Database System for E-Commerce Retail Company

Group 14

Ge Wentao	119010080
Zheng Jin	119020078
Dong Jingqi	119010060
Song Wenxin	120090625
Wang Yuancheng	119010319
Zhao Yuzheng	119020077

Database

Database System for E-Commerce Retail Company

Content

Part 1 — Introduction and Motivation

Part 2 — Database Design and Implementation

Part 3 — Database Optimization

Part 4 — Data Implementations

Part 5 — Data Analysis and Data Mining

Part 6 — Conclusions and Self-Evaluation
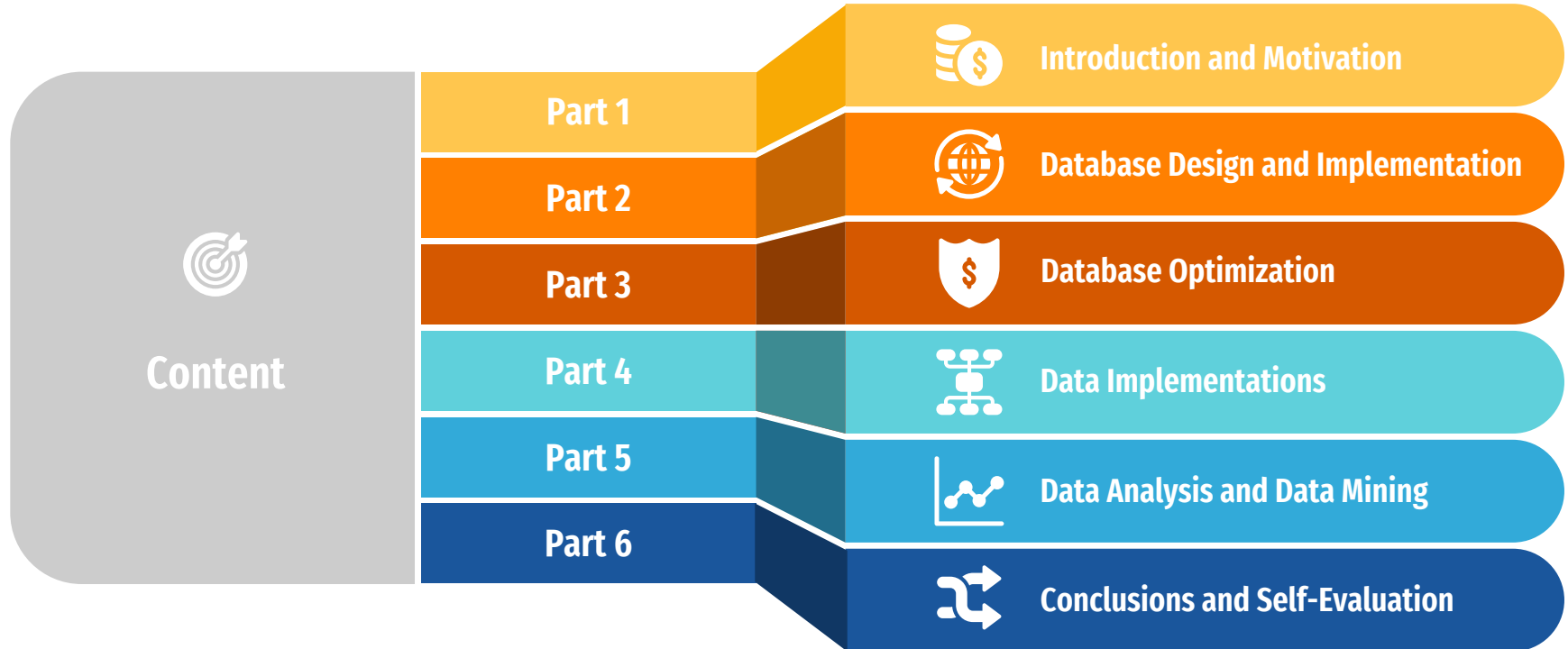
Part 1 - Why E-Commerce?

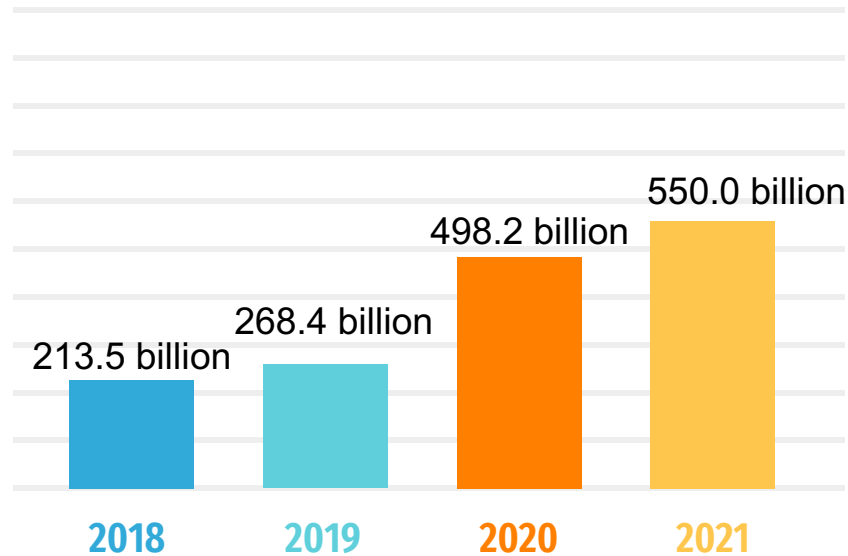2:00 AM

On BED

Not limited

Store A

Store B

20 CNY

5.9 CNY

More choices

Part 1 - Turnover for Double Eleven (11·11)

It's tough to manage huge amount of information

Part 1 - Our goal

Data Organization
The system can classify and organize scattered data for better management

Data Storage
The system can store transaction records for a certain period of time to prepare for data analysis

Data Analysis
The system can determine the trend of product sales based on a large amount of data
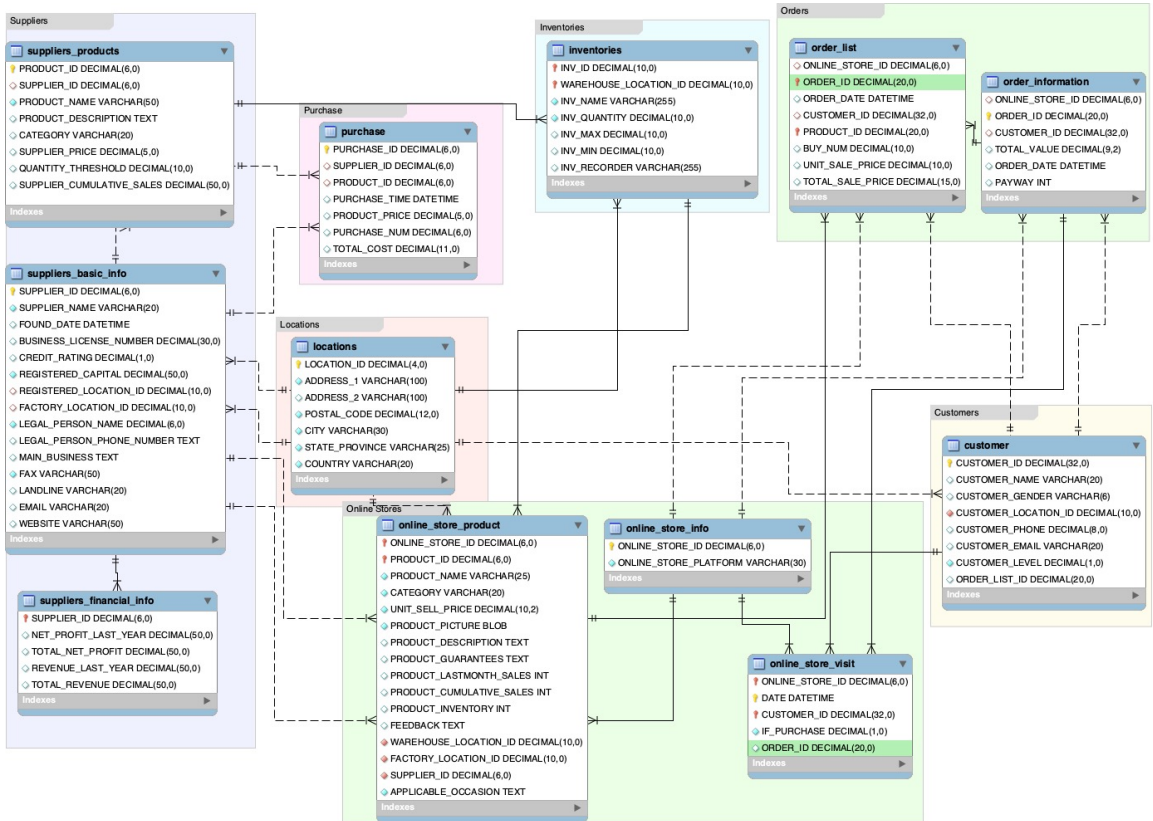
Decision Friendly
The system can give visual information to help people who are not computer literate to manage the store

Part 2 - Database Design

Entity-Relationship Model

- 12 entities and 23 relationships

- Entities are divided into 7 parts
  - Suppliers
  - Purchase
  - Locations
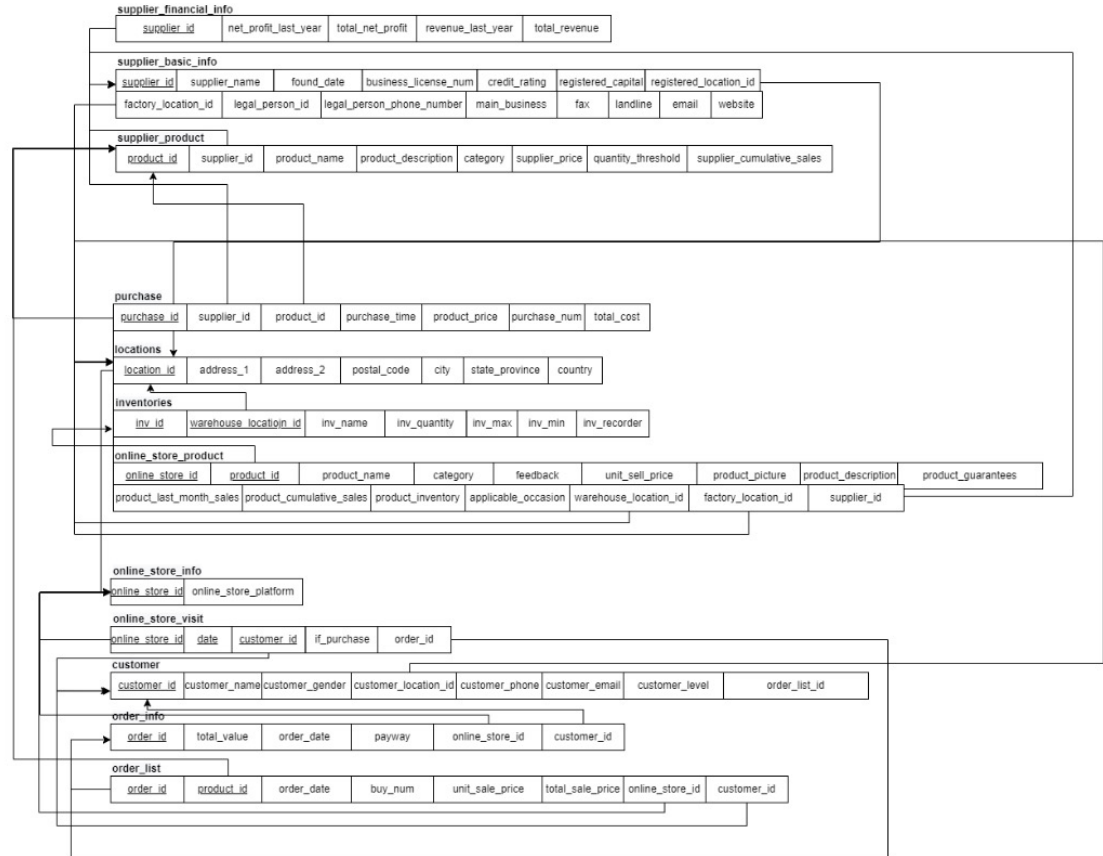 - Inventories
  - Online Store
  - Orders
  - Customers

Suppliers

suppliers_products
- PRODUCT_ID DECIMAL(6,0)
- SUPPLIER_ID DECIMAL(6,0)
- PRODUCT_NAME VARCHAR(50)
- PRODUCT_DESCRIPTION TEXT
- CATEGORY VARCHAR(20)
- SUPPLIER_PRICE DECIMAL(5,0)
- QUANTITY_THRESHOLD DECIMAL(10,0)
- SUPPLIER_CUMULATIVE_SALES DECIMAL(50,0)
- Indexes

suppliers_basic_info
- SUPPLIER_ID DECIMAL(6,0)
- SUPPLIER_NAME VARCHAR(20)
- FOUND_DATE DATETIME
- BUSINESS_LICENSE_NUMBER DECIMAL(30,0)
- CREDIT_RATING DECIMAL(1,0)
- REGISTERED_CAPITAL DECIMAL(50,0)
- REGISTERED_LOCATION_ID DECIMAL(10,0)
- FACTORY_LOCATION_ID DECIMAL(10,0)
- LEGAL_PERSON_NAME DECIMAL(6,0)
- LEGAL_PERSON_PHONE_NUMBER TEXT
- MAIN_BUSINESS TEXT
- FAX VARCHAR(50)
- LANDLINE VARCHAR(20)
- EMAIL VARCHAR(20)
- WEBSITE VARCHAR(50)
- Indexes

suppliers_financial_info
- SUPPLIER_ID DECIMAL(6,0)
- NET_PROFIT_LAST_YEAR DECIMAL(50,0)
- TOTAL_NET_PROFIT DECIMAL(50,0)
- REVENUE_LAST_YEAR DECIMAL(50,0)
- TOTAL_REVENUE DECIMAL(50,0)
- Indexes

Purchase

purchase
- PURCHASE_ID DECIMAL(6,0)
- SUPPLIER_ID DECIMAL(6,0)
- PRODUCT_ID DECIMAL(6,0)
- PURCHASE_TIME DATETIME
- PRODUCT_PRICE DECIMAL(5,0)
- PURCHASE_NUM DECIMAL(6,0)
- TOTAL_COST DECIMAL(11,0)
- Indexes

Locations

locations
- LOCATION_ID DECIMAL(4,0)
- ADDRESS_1 VARCHAR(100)
- ADDRESS_2 VARCHAR(100)
- POSTAL_CODE DECIMAL(12,0)
- CITY VARCHAR(30)
- STATE_PROVINCE VARCHAR(25)
- COUNTRY VARCHAR(20)
- Indexes

Inventories

inventories
- INV_ID DECIMAL(10,0)
- WAREHOUSE_LOCATION_ID DECIMAL(10,0)
- INV_NAME VARCHAR(255)
- INV_QUANTITY DECIMAL(10,0)
- INV_MAX DECIMAL(10,0)
- INV_MIN DECIMAL(10,0)
- INV_RECORDER VARCHAR(255)
- Indexes

Online Stores

online_store_product
- ONLINE_STORE_ID DECIMAL(6,0)
- PRODUCT_ID DECIMAL(6,0)
- PRODUCT_NAME VARCHAR(25)
- CATEGORY VARCHAR(20)
- UNIT_SELL_PRICE DECIMAL(10,2)
- PRODUCT_PICTURE BLOB
- PRODUCT_DESCRIPTION TEXT
- PRODUCT_GUARANTEES TEXT
- PRODUCT_LASTMONTH_SALES INT
- PRODUCT_CUMULATIVE_SALES INT
- PRODUCT_INVENTORY INT
- FEEDBACK TEXT
- WAREHOUSE_LOCATION_ID DECIMAL(10,0)
- FACTORY_LOCATION_ID DECIMAL(10,0)
- SUPPLIER_ID DECIMAL(6,0)
- APPLICABLE_OCCASION TEXT
- Indexes

online_store_info
- ONLINE_STORE_ID DECIMAL(6,0)
- ONLINE_STORE_PLATFORM VARCHAR(30)
- Indexes

online_store_visit
- ONLINE_STORE_ID DECIMAL(6,0)
- DATE DATETIME
- CUSTOMER_ID DECIMAL(32,0)
- IF_PURCHASE DECIMAL(1,0)
- ORDER_ID DECIMAL(20,0)
- Indexes

Orders

order_list
- ONLINE_STORE_ID DECIMAL(6,0)
- ORDER_ID DECIMAL(20,0)
- ORDER_DATE DATETIME
- CUSTOMER_ID DECIMAL(32,0)
- PRODUCT_ID DECIMAL(20,0)
- BUY_NUM DECIMAL(10,0)
- UNIT_SALE_PRICE DECIMAL(10,0)
- TOTAL_SALE_PRICE DECIMAL(15,0)
- Indexes

order_information
- ONLINE_STORE_ID DECIMAL(20,0)
- ORDER_ID DECIMAL(20,0)
- CUSTOMER_ID DECIMAL(32,0)
- TOTAL_VALUE DECIMAL(9,2)
- ORDER_DATE DATETIME
- PAYWAY INT
- Indexes

