

# **The Database Proposal for a Beverage Chain Store in Ningbo**

Linyan Li

May 2024

## CONTENT

<b>1. Introduction</b> .....	3
<b>2. Data design document and data dictionary</b> .....	3
<b>2.1 Business rule &amp; Assumption</b> .....	3
<b>2.2 System definition</b> .....	4
<b>2.3 ER diagram</b> .....	5
<b>2.4 Data schema</b> .....	7
<b>2.5 Data dictionary</b> .....	8
<b>3. SQL queries</b> .....	12
<b>3.1 Customer</b> .....	12
<b>3.2 Staff</b> .....	13
<b>3.3 Suppliers</b> .....	15
<b>3.4 Third-party logistics provider</b> .....	15
<b>4. Visualization</b> .....	16
<b>4.1 Material supply situation</b> .....	16
<b>4.2 Customer portrait</b> .....	18
<b>4.3 Sales analysis</b> .....	20
<b>4.3.1 Store performance</b> .....	20
<b>4.3.2 Product performance</b> .....	21
<b>5. Business intelligence</b> .....	22
<b>5.1 Regression</b> .....	22
<b>5.2 Classification</b> .....	25
<b>6. Reflection: Advantage &amp; Limitation</b> .....	26
<b>7. Conclusion</b> .....	26
<b>Appendix</b> .....	27
<b>1. Meeting minutes</b> .....	Error! Bookmark not defined.
<b>2. Python code for BI regression</b> .....	27

## **1. Introduction**

As the public lifestyle is becoming ever-increasingly youthful, coffee, milk tea and other beverages are gradually popularized. More and more brands of chain beverage stores are appearing, and various categories of drinks are emerging. Waking up early for a cup of coffee seems to have become a new trend. To support the operation of a brand chain beverage shop in Ningbo, this report provides an exhaustive description of a specially designed database from the design purpose to using process.

## **2. Data design document and data dictionary**

### **2.1 Business rule & Assumption**

The company is a start-up beverage chain store headquartered in Ningbo. Although at an initial stage, a relatively well-developed company structure and stable supply chain have been established, which are essential for future expansion. For the raw materials, the company builds close cooperation relationships with 50 suppliers, each of them providing less than 5 kinds of materials and one kind of material is provided by one supplier. Thus, each material shipped in could be inspected one-on-one by simply tracking the unique supplier for it, which guarantees a relatively high-quality level. The transportation of material between suppliers, warehouses, and stores is handled by a specialized transportation company. The processes of materials in and out of storage are recorded in inventory logs. As for personnel assignment, general staff and interns are reasonably allocated to 10 chain stores and each store is managed by a high-level staff.



**Figure 2.1**

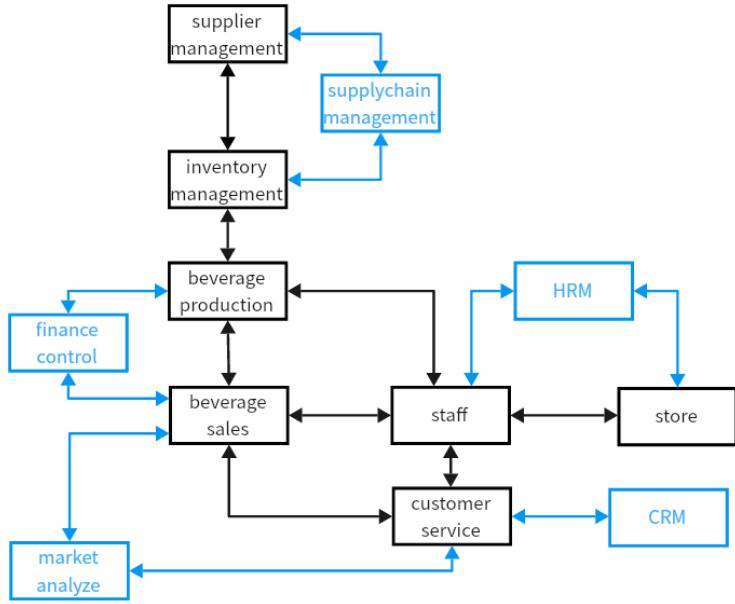
The company offers 30 differentiated products, mostly using raw materials from abroad. Targeted customers are the younger generations, such as white-collar workers who are aspiring to a classy lifestyle and students who are keen

on exploring novel experiences. To satisfy segmented demand, different channels for ordering are provided (online+delivery, online+pickup, offline+delivery, offline+pickup). The business focuses on maintaining a stable customer base and expanding by encouraging repeated purchases and strengthening links with customers. Thus, a loyalty points redemption mechanism and review scoring mechanism are added.

In the next 3 years, expanding the customer base and product category, increasing sales in Ningbo and gaining a foothold in the medium-to-high ends beverage market will be under consideration. After that, relying on the birthplace, the company aims to expand its footprint across more city branches, radiation in East China then finally the whole country. Beyond a commercial objective, the company is also committed to promoting sustainable development by improving operation processes and minimizing environmental footprint. To achieve this goal, sustainability is assessed and recorded.

## **2.2 System definition**

By analyzing the data collected so far, the boundary of the database system can be defined. The graph below shows the function of the database and aspects of the company's other operations that can be supported by the database. Supply chain management could analyze supplier's overview and performance based on analyzing data such as material changed quantity and corresponding suppliers, collected from the inventory and supplier management department. Products' cost, selling price, and sales quantity can be collected by the staff of the finance controlling department, contributing to the cost and profit analysis of categories of product. Customer relationship management organizes customer data and shares the data with market analysis management. Market analysis management department discovers the linkage of beverage sales data with customer data and makes user portraits to do market promotion targeting certain category of customers and then do well-directed market penetration. Besides, human resource management department gains staff identity information and dispatches staff amongst stores according to need.



**Figure 2.2**

### 2.3 ER diagram

By analyzing users of the database, such as staff and suppliers, we find out data categories that are valuable to them and identify entities and analyze their mutual relationships. The type of relationship that occurred frequently in ER diagram is many-to-many. To address this, we create an extra entity between the two entities that have a many-to-many relationship and change the relationship from many-to-many to one-to-many. Also, we find some relationships between entities are one-to-one type, thus we transform some of the entities to be the foreign key of the other to merge unnecessary entities. After that, we determine the primary keys in each entity, then post a copy of the parent entities' primary keys to child entities to act as foreign keys. We adjust the ER diagram to ensure it meets integrity constraints. For instance, we create an "Inventory log" entity to record every material quality change in inventory and link it with the "Material" entity to trace the changed material's other attribute. One characteristic of the shops is customers can be divided into three categories which are online, offline and cross-channel, to deