

GROUP 3:

Agad, Jiro Laurenz S.

Aguirre, Maria Angeline

Arcibal, Kim Justine S.

Cheng, Xian Hui B.

Tingal, Franchezka Faith E.

Activity #1: Leveraging Big Data and Data Science

Scenario: Your organization, RetailCo, a mid-sized retail company, is preparing to implement a new analytics platform aimed at utilizing big data and data science. The objective is to enhance decision-making, optimize inventory management, and boost customer engagement through insights derived from data.

A. Define what big data is and explain its importance for RetailCo. What are the three primary characteristics of big data—volume, velocity, and variety—and how do they relate to RetailCo's operations? Provide examples of the different types of data RetailCo might gather (such as customer transactions, social media activity, and inventory data) and discuss how each type can influence business strategies.

Big Data Overview:

Big Data refers to a large volume of data that is processed very quickly (velocity) and comes in various formats (variety). In short, it refers to a large, complex dataset that cannot be processed with traditional database management tools due to its volume, velocity, and variety.

Importance:

Big Data is important in RetailCo's operation because it allows the company to collect huge amounts of data that can help optimise their inventory, analyse customer behaviour, and customise their marketing strategies to target specific customers. In short, big data is important for improving their business performance.

Primary Characteristics of Big Data:

1. Volume

Refers to the amount of data that is collected. This is used to store, manage and analyse huge amounts of data.

Relation to RetailCo's Operation:

RetailCo can collect hundreds of data daily, including transaction records, sales records, website logs, and social media interactions. By collecting hundreds of data, the company can analyse and predict customer preferences and satisfaction.

2. Velocity

Refers to the speed at which the data is collected, generated, and processed. This is used to analyse fast moving data like real time updates, enabling timely decisions.

Relation to RetailCo's Operation:

In RetailCo's operation, this includes data generated in real time, such as live sales updates, customer interactions, and real-time inventory updates. This real time analytics can help the company adjust prices based on demand, prevent overstocking and stockouts, and respond to customer reviews and feedback immediately.

3. Variety

Refers to the different types of data that are collected, including structured (organised and categorised) and unstructured (no specific format) datasets. This is used to help users combine and analyse different sources of information for more comprehensive understanding.

Relation to RetailCo's Operation:

In RetailCo, structured data can include transaction and inventory data, while unstructured data can include customer reviews and social media comments. These datasets can help the company identify market trends, increase customer satisfaction, and boost company profits.

Data Types Collected:

1. Customer purchase history

Type of Data: Transaction data

- This is a structured data that records all customer transaction like purchased items, prices, quantities, payment methods, and timestamp

Impact on Business Strategies:

RetailCo can use past purchase data by a customer to send personalised recommendations and targeted promotions. Customers can receive special vouchers

and information about upcoming releases based on their past purchases. For example, if a customer often buys school supplies, RetailCo can send them offers related to school supplies, like new releases of backpacks, boosting RetailCo's engagement and sales.

2. Social media interactions and engagement

Type of Data: Text, image, and video data

- This is a unstructured data from social media platforms like facebook, instagram, twitter, and tiktok that records post, comments, feedback, shares, likes, and reviews that is related to RetailCo

Impact on Business Strategies:

Engaging with customers can help RetailCo understand what their consumers think about them, see how the public perceives their brand, and spot new trends. This data can help RetailCo create marketing plans that match their customer expectations and wants. For example, if a customer posts a review or question about the company on TikTok, RetailCo can respond with helpful information or provide discount codes, creating a positive brand experience.

3. Website and app activity logs

Type of Data: Clickstream data

- This is a semi structured data that records the logs of customer interaction on the website and app of RetailCo. These records can include page visits, items in the cart, time spent in the website/app, and search queries.

Impact on Business Strategies:

This helps RetailCo understand how customers behave on the website and how the website is performing. This data can help RetailCo improve user experience and optimise the web design. For example, RetailCo can track the number of visits per page, and if the data shows that new arrivals are not often visited, RetailCo can change the layout of the website to highlight the new arrivals more and increase sales.

4. Inventory and supply chain data

Type of Data: Operational Data

- This is structured data that records the inventory and supplies of RetailCo. These records can include stock levels, supplier details, delivery schedules, logistics and order history.

Impact on Business Strategies:

Having accurate inventory and supply data helps RetailCo manage their stocks. This data ensures that popular products are always available for customers, reducing

situations where products are out of stock that could cause customer dissatisfaction. For example, Christmas decors are in high demand during a specific period of time, so RetailCo can adjust their supplies depending on the time and date to avoid overstocking and shortages.

5. Product performance

Type of Data: Product Data

- This is a structured data that records sales volume, return rates, and customer satisfaction.

Impact on Business Strategies:

Understanding product performances and how they are selling helps RetailCo save money. The data can help them decide when to stop selling products, promote products, and adjust prices. For example, based on the data, if brand X is not generating much revenue, RetailCo can discontinue selling brand X and focus on the products that generate more revenue.

6. Employee performance

Type of Data: Human Resource Data

- This is structured data that records the employees sales, customer service rating, and training completion rates.

Impact on Business Strategies:

Knowing employee performance helps RetailCo identify top employees and areas for improvement. This data can help RetailCo understand areas where some employees need more training to improve. For example, if employee performance shows that some employees struggle with upselling, RetailCo can provide training to enhance their skills, which can lead to increased sales.

7. Customer loyalty data

Type of Data: Customer Behaviour Data

- This is structured data that records information about the customer participation in loyalty programs, these records can include points earned, tier earned, and redeemed points.

Impact on Business Strategies:

Loyalty data helps RetailCo group their customers based on their tiers and loyalty points, helping them identify their most valuable customers and create strategies to keep them. For example, a customer with high loyalty points can receive early access to new collections, keeping them engaged and improving their overall experience.

B. Describe how data science can convert big data into actionable insights for RetailCo. Identify at least three specific data science methods (such as predictive analytics, clustering, and sentiment analysis) that could be utilized. For each method, explain how it can be applied in a retail setting and what potential insights or outcomes it might produce.

Overview

Since RetailCo collects large amounts of data from many sources, such as sales transactions, customer behaviours, website activity, social media, and inventory records - this huge amount of data is big data. Without a way to analyse it, big data is just raw information. Data science comes in to process this data, extract useful insights, and make it easier for RetailCo to act on these insights.

Process:

Data science uses algorithms and statistical analysis to turn big data into insights that RetailCo can act on. For example, by analyzing past sales data, RetailCo can predict a summer spike in sportswear demand. This helps the company:

- Optimise inventory by predicting which products will sell more.
- Avoid stock issues by adjusting inventory levels.
- Improve marketing by understanding customer preferences and targeting customers with relevant offers.
- Personalise promotions and offers to boost customer satisfaction and sales.

5 Data Science Methods to utilise:

1. Predictive Analytics

Uses historical data and statistical models to forecast future outcomes by using patterns found in historical and transactional data to identify risks and opportunities.

Application In Retail:

RetailCo can use predictive analytics to forecast customer demand, manage inventory, and predict customer behaviour. By analysing past sales, trends, and external factors, the company can stock popular items while reducing overstock of products that don't sell well.

Potential Insights/Outcomes:

Predictive analytics helps RetailCo reduce stockouts, improve cash flow, and avoid excess inventory. It also enhances marketing by predicting customer preferences, leading to personalised promotions.

2. Clustering

Data mining technique that groups data points (like customers) based on similarities into similar clusters. These groups are distinct from other clusters, which makes it easier to analyse the data and make targeted decisions.

Application In Retail:

RetailCo can use clustering to segment customers based on their purchasing habits, preferences, and demographics. For example, customers could be grouped into high-spenders, budget-conscious shoppers, or frequent buyers. This segmentation can help RetailCo tailor its marketing campaigns and promotions to the needs of specific customer groups.

Potential Insights/Outcomes:

Through clustering, RetailCo can develop targeted marketing strategies that resonate with different customer segments. It can also refine inventory management by stocking products that cater to specific groups, improving product availability.

3. Sentiment Analysis (Opinion Mining)

Uses natural language processing (NLP) to analyse customer opinions from text data in order to help organisations understand how customers feel about their products and services.

Application In Retail:

RetailCo can apply sentiment analysis to evaluate customer opinions in reviews, social media mentions, or surveys. By understanding whether customers have positive, negative, or neutral feelings toward products or services, they can modify it based on this.

Potential Insights/Outcomes:

Sentiment analysis gives real-time insights, enabling RetailCo to quickly address negative feedback and improve customer experience. It also highlights

product strengths and weaknesses, helping to fine-tune product lines and marketing strategies.

4. Market Basket Analysis

Data mining technique used to identify product associations based on customers' purchasing behaviour. It helps retailers understand which items are frequently bought together, often referred to as "affinity analysis."

Application In Retail:

RetailCo can use market basket analysis to discover relationships between products frequently bought together. For example, if customers tend to buy batteries with electronics, RetailCo can use this insight for cross-selling or creating product bundles. This can also be applied to online recommendation systems.

Potential Insights/Outcomes:

This strategy boosts sales through cross-selling and upselling, optimises product placement, and increases average transaction value. It also improves inventory management and enhances the shopping experience with personalised product recommendations.

5. Churn Prediction

Use machine learning to predict which customers are likely to leave based on behaviour patterns. It identifies patterns in customer behaviour, such as reduced purchases or lower interaction with marketing materials (potential churn)

Application In Retail:

RetailCo can use churn prediction to spot customers who may stop shopping with the brand by analysing buying habits, loyalty program participation, and engagement with promotions. This enables RetailCo to offer timely incentives or personalised service to encourage repeat purchases.

Potential Insights/Outcomes:

Churn prediction helps RetailCo target at-risk customers with tailored offers, leading to better customer retention and increased sales. This approach improves long-term customer loyalty, raises average lifetime spending, and stabilises revenue streams for the business.

C. Identify the potential challenges RetailCo might encounter when implementing the analytics platform. Highlight at least three challenges (such as issues with data quality, difficulties integrating with existing systems, and resistance from employees) and, suggest practical solutions for overcoming each of these challenges.

Potential Challenges:

1. Data Quality Issues

Introduction

Data quality in data analysis is its state being accurate, comprehensive, consistent and relevant. Ensuring high-data quality for RetailCo is essential as having accurate analytics lead to smart business decisions. While the low-quality data can lead to revenue loss due to bad business decisions. It can also significantly affect RetailCo's ability to understand customer preferences, optimise inventory and forecast demand because of the insufficient quality of data.

Poor data quality can lead to inaccurate analysis and poor business decisions because it provides insufficient or incorrect information, which should otherwise serve as the foundation for informed decision-making. When data is unreliable, it prevents businesses from drawing accurate insights, resulting in misguided strategies and inefficiencies in operations. For example, RetailCo may misunderstand their target market and create ineffective marketing strategies if they have incomplete customer data. Inaccurate inventory data may also lead to situations of overstock or stockout, which would decrease revenue and consumer satisfaction.

Common Data Quality Issues they might face:

a. Duplicate Data

- **Situation:** Several records of the same client because of different spellings of their names or different purchase entries.
- **How it Affects RetailCo:** Decisions about marketing and inventories may be impacted by inaccurate analysis of consumer purchasing patterns brought on by duplicate data.
- **Solution:** Make that every new record has been verified against existing records before being added by implementing real-time data validation checks at the point of entry. This stops duplicate entries from ever being created.

b. Incorrect Data Entries

- **Situation:** Customer support staff may record customer complaints or questions incorrectly, or RetailCo employees may manually input wrong product pricing.

- **How it Affects RetailCo:** Product listings that are priced incorrectly may result in lost sales or dissatisfied consumers. In the same way, insufficient replies to customer complaints that are not properly observed could negatively affect connections with customers.
- **Solution:** Validate data automatically at the time of entry. Make use of mandatory fields, dropdown menus, and confirmation checks to guarantee accuracy. RetailCo could also use staff instruction to reduce human error.

c. Incomplete Data

- **Situation:** RetailCo may have client records that lack important information like product descriptions or measurements, or customer data that lack fields like addresses.
- **How it Affects RetailCo:** This missing information may result in unsuccessful advertisements or mistakes in order fulfilment, which would leave customers dissatisfied.
- **Solution:** To guarantee that crucial information is recorded at the point of entry, use necessary fields and automatic data validation methods.

2. Data Privacy Concerns

Introduction

Data privacy and security refer to the protection of sensitive personal information from data breaches and illegal access. It is important to maintain the customer satisfaction and protect the company's reputation by securing the customer data such as payment information and personal details.

The importance of data privacy and security to RetailCo is because this company handles large amounts of customer data, including payment information that would have a financial impact for the customers. Protecting this data is essential for maintaining customer trust and preventing expensive data breaches, in addition to being required by law under laws like the CCPA and GDPR.

Common Data Privacy Concerns they might face:

a. Data Breaches

- **Situation:** RetailCo could be the subject of cyberattacks, which might expose personal data or financial information belonging to customers.
- **How it Affects RetailCo:** A data breach would cause financial losses, severely damage trust among consumers, and may result in fines for violating privacy laws.

