 **ASSIGNMENT 2**

|  |  |  |  |
| --- | --- | --- | --- |
| **Qualification** | **BTEC Level 5 HND Diploma in Computing** | | |
| **Unit number and title** | **Unit 17: Business Process Support** | | |
| **Submission date** | 10/4/2024 | **Date Received 1st submission** |  |
| **Re-submission Date** | 15/4/2024 | **Date Received 2nd submission** |  |
| **Student Name** | Tran Cong Hoang | **Student ID** | BH00317 |
| **Class** | IT0601 | **Assessor name** | Dinh Van Dong |
| **Student declaration**  I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice. | | | |
|  |  | **Student’s signature** | Conghoang |

**Grading grid**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **P5** | **P6** | **P7** | **M3** | **M4** | **D2** | **D1** |
|  |  |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| **❒ Summative Feedback: ❒ Resubmission Feedback:** | | |
| **Grade:** | **Assessor Signature:** | **Date:** |
| **Internal Verifier’s Comments:** | | |
| **Signature & Date:** | | |

|  |
| --- |
| Table of Contents  [I. Introduction 1](#_Toc164185071)  [II. Contents 2](#_Toc164185072)  [P5: Discuss how tools and technologies associated with data science are used to support business processes and inform decisions. 2](#_Toc164185073)  [1) Introduction 2](#_Toc164185074)  [2) Data Collection and Integration 3](#_Toc164185075)  [3) Data Exploration, Visualization, and Predictive Analytics 4](#_Toc164185076)  [4) Process Optimization and Customer Segmentation 6](#_Toc164185077)  [5) Fraud Detection, Risk Management, and Real-time Decision Support 7](#_Toc164185078)  [6) Big Data and Cloud Technologies 8](#_Toc164185079)  [7) Implementation Challenges and Considerations 9](#_Toc164185080)  [8) Case Studies and Success Stories 10](#_Toc164185081)  [9) Future Directions and Conclusion 11](#_Toc164185082)  [P6: Design a data science solution to support decision-making related to a real-world problem. 13](#_Toc164185083)  [1) Problems Faced by ABC Manufacturing 13](#_Toc164185084)  [2) Proposed Data Science Solution 17](#_Toc164185085)  [3) High-Level Architecture Overview 18](#_Toc164185086)  [4) Recommendations 20](#_Toc164185087)  [5) Conclusion 20](#_Toc164185088)  [P7: Implement a data science solution to support decision-making related to a real-world problem. 21](#_Toc164185089)  [1) Set up the environment 21](#_Toc164185090)  [2) Data Collection 23](#_Toc164185091)  [3) Data Cleaning and Preprocessing 24](#_Toc164185092)  [4) Github 38](#_Toc164185093)  [III. Conclusion 39](#_Toc164185094)  [IV. References 40](#_Toc164185095) |

|  |
| --- |
| List of Figures  [Figure 1: Data science process architecture 19](#_Toc164184352)  [Figure 2: Download Visual Studio Code 21](#_Toc164184353)  [Figure 3: Install Python 22](#_Toc164184354)  [Figure 4: Install Jupyter 22](#_Toc164184355)  [Figure 5:Install library matplotlib 23](#_Toc164184356)  [Figure 6: Install library pandas 23](#_Toc164184357)  [Figure 7: Raw data 23](#_Toc164184358)  [Figure 8: Import library 25](#_Toc164184359)  [Figure 9: Data cleaning 25](#_Toc164184360)  [Figure 10: Data cleaning 27](#_Toc164184361)  [Figure 11: Data clean 30](#_Toc164184362)  [Figure 12: Process and analyze data 30](#_Toc164184363)  [Figure 13: Desired data 31](#_Toc164184364)  [Figure 14: Process and analyze data 31](#_Toc164184365)  [Figure 15: Desired data 32](#_Toc164184366)  [Figure 16: : Process and analyze data 32](#_Toc164184367)  [Figure 17: Desired Data 33](#_Toc164184368)  [Figure 18: Desired Data 33](#_Toc164184369)  [Figure 19: Process and analyze data 34](#_Toc164184370)  [Figure 20: Process and analyze data 35](#_Toc164184371)  [Figure 21: Desired data 35](#_Toc164184372)  [Figure 22: Desired data 36](#_Toc164184373)  [Figure 23: Process and analyze data 37](#_Toc164184374)  [Figure 24: Desired data 38](#_Toc164184375) |

|  |
| --- |
| List of Table  [Table 1: Role of data science tools 3](#_Toc164184376) |

# Introduction

In today's dynamic business landscape, organizations face multifaceted challenges in managing their supply chains efficiently. The advent of Industry 4.0, characterized by automation, data exchange, and smart manufacturing technologies, has transformed traditional supply chain management. As a Junior Analyst at ABC Manufacturing, your role involves analyzing how data science can enhance the company's operations and support decision-making. Let's delve into the key aspects of data-driven supply chain management and explore how it can address real-world problems faced by the Operation Director.

Supply chain management encompasses the end-to-end flow of goods, services, and information from origin to customers. Traditionally, Manufacturing Execution Systems (MES) collected data for periodic reports, providing insights into past events. However, the shift toward real-time data analytics has revolutionized supply chain practices. Real-time data enables up-to-date reporting, proactive decision-making, and optimization of production processes.

Demand forecasting plays a pivotal role in addressing uncertainties within supply chains. Accurate predictions are essential for production scheduling, inventory planning, and risk mitigation. Leveraging data science techniques, such as predictive analytics and machine learning, can significantly enhance demand forecasting accuracy. By analyzing historical sales data, market trends, and customer preferences, organizations can anticipate future demand patterns and optimize resource utilization.

As we explore data science applications, we'll focus on designing solutions that support decision-making. Our approach involves integrating real-time data from sensors, IoT devices, and production facilities. By monitoring equipment performance, identifying bottlenecks, and streamlining logistics operations, ABC Manufacturing can achieve cost savings, improve product quality, and enhance overall supply chain performance.

While data science offers immense potential, challenges persist. Integrating multiple autonomous processes, managing real-time data, and ensuring cross-functional involvement are critical. Additionally, we'll explore how data science can address closed-loop supply chains (CLSCs) and propose a visionary functional software architecture.

In this report, we'll delve into practical use cases, share experiences, and evaluate the impact of data science techniques on ABC Manufacturing's supply chain. Let's unlock the power of data-driven decision-making and drive operational excellence in the ever-evolving landscape of supply chain management.

# Contents

## P5: Discuss how tools and technologies associated with data science are used to support business processes and inform decisions.

### Introduction

#### Overview of data science in business processes and decision-making

In today's data-driven world, data science plays a critical role in optimizing business processes and informing decision-making. Data science involves the extraction, analysis, and interpretation of large volumes of data to extract valuable insights and drive strategic actions. By harnessing the power of data, organizations can gain a competitive edge, improve operational efficiency, and make informed decisions based on evidence rather than intuition.

Data science encompasses a wide range of techniques and methodologies, including statistical analysis, machine learning, data visualization, and predictive modeling. These approaches enable businesses to uncover patterns, trends, and correlations hidden within their data, providing valuable insights for driving growth, identifying opportunities, and mitigating risks.

#### Importance of tools and technologies in supporting data science

The success of data science heavily relies on the availability of advanced tools and technologies that assist in processing and analyzing large volumes of data efficiently. These tools and technologies enable data scientists and analysts to extract meaningful information from complex datasets, explore data relationships, and generate actionable insights.

Data science tools provide functionalities such as data preprocessing, exploratory data analysis, machine learning algorithm implementation, and data visualization. These tools simplify and automate various stages of the data science workflow, allowing practitioners to focus on the interpretation and application of the results.

Additionally, technologies like cloud computing, distributed computing frameworks, and scalable storage solutions have revolutionized data science by enabling the processing and analysis of massive datasets in a cost-effective manner. These technologies provide the infrastructure and computing power necessary to handle the computational demands of data-intensive tasks, allowing organizations to leverage big data for decision-making. In conclusion, data science tools and technologies are instrumental in supporting the data-driven decision-making process. By leveraging these tools, businesses can derive valuable insights from their data, optimize operations, and gain a competitive advantage in today's data-centric landscape.

### Data Collection and Integration

#### Role of Data Science Tools

Data science tools play a pivotal role in collecting, integrating, and consolidating data from diverse sources within ABC Manufacturing. These sources include transactional systems (e.g., ERP systems), customer interactions (e.g., CRM data), social media (e.g., sentiment analysis), and IoT devices (e.g., sensors on production lines). By using tools like ETL pipelines, data connectors, and APIs, ABC Manufacturing can efficiently gather data and create a unified view for analysis. See the following table for better understanding role of data science tools

|  |  |  |  |
| --- | --- | --- | --- |
| **System** | **Role** | **Data Collection** | **Integration** |
| ERP Systems | Manage critical business processes (e.g., finance, inventory) | Extract sales transactions, inventory levels, order processing data | Combine ERP data with other sources (e.g., CRM, IoT) for a holistic view |
| CRM Data | Track customer interactions | Capture customer behavior, preferences, feedback | Personalize marketing efforts and enhance satisfaction |
| Social Media (Sentiment Analysis) | Provide insights into brand perception, trends, and sentiment | Analyze social media posts, reviews, comments | Integrate sentiment analysis results with internal data |
| IoT Devices (Sensors) | Collect real-time data from machinery, equipment, production lines | Process sensor data (temperature, pressure, energy consumption) | Combine IoT data with other operational data for optimization |
| ETL Pipelines | Extract, transform, and load data from various sources | Extract data from files, databases, APIs | Transform and clean data for consistency; load into a central repository |
| Data Connectors and APIs | Connect to external systems (e.g., third-party vendors, partners) | Retrieve relevant data for integration | Ensure seamless data flow across systems |

Table 1: Role of data science tools

#### Benefits of Data Collection and Integration

A unified data view enables better decision-making by providing a holistic understanding of the organization's operations. Integrated data allows for cross-functional insights, such as correlating sales data with production efficiency or customer feedback with inventory levels. Benefits of Data Collection and Integration are:

* Holistic Decision-Making:
* A unified view of data enables decision-making (including the Operation Director) to understand the entire organization’s landscape.
* Cross-functional insights allow for informed decisions that consider multiple aspects (e.g., sales, production, customer satisfaction).
* Efficiency and Accuracy:
* Integrated data reduces manual data entry and minimizes errors.
* Automation streamlines processes, saving time and resources.
* Real-Time Insights:
* Real-time integration provides up-to-date information for timely decision-making.
* For example, monitoring production line sensors in real time can prevent equipment failures.
* Strategic Planning and Forecasting:
* Integrated historical data supports trend analysis, demand forecasting, and capacity planning.
* ABC Manufacturing can allocate resources effectively based on data-driven insights.

In summary, data collection and integration are foundational steps for leveraging data science effectively. Help we can harness these tools to optimize operations, enhance customer experiences, and drive competitive advantage.

### Data Exploration, Visualization, and Predictive Analytics

In ABC Manufacturing, data science tools play a crucial role in performing exploratory data analysis, data visualization, and predictive analytics. These tools enable the company to gain valuable insights from their data and make informed decisions. Here's a detailed explanation of how these tools are used:

#### Exploratory Data Analysis

**Python Libraries**: ABC Manufacturing utilizes popular Python libraries such as Pandas, Matplotlib, and Seaborn for exploratory data analysis. Pandas provides powerful data manipulation and analysis capabilities, allowing analysts to clean, transform, and preprocess data efficiently. Matplotlib and Seaborn enable the creation of various visualizations, including scatter plots, histograms, and box plots, to understand the distribution, relationships, and anomalies within the data.

**Pandas**: ABC Manufacturing can use Pandas for data manipulation, cleaning, and basic statistical analysis.

**Matplotlib and Seaborn**: These libraries allow for creating various types of plots (e.g., scatter plots, histograms, box plots) to visualize data distributions and relationships.

**Benefits**: By performing exploratory data analysis, ABC Manufacturing can identify patterns, correlations, and outliers in their data. This helps uncover hidden insights and discover factors that may impact their business operations and performance.

#### Data Visualization

**Tableau and Power BI**: ABC Manufacturing employs data visualization tools like Tableau and Power BI to create interactive and visually appealing dashboards and reports. These tools allow users to combine multiple data sources, create dynamic visualizations, and share insights across the organization.

**Benefits**: Data visualization enhances understanding and communication of complex information. ABC Manufacturing can present data in a clear and intuitive manner, enabling stakeholders to grasp trends, outliers, and patterns quickly. Visualizations facilitate data-driven decision-making by providing a comprehensive view of key performance indicators and metrics.

#### Predictive Analytics

**Machine Learning and Statistical Algorithms**: ABC Manufacturing leverages machine learning algorithms and statistical models to perform predictive analytics. Popular libraries like Scikit-learn, TensorFlow, and PyTorch provide a wide range of algorithms for tasks such as regression, classification, clustering, and recommendation.

**Benefits**: Predictive analytics enables ABC Manufacturing to forecast future outcomes, identify trends, and make data-driven predictions. By building predictive models, the company can anticipate demand, optimize inventory, improve forecasting accuracy, and make informed decisions regarding pricing, production, and resource allocation.

Overall, data exploration, visualization, and predictive analytics empower ABC Manufacturing to identify patterns, trends, and anomalies in their data. These insights help the company gain a competitive edge, optimize operations, enhance customer experiences, and drive strategic decision-making. By leveraging data science tools and techniques, ABC Manufacturing can unlock the full potential of their data and capitalize on the power of data-driven decision-making failures.

### Process Optimization and Customer Segmentation

#### Process Mining

* Role of Process Mining

Process mining is a data-driven technique that extracts insights from event logs to improve business processes.

It involves analyzing historical data (e.g., transaction records, production logs) to understand how processes are executed.

* Identifying Bottlenecks and Inefficiencies

By examining event sequences, process mining reveals bottlenecks, delays, and deviations.

ABC Manufacturing can pinpoint areas where efficiency can be enhanced.

* Improving Process Efficiency

ABC Manufacturing can streamline workflows, reduce cycle times, and optimize resource allocation.

For example, identifying delays in order processing or inventory management can lead to faster response times.

#### Customer Segmentation

* Role of Customer Segmentation

Customer segmentation divides the customer base into distinct groups based on shared characteristics.

It allows ABC Manufacturing to tailor strategies for each segment.

* Clustering Algorithms (e.g., k-means)

ABC Manufacturing can group customers based on behavior, demographics, purchase history, or preferences.

Segments may include loyal customers, occasional buyers, high spenders, etc.

* Benefits of Customer Segmentation

**Targeted Marketing**: ABC Manufacturing can create personalized campaigns for each segment.

**Product Customization**: Tailor products or services to specific customer needs.

**Pricing Strategies**: Set pricing based on segment preferences and willingness to pay.

In summary, process optimization and customer segmentation empower ABC Manufacturing to enhance operational efficiency and deliver personalized experiences.

### Fraud Detection, Risk Management, and Real-time Decision Support

#### Fraud Detection

Fraud detection is a critical area where data science tools play a vital role in ABC Manufacturing. By applying machine learning models, such as anomaly detection algorithms and random forests, the company can identify unusual patterns and behaviors that are indicative of fraudulent activities. These models analyze various data sources, including transactional records, customer interactions, and historical fraud cases, to detect anomalies and flag potentially fraudulent transactions or activities. By detecting fraud early on, ABC Manufacturing can prevent financial losses, protect its assets, and maintain the trust of its customers and partners.

#### Risk Management

Data science tools are also instrumental in risk management for ABC Manufacturing. By leveraging historical data and statistical analysis, these tools assess and quantify risks associated with various aspects of the business, such as supply chain disruptions, financial risks, and compliance issues. Risk models are developed to analyze past incidents, identify trends and patterns, and predict potential risks and their potential impact. By understanding and quantifying risks, ABC Manufacturing can make informed decisions regarding risk mitigation strategies, resource allocation, and contingency planning, ultimately minimizing the potential negative impact on the business.

#### Real-time Analytics and Dashboards

Real-time analytics and dashboards provide ABC Manufacturing with timely insights and support real-time decision-making. By leveraging data science tools and techniques, the company can process and analyze data in real-time, enabling them to monitor key performance indicators (KPIs), production metrics, and inventory levels on the fly. Real-time dashboards provide a visual representation of these metrics, allowing the Operations Director and other stakeholders to monitor the current state of operations, identify potential issues or bottlenecks, and make data-driven decisions promptly. This real-time decision support empowers ABC Manufacturing to respond quickly to changing market conditions, optimize production processes, and ensure efficient resource utilization.

By applying data science tools in fraud detection, risk management, and real-time decision support, ABC Manufacturing can safeguard its business operations, mitigate potential risks, and make informed decisions based on timely insights. These capabilities not only protect the company's assets but also contribute to its overall operational efficiency, customer satisfaction, and long-term success.

### Big Data and Cloud Technologies

#### Big Data Handling

Big data refers to large and complex datasets that exceed the capabilities of traditional data processing tools. ABC Manufacturing can leverage big data to gain insights, optimize processes, and improve decision-making. To handle big data efficiently, ABC Manufacturing utilizes the following technologies:

* **Hadoop**: Hadoop is an open-source framework for distributed storage and processing of large datasets. It utilizes the Hadoop Distributed File System (HDFS) for storing data across multiple nodes and the MapReduce programming model for parallel computation. ABC Manufacturing can leverage Hadoop to store and process massive amounts of data, enabling scalable and efficient data analysis.
* **Spark**: Spark is a fast and flexible big data processing engine. It provides APIs for batch processing, real-time streaming, machine learning, and graph processing. ABC Manufacturing can leverage Spark's in-memory processing capabilities to perform complex analytics on large datasets, enabling faster data processing and iterative analysis.

#### Cloud Solutions

Cloud platforms offer on-demand access to computing resources over the internet, providing ABC Manufacturing with cost-effective and scalable solutions for data storage, processing, and analytics. ABC Manufacturing utilizes cloud services from providers such as AWS, Google Cloud, or Azure for their data-related needs. Key benefits of cloud solutions include:

* **Cost-Effectiveness**: Cloud platforms allow ABC Manufacturing to pay only for the resources they use, avoiding upfront infrastructure costs. This cost-effective model is particularly beneficial when dealing with large-scale data processing and storage requirements.
* **Scalability**: Cloud platforms offer the ability to scale computing resources up or down based on demand. ABC Manufacturing can easily handle fluctuations in data processing needs, ensuring optimal performance without investing in and managing physical infrastructure.
* **Global Accessibility**: Cloud solutions enable ABC Manufacturing to access data and applications from anywhere with an internet connection. This accessibility promotes collaboration and facilitates real-time decision-making, even across geographically dispersed teams.

By embracing big data and cloud technologies, ABC Manufacturing can efficiently handle large datasets, make data-driven decisions, and enhance overall supply chain performance. The scalability, parallel processing capabilities, cost-effectiveness, and global accessibility provided by these technologies enable ABC Manufacturing to leverage data as a valuable asset and drive business success.

### Implementation Challenges and Considerations

#### Data Quality

Ensuring data accuracy, consistency, and completeness is crucial for ABC Manufacturing's data science initiatives. They should implement the following measures:

* **Data Validation**: Validate data at the source to prevent errors and inconsistencies. Implement automated checks to identify anomalies or discrepancies in the data.
* **Data Cleansing**: Utilize data cleansing techniques such as removing duplicates, handling missing values, and standardizing formats to maintain high-quality data.
* **Monitoring**: Regularly monitor data quality to promptly identify and address any issues that may arise.

#### Resource Allocation

Proper allocation of resources is essential for the success of data science initiatives. ABC Manufacturing should consider the following:

* **Skilled Personnel**: Allocate skilled data scientists, analysts, and domain experts who can effectively analyze and interpret the data. Provide training and upskilling opportunities to existing staff to enhance their capabilities.
* **Computing Resources**: Allocate sufficient computational resources, whether through cloud services or on-premises infrastructure, to handle the processing and analysis requirements of big data.
* **Cost-Benefit Analysis**: Conduct a thorough cost-benefit analysis to ensure that the allocated resources align with the expected benefits and the strategic goals of ABC Manufacturing.

#### Data Privacy and Security

Protecting sensitive information and ensuring compliance with regulations are crucial considerations. ABC Manufacturing should take the following steps:

* **Protecting Sensitive Information**: Implement access controls, encryption, and secure protocols to safeguard proprietary data, customer records, and financial information.
* **Compliance with Regulations**: Ensure compliance with relevant regulations, such as the General Data Protection Regulation (GDPR), when handling personal data. Implement practices like data anonymization, consent management, and privacy impact assessments as required.
* **Ethical Considerations**: Address ethical concerns related to data usage, such as bias and fairness. Transparently communicate data practices to stakeholders and ensure responsible and ethical handling of data.

By addressing these implementation challenges and considerations, ABC Manufacturing can ensure the successful implementation of data science solutions while maintaining data integrity, privacy, and ethical practices. This will enable them to derive valuable insights and make informed decisions to drive business growth and success.

### Case Studies and Success Stories

#### Example 1: Inventory Optimization with Predictive Demand Forecasting

**Challenge**: ABC Manufacturing faced excess inventory, leading to storage costs, obsolescence, and tied-up capital. Traditional inventory management methods were inefficient and reactive.

**Solution**: ABC Manufacturing implemented predictive demand forecasting models. Leveraging historical sales data, seasonality patterns, and external factors (e.g., promotions, market trends), the models predicted future demand.

**Results**:

* 20% Reduction in Excess Inventory: By accurately anticipating demand fluctuations, ABC Manufacturing reduced excess stock.
* Cost Savings: Lower storage costs and minimized waste due to inventory obsolescence.
* Improved Customer Satisfaction: ABC Manufacturing consistently met customer demands, enhancing satisfaction.

#### Example 2: Real-Time Equipment Monitoring for Operational Efficiency

**Challenge**: Frequent equipment breakdowns disrupted production schedules. Downtime led to lost productivity, maintenance costs, and missed delivery deadlines.

**Solution**: ABC Manufacturing implemented real-time monitoring using IoT sensors on production lines. Data on equipment performance, temperature, vibrations, and other parameters were collected continuously.

**Results**:

* Prevented Equipment Failures: Real-time alerts allowed proactive maintenance, preventing breakdowns.
* Reduced Downtime: ABC Manufacturing minimized unplanned downtime, improving overall operational efficiency.
* Cost Savings: Lower maintenance costs and increased production uptime.

In summary, these case studies demonstrate how data-driven approaches can lead to tangible benefits for ABC Manufacturing. By implementing predictive demand forecasting and real-time equipment monitoring, they achieved inventory optimization, improved operational efficiency, cost savings, and enhanced customer satisfaction. These successes showcase the power of data science and analytics in addressing specific challenges and driving positive outcomes for businesses.

### Future Directions and Conclusion

#### Emerging Trends

ABC Manufacturing should consider the following emerging trends to further enhance their data-driven operations:

* **AI-Driven Decision Support Systems**: Exploring AI-powered tools and algorithms can assist decision-making processes. Predictive maintenance algorithms and automated demand forecasting systems are examples that can optimize processes and provide valuable insights.
* **Edge Computing**: Implementing edge computing solutions can bring data processing closer to the source, enabling real-time analytics, reduced latency, and improved efficiency. For instance, utilizing edge devices on production lines to detect anomalies instantly can enhance operational effectiveness.
* **Blockchain for Supply Chain Transparency**: Integrating blockchain technology in the supply chain can ensure transparency, traceability, and security. ABC Manufacturing can leverage blockchain to track raw materials, verify authenticity, and establish trust with partners, enhancing overall supply chain management.

#### Conclusion

ABC Manufacturing should recognize data science as a strategic asset, going beyond its technical aspects. It is a driver of efficiency, informed decision-making, and competitive advantage. To thrive in the ever-evolving business landscape, the company should embrace continuous adaptation.

Regular evaluation of data science strategies, adoption of new technologies, and maintaining an agile mindset are crucial. By staying ahead of emerging trends and proactively embracing new opportunities, ABC Manufacturing can position itself for success in the dynamic world of data-driven operations.

By combining the exploration of emerging trends with a strategic approach to data science, ABC Manufacturing can unlock new opportunities, drive innovation, and maintain a competitive edge in their industry.

## P6: Design a data science solution to support decision-making related to a real-world problem.

### Problems Faced by ABC Manufacturing

ABC Manufacturing is facing some problems in the data collection process. Here is a description of the problems they are experiencing:

* **Unstructured data sources**: ABC Manufacturing's data is scattered across various sources such as spreadsheets, databases, and legacy systems. This makes finding and using data difficult.
* **Data quality issues**: Inaccurate, incomplete, or inconsistent data can affect decision-making. For example, data may have missing values, contain noise, or not conform to the same standard format.
* **Lack of data management**: ABC Manufacturing has not established clear data management policies and processes. This leads to a lack of data control and monitoring, making it difficult to ensure data consistency and reliability.

The above issues can reduce the efficiency and effectiveness of data use in the decision process. To solve these problems, ABC Manufacturing needs to identify and implement solutions such as building a centralized data warehouse, establishing data quality control processes, and developing data management policies.

#### Disorganized Data Collection

Different departments within the company may store data in their own ways, leading to inefficiencies in data organization and utilization.

**Problem**: The lack of standardized data collection processes and standards across different departments can result in difficulties in aggregating, analyzing, and utilizing data effectively. It can lead to inconsistencies, redundancies, and inefficiencies in data management.

**Impact**: Disorganized data collection can hinder decision-making processes, as relevant data may be scattered, difficult to locate, or not easily accessible. It can also lead to delays in information retrieval and increase the chances of errors or inaccuracies in data analysis.

**Solution**: To address this challenge, ABC Manufacturing should establish clear data collection processes and standards across the organization. The following steps can be taken:

* Develop a standardized data collection framework: Define a set of guidelines, templates, and protocols for data collection across different departments. This framework should specify the types of data to be collected, the format in which it should be recorded, and the frequency of data updates.
* Implement data collection tools and systems: Provide departments with tools and systems that facilitate consistent data collection and storage. This could include centralized databases, data entry forms, or automated data capture solutions.
* Conduct training and awareness programs: Train employees on the importance of standardized data collection practices and provide them with the necessary skills and knowledge to adhere to the established processes.
* Assign data stewards: Designate individuals or teams responsible for overseeing data collection activities, ensuring compliance with the established standards, and resolving any issues or inconsistencies.
* Regularly review and update data collection processes: Continuously monitor and improve the data collection processes based on feedback, emerging technologies, and evolving business needs.

By establishing clear data collection processes and standards, ABC Manufacturing can ensure that data is consistently and accurately captured, making it easier to aggregate, analyze, and utilize for optimizing inventory levels and making informed business decisions.

#### Lack of Adequate Storage Infrastructure

ABC Manufacturing might lack a suitable data storage infrastructure. Data could be stored on personal computers or in isolated systems that don't interact effectively.

**Problem**: The absence of a robust and centralized data storage infrastructure can pose several challenges. Data stored on personal computers or isolated systems may lead to data silos, where information is fragmented and not easily accessible to all relevant stakeholders. This can result in inefficiencies, limited data sharing, and difficulties in performing comprehensive data analysis.

**Impact**: The lack of adequate data storage infrastructure can impede the optimization of inventory levels. It may hinder the ability to integrate data from different sources, hindering accurate demand forecasting and inventory management. It can also limit the scalability and flexibility of data analysis efforts.

**Solution**: To address this challenge, ABC Manufacturing should implement a robust data storage system, such as a data warehouse or data lake, to centralize and manage data efficiently. The following steps can be taken:

* + Assess data storage requirements: Identify the types of data that need to be stored, the volume of data, and the expected growth in data over time. Consider factors such as data security, accessibility, and performance requirements.
  + Choose an appropriate data storage solution: Evaluate different options, such as data warehouses, data lakes, or a combination of both, based on the organization's specific needs. Data warehouses are suitable for structured and pre-defined data, while data lakes can handle structured, semi-structured, and unstructured data.
  + Design the data storage architecture: Define the structure and hierarchy of the data storage system. Determine how data will be organized, categorized, and accessed. Consider factors such as data partitioning, indexing, and data retention policies.
  + Implement data integration processes: Establish mechanisms to extract, transform, and load (ETL) data from various sources into the centralized data storage system. This includes defining data ingestion pipelines, data cleansing procedures, and data transformation workflows.
  + Ensure data security and governance: Implement appropriate security measures to protect data, including access controls, encryption, and regular backups. Establish data governance policies and procedures to ensure data quality, compliance, and data lifecycle management.
  + Provide data accessibility and analytics capabilities: Enable authorized users to access and retrieve data efficiently. Implement data visualization tools, query engines, and analytics platforms to facilitate data analysis and reporting.

By implementing a robust data storage system, ABC Manufacturing can overcome the challenges associated with the lack of adequate storage infrastructure. This enables efficient data integration, analysis, and utilization for optimizing inventory levels and making informed business decisions.

#### Inconsistent Data Formats

Data collected from various sources may not adhere to a consistent format, making data aggregation and analysis challenging.

**Problem**: Inconsistent data formats pose challenges when attempting to aggregate and analyze data from different sources. Data collected from various departments, systems, or external sources may have different structures, naming conventions, or units of measurement. This inconsistency can hinder data integration and lead to errors or inaccuracies in analysis.

**Impact**: Inconsistent data formats can impede the accuracy and reliability of demand forecasting and inventory optimization efforts. It can lead to difficulties in combining data from multiple sources, increasing the time and effort required for data preprocessing and transformation. Inaccurate or incompatible data formats can also result in flawed analysis and suboptimal decision-making.

**Solution**: To address this challenge, ABC Manufacturing should standardize data formats during data ingestion and apply necessary data transformations as needed. The following steps can be taken:

* + Define data format standards: Establish clear guidelines and standards for data formats, including naming conventions, data types, units of measurement, and date/time formats. Ensure that these standards are communicated and understood across the organization.
  + Implement data transformation processes: Develop data transformation pipelines or scripts that convert the incoming data into the standardized format. This may involve data cleaning, restructuring, or normalization techniques to align the data with the desired format.
  + Utilize data integration tools: Employ data integration tools or platforms that support data format conversions and data mapping. These tools can help automate the process of transforming and harmonizing data from different sources.
  + Conduct data validation and quality checks: Implement data validation processes to identify and rectify any inconsistencies or errors in the data. This can include checks for missing values, outliers, or data inconsistencies between different data sources.
  + Collaborate with data providers: Establish communication channels with data providers to ensure that data is collected and delivered in the standardized format. Provide clear instructions and guidelines for data submission to ensure consistency.

By standardizing data formats during ingestion and applying necessary data transformations, ABC Manufacturing can overcome the challenges associated with inconsistent data formats. This enables effective data aggregation, analysis, and utilization for optimizing inventory levels and making data-driven decisions.

#### Presence of Noisy Data

During data collection, inaccurate, incomplete, or unnecessary data (noise) can affect decision-making.

**Problem**: Noisy data refers to data that is inaccurate, inconsistent, or irrelevant to the analysis at hand. It can arise due to human errors, technical glitches, or data entry mistakes. Noisy data can mislead decision-making processes and lead to suboptimal inventory optimization outcomes.

**Impact**: Noisy data can distort analysis results and misrepresent the true state of inventory levels. It can lead to incorrect demand forecasting, inaccurate inventory planning, and unreliable insights. Decision-making based on noisy data may result in excessive or insufficient inventory, increased holding costs, stockouts, or missed sales opportunities.

**Solution**: To address this challenge, ABC Manufacturing should implement measures to ensure data quality and remove noise. The following steps can be taken:

* Implement data quality checks: Define validation rules and checks to identify potential issues in the data. This can include checks for missing values, outliers, inconsistencies, or data integrity problems. Automated data quality tools or scripts can be utilized to streamline this process.
* Cleanse and preprocess data: Develop data cleansing procedures to remove or correct inaccurate or inconsistent data. This can involve techniques such as imputation, outlier removal, or data normalization. Preprocessing steps may also include data integration, data deduplication, or data transformation to improve data quality.
* Define data relevancy criteria: Identify and eliminate unnecessary or irrelevant data that does not contribute to the inventory optimization process. This can involve filtering out data that is outdated, redundant, or unrelated to the analysis goals.
* Involve domain experts: Collaborate with subject matter experts who possess a deep understanding of the data and its context. They can provide valuable insights to identify and handle noisy data effectively.
* Continuously monitor data quality: Establish processes to regularly monitor and maintain data quality. This includes periodic data audits, performance tracking of data collection processes, and feedback loops to address data quality issues.

By implementing data quality checks, cleansing procedures, and preprocessing techniques, ABC Manufacturing can mitigate the impact of noisy data. This ensures that accurate and reliable data is utilized for inventory optimization, leading to better decision-making and improved business outcomes.

### Proposed Data Science Solution

The proposed data science solution aims to address the challenges faced by ABC Manufacturing, including the following key issues:

* **Disorganized Data Collection**: To tackle this challenge, the solution will include a structured data ingestion process. It will involve collecting data from various sources, such as the Customer table, WebsiteAccessHistory table, Sale History table, and Product table. The data will be extracted, transformed, and loaded into a centralized repository for further processing.
* **Lack of Adequate Storage Infrastructure**: To overcome the issue of inadequate storage infrastructure, the solution will consider implementing a scalable and efficient storage system. This may involve utilizing cloud-based storage solutions or upgrading the existing infrastructure to accommodate the growing volume of data.
* **Inconsistent Data Format**: The solution will incorporate data preprocessing techniques to address the problem of inconsistent data format. This may involve standardizing the data format across different tables and performing data cleansing tasks to remove any inconsistencies or errors. Techniques such as data normalization, data imputation, and data validation will be applied to ensure data consistency and quality.
* **Presence of Noisy Data**: To handle noisy data, the solution will employ data cleaning and outlier detection methods. These techniques will help identify and remove any outliers, errors, or inconsistencies present in the data. This will ensure that the data used for analysis and modeling is accurate and reliable.

In addition to addressing these challenges, the proposed solution will prioritize demand forecasting as a key objective. By leveraging techniques such as time series analysis and machine learning, the solution will develop models to forecast demand accurately. These models will utilize historical sales data, customer information, website access history, and product details to make reliable predictions.

By focusing on demand forecasting and addressing the challenges of disorganized data collection, lack of adequate storage infrastructure, inconsistent data format, and noisy data, the proposed data science solution will help ABC Manufacturing optimize its operations, maximize revenue, and make data-driven decisions.

### High-Level Architecture Overview

The high-level architecture overview of the proposed solution can be structured 6 layers as follows:

* Data Sources Layer

This layer represents the various sources from which the data is obtained. In the case of ABC Manufacturing, the data sources could include the Customer table, WebsiteAccessHistory table, Sale History table, and Product table. These tables contain valuable information about customers, website access history, sales, and products.

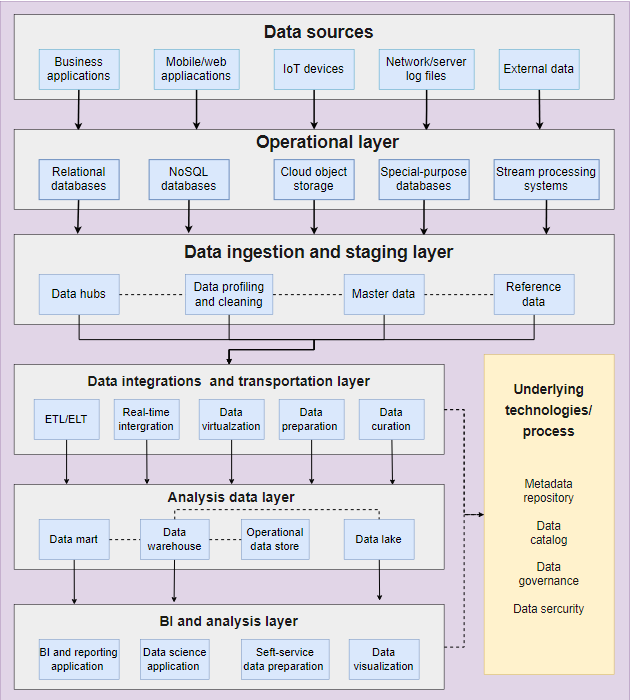
* **Operational Layer**: The operational layer represents the business processes and systems that generate and capture the data in real-time. It includes the activities related to customer interactions, website access, sales transactions, and product management. This layer ensures that the operational data is recorded and available for further processing.
* **Data Ingestion and Staging Layer**: The data ingestion and staging layer is responsible for collecting and consolidating the data from various sources. It involves extracting the data from the operational layer, transforming it into a consistent format, and loading it into a staging area or a centralized repository. This layer ensures that the data is ready for further processing and analysis.
* **Data Integrations and Transportation Layer**: The data integrations and transportation layer focuses on integrating data from different sources and transporting it to the appropriate destinations. It involves tasks such as data integration, data cleansing, data validation, and data enrichment. This layer ensures that the data is accurate, complete, and reliable before it moves to the next stages.

Figure 1: Data science process architecture

Link data science process architecture: <https://drive.google.com/file/d/1pMVU76wfDSWRb10i3r1ehBXJxdoiDIvH/view?usp=sharing>

* **Analysis Data Layer**: The analysis data layer is where the processed and integrated data resides for further analysis. It includes a storage system or a data warehouse that holds the structured and transformed data. This layer provides a foundation for performing advanced analytics, data mining, and modeling tasks. It enables efficient data retrieval and supports the generation of insights and reports.
* **BI and Analysis Layer**: The BI and analysis layer represents the front-end tools and applications that enable users to visualize, analyze, and derive insights from the data. It includes business intelligence (BI) tools, dashboards, and reporting systems. This layer allows stakeholders to explore the data, generate meaningful reports, and make informed decisions based on the insights gained from the analysis.

### Recommendations

Based on the data collected, here are some recommendations for ABC Manufacturing:

* Customer Segmentation: Analyze customer data to identify different customer segments based on demographics, purchase history, and website access patterns. This can help tailor marketing strategies and product offerings to specific customer groups.
* Personalized Marketing: Utilize customer data, such as purchase history and website access history, to personalize marketing campaigns and recommend products based on customer preferences and behavior.
* Pricing Optimization: Analyze product data and market trends to optimize pricing strategies. Consider factors like product category, customer segments, and competitive pricing to maximize revenue and profitability.

### Conclusion

The proposed data science solution for ABC Manufacturing focuses on demand forecasting for their technology products. By leveraging data analysis, time series analysis, and machine learning techniques, ABC Manufacturing can accurately forecast demand, optimize inventory levels, and make informed business decisions. Customer segmentation, personalized marketing, and pricing optimization can further enhance their operations and maximize revenue. The high-level architecture overview suggests a structured approach to data ingestion, processing, model development, and real-time monitoring. By implementing these recommendations and leveraging data-driven insights, ABC Manufacturing can gain a competitive edge in the technology product market.

## P7: Implement a data science solution to support decision-making related to a real-world problem.

### Set up the environment

Below are details on the steps to set up a working environment on Visual Studio Code with Python, download extensions, create projects and install libraries:

#### Step 1: Download and install Visual Studio Code

1. A screenshot of a computer

   Description automatically generatedVisit the official Visual Studio Code website at <https://code.visualstudio.com/>.

Figure 2: Download Visual Studio Code

2. Download the version of **Visual Studio Code** appropriate for your operating system.

3. Run the downloaded installation file and follow the instructions to complete the installation process.

#### Step 2: Download Extension

1. Open Visual Studio Code.

2. Access "Extensions" by clicking the square box icon located on the left sidebar of the user interface.

3. Search and install the following extensions:

- **Python**: Developed by Microsoft, this extension provides Python support for **Visual Studio Code**.A screenshot of a computer

Description automatically generated

Figure 3: Install Python

A screenshot of a computer

Description automatically generated - **Jupyter**: Developed by Microsoft, this extension allows you to create and work with **Jupyter** notebooks directly in Visual Studio Code.

Figure 4: Install Jupyter

#### Step 3: Create Project

1. Create a new folder on your computer to contain your project.

2. Open **Visual Studio Code** and select "**Open Folder**" from the "**File**" menu.

3. Select the project folder you just created and click the "Open" button to open the project folder in **Visual Studio Code**.

#### Step 4: Install Python libraries

1. Open **Terminal** in **Visual Studio Code** by pressing “**Ctrl +**” (or go to **View** > **Terminal** from the menu bar).

2. In **Terminal**, install the **Python** libraries by running the following commands:



Figure 5:Install library matplotlib

Figure 6: Install library pandas

### Data Collection

Data science tools enable the collection and integration of vast amounts of data from various sources. This includes structured data from databases, unstructured data from text documents and social media, and real-time data from sensors and IoT devices. By aggregating and organizing these diverse data sets, organizations can gain a comprehensive view of their operations and make data-driven decisions.

A screenshot of a computer

Description automatically generatedFor ABC Manufacturing's requirements, data collection will focus on the following aspects:

Figure 7: Raw data

* **Customer Data**: Collect customer personal and contact information to analyze purchasing behavior and improve marketing goals. This data can be collected through a CRM system when a customer signs up, makes a purchase, or interacts with customer service.
* **Website Access History**: Record customers' website access activities to better understand online preferences and behavior. This data may be collected through cookies and web analytics tools.
* **Sales History**: Records information about sales transactions, including product, quantity, and revenue. This data is often collected automatically from POS and sales management systems.
* **Product Information**: Manage detailed product information, including name, category, and price. This data is managed through a product management system and is regularly updated to reflect any product or price changes.

How to Collect Data:

* **ChatGPT**: Uses artificial intelligence (AI) to generate data from different sources.
* **Automation**: Use automated systems to collect data from various customer touchpoints, such as websites, mobile applications, and POS systems.
* **Web Analytics**: Apply web analytics tools to track user behavior on the website and collect data about the pages they visit, how long they stay, and the actions they take.
* **Sales Management System**: Integrate the sales management system with the database to automatically record every transaction and interaction with customers.
* **Security and Privacy**: Ensure that the data collection process complies with data security and privacy regulations, protecting customers' personal information.

This data allows you to store information about customers, website visit history, transaction history, and product details. By analyzing this data, you can learn about your customers' purchasing behavior, popular products, revenue per transaction, and more. Collecting and analyzing this data will help ABC Manufacturing make more accurate business decisions, from forecasting demand, managing inventory, to providing better customer service.

### Data Cleaning and Preprocessing

Using Python, the Pandas library, and the matplotlib library, perform the following steps to clean and preprocess the data:

* Delete unnecessary columns.
* Handle missing values by filling in or removing.
* Standardize data to ensure consistency.
* A screen shot of a computer screen

  Description automatically generatedEncode non-numeric data into numeric form for analysis..

Figure 8: Import library

impimport pandas as pd

The Pandas library is a powerful tool in Python for analyzing and processing tabular data. Importing as pd means you can access functions and classes in Pandas through the pd prefix.

import os

The os library allows you to interact with the operating system, such as changing the working directory, listing directory contents, and manipulating file paths.

import csv

The csv library is used to read and write CSV (Comma-Separated Values) files, a popular format for storing tabular data.

import matplotlib.pyplot as plt

A screen shot of a computer program

Description automatically generatedMatplotlib is a plotting library in Python. pyplot is a module in Matplotlib that provides a MATLAB-like interface, and plt is a common alias for calling plotting functions from pyplot.

Figure 9: Data cleaning

path = r"C:\BTEC\_FPT\IT0602-Business Process Support\ASM\DataASMBPS"

This is a string representing the directory path containing the files.

This string is enclosed in double quotes ("") and prefixed with r to indicate that this is a raw string, in which special characters such as \ are not considered escape sequences.

filepaths = []

This is an empty list variable ([ ]) named filepaths. This variable is used to store the complete paths of the files.

# Collect file paths

for file in os.listdir(path):

    filepath = os.path.join(path, file)

    if file.endswith(".csv"):

        filepaths.append(filepath)

The above code has the function of collecting the paths of CSV files from the directory specified by the path variable. Here is a line-by-line explanation of the code:

for file in os.listdir(path)

This loop runs through each file or directory in the directory specified by the path variable.

The os.listdir(path) function returns a list of files and directories in the path.

filepath = os.path.join(path, file)

This line of code creates the complete path for each file or directory in the loop.

The os.path.join(path, file) function combines the directory path (path) with the file or directory name (file) to create the full path.

if file.endswith(".csv"):

This condition checks whether the file name ends with the ".csv" extension.

If the file ends with ".csv", meaning it is a CSV file, the code will continue to process the file.

filepaths.append(filepath)

If the file meets the above condition, the complete path of that file (filepath) will be added to the filepaths list. This ensures that only the paths of the CSV files are collected and stored in the filepaths list.

A screen shot of a computer

Description automatically generatedIn short, this code iterates through the files and folders in the specified directory, creates the complete path for each file, and checks whether they are CSV files or not. If it is a CSV file, the full path of that file will be added to the filepaths list for use in the next processing steps.

The above code has the function of reading CSV files and storing data from those files into DataFrames. Here is a line-by-line explanation of the code:

dataframes = []

- This is an empty list (`[]`) named `dataframes`.

- This list will be used to store the DataFrames after they are read from CSV files.

for filepath in filepaths:

    df = pd.read\_csv(filepath)

    dataframes.append(df)

- This loop runs through each CSV file path in the `filepaths` list.

- This line of code uses the pandas library's `pd.read\_csv()` function to read data from the CSV file specified by the `filepath` variable.

- Data is read into a new frame named `df`.

Figure 10: Data cleaning

- DataFrame (`df`) containing data from the current CSV file is added to the `dataframes` list.

- This allows storing all DataFrames created from CSV files in the `dataframes` list.

In short, this code iterates through a list of CSV file paths (`filepaths`), reads data from each CSV file using `pd.read\_csv()`, and then adds a DataFrame containing the data from that file to the `dataframes` list. When completed, the `dataframes` list will contain all the DataFrames that were created from the CSV files.

# Process the dataframes

for i, df in enumerate(dataframes):

    print(f"Dataframe {i+1}:")

    print(df.head())

    print("-------------------")

for i, df in enumerate(dataframes): The for loop will iterate through the list of dataframes, with enumerate providing both the index of each element (starting at 0) and the element itself (in this case, each dataframe ).

print(f"Dataframe {i+1}:"): This line will print a string, for example “Dataframe 1:”, “Dataframe 2:”, etc., based on the index of the dataframe in the list. i+1 is used so that the index starts from 1 instead of 0.

print(df.head()): The head() method of dataframe df will return the first 5 lines of that dataframe. This line will print those 5 lines so the user can preview the data.

print("-------------------"): This line prints a series of dashes to separate the information of different dataframes, making it easier to Easy to recognize when viewing results.

This code is useful when you want to quickly see the contents of multiple dataframes, especially when you are working with data from many different sources and want to quickly check them without having to write many separate print commands.

After running the code, we will have the following data: Customer, WebsiteAccessHistory, Sale History, and Product. Each table contains different columns of data to store specific information. Let's look at each table to better understand the data that was generated:

1. Customer table:

- CustomerID: This is a unique number assigned to each customer.

- FirstName: This is the customer's first name.

- LastName: This is the customer's last name.

- Gender: This is the customer's gender, it can be "Male" or "Female" or any gender.

- Email: This is the customer's email address.

- Phone: This is the customer's contact phone number.

- Address: This is the customer's address.

1. WebsiteAccessHistory table:

- AccessID: This is a unique number assigned to each website visit.

- CustomerID: This is a foreign key associated with the Customer table, to determine which customer made the access.

- AccessDate: This is the date and time the website was accessed.

- PageVisited: This is the website that the customer has visited.

1. Sales History table:

- SaleID: This is a unique number assigned to each purchase transaction.

- CustomerID: This is a foreign key associated with the Customer table, to determine which customer made the transaction.

- ProductID: This is a foreign key associated with the Product table, to determine which product has been purchased.

- SaleDate: This is the date and time of the purchase transaction.

- Quantity: This is the number of products purchased in the transaction.

- Revenue: This is the revenue earned from transactions.

1. Product table:

- ProductID: This is a unique number assigned to each product.

- ProductName: This is the name of the product.

- Category: This is the category or type of product, for example: mobile phones, laptops, headphones, etc.

- Price: This is the price of the product.

This data allows you to store information about customers, website visit history, transaction history, and product details. By analyzing this data, you can learn about your customers' purchasing behavior, popular products, revenue per transaction, and more.

After processing and having the necessary data, we can perform data analysis to help ABC make appropriate business decisions for each time and each object.

A screen shot of a computer code

Description automatically generatedA screenshot of a computer

Description automatically generatedRemember, data science isn't just about algorithms; it's about empowering decision-makers with actionable intelligence. As ABC Manufacturing evolves, so does the landscape of supply chain management. Let's stay curious, adaptive, and committed to excellence

Figure 11: Data clean

Figure 12: Process and analyze data

The code above reads a CSV file into a dataframe and then calculates the number of customers by gender.

* + `pd.read\_csv('Customer.csv')`: This function is used to read a CSV file named 'Customer.csv' and convert it into a dataframe. A dataframe is a 2-dimensional data structure in Pandas, like a table in a database or a spreadsheet, with rows and columns.
  + `df['Gender'].value\_counts()`: The `value\_counts()` method is used to count the number of unique values in the 'Gender' column of the DataFrame `df`. The returned result will be a new Series with unique values as indices and corresponding quantities as values.
  + `print(customer\_count)`: Finally, this line will print the Series `customer\_count`, which tells you the number of male and female customers (or other genders if applicable) in the CSV file.

A screenshot of a computer

Description automatically generatedThe results of this code will provide useful information about the gender distribution of customers, which can help businesses better understand their customer structure.

Figure 13: Desired data

Running the above code we will get the following information:

A computer screen with text on it

Description automatically generated

Figure 14: Process and analyze data

This code parses customer data from a CSV file and then visualizes that information through a pie chart. Specifically, it reads data from the file 'Customer.csv', calculates the number of customers by gender, and draws a pie chart to show the percentage of each gender. This chart gives viewers a visual insight into the gender structure of the customer base. Finally, the chart is displayed on the screen. This is a useful tool in A pie chart with numbers and text

Description automatically generateddata analysis and data visualization to easily understand and share information.

Figure 15: Desired data

A screen shot of a computer code

Description automatically generated

Figure 16: : Process and analyze data

The above code does the following in Python using the pandas and matplotlib libraries:

* Convert column “**SaleDate**” to date data type using **pd.to\_datetime** function, date format specified as **%m/%d/%Y**.
* Create a new column “**Month**” to store information about the month from the “**SaleDate**” column.
* Group the data by month and calculate total revenue for each month using the groupby and sum methods.
* Create a list of month names corresponding to month numbers 1 to 12.
* Draw a column chart to display total revenue by month using the **matplotlib** library, with the x-axis representing the months and the y-axis representing the revenue.
* The column chart will display total revenue for each month, helping ABC Manufacturing easily identify revenue trends over time during the year to make appropriate business decisions. After running the above code, we will have the following data: A graph of blue bars

  Description automatically generated

Figure 17: Desired Data

Figure 18: Desired Data

The code does the following with sales data:

* Reading CSV file: First, it reads data from 'SaleHistory.csv' file and creates a dataframe from pandas library.
* A screen shot of a computer program

  Description automatically generatedProcess the 'Revenue' column: It then removes the dollar character from the 'Revenue' column and converts the value to a float.

Figure 19: Process and analyze data

* Calculate total revenue: The code calculates the total revenue from the 'Revenue' column that has been processed and prints the total revenue.
* Count transactions: It calculates the number of transactions by counting the number of rows in the dataframe and printing out the number.
* Calculate the total sales quantity of each product: It groups the data by 'ProductID' and calculates the total sales quantity for each product.
* Identify best-selling products: Find the product with the highest sales volume and print out that product's information.
* Calculate the total purchase amount per customer: Group the data by 'CustomerID' and calculate the total purchase amount per customer.
* Identify customers who buy the most: Find the customer with the most purchases and print out that customer's information.
* Convert column 'SaleDate' to date data type: Code converts column 'SaleDate' to date data type.
* Group by date and calculate daily revenue: Finally, it groups the data by date and calculates the total revenue for each day, then prints out the daily revenue.

A screen shot of a computer code

Description automatically generatedA screenshot of a computer

Description automatically generatedAfter running the above code, we will have the following data:

Figure 20: Process and analyze data

Figure 21: Desired data

The code is used to draw a bar chart showing the quantities of the most popular products. Here is a step-by-step explanation:

* **Initialize chart**: `**plt.figure(figsize=(10, 6))`** initializes a new chart with dimensions defined as 10 inches wide and 6 inches tall.
* **Draw a column chart**: `**popular\_products[:5].plot(kind='bar'**)` draws a column chart for the 5 most popular products from the `**popular\_products**` data. The **`[:5]**` section only retrieves the first 5 products from the sorted data.
* **Set X-axis label**: `**plt.xlabel**(**'Product Name'**)` labels the X-axis "**Product Name**".
* **Set the Y-axis label**: `**plt.ylabel**(**'Quantity'**)` labels the Y-axis "**Quantity**", representing the product quantity.
* **Set title**: `**plt.title**('The most popular products')` titles the chart "The most popular products".
* **Show graph**: `**plt.show**()` displays the drawn graph on the screen.

This chart helps viewers easily identify the most popular products by comparing the number of sales between them. Running the above code, we will have the following data chart:

A graph of blue rectangular bars

Description automatically generated with medium confidence

Figure 22: Desired data

The code analyzes product data from a CSV file. Here are the steps:

* A screen shot of a computer code

  Description automatically generatedReading CSV file: The code starts by reading data from the file **'SaleHistory.csv'** and creates a dataframe `**df\_products**`.

Figure 23: Process and analyze data

* Statistics on the number of products by category: Use the function `**value\_counts**()` to count the number of products in each category and store it in the variable `**category\_counts**`.
* Drawing a column chart: Next, the code uses the **matplotlib** library to draw a column chart, showing the number of products in each category.
* Convert column **'Price'** to numeric: Column 'Price' is converted from string to float, after removing dollar character and comma.
* Product value statistics: Code calculates basic statistics about product value such as average price, highest price, lowest price, etc., and prints the results.
* Discover popular products: Finally, the code counts the number of each product that appears and prints a list of popular products.

This code is useful for analyzing product data, helping users better understand product distribution by category and price, as well as identify popular products. Charts and statistics generated from this code can provide an intuitive and easy-to-understand view of product data. After running the above code, we will have the desired information as follows:

A graph of products in each category

Description automatically generated

Figure 24: Desired data

### Github

Link: <https://github.com/hoaanngg2003/Python_ASM/tree/main/ASM_CodePython>

# Conclusion

In our exploration of data science's impact on supply chain management, we've uncovered a wealth of opportunities and challenges. As we conclude this report, let's recap the key takeaways and emphasize the transformative potential of data-driven decision-making:

* Industry 4.0 Revolution: The convergence of automation, data exchange, and smart technologies has ushered in a new era for supply chain management. ABC Manufacturing stands at the forefront of this revolution, leveraging data science to optimize processes and enhance efficiency.
* Real-Time Insights: Traditional Manufacturing Execution Systems (MES) provided historical insights, but real-time data analytics now empowers proactive decision-making. By monitoring production processes, equipment performance, and logistics in real time, ABC Manufacturing gains a competitive edge.
* Demand Forecasting Precision: Accurate demand forecasting is the bedrock of effective supply chain planning. Predictive analytics and machine learning allow ABC Manufacturing to anticipate market trends, customer preferences, and future demand patterns. This precision drives resource optimization and cost savings.
* Integrated Solutions: Our focus on designing data science solutions aligns with ABC Manufacturing's goals. Integrating data from sensors, IoT devices, and production facilities enables informed decision-making. Identifying bottlenecks and streamlining logistics ensures seamless operations.
* Challenges and Vision: While data science offers immense potential, challenges persist. ABC Manufacturing must navigate the complexities of integrating autonomous processes, managing real-time data, and fostering cross-functional collaboration. Our proposed functional software architecture charts a visionary path forward.
* Practical Implementation: In the coming months, ABC Manufacturing will implement data-driven solutions. We'll evaluate their impact, measure success, and continuously refine our approach. The journey toward operational excellence is ongoing.

As the Junior Analyst, your role is pivotal. By unlocking the power of data-driven insights, you contribute to ABC Manufacturing's resilience, agility, and growth. Let's embrace this transformative journey and shape the future of supply chain management together.

.

# References

**There are no sources in the current document.**