

CMPG 321: GROUP PROJECT

PROJECT TITLE:

**ClearVue Sales Intelligence
System**

**Requirements analysis and conceptual
design**

Group Name and number:

Group29- SQL Slashers

Table of Contents

Table of Contents	1
Group Members	2
Executive Summary.....	4
Requirements Definition Approach	5
Processes for requirements gathering and data	5
Stakeholders considered in this analysis.....	5
Business requirements.....	6
Sales Reporting and Analysis	6
Product Sales Reporting.....	6
Customer Data Reporting.....	6
Supplier Focused Analytics and Reporting	6
Payment and Financial Reporting.....	6
High-Level Business Objectives	7
Analytical and information requirements	8
Sales and Profitability Analysis	8
Product and Inventory Analysis.....	8
Financial and Customer Analysis	8
Preliminary Source System Analysis	9
Analysis of Data Sets.....	9
Relationships and Hierarchies	10
Preliminary Success Criteria:.....	11
Conceptual NoSQL Design	12
The SALES Collection	12
The CUSTOMER Collection	13
The SUPPLIER Collection	14
The FINANCE Collection	16
AI Usage Log	17
AI prompts and output used.....	17
Explanation of validations	17
BIBLIOGRAPHY/REFERENCES	18

Group Members

Group Leader:



Name and Surname: Kevin Mbolongwe

Student number: 41503619

Contact info: Tel no. :+27 81 716 4849

Other group members:



Name and Surname: Jade Loggenberg

Student number: 41556763

Contact info: Tel no. :+27 62 447 8387



Name and Surname: Thakgalo Sehlohla

Student number: 41584333

Contact info: Tel no. :+27 69 328 9696



Name and Surname: Nandi Mahlangu

Student number: 40983889

Contact info: Tel no. :+27 62 104 4557



Name and Surname: LR MAGANO

Student number: 40970507

Contact info: Tel no. :+27 61 494 0471



Name and Surname: Ishmael Ibrahim

Student number: 45737185

Contact info: Tel no. :+27 67 806 4752

Executive Summary

The purpose of this project is to upgrade ClearVue Ltd.'s existing Business Intelligence (BI) system, which currently relies on an outdated relational database, to a more agile and scalable solution. The current system, primarily based on the provided MS Access data, struggles to provide timely, comprehensive reporting for management.

Our proposed solution is a modern, flexible technology stack featuring MongoDB and Power BI. MongoDB will serve as the core NoSQL database, ingesting and transforming raw data from the provided spreadsheets. Its flexible, document-based design is uniquely suited to handle hierarchical data and will eliminate the need for complex joins, directly addressing the key project success criterion of achieving a maximum five-second query execution time for sales reports.

Power BI will be used for visualisation and reporting, providing an intuitive dashboard that enables senior management to independently access and analyse data. The system is designed to provide actionable insights into sales, product, customer, and supplier performance, supporting a strategic pivot toward supplier-focused analytics and ensuring data accuracy aligned with ClearVue's unique financial calendar. By migrating to this new system, ClearVue Ltd. will gain a scalable, high-performance BI capability that drives informed strategic decision-making and enhances overall organisational performance.

Requirements Definition Approach

The requirements for this project were defined through a comprehensive, multi-faceted approach to ensure alignment between business needs and the technical solution. Our methodology went beyond a simple data review to proactively identify opportunities for value creation and to address underlying data quality issues.

Processes for requirements gathering and data

- **Source System Data Analysis:** We conducted a detailed analysis of all provided Excel files to understand the existing relational structure, key fields, and inherent data relationships.
- **Conceptual Modelling and Stakeholder Alignment:** Based on the data analysis, we developed conceptual data models, such as the CUSTOMER collection and the SUPPLIER collection. This approach allowed us to directly translate business requirements into technical designs.
- **Data Quality Assessment:** As part of our analysis, we performed a preliminary data quality assessment. This included identifying and documenting several data cleaning rules.
- **Do background research on the company and industry standards using the internet**
- **Documenting and validating requirements through questionnaires and surveys**
- **Gathered direct input from stakeholders through interviews**
- **Prototyping the new system so that users can have an idea of what the next solution could look like**
- **Observing the users of the old system**

Stakeholders considered in this analysis

This project's success is defined by its ability to meet the diverse needs of key stakeholder groups within ClearVue Ltd. By understanding their specific requirements, we have designed a solution that provides targeted, actionable insights.