- **Solution:** Invest in cutting-edge security tools like multi-factor authentication and encryption, and perform frequent security assessments.

b. Compliance with Regulations

- **Situation:** RetailCo must abide by a number of data privacy rules, including the CCPA in California and the GDPR in Europe, because it operates in numerous locations.
- **How it Affects RetailCo:** Serious legal consequences, fines, and operational limitations might result from noncompliance.
- **Solution:** Implement compliance procedures, such as getting consumer consent for data usage and making sure data management methods are transparent, and stay informed of developing data protection laws.

c. Internal Data Misuse

- **Situation:** RetailCo employees may have access to private customer data and misuse it for their own benefit or disclose it to third parties.
- **How it Affects RetailCo:** Employee misuse of client data may lead to loss of trust among customers, legal penalties, and negative effects on the company's brand.
- **Solution:** To identify any misuse, establish strict access restrictions, instruct employees on data ethics, and keep monitoring out for unusual access patterns.

3. System Integration Problems

Introduction

System Integration is a process of connecting different sub-systems such as software applications, hardware systems, and data sources to function as one. It ensures that all data entered by these systems is automatically shared across them. This is achieved through Applications Programming Interfaces (APIs), which allow software programs to communicate using rules and protocols. Middleware, a layer of software, acts as a bridge between different systems, allowing real-time data exchange, making the updates made in one system reflect instantly on other connected systems. This eliminates the need for manual data transfers, which can lead to errors and inefficient results.

RetailCo operates multiple systems to ensure smooth operations, including point-of-sale (POS), inventory, customer relationship management (CRM), and new analytics platforms. These systems track crucial information such as sales data, stock

levels, and customer data. Without proper integration, data would disintegrate within each system, complicating RetailCo's access to complete and accurate operational results.

Common System Integration Problems they might face:

a. Legacy System Incompatibility

- **Situation:** The POS systems of RetailCo's are outdated and compatible only with earlier software versions.
- **How It Affects RetailCo:** It could be more difficult for RetailCo's to obtain real-time data, and they may need to manually transfer data which will be time consuming and might cause errors.
- **Solution:** RetailCo should be thinking about upgrading the systems to modern versions which are compatible with the current analytics platform.

b. Data Silos Across Departments:

- **Situation:** The departments of RetailCo (sales, inventory management, customer service, and marketing) use different systems that do not communicate well with one another.
- **How It Affects RetailCo:** There could be disparities and contradictions in the data collected because different departments may have different versions of it.
- **Solution:** RetailCo can use integration tools like middleware so that it could link existing systems. These tools can ensure that all departments have access to the same real-time data and will reduce silos.

c. Real-Time Data Integration:

- **Situation:** RetailCo should have data throughout its system to react quickly to changes in inventory levels.
- **How It Affects RetailCo:** Their inventory may be inaccurate where the stocks levels are not updated instantly after a sale.
- **Solution:** Utilize APIs to implement real-time system integration that allows for immediate transfer of data between the point-of-sale systems

4. Scalability and Performance Issues

Introduction

Scalability is the ability of a company to operate under growing and advancing data volumes and workloads. In the case of a mid-size company such as RetailCo, with the expansion of the organization the system will generate more data, its systems should be able to accommodate this growing data without system breakdowns.. Planning for scalability ensures that the systems of RetailCo can handle an increasing number of customers, larger data volumes, higher transactional activities, all while making sure that operations run smoothly.

Common Scalability and Performance Issues they might face:

a. Handling Growing Data Volumes

- **Situation:** Increased activities of RetailCo have led to a rapid buildup of data from sales, customer interactions, and stock management.
- **How It Affects RetailCo:** There might be a possibility of lag or delay in RetailCo's systems that will in turn lead to the slow processing of transactions or reports.
- **Solution:** RetailCo should try storage and computing platforms that are cloud-based and can scale elastically in order to handle the increasing data volumes without any decline in performance.

b. System Performance

- **Situation:** In case of flash sales, the performance of the RetailCo system drops and it causes a delay in the processing of payments, updating the inventory and even logging in customers which hinders user experience across different platforms
- **How It Affects RetailCo:** When systems tend to lag at the checkout of the transaction, the buyers may abandon their carts which may lead to loss of sales and a negative impact on the brand.
- **Solution:** Use load balancing to route incoming traffic to several servers in such a way that no single server becomes overloaded during peak hours. It helps balance between the traffic load and prevent the system from crashing out due to an overload.

c. Scalability of Machine Learning Models

- **Situation:** The current machine learning models within RetailCo may not cope with the increasing volumes of data that the company experiences as it expands. This in turn may lead to incorrect models and ultimately an unpleasant experience to the customers.
- **How It Affects RetailCo:** As the data increases, the models may become less effective if they cannot be regularly retrained with the

most up-to-date information, resulting in outdated or less accurate predictions

- **Solution:** Retailco should adopt Automated Machine Learning since it makes machine learning easier by automating system operations like product recommendation and forecasting demands of the customers

5. Employee Resistance

Introduction

Employees display resistance to change when they are hesitant or unwilling to adopt new technology or methods presented in the workplace. This might occur for a variety of reasons. For example, employees may be concerned that new technology may replace their occupations, resulting in job insecurity. Others may believe they lack the necessary skills to manage the new tools or systems, making them concerned about their capacity to adapt. In addition, learning new workflows or procedures might be difficult, especially if people have been doing things the same way for a long time.

Given that employees are the primary users of the analytics platform, their commitment to it is critical to RetailCo's success in implementing it. Without their assistance, the platform may be underutilised, resulting in lower ROI and efficiency. Ensuring their acceptance through communication and training makes adoption easier and maximises the platform's advantages.

Common Employee Resistance Problems they might face:

a. Fear of Job Loss

- **Situation:** Some RetailCo employees may fear automation will replace their roles, especially in manual decision-making processes.
- **How It Affects RetailCo:** This fear can lead to low morale, disengagement, and even staff turnover, slowing platform adoption.
- **Solution:** Clear communication about job security and offering training programs can reassure employees and help them transition to new roles supported by technology.

b. Lack of Data Literacy

- **Situation:** Many RetailCo employees may lack the ability to interpret and work with data, making it difficult to make use of the platform's capabilities.
- **How It Affects RetailCo:** Low data literacy limits the ability to make data-driven decisions, lowering the platform's value.

- **Solution:** RetailCo should invest in data literacy training programs to prepare employees to use the analytics platform effectively.

c. Resistance to Changing Workflows

- **Situation:** Long-term employees may be resistant to replacing familiar workflows with new technology-driven processes.
- **How It Affects RetailCo:** Resistance to workflow changes can cause delays in adoption and inefficiencies, reducing the platform's expected gains.
- **Solution:** Involving employees in the design and implementation of new workflows, as well as providing ongoing support, can make the transition easier and increase buy-in.

6. Data Integration Issues

Introduction

To develop a new analytics platform, RetailCo. must combine a wide range of data sources including inventory management systems, CRM software, and point of sale (POS) systems. To accomplish successful data use, a variety of challenges faced by this integration must be overcome.

Common Data Integration Problems they might face:

a. Fragmented Data Sources

- **Situation:** RetailCo has various data groups, including customer data in the CRM, sales data in POS systems, and inventory data in separate management software. Each system runs autonomously, making it impossible to gain a comprehensive understanding of business processes.
- **How It Affects RetailCo:** Fragmented data produces inconsistent information, making it difficult to analyze client behavior and inventory levels. This unpredictability may hinder strategic decision-making, resulting in missed sales opportunities and inefficient inventory management.
- **Solution:** Create a centralised data warehouse that combines information from all sources. Use the extract, transform, and load processes to ensure that data from different systems is consistent, allowing for more thorough analytics and reporting

b. Real-time Data Processing

- **Situation:** RetailCo needs quick information for flexible inventory management and customer engagement. However, present systems are

not designed to handle data in real time, resulting in delays in decision-making.

- **How It Affects RetailCo:** Without real-time data processing, RetailCo may overlook crucial sales trends, resulting in stock outs or overstocks. This gap might reduce client satisfaction while increasing operating costs.
- **Solution:** Adopt streaming data technologies for real-time data gathering and analysis. This might involve utilising cloud-based solutions that provide the infrastructure required for real-time data processing, ensuring that insights are immediately usable.

c. Data Duplication Across Systems

- **Situation:** RetailCo's current systems frequently result in data duplication, in which the same customer or product information is recorded in various locations. This redundancy complicates data collection and analysis.
- **How It Affects RetailCo:** Data duplication might result in wrong reporting and insights, influencing inventory decisions and marketing tactics. It can also lead to a confused experience for customers who receive inconsistent messaging.
- **Solution:** Implement data governance policies and tools to encourage data cleaning and deduplication. Regular audits of data quality should be performed to identify and correct duplication, ensuring that all systems reflect accurate and consistent information.