Customers

customer
- CUSTOMER_ID DECIMAL(32,0)
- CUSTOMER_NAME VARCHAR(20)
- CUSTOMER_GENDER VARCHAR(6)
- CUSTOMER_LOCATION_ID DECIMAL(10,0)
- CUSTOMER_PHONE DECIMAL(8,0)
- CUSTOMER_EMAIL VARCHAR(20)
- CUSTOMER_LEVEL DECIMAL(1,0)
- ORDER_LIST_ID DECIMAL(20,0)
- Indexes

Part 2 - Database Design

Entity-Relationship Model

- 12 entities and 23 relationships

- Entities are divided into 7 parts
  - Suppliers
  - Purchase
  - Locations
  - Inventories
  - Online Store
  - Orders
  - Customers

Table 1: Entities in the database

Category	Entity	Description
Suppliers	suppliers_basic_info	Contain the basic information of all suppliers
	suppliers_financial_info	Contain the financial information of all suppliers
	suppliers_products	Contain the products that each supplier supplies
Purchase	purchase	Contain the information about purchasing product from suppliers
Inventories	inventories	Contain the inventories of products that each supplier possess
Locations	locations	Contain the location information
Online Store	online_store_info	Contain the information about the online store
	online_store_product	Contain the products and product information that are in the online store
	online_store_visit	Contain the customer access (to the store) information
Customers	customers	Contain the customer information
Orders	order_information	Contain the detailed information about orders
	order_list	Contain the list of orders in each online store

Part 2 - Database Design

Relational Schema

- 12 entities and 23 relationships

- Entities are divided into 7 parts
  - Suppliers
  - Purchase
  - Locations
  - Inventories
  - Online Store
  - Orders
  - Customers

Part 3 - Database Optimization

Indexing
a powerful structure which can make searching more effective

Norm form
reduce the duplication of data and data anomalies, ensure referential integrity

Further optimization

Other optimization methods

Part 3 - Database Optimization

B+ Tree

O(n) → O(lgn)

makes sorting more efficient

Stable query time

Hashing

O(n) → O(1)

collisions

more disk storage to save the values

Part 3 - Database Optimization

Rules for creating indexes

frequently searched and infrequently changed

```
----------Create index----------
▶ Run SQL
CREATE INDEX SUPPLIER_NAME_INDEX ON `project`.`suppliers_basic_info`(SUPPLIER_NAME);
▶ Run SQL
CREATE INDEX PURCHASE_TIME_INDEX ON `project`.`purchase`(PURCHASE_TIME);
▶ Run SQL
CREATE INDEX PRODUCT_NAME_INDEX ON `project`.`suppliers_products`(PRODUCT_NAME);
▶ Run SQL
CREATE INDEX CUSTOMER_NAME_INDEX ON `project`.`customer`(CUSTOMER_NAME);
▶ Run SQL
CREATE INDEX ORDER_DATE_INDEX ON `project`.`order_list`(ORDER_DATE);
```

Part 3 - Database Optimization

1.

Reduce the duplication of data

2.

Ensure referential integrity

4.

Avoid data anomalies

3.

Simplify data management

01

02

03

04

3NF

Part 4 - Data Generation

Inventory/Product
The names and basic information such as category and price are from amazon.com

Suppliers, customers, & locations
Data of suppliers, customers, and locations are from data.world

Purchase
Purchase data are from the combination of suppliers & inventories

Order
Order data are from the combination of customers & inventories

Part 4 – Database Implementations

Standerdize the format

Add Primary Keys

Add Foreign Keys
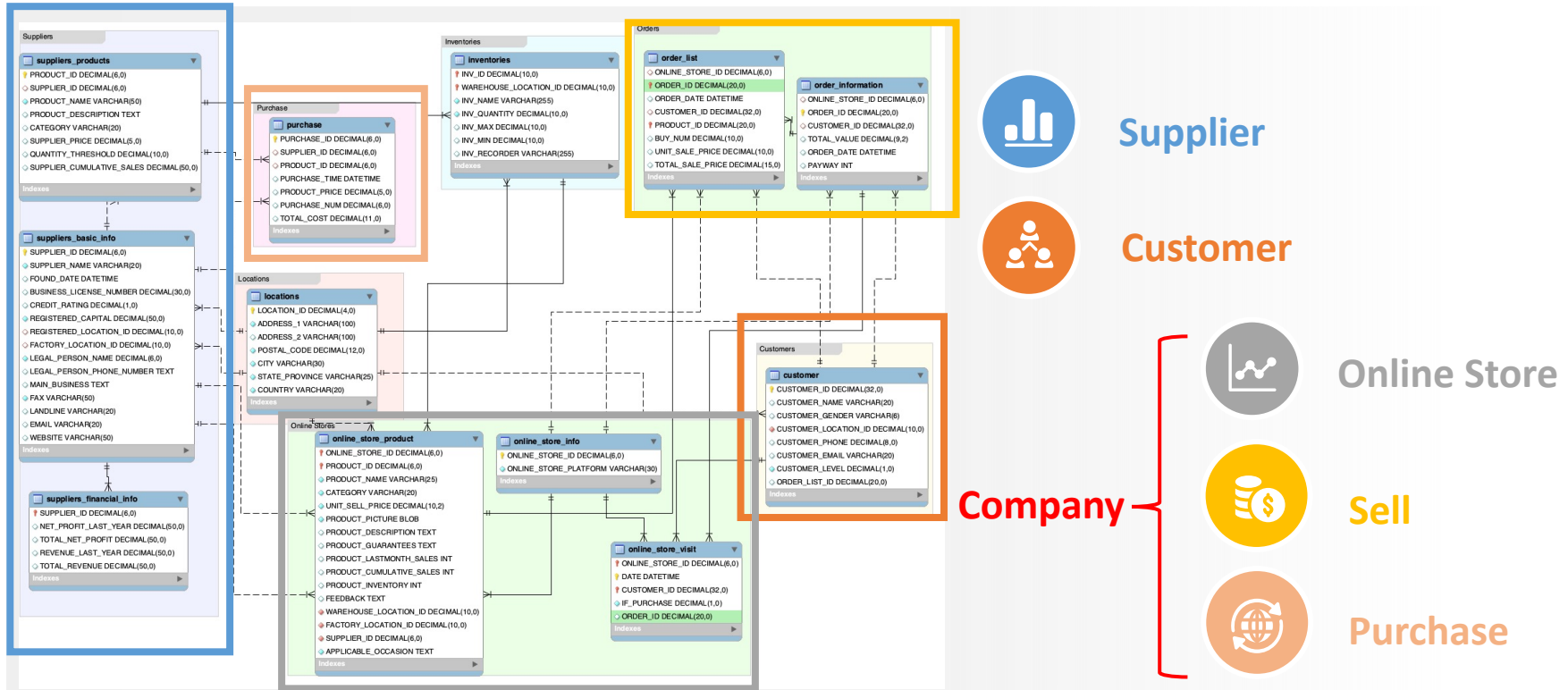
Part 5 - Basic Functions and Result

Supplier

Profit revenue.

SUPPLIER_ID	NET_PROFIT_LAST_YEAR	TOTAL_NET_PROFIT
10	878756	3878756
9	3454576	83454576
6	4324324	54324324
4	4342341	44342341
7	4554554	64554554
20	5452663	35452663
15	6245692	62456921
11	6567665	56567665
12	7675768	67675768

Ordered by: 1.net profit last year
2.total net profit.

1 → recent performance
2 → strong management ability
→→ decide the proportion of goods

The minimum purchasing cost for the same product.

PRODUCT_ID	SUPPLIER_ID	PRODUCT_NAME	CATEGORY	MINIMUM_PRICE
1	1	ANCO MILK	DRINK	3
2	1	ASIA ORANGE	FRUIT	4
3	6	BOY LONDON SPRING SHORT	CLOTH	99
4	7	SONY PS5	ENTERTAINMENT	800
5	8	NICE SHAMPOO	DAILY USE	20
4	7	SONY PS5	ENTERTAINMENT	800
5	8	NICE SHAMPOO	DAILY USE	20
11	17	MACBOOK PRO M1 32+1TB	ELECTRONIC	2999

Eg.

SUPPLIER_ID	PRODUCT_NAME	SUPPLIER_PRICE
1	ANCO MILK	3
2	ANCO MILK	3
3	ANCO MILK	3.5

→→ essential when deciding the selling price.

Part 5 - Basic Functions and Result

Purchase

The supplier that this company order the most from.

SUPPLIER_ID	SUM(TOTAL_COST)
1	64
2	114.6
3	198
5	200

The product that this company order the most.

