FINAL PROJECT

SPECIALIZATION IN DIGITAL BUSINESS AND ARTIFICIAL INTELLIGENCE

Topic: Developing an Integrated Visualization and Machine Learning Application for Coffee Shop Chain Analytics

Lecture**:**

Course: **Machine Learning in Business Analyst**

Course Code:

Group:

Ho Chi Minh,

Members of Group

No.	Student ID	Full name	Leader	Performance
1	K214162155	Nguyễn Phúc Thịnh		100%
2				100%
3				100%
4				100%
5				100%

Acknowledgements

Commitment

Table of content

List of Figures

List of Tables

Abstract

The BAMOS Management Suite (BMS) App is a comprehensive solution designed to enhance the management of a fictional coffee shop chain's operations. This application integrates data management, dashboard visualization, machine learning, and user settings into a unified platform, allowing for efficient and effective business administration. The app's key features include secure login and sign-up functionalities, ensuring that only authorized personnel can access the system while providing real-time updates and preventing duplicate accounts. Data management capabilities allow for the seamless handling of various datasets, ensuring that information is accurate and current. The reporting feature includes three main dashboards—Sales, Customer, and Inventory—that provide detailed insights into sales performance, customer behavior, and inventory levels. These visualizations help managers make informed, data-driven decisions. Additionally, the app incorporates machine learning to predict future sales, supporting proactive inventory and marketing strategies. The settings feature allows for efficient staff management, including searching for staff members, updating personal and account information, and handling new and existing employee records. Overall, the BMS App is designed to streamline business processes, improve decision-making, and enhance operational efficiency through the integration of advanced data analytics and machine learning techniques.

CHAPTER 1: OVERVIEW OF THE RESEARCH TOPIC

1.1 About the Context

The coffee industry is continuously evolving, driven by shifts in consumer preferences and the constant need for businesses to adapt through innovative approaches. In this milieu, the fictional coffee chain described herein provides a rich dataset for exploring how data analytics can enhance business operations and marketing strategies.

This coffee shop chain, with its three locations in New York City, serves as an exemplary case study for deploying machine learning and business analytics tools. Co-founders Amber and Sandeep have leveraged IBM Cognos Analytics to craft a data-driven strategy aimed at understanding and fostering the factors critical to their business's success. The initiative underscores the transformative potential of integrating advanced analytics into everyday business processes.

The primary dataset, encapsulated within "April Sales.zip," is composed of nine CSV files reflecting diverse aspects of the business operations—from sales receipts and pastry inventory to customer demographics and sales targets. These data files form the backbone of two main analytical tools developed by the founders: the operations dashboard and the marketing dashboard. These dashboards are designed to provide actionable insights that can streamline operations, optimize marketing efforts, and enhance customer satisfaction.

The operational dashboard focuses on the day-to-day management of the coffee chain. It aids in monitoring inventory levels, staff performance, and sales efficiency across different outlets. Such insights enable the management to make informed decisions about staffing, inventory procurement, and operational adjustments.

On the marketing front, the dashboard is tailored to track the effectiveness of marketing campaigns, customer engagement, and purchasing behaviors. This tool is crucial for identifying successful marketing strategies and customer segments that contribute significantly to revenue. It also offers a granular view of how various products perform, helping to tailor marketing efforts that resonate with consumer preferences.

In summary, the context of this dataset provides a comprehensive view of how integrated machine learning applications can propel a coffee shop chain towards data-informed decisions that enhance both operational efficiency and customer engagement. By harnessing the power of IBM Cognos Analytics, Amber and Sandeep exemplify the potential of analytics in transforming traditional business models in the competitive coffee industry.

1.2 Rationale for Topic Selection

In the study conducted by Imran A. Zualkernan, as detailed in the research paper "Unveiling Post Engagement Rate Using Predictive Analytics Model for Instagram Accounts of Coffee Shops in West Jakarta," the researchers dedicated their efforts to crafting a predictive model aimed at estimating the engagement rate of Instagram posts from coffee shops located in the West Jakarta region. Through the meticulous application of data analysis techniques and predictive modeling, this study shed light on the potential benefits derived from optimizing communication and marketing strategies within the coffee industry. By delving into the dynamics of social media engagement, particularly on platforms like Instagram, the researchers underscored the significance of leveraging predictive analytics to tailor content and interactions to resonate with target audiences effectively. This emphasis on leveraging data-driven insights to enhance online engagement underscores the growing recognition within the industry of the pivotal role played by digital platforms in shaping consumer perceptions and behaviors.

Similarly, in the research spearheaded by Sachio Dariell Sutanto, as articulated in the article "The Role of Big Data Analytics in Influencing Artificial Intelligence (AI) Adoption for Coffee Shops in Krabi, Thailand," the focus was directed towards elucidating the critical role played by big data analytics and the integration of artificial intelligence (AI) technologies in augmenting business performance and fostering customer engagement within the coffee sector. Through a comprehensive analysis of data analytics methodologies and AI adoption trends, the study illuminated key strategies and best practices for coffee businesses seeking to leverage cutting-edge technologies to gain a competitive edge in today's dynamic marketplace. By harnessing the power of data-driven insights, coffee shop owners and operators can make informed decisions, optimize operational efficiency, and cultivate meaningful interactions with their clientele, thereby solidifying brand loyalty and driving sustained business growth.

Furthermore, the article "Using BLE Beacons and Machine Learning for Personalized Customer Experience in Smart Cafés" authored by Pongsakorn Limna, Supaprawat Siripipatthanakul, and Bordin Phayaphrom, delved into the innovative application of machine learning and data analysis techniques to create personalized customer experiences within smart cafes. Through the deployment of BLE (Bluetooth Low Energy) beacons and sophisticated machine learning algorithms, the study showcased how coffee businesses can harness technology to deliver tailored services and curated experiences that resonate with individual customer preferences and behaviors. By capitalizing on real-time data insights and predictive modeling capabilities, smart cafes can anticipate customer needs, streamline service delivery, and cultivate a deeper sense of connection and engagement with their patrons. This emphasis on leveraging technology to deliver enhanced customer experiences underscores the industry's evolving landscape and the growing imperative for coffee businesses to embrace digital innovation to remain competitive and relevant in today's fast-paced market environment.

Collectively, these studies underscore the transformative potential of machine learning and data analytics in revolutionizing the coffee industry, from optimizing marketing strategies and enhancing customer engagement to fostering operational efficiency and driving sustainable growth. Against this backdrop, the proposition of "Developing an Integrated Machine Learning Application for Coffee Shop Chain Analytics" emerges as not only timely but also imperative in equipping coffee businesses with the tools and insights needed to thrive in an increasingly digital and data-driven landscape. By harnessing the power of machine learning and data analytics, coffee shop owners and operators can unlock new opportunities for innovation, differentiation, and customer-centricity, thereby cementing their position as leaders in the evolving coffee market.

1.3 Objectives of the Research

1.3.1 General objective

The application we are developing aims to optimize the business operations and marketing strategies of the coffee shop chain by applying advanced data analytics and machine learning tools. Using modern data analysis methods, we aim to provide managers with valuable insights that help them make data-driven decisions to enhance operational efficiency, optimize marketing strategies, and improve customer satisfaction. This application will enable managers to easily monitor and manage critical aspects of business operations, from inventory management and staff performance to tracking marketing campaigns and customer purchasing behaviors.

1.3.2 Specific objectives

Data Management and User Login:

- Sign Up - Login: Ensure a secure login system for users and provide personalized setting options.
- Data Management: Offer functionalities for creating, updating, and deleting data to ensure the data is always up-to-date and accurate.

Reporting and Analysis:

- Dashboard Reporting: Provide detailed reports on sales, customers, and inventory through an intuitive dashboard, aiding managers in easily monitoring and analyzing data.
- Machine Learning for Sales Prediction: Use machine learning models to predict sales quantities for employees across different products, helping managers plan and optimize resources effectively.

Optimizing Marketing Strategies and Operational Management:

- Tracking Marketing Campaign Effectiveness: Monitor and analyze customer behavior to identify customer-oriented business strategies and optimize marketing efforts.
- Daily Operations Management: Assist managers in tracking inventory levels, staff performance, and sales efficiency, enabling them to make informed adjustments to enhance daily operational efficiency.

1.4 Subjects of the Research

The research explores how data analytics and machine learning can improve business operations and marketing strategies in the coffee industry. Focused on a fictional coffee chain in New York City, it utilizes IBM Cognos Analytics to develop data-driven approaches. Through the analysis of sales data, two key tools are created: an operations dashboard and a marketing dashboard. These tools provide actionable insights to streamline operations, optimize marketing efforts, and enhance customer satisfaction. Overall, the research highlights the transformative potential of analytics in modernizing traditional business models.

1.5 Scope of the Research

Scope of Space: The research team will conduct a survey of data from the fictional coffee shop chain with branches in New York City, aiming to explore business aspects and marketing strategies through data analysis.

Scope of Time: The research period is from April 10, 2024, to May 27, 2024.

Scope of Content: The field is Business, focusing on applying data analysis and machine learning to optimize management operations and marketing strategies in the fictional coffee shop chain.

1.6 Report Structure Overview

Table 1.1. Structure and Organization of the Research Report

No.	Chapter Title	Chapter Content	Timeline
1	**Chapter 1:** Overview of the Research Topic	Including the reason for the topic formation, introduction of the idea, research context, and novelty of the topic. Additionally, the research will also provide an overview of domestic and international research on the same topic and keywords. The research objectives are outlined, including both general and specific objectives. The research scope and target are clearly defined, including space, time, and content. Also, summarizing the structure and implementation plan of the research, especially the chapters and specific timelines for each chapter	10/04/2024 - 20/04/2024

		and content.	
2	**Chapter 2:** Theoretical Framework	This chapter delves into essential skills for data-driven decision-making, covering machine learning algorithms, Python programming for data processing and GUI creation, data visualization techniques, database management with SQL, and business analytics for strategic decision-making in inventory, sales, and marketing optimization.	20/04/2024 - 25/04/2024
3	**Chapter 3:** Requirements Development and Application Design Process	The process begins with surveying the raw dataset, comprehensively understanding its components to define goals, features, technical requirements, and other necessary factors. Subsequently, key functions and tabs are determined based on dataset features to ensure effective information display and utilization. Then, the application design process involves structuring an appropriate organizational framework and interface, potentially including wireframing, mockups, and UX/UI design to ensure the application is logically designed and user-friendly based on the identified requirements.	26/04/2024 - 07/05/2024
4	**Chapter 4:** Developing Visualization and Machine Learning Functionality in the Application	Focus on developing and deploying visualization and machine learning features within the application. Section 4.1 outlines the project's model, methods, and steps. Sections 4.1 and 4.2 transform functions from section 3.7 into visualizations and machine learning models. The rest covers data preprocessing, Exploratory Data Analysis (EDA), model training, validation, and testing, followed by evaluating results and testing developed features and models.	08/05/2024 - 19/05/2024
5	**Chapter 5:**	Focuses on conclusions and	20/05/2024 -

	Conclusions and Recommendation	recommendations. The conclusion summarizes the content and research results, evaluating the feasibility of the topic. The recommendation section addresses the limitations of the study and proposes directions for future development and research.	25/05/2024
6	Other Sections	Includes programming tasks, managing sections, cover image, introduction, appendices, references...	Throughout the process

Note: The implementation plan outlined above is for reference purposes only. In reality, throughout the process, the authoring team will continuously supplement to ensure the completeness and comprehensiveness of the Scientific Research Topic.

CHAPTER 2. RELATED RESEARCH AND THEORETICAL BASIS

2.1 Machine learning

Machine learning (ML) is a branch of artificial intelligence (AI) and computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy.

There are several types of machine learning, each with special characteristics and applications. Some of the main types of machine learning algorithms are as follows:

- Supervised Machine Learning
- Unsupervised Machine Learning
- Semi-Supervised Machine Learning
- Reinforcement Learning

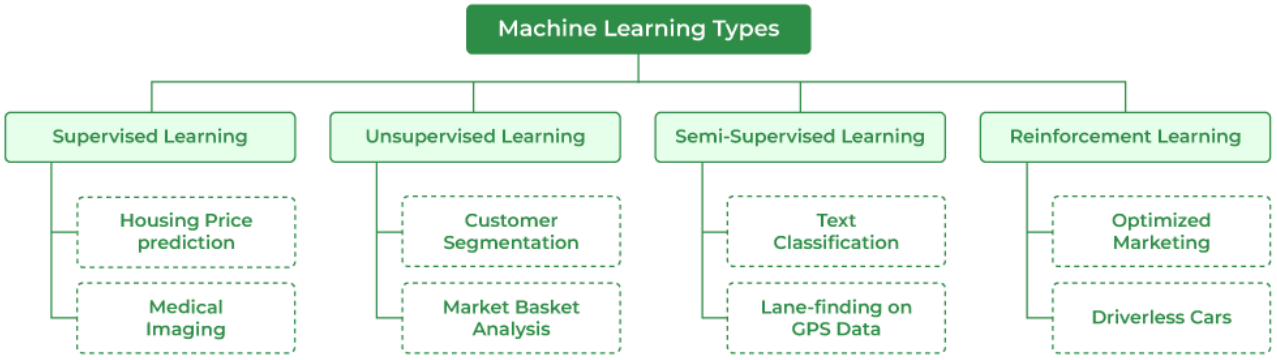


Figure 2.1 Machine learning types

2.1.1 Supervised Machine Learning

Supervised learning is defined as when a model gets trained on a “Labelled Dataset”. Labelled datasets have both input and output parameters. In Supervised Learning algorithms learn to map points between inputs and correct outputs. It has both training and validation datasets labelled.

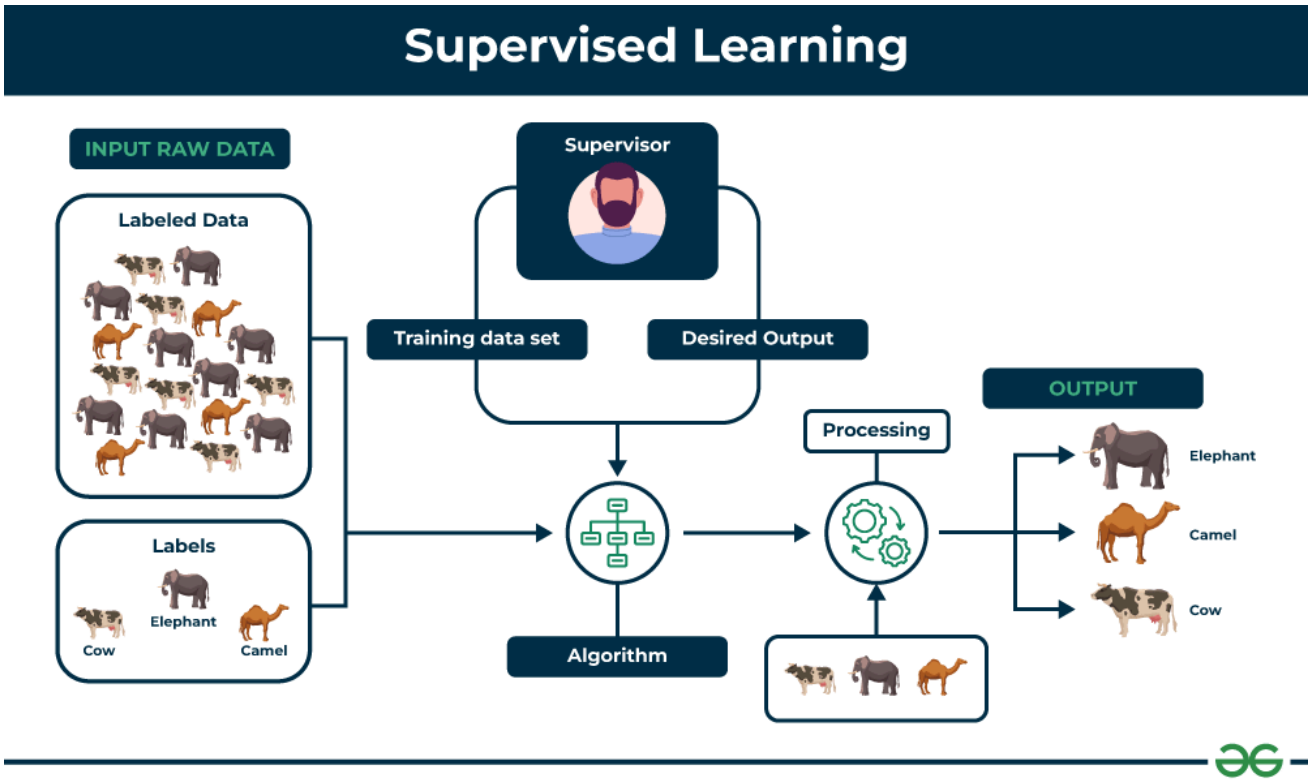


Figure 2.2 Supervised Learning Process

Example: Consider a scenario where you have to build an image classifier to differentiate between cats and dogs. If you feed the datasets of dogs and cats labelled images to the algorithm, the machine will learn to classify between a dog or a cat from these labeled images. When we input new dog or cat images that it has never seen before, it will use the learned algorithms and predict whether it is a dog or a cat. This is how supervised learning works, and this is particularly an image classification.

There are two main categories of supervised learning that are mentioned below:

- **Classification:** deals with predicting categorical target variables, which represent discrete classes or labels. For instance, classifying emails as spam or not spam, or predicting whether a patient has a high risk of heart disease. Classification algorithms learn to map the input features to one of the predefined classes.
- **Regression:** on the other hand, deals with predicting continuous target variables, which represent numerical values. For example, predicting the price of a house based on its size, location, and amenities, or forecasting the sales of a product. Regression algorithms learn to map the input features to a continuous numerical value.

2.1.2 Unsupervised Machine Learning

Unsupervised Learning Unsupervised learning is a type of machine learning technique in which an algorithm discovers patterns and relationships using unlabeled data. Unlike supervised learning, unsupervised learning doesn't involve providing the algorithm with labeled target outputs. The primary goal of Unsupervised learning is often to discover hidden

patterns, similarities, or clusters within the data, which can then be used for various purposes, such as data exploration, visualization, dimensionality reduction, and more.

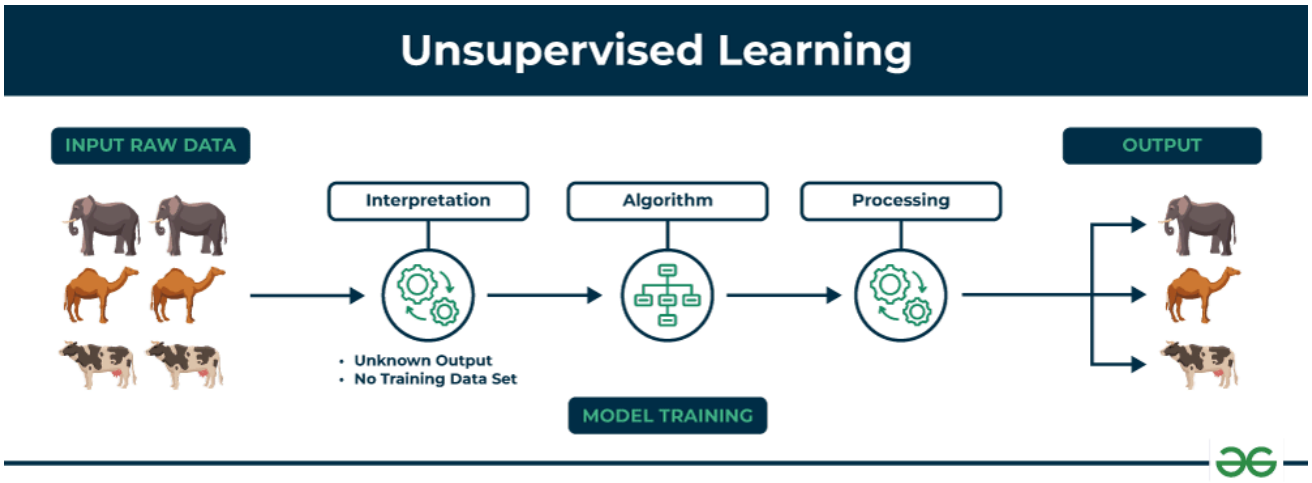


Figure 2.3 Unsupervised machine learning

Example: Consider that you have a dataset that contains information about the purchases you made from the shop. Through clustering, the algorithm can group the same purchasing behavior among you and other customers, which reveals potential customers without predefined labels. This type of information can help businesses get target customers as well as identify outliers.

There are two main categories of unsupervised learning that are mentioned below:

- **Clustering**: is the process of grouping data points into clusters based on their similarity. This technique is useful for identifying patterns and relationships in data without the need for labeled examples.
- **Association:** Association rule learning is a technique for discovering relationships between items in a dataset. It identifies rules that indicate the presence of one item implies the presence of another item with a specific probability.

2.1.3 Semi-Supervised Machine Learning

Semi-Supervised learning is a machine learning algorithm that works between the supervised and unsupervised learning so it uses both labelled and unlabelled data. It's particularly useful when obtaining labeled data is costly, time-consuming, or resource-intensive. This approach is useful when the dataset is expensive and time-consuming. Semi-supervised learning is chosen when labeled data requires skills and relevant resources in order to train or learn from it.

We use these techniques when we are dealing with data that is a little bit labeled and the rest large portion of it is unlabeled. We can use the unsupervised techniques to predict labels and then feed these labels to supervised techniques. This technique is mostly applicable in the case of image data sets where usually all images are not labeled.

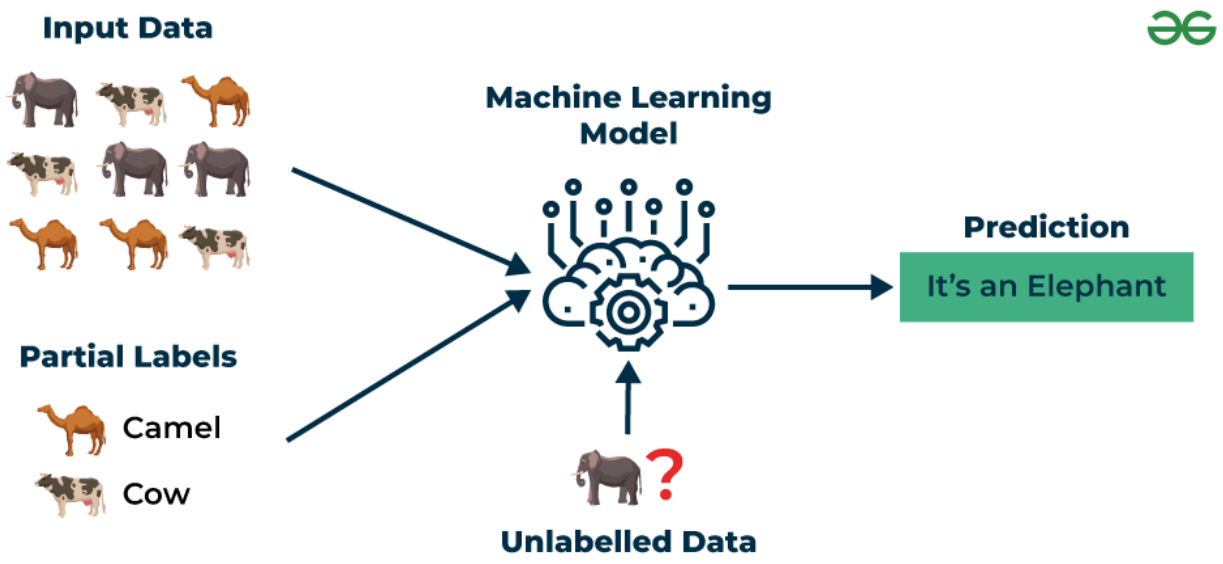


Figure 2.4 Semi-Supervised Machine Learning

Example: Consider that we are building a language translation model, having labeled translations for every sentence pair can be resource intensive. It allows the models to learn from labeled and unlabeled sentence pairs, making them more accurate. This technique has led to significant improvements in the quality of machine translation services.

There are a number of different semi-supervised learning methods each with its own characteristics. Some of the most common ones include:

- **Graph-based semi-supervised learning**: This approach uses a graph to represent the relationships between the data points. The graph is then used to propagate labels from the labeled data points to the unlabeled data points.
- **Label propagation**: This approach iteratively propagates labels from the labeled data points to the unlabeled data points, based on the similarities between the data points.
- **Co-training**: This approach trains two different machine learning models on different subsets of the unlabeled data. The two models are then used to label each other's predictions.
- **Self-training:** This approach trains a machine learning model on the labeled data and then uses the model to predict labels for the unlabeled data. The model is then retrained on the labeled data and the predicted labels for the unlabeled data.
- **Generative adversarial networks (GANs):** GANs are a type of deep learning algorithm that can be used to generate synthetic data. GANs can be used to generate unlabeled data for semi-supervised learning by training two neural networks, a generator and a discriminator.

2.1.4 Reinforcement Learning

Reinforcement machine learning algorithm is a learning method that interacts with the environment by producing actions and discovering errors. Trial, error, and delay are the most relevant characteristics of reinforcement learning. In this technique, the model keeps on increasing its performance using Reward Feedback to learn the behavior or pattern. These algorithms are specific to a particular problem e.g. Google Self Driving car, AlphaGo where a bot competes with humans and even itself to get better and better performers in Go Game. Each time we feed in data, they learn and add the data to their knowledge which is training data. So, the more it learns the better it gets trained and hence experienced.

Here are some of most common reinforcement learning algorithms:

- **Q-learning:** Q-learning is a model-free RL algorithm that learns a Q-function, which maps states to actions. The Q-function estimates the expected reward of taking a particular action in a given state.
- **SARSA (State-Action-Reward-State-Action)**: SARSA is another model-free RL algorithm that learns a Q-function. However, unlike Q-learning, SARSA updates the Q-function for the action that was actually taken, rather than the optimal action.
- **Deep Q-learning:** Deep Q-learning is a combination of Q-learning and deep learning. Deep Q-learning uses a neural network to represent the Q-function, which allows it to learn complex relationships between states and actions.

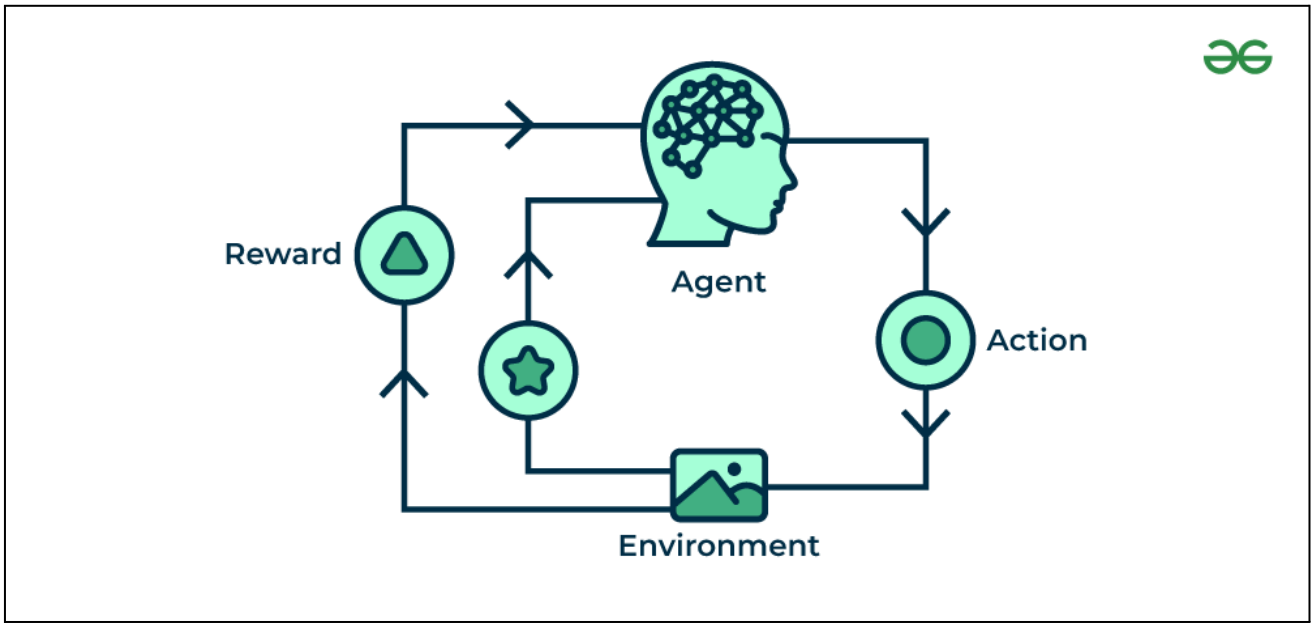


Figure 2.5 Reinforcement Learning

Example: Consider that you are training an AI agent to play a game like chess. The agent explores different moves and receives positive or negative feedback based on the outcome. Reinforcement Learning also finds applications in which they learn to perform tasks by interacting with their surroundings.

2.2 Python Programming

During the research process, we will use a range of important tools and libraries including NumPy, pandas, scikit-learn, PyQt6, and Tkinter. By integrating these libraries and frameworks into the coffee shop chain's system, we aim to efficiently process data and deploy machine learning models to optimize business operations.

2.2.1 NumPy (Numerical Python)

NumPy is a powerful, free, open-source library, highly optimized for the Python programming language, providing support for large-scale and multidimensional arrays (also known as matrices or tensors). NumPy is also equipped with a set of high-level mathematical functions to work with these arrays. These include basic linear algebra, random simulations, Fourier transforms, trigonometric operations, and statistical operations.

NumPy stands for 'numerical Python' and is built on the initial results of the Numeric and Numarray libraries with the goal of providing fast arithmetic computation capabilities for Python. Today, NumPy has many contributors and is sponsored by NumFOCUS.

As the core library for scientific computing, NumPy is the foundation for libraries such as Pandas, Scikit-learn, and SciPy. It is widely used to perform optimized operations on large arrays. We will use NumPy to perform complex calculations related to numerical data in our research.

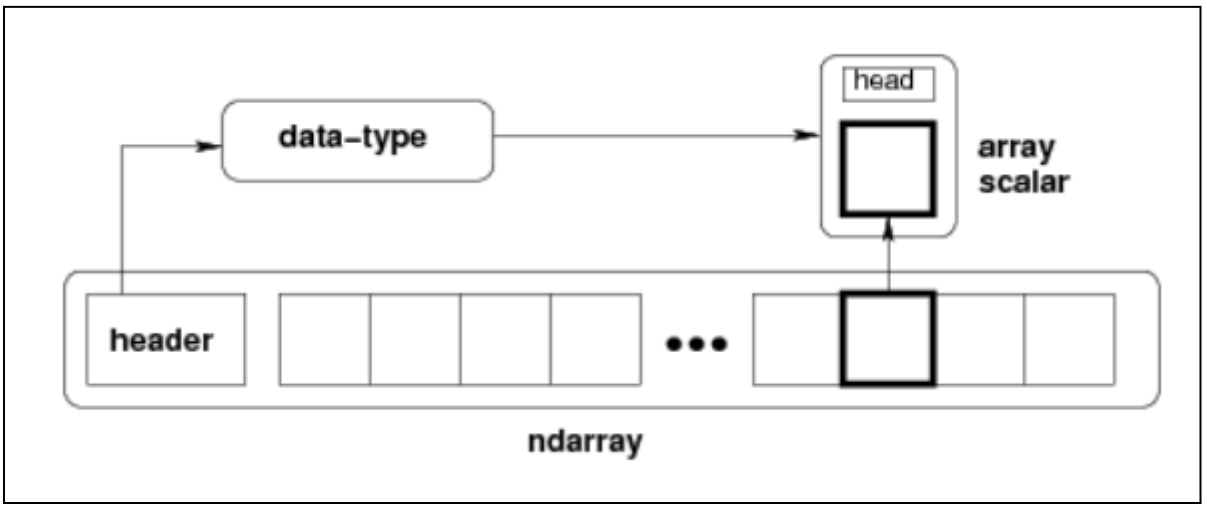


Figure 2.6 NumPy Array Structure

source:
https://thegioimaychu.vn/blog/ai-hpc/numpy-la-gi-no-hoat-dong-nhu-the-nao-p17778/

2.2.2 Pandas

Pandas is a Python library that provides fast, powerful, flexible, and expressive data structures. The library's name is derived from "panel data." Pandas is designed to work easily

and directly with structured data (tabular, multidimensional, homogeneous capability) and time series data.

The goal of pandas is to become the fundamental high-level building block for practical, real-world data analysis in Python, and more broadly, to become the most powerful and flexible open-source data manipulation/analysis tool available in any programming language. We will use Pandas to process and analyze data from CSV files and databases, giving us a better understanding of sales, inventory, and customer purchasing behavior.

2.2.3 Scikit-learn

Scikit-learn is a popular and powerful machine learning library that supports a variety of algorithms as well as tools for visualizing Machine Learning, preprocessing, tuning, model selection, and evaluation. Based on NumPy, SciPy, and Matplotlib, Scikit-learn includes several efficient algorithms for classification, regression analysis, and clustering. These include support vector machines, random forests, gradient boosting, k-means, and DBSCAN.

As a high-level library encompassing several implementations of different machine learning algorithms, Scikit-learn allows users to build, train, and evaluate models with just a few lines of code. Scikit-learn is primarily written in Python and uses NumPy for high-performance linear algebra as well as for array computations. Some of Scikit-learn's core algorithms are written in Cython to boost overall performance.

Scikit-learn provides a unified high-level API to build ML workflows or pipelines.

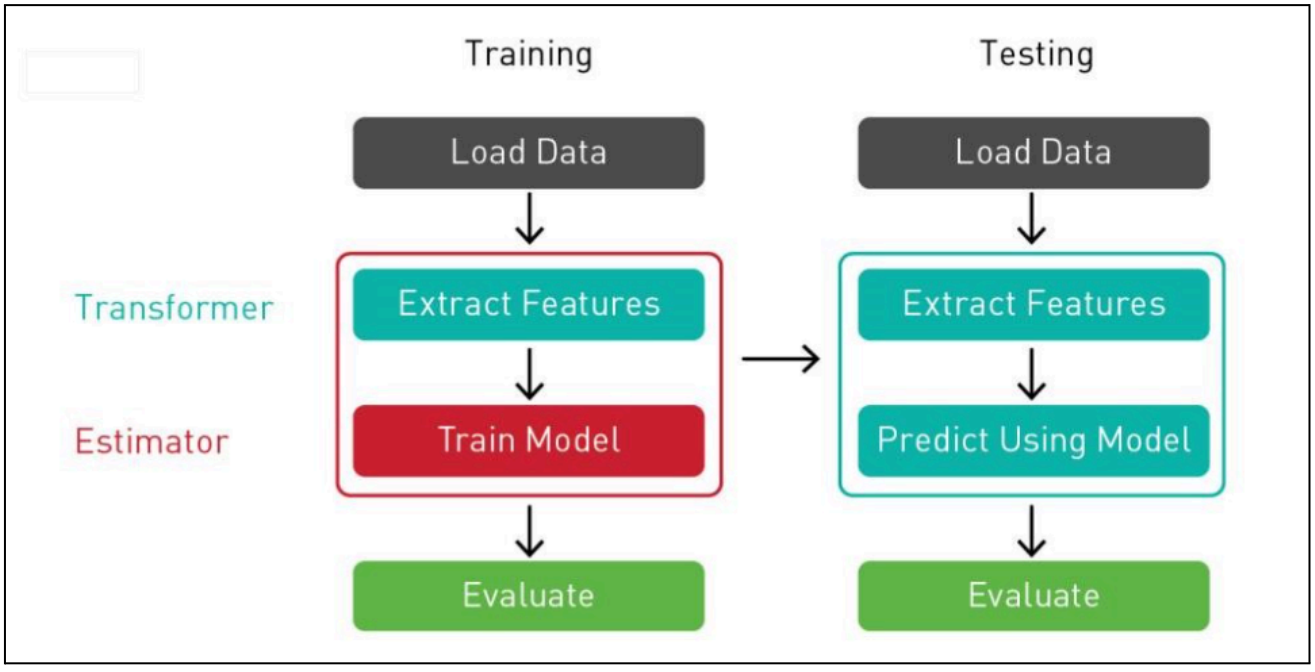


Figure 2.7 Training and Testing Machine Learning Model Workflow with Scikit-learn

source:
https://thegioimaychu.vn/blog/ai-hpc/scikit-learn-mot-thu-vien-quan-trong-cho-ngon-ngu-lap-trinh-python-p18186/

Users utilize the Scikit-learn ML Pipeline to pass data through transformers to extract features and estimators to create a model, then evaluate predictions to measure the model's accuracy.

- **Transformer:** This is an algorithm that transforms or inputs data for processing.
- **Estimator:** This is a machine learning algorithm that trains or fits data to build a model, which can be used for prediction.
- **Pipeline:** A pathway that connects Transformers and Estimators together to specify the ML workflow.

We will use Scikit-learn to develop predictive models of customer purchasing behavior and make strategic recommendations. The aim is to analyze data on sales, inventory, and customer purchasing behavior to make effective strategic decisions, from improving sales performance to optimizing inventory management processes.

2.2.4 PyQt6

PyQt6 is a software development kit (SDK) used to create graphical user interface (GUI) applications using the Python programming language. It is the latest version of the PyQt library, based on the Qt library, one of the most popular GUI application development frameworks for platforms including Windows, macOS, and Linux.

Additionally, we will focus on developing a graphical user interface (GUI) using PyQt6 or other tools like Tkinter. This interface will allow store managers and customers to interact intuitively with the data and analysis results. The purpose of this is to provide an easy and convenient approach, helping employees and customers easily monitor business performance, perform management tasks, and enhance the shopping experience.

2.3 Data Visualization

Data visualization is the process of converting data into charts, graphs, and other visual representations to facilitate understanding and analysis. The goal of data visualization is to help viewers easily grasp information, identify trends, patterns, and outliers in the data.

Popular tools and libraries:

- **Matplotlib:** A popular Python library for creating static charts, well-suited for simple visualizations.
- **Seaborn:** A library built on Matplotlib, offering more attractive and easier-to-use charts.
- **Plotly**: A powerful library for interactive charts.

In this project, we will use libraries like PyQtGraph, Matplotlib, and Seaborn to create charts and graphs to visualize business data and sales performance. Using these libraries helps us

easily present important information, detect trends and patterns in the data, and support data-driven decision-making effectively. Below is a detailed explanation of PyQtGraph, Matplotlib, and Seaborn to help you better understand how we apply these tools in the project.

2.3.1 Matplotlib

Matplotlib is one of the most popular Python libraries used for data visualization. It is a cross-platform library for creating 2D plots from data arrays. Matplotlib is written in Python and utilizes NumPy, Python's mathematical extension. It provides an object-oriented API for embedding plots in applications and uses Python GUI toolkits like PyQt, WxPython, or Tkinter. Additionally, it can be used within the Python and IPython shell, Jupyter Notebook, and web servers.

Matplotlib has an interface named Pylab, designed to resemble MATLAB - a proprietary programming language developed by MathWorks. Matplotlib, along with NumPy, can be considered as an open-source alternative to MATLAB.

Matplotlib was initially developed by John D. Hunter in 2003. The current stable version is 2.2.0, released in January 2018.

2.3.2 Seaborn

Seaborn is a plotting library based on Matplotlib, designed to help users create complex charts more easily. Seaborn allows users to create scatter plots with regression lines and distribution plots with fitted curves. The library can also be used to draw box plots and line plots.

We will use Matplotlib and Seaborn to design a user interface that is user-friendly and visually appealing, helping users understand analytical information more easily. These libraries will assist us in creating attractive and intuitive charts and graphs, enabling users to quickly grasp important trends and data patterns. This not only aids in presenting information clearly and vividly but also facilitates user interaction with the data. As a result, they can analyze sales performance, assess key business metrics, and make accurate and timely business decisions. The combination of Matplotlib and Seaborn will ensure that the user interface is both visually appealing and effective in conveying information.

2.3.3 PyQtGraph

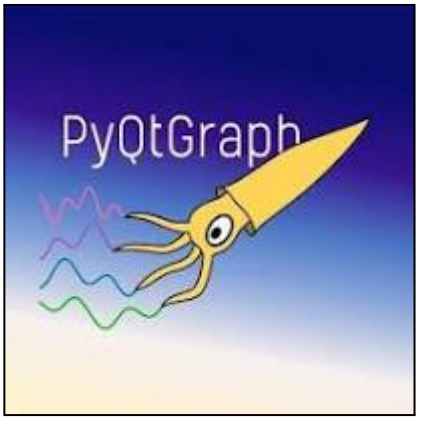


Figure 2.8 PyQtGraph Logo

source: https://x.com/pyqtgraph

PyQtGraph is another graphics library we use in this project, particularly useful for presenting real-time data and high interactivity. Unlike Matplotlib and Seaborn, this library is designed to optimize performance with large datasets, providing the ability to display and interact with data instantly.

PyQtGraph is a graphics and user interface library for Python, primarily aimed at scientific and technical applications. The library aims to fulfill two main needs: firstly, providing fast interactive graphics display for data types such as graphs, videos, and secondly, supporting rapid application development with tools like the property trees used in Qt Designer.

PyQtGraph leverages the Qt GUI platform technology through PyQt or PySide to achieve high performance in graphics processing, especially with large-scale data. This library uses the Qt GraphicsView framework, a powerful graphics system, and introduces optimized and simplified graphics primitives to enhance data visualization with less effort.

PyQtGraph is capable of running on operating systems such as Linux, Windows, and OSX, ensuring wide application for developers on various platforms.

Using PyQtGraph in this project allows us to provide a dynamic graphical user interface (GUI) where users can interact directly with charts and adjust parameters to analyze data instantly. This is especially suitable when quick analysis and decision-making are needed, enhancing work efficiency and providing the ability to gain both an overview and detailed insight into sales performance and business data.

2.4 Database Management

Database management is a fundamental aspect of data processing within an organization. It involves using software to store and manage data in a structured manner, ensuring that data is always available, reliable, and secure. Efficient database management allows organizations to maintain data integrity, improve data retrieval efficiency, and support business processes and decision-making. To effectively manage databases, understanding relational databases is crucial.

Understanding relational databases is essential for effective database management. A relational database is a type of database that stores data in tables, organized into rows and columns. Each table represents a different entity, and the relationships between these tables are established through foreign keys. This structure enables complex queries and data manipulation, providing a powerful way to manage large amounts of data and relationships. Relational databases support data consistency and minimize redundancy by ensuring that each piece of data is stored only once and referenced when needed. To fully harness the potential of these databases, using SQL is indispensable.

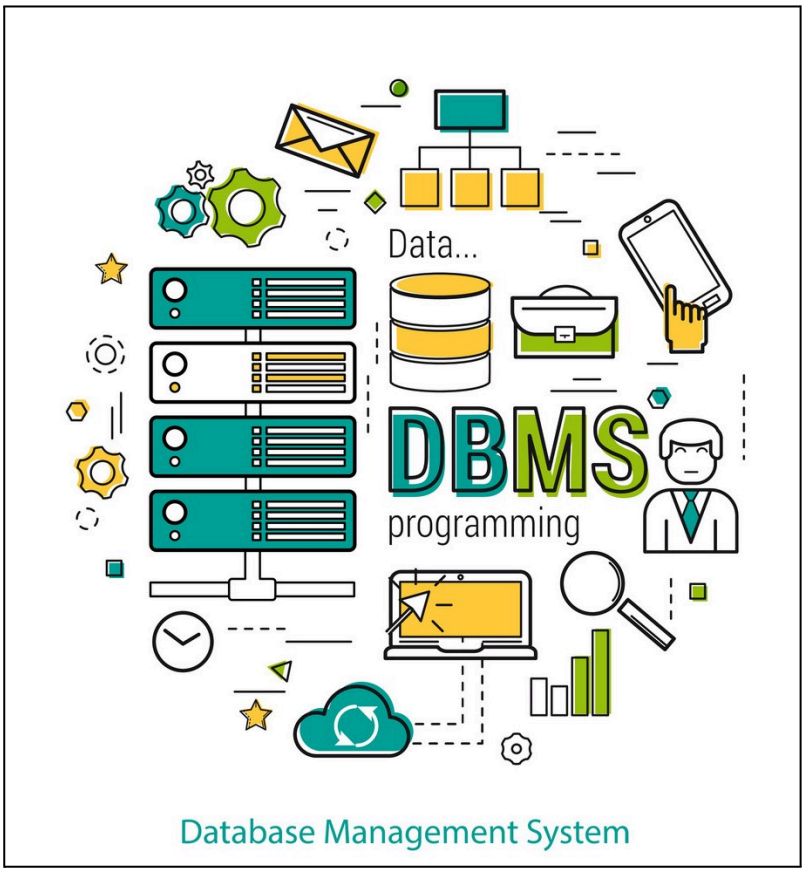


Figure 2.9 Database Management System Concept

source: https://dean2020.edu.vn/dbms/

Using SQL (Structured Query Language) is essential for querying and managing data in relational databases. SQL is the standard language for interacting with relational databases, allowing users to perform various operations such as inserting, updating, deleting, and retrieving data. By using SQL, users can write queries to extract specific data from one or multiple tables, connect tables based on relationships, and perform aggregation and computation. This capability enables efficient data analysis and reporting, helping businesses make decisions based on accurate and up-to-date information.

For a coffee shop chain, storing and managing data requires an efficient and secure database system. Using a relational database combined with SQL helps organize data in a structured manner and facilitates easy retrieval. Implementing security measures and regular data backup strategies ensures data integrity and safety, thereby supporting business

decision-making and improving operational performance of the shop. Careful organization and security standards along with regular backups are key factors in ensuring that data is always protected and ready for management and performance analysis processes.

2.4.1 MySQL Workbench

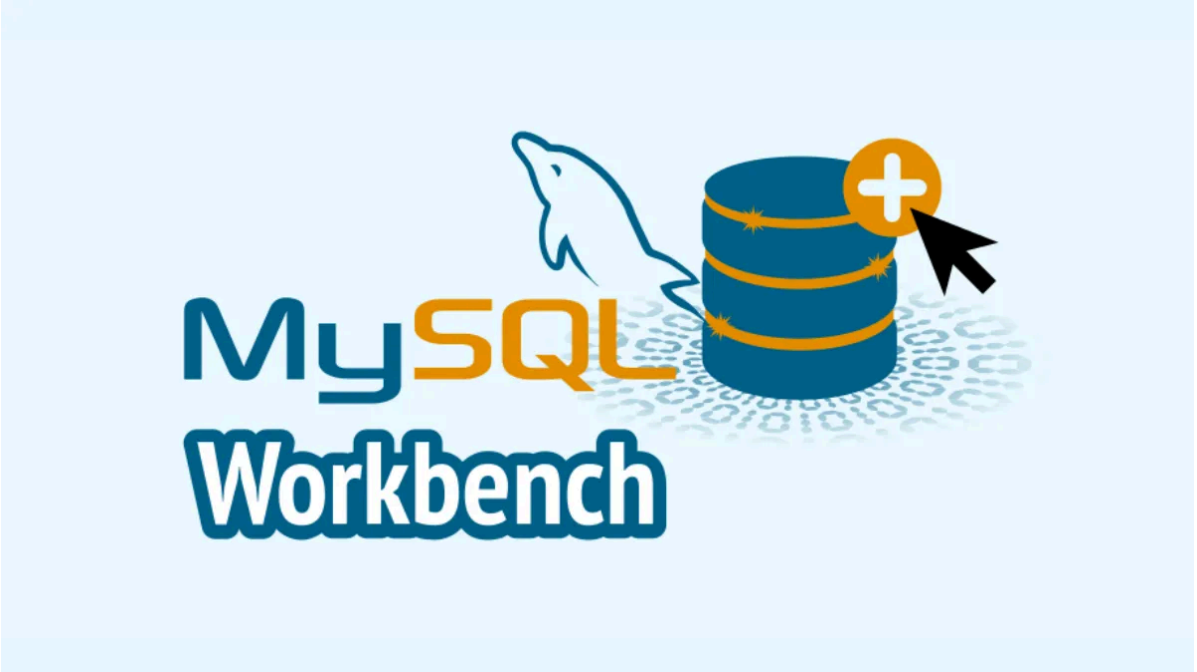


Figure 2.10 Logo MySQL Workbench

source: https://vietnix.vn/mysql-workbench/

In today's context, utilizing professional tools for managing relational databases is highly important. MySQL Workbench stands out as one of the robust tools aiding users and developers in interacting with MySQL databases through a graphical interface. It's a comprehensive tool enabling efficient database design, modeling, creation, and management. Leveraging MySQL Workbench not only enhances data management capabilities but also provides robust support for data analysis and maintenance, thereby facilitating organizations to easily execute tasks such as database design, SQL query creation, and data backup.

Integrating MySQL Workbench into the database management system of a coffee shop chain simplifies database management operations through its intuitive interface. This tool provides robust support for creating and managing relationships between tables, as well as optimizing queries, thereby enhancing data retrieval and analysis efficiency. This not only ensures data consistency and reliability but also supports decision-making processes based on updated and accurate data.

2.5 Business Analytics

2.5.1 Business Analytics

Business Analytics, also known as business analysis, is the process of analyzing and processing data from business activities to make decisions based on information to enhance efficiency and increase profitability. This process involves not only collecting data and presenting figures but also focuses on applying advanced methods such as data collection, data mining, sequence identification, text mining, forecasting, and predictive analysis. This helps understand current activities, forecast future trends, and optimize business strategies.

The business analytics process consists of three main steps: data processing and analysis, drawing conclusions from analyzed data, and finally proposing actionable decisions. Business Intelligence, an important component of Business Analytics, utilizes technologies such as machine learning and artificial intelligence to analyze and provide detailed information, supporting daily business decisions.

Figure 2.11 Relationship Between Predictive, Prescriptive, and Descriptive Analytics in Business Analysis source: https://www.predictiveanalyticstoday.com/business-analytics/

2.5.2 Business Analytics in the project

In our project, we have applied the principles of Business Analytics to study data from a hypothetical coffee shop chain in New York City. The goal is to improve strategy and optimize business operations through detailed data analysis of revenue, inventory, and

customer demographic information. This data has helped us understand trends and develop appropriate strategies to increase revenue and operational efficiency.

Data analysis has enabled us to identify top-selling products, discover effective operating hours, and gain a better understanding of changes in consumer behavior. From there, we can determine optimal pricing strategies, products, and selling times to accurately meet customer demand. Additionally, inventory management has also been improved through the use of historical data, allowing us to provide sufficient top-selling products while minimizing excess and waste.

Thanks to the application of Business Analytics, our project has not only enhanced our understanding of business operations but also established a solid foundation for making accurate strategic decisions, thereby optimizing management, sales, and marketing strategies.

CHAPTER 3. BUSINESS REQUIREMENTS ANALYSIS

3.1 Introducing the Company

Amber and Sandeep, the co-founders of a fictional coffee chain, have established three vibrant coffee shops located in New York City. Their business vision focuses on providing high-quality coffee and pastries, coupled with a unique and welcoming atmosphere that appeals to a diverse clientele. To support their business growth and ensure they make data-driven decisions, they have invested in IBM Cognos Analytics.

The coffee chain's operations are meticulously documented in a series of spreadsheets, which include data on sales receipts, pastry inventory, sales targets, customer demographics, dates, products, sales outlets, staff, and generational trends. This comprehensive data collection forms the backbone of their analytical efforts, enabling them to identify key factors that contribute to their success.

To leverage this data, Amber and Sandeep need to developed three main dashboards: Sales, Inventory, and Customer. The Sales Dashboard provides detailed insights into sales performance, helping them understand trends and patterns over time. The Inventory Dashboard assists in managing stock levels and reducing waste by providing a clear view of inventory status and sales data. The Customer Dashboard offers valuable information on customer behavior and demographics, enabling the business to tailor marketing efforts and improve customer satisfaction.

These dashboards are integral to the coffee chain's strategy, allowing Amber and Sandeep to make informed decisions that drive business growth and operational efficiency. Through their commitment to data analytics and a customer-centric approach, they continue to enhance the overall experience for their patrons, ensuring the coffee chain remains a popular choice in New York City's bustling coffee scene.

3.2 Data Description

3.2.1 Product

Purpose: Manage product information.

- Store detailed information about the product, including ID, product group, product type, description, unit of measure, wholesale and retail prices, and promotional information.
- Support product catalog management, pricing, and promotional campaigns.

Product data table with the following columns:

- product_id (INTEGER, NOT NULL, PRIMARY KEY): The primary key of the table, each value in this column is unique and not null, uniquely identifying each product.
- product_group (VARCHAR(15), NOT NULL): Product group.
- product_category (VARCHAR(18), NOT NULL): Product category.
- product_type (VARCHAR(21), NOT NULL): Product type.
- product (VARCHAR(28), NOT NULL): Product name.
- product_description (VARCHAR(95), NOT NULL): Product description.
- unit_of_measure (VARCHAR(6), NOT NULL): Unit of measure.
- current_wholesale_price (NUMERIC(5,2), NOT NULL): Current wholesale price.
- current_retail_price (NUMERIC(5,2), NOT NULL): Current retail price.
- tax_exempt_yn (BIT, NOT NULL): Whether the product is tax exempt or not.
- promo_yn (BIT, NOT NULL): Whether the product is on promotion or not.
- new_product_yn (BIT, NOT NULL): Whether the product is new or not.
- metric (VARCHAR(9), NOT NULL): Metric unit of measure.

Table 3.1 Product data description table

Data	**Data type**	**Keys**	**Required?**	**Validation**
product_id	INTEGER	PK	Yes	
product_group	VARCHAR(15)		Yes	
product_category	VARCHAR(18)		Yes	
product_type	VARCHAR(21)		Yes	
product	VARCHAR(28)		Yes	
product_description	VARCHAR(95)		Yes	

unit_of_measure	VARCHAR(6)		Yes	
current_wholesale_price	NUMERIC(5,2)		Yes	
current_retail_price	NUMERIC(5,2)		Yes	
tax_exempt_yn	BIT		Yes	
promo_yn	BIT		Yes	
new_product_yn	BIT		Yes	
metric	VARCHAR(9)		Yes	Ví dụ: "USD", "12 oz", "1 lb"

In this table:

- product_id is the primary key (Primary Key) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.2 Customer

Purpose: Manage customer information.

- Store personal information of customers such as ID, name, email, registration date, loyalty card number, birthdate, and gender.
- Support marketing activities, customer care, and loyalty programs.

Customer data table with the following columns:

- customer_id (INTEGER, NOT NULL, PRIMARY KEY): The primary key of the table, each value in this column is unique and not null, uniquely identifying each customer.

  home_store (INTEGER, NOT NULL): Code of the store where the customer frequently shops.
- customer_first_name (VARCHAR(20), NOT NULL): Customer's first name.
- customer_email (VARCHAR(25), NOT NULL): Customer's email.

- customer_since (DATE, NOT NULL): The date the customer started shopping at the store.
- loyalty_card_number (VARCHAR(12), NOT NULL): Loyalty card number of the customer.
- birthdate (DATE, NOT NULL): Customer's birthdate.
- gender (VARCHAR(1), NOT NULL): Customer's gender (M or F).
- birth_year (INTEGER, NOT NULL): Customer's birth year.

Table 3.2 Customer data description table

Data	**Data type**	**Keys**	**Required?**	**Validation**
customer_id	INTEGER	PK	Yes	
home_store	INTEGER		Yes	
customer_first_name	VARCHAR(20)		Yes	
customer_email	VARCHAR(25)		Yes	
customer_since	DATE		Yes	
loyalty_card_number	VARCHAR(12)		Yes	
birthdate	DATE		Yes	
gender	VARCHAR(1)		Yes	
birth_year	INTEGER		Yes	

In this table:

- Customer_id is the primary key (Primary Key) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.3 Pastry Inventory

Purpose: Manage pastry inventory.

- Track the quantity of pastries at the start of the day, the quantity sold, the quantity spoiled, and the spoilage percentage.
- Help manage inventory effectively, reduce waste, and optimize supply.

*** Pastry inventory data table with the following columns:**

- sales_outlet_id (INTEGER, NOT NULL): Code of the sales outlet.
- transaction_date (DATE, NOT NULL): Transaction date.
- product_id (INTEGER, NOT NULL): Product code.
- start_of_day (INTEGER, NOT NULL): Quantity of product at the start of the day.
- quantity_sold (INTEGER, NOT NULL): Quantity of product sold.
- waste (INTEGER, NOT NULL): Quantity of product wasted.
- percent_waste (INTEGER, NOT NULL): Percentage of product wasted.

Table 3.3 Pastry inventory data description table

Data	Data type	Keys	Required?	Validation
sales_outlet_id	INTEGER	PK	Yes	
transaction_date	DATE		Yes	
product_id	INTEGER		Yes	
start_of_day	INTEGER		Yes	
quantity_sold	INTEGER		Yes	
waste	INTEGER		Yes	
percent_waste	INTEGER		Yes	

In this table:

- sales_outlet_id is designated as the primary key (Primary Key).
- All columns are mandatory (Required).

3.2.4 Sales Outlet

Purpose: Manage sales outlet information.

- Store detailed information about sales outlets, including store ID, store type, square footage, address, phone number, postal code, longitude, and latitude.
- Support the management and analysis of sales outlet operations.

*** Sales outlet data table with the following columns:**

- sales_outlet_id (INTEGER, NOT NULL, PRIMARY KEY, AUTO_INCREMENT): The primary key of the table, each value in this column is unique and not null, uniquely identifying each sales outlet.
- sales_outlet_type (VARCHAR(9), NOT NULL): Type of sales outlet (e.g., warehouse, retail).
- store_square_feet (INTEGER, NOT NULL): Area of the store in square feet.
- store_address (VARCHAR(18), NOT NULL): Store address.
- store_city (VARCHAR(16), NOT NULL): City where the store is located.
- store_state_province (VARCHAR(2), NOT NULL): State or province where the store is located.
- store_telephone (VARCHAR(12), NOT NULL): Store telephone number.
- store_postal_code (INTEGER, NOT NULL): Postal code of the store.
- store_longitude (NUMERIC(10,6), NOT NULL): Longitude of the store.
- store_latitude (NUMERIC(9,6), NOT NULL): Latitude of the store.
- manager (NUMERIC(4,1)): Manager code of the store.
- Neighborhood (VARCHAR(17), NOT NULL): Neighborhood where the store is located.

