Project: Data Analysis in Mint Classic Company Database

About this Project

In this project, I will perform exploratory data analysis for Mint Classics Company, a retailer of model cars. The company is looking to close one of its storage facilities. The objective is to recommend inventory reduction strategies that will not negatively impact customer service.

Database overview:

The database is a well-structured relational database the combines all operations of Mint Classics company. The database consists of multiple tables that are interrelated and represents entire operations of Mint classics Company. Each table contains specific information about the company operations and it is related with one or more tables using distinct Keys (Primary & foreign Keys).

The database schema is as follows:

Table Name	Containing Columns	Information
products	productCode (Int) PK, productName (Varchar), productLine (Varchar), quantityInStock (Int), buyPrice (Decimal (10,2)), MSRP (Decimal (10,2))	Related to available products and their supply line, how much quantity is there in stock, buying and MSRP prices

Table Name	Containing Columns	Information
productlines	productLine (varchar (50)) PK, textDescription (varchar (4500)), htmlDescription (mediumtext), image (mediumblob)	Explains each product line producing products with link and images

Table Name	Containing Columns	Information
warehouses	warehouseCode (varchar (10)) PK, warehouseName (varchar (45)), warehousePctCap (varchar (50))	Contain information about warehouse code, names and how much their storage is full

Table Name	Containing Columns	Information
orders	orderNumber (int) PK, orderDate (date), requiredDate (date), shippedDate (date), status (varchar (20)), comments (text), customerNumber (int) FK	Stores information about orders, dates of orders, shipment and status of shipment and customer Numbers

Table Name	Containing Columns	Information
orderdetails	orderNumber (int) PK, productCode (varchar (20)) PK, quantityOrdered (int), priceEach (decimal (10,2)), orderLineNumber (smallint)	Stores info about orders, products, quantity ordered and price of products purchased and orderline

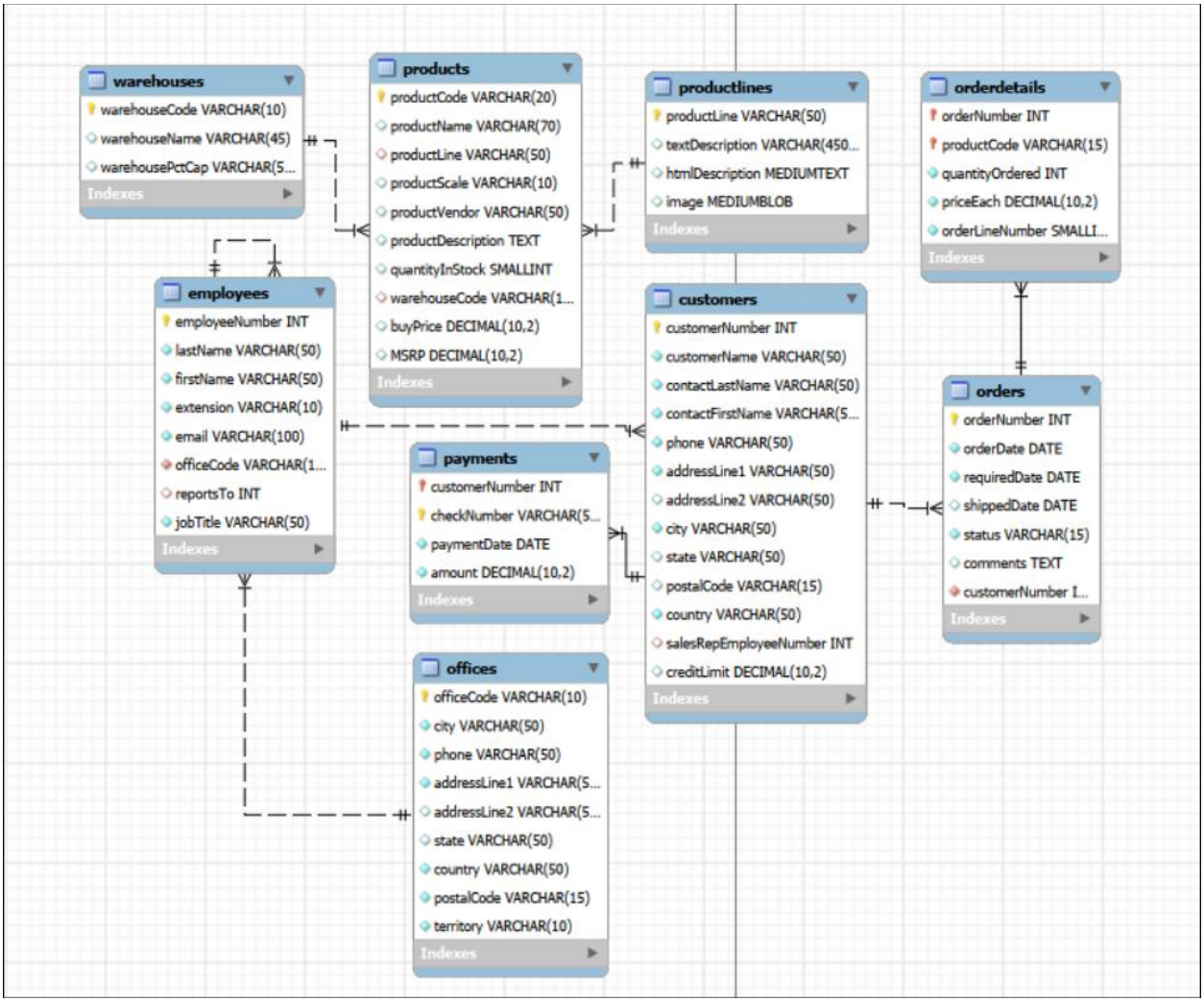
Table Name	Containing Columns	Information
customers	customerNumber (int) PK, customerName (varchar(50)), contactLastName (varchar(50)), contactFirstName (varchar(50)), phone (varchar(50)), addressLine1 (varchar(50)), addressLine2 (varchar(50)), city (varchar(50)), state (varchar(50)), postalCode (varchar(50)), country (varchar(50)), salesRepEmployeeNumber (int) FK, creditLimit (decimal(10,2))	Stores information about customers and their contact information and geographic distribution, credit limits, and their assigned sales rep