- **Senior Management:** The primary stakeholders, who require the most critical insights to drive strategy. Their needs include intuitive dashboards and enhanced reporting capabilities.
- **Finance Officers:** This team requires detailed financial reporting to manage the company's fiscal health. Their core analytical needs include age analysis, payment tracking, and product sales reports.
- **Sales Representatives:** Sales representatives require reports to understand their individual performance. They need access to key metrics on sales commissions, customer activity, and sales trends.
- **Inventory and Purchasing Department:** This department's needs are focused on the supply chain. They require insights into inventory inflow and supplier performance, which will be provided by the supplier-focused analytics.

- Project managers: Who will oversee data warehousing projects, and ensure that the planning, procurement and execution of projects meet the defined scope. Project managers would need the data to ensure that the projects have a defined start and finish.

This rigorous process ensures that every requirement is tied to a specific business need, is supported by the source data, and addresses potential data integrity challenges, setting a solid foundation for the project's success.

Business requirements

The business requirements for the new BI system are derived from the need to transform existing operational data into actionable insights for strategic decision-making. The system must fulfil the following core requirements, with additional details and context provided to ensure comprehensive coverage.

Sales Reporting and Analysis

- The system must provide real-time and historical sales data analysis.
- This includes generating key reports (daily, weekly, annual, historical, etc.)

Product Sales Reporting

- The new system must support detailed analysis of product performance.
- This will be achieved by using a NoSQL model that embeds rich product information (e.g., GENDER, MATERIAL, STYLE) from Products.xlsx and Products Styles.xlsx directly within sales documents

Customer Data Reporting

- The CUSTOMER collection will be designed to embed dimensional data from lookup tables such as Customer Account Parameters, Customer Categories, and Customer Regions.
- This will also include products most bought by customers according to categories

Supplier Focused Analytics and Reporting

- A key business objective is to pivot toward analysing supplier-side data
- The new system must support this by consolidating data from Suppliers.xlsx, Purchases Headers.xlsx, and Purchases Lines.xlsx into a single, comprehensive SUPPLIER collection.

Payment and Financial Reporting

- The system must enable a clear view of the company's financial health.
- It must provide reporting on payments from the Payment Header and Payment Lines.
- A crucial requirement is the ability to perform a detailed **age analysis** of outstanding payments, which is supported by the AGE_ANALYSIS dataset and is a key metric for finance officers to measure overdue payments and assess risk.

High-Level Business Objectives

- **Improve Strategic Decision-Making:** The new system will provide senior management with quick, actionable insights. A key measure of success is query performance, which must achieve maximum execution for all key sales reports. This directly supports the need for timely information to guide strategic business decisions.
- **Demonstrate NoSQL Benefits:** The project aims to prove the superiority of a NoSQL solution over a traditional relational database for this use case. Success will be shown by exemplifying how the flexible, document-based data model can manage hierarchical data and support increased scalability for future growth.
- **Ensure Data Accuracy and Time Alignment:** The system must accurately reflect business performance based on ClearVue Ltd.'s unique financial calendar. All reports and visualisations will be aligned with the company's financial periods , ensuring data integrity and consistency.
- **Support Scalability and Flexibility:** The conceptual NoSQL data model must be scalable and adaptable. A key success factor is the system's ability to integrate and analyse supplier data seamlessly. This supports the company's planned strategic shift toward more comprehensive supplier-focused analytics, enabled by consolidating purchasing and supplier information into a single collection.
- **Improve User Experience and Adoption:** The final Power BI dashboard must be intuitive and easy for senior management to operate, enabling them to independently query and visualise sales data without relying on IT.

Analytical and information requirements

The following analytical and information requirements have been identified for the new Business Intelligence system. They are designed to provide key stakeholders with the insights needed to make informed decisions.

Sales and Profitability Analysis

- *Sales Performance by Financial Period:* Generate daily, weekly, and monthly reports that track total sales and quantity sold.
- *Profit Trend Analysis:* Create a report that shows total profit over time. This will be made possible by a calculated field ($TOTAL_PROFIT = TOTAL_LINE_PRICE - LAST_COST$) embedded within each sales document, allowing for a fast and direct query to track profit trends by financial period
- *Sales Performance by Customer and Representative:* Analyse sales data by customer, representative, and region. This report will combine `TOTAL_LINE_PRICE` from the Sales Line collection with `CUSTOMER_NUMBER`, `REP_CODE`, and `REGION_CODE` from the Sales Header and embedded Customer documents, providing a comprehensive view of performance by key stakeholders.

Product and Inventory Analysis

- *Top-Selling Products by Attribute:* Identify top-performing products based on attributes like brand, category, gender, and style. The proposed NoSQL design embeds product information within each sales line item.
- *Inventory Inflow and Supplier Performance:* Provide detailed reports on purchase trends and supplier performance. This is supported by the Supplier collection, which consolidates data from Suppliers and the Purchases files. This allows for reports that track total purchased quantity, total cost, and supplier terms, enabling effective supply chain management.

Financial and Customer Analysis

- *Payment and Age Analysis:* Generate reports that track payments and outstanding invoices. This will use data from Payment Lines and `AGE_ANALYSIS`. The reports will provide a clear view of financial health, including overdue payments and accounts at risk, which is a critical requirement for the finance team.
- *Customer Segmentation and Performance:* Analyse customer performance based on their segment and region. The report can show which customer types and regions are most profitable, guiding marketing and sales strategies.

Preliminary Source System Analysis

Analysis of Data Sets

- **AGE_ANALYSIS:** This dataset documents outstanding customer invoices and their due dates in 30-day increments.
- **CUSTOMER:** This file contains dimensional data about each customer, including CUSTOMER_NUMBER, CCAT_CODE, REGION_CODE, and CREDIT_LIMIT. This file serves as the basis for the CUSTOMER collection model.
- **Customer Account Parameters:** This dataset provides the type of account parameter (e.g., 'Closed', 'Consignment', 'Trading Account') for each CUSTOMER_NUMBER.
- **Customer Categories:** This lookup table provides a category description (CCAT_DESC) for each CCAT_CODE.
- **Customer Regions:** This lookup table provides a region description (REGION_DESC) for each REGION_CODE.
- **Payment Header:** This transactional dataset holds the deposit reference for each customer payment.
- **Payment Lines:** This transactional dataset provides detailed information for each deposit, including the date, bank amount, discount, and the total payment.
- **Products:** This dataset contains dimensional data about the company's inventory, with fields like INVENTORY_CODE and PRODCAT_CODE.
- **Product Brands:** A lookup table providing a description (PRODBRA_DESC) for each PRODBRA_CODE.
- **Product Categories:** This lookup table links PRODCAT_CODE to a description, a BRAND_CODE, and a PRAN_CODE.
- **Product Ranges:** A lookup table providing a description (PRAN_DESC) for each PRAN_CODE.
- **Product Styles:** This dataset contains descriptive details for each INVENTORY_CODE, such as GENDER and MATERIAL.
- **Purchases Headers:** This transactional dataset contains purchase order information, including SUPPLIER_CODE, PURCH_DOC_NO, and PURCH_DATE.
- **Purchases Lines:** This transactional dataset provides detailed line-item information for each purchase document. Key fields include PURCH_DOC_NO, INVENTORY_CODE, QUANTITY, UNIT_COST_PRICE, and TOTAL_LINE_COST.
- **Representatives:** This is a lookup table providing a description (REP_DESC) for each REP_CODE, as well as COMM_METHOD and COMMISSION.
- **Sales Header:** This transactional dataset contains key sales document information.

- **Sales Line:** This transactional dataset provides detailed line-item information for each sales document.
- **Suppliers:** This dataset contains dimensional data about each supplier, with fields like SUPPLIER_CODE, SUPPLIER_DESC, NORMAL_PAYTERMS, and CREDIT_LIMIT. This file serves as the basis for the conceptual SUPPLIER collection model.
- **Trans Types:** This is a lookup table that provides descriptions for transaction type codes.

Relationships and Hierarchies

- The CUSTOMER dataset links to Customer Account Parameters, Customer Categories, and Customer Regions via CUSTOMER_NUMBER, CCAT_CODE, and REGION_CODE, respectively. It also links to Sales Header and Payment Header via CUSTOMER_NUMBER.
- The SALES HEADER dataset links to Sales Line via DOC_NUMBER (1:M relationship) and to Trans Types via TRANSTYPE_CODE (1:M relationship).
- The SALES HEADER also links to Representatives via REP_CODE, which is an excellent candidate for embedding.
- The PURCHASES HEADERS dataset links to Purchases Lines via PURCH_DOC_NO (1:M relationship) and to Suppliers via SUPPLIER_CODE.
- The PRODUCTS dataset links to Product Categories via PRODCAT_CODE and to Product Styles via INVENTORY_CODE.
- The AGE_ANALYSIS dataset links to CUSTOMER via CUSTOMER_NUMBER and has 1:M relationship(one customer has many age analysis).
- The PAYMENT_LINE dataset links to CUSTOMER and PAYMENT_HEADER ,and has a 1:M relationship (one customer hand has many payments)

Preliminary Success Criteria:

Query performance: The system must be able to generate all key sales reports (daily, weekly, annual, historical, etc) with a maximum query execution time of five seconds. This will ensure management can access timely insights for strategic decision-making.

Demonstration of NoSQL Advantages: This prototype will be considered successful if it demonstrates the superiority of a NoSQL-based solution compared to traditional relational databases. This includes showcasing its enhanced flexibility in handling hierarchical data and its greater scalability.

Data Accuracy and Temporal Alignment: All reports and visualisations must accurately reflect sales data as defined by ClearVue Ltd's unique financial year structure. The system should precisely align data with the financial month.

Scalability and Flexibility: The conceptual NoSQL data model must demonstrate its ability to scale for future growth and adapt to new data types. A key success factor is the system's readiness to integrate and analyse supplier data, aligning with the company's anticipated strategic shift.

User Experience and Adoption: The BI dashboard must be intuitive and easy for ClearVue's senior management to navigate, enabling them to query and visualise sales data independently. The insights provided should be actionable, supporting informed strategic business decisions.

Conceptual NoSQL Design

The SALES Collection

The design embeds all line-item details, including product dimensions and measures, directly within a single sales document. This makes it possible to retrieve all information about a sale with a single, fast query

- **MongoDB JSON Example:**

```
{
  "_id": "DC700467",
  "trans_type_code": "2",
  "trans_type_desc": "CREDIT NOTE",
  "customer_number": "ESP100",
  "rep_code": "02JUL",
  "trans_date": "2019-03-25",
  "fin_period": "201901",
  "total_revenue": 1000.0,
  "total_cost": 500.0,
  "line_items": [
    {
      "inventory_code": "123ABC",
      "quantity": 2,
      "unit_sell_price": 500.0,
      "unit_cost": 250.0,
      "total_line_price": 1000.0,
      "profit": 500.0
    },
    {
      "inventory_code": "456DEF",
      "quantity": 1,
      "unit_sell_price": 0.0,
      "unit_cost": 0.0,

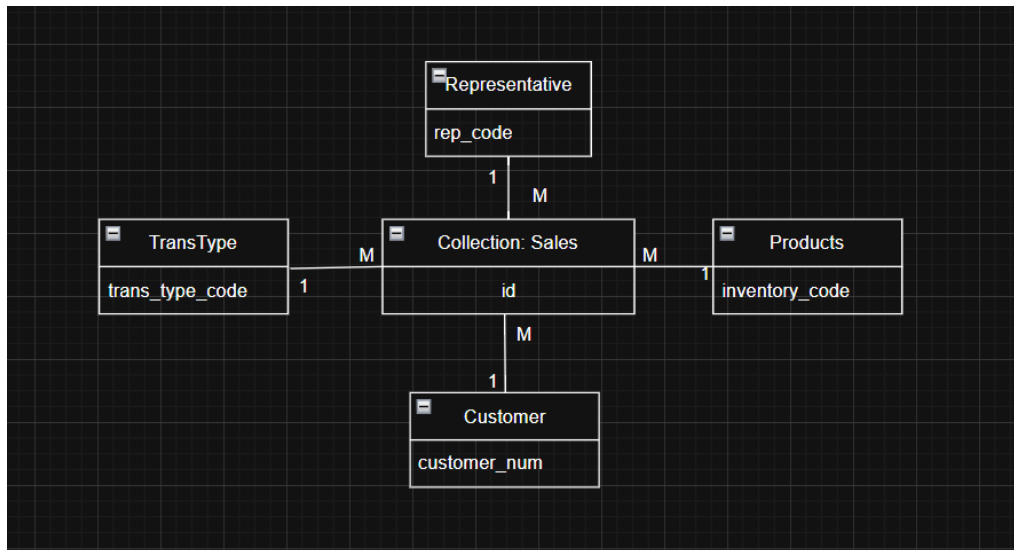
```

```

    "total_line_price": 0.0,
    "profit": 0.0
  }
]
}

```

- **Diagram of Sales Collection:**



The CUSTOMER Collection

This collection acts as a master dimension for all customer information. The design embeds lookup table data directly into the customer document to avoid joins and facilitate fast queries.

- **MongoDB JSON Example:**

```

{
  "_id": "AKRA01",
  "customer_name": "Akra Trading Ltd",
  "customer_categories": {
    "ccat_code": 27,
    "ccat_desc": "Rustenburg Brits"
  },
  "region": {
    "region_code": "1a",
    "region_desc": "Pretoria Central"
  },
}

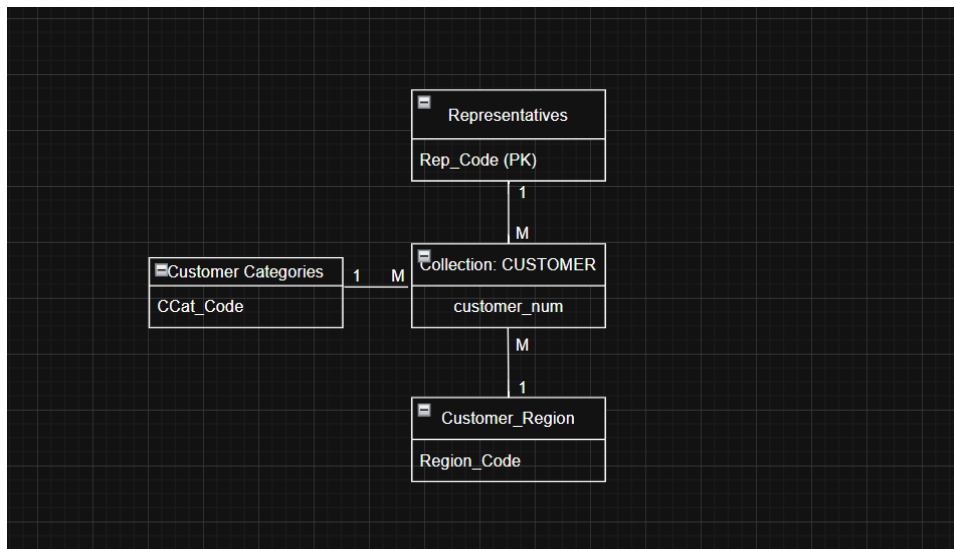
```

```

"rep_code": "010",
"credit_limit": 3000.0,
"settle_terms": 0,
"normal_payterms": 120,
"discount": 0.0,
"status": "active"
}

```

- **Diagram of Customer Collection**



The SUPPLIER Collection

This collection is structured to provide comprehensive supplier and purchasing data in a single document, perfect for supply chain analytics. It organizes all purchase orders and their line items as a single document, which is ideal for analysing inventory inflow and supplier performance.