Table 3.4 Sales outlet data description table

Data	**Data type**	**Keys**	**Required?**	**Validation**
sales_outlet_id	INTEGER	PK	Yes	Auto_increment
sales_outlet_type	VARCHAR(9)		Yes	

store_square_feet	INTEGER		Yes	
store_address	VARCHAR(18)		Yes	
store_city	VARCHAR(16)		Yes	
store_state_province	VARCHAR(2)		Yes	
store_telephone	VARCHAR(12)		Yes	
store_postal_code	INTEGER		Yes	
store_longitude	NUMERIC(10,6)		Yes	
store_latitude	NUMERIC(9,6)		Yes	
manager	NUMERIC(4,1)			
Neighborhood	VARCHAR(17)			

In this table:

- sales_outlet_id is the primary key (Primary Key) and mandatory (Required).
- All other columns are also mandatory (Required) except for manager.

3.2.5 Generations

Purpose: Identify customer generation based on birth year.

- Classify customers into different generations (Baby Boomers, Gen X, etc.) based on birth year.
- Assist in demographic analysis and targeted marketing by generation.

*** Generations data table with the following columns:**

- birth_year (INTEGER, NOT NULL, PRIMARY KEY, AUTO_INCREMENT): The primary key of the table, each value in this column is unique and not null, identifying the birth year of the generation.

- generation (VARCHAR(19), NOT NULL): Name of the generation (e.g., Baby Boomers, Gen X).

Table 3.5 Generations data description table

Data	Data Type	Keys	Required	Validation
birth_year	INTEGER	PK	Required	Auto_increment
generation	VARCHAR(19)		Required	

In this table:

- birth_year is the primary key (Primary Key) and mandatory (Required).
- generation is a mandatory (Required) column.

3.2.6 Dates

Purpose: Manage date information.

- Store information about transaction dates, including day, week, month, quarter, and year.
- Support data analysis over different time periods, helping to track trends and performance by week, month, quarter, and year.

*** Dates data table with the following columns:**

- transaction_date (DATE, NOT NULL, PRIMARY KEY): The primary key of the table, each value in this column is unique and not null, identifying the transaction date.
- Date_ID (DATE, NOT NULL): Date ID, often formatted as a number.
- Week_ID (INTEGER, NOT NULL): Week ID.
- Week_Desc (VARCHAR(7), NOT NULL): Description of the week (e.g., "Week 14").
- Month_ID (INTEGER, NOT NULL): Month ID.
- Month_Name (VARCHAR(5), NOT NULL): Name of the month (e.g., "April").
- Quarter_ID (INTEGER, NOT NULL): Quarter ID.
- Quarter_Name (VARCHAR(2), NOT NULL): Name of the quarter (e.g., "Q2").

Year_ID (INTEGER, NOT NULL): Year ID.

Table 3.6 Dates data description table

Data	Data Type	Keys	Required	Validation
Date_ID	DATE	PK	Required	Auto_increment
transaction_date	DATE		Required	
Week_ID	INTEGER		Required	
Week_Desc	VARCHAR(7)		Required	
Month_ID	INTEGER		Required	
Month_Name	VARCHAR(5)		Required	
Quarter_ID	INTEGER		Required	
Quarter_Name	VARCHAR(2)		Required	
Year_ID	INTEGER		Required	

In this table:

- transaction_date is the primary key (Primary Key) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.7 Staff

Purpose: Manage staff information.

- Store personal information of employees, including ID, first name, last name, position, start date, and work location.
- Support human resource management and track employees' work history.

*** Staffs data table with the following columns:**

- staff_id (INTEGER, NOT NULL, PRIMARY KEY, AUTO_INCREMENT): The primary key of the table, each value in this column is unique and not null, uniquely identifying each staff member.
- first_name (VARCHAR(8), NOT NULL): Staff member's first name.
- last_name (VARCHAR(8), NOT NULL): Staff member's last name.

- position (VARCHAR(15), NOT NULL): Staff member's position.
- start_date (DATE, NOT NULL): Staff member's start date.
- location (VARCHAR(2), NOT NULL): Staff member's work location.

Table 3.7 Staff data description table

Data	Data Type	Keys	Required	Validation
staff_id	INTEGER	PK	Yes	Auto_Increment
first_name	VARCHAR(8)		Yes	
last_name	VARCHAR(8)		Yes	
position	VARCHAR(15)		Yes	
start_date	DATE		Yes	
location	VARCHAR(2)		Yes	

In this table:

- staff_id is the primary key (Primary Key) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.8 Sales Targets

Purpose: Manage sales targets.

- Store sales goals for each store by product category (coffee beans, beverages, food, merchandise) and total sales.
- Help track and evaluate performance against set targets.

* **Sales targets data table with the following columns:**

- sales_outlet_id (INTEGER, NOT NULL): Code of the sales outlet.
- beans_goal (INTEGER, NOT NULL): Sales target for coffee beans.
- beverage_goal (INTEGER, NOT NULL): Sales target for beverages.
- food_goal (INTEGER, NOT NULL): Sales target for food.
- merchandise_goal (INTEGER, NOT NULL): Sales target for merchandise.

- total_goal (INTEGER, NOT NULL): Total sales target.
- Date_ID (INTEGER, NOT NULL): Date code.
- Year_ID (INTEGER, NOT NULL): Year code.

Table 3.8 Sales Targets data description table

Data	**Data Type**	**Keys**	**Required**	**Validation**
sales_outlet_id	INTEGER	FK	Required	References `sales_outlet(sales_outlet_id)`
beans_goal	INTEGER		Required	
beverage_goal	INTEGER		Required	
food_goal	INTEGER		Required	
merchandise_goal	INTEGER		Required	
total_goal	INTEGER		Required	
Date_ID	INTEGER		Required	
Year_ID	INTEGER	FK	Required	`References dates(Year_ID)`

In this table:

- sales_outlet_id and Year_ID are foreign keys (Foreign Keys) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.9 Sales

Purpose: Manage sales transaction information.

- Record details of sales transactions, including transaction ID, date and time of transaction, store ID, staff ID, customer ID, product information, and quantity.
- Support sales analysis, track sales performance, and shopping trends.

*** Sales data table with the following columns:**

- transaction_id (INT): Transaction ID.
- transaction_date (DATE): Transaction date.
- transaction_time (TIME): Transaction time.
- sales_outlet_id (INT): Sales outlet code.
- staff_id (INT): Staff code.
- customer_id (INT): Customer code.
- instore_yn (BOOLEAN): Whether the transaction was made in-store.
- order_id (INT): Order ID.
- line_item_id (INT): Line item ID.
- product_id (INT): Product code.
- quantity (INT): Product quantity.
- line_item_amount (FLOAT): Line item amount.
- unit_price (FLOAT): Unit price of the product.
- promo_item_yn (BOOLEAN): Whether the product is part of a promotion.

Table 3.9 Sales data description table

Data	**Data type**	**Keys**	**Required**	**Validation**
transaction_id	INTEGER		Yes	
transaction_date	DATE		Yes	
transaction_time	TIME		Yes	
sales_outlet_id	INTEGER	FK	Yes	References sales_outlet(sales_outlet_id)
staff_id	INTEGER	FK	Yes	References staffs(staff_id)
customer_id	INTEGER	FK	Yes	References customer(customer_id)
instore_yn	BOOLEAN		Yes	
order_id	INTEGER		Yes	
line_item_id	INTEGER		Yes	
product_id	INTEGER	FK	Yes	References `product(product_id)`
quantity	INTEGER		Yes	
line_item_amount	FLOAT		Yes	

unit_price	FLOAT		Yes	
promo_item_yn	BOOLEAN		Yes	

In this table:

- sales_outlet_id, staff_id, customer_id, and product_id are foreign keys (Foreign Keys) and mandatory (Required).
- All other columns are also mandatory (Required).

3.2.10 Transaction_times

Purpose: Manage transaction time information.

- Store specific times of transactions and categorize them by part of the day (morning, afternoon, evening).
- Support transaction data analysis by time of day, helping to optimize service times and sales strategies.

* **Transaction times data table with the following columns:**

- transaction_id (INT): Transaction ID.
- transaction_time (TIME): Transaction time.
- part_of_day (VARCHAR(20)): Part of the day when the transaction occurred (e.g., Morning, Afternoon, Evening).

Table 3.10 Transaction_times data description table

Data	**Data type**	**Keys**	**Required?**	**Validation**
transaction_id	INTEGER	FK	Yes	References `sales(transaction_id)`
transaction_time	TIME		Yes	

part_of_day	VARCHAR(20)		Yes	

In this table:

- sales_outlet_id, staff_id, customer_id, and product_id are foreign keys (Foreign Keys) and mandatory (Required).
- All other columns are also mandatory (Required).

3.3 Data processing

In this study, data processing is a critical step to ensure the quality and usability of the dataset for further analysis. We utilized the Python library pandas for this purpose. Pandas is a powerful and flexible tool for data manipulation and analysis, widely used in data science and machine learning projects. It provides data structures and functions needed to efficiently clean, transform, and analyze data.

The primary reasons for choosing pandas include its ease of use, performance, and the extensive range of functionalities it offers. With pandas, we can handle large datasets, perform complex operations with simple commands, and seamlessly integrate with other libraries such as NumPy and Matplotlib.

In our data processing workflow, pandas was employed to:

- Load the Dataset: Import data from various file formats, including CSV and Excel, into a structured DataFrame.
- Data Cleaning: Handle missing values, remove duplicates, and correct inconsistencies in the data.
- Data Transformation: Convert data types, normalize values, and perform aggregations to prepare the data for analysis.
- Data Exploration: Perform exploratory data analysis (EDA) to understand the dataset's structure and identify patterns and anomalies.
- Feature Engineering: Create new features or modify existing ones to enhance the predictive power of our models.

By using pandas, we ensured that the data was in an optimal state for the subsequent steps of our analysis, thereby enhancing the accuracy and reliability of our results.

3.4 Converting the Dataset from CSV to SQL

To facilitate more efficient querying and data management, we converted our dataset from a CSV file to an SQL format. This step was essential for leveraging the powerful data handling capabilities of relational databases, particularly MySQL.

We utilized the online tool provided by ConvertCSV.com (https://www.convertcsv.com/csv-to-sql.htm) for this conversion process. This tool offers a straightforward and user-friendly interface for transforming CSV data into SQL commands suitable for database insertion.

The conversion process involved the following steps:

- Upload CSV File: The CSV file containing the processed data was uploaded to the ConvertCSV.com website.
- Configure Settings: The tool provided options to specify the table name, column types, and other SQL-specific settings. These configurations ensured that the resulting SQL file would be compatible with our MySQL database schema.
- Generate SQL File: After configuring the settings, the tool generated an SQL file with the appropriate CREATE TABLE and INSERT INTO statements.
- Download and Import: The generated SQL file was downloaded and then imported into our MySQL database using standard MySQL commands or tools like phpMyAdmin.
- By converting our dataset to SQL, we were able to take advantage of MySQL's robust querying capabilities, enabling more complex data analysis and manipulation tasks. This conversion also facilitated better data integrity, easier data sharing, and more efficient data management within our research project.

By converting our dataset to SQL, we were able to take advantage of MySQL's robust querying capabilities, enabling more complex data analysis and manipulation tasks. This conversion also facilitated better data integrity, easier data sharing, and more efficient data management within our research project.

3.5 Connecting to MySQLite

To effectively manage and analyze our dataset, we utilized SQLite, a lightweight and efficient database management system. SQLite is an ideal choice for applications that require a self-contained, serverless, and zero-configuration database engine. It is widely used for its simplicity, reliability, and extensive support across various programming environments.

Database Management with SQLite:

- Setting Up the Database: We initialized an SQLite database to store our converted SQL data. This involved creating a new database file and defining the necessary schema to accommodate our dataset.
- Importing Data: Using the SQL file generated from the CSV conversion process, we imported the data into SQLite. This step was crucial for ensuring that our data was structured correctly and ready for querying.

- Database Connection: To interact with the SQLite database, we established a connection using Python's sqlite3 library. This library provides a straightforward API for executing SQL commands and retrieving results within a Python environment.
- Executing Queries: With the connection established, we performed various SQL queries to manipulate and analyze the data. This included data retrieval, updates, deletions, and complex joins, enabling a comprehensive examination of our dataset.
- Data Integrity and Security: SQLite ensures data integrity through its ACID-compliant transaction model. This means that all database transactions are processed reliably, safeguarding our data against corruption. Additionally, SQLite's simplicity and self-contained nature enhance security by reducing the attack surface compared to more complex database systems.

Using SQLite offers several key benefits. Its ease of use, with minimal configuration and setup, integrates smoothly into our data processing workflow. The portability of storing the entire database in a single file simplifies transferring and sharing across systems, essential for collaborative projects. Optimized for fast read operations, SQLite enhances performance for large dataset queries, ensuring efficient data analysis. It scales well for single-user or small-to-medium applications, fitting our research project's needs. Moreover, SQLite's ACID-compliant transaction model guarantees data integrity and reliability, while its simplicity and self-contained nature enhance security by reducing the attack surface compared to more complex systems. These features collectively make SQLite a robust and efficient choice for managing our dataset.

3.6 Business Requirements

Based on the provided dataset and context of the fictional coffee chain, the following business requirements have been identified and grouped into four main categories.

3.6.1 Data Integration and Accessibility

The first requirement is to integrate data from multiple sources, ensuring that all ten CSV files (Sales Receipts, Pastry Inventory, Sales Targets, Customer, Dates, Product, Sales Outlet, Staff, Generation, Transactions Time) are combined into a coherent data module. This integration aims to enable seamless access to comprehensive data for analysis and reporting purposes. The success of this requirement will be measured by the successful integration and verification of data from all sources without data loss or inconsistencies.

3.6.2 Dashboard Development

The coffee chain has developed three main dashboards to monitor different aspects of its operations. The Sales Dashboard provides insights into sales performance, including total sales by date, sales growth rates by different factors, and sales distribution across outlets. The Inventory Dashboard tracks inventory levels, total quantity sold, waste percentages, and sales targets to ensure efficient inventory management and reduce waste. The Customer Dashboard

analyzes customer behavior, segmenting data by gender, age group, and purchase patterns to inform marketing strategies and improve customer satisfaction. These dashboards will help in making informed decisions by providing real-time, accurate data visualizations.

3.6.3 Machine Learning for Sales Prediction

A significant requirement is to develop and deploy machine learning models to predict future sales based on historical data. This will enhance sales forecasting accuracy, supporting inventory planning and marketing strategies. The effectiveness of this requirement will be measured by the accuracy and reliability of the sales predictions, determined by comparing the predicted values with actual sales data.

3.6.4 Customer and Performance Analysis

Customer segmentation and analysis using data from the Customer table are essential for personalizing marketing efforts and improving customer satisfaction. Understanding different customer segments will help in tailoring marketing campaigns, with success metrics including an increase in targeted marketing campaign success rate and customer satisfaction scores. Defining and implementing key performance indicators (KPIs) for both operations and marketing dashboards is also crucial. These KPIs will provide measurable benchmarks to assess business performance, with timely achievement and regular review of KPIs being the main metrics.

Lastly, ensuring robust data security measures to protect sensitive customer and business information is paramount. This requirement aims to maintain customer trust and comply with relevant data protection regulations, with success measured by the absence of data breaches or unauthorized access incidents and compliance with data protection standards.

These consolidated business requirements will guide the development and implementation of data-driven solutions for the fictional coffee chain, enhancing their operational efficiency, marketing effectiveness, and predictive capabilities.

CHAPTER 4. DEVELOPING VISUALIZATION AND MACHINE LEARNING FUNCTIONALITY IN THE APPLICATION

4.1 About BMS - BAMOS Management Suite App

The BAMOS Management Suite (BMS) App is a comprehensive solution designed to streamline and enhance the management of Amber and Sandeep's coffee chain operations. This robust application addresses various business needs through its multifaceted features, ensuring efficient and effective management of the coffee shops.

Key Features of BMS App:

- Login: The BMS App includes a secure login feature that ensures only authorized personnel can access the system. This feature is crucial for maintaining data security and confidentiality, protecting sensitive business information from unauthorized access.

- Data Management: The data management feature allows users to manage and organize data efficiently. This includes the ability to upload, update, and maintain various datasets related to sales receipts, inventory, customer demographics, and more. This feature ensures that all data is up-to-date and readily available for analysis and decision-making.

- Report Feature - Dashboard Visualization: The app offers powerful dashboard visualization capabilities, enabling users to create and view interactive dashboards. These dashboards provide real-time insights into various aspects of the business, such as sales performance, inventory levels, and customer behavior. The visualizations help in identifying trends, making data-driven decisions, and monitoring the overall health of the business.

- Machine Learning: The machine learning feature of the BMS App is designed to enhance predictive analytics capabilities. By leveraging historical data, the app can generate accurate sales forecasts, predict inventory needs, and identify potential market opportunities. This feature helps in proactive planning and optimizing business operations.

- Settings: The settings feature allows users to customize the app according to their preferences and business needs. This includes configuring user roles and permissions, setting up notification alerts, and adjusting various system parameters to ensure the app operates optimally for the business.

Through these comprehensive features, the BMS App provides a centralized platform for managing the coffee chain's operations. By integrating secure access, efficient data management, insightful dashboard visualizations, advanced machine learning capabilities,

and customizable settings, the app empowers Amber and Sandeep to make informed decisions and drive their business forward effectively.

4.2 Proposal Model

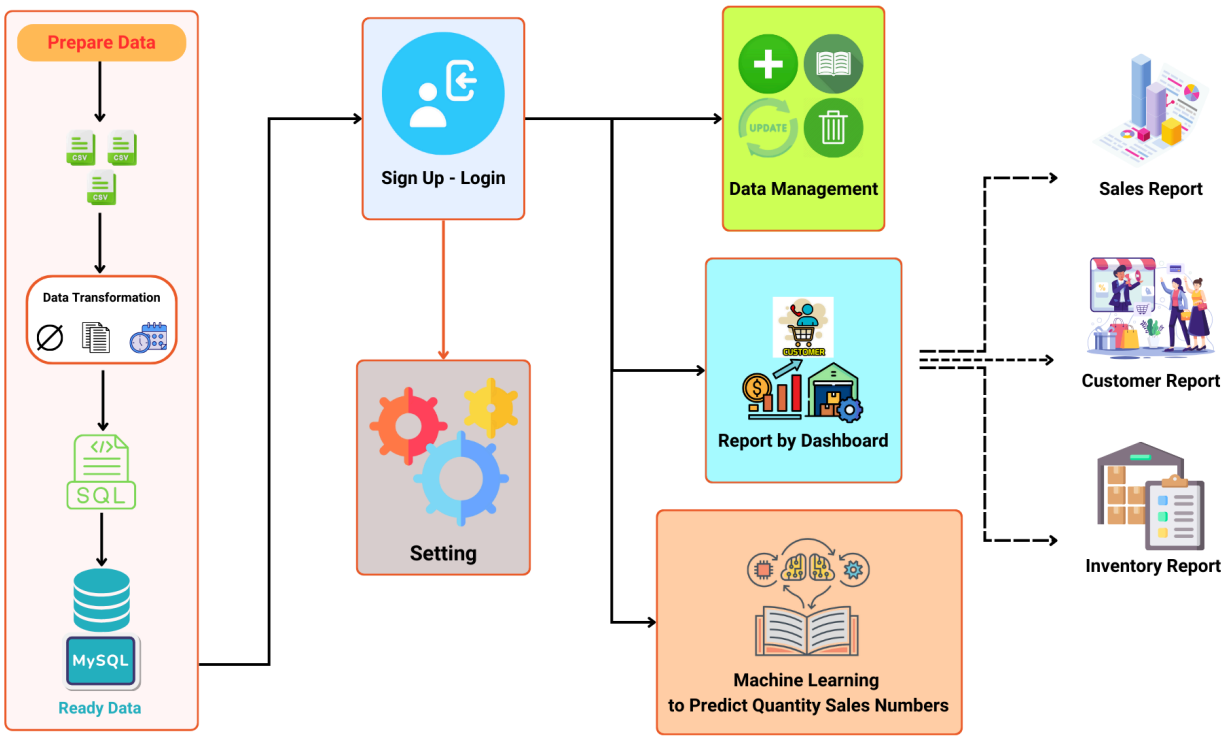


Figure 4.1 - The BAMOS Management Suite (BMS) App Proposal Model

The BAMOS Management Suite (BMS) App is designed to provide a comprehensive solution for managing the various operations of a coffee shop chain. This proposal model outlines the key components and functionalities of the app, highlighting how they interconnect to streamline business processes and improve decision-making.

4.2.1 Data Preparation and Integration

The foundation of the BMS App is built on robust data preparation and integration processes. The initial step involves collecting data from various sources, typically in CSV format. This raw data undergoes a transformation process to ensure it is clean, consistent, and ready for analysis. Data transformation includes tasks such as removing duplicates, handling missing values, and formatting data correctly. Once transformed, the data is stored in a MySQL database, making it accessible for further processing and analysis.

4.2.2 User Authentication and Access Control

The next critical component of the BMS App is the user authentication and access control system, which includes a secure Login - Sign Up feature. This system ensures that only authorized users can access the app. Users can log in using their credentials, while new users can sign up and verify their accounts through email. The system supports role-based access control, assigning different levels of access and permissions based on user roles (e.g., admin,

manager, staff). This feature ensures secure and efficient management of user access to the system.

4.2.3 Setting Feature for Staff Management

The Setting tab within the BMS App is dedicated to managing staff members. This feature allows managers to search for staff members by ID, view and update personal and account information, add new staff members, and delete records of those no longer employed. By providing a user-friendly interface for these tasks, the Setting tab helps maintain an organized and up-to-date record of all employees, ensuring accurate staff management.

4.2.4 Data Management and Reporting

The core functionality of the BMS App revolves around data management and reporting. The data management feature allows users to add, update, and maintain datasets related to sales, inventory, and customer information. This organized data is then used to generate detailed reports and dashboards. The app includes three main dashboards: Sales, Inventory, and Customer. These dashboards provide real-time insights into sales performance, inventory levels, and customer behavior, respectively. Managers can use these insights to make data-driven decisions that enhance business operations.

4.2.5 Machine Learning for Sales Prediction

To further enhance decision-making, the BMS App incorporates machine learning capabilities. The app uses historical data to develop predictive models that forecast future sales. These predictions help in planning inventory, optimizing stock levels, and identifying potential market opportunities. The machine learning feature is integral for proactive business planning and improving operational efficiency.

The BAMOS Management Suite (BMS) App integrates data preparation, user authentication, staff management, data management, reporting, and machine learning into a cohesive platform. This proposal model demonstrates how each component contributes to the overall functionality of the app, providing a comprehensive solution for managing the coffee shop chain's operations. By leveraging the BMS App, Amber and Sandeep can make informed decisions, streamline processes, and drive their business towards greater success.

4.3 Login - Sign Up Feature

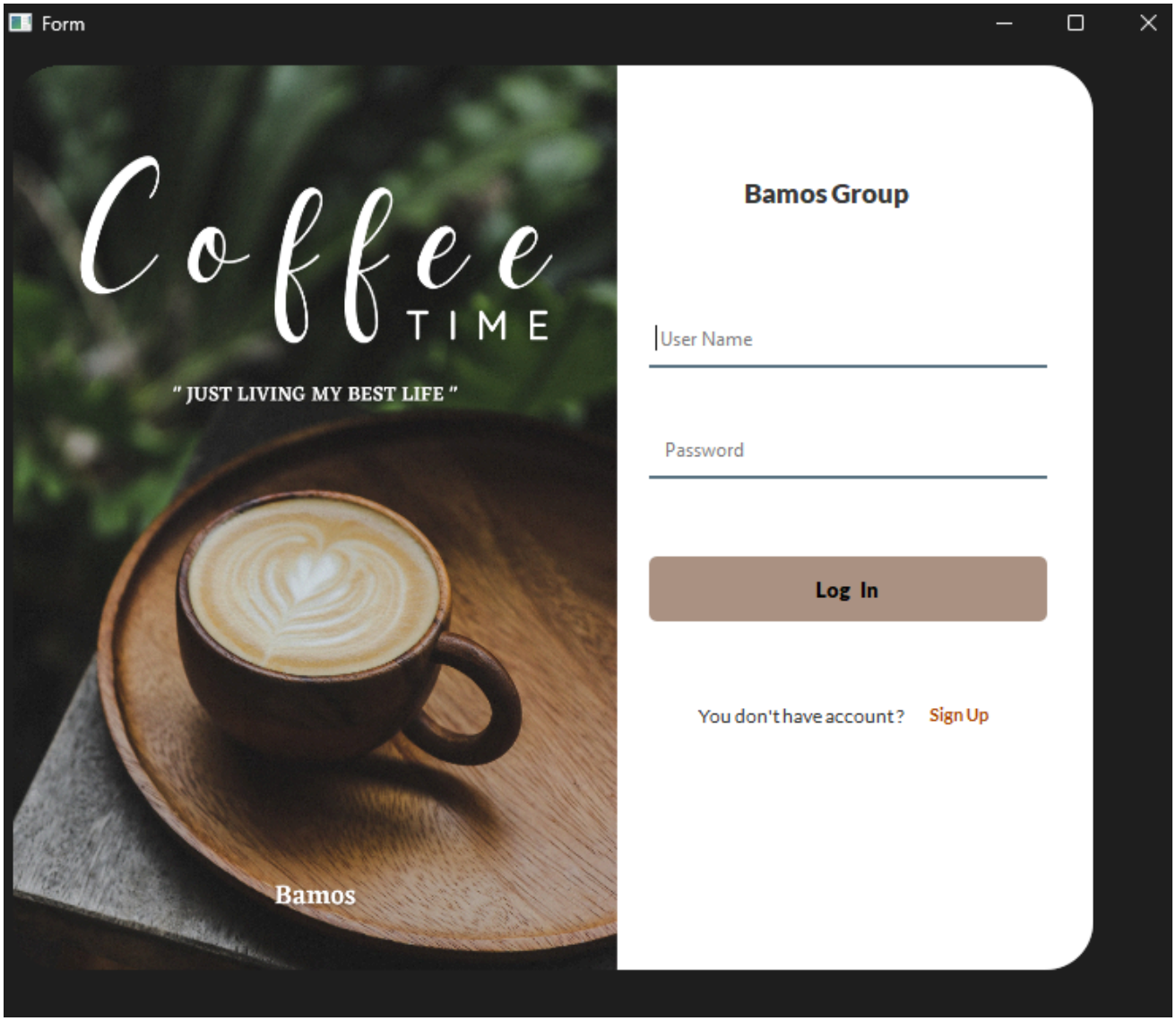


Figure 4.2 - Login - Sign Up Screen

The Login - Sign Up feature of the BAMOS Management Suite (BMS) App is designed to provide secure and user-friendly access control to the system. This feature ensures that only authorized users can access the application while also allowing new users to register and gain access efficiently.

The user authentication component includes a secure login process where existing users can log in using their credentials (username and password). The system employs encrypted communication to protect user credentials during the login process, ensuring the confidentiality of sensitive information. Additionally, the app offers a "Forgot Password" feature, allowing users to reset their passwords securely. This process involves verifying their identity through email or SMS verification codes before permitting a password reset.

For user registration, new users can create an account by providing necessary information such as their name, email address, and password. The registration process includes email

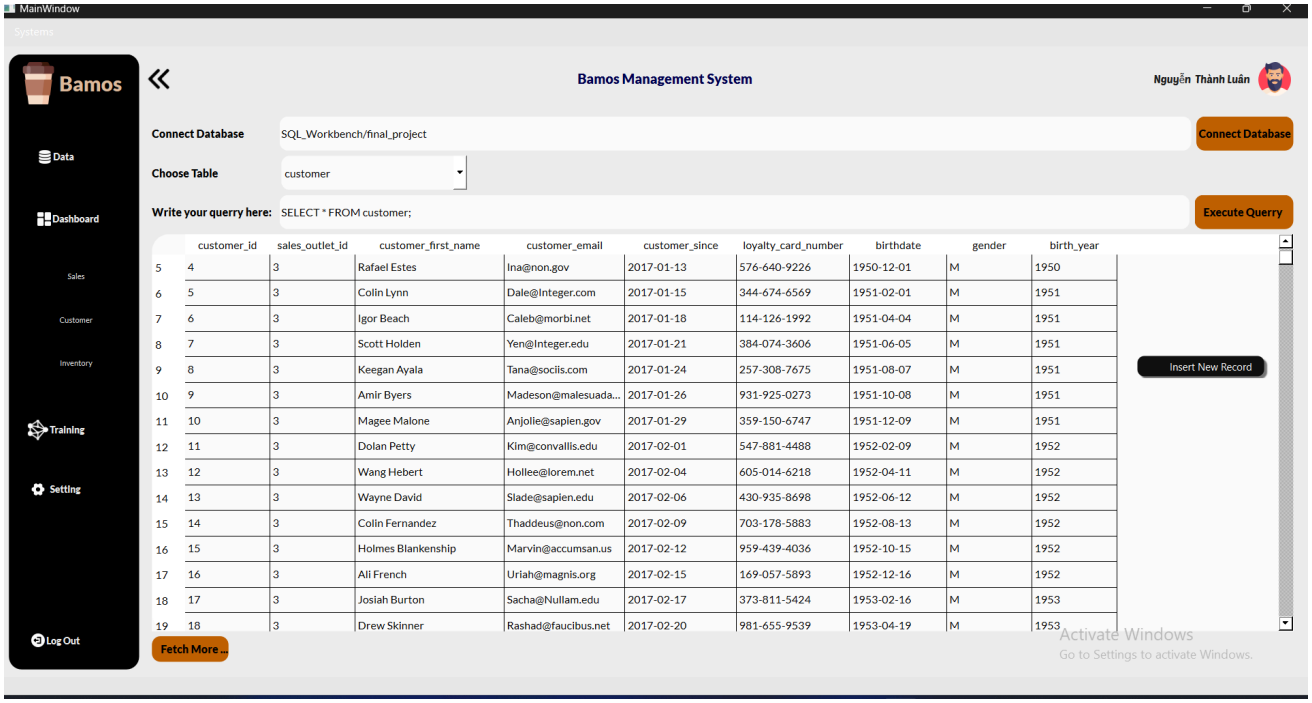
verification to confirm the user's identity and ensure the validity of the provided email address. During registration, users can be assigned specific roles (e.g., admin, manager, staff) that determine their access level and permissions within the app. This role-based access control ensures that users only have access to the features and data relevant to their responsibilities. When a new account is created, the information is automatically updated in the database in real time. If a duplicate account is detected, the system will notify the user, preventing multiple accounts for the same individual.

Security measures within the Login - Sign Up feature include encryption of all sensitive data, such as passwords, using industry-standard encryption algorithms. To enhance security further, the BMS App can incorporate multi-factor authentication (MFA), requiring users to verify their identity through a second factor, such as a code sent to their mobile device, in addition to their password.

User management capabilities within the app allow users to manage their profiles, update personal information, and change their passwords. This ensures that user information remains accurate and up-to-date. Additionally, the app maintains logs of user access and activities, providing administrators with insights into who accessed the system and what actions were performed. This audit trail helps in monitoring security and identifying any unauthorized access attempts.

By incorporating these components, the Login - Sign Up feature of the BMS App ensures secure, efficient, and user-friendly access to the system. It provides a robust foundation for managing user authentication and authorization, contributing to the overall security and integrity of the BAMOS Management Suite.

4.4 Data Management Feature



	customer_id	sales_outlet_id	customer_first_name	customer_email	customer_since	loyalty_card_number	birthdate	gender	birth_year
5	4	3	Rafael Estes	lna@non.gov	2017-01-13	576-640-9226	1950-12-01	M	1950
6	5	3	Colin Lynn	Dale@Integer.com	2017-01-15	344-674-6569	1951-02-01	M	1951
7	6	3	Igor Beach	Caleb@morbi.net	2017-01-18	114-126-1992	1951-04-04	M	1951
8	7	3	Scott Holden	Yen@Integer.edu	2017-01-21	384-074-3606	1951-06-05	M	1951
9	8	3	Keegan Ayala	Tana@soclis.com	2017-01-24	257-308-7675	1951-08-07	M	1951
10	9	3	Amir Byers	Madeson@malesuada...	2017-01-26	931-925-0273	1951-10-08	M	1951
11	10	3	Magee Malone	Anjolie@sapien.gov	2017-01-29	359-150-6747	1951-12-09	M	1951
12	11	3	Dolan Petty	Kim@convallis.edu	2017-02-01	547-881-4488	1952-02-09	M	1952
13	12	3	Wang Hebert	Hollee@lorem.net	2017-02-04	605-014-6218	1952-04-11	M	1952
14	13	3	Wayne David	Slade@sapien.edu	2017-02-06	430-935-8698	1952-06-12	M	1952
15	14	3	Colin Fernandez	Thaddeus@non.com	2017-02-09	703-178-5883	1952-08-13	M	1952
16	15	3	Holmes Blankenship	Marvin@accumsan.us	2017-02-12	959-439-4036	1952-10-15	M	1952
17	16	3	Ali French	Uriah@magnis.org	2017-02-15	169-057-5893	1952-12-16	M	1952
18	17	3	Josiah Burton	Sacha@Nullam.edu	2017-02-17	373-811-5424	1953-02-16	M	1953
19	18	3	Drew Skinner	Rashad@faucibus.net	2017-02-20	981-655-9539	1953-04-19	M	1953

Figure 4.3 Data Management Feature

The Data Management feature of the BAMOS Management Suite (BMS) App is designed to provide a robust and user-friendly interface for managing various data aspects of the coffee shop chain. This feature is essential for maintaining accurate and up-to-date records across multiple tables within the database.

Key Components of the Data Management Feature:

- Database Connection: The Data Management tab allows users to connect to the relevant database. Users can select the specific database (e.g., SQL_Workbench/final_project) to ensure they are working with the correct dataset.
- Table Selection: Users can choose from a dropdown menu to select the specific table they wish to manage. The available tables include customer, dates, pastry inventory, product, sales, sales outlet, sales targets, staff, and transaction times. This flexibility allows managers to focus on the data relevant to their current tasks.
- Data Querying: The feature includes a query input field where users can write and execute SQL queries to retrieve specific data from the selected table. This functionality enables customized data retrieval, allowing users to filter and sort data based on specific criteria.
- Viewing and Editing Data: The Data Management tab displays the retrieved data in a table format, providing a clear and organized view of the information. Users can perform various actions, such as inserting new rows, updating existing records, or removing selected rows. The interface supports real-time data updates, ensuring that any changes made are immediately reflected in the database.

- Insert New Records: Users can add new records to the selected table by entering the required information and choosing the appropriate insert option (e.g., Insert First, Insert Last, Insert Above, Insert Below). This functionality ensures that new data is promptly added to the database, keeping records current and comprehensive.
- Update Existing Records: Managers can update the details of existing records by selecting a row and modifying the relevant fields. Once the changes are made, clicking the Update button ensures that the database is accurately updated in real-time.
- Delete Records: The Data Management feature allows users to delete records that are no longer needed. By selecting the desired row and clicking the Delete button, managers can remove outdated or incorrect data, ensuring that the database remains clean and relevant.
- Real-Time Synchronization: All actions performed within the Data Management tab are synchronized in real-time with the database. This ensures that any additions, updates, or deletions are immediately reflected across all connected systems and dashboards.

User Interface Elements:

- Connect Database Button: Enables users to connect to the specified database.
- Choose Table Dropdown: Allows selection of the table to manage.
- Query Input Field and Execute Button: For writing and executing SQL queries.
- Data Table View: Displays the retrieved data in a tabular format for easy viewing and editing.
- Action Buttons (Insert, Update, Delete): Facilitates data manipulation tasks, ensuring streamlined and efficient data management.

By incorporating these components, the Data Management feature of the BMS App provides a comprehensive solution for managing the coffee shop chain's data. This feature enhances the accuracy and efficiency of data handling, allowing managers to focus on making informed decisions based on reliable and up-to-date information.

4.5 Report Feature - Dashboard Visualization

The Report Feature of the BAMOS Management Suite (BMS) App provides comprehensive dashboard visualizations, offering insightful and actionable data to manage the coffee shop's operations effectively. The application includes three main dashboards: Sales, Customer, and Inventory, each serving distinct purposes to aid in data-driven decision-making.

4.5.1 Sales Dashboard

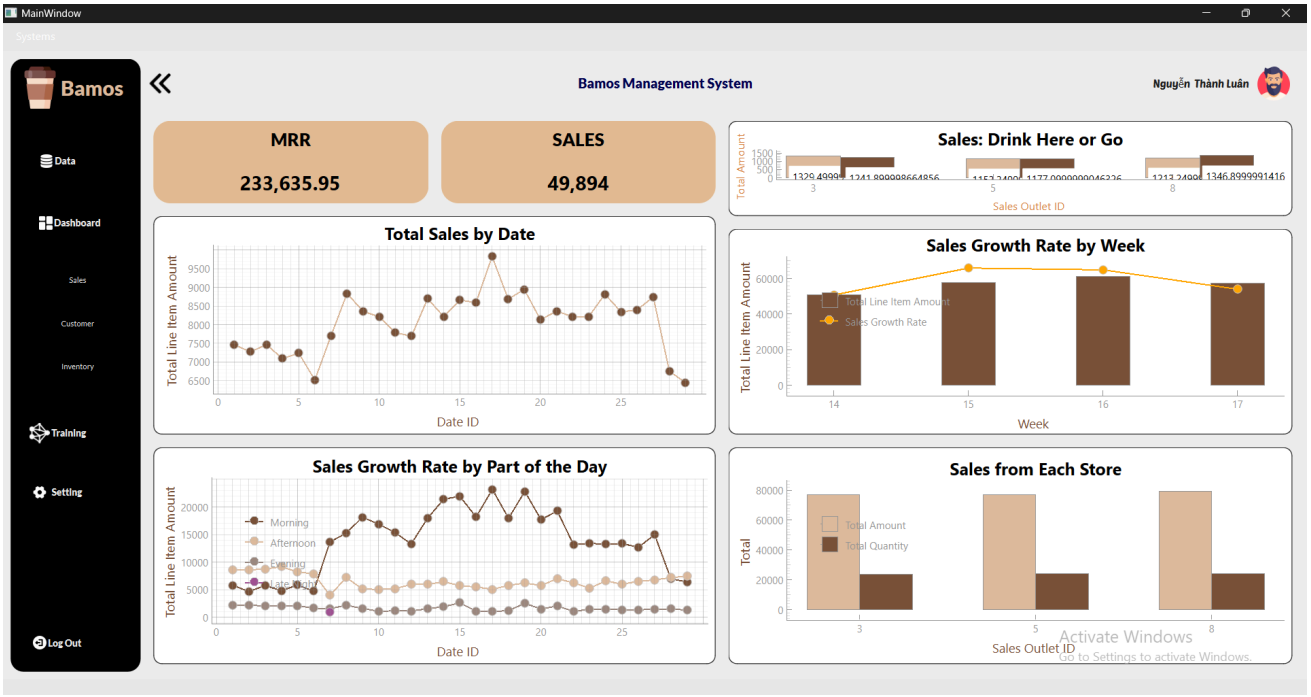


Figure 4.4 Sales Dashboard

The Sales Dashboard delivers detailed insights into the coffee shop's sales performance. It features various visualizations such as total sales by date, sales growth rate by part of the day, and a comparison of in-store versus takeaway sales. These charts help identify sales trends, peak sales periods, and the performance of different sales outlets. Additionally, the dashboard provides a weekly sales growth rate and a breakdown of sales from each store, enabling managers to track sales progress and make informed decisions to optimize revenue.

