RELATIONAL DATABASE PROJECT

INVENTORY CONTROL SYSTEM FOR E-COMMERCE

Kieu Van Dang Email: dang.v@northeastern.edu

LinkedIn: https://www.linkedin.com/in/kirudang/ Phone: +1 647 782 4558

Data science enthusiast with a great passion for data pre-processing and prediction with a strong background in business. Relevant skills include machine learning, statistics, problem-solving, programming (including SQL, Python, and R), and critical & creative thinking.

INDEX

A. Project Introduction

1. Overall business context

Given the context of the relentless development of technology nowadays and the huge transactions of e-commerce in recent times due to the Covid-19 outbreak, Canada e-commerce surged up to 75% last year in 2020. Nevertheless, the penetration of online retail of Canada, for a country with 80% living in the urban areas, was still modest at 12% of total retail. This figure is relatively humble compared to its counterparts in Asia where e-commerce penetration is roughly greater than 30%.

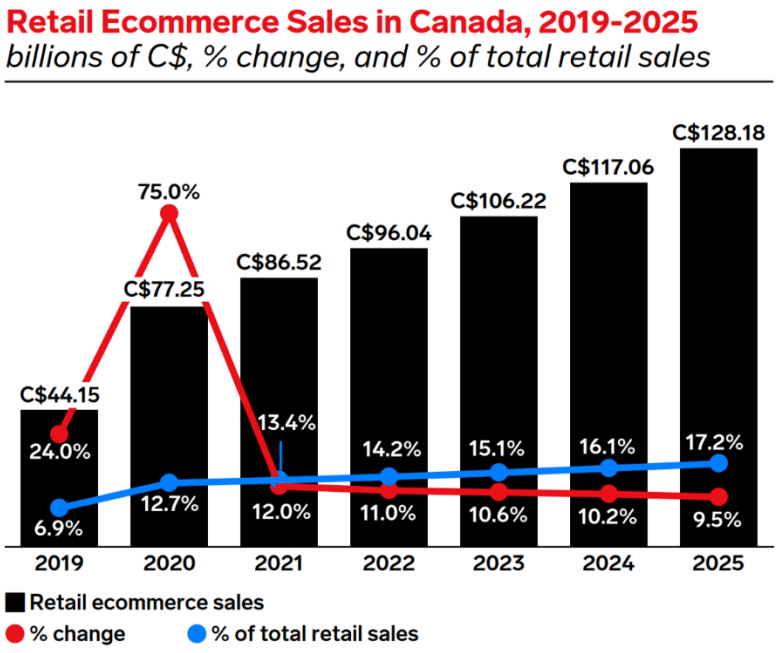


Figure 1: *E-commerce sales in Canada, 2019 - 2025*

According to ITBusiness Canada, one of the reasons for Canada's lagging e-commerce sector is that the country has invested far less in technologies than our large foreign competitors. From my perspective, this slow development mainly comes from unevenly supported logistic infrastructures and delivery issues. Since I have been working in the retail industry for the past 6 years and I am currently pursuing a master's in Data Analytics to transform myself into a Data Scientist in such an industry, these concerns have become my keen interest. For that reason, I come up with a simple **Inventory Control System** in the e-

commerce sector to see how this industry has coped with such barriers. Hopefully, I can use this project for my further career development in the retail domain.

2. Use-case scenario and the application solution

Inventory control is one of the most challenging jobs in the industry, especially for online retailers. This system allows many related stakeholders within the company to get all of the information they need in one go. The details related to the sales along with stocks can be got in just one stretch through this application. In the real-world setting, Inventory Control System will help the category employees to decide which levels of stock they should keep to maximize profitability. The system also facilitates the warehouse workers to manage the stock well and inform promptly to related parties such as purchasing department for goods influx by new purchases from suppliers.

3. Relevant data sources

This project began as a result of my initial curiosity about the e-commerce industry without a publicly available dataset. As a result, I have gathered information from various internet sources to create a fictitiously unified dataset that could serve the purpose of this system. All data sources will be cited in the bibliography at the end of this report.

B. Database System Analysis and Design

1. User Profiles

Inventory management systems simplify and consolidate the process of controlling stock drift and maintaining stock levels to ensure that the right amount of stock is available at the right time. Therefore, it will benefit the following stakeholders within various company departments:

Category management is one of the most frequent users of this system. Generally, they will refer to four entities in the table which are Item_Orders, Order_Addresses, Orders, and Item_Availabilities. By looking thoroughly at these metrics, category managers could infer the insights for categories that they manage. *The first observation is goods movement* of a specific product or variant within a category,

through order quantity per a fixed period compared to that of other items. There are two significant scenarios in this insight, which are slow product movement and overselling products. As per low selling items, category management may create a promotion to clear spoilage or dead stocks and then notify to purchasing department to reduce the stock level for product suppliers. Conversely, overselling is a common issue in e-commerce, especially in mega promotion campaigns and festival seasons such as Single day 11.11, Black Friday, Boxing Day, and so on. Overselling will harm the company's reputation and overall e-commerce segment because buyers must wait too long for an ordered item, leaving a bad imprint in their mindset about e-commerce operation as well as inflicting negative sentiment towards the company. *The second advancement is buyer demographic* resulting from the order address table. With this table, category employees and managerial levels could depict where the buyers come from, what they buy in that region, and further descriptive as well as inferential analytics. *The third importance is time series analysis* from the information of order table, in which we have data from year, month and date. Associating with data retrieved from other tables, we could draw several key findings about time series analysis. *Another advantage is the operational analysis* such as order level or cancelation rate. Computing order status to see order completion rate or checking stocks on hand to see if they are below the set minimum level or not could be telling examples in this case.

Warehouse top-level and employees are the next beneficial parties of the system. The major application of warehouse employees is *warehouse cost savings* by overseeing Item_Availabilities and Warehouses entity. In e-commerce, apart from e-commerce marketplaces, in which products will be shipped directly from suppliers' warehouses, e-commerce retailers operate in a different way. They always have their own warehouses located in cities' suburbs to meet both inventory storage and fastest delivery. Every retailer also possesses various warehouses across the nation. Therefore, a well-designed inventory system will cut costs in storage by improving inventory levels and making better use of storeroom space

by answering when to order, which products are out of stock or where locations of warehouse should be optimal.