Table Name	Containing Columns	Information
payments	customerNumber (int) FK, checkNumber (varchar(50)) PK, paymentDate (date), amount (decimal(10,2))	Stores info related to payments amount, date of payment and the customer number who made the payment and details of check

Table Name	Containing Columns	Information
employees	employeeNumber (int) PK, lastName (varchar(50)), firstName (varchar(50)), extension (varchar(50)), email (varchar(100)), officeCode (varchar(10)), reportsTo (int), jobTitle (varchar(50))	Store employee names, contact info, reporting person details and their designation in company

Table Name	Containing Columns	Information
offices	officeCode (varchar(50)) PK, city (varchar(50)), phone (varchar(50)), addressLine1 (varchar(50)), addressLine2 (varchar(50)), state (varchar(50)), country (varchar(50)), postalCode (varchar(15)), territory (varchar(10))	Store office codes, contact info and geographic distribution

Database Model ERR (made using MySQL workbench)



Methodology:

This data analysis was done using structured, clustered approach by creating clusters based on the core functionality of the Mint Classic Company. The database contains nine related table datasets describing the overall operations of the company. Using these tables and their stored information and key relation with other tables, I have identified 3 cluster views. These views created all the information related to each core functionality of company i.e., customer, products (Inventory), employees. List of cluster views created mentioned below along with their table sources.

Cluster view	Tables used in creating views
customers_cluster_view	Customers, orders, orderdetails, payments, employees, products
employees_cluster_view	Employees, offices, customers
product_cluster_view	Products, warehouses, productlines, orders, orderdetails

DATA CLEANING:

- **COALESCE** function was used, wherever required, while creating clusters to get clear data without 'null' values
    - In customers cluster, **COALESCE** was used to get customers without orders values as '0' instead of null for accurate analysis of customer orders and sale value
    - In products cluster, **COALESCE** was used to get products without orders values as '0' instead of null for accurate analysis of products orders and sale value
    - In employee cluster, **COALESCE** was used to get employees without assigned customers values as '0' instead of null for accurate analysis of employee assigned customers and sale value