- **MongoDB JSON Example:**

```

{
  "_id": "PN000427",
  "purchaseDate": "2018-03-08",
  "financialPeriod": "201803",
  "supplier": {
    "supplierID": "006",
    "name": "DR Supplier",
    "excludesVAT": true,

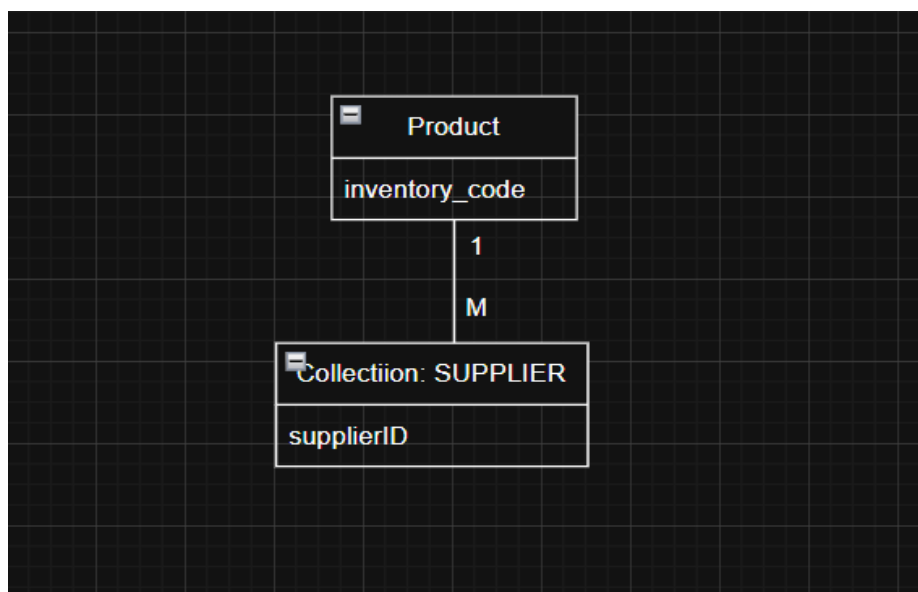
```

```

    "paymentTerms": 90
  },
  "lineItems": [
    {
      "productID": "8176MCO568X",
      "quantity": 10,
      "unitCost": 179.23,
      "totalCost": 1792.30
    },
    {
      "productID": "8180MCO568",
      "quantity": 10,
      "unitCost": 163.70,
      "totalCost": 1637.00
    }
  ],
  "totalPurchaseCost": 57000.01
}

```

- **Diagram of Supplier Collection**



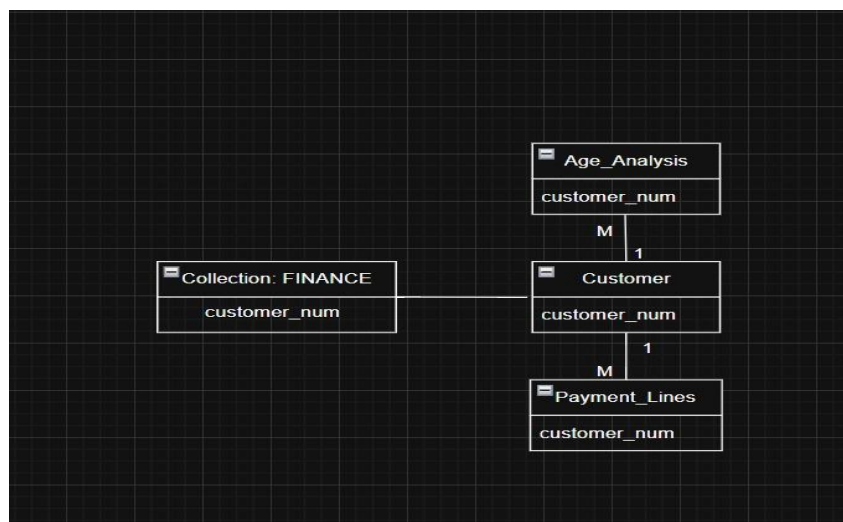
The FINANCE Collection

This collection, derived from the model, is designed for efficient payment and age analysis, consolidating key financial data.