Furthermore, there are several user profiles for the system such as **purchasing employees or company management levels**. Combination of products and stock availabilities will help purchasing staff gain power in business negotiation with suppliers meanwhile understanding supply and demand principles will facilitate top management in business decisions. These business profiles will rely more on major entities of this system such as Orders, Warehouses or Suppliers.

2. General business rules

Before drawing the Entity Relationship Diagram (ERD), all business rules should be clearly regulated to target for a smooth system. Here are several rules to shape the system and associate with profile users above.

- Shoppers could order products.
- Shoppers could cancel the order in which products are not delivered yet.
- One order could contain many products, which are from different warehouses.
- Shoppers could hold products in their cart by keeping for one day without check out, after that stocks will be released if there is no ongoing transaction.
- Shoppers could choose normal shipping or express shipping.
- Shopper can store maximum 4 different addresses in their account.
- One order has many statuses such as Unpaid, Confirmed, Shipped, Delivered, Returning, Returned.
- Sales is only recorded when status turns “Delivered”.
- Deliver will be processed from the nearest warehouse, which is enough stocks, and primarily based on the shipping address of the order.
- Shoppers could return products within 15 days if they are not satisfied with the products.
- The product must return from the shipping address only.

- Category Manager could set minimum and maximum stocks for the warehouse.
- Category, purchasing employees could access the level of stock for each product.
- Purchasing employee could see order address summary.

3. Database Design and Entity Relationship Diagram

Within the scope of this project, the total entities of the system are 10, ranging from SHOPPERS to ORDERS as well as WAREHOUSES and SUPPLIERS. Four primary entities include **ORDERS**, **SHOPPERS**, **PRODUCTS**, and **WAREHOUSES**, which carry fundamental information for system baseline and meet the business queries from our major user profiles as stated above. There are two association entities inferred from these four foundation entities: **ITEM_ORDERS**, which provides the outflow of goods, and **ITEM_AVAILABILITIES**, which refers to stock on hand. The remaining entities are created to meet the Third Normal Form of data optimization such as **PRODUCT_CATEGORIES**, **SHIPPERS** and **SUPPLIERS**. Relationships among and between entities are stated in Figure 2.

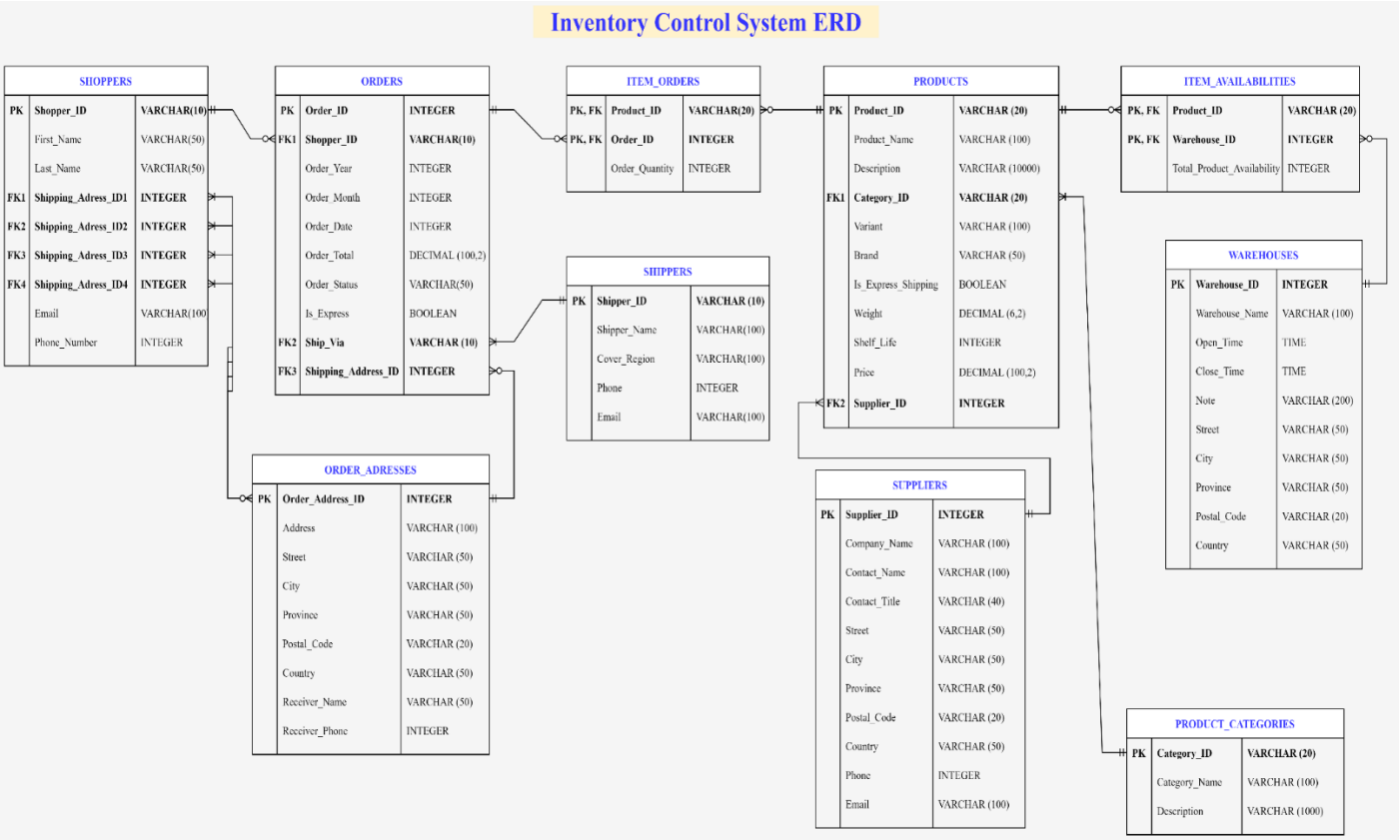


Figure 2: *Inventory Control System ERD*

C. Database Implementation

In real-world settings, many use cases may be reflected in the data queries. Since there are three major user profiles of this system, it could benefit them by addressing the following business concerns. My business queries for this system will be showcased from basic to advanced levels, delivering general requests from a real business context. Please note that for display purpose, rows of results are limited to 20 observations only.