DATA ANALYSIS AND INTERPRETATION (Cluster Wise)

A. product_cluster_queries:

1. Inventory overview

a. Total products – A total of 110 products were produced across all productlines.

b. Total Stock – A total of 5,55,131 stock units were there across all productlines.

c. ProductLine Assignment – All products were assigned across 7 product lines and there is no single product without a product line.

d. Warehouse Storage - All products were stored across 4 warehouses and there is no single product stored without a warehouse.

e. Unsold products: The product '1985 Toyota Supra' with code 'S18_3233' with 7,733 stock units is being unsold by company and is stored at East Warehouse (code-b)

f. Timerange of product inventory storage: The data time range is in between '2003-01-06' and '2005-05-31'.

2. Warehouse stock and space utilization:

- Warehouse 'East (b)' has more stocks (219183) stored with 67% storage filled.
- Warehouse 'South (d)' has less stock stored (79380) with highest filled percent capacity and low left-over storage space
- Among all warehouses, warehouse 'West (c)' has more left-over storage space and low stock stored compared to a & b warehouses

3. Product line wise orders and revenue

- Productline classic Cars has more orders (33,817) and more ordervalue ($ 36,70,560) followed by vintage cars and motor cycles
- Productline Trains has lowest orders (2651) and more ordervalue ($ 1 75,030) followed by vintage cars and motor cycles
- Eventhough warehouse c has only one productline and product i.e., vintage cars, the orders and order value is second in poisition.

4. Inventory Turnover Rates

A. Product Turnover Rate

- the products with highest turnover ratio are '1960 BSA Gold Star DBD34 (68%)','1968 Ford Mustang (13%)','1928 Ford Phaeton Deluxe (6%) ','1997 BMW F650 ST' (5%)
- the higher turnover ratio indicates higher demand for those products
- 89 products are sitting idle without any orders and wasting warehouse space.

B. Warehouse Turnover Rate

- Product line motorcycles have highest turnover rate (77%) indicating that motorcycles have highest customer base
- Lowest turnover rate is from ships, truck & buses and trains product line, which belongs to warehouse d

5. Products for discontinuation:

Products with low sales, low demand and high stock were suitable for discontinuation

- Five products for discontinuation:
    1. '1966 Shelby Cobra 427 S/C'
    2. '1939 Chevrolet Deluxe Coupe'
    3. '1982 Lamborghini Diablo',
    4.'1982 Ducati 996 R'
    5.'1950's Chicago Surface Lines Streetcar'

6. Products for stock increase:

Products with high sales, high demand and low stock were suitable for discontinuation

- Five products for stock increase:
    1. '1968 Ford Mustang'
    2.'1962 Volkswagen Microbus'
    3.'1958 Setra Bus'
    4. '1969 Ford Falcon'
    5.'1957 Corvette Convertible'

7. Outliers detection

Outliers were identified by calculating interquartile range (IQR) for stock, orders and sales and products falling outside these ranges are outliers and need attention.

- 4 outliers were identified

    1. '1985 Toyota Supra' - for zero orders and zero sales

    2. '1957 Ford Thunderbird' - for very small orders

    3. '2001 Ferrari Enzo' - for very large total Sales

    4. '1992 Ferrari 360 Spider red' - for very large total orders and total Sales

Insights Summary for Product Management:

1. High ordered, sales and low stock items identified should be increased for company revenue growth

2. Low ordered, sales and low stock items identified should be discontinued for replacing with high value goods

B. customers_cluster_queries:

1. Customers overview;

- Total number of customers is 122

- 98 customers are with active orders

- 24 customers are with no active orders

2. Top customers with high payments

customerNumber	customerName	totalPaid
141	Euro+ Shopping Channel	715738.98
124	Mini Gifts Distributors Ltd.	584188.24
114	Australian Collectors, Co.	180585.07
151	Muscle Machine Inc	177913.95
148	Dragon Souveniers, Ltd.	156251.03
323	Down Under Souveniers, Inc	154622.08
187	AV Stores, Co.	148410.09
276	Anna's Decorations, Ltd	137034.22
321	Corporate Gift Ideas Co.	132340.78
146	Saveley & Henriot, Co.	130305.35

- These customers should be given priority and special attention for company growth and revenue

2. Top customers with lowest payments (Exclusing Non-Active customers)

customerNumber	customerName	totalPaid
219	Boards & Toys Co.	7918.60
198	Auto-Moto Classics Inc.	21554.26
103	Atelier graphique	22314.36
473	Frau da Collezione	25358.32
381	Royale Belge	29217.18
456	Microscale Inc.	29230.43
489	Double Decker Gift Stores, Ltd	29586.15
415	Bavarian Collectables Imports, Co.	31310.09
173	Cambridge Collectables Co.	32198.69
362	Gifts4AllAges.com	33533.47

- These customers need to be studied and analyzed for getting better orders from them and motivate them to order more from the company there by increasing company growth

3. Customers ordering pattern and retention by company

a. Ordering Pattern:

- Customer 'Euro+ Shopping Channel' has highest orders (24 No)

- Customer ' Bavarian Collectables Imports, Co.' has lowest orders (1 No)

b. Customers retention by company:

- Compared to 2005 most of the customers are retained in 2004 from 2003

- The drop in retaining customers from 2004 to 2005 need to be studied

4. Geographic and Credit distribution

a. Country wise total payments

- country USA highest total payments (' $ 3040029.52') followed by UK, Switzerland

- This suggests that, USA is company's biggest market

- Countries like Israel, Netherland, Poland, Portugal, Russia and South Africa were not yet started their business with the company.

- In order to get orders from the above non active countries, customer consumption patters must be studies with relevance to company products .

b. Total customers from each country

- Most of the customers were from USA (36) followed by Germany (13) and France (12)

- Hence, the customers from the USA be prioritized for increasing orders.

c. Total payments by Credit limit range

- most of the payments were done by customers with high credit limits

- this indicates that with increasing credit limit, customers are interested to buy more from the company.

5. Calculating sales representatives' performance over the ordering period

a. Total sales made by company during the period : $ 88,53,839.23/-

b. Distribution of sales by sales reps:

- Employees 'Gerard Hernandez' and 'Leslie Jennings' made highest sales ($ '1112003.81' and $ '989906.55')

- Employee 'Leslie Thompson' made lowest sales ($ '347533.03').

- Employees with lowest sales need training and improvement.

6. Shipping and Orders trend

a. The average delivery time: 3.7 Days

b. Year wise Orders traffic:

- Highest orders were placed in year 2004, followed by 2003 and least orders were in 2005.

c. Order status distribution across year and month

- Year end orders were lot of completed orders due to holiday season.

- 2nd and 3rd quartile orders were mostly cancelled.

7. Payment's trend

- In 2003 and 2004, highest payments were made during November and December indicating that, holidays are most important months to be covered.

- increasing stock and putting up new discounts and offers during theses seasons will help in getting maximum sales for the company.

Summary of recommendations using customers cluster analysis:

- High value customers should be given priority, personalized services and loyalty rewards for continuing their relations with the company.
- Expanding into low sales Regions with different product lines may increase the orders and customer base.
- Sales representatives with low sales need to be motivated and trained in selling efficiently.
- Shipping efficiency should be increased by decreasing the number of days to delivery from 3.7 days to 1 day.
- Significant sales were observed during holiday seasons so we should increase the stock and put up price discounts and offers to increase total sales and tap the season

C. Employee cluster analysis

1.. Employee overview

- **Total employees:** Mint Classics company has a total of 108 employees
- **Job titles:** A total of seven designations were found (President, VP sales, VP marketing, sales manager (APAC), sales manager (EMEA), Sales manager (NA), sales representatives)
- **Employees distribution:** Most of the employees were sales representatives Indicating the company mostly invested in selling their products.

2. Employees without assigned customers

- Eight employees were without customers among which six are in manager or related reporting designation

- The remaining employees without customers assigned were only two Namely Tom King Yoshimi Kato, which indicates that these employees are either New To the organization or There has been a recent change of responsibilities for them.

3. Distribution of employees across offices and Location:

a. Employee distribution by office: All employees were Assigned with an office and there is no employee without an office.

b. Employees distribution by country:

- The United has a greater number of employees in minty classics company with 43 number followed by France and Japan

- This indicating the high potential United States customer base which needs more employees for providing services.

c. Customers managed by offices:

- Office located in Paris of country France has more customers followed by London Office in UK.
- By combining three offices that is NYC, San Francisco Boston in country USA a total of 39 customers were there
- Office in Tokyo of Japan has lowest customer base that is 5, indicating more market penetration and more product lines distribution

4. Managerial structure and reports

a. Top manager with number of reporters:

- Manager Gerard Bondur, sales manager for EMEA region has highest number repartees (46) is followed by Anthony Bow (39)

- The higher number of reportees is for Gerard Bondur indicates that he is very efficient in managing sales representatives and is work efficiency is also high compared to others.

b. Customers managed by sales representatives

- Among the employees Pamela Castillo Manages more customers that is 9, indicating this superior efficiency and skills in sales.

- On the other hand, Employees Andy Fixster, Peter Marsh Mami Nishi are managing lower customers that is 5 Indicating improvement in customer relationships and efficiency

Summary of Insights using employee cluster analysis:

- Employees with no assigned customers need to be assigned with customers for decreasing customer load for others
- Increasing the workforce in low employee regions like Sydney and Tokyo helps business growth by creating new customer base.
- Managers like Grad Bundur with higher number of report needs to be supported with all resources and autonomy for strengthening the company feature
- Employees with low customer base needs to be trained in customer relationship and improve efficiencies

Inter cluster analysis

This intercluster analysis involves integrating data from three clusters customers Products and employ by doing so this analysis offers a holistic way of mental classics operation

1. Distribution of product revenue across sales representatives

- Employee Gerard Hernandez made highest sales ($ 132 197 88. 60) Followed by Leslie Jennings ($ 12765292. 92) Indicating higher bonus for these employees and promotion also
- Employee like Leslie Thompson Julie Ferrelli and Martin Gerard made lowest sales compared to others indicating these employees need to be trained and provide with proper resources

2. Region wise warehouse utilization

- Warehouse East (code b) has more customer base (94)and more stocks ordered (3,97,356)
- Warehouse South (code d) Has lower stocks added Indicating Lower customer preferred products.
- Hence the stock from the warehouse South needs to be replaced with more customer preferred products for increasing sales

3. Region wise popular products

- Regions like France, UK, USA and Australia showed more interest in classic cars, vintage cars Product line
- Trains and ships product lines are less interested for most of regions

4. Customer retention by product line

- Product line classic cars Vintage cars has higher customer retention followed by motorcycle and planes
- Lowest customer retention is by the product line trains which need to be replaced with more customer-based product lines for increasing company's profits and sales

5. Non completed and disputed orders

- Orders from customers with low credit limit has been on hold due to
- The main reason for disputed Status in orders was due to damaged product Or not meeting the customers' expectations

- The main reasons for order cancelations is either due to mistakenly placed order or sometimes customer found better offer

6. Sales representatives and their customer relationships

- Employ Pamela Casillo has more customers Followed by Barry Jones And some others
- Lowest customer base was found with employees Andy fixture, Peter Marsh, Mami Nishi and Martin Gerard.

Addressing the original business problem : warehouse closure

After analyzing products, customers, employees' clusters, here are some strategic recommendations to address the business problem in concern i.e., closure of a warehouse.

1. Warehouse Closure:

Closing warehouse South (D) is optimal decision due to its low orders, high stock, and moderate value products.

2. Inventory Management

Low demand or idle stock like 1985 Toyota Supra, 1966 Shelby Cobra 427/c should be replace with high demand fast selling products like 1968 Ford Mustang and 1960 Gold star DBD34.

3. Warehouse storage optimization:

The warehouse b stores more products compared to others. Hence, stock from this warehouse should be distributed to low traffic warehouse like South (b) for optimum space utilization.

4. Customers retention strategy

- Special focus should be given to high values customers and countries with most customers for better customer experience

- We must find better ways to sell product in regions with low customer base.

5. Sales representatives training and improvement

- Incentives and bonus should be given to high performing sales rep with higher customer base and high sale value.

- Trainings and performance improvement programmes should be arranged for low efficiency sales reps.