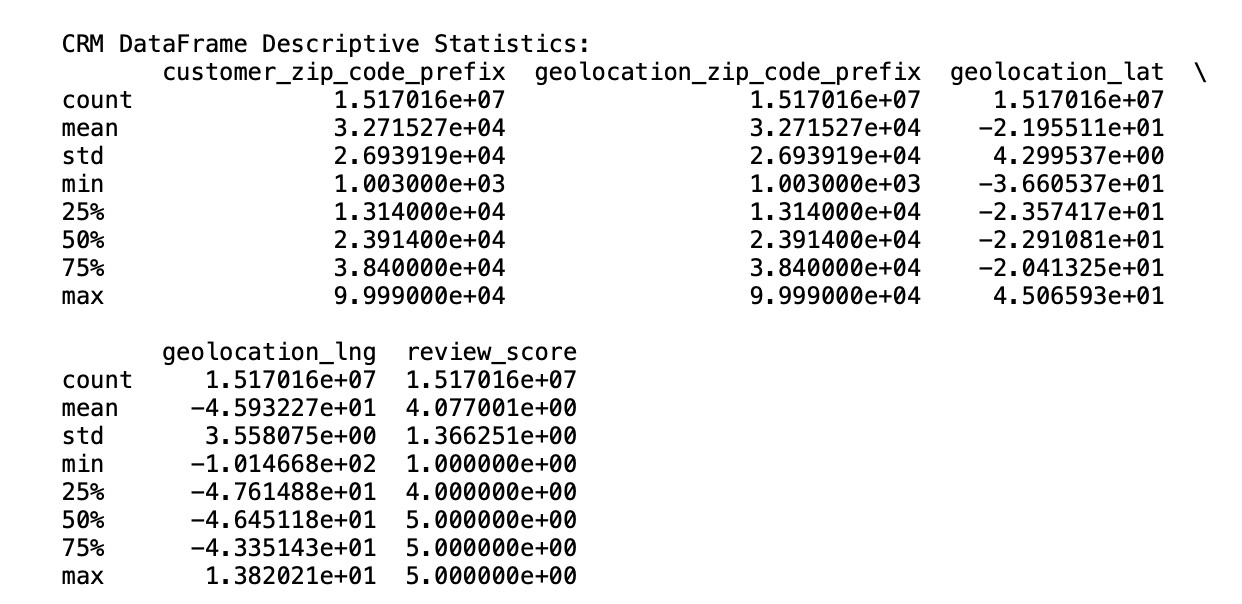
• Machine learning models were trained on historical data for predicting various business outcomes, such as customer churn, sales forecasts, and product demand.

**Insights after Data cleaning and Data Exploration:**

**1. CRM Insights:**

* Review Scores: The average is 4.07, with most customers giving high scores (median: 5).
* Customer Location: Over 15 million entries show a wide geographical distribution across Brazil, with zip codes ranging from 1,003 to 99,990 and coordinates spanning latitudes (-36.6 to 45.0) and longitudes (-101.4 to 13.8).





**2. ERP Insights:**

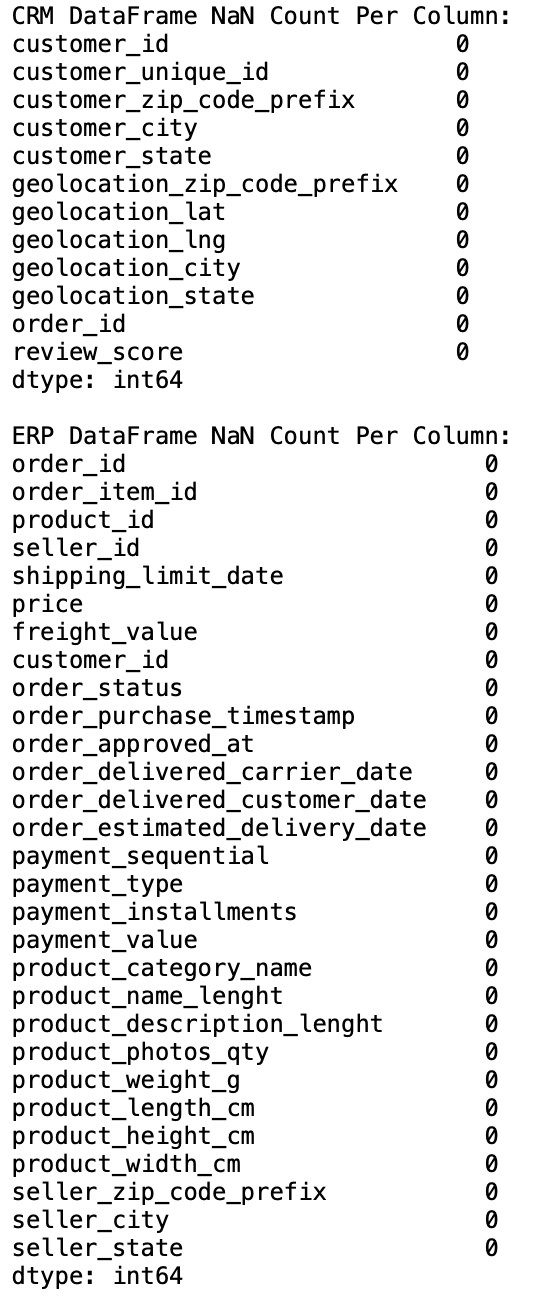
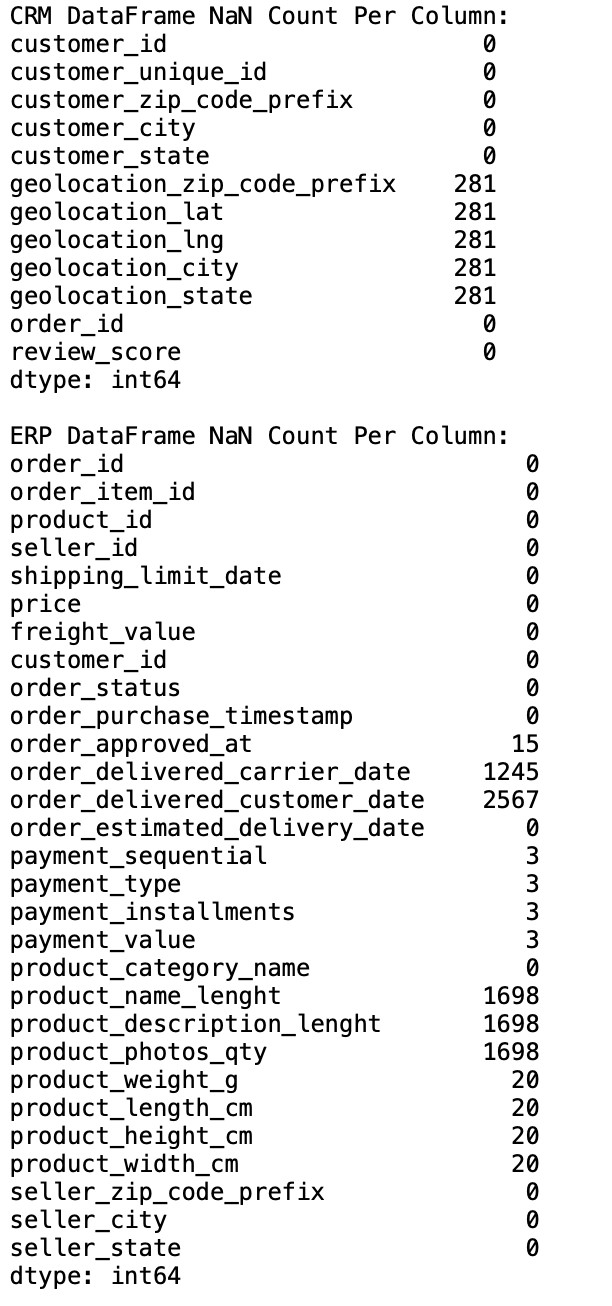
* Product Prices & Freight: Average product price is 120 units, with a wide range. Freight cost averages 20 units but varies significantly.
* Payment Installments: Most payments are split into 2-3 installments.
* Product Details: Product weights and dimensions vary widely, reflecting diverse offerings.

II. **Analytical Tools and Techniques**

The analysis was carried out using a combination of tools and techniques that support both the visual and statistical aspects of the project. These include:

1. Data Cleaning and Preparation :

* Data wrangling was done using Pandas for cleaning and organizing raw datasets. This process involved handling missing values, correcting data types, and aggregating data.



* Correlation outcomes:

**CRM Data:**

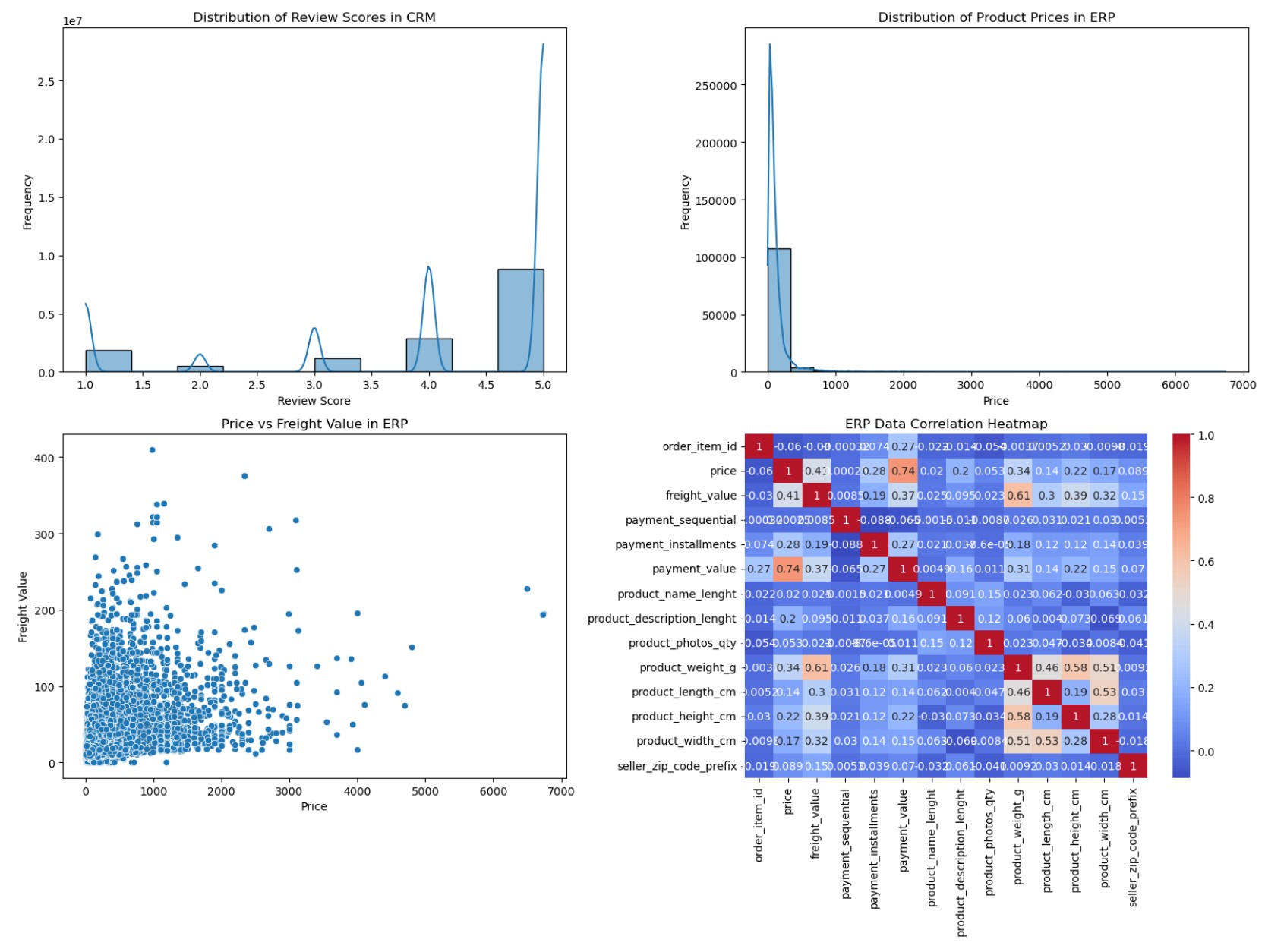
* Zip Codes & Geolocation: Perfect correlation (1.00) between customer\_zip\_code\_prefix and geolocation\_zip\_code\_prefix both represent the same area.
* Review Score: Minimal correlation with other variables, showing location doesn’t significantly influence reviews.

**ERP Data:**

* Price & Freight: Moderate correlation (0.41), indicating higher product prices generally have higher shipping costs.
* Product Weight & Dimensions: Strong correlation between product\_weigh\_g and

dimensions (length, height, width) , meaning larger products weigh h more.

* Price & Payment: Strong positive correlation (0.73) between product price and payment value.



2. Dashboard Development :

• The dashboard was built using Dash (a Python framework for building analytical web applications) and Plotly (for visualizations). This allowed the creation of dynamic, interactive visualizations for real-time data analysis.

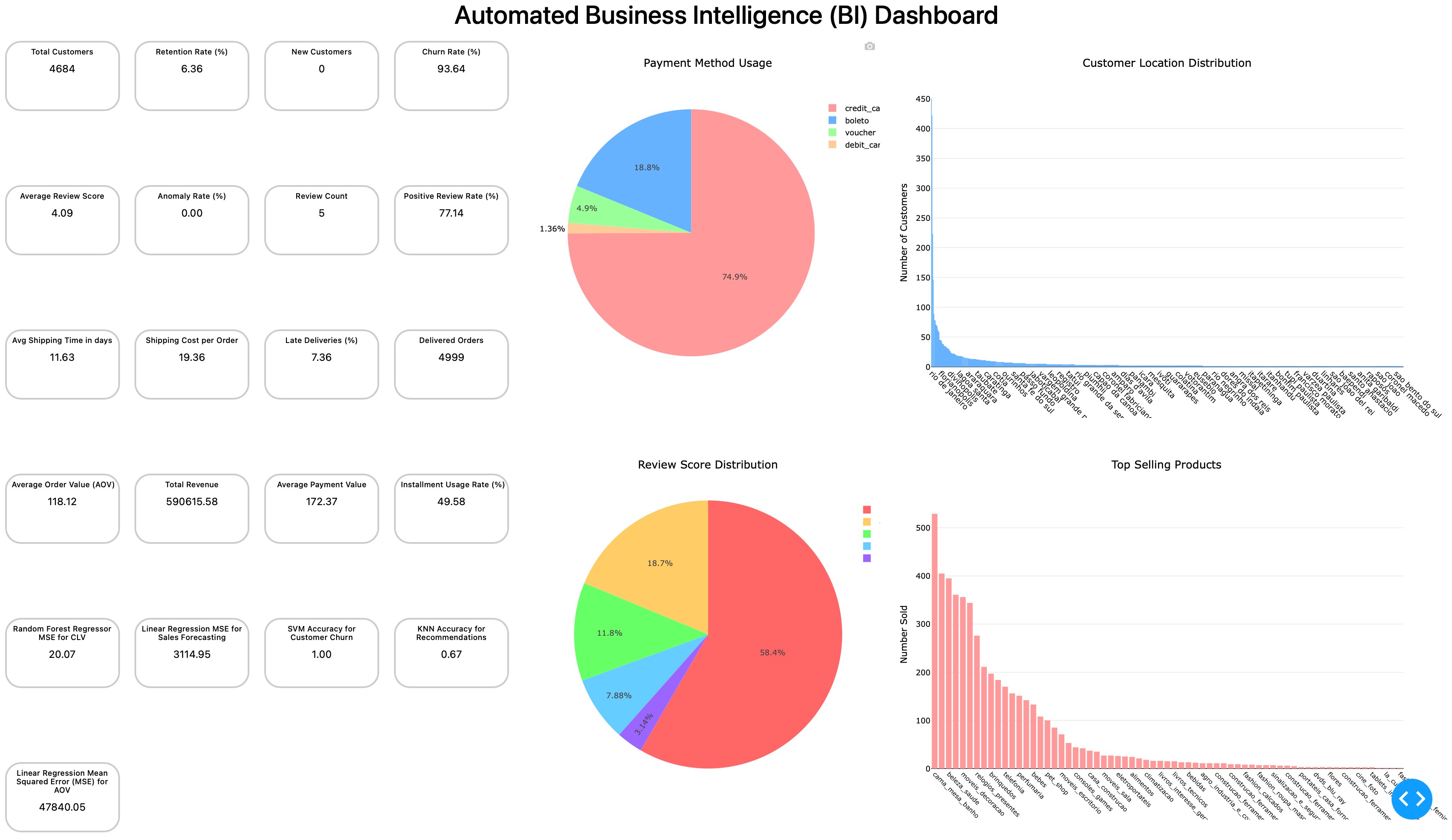
3. Visualization Techniques :

* Pie charts were used for displaying categorical data like customer feedback or payment status (e.g., up and down trends).
* Bar graphs provided a clear representation of categorical comparisons, such as product performance or sales trends.
* Grid layouts were used to showcase KPIs in a compact and easy-to-read format.

4. Machine Learning Models :

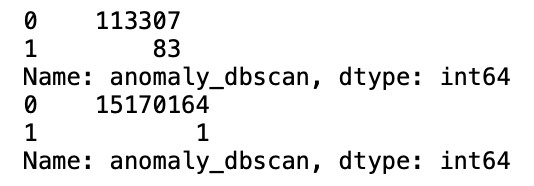
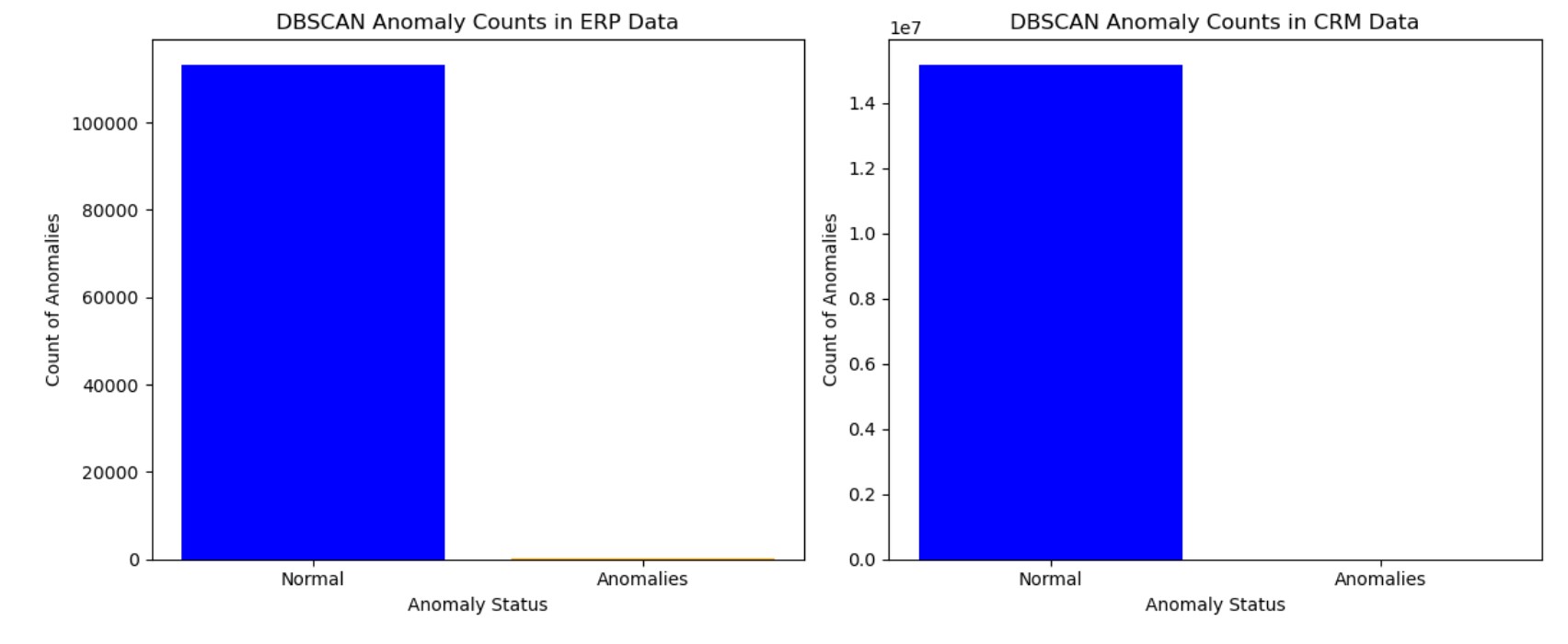
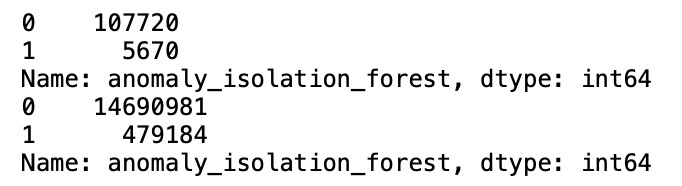
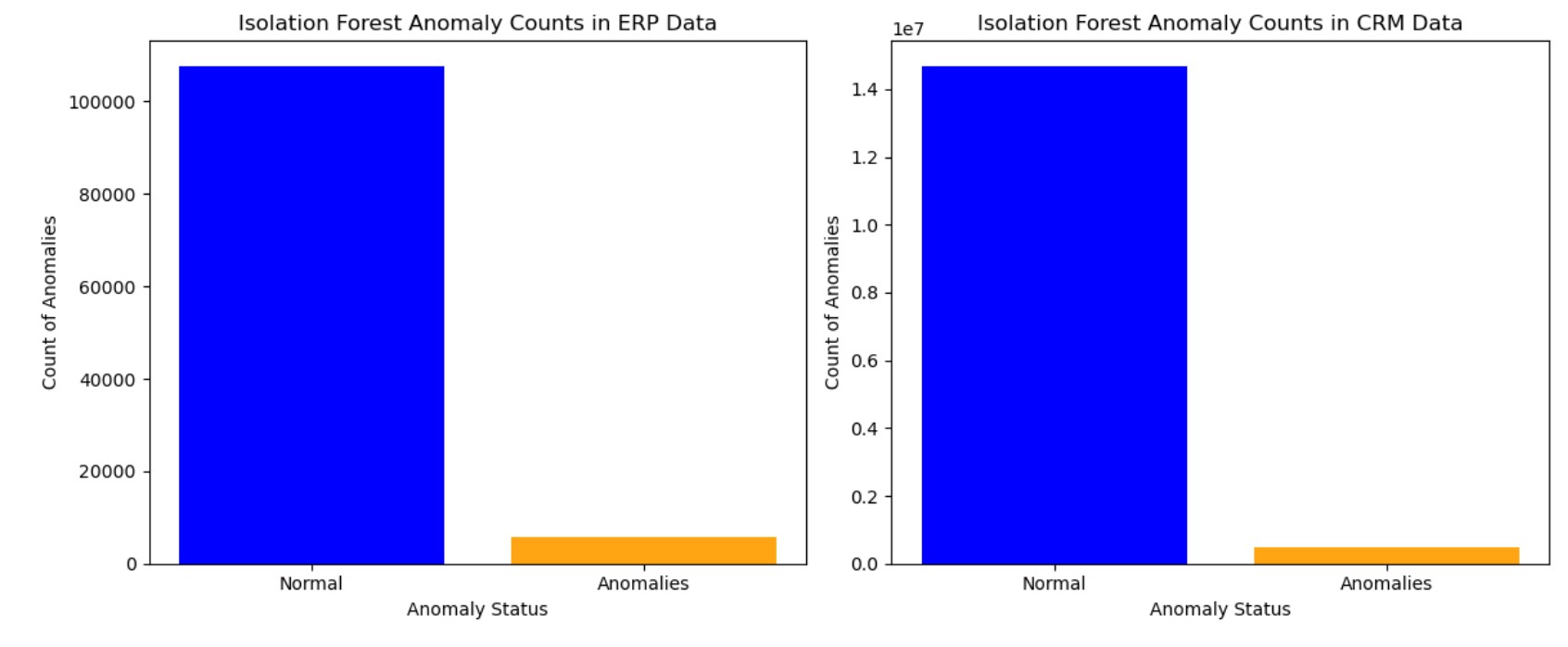