PRODUCT_ID	SUM(TOTAL_COST)
17	16.5
16	180
20	180

For Purchasing Department:

Seek cooperation

⬇

reduce purchase cost

⬇

boost profits

Sell

The selling records for different months.

PRODUCT_ID	ORDER_MONTH	SALES
1	3	45
2	3	2599
1	4	55
2	4	5198

Grouped by: months.
The most popular items for each month or different seasons.

The retailing department:
adjust the purchase quantity for some items in time

Part 5 - Basic Functions and Result

Customer

The detailed information of order.

ORDER_ID	ONLINE_STORE_ID	CUSTOMER_ID	PRODUCT_ID	BUY_NUM	UNIT_SALE_PRICE	TOTAL_SALE_PRICE
1	1	10001	1	2	5	10
1	1	10001	2	2	7	14
1	1	10001	3	1	499	499

By order id.
Help with after sales problems.
Customer server: know better of the products the customer has bought

The total amount of consumption for different customers.

CUSTOMER_ID	TOTAL_VALUE
1	523
2	2599
3	69

grade of membership

⬆⬇

Consumption

Offer different account

Online store

The average conversion rate.

DATE	Daily_Conversion_Rate
2022/3/17	33.30%
2022/3/18	50%
2022/3/19	100.00%
2022/3/20	66.70%

The conversion rate=
$$\frac{\text{\# of records of transaction}}{\text{\# of visit records.}}$$

Help to check:
Shopping Interface:
        attractive enough?
display order:
        reasonable.?

Part 5 - Transaction Update Function

Original

INV_ID	WAREHOUSE_LOCATION_ID	INV_NAME	INV_QUANTITY	INV_MAX	I
1	1000000001	ANCO MILK	500	1000	0
2	1000200002	ASIA ORANGE	1000	2000	0
3	1234231558	BOY LONDON SPRING SHORT SLEEVE T-SHIRT	250	2000	0
4	1000000003	SONY PS5	200	500	0
5	1000000004	NICE SHAMPOO	1000	5000	0
6	1000000005	YIBAO WATER	2000	10000	0

ANCO MILK

Purchase 100 boxes from the suppliers

INV_ID	WAREHOUSE_LOCATION_ID	INV_NAME	INV_QUANTITY	INV_MAX	I
1	1000000001	ANCO MILK	600	1000	0
2	1000200002	ASIA ORANGE	1000	2000	0
3	1234231558	BOY LONDON SPRING SHORT SLEEVE T-SHIRT	250	2000	0
4	1000000003	SONY PS5	200	500	0
5	1000000004	NICE SHAMPOO	1000	5000	0
6	1000000005	YIBAO WATER	2000	10000	0

Make change on table
purchase **and** *inventories*.

Sell 1 box to the customers

INV_ID	WAREHOUSE_LOCATION_ID	INV_NAME	INV_QUANTITY	INV_MAX	I
1	1000000001	ANCO MILK	599	1000	0
2	1000200002	ASIA ORANGE	999	2000	0
3	1234231558	BOY LONDON SPRING SHORT SLEEVE T-SHIRT	250	2000	0
4	1000000003	SONY PS5	200	500	0
5	1000000004	NICE SHAMPOO	1000	5000	0
6	1000000005	YIBAO WATER	2000	10000	0

Make change on table
order_information,
order_list, **and** *inventories*.

Part 5 - Data Mining

User Retention Rate Analysis

Day-N Retention Rate = $\dfrac{R_{t_N}(U_{t_0})}{U_{t_0}}$

u_{t_0} - the number of **new users at day t_0**

$R_{t_N}(U_{t_0})$ - the number of new users registering at day t_0 and **visiting the website at day t_N**
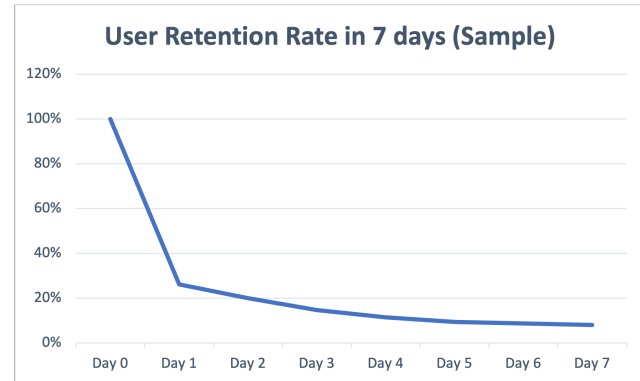
Observation