For Category management, they want to produce more insights about business performance by evaluating product movement, customer demographic, the effectiveness of delivery, or the best time stone to sell products via the platform.

1. *Movements of products with their general information by quantity. The output should be sorted in descending order and only display the first and last 5 results to see which one is the best-selling and which one has bad performance among the assortments.*

```
WITH Sales_Table AS
(SELECT *, RANK () OVER (ORDER BY Total_Quantity DESC) Sales_Rank FROM
(SELECT p.Product_Name, SUM (Order_Quantity) AS Total_Quantity FROM ITEM_ORDERS i
NATURAL JOIN PRODUCTS p GROUP BY i.Product_ID ORDER BY Total_Quantity))

SELECT * FROM Sales_Table WHERE Sales_Rank <= 5
UNION
SELECT * FROM Sales_Table WHERE Sales_Rank >= ((SELECT MAX(Sales_Rank) FROM Sales_Table)-5)
ORDER BY Total_Quantity DESC;
```

11 records

Product_Name	Total_Quantity	Sales_Rank
Vintage Parts 14558 3" Green Fuzzy Dice with White Dots - Pair	305	1
Lionel Racing, Paul Menard, Motorcraft, 2019, Ford Mustang, NASCAR Diecast 1: 64 Scale	187	2
A&R Stryker Waxed Hockey Skate Laces	178	3
Funko Pop Keychain: Toy Story - Woody	178	3
Funny Sticker World Alpaca and Sheep Stickers	178	3
Harry Potter Novelty Magnet, Multi Color	114	104
Jumbo Oversize Playing Cards 4.5"x7" by Midway Monsters	114	104

2. *a. Movements of products by category by total amount with their general information. The output should be sorted in descending order to see which one is top 3 best-selling items in each category in terms of revenue. Revenue should be counted only status is delivered.*

```
WITH TrackSales AS (
SELECT p.Product_Name, pc.Category_ID, pc.Category_Name, SUM(i.Order_Quantity*p.Price) AS Revenue FROM ITEM_ORDER
S i NATURAL JOIN PRODUCTS p NATURAL JOIN ORDERS o NATURAL JOIN PRODUCT_CATEGORIES pc WHERE o.Order_Status == "Del
ivered" GROUP BY i.Product_ID)

SELECT * FROM (SELECT *, RANK() OVER ( PARTITION BY Category_ID ORDER BY Revenue DESC) Sales_Rank FROM TrackSale
s) WHERE Sales_Rank <= 3 ORDER BY Category_ID, Sales_Rank ASC LIMIT 20;
```

20 records

Product_Name	Category_ID	Category_Name	Revenue	Sales_Rank
Rubies Black Eyelashes and Adhesive	BPC	Beauty & Personal Care	822.25	1
Disney Mickey Mouse Ice Cream D-Lish Treats Phone Charm	CPA	Cell Phones & Accessories	899.25	1
All About Details CAT80YL 80-Years-Loved Cake Topper,1PC, 80th Birthday, Party Decor, Glitter Gold (Gold & Black), 6 x 8,	GGF	Grocery & Gourmet Food	2242.68	1
amscan Gold Plastic Forks | Party Supply | 240 ct.	HH	Health & Household	2578.50	1
Team Associated 25202 M3 x 10mm Flat Head Hex Screw	HO	Hobbies	687.70	1
Educational Insights Fluorescent Light Filters (Whisper White), Set of 4	OP	Office Products	3478.14	1
REALBUG Dragonfly Paperweight(3x3x1)	OP	Office Products	2278.80	2

b. For the top 3 products by revenue above, the company wants to discount selling prices for them by 10% for those ranking 1, 5% for those ranking 2 and 3% for those ranking 3 by each category, the rest products will keep the same price for next year. Price will be rounded up to 2 decimal numbers. Display only products with price change. Revenue should be counted only status is delivered.

```
WITH TrackSales AS (
SELECT p.Product_Name, p.Price, pc.Category_ID, pc.Category_Name,  SUM(i.Order_Quantity*p.Price) AS Revenue FROM
 ITEM_ORDERS i
NATURAL JOIN PRODUCTS p NATURAL JOIN ORDERS o NATURAL JOIN PRODUCT_CATEGORIES pc
WHERE o.Order_Status == "Delivered"
GROUP BY i.Product_ID),
Top3 AS (SELECT * FROM (SELECT *, RANK() OVER ( PARTITION BY Category_ID ORDER BY Revenue DESC) Sales_Rank FROM T
rackSales) WHERE Sales_Rank <= 3 ORDER BY Category_ID, Sales_Rank ASC)

SELECT *, CASE WHEN Sales_Rank =1 THEN ROUND(Price*0.9,2)
          WHEN Sales_Rank =2 THEN ROUND(Price*0.95,2) ELSE ROUND(Price*0.97,2)
END AS New_Price FROM Top3 LIMIT 20;
```

20 records

Product_Name	Price	Category_ID	Category_Name	Revenue	Sales_Rank	New_Price
Rubies Black Eyelashes and Adhesive	7.15	BPC	Beauty & Personal Care	822.25	1	6.44
Disney Mickey Mouse Ice Cream D-Lish Treats Phone Charm	8.25	CPA	Cell Phones & Accessories	899.25	1	7.43
All About Details CAT80YL 80-Years-Loved Cake Topper,1PC, 80th Birthday, Party Decor, Glitter Gold (Gold & Black), 6 x 8,	16.99	GGF	Grocery & Gourmet Food	2242.68	1	15.29
amscan Gold Plastic Forks | Party Supply | 240 ct.	34.38	HH	Health & Household	2578.50	1	30.94
Team Associated 25202 M3 x 10mm Flat Head Hex Screw	5.98	HO	Hobbies	687.70	1	5.38
Educational Insights Fluorescent Light Filters (Whisper White), Set of 4	30.78	OP	Office Products	3478.14	1	27.70
REALBUG Dragonfly Paperweight(3x3x1)	18.99	OP	Office Products	2278.80	2	18.04
Pacon Sketch Pads and Drawing Paper (PAC4712)	20.22	OP	Office Products	1981.56	3	19.61