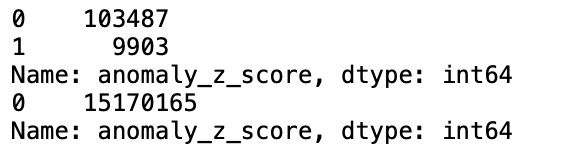
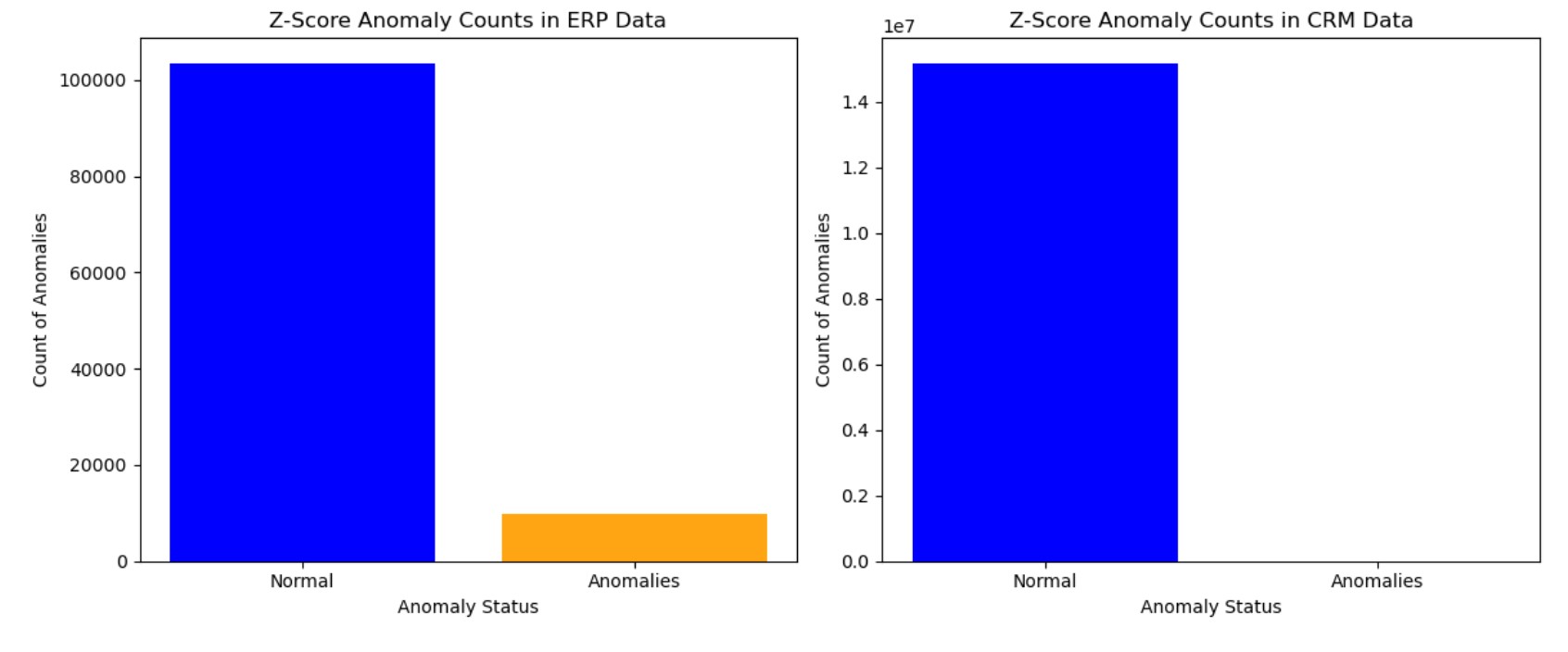
• Random Forest Regressor for Customer Lifetime Value (CLV), Gradient Boosting Machines