- **MongoDB JSON Example:**

```
{
  "customer_number": "AACJ01",
  "fin_period": "201903",
  "total_due": 470.64,
  "amt_current": 7.32,
  "days_due": {
    "30": 7.32,
    "150": 456,
  },
  "payment_lines": [
    {
      "deposit_date": "2019-08-30",
      "deposit_ref": "DB06-176",
      "bank_amount": -860.93,
      "discount": -45
    }
  ]
}
```

- **Diagram of Finance Collection**



AI Usage Log

This log documents the specific ways AI such as ChatGPT ,Copilot and Grok was used throughout the project development process, to draft business requirements and schema examples.

AI prompts and output used

- Customer collection and business objectives:
<https://copilot.microsoft.com/shares/Dd5861fxSSFj67nz5nmuv>
- Structure business requirements: <https://chatgpt.com/share/68b18588-4a40-8012-a708-efa0da719961>
- Age_Analysis table: <https://chatgpt.com/share/68af355e-fd50-8003-9908-334572662633>
- Requirements gathering process: <https://chatgpt.com/share/68af370d-006c-8003-8d4b-d2ba98609624>
- Payment_Lines Table: <https://chatgpt.com/share/68af3abe-8d34-8003-b291-b6aafce9533a>
- Sales collection and overall business profile :
https://grok.com/share/c2hhcmQtNA%3D%3D_c5e1f5f4-013b-4edb-a04c-41034476c04c
- Analytical requirements example: <https://chatgpt.com/share/68b1b447-7670-8011-b54a-aa432c9e7d71>

Explanation of validations

- **Executive Summary:** AI was used to draft and refine the initial executive summary, ensuring it was professional, concise, and highlighted the core project objectives.
- **Requirements Definition:** AI was instrumental in structuring the requirements document, providing a logical flow for the business requirements and separating them into distinct sections (e.g., stakeholders, analytical requirements).
- **Conceptual NoSQL Design:** AI synthesized the individual conceptual designs into a single, cohesive database design. This involved identifying commonalities and ensuring the combined model was logical and optimized for MongoDB. AI also provided the example JSON code snippets for each collection to illustrate the final design.
- **Documentation Refinement:** AI reviewed and enhanced the language, formatting, and clarity of the entire project documentation, including sections like "High-Level Business Objectives" and "Preliminary Success Criteria," making the final document more professional and academically sound.

BIBLIOGRAPHY/REFERENCES

1. Chodorow, K. (2013). MongoDB: The Definitive Guide. O'Reilly Media.

This classic guide provides foundational knowledge on MongoDB's architecture and schema design, which is highly relevant to this project's data modelling approach.

2. Redmond, E. & Wilson, J. (2012). Seven Databases in Seven Weeks: A Guide to Modern Databases and the NoSQL Movement. The Pragmatic Bookshelf.

This book offers a broad perspective on the benefits of NoSQL databases and their use cases, providing theoretical support for the project's shift away from a traditional relational model.

3. Sadalage, M. & Fowler, M. (2012). NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Addison-Wesley Professional.

A key resource for understanding the core principles behind NoSQL databases, including document-based modelling, which is the cornerstone of the conceptual design presented here.

4. 11 Requirements Gathering Techniques for Agile Product Teams :

<https://www.jamasoftware.com/requirements-management-guide/requirements-gathering-and-management-processes/11-requirements-gathering-techniques-for-agile-product-teams/>