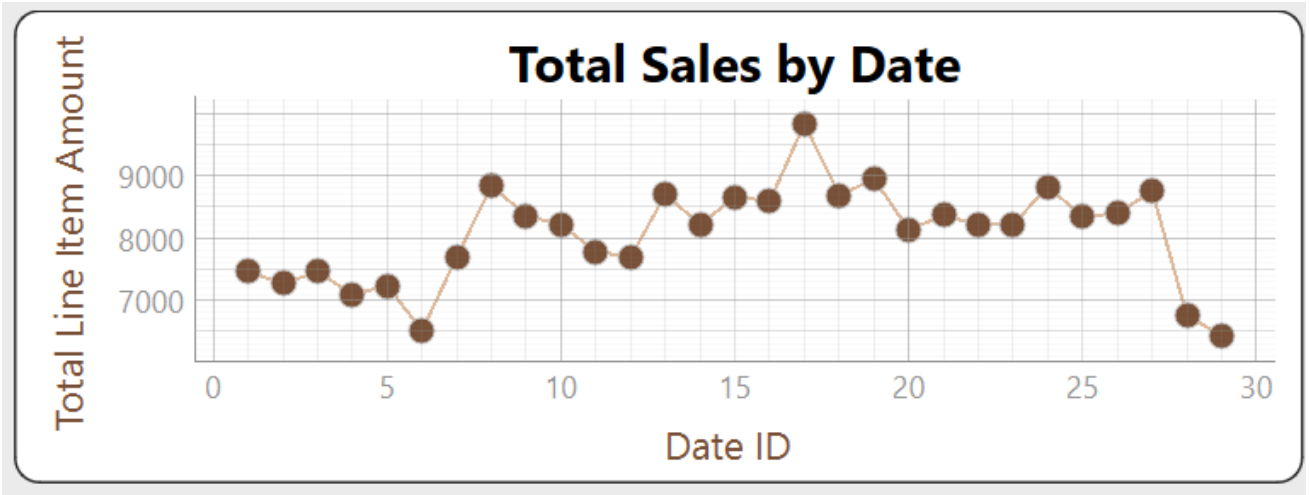


Figure 4.5 Total Sales by Date in April 2019

The line chart in the sales dashboard shows key trends for April 2019. Managers can see a sharp rise mid-month, with the highest sales around the 14th-15th, and significant drops around the 5th and 30th. This helps in identifying peak sales periods and potential issues.

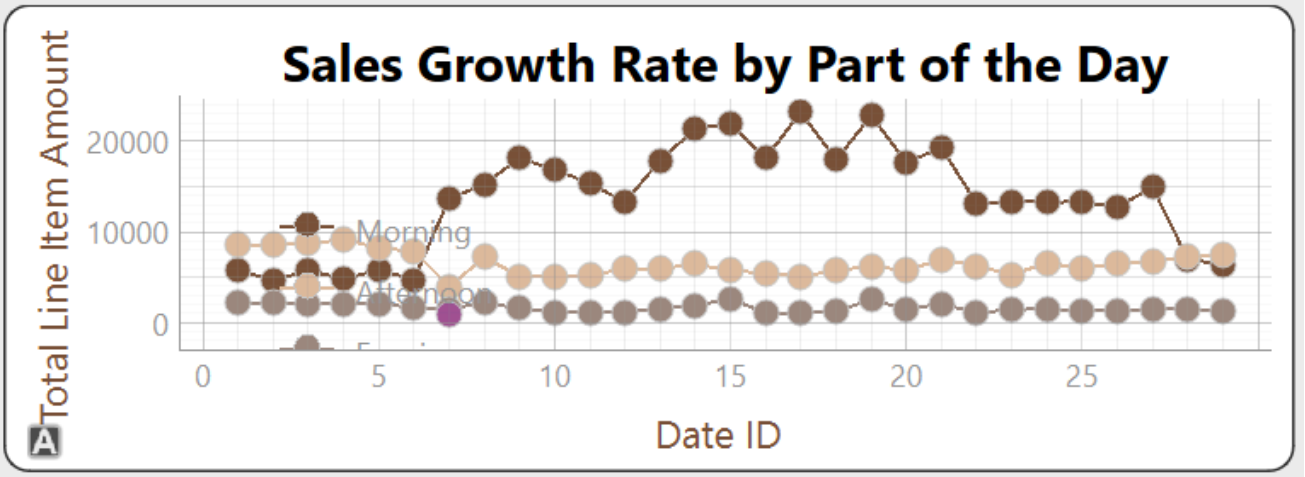


Figure 4.6 Sales Growth Rate by Part of the Day

The chart for sales growth rate by part of the day in April 2019 highlights that morning sales are significantly higher, peaking around the 15th. Afternoon and evening sales remain stable but lower, while night sales are the lowest. This chart is crucial in a sales dashboard, helping managers identify the most profitable times of the day and optimize staffing and marketing efforts accordingly.

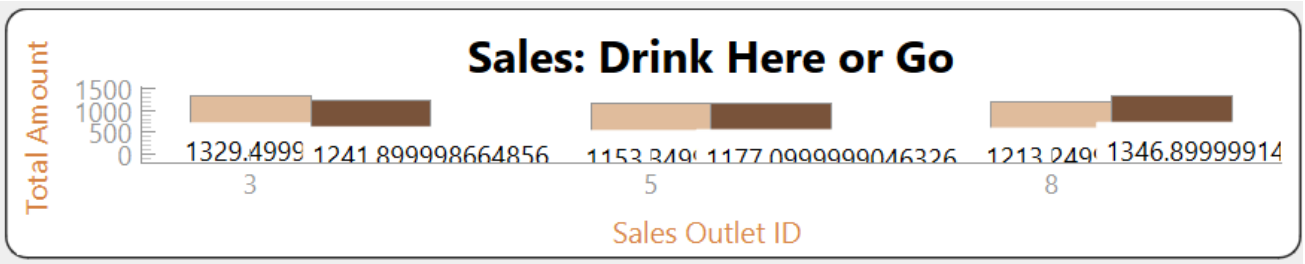


Figure 4.7 Sales: Drink Here or Go

The sales chart comparing in-store vs. takeaway highlights that takeaway sales generally surpass in-store sales, with notable differences across outlets. Outlet 3 shows higher takeaway sales, while Outlet 8 leads in in-store sales. This chart's inclusion in the sales dashboard is vital for managers to understand customer preferences, tailor services, and enhance sales strategies across different outlets.

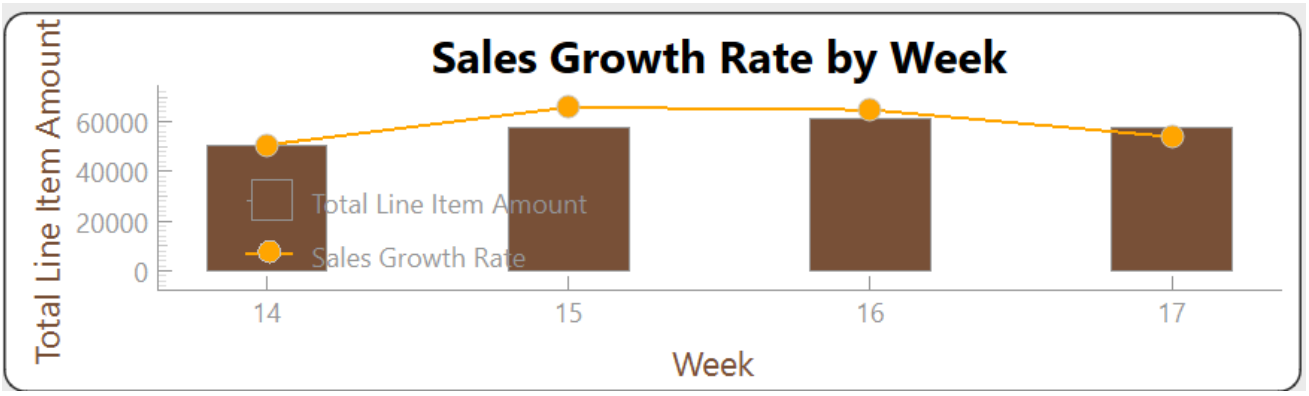


Figure 4.8 Sales Growth Rate by Week

The chart for sales growth rate by week shows total sales amounts and growth rates across weeks 14 to 17. Week 16 records the highest sales, while weeks 14 and 17 have lower totals. This chart is vital for the sales dashboard, allowing managers to track weekly performance trends and identify peak sales periods for better resource allocation and strategic planning.

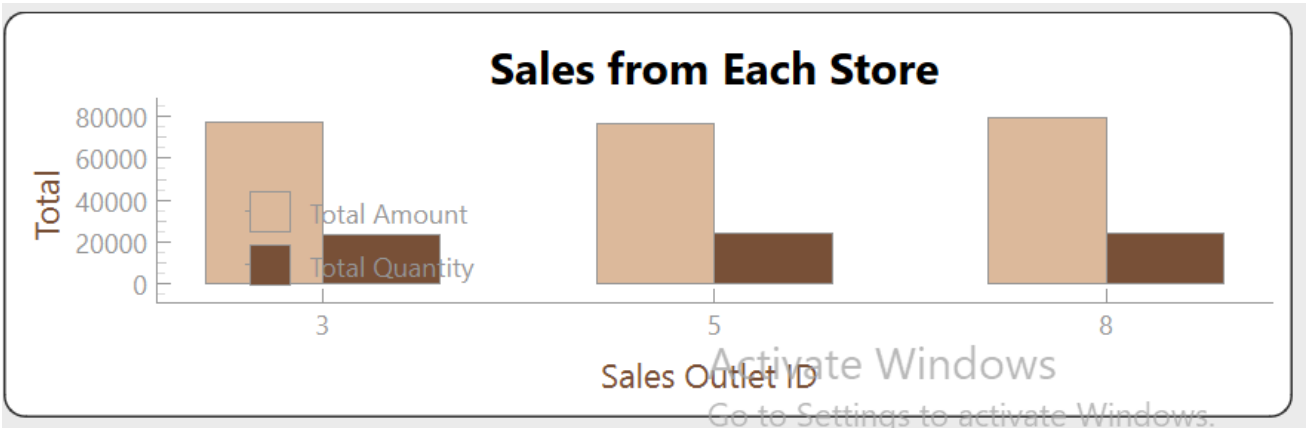


Figure 4.9 Sales from Each Store

The sales chart for each store compares total sales amount and quantity across outlets 3, 5, and 8. Outlet 5 leads in both metrics, while outlets 3 and 8 show similar performance. Including this chart in the sales dashboard is essential for managers to assess store performance, understand customer demand, and optimize inventory and marketing strategies for each location.

4.5.2 Customer Dashboard

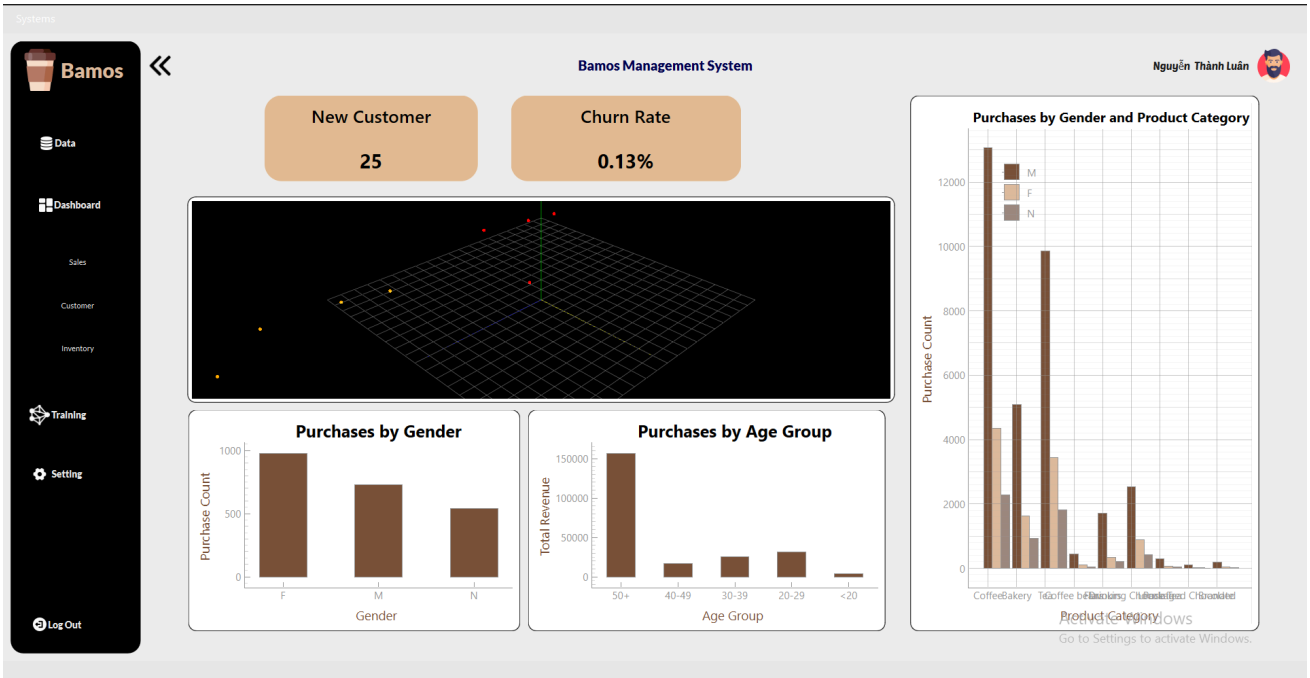


Figure 4.10 Customer Dashboard

The Customer Dashboard focuses on analyzing customer behavior and demographics. It includes visualizations for purchases by gender and age group, churn rate, and purchases by gender and product category. This dashboard leverages k-means clustering for customer segmentation, allowing managers to identify distinct customer groups and tailor marketing efforts accordingly. The customer segmentation insights help in personalizing customer experiences, improving satisfaction, and increasing retention rates. The dashboard also tracks new customer counts, providing a clear picture of customer acquisition trends. A dedicated k-means clustering chart is included to visually represent the different customer segments, helping managers quickly identify and analyze distinct clusters of customer behavior and preferences.

Figure 4.11 Purchases by Gender

The chart shows a roughly equal distribution of purchases between male and female customers, with a notably smaller number from the 'N' category. This suggests the necessity for targeted marketing strategies across different gender demographics to enhance engagement and sales.

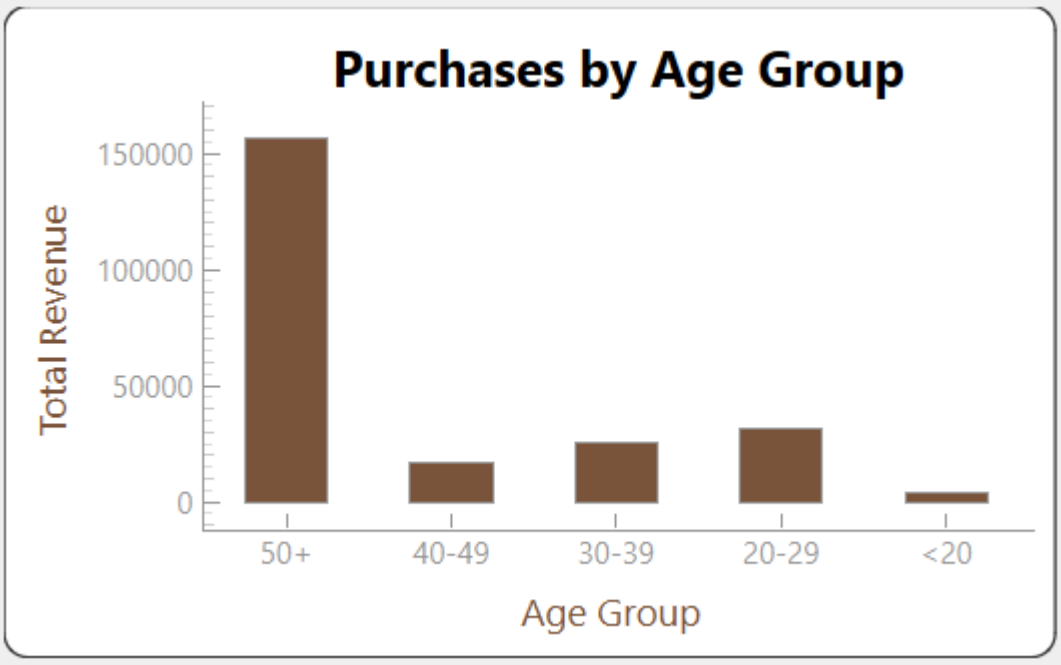


Figure 4.12 Purchases by Age Group

The data reveals that the 50+ age group significantly dominates revenue, indicating their high purchasing power. Targeted strategies to cater to this demographic could include specialized products and marketing campaigns, while also exploring ways to increase engagement with younger consumers.

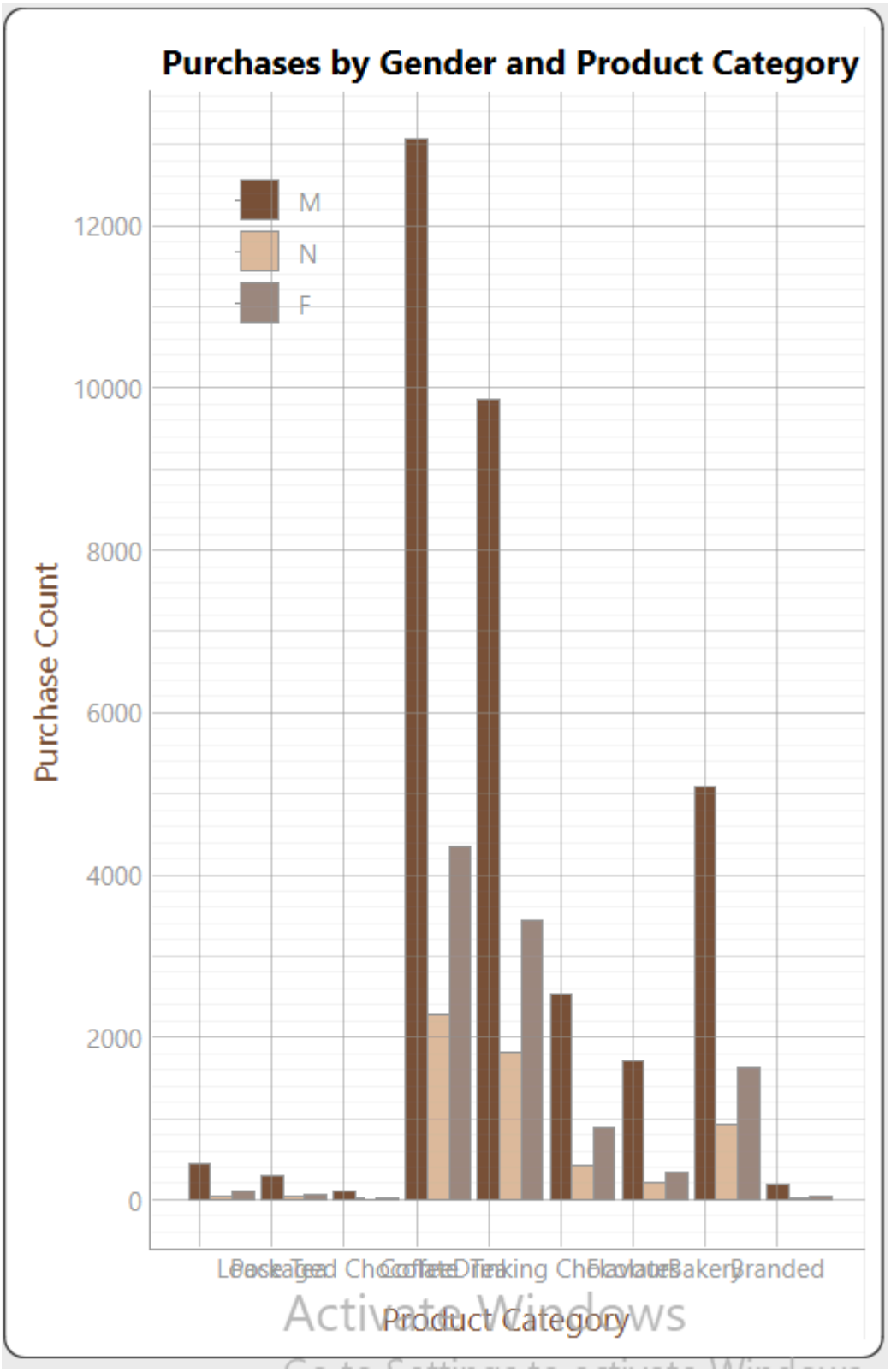


Figure 4.13 Purchase by Gender and by Product Category

This chart identifies purchasing preferences across different product categories by gender, highlighting opportunities to tailor marketing strategies and product offerings. For instance, males show a strong preference for beverages, while females dominate in the bakery category. This insight enables more effective product positioning and targeted promotions to boost sales in each category.

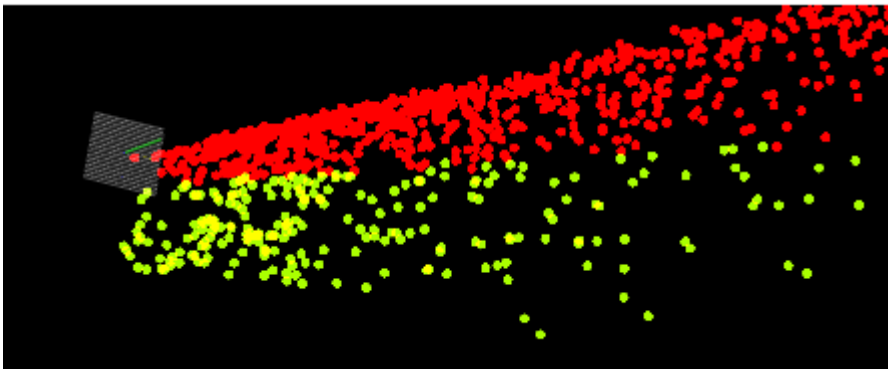


Figure 4.14 KMeans

This visualization from the KMeans clustering analysis shows the segmentation of data into three distinct groups, represented in yellow, red and green colors. This segmentation enables precise targeting and customization of marketing strategies, product development, and customer service enhancements based on the distinct characteristics and behaviors of each group. By understanding these segments, businesses can allocate resources more efficiently and improve customer satisfaction and engagement.

4.5.3 Inventory Dashboard

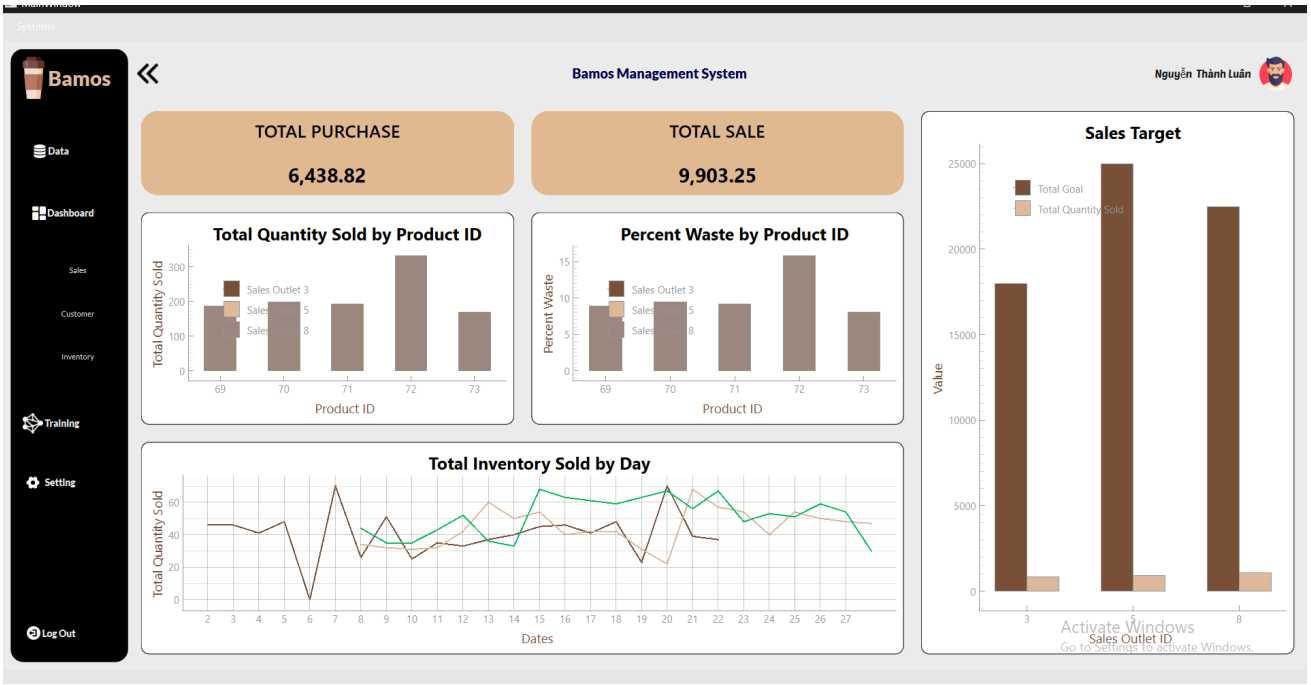


Figure 4.15 Inventory Dashboard

The Inventory Dashboard is designed to manage and monitor the coffee shop's inventory effectively. It includes charts for total quantity sold by product ID, percent waste by product ID, and total inventory sold by day. These visualizations help in understanding product performance, identifying waste patterns, and ensuring optimal inventory levels. The dashboard also features a sales target comparison, showing actual sales against predefined targets. This helps in evaluating inventory turnover and making necessary adjustments to inventory management strategies.

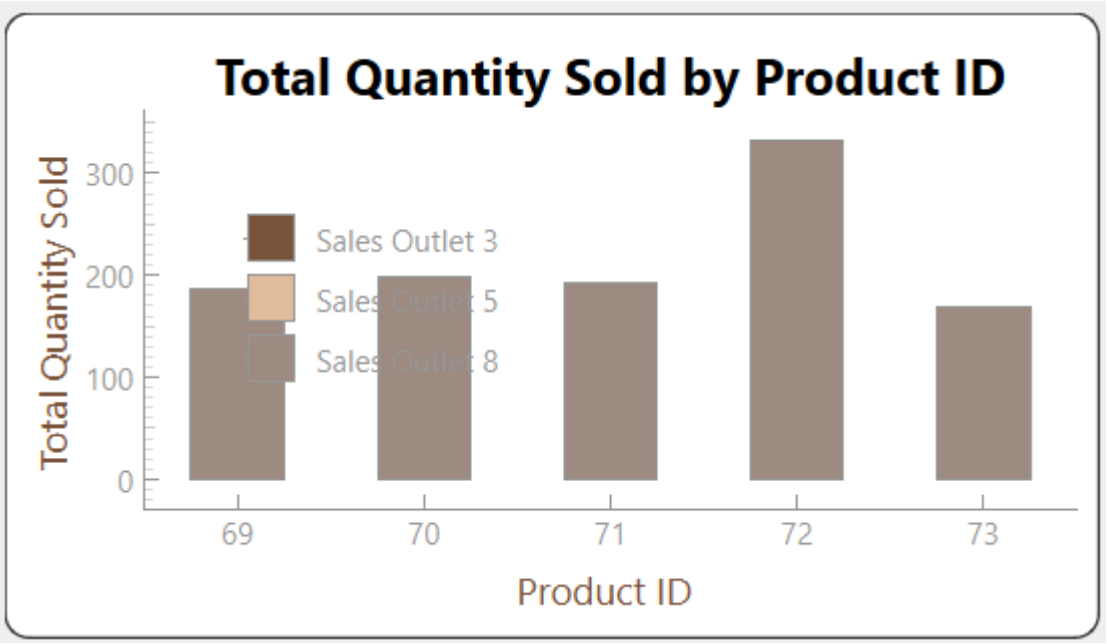


Figure 4.16 Total Quantity Sold by Product

This chart illustrates the total quantity sold for various products across different sales outlets. It highlights that certain products, like Product ID 72, have consistent high sales across all outlets, indicating strong market acceptance and demand. This insight can guide inventory management, ensuring optimal stock levels to prevent shortages and maximize sales.

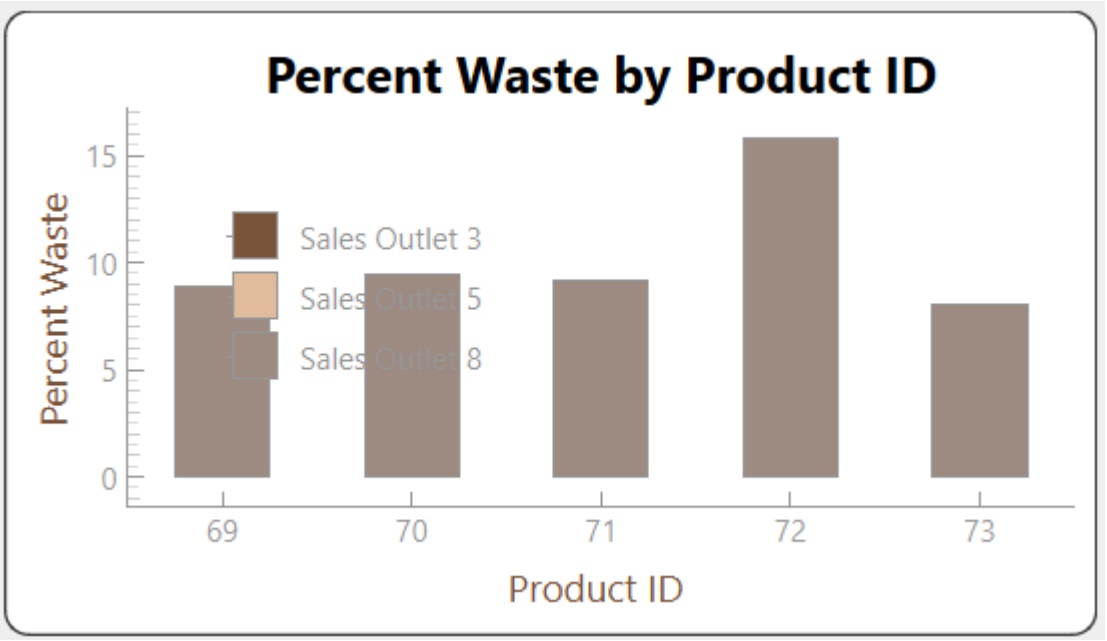


Figure 4.17 Percent Waste by Product ID

The second chart details the percentage of waste by product across the same outlets. It reveals that Product ID 70 experiences significantly higher waste, especially in Sales Outlet 5. This information is crucial for improving supply chain efficiency and reducing losses by addressing the causes of high waste in specific products and outlets, potentially through better inventory control or revising sales strategies.

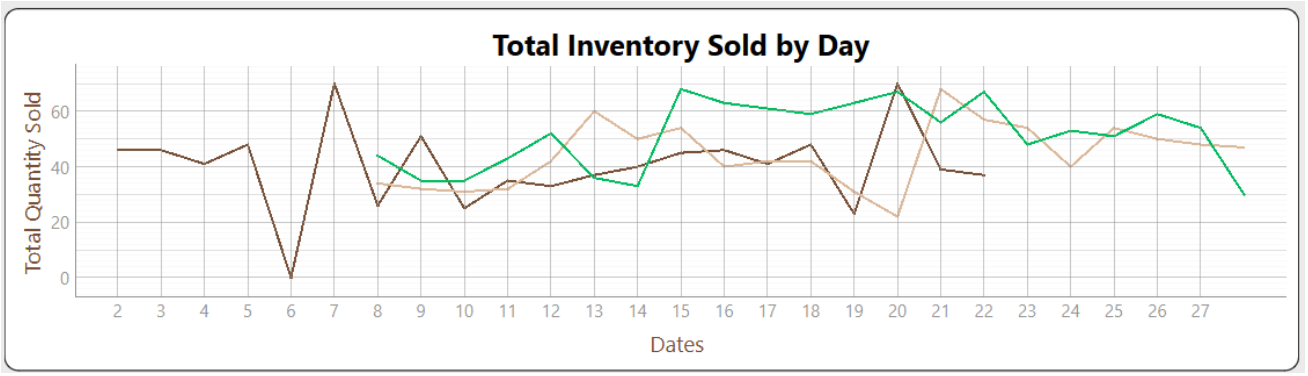


Figure 4.18 Total Inventory Sold by Day

This graph displays daily sales trends across different sales outlets over the month. The data shows variability in sales, with certain days and outlets experiencing peaks that could correlate with promotional events or specific market dynamics. This insight helps in planning better stock management and marketing strategies to leverage days with high sales potential.

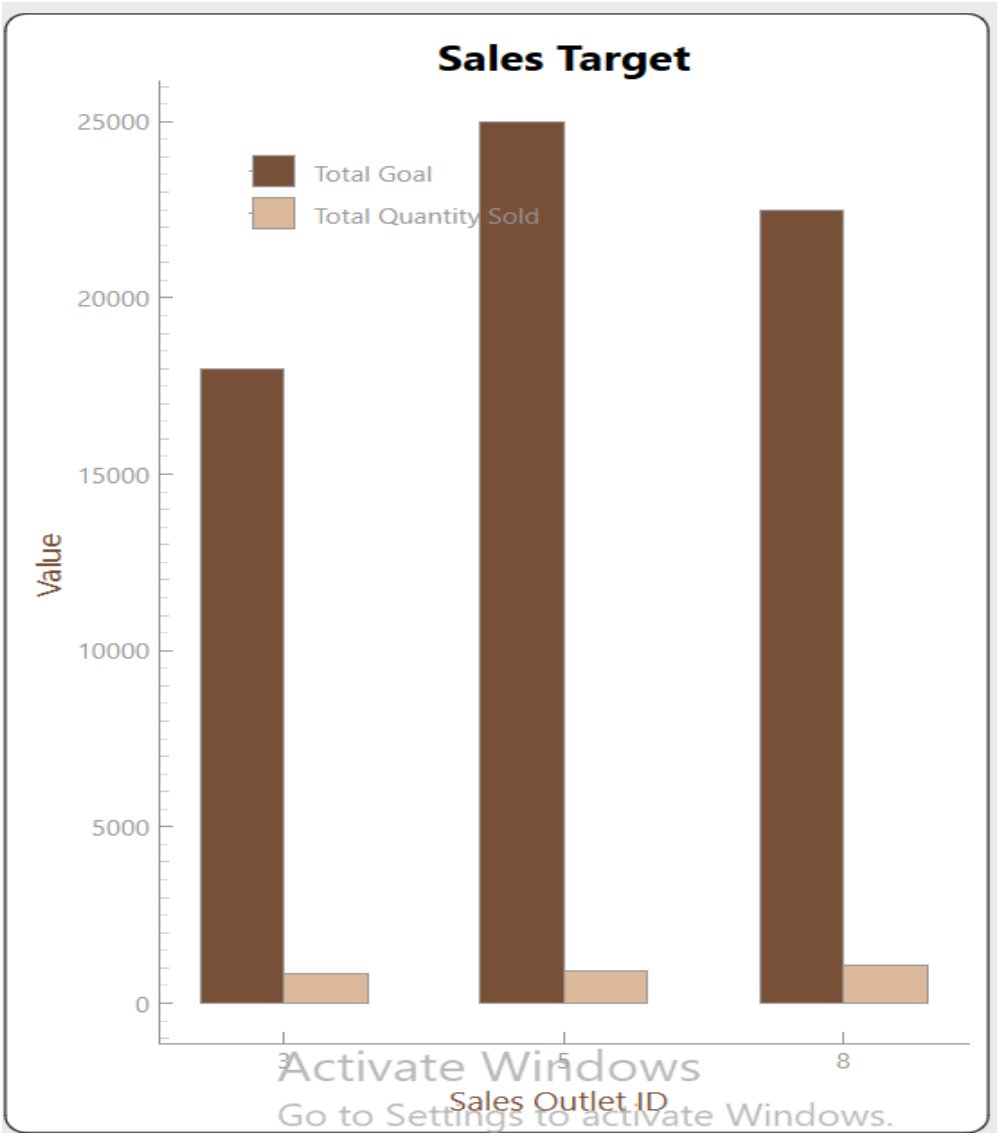


Figure 4.19 Sales Target vs Actual Sales by Sales Outlet ID

The chart compares the actual sales to the sales targets for each outlet, revealing significant discrepancies. Outlets 3 and 8 are particularly underperforming against set targets, while Outlet 5 slightly exceeds its goal. This analysis is crucial for addressing performance issues, possibly adjusting sales targets, or implementing specific strategies to boost sales in underperforming outlets.

4.5.4 Key Features

- Real-Time Data: All dashboards are updated in real-time, ensuring that the displayed data is current and reflective of the latest business activities.
- Interactive Visualizations: Users can interact with the charts and graphs to drill down into specific data points, providing a deeper understanding of the underlying trends and patterns.
- Customizable Views: The dashboards can be customized to display the most relevant metrics and KPIs for different user roles, ensuring that each user has access to the information they need.
- Integrated Machine Learning: The Customer Dashboard's k-means clustering feature integrates machine learning to provide advanced customer segmentation and analysis.

By incorporating these dashboards, the BMS App enables managers to make data-driven decisions that enhance operational efficiency, optimize sales strategies, and improve customer satisfaction. The comprehensive and interactive nature of these visualizations ensures that all aspects of the coffee shop's operations are monitored and managed effectively.

4.6 Machine Learning Feature

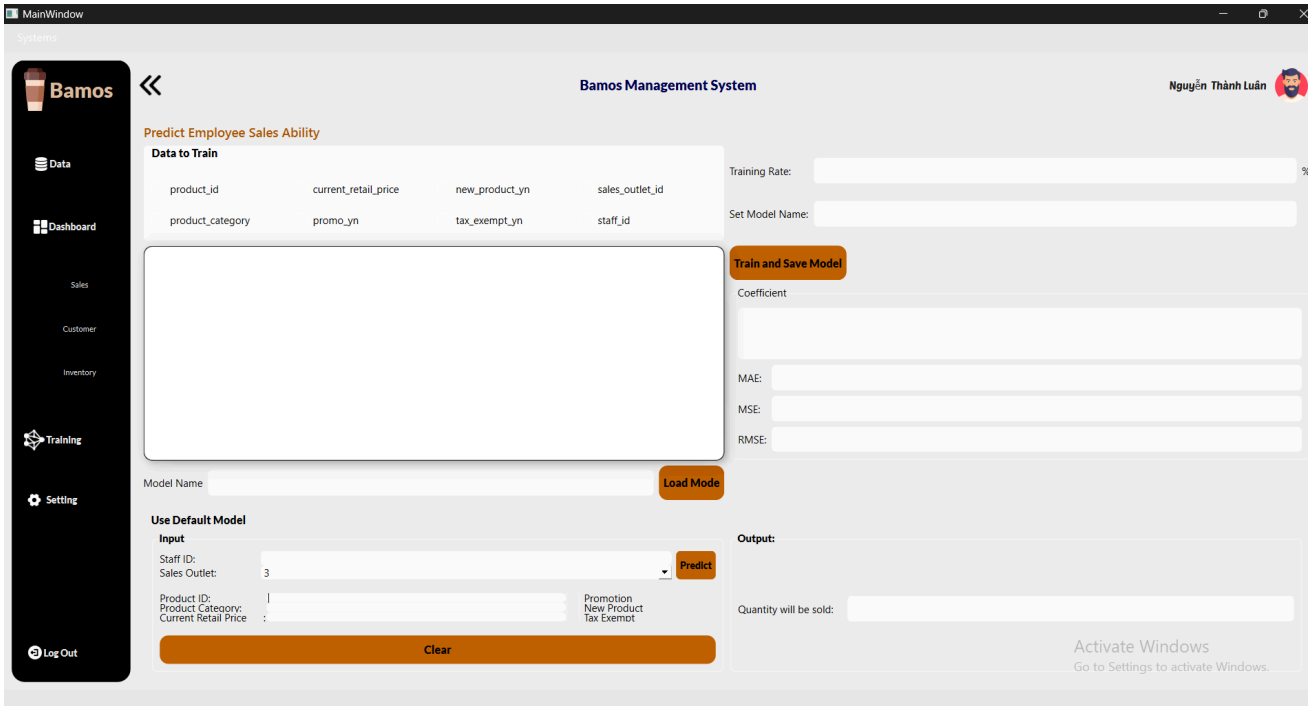


Figure 4.20 Machine Learning Feature

The Machine Learning feature of our coffee shop management application is designed to provide users with advanced predictive capabilities, enhancing decision-making processes and operational efficiency. This feature allows users to create, upload, and use machine learning models tailored to their specific needs.

Users can select input attributes by clicking on the desired attributes displayed on the screen. Then, users need to enter the Training Rate to determine the size of the dataset to train the model. Finally, users will name the model and click the "Train and Save Model" button to start the training and saving process. This ensures that the model is trained with specific data and ready for future predictions.

The next feature is loading a model: users will click the "Load Model" button. A file browsing window will appear, and users can navigate to where the model is stored and select it. The path to the model will be displayed in the lineEdit input field, along with performance metrics such as MAE (Mean Absolute Error), MSE (Mean Squared Error), and R2 (R-squared). Additionally, a chart comparing actual values (actual - y_test) with predicted values (y_pred) will be displayed, helping users better understand the model's performance and prediction accuracy.

The final feature allows users to use a default model pre-trained by us. Users need to fill in all input variables such as staff_id, sales_outlets, product category, and other relevant factors. Once the "Predict" button is clicked, the model will use this input data to predict the number of products that a staff member can sell in a day. This prediction is crucial for sales planning, employee performance evaluation, and inventory management. Accurate predictions help managers allocate resources effectively, identify outstanding employees and areas needing improvement, and plan inventory appropriately to minimize the risk of overstocking or stockouts.

Currently, user-trained models cannot be directly utilized within the application as we have not yet developed this functionality fully. This is an area for improvement in the future, enabling users to fully leverage the machine learning models they have created, thereby further enhancing their predictive capabilities and coffee shop management.

By integrating these machine learning features, our application provides a powerful platform for data-driven decision-making, enhancing the efficiency and effectiveness of coffee shop management operations.

4.7 Setting Feature

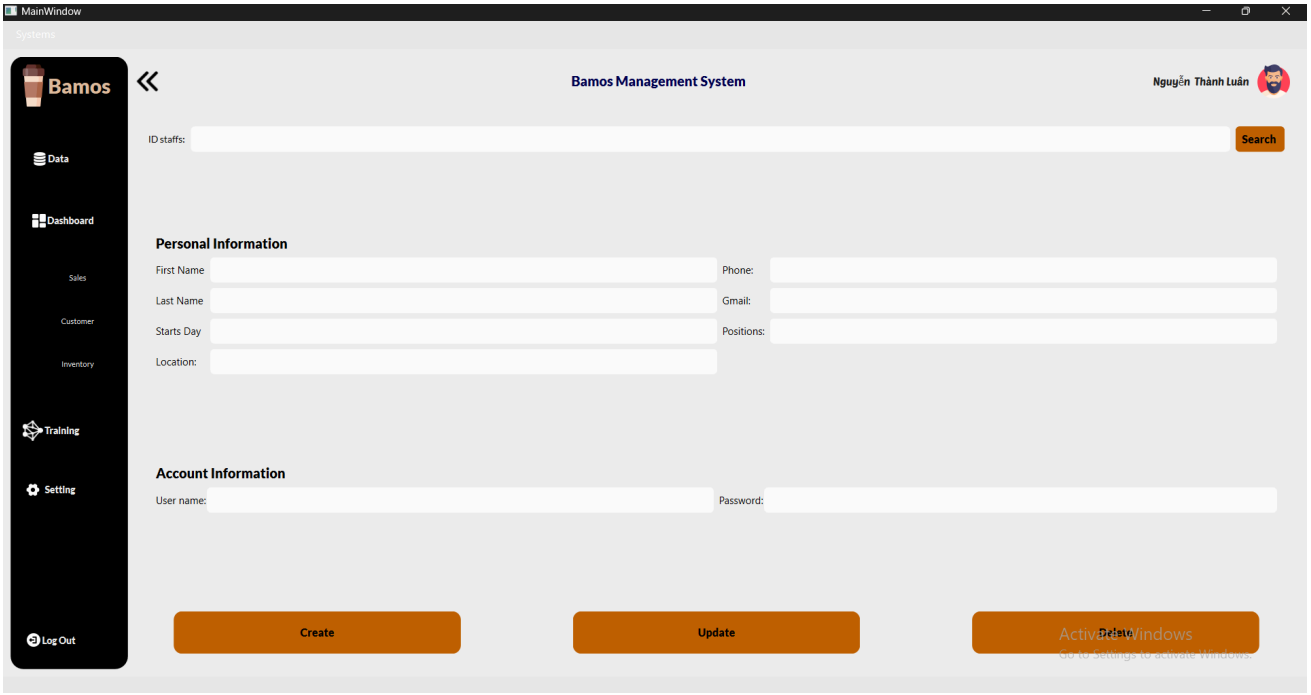


Figure 4.21 Setting Feature

The Setting feature of our coffee shop management application is designed to facilitate the efficient management of staff members. This functionality is crucial for maintaining an organized and up-to-date record of all employees, which is stored in the staffs table within our database. The Setting tab provides a user-friendly interface for performing various staff management tasks.

The Setting tab includes a search function, allowing users to search for staff members by entering their unique ID in the search bar. This functionality enables managers to quickly locate and review the details of a specific employee, ensuring easy access to individual records. Additionally, the tab displays personal information fields including First Name, Last Name, Start Date, Location, Phone, Email, and Position. This information is crucial for identifying and managing staff members, and managers can view and edit these details to keep the staff records accurate and current.

In addition to personal information, the Setting tab also includes fields for managing user account details, such as the Username and Password. This feature ensures that each staff member's account credentials are securely maintained and can be updated as needed. Managers can also create new staff members by filling out the required fields and clicking the Create button. This process ensures that all new hires are promptly recorded and can begin using the system.

Updating staff information is made easy with the Setting tab, as managers can make changes to the displayed fields and click the Update button to reflect these changes in the database. This functionality ensures that any changes in personal or account information are accurately updated. Furthermore, the Setting tab provides an option to delete staff members from the

database. By selecting the desired staff member and clicking the Delete button, managers can remove employees who are no longer part of the team, keeping the staff records current.

The user interface elements of the Setting tab include a search bar and button located at the top of the tab, enabling the search functionality based on staff ID. Personal and account information fields are arranged in a form-like layout, making it easy to view and edit employee details. Action buttons for Create, Update, and Delete are positioned at the bottom of the tab, facilitating these respective actions and ensuring a streamlined and efficient management process.

Overall, the Setting feature enhances the efficiency of staff management within our coffee shop by providing a centralized and accessible platform for managing employee records. This functionality not only improves the accuracy of staff information but also streamlines administrative tasks, allowing managers to focus on other critical aspects of running the coffee shop.

4.8 Results of Machine Learning

Our machine learning models have been trained to predict the daily product sales volume of an employee at a coffee shop, using a diverse set of features: product_id, product_category, current_retail_price, promo_yn, new_product_yn, tax_exempt_yn, sales_outlet_id, and staff_id, with the output variable being total_quantity.

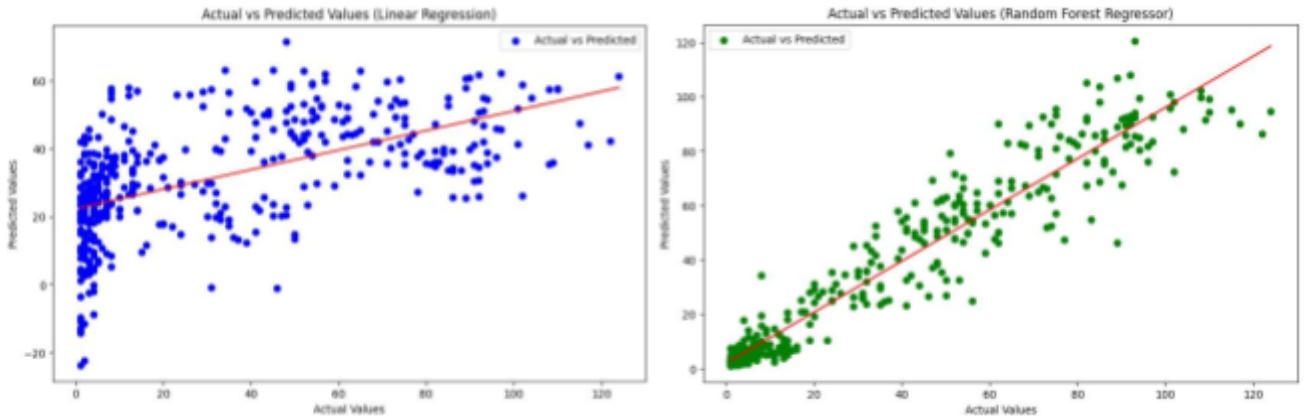


Figure 422 Results of Machine Learning

To assess the performance of these models, we deployed and compared two popular algorithms: Linear Regression and Random Forest Regressor. The results are as follows:

- Linear Regression Model
  - Mean Squared Error (MSE): 746.5459563850056
  - Mean Absolute Error (MAE): 22.231807131112003
  - R-squared (R2): 0.31937799220123075
- Random Forest Regressor Model
  - Mean Squared Error (MSE): 84.22704964539007
  - Mean Absolute Error (MAE): 6.083309692671394

- R-squared (R2): 0.9232106434301712

The superior performance of the Random Forest Regressor is evident through significantly lower MSE and MAE values, along with a high R-squared value. This indicates that the Random Forest model is much more accurate in predicting the total daily product sales volume per employee.

The scatter plot above shows the relationship between the actual values and the predictions from the Random Forest Regressor model. The red line represents the perfect prediction line where the actual values equal the predicted values. The proximity of the data points to this line indicates the high accuracy of the Random Forest model.

Performance Metrics

The following performance metrics provide further insights into the effectiveness of the model:

- Mean Absolute Error (MAE): The average magnitude of the errors in predictions, irrespective of their direction. For the Random Forest model, the MAE is 6.0833, significantly lower than the MAE for the Linear Regression model, which is 22.2318.
- Mean Squared Error (MSE): The average of the squares of the differences between the predicted values and the actual values. A lower MSE indicates a better fit model. The MSE for the Random Forest model is 84.2270, significantly lower than the MSE for the Linear Regression model, which is 746.5459.
- R-squared (R2): The proportion of variance in the dependent variable that can be predicted from the independent variables. The Random Forest model has an R2 of 0.9232, indicating that it can explain 92.32% of the variance in the target variable, compared to just 31.94% for the Linear Regression model.

In conclusion, the Random Forest Regressor model has proven to be exceptionally effective for our dataset. The model's ability to handle a diverse, non-uniform set of input features contributes to its superior performance. This highlights the flexibility and strength of the Random Forest algorithm in managing complex and heterogeneous data.

By leveraging this model, the management of our coffee shop can make highly accurate sales predictions, optimize inventory, and enhance employee performance evaluation, thereby driving operational excellence and customer satisfaction.

In the future, we will continue to refine our models and explore additional machine learning techniques to continuously improve our predictive capabilities and operational efficiency.

CHAPTER 5. CONCLUSION

5. 1 Conclusion

5.1.1 Summary and Research Results

Our project has successfully developed a fully functional café management application, combining a user-friendly and attractive interface. This application provides a range of essential features for efficient management, from tracking sales figures to inventory and staff management. Particularly, the app is designed with a beautiful interface, ensuring that users can navigate and use it easily without extensive training. Emphasizing user experience has ensured that both newcomers and experienced managers can maximize the efficiency of using the application.

5.1.2 Feasibility Assessment

Our application demonstrates high practicality, built on a modern technological platform with strong integration capabilities with other systems. The feasibility of the application relies not only on its current features but also on its potential for future development and expansion. With a flexible database and well-structured data model, the application can easily adapt to the changing requirements of cafés and scale up to serve larger chains or even franchise business models. This creates a solid foundation for continuous growth and innovation, increasing long-term value for investors and stakeholders.

5. 2 Recommendation

5.2.1 Limitations

Our team has developed an application for managing a coffee shop. While the application meets many of our initial goals, there are several limitations that need to be addressed to improve its functionality and user experience.

Role-Based Access Control: One of the major limitations of our application is the absence of a role-based access control (RBAC) system. Currently, the application does not differentiate between various user roles such as employees, managers, the CEO, and other positions within the coffee shop. This lack of granularity in permissions means that all users have the same level of access to the application's features and data, which can lead to security vulnerabilities and operational inefficiencies. For example, sensitive information and critical functions that should be restricted to management are accessible to all users. Implementing an RBAC system would enhance security by ensuring that users can only access the features and data relevant to their roles.

Dependency on MySQL: Another significant limitation is the application's reliance on MySQL being installed on the user's machine. Users are required to have MySQL installed and properly configured, along with the necessary database setup on their local systems. This dependency can pose a challenge for users who are not technically proficient or familiar with

database management. It also limits the portability and ease of deployment of the application. Potential users might face difficulties during installation and setup, which could hinder the adoption of the application. Addressing this limitation could involve exploring options such as bundling the database with the application or utilizing a cloud-based database service.

Machine Learning Model: The current implementation of the machine learning component in our application is limited to a single model. This restriction reduces the flexibility and adaptability of the application in handling various scenarios and data patterns. For instance, the model may not perform optimally under different operational conditions or with diverse datasets. Utilizing multiple machine learning models or an ensemble approach could enhance the accuracy and robustness of the predictions and recommendations provided by the application. Expanding the machine learning capabilities would allow the application to cater to a wider range of use cases and improve its overall performance.

These limitations highlight important areas for future development to enhance the functionality, usability, and security of our coffee shop management application. Addressing these issues will help in creating a more robust and user-friendly application that can better meet the needs of its users. By implementing role-based access control, reducing dependency on specific database setups, and expanding the machine learning capabilities, we can significantly improve the effectiveness and appeal of our application.

5.2.2 Future Work

In our Future Work section, we outline several enhancements to further develop and refine our application, emphasizing security, performance, and user engagement. Firstly, we plan to implement a Role-Based Access Control (RBAC) system to enhance security measures and efficiently manage user permissions according to their organizational roles. Secondly, to improve scalability and alleviate potential bottlenecks, we aim to reduce our dependency on MySQL by exploring alternative databases and data storage solutions. This will enhance performance and reliability, particularly under high load conditions.

Additionally, we intend to expand our machine learning capabilities by incorporating more advanced algorithms and models, including enhancing our existing KMeans clustering and forecasting models like Prophet and ARIMA. This expansion will enable a deeper understanding of customer behaviors and preferences. To ensure the application operates smoothly and efficiently, we will continue to optimize the source code, improve scalability, and conduct regular testing and tuning. Load balancing and resource management will also be prioritized to efficiently handle increased traffic and data volumes.

Finally, to attract and retain users, we plan to upgrade the user interface and introduce new features that enhance user engagement and interaction. These upgrades will include designing more intuitive and responsive layouts and improving overall navigability and usability. By focusing on these strategic enhancements, we aim to create a robust, scalable, and user-friendly application that not only meets current market needs but also adapts to future demands and opportunities.

References

[1]. The Role of Big Data Analytics in Influencing Artificial Intelligence (AI) Adoption for Coffee Shops in Krabi, Thailand, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3944075

[2]. Using BLE beacons and Machine Learning for Personalized Customer Experience in Smart Cafés, https://ieeexplore.ieee.org/abstract/document/9051187

[3] Unveiling Post Engagement Rate Using Predictive Analytics Model for Instagram Accounts of Coffee Shops in West Jakarta, https://jurnal.integrasisainsmedia.co.id/index.php/JIMS/article/view/97

[4] What is machine learning (ML)?, https://www.ibm.com/topics/machine-learning

[5] Machine Learning Algorithms, https://www.geeksforgeeks.org/machine-learning-algorithms/

[6] Types of Machine Learning, https://www.geeksforgeeks.org/types-of-machine-learning/

[7] NumPy là gì? Nó hoạt động như thế nào?, https://thegioimaychu.vn/blog/ai-hpc/numpy-la-gi-no-hoat-dong-nhu-the-nao-p17778/

[8] Xử lý dữ liệu với Pandas trong Python, https://codelearn.io/sharing/xu-ly-du-lieu-voi-pandas-trong-python

[9] Scikit-learn một thư viện quan trọng cho ML trên ngôn ngữ lập trình Python, https://thegioimaychu.vn/blog/ai-hpc/scikit-learn-mot-thu-vien-quan-trong-cho-ngon-ngu-lap-trinh-python-p18186/

[10] Bài 1: Giới thiệu Matplotlib - Matplotib Cơ Bản, https://vncoder.vn/bai-hoc/gioi-thieu-matplotlib-492

[11] PYTHON SEABORN - NẮM RÕ CÁCH SỬ DỤNG THƯ VIỆN SEABORN CỦA PYTHON, https://t3h.com.vn/tin-tuc/python-seaborn

[12] What is pyqtgraph?, https://pyqtgraph.readthedocs.io/en/latest/getting_started/introduction.html

[13] Tất tần tật về SQL - SQL là gì? Sử dụng như thế nào?, https://intech.vietnamworks.com/article/tat-tan-tat-ve-sql-sql-la-gi-su-dung-nhu-the-nao

[14] MySQL Workbench Enhanced Data Migration, https://www.mysql.com/products/workbench/

[15] Business Analytics (BA) Là Gì: Các Nguyên Tắc Cơ Bản, https://www.vietdemy.vn/blog/business-analytics-ba-la-gi-cac-nguyen-tac-co-ban/

[16] Business Analytics là gì? Cách phân biệt Business Analytics với Data Analytics và Business Intelligence, https://www.brandcamp.asia/blog/213-Business-Analytics-la-gi-Cach-phan-biet-Business-Analytics-voi-Data-Analytics-va-Business-Intelligence

[17] BUSINESS ANALYSIS VÀ BUSINESS ANALYTICS, https://www.bacs.vn/vi/blog/kien-thuc/business-analysis-va-business-analytics-4918.html