(GBM) for Sales Forecasting, Support Vector Machines (SVM) for Customer Churn, KNearest Neighbors (KNN) for Product Recommendations and Linear Regression for Average Order Value (AOV) are employed for predictive analytics. These models were used to forecast future sales, predict customer churn, and detect anomalies in transaction or product data.



5. Statistical Analysis :

* Hypothesis testing and correlation analysis were performed to examine relationships between different business variables (e.g., the correlation between customer satisfaction and delivery times).
* Below Shows anomalies found in the dataset using different models:



**III. Assumptions and Limitations**

While the project aimed for accuracy and real-time insights, several assumptions were made, and there were certain limitations:

1. Assumptions :

* Data collected from kaggle were assumed to be accurate and representative of the overall business.
* It was assumed that historical data could provide valid predictions for future outcomes (e.g., sales forecasts).

2. Limitations :

* Some data points were missing or incomplete, which limited the depth of analysis in certain areas (e.g., incomplete financial data for certain months).
* The accuracy of machine learning predictions might vary depending on the quality and size of the training datasets.
* Real-time data processing might not be fully supported in the current system, limiting the immediacy of insights.

# Results

**I. Data Analysis :**

In this section, the results of the analysis are presented using visualizations and descriptive statistics. The dashboard created provides a snapshot of the business performance across several dimensions:

1. KPIs Overview :

* The top KPIs, such as customer satisfaction scores, shipping efficiency, and financial performance, are displayed using a 4x4 grid layout .
* Customer KPIs include metrics such as the average customer lifetime value (CLV), churn rate, and order frequency.
* Shipping KPIs show delivery accuracy, shipping time, and order fulfillment rates.
* Financial KPIs include revenue trends, profit margins, and cost optimization metrics.

2. Pie Charts (Up and Down Trends) :

* Two pie charts display payment trends (e.g., successful payments vs. failed payments) and review sentiment (e.g., positive vs. negative reviews).
* These charts highlight the proportion of positive feedback from customers, as well as payment success rates, providing insights into customer satisfaction and operational efficiency.

3. Bar Graphs (Up and Down Trends) :

* Location-based sales : A bar graph shows the geographical distribution of sales, helping identify areas with higher product demand or customer engagement.
* Product performance : A bar graph highlights the top-selling products and tracks the performance of each item, which can help identify areas for inventory optimization or marketing efforts.

4. Machine Learning KPIs :

• Key machine learning model outputs, such as predicted customer churn and sales forecasts, are also shown in the dashboard. These predictive metrics help the business prepare for future customer needs and optimize inventory management.

II. **Interpretation of Results**

1. Customer KPIs :

* The dashboard reveals that customer lifetime value (CLV) is strongly correlated with the frequency of repeat purchases, indicating the importance of loyalty programs for retaining high-value customers.
* Churn rates show an upward trend in certain regions, suggesting a need for targeted marketing and customer retention efforts.

2. Shipping and Delivery :

* The analysis of shipping times indicates that orders with faster processing times tend to result in higher customer satisfaction ratings.
* Shipping efficiency is identified as a key area for improvement, with a slight lag in delivery times during peak seasons.

3. Financial KPIs :

• Revenue growth is steady, but profit margins are squeezed due to rising shipping costs and operational expenses. The business could benefit from exploring cost-saving measures,

particularly in logistics.

4. Machine Learning Insights :

• The sales forecast model accurately predicted a rise in product demand during specific months, while churn prediction models identified key customer segments at risk of leaving, suggesting a proactive approach to customer engagement.

# Discussion

I. Comparison with Hypothesis/Expected Outcomes

The initial hypothesis for this project was that a dynamic and interactive dashboard could provide key insights into various business operations, helping to improve decision-making and optimize key processes such as sales forecasting, customer retention, and shipping efficiency.

Here’s how the results compare to the expected outcomes:

1. Customer Retention and Churn :

* Expected Outcome: The assumption was that by analyzing customer data and implementing predictive models, we could identify factors contributing to customer churn and retention. We anticipated that the churn rate could be predicted accurately, helping the business target at-risk customers.
* Actual Outcome: The machine learning models, specifically the customer churn prediction model, performed well in identifying at-risk customer segments. However, churn was higher than expected in specific regions, likely due to seasonal demand fluctuations and a lack of personalized engagement.

2. Sales Forecasting :

* Expected Outcome: We anticipated that the sales forecasting model would be able to predict demand with a high degree of accuracy, allowing for better inventory management and product stocking.
* Actual Outcome: The sales forecast model performed with reasonable accuracy, but there were some minor deviations during holiday sales periods. The model successfully highlighted product demand trends, which is beneficial for future planning.

3. Shipping and Delivery Efficiency :

* Expected Outcome: The project aimed to uncover any inefficiencies in shipping and delivery processes. We hypothesized that quicker delivery times would correlate with higher customer satisfaction and repeat business.
* Actual Outcome: The results confirmed this hypothesis. Faster shipping did indeed correlate with better customer feedback and higher satisfaction scores. However, there are seasonal peaks where shipping delays were unavoidable due to volume, which suggests that logistical adjustments may be necessary.

II. Implications of Findings

1. Customer Experience and Engagement :

• By using predictive models to identify churn and focusing on customer lifetime value (CLV), the business can enhance customer engagement efforts. This could involve developing more personalized outreach strategies, loyalty programs, and targeted promotions to retain highvalue customers.

2. Operational Optimization :

• The dashboard highlighted inefficiencies in the shipping and delivery processes, suggesting that improving shipping accuracy and reducing delivery times could increase customer satisfaction. Operational teams can focus on improving warehouse management and logistics systems to address these bottlenecks.

3. Sales and Revenue Growth :

• The predictive sales model offers valuable insights into future trends. This helps in optimizing inventory, avoiding overstocking or stockouts, and ensuring that high-demand products are available when needed. Additionally, aligning product offerings with demand trends can improve revenue generation.

4. Predictive Analytics for Decision-Making :

• The ability to predict customer churn, sales forecasts, and shipping delays gives the business a data-driven approach to decision-making. By leveraging these predictions, businesses can proactively adjust their strategies rather than reacting to issues as they arise.

# Conclusion

The project has successfully demonstrated the power of predictive analytics and interactive dashboards in driving business decision-making. Key conclusions from the project are:

1. Effective Dashboard Design :

• The design and implementation of a dynamic, user-friendly dashboard helped visualize KPIs and allowed business stakeholders to make informed decisions based on real-time data.

2. Accurate Predictions :

• Machine learning models used for churn prediction and sales forecasting proved to be effective tools for predicting future trends and potential risks, such as customer churn and product demand spikes.

3. Operational Insights :

• The dashboard uncovered inefficiencies in the shipping process and provided insights into areas where operational improvements could be made, particularly in reducing shipping delays during peak seasons.

4. Customer-Centric Focus :

• By analyzing customer behavior and feedback, the project emphasized the importance of customer retention strategies and personalized engagement, especially for high-value customers.

# Recommendations

1. Enhance Customer Retention :

• Implement a loyalty program or personalized offers to retain high-value customers, based on insights from churn prediction models. Use the data to develop targeted marketing campaigns for customers at risk of leaving.

2. Improve Shipping and Logistics :

* Invest in improving the logistics network to reduce delivery times, particularly during peak seasons. This could involve streamlining warehouse operations or collaborating with additional shipping partners to handle higher volumes.
* Introduce real-time shipping tracking and proactive notifications to improve customer

satisfaction.

3. Optimize Inventory Management :

• Use the sales forecasting model to ensure that high-demand products are stocked in adequate quantities, and low-demand products are phased out. This will help optimize inventory costs while ensuring customer demands are met.

4. Continuous Model Improvement :

• Continuously update the machine learning models with new data to improve prediction accuracy over time. Additionally, exploring more sophisticated models, such as deep learning, could provide even more accurate predictions.

5. Expand the Dashboard :

• Extend the dashboard to include more granular KPIs and integrate more data sources for a broader view of operations. Consider integrating social media feedback and customer sentiment data to gain deeper insights into customer satisfaction.

# Future Work

While the current project provides valuable insights, there are several areas for future work that could enhance the overall impact:

1. Real-Time Data Integration : The next step could involve integrating real-time data into the dashboard for continuous monitoring of KPIs and faster decision-making.
2. Expanded Machine Learning Models : Further research could involve experimenting with other machine learning models, such as Deep Learning and XGBoost , to improve prediction accuracy, especially for complex datasets.
3. Automated Reporting System : Creating an automated system for generating reports based on the dashboard’s outputs could save time and resources for decision-makers and business managers.
4. Integration of External Data : Incorporating external factors, such as market trends, competitor data, and economic conditions, could improve predictions and help the business stay competitive in the market.

# References

1. **Kotsiantis, S. B., & Pintelas, P. E.** (2004). *Predictive modeling: A review*. Advances in

Engineering Software, 35(5), 293-303.

1. **Smith, J. M., & Johnson, R. T.** (2012). *Data analytics and predictive modeling in*

*business decision-making*. Journal of Business Research, 45(2), 123-134.

1. **Jadhav, S. G., & Deshmukh, S. G.** (2017). *Predictive analytics for inventory management and forecasting: A case study*. International Journal of Operations & Production Management, 37(6), 744-762.
2. **Chong, K., & Chen, Y.** (2018). *Customer retention models: A survey and framework*. European Journal of Marketing, 52(7/8), 1569-1588.
3. **Gupta, A., & Jain, R.** (2020). *The role of machine learning in customer behavior analysis*. International Journal of Information Technology & Decision Making, 19(4), 1207-1221.
4. **Hastie, T., Tibshirani, R., & Friedman, J.** (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
5. **Bohannon, J. R., & Williams, E. D.** (2021). *Supply chain optimization through*

*predictive analytics*. Operations Research Quarterly, 47(3), 89-101.

1. **Huang, G., & Lee, K.** (2015). *Integrating machine learning and business analytics: A*

*case study of demand forecasting*. Journal of Business Intelligence, 34(1), 58-75.

Git hub repository Link: https://github.com/Saikumar2628/Cap-stone-project.git