3. *Customer demographic: Which countries have the most orders from the site?*

```
SELECT od.Country, COUNT(o.Order_ID) AS OrderCount, ROUND(COUNT(od.Country)*100/(SELECT COUNT(*) FROM ORDERS),2)
AS Percentage FROM ORDERS o
INNER JOIN ORDER_ADDRESSES od On o.Shipping_Address_ID = od.Order_Address_ID
GROUP BY od.Country ORDER BY OrderCount DESC LIMIT 20;
```

19 records

Country	OrderCount	Percentage
USA	97	37
France	30	11
Spain	27	10
Australia	15	5

4. *Customer demographic: Detail to location of province of for top 1 country having the most orders.*

```
SELECT od.Country, od.Province, COUNT(o.Order_ID) AS OrderCount, ROUND(COUNT(od.Province)*100/t.OrderCount,2) AS
 Percentage FROM ORDERS o
INNER JOIN ORDER_ADDRESSES od On o.Shipping_Address_ID = od.Order_Address_ID
INNER JOIN (
              SELECT od.Country, COUNT(o.Order_ID) AS OrderCount FROM ORDERS o
              INNER JOIN ORDER_ADDRESSES od On o.Shipping_Address_ID = od.Order_Address_ID
              GROUP BY od.Country ORDER BY OrderCount DESC LIMIT 1 ) t
ON od.Country = t.Country
GROUP BY od.Country, od.Province ORDER BY OrderCount DESC;
```

Country	Province	OrderCount	Percentage
USA	CA	35	36
USA	MA	21	21
USA	NY	18	18
USA	PA	8	8

5. *Operational analysis: The proportion of Order status? This is to know if the cancellation rate is less than 0.5% compared to the average of the industry.*

```
SELECT Order_Status, COUNT(Order_Status) AS StatusCount,
ROUND(COUNT(Order_Status)*100/(SELECT COUNT(*) FROM ORDERS),2) AS Proportion FROM ORDERS
GROUP BY Order_Status ORDER BY StatusCount DESC;
```

5 records

Order_Status	StatusCount	Proportion
Delivered	242	93
Cancelled	9	3
Returned	6	2
Shipped	2	0
On Hold	1	0

6. *a. Time series analysis: Which month of the year and which date of that month the selling peaks at?*

```
SELECT o.Order_Month, o.Order_Date, COUNT(o.Order_ID) AS "Date_Order", SUM(o.Order_Total) AS "Revenue_Per_Date" F
ROM ORDERS o
WHERE o.Order_Month == (SELECT temp.Order_Month FROM (SELECT o.Order_Month, SUM(o.Order_Total) AS "Monthly_Revenu
e" FROM ORDERS o GROUP BY o.Order_Month ORDER BY "Monthly_Revenue" DESC LIMIT 1) temp)
GROUP BY o.Order_Date ORDER BY "Revenue_Per_Date" DESC LIMIT 1;
```

1 records

Order_Month	Order_Date	Date_Order	Revenue_Per_Date
11	5	6	2506.43

b. Time series analysis: Which quarter of the year, the selling peaks at?

```
SELECT CASE
WHEN Order_Month <4 THEN "Q1" WHEN Order_Month >=4 AND Order_Month <7 THEN "Q2"
WHEN Order_Month >=7 AND Order_Month <10 THEN "Q3" ELSE "Q4" END AS "Quarter", COUNT(Order_ID) AS "Count", SUM(Or
der_Total) AS "Revenue_Per_Quarter"
FROM ORDERS GROUP BY "Quarter" ORDER BY "Revenue_Per_Quarter"  DESC;
```

4 records

Quarter	Count	Revenue_Per_Quarter
Q4	107	48408.80
Q3	51	22030.74
Q1	55	21948.08
Q2	47	15577.31

For warehouse top-level and employees, they also acquire findings for stock level, stock distribution to meet demands from commercial departments.

7. *Total product availabilities per item and its information: name and category, sorted by descending order. Also display stock ranking within category and within total available products.*

```
WITH Stock AS (SELECT p.Product_Name, pc.Category_ID, pc.Category_Name,
SUM(iv.Total_Product_Availability) AS InStock FROM PRODUCTS p NATURAL JOIN ITEM_AVAILABILITIES iv NATURAL JOIN PR
ODUCT_CATEGORIES pc
GROUP BY p.Category_ID, iv.Product_ID)

SELECT * , RANK() OVER (ORDER BY InStock DESC) Stock_Rank,
RANK() OVER (PARTITION BY Category_ID ORDER BY InStock DESC) Category_Rank
FROM Stock ORDER BY Category_ID LIMIT 20;
```

20 records

Product_Name	Category_ID	Category_Name	InStock	Stock_Rank	Category_Rank
Rubies Black Eyelashes and Adhesive	BPC	Beauty & Personal Care	12289	7	1
Disney Mickey Mouse Ice Cream D-Lish Treats Phone Charm	CPA	Cell Phones & Accessories	9020	86	1
All About Details CAT80YL 80-Years-Loved Cake Topper,1PC, 80th Birthday, Party Decor, Glitter Gold (Gold & Black), 6 x 8,	GGF	Grocery & Gourmet Food	8840	92	1
amscan Gold Plastic Forks | Party Supply | 240 ct.	HH	Health & Household	10160	48	1
Team Associated 25202 M3 x 10mm Flat Head Hex Screw	HO	Hobbies	9571	69	1
Educational Insights Fluorescent Light Filters (Whisper White), Set of 4	OP	Office Products	10747	31	1
REALBUG Dragonfly Paperweight(3x3x1)	OP	Office Products	10670	33	2
Pacon Sketch Pads and Drawing Paper (PAC4712)	OP	Office Products	9594	68	3

8. *a. For the top 5 best-selling product by quantity above, category manager sets minimum of 40 units per product per warehouse, is there any warehouses that violate this rule?*

```
SELECT w.Warehouse_ID, w.Warehouse_Name, p.Product_Name, SUM(iv.Total_Product_Availability) AS InStock FROM WAREH
OUSES w NATURAL JOIN ITEM_AVAILABILITIES iv NATURAL JOIN PRODUCTS p
WHERE p.Product_ID IN (SELECT k1.Product_ID FROM
        (SELECT p.Product_ID, p.Product_Name, SUM(i.Order_Quantity) AS Total_Quantity
        FROM ITEM_ORDERS i NATURAL JOIN PRODUCTS p
        GROUP BY p.Product_ID
        ORDER BY Total_Quantity DESC
        LIMIT 5) k1)
GROUP BY w.Warehouse_ID, p.Product_ID HAVING InStock < 40 ORDER BY InStock DESC;
```

3 records

Warehouse_ID	Warehouse_Name	Product_Name	InStock
19	Henningsen Cold Storage Co.	Lionel Racing, Paul Menard, Motorcraft, 2019, Ford Mustang, NASCAR Diecast 1: 64 Scale	34
12	United Cold Storage	Vintage Parts 14558 3" Green Fuzzy Dice with White Dots - Pair	32
16	Flagship Logistics Group	Funny Sticker World Alpaca and Sheep Stickers	5

b. For top 5 best-selling product by quantity above, category manager sets minimum 40 units per product per warehouse. If stock is < 40 units, then set status: Order Immediately; if stock from 40 to 100, status: Order Soon; otherwise set status: Safe Level. Display products in urgent level.

```
WITH Status AS (SELECT w.Warehouse_ID, w.Warehouse_Name, p.Product_Name, SUM(iv.Total_Product_Availability) AS In
Stock FROM WAREHOUSES w NATURAL JOIN ITEM_AVAILABILITIES iv NATURAL JOIN PRODUCTS p
WHERE p.Product_ID IN (SELECT k1.Product_ID FROM
        (SELECT p.Product_ID, p.Product_Name, SUM(i.Order_Quantity) AS Total_Quantity
        FROM ITEM_ORDERS i NATURAL JOIN PRODUCTS p
        GROUP BY p.Product_ID
        ORDER BY Total_Quantity DESC
        LIMIT 5) k1) GROUP BY w.Warehouse_ID, p.Product_ID)

SELECT * , CASE WHEN InStock < 40 THEN "Order Immerdiately"
WHEN InStock >= 40 AND InStock <=100 THEN "Order Soon" ELSE "Safe Level"
END AS Stock_Status FROM Status ORDER BY InStock ASC LIMIT 20;
```

Warehouse_ID	Warehouse_Name	Product_Name	InStock	Stock_Status
16	Flagship Logistics Group	Funny Sticker World Alpaca and Sheep Stickers	5	Order Immerdiately
12	United Cold Storage	Vintage Parts 14558 3" Green Fuzzy Dice with White Dots - Pair	32	Order Immerdiately
19	Henningsen Cold Storage Co.	Lionel Racing, Paul Menard, Motorcraft, 2019, Ford Mustang, NASCAR Diecast 1: 64 Scale	34	Order Immerdiately
13	H&M Bay, Inc.	A&R Stryker Waxed Hockey Skate Laces	54	Order Soon
3	Americold Logistics	Funny Sticker World Alpaca and Sheep Stickers	59	Order Soon

9. *Is there any warehouse, which has out-of-stock items? Display warehouse necessary information and the number of out-of-stock item it has.*

```
SELECT w.Warehouse_ID, w.Warehouse_Name, w.City, w.Country, COUNT (w.Warehouse_ID) AS "OOS_SKU" From WAREHOUSES w
 NATURAL JOIN ITEM_AVAILABILITIES iv
WHERE iv.Total_Product_Availability = 0 GROUP BY w.Warehouse_ID;
```

3 records

Warehouse_ID	Warehouse_Name	City	Country	OOS_SKU
9	Preferred Freezer Services	Hialeah Gardens	UNITED STATES	2
14	Subzero Logistics and Cold Storage Warehousing	St. Louis	UNITED STATES	1
18	United States Cold Storage, Inc.	Dallas	UNITED STATES	1

10. *Location of the warehouse with the most inventory, does that align with sales performance?*

```
SELECT DISTINCT w.Warehouse_ID, w.Warehouse_Name,w.Country, w.Province, SUM(iv.Total_Product_Availability) AS Tot
alInventory FROM WAREHOUSES w
NATURAL JOIN ITEM_AVAILABILITIES iv
GROUP BY w.Warehouse_ID ORDER BY TotalInventory DESC LIMIT 5;
```

5 records

Warehouse_ID	Warehouse_Name	Country	Province	TotalInventory
5	Lineage Logistics	UNITED STATES	CA	58529
18	United States Cold Storage, Inc.	UNITED STATES	TX	58488
20	North American Cold Storage, Inc.	UNITED STATES	IN	57941

For purchasing employees and top management, increasing supply chain power, and providing information for business decision are top queries from this system.

11. *Top 5 suppliers with the most products contributed?*

```
SELECT s.Company_Name, COUNT(p.Product_ID) AS "Total_Number"  From PRODUCTS p
INNER JOIN SUPPLIERS s On s.Supplier_ID = p.Supplier_ID
GROUP BY s.Company_Name ORDER BY "Total_Number" DESC LIMIT 5;
```

5 records

Company_Name	Total_Number
WALGREEN CO	129
WAL-MART STORES EAST, LP	44
WAL-MART STORES INC	23
LUXOTTICA RETAIL NORTH AMERICA INC	21
TARGET CORPORATION	18

Please note that for this query, I group by Company name instead of Company ID because one company have many different branches by address and each branch associate with a unique ID.

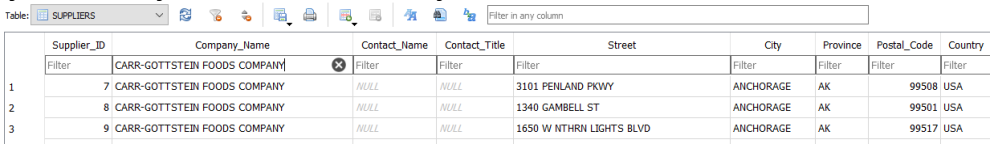


Table: SUPPLIERS

	Supplier_ID	Company_Name	Contact_Name	Contact_Title	Street	City	Province	Postal_Code	Country
	Filter	CARR-GOTTSTEIN FOODS COMPANY	Filter	Filter	Filter	Filter	Filter	Filter	Filter
1	7	CARR-GOTTSTEIN FOODS COMPANY	NULL	NULL	3101 PENLAND PKWY	ANCHORAGE	AK	99508	USA
2	8	CARR-GOTTSTEIN FOODS COMPANY	NULL	NULL	1340 GAMBELL ST	ANCHORAGE	AK	99501	USA
3	9	CARR-GOTTSTEIN FOODS COMPANY	NULL	NULL	1650 W NTHRN LIGHTS BLVD	ANCHORAGE	AK	99517	USA

12. *Top sold products belong to which suppliers?*

```
SELECT s.Supplier_ID, s.Company_Name, p.Product_Name, SUM(i.Order_Quantity) AS "Total_Order"  From ITEM_ORDERS i
NATURAL JOIN PRODUCTS p NATURAL JOIN SUPPLIERS s
GROUP BY s.Supplier_ID, p.Product_ID ORDER BY "Total_Order" DESC LIMIT 5;
```

5 records

Supplier_ID	Company_Name	Product_Name	Total_Order
18744	PUBLIX SUPER MARKETS INC	Vintage Parts 14558 3" Green Fuzzy Dice with White Dots - Pair	305
51861	TAMARKIN COMPANY	Lionel Racing, Paul Menard, Motorcraft, 2019, Ford Mustang, NASCAR Diecast 1: 64 Scale	187
25601	DANIEL C EBY D.O. PC	Funny Sticker World Alpaca and Sheep Stickers	178

13. *In each product category, show the top 3 products that have the most sales increase between the most recent year and the previous year. For each record, display category name, product name, brand, supplier name, supplier country, current year sales, last year sales, variance (%).*

```
WITH Sales AS (SELECT p.Product_Name, p.Category_ID, p.Supplier_ID, p.Brand, o.Order_Year,      i.Order_Quantity,
 p.Price, (i.Order_Quantity * p.Price) AS Sales
        FROM ITEM_ORDERS i NATURAL JOIN ORDERS o NATURAL JOIN PRODUCTS p
        WHERE o.Order_Year IN ((SELECT MAX(Order_Year) AS Current_Year FROM ORDERS),
        (SELECT MAX(Order_Year)-1 AS Last_Year FROM ORDERS))),
Sales_Sum AS (SELECT Product_Name, Category_ID, Supplier_ID, Brand,
    SUM(IIF(Order_Year = (SELECT MAX(Order_Year) AS Current_Year FROM Orders), Sales, 0)) AS     Current_Year_Sal
es,
        SUM(IIF(Order_Year = (SELECT MAX(Order_Year)-1 AS Last_Year FROM Orders), Sales, 0)) AS     Last_Year_Sal
es FROM Sales      GROUP BY Product_Name, Category_ID, Supplier_ID, Brand
        HAVING Last_Year_Sales <> 0),
Variance AS (SELECT     Product_Name, Category_ID, Supplier_ID, Brand, Current_Year_Sales,        Last_Year_Sale
s, ROUND(((Current_Year_Sales - Last_Year_Sales)/Last_Year_Sales)*100,2)      AS Sales_Variance, ROW_NUMBER() OVE
R (PARTITION BY Category_ID ORDER BY (Current_Year_Sales - Last_Year_Sales)/Last_Year_Sales DESC) AS Variance_Ran
k
        FROM Sales_Sum)

SELECT  c.Category_Name, v.Product_Name, s.Company_Name AS Supplier_Name, v.Current_Year_Sales, v.Last_Year_Sale
s, v.Sales_Variance
FROM Variance v NATURAL JOIN SUPPLIERS s NATURAL JOIN PRODUCT_CATEGORIES c
WHERE Variance_Rank <= 3 AND v.Sales_Variance > 0
ORDER BY c.Category_Name, v.Sales_Variance DESC;
```

8 records

Category_Name	Product_Name	Supplier_Name	Current_Year_Sales	Last_Year_Sales	Sales_Variance
Grocery & Gourmet Food	All About Details CAT80YL 80-Years-Loved Cake Topper,1PC, 80th Birthday, Party Decor, Glitter Gold (Gold & Black), 6 x 8,	WALGREEN CO	152.91	101.94	50.00
Office Products	REALBUG Dragonfly Paperweight(3x3x1)	CENTRACARE CLINIC	284.85	189.90	50.00
Pet Supplies	Rubies Costume Star Wars Collection Pet Costume, Princess Leia	LUXOTTICA RETAIL NORTH AMERICA INC	113.16	18.86	500.00
Remote & App Controlled Vehicle Parts	Redcat Racing BS810-045 Drive Cups for Differential (Fits Center/Front/Rear, 2Piece)	HOVEROUND CORPORATION	268.00	33.50	700.00
Sports & Outdoors	Funko Pop Keychain: Toy Story - Woody	FAMILY EYE CARE OF OFALLON PC	34.93	14.97	133.33
Toys & Games	Sweet Safari Girl Giant Sign Banner	LINCARE INC	86.97	6.69	1200.00
Toys & Games	Funko Dorbz: The Flash - Zoom Action Figure	HIGHLAND PARK CVS LLC	84.00	12.00	600.00

14. *Logistic management: As the demand for pet care is increasing significantly, the company wants to check if every product with "Pet" in its name and within the Pet Supplies and Toys & Games category but not from other categories (using except for negative subquery or condition for the most convenience) has grown significantly and is offered shipping express or not. Display sales increase growth rate from 2016 to 2019 and its shipping method.*

```
SELECT Product_Name, Category_Name, Is_Express_Shipping, ROUND(((FY17-FY16)/FY16)*100,2) AS "YoY1617",ROUND(((FY1
8-FY17)/FY17)*100,2) AS "YoY1718",ROUND(((FY19-FY18)/FY18)*100,2) AS "YoY1819" FROM (SELECT p.Product_Name, pc.Ca
tegory_Name, p.Is_Express_Shipping,
SUM(CASE WHEN o.Order_Year =2016 THEN p.Price*i.Order_Quantity ELSE 0 END) AS FY16,
SUM(CASE WHEN o.Order_Year =2017 THEN p.Price*i.Order_Quantity ELSE 0 END) AS FY17,
SUM(CASE WHEN o.Order_Year =2018 THEN p.Price*i.Order_Quantity ELSE 0 END) AS FY18,
SUM(CASE WHEN o.Order_Year =2019 THEN p.Price*i.Order_Quantity ELSE 0 END) AS FY19
FROM ORDERS o NATURAL JOIN ITEM_ORDERS i NATURAL JOIN PRODUCTS p NATURAL JOIN PRODUCT_CATEGORIES pc
WHERE o.Order_Status == "Delivered" AND p.Product_Name LIKE "%pet%" AND pc.Category_Name IN ('Toys & Games','Pet
 Supplies') GROUP BY i.Product_ID)
```

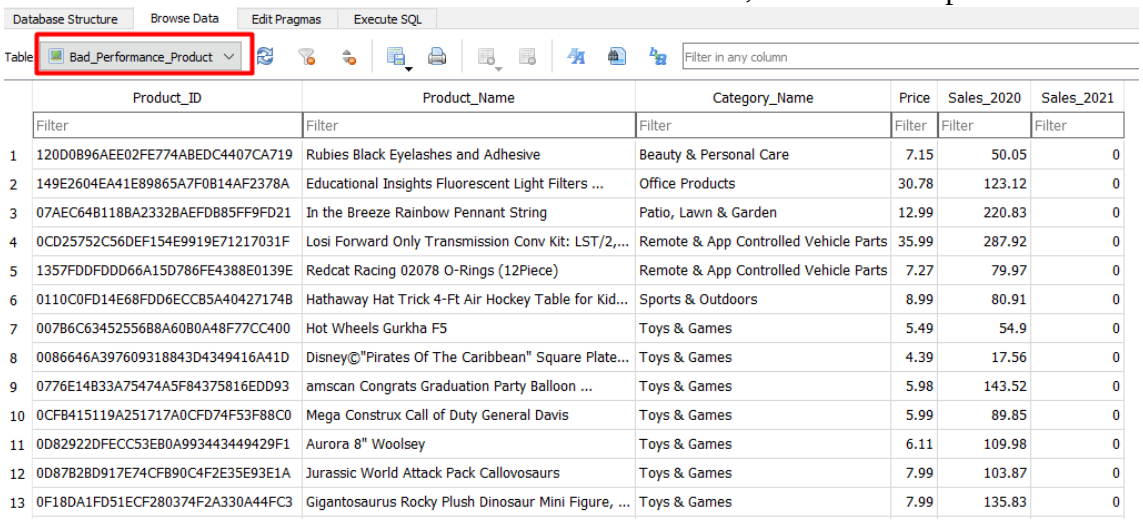
2 records

Product_Name	Category_Name	Is_Express_Shipping	YoY1617	YoY1718	YoY1819
Rubie's Pet Costume	Pet Supplies	1	NA	55.56	0
Rubies Costume Star Wars Collection Pet Costume, Princess Leia	Pet Supplies	1	-100	NA	NA

15. *The marketing department wants to investigate about bad performance of some products that had sales in 2020 but not in 2021. Create a new table displaying all related information: price, sales in 2020 for those by category for later use. (Deploy Left Join, View and If Null function to replace Null/ NA value to 0 for Sales in 2021 for this query)*

```
DROP VIEW IF EXISTS Bad_Performance_Product;
CREATE VIEW Bad_Performance_Product AS
SELECT t1.Product_ID, t1.Product_Name, t1.Category_Name, t1.Price, t1.Sales_2020, IFNULL(t2.Sales_2021, 0) AS Sal
es_2021
FROM (
SELECT p.Product_ID, p.Product_Name, c.Category_Name, p.Price,SUM(i.Order_Quantity * p.Price) AS "Sales_2020"
        FROM ITEM_ORDERS i NATURAL JOIN ORDERS o NATURAL JOIN PRODUCTS p NATURAL JOIN PRODUCT_CATEGORIES c
        WHERE o.Order_Year = 2020 GROUP BY p.Product_ID ) t1
LEFT JOIN (
SELECT p.Product_ID, p.Product_Name, c.Category_Name, p.Price,SUM(i.Order_Quantity * p.Price) AS "Sales_2021"
        FROM ITEM_ORDERS i NATURAL JOIN ORDERS o NATURAL JOIN PRODUCTS p NATURAL JOIN PRODUCT_CATEGORIES c
        WHERE o.Order_Year = 2021 GROUP BY p.Product_ID ) t2
ON t1.Product_ID = t2.Product_ID
WHERE t2.Product_ID IS NULL
ORDER BY t1.Category_Name
```

Because we use View to create a virtual table for later use, the result is captured from database.



	Product_ID	Product_Name	Category_Name	Price	Sales_2020	Sales_2021
	Filter	Filter	Filter	Filter	Filter	Filter
1	120D0B96AEE02FE774ABEDC4407CA719	Rubies Black Eyelashes and Adhesive	Beauty & Personal Care	7.15	50.05	0
2	149E2604EA41E89865A7F0B14AF2378A	Educational Insights Fluorescent Light Filters ...	Office Products	30.78	123.12	0
3	07AEC64B118BA2332BAEFDB85FF9FD21	In the Breeze Rainbow Pennant String	Patio, Lawn & Garden	12.99	220.83	0
4	0CD25752C56DEF154E9919E71217031F	Losi Forward Only Transmission Conv Kit: LST/2,...	Remote & App Controlled Vehicle Parts	35.99	287.92	0
5	1357FDDFDDD66A15D786FE4388E0139E	Redcat Racing 02078 O-Rings (12Piece)	Remote & App Controlled Vehicle Parts	7.27	79.97	0
6	0110C0FD14E68FDD6ECCB5A404271748	Hathaway Hat Trick 4-Ft Air Hockey Table for Kid...	Sports & Outdoors	8.99	80.91	0
7	007B6C63452556B8A60B0A48F77CC400	Hot Wheels Gurkha F5	Toys & Games	5.49	54.9	0
8	0086646A397609318843D4349416A41D	Disney©"Pirates Of The Caribbean" Square Plate...	Toys & Games	4.39	17.56	0
9	0776E14B33A75474A5F84375816EDD93	amscan Congrats Graduation Party Balloon ...	Toys & Games	5.98	143.52	0
10	0CFB415119A251717A0CFD74F53F88C0	Mega Construx Call of Duty General Davis	Toys & Games	5.99	89.85	0
11	0D82922DFECC53EB0A993443449429F1	Aurora 8" Woolsey	Toys & Games	6.11	109.98	0
12	0D87B2BD917E74CFB90C4F2E35E93E1A	Jurassic World Attack Pack Callovosaurs	Toys & Games	7.99	103.87	0
13	0F18DA1FD51ECF280374F2A330A44FC3	Gigantosaurus Rocky Plush Dinosaur Mini Figure, ...	Toys & Games	7.99	135.83	0

D. Analytics and Reports

With only several analyses from our populated database, the system could deliver very insightful findings for user profiles. Below are key take-aways that we could acquire after conducting business queries from basic to advanced level.

Product movement: Top and bottom selling items are determined differently by conditions of either the number of units sold or total revenue. By sold unit, top selling is *Vintage Parts 14558 3" Green Fuzzy Dice with White Dots – Pair* from Toys & Game category, while by revenue, it is *Educational Insights Fluorescent Light Filters (Whisper White), Set of 4* from Office Products category due to its high value per item. Since there is difference in terms of price per unit, when ranking products by category, it should be done by category primarily for fair assessments. Based on insights derived from queries, discount will be applied for top-selling items by units sold and the level of discount depends on sale ranking from historical data.

Customer demographic: The United States leads in sales with 37%, followed by France and Spain with 11% and 10%, respectively. This information accurately represents the reality that all warehouses are situated in the United States, where local buyers will most likely benefit from faster delivery, which will improve regional sales. If we dig deeper into the United States, the three states with the most orders are California, Massachusetts, and New York. The company's senior management will focus on these critical areas since this region accounts for 74% of its customers within the country, and it is recommended that more warehouses be built in order to preserve stock safety and achieve a better customer satisfaction measure.

Operational analysis: The cancellation rate remains extraordinarily high, at around 3%, which is six times higher than the industry average. This raises the ire of the operational team as well as customer service, and the corporation should investigate the causes further by consulting a database of "reasons to cancel."

Time series analysis: November is the busiest shopping month of the year, owing to large-scaled sales festivals such as Black Friday and Single Day Sales (11.11). Surprisingly, the 5th is the top-selling date in the dataset within the month. This means that rigorous stock planning and campaign management should be done ahead of time.

Inventory management: There are three warehouses violating the minimum safety inventory level when the stock is less than 40 units for top-selling products by unit sold. To that end, the Stock status column is created to remind purchasing and warehouse employees which products should be ordered immediately to meet market demand since stockouts hurt business detrimentally. Furthermore, *Preferred Freezer Services (Warehouse ID: 09)* is the warehouse with 2 products currently out of stock. This warehouse should be investigated for the causes of poor management. Finally, as per locations of warehouse, the one with the most inventory is located at CA, USA, corresponding well with the insight from Customer Demographic Analysis that the majority of orders also come from CA, USA. This indicates that we have well-managed distribution across numerous warehouse sites to fulfill market demand.

General management and supplier negotiation: For supplier *ID 18744 - PUBLIX SUPER MARKETS INC* currently supplies the best-selling item by quantity, top management should utilize this information to deal with the profit margin of products. It appears that Toys & Games is the most decent category at the moment with the highest year-on-year sale growth rate. While Pet Supplies products are always available with express shipping options, their sales are not prominent as expected.

In conclusion, even with the modest amount of data populated, the system functions effectively and still delivers helpful insights for its users. It can also generate a virtual database with a View function for later use. From this baseline, we can further develop a more intense system or application reaching beyond a scope of 10 entities. Certainly, I can leverage this mini project for my future career development in the retail domain.

REFERENCES

Kabacoff, R. I. (2015). *R in Action*.

data.world. *Amazon and Best Buy Electronics*. https://data.world/datafiniti/amazon-and-best-buy-electronics

Nelson, G. *Main subject Area ERD for the OLTP Database* https://www.researchgate.net/figure/Main-subject-Area-ERD-for-the-OLTP-Database-Source_fig1_237694304

Tutorialspoint. *Data Warehousing - Quick Guide*. https://www.tutorialspoint.com/dwh/dwh_quick_guide.html

Ung, D. (Apr 8, 2019). *Data Warehousing: Basics of Relational Vs Star Schema Data Modeling*. https://medium.com/@daryl.ung/data-warehousing-basics-of-relational-vs-star-schema-data-modeling-75a68eeaf0e3

Valverde, R. *Sample ERD diagram of a typical order processing system* https://www.researchgate.net/figure/Sample-ERD-diagram-of-a-typical-order-processing-system_fig1_3156662555

Williams, B. (June 24th. 2012). *Data Model for Customers and Inventory Management*. http://www.databaseanswers.org/data_models/customers_and_inventory_mgt/index.html

yugabyteDB. *Northwind sample database*. https://docs.yugabyte.com/latest/sample-data/northwind/

Briggs, P. (Jul 13, 2021). *Canada Ecommerce Forecast 2021*. https://www.emarketer.com/content/canada-ecommerce-forecast-2021

Linnworks. *The Benefits Of Using An Inventory Management System (2021)*. https://www.linnworks.com/blog/inventory-management-benefits

Scientist, B. a. D. (Jan 14, 2020). *Combine SQL Queries With UNION, INTERSECT, EXCEPT*. https://www.youtube.com/watch?v=krnAfIHqGzI

Doug, D. b. (Oct 12, 2017). *SQL Ranking Functions: Part 1 The Over Clause*. https://www.youtube.com/watch?v=YdxyTMjpMMs&t=23s