- Convex Curve
- Marginal user retention rate decreases
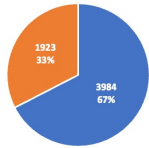
Suggestion

- Justify the company's operation strategies

Cohort of App Launched > App Launched — Daily — Users who came back on →

Cohort	New Users	Day 0	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
All	134,529	100.0%	26.1%	20.0%	14.8%	11.5%	9.4%	8.8%	8.1%
Aug 30	32,961	100.0%	31.3%	26.8%	20.2%	13.4%	10.2%	9.7%	8.1%
Aug 31	20,144	100.0%	21.6%	17.2%	13.7%	9.8%	8.4%	7.9%	
Sep 01	17,772	100.0%	22.0%	19.8%	14.2%	11.2%	9.6%		
Sep 02	21,455	100.0%	27.4%	21.3%	15.8%	11.7%			
Sep 03	18,566	100.0%	24.5%	16.4%	10.3%				
Sep 04	10,656	100.0%	29.2%	18.6%					
Sep 05	12,975	100.0%	26.8%						
Sep 06	5,651	100.0%							

User Retention Rate in 7 days (Sample)

Part 5 - Data Mining

Gender Analysis

Gender of Customers

- 1923 33% (male/orange)
- 3984 67% (female/blue)
- female ■ male

Total Consumption towards Different Gender
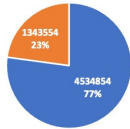
- 1343554 23%
- 4534854 77%
- female ■ male

Average Consumption towards Different Gender
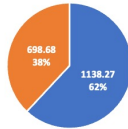
- 698.68 38%
- 1138.27 62%
- female ■ male

Observation

Women are the main force of consumption

Suggestion
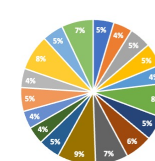
Targeted recommendations

State Analysis

Number of Customers

15%, 7%, 11%, 6%, 11%, 8%, 4%, 2%, 2%, 2%, 5%, 6%, 4%, 1%, 2%, 8%

Total Consumption

14%, 6%, 10%, 8%, 9%, 11%, 4%, 6%, 9%, 4%, 2%, 2%, 4%, 5%, 4%, 1%, 2%, 3%

Average Consumption

5%, 4%, 5%, 4%, 8%, 5%, 6%, 7%, 9%, 5%, 4%, 5%, 4%, 5%, 8%, 4%, 7%, 5%
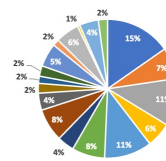
- ■ CALIFORNIA ■ NEW YORK ■ MISSISSIPPI ■ OREGON ■ TEXAS ■ WISCONSIN ■ UTAH ■ SOUTH CAROLINA ■ INDIANA
- ■ COLORADO ■ MINNESOTA ■ FLORIDA ■ MASSACHUSETTS ■ PENNSYLVANIA ■ NEBRASKA ■ NEVADA ■ TENNESSEE ■ ALABAMA
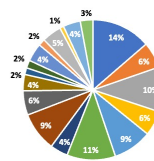
Observation

- California
- Colorado

Suggestion

- Customer events
- Marketing

Part 5 - Data Mining
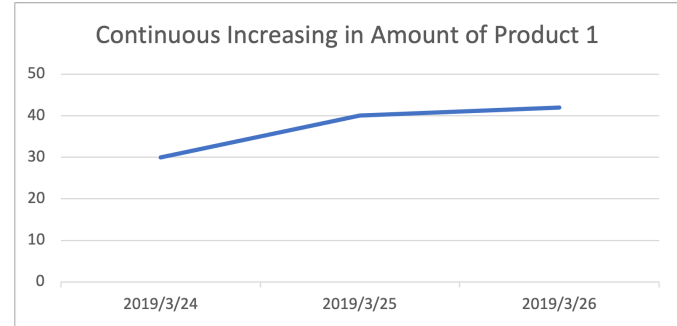
Return the time period where the product has continuous increasing in selling amount

- Reflect the popularity of the product
- Help check the peak season for products
- Conduct the marketing of different products according to the season

Part 6 – Conclusion and Self-Evaluation

Conclusion — A Database System for E-commerce Retail Company

Conclusion — Recognize Entities and Relationship between Them

Conclusion — Data Generation and Data Analysis / Mining

Self-Evaluation — A Well-developed E-commerce Data Management System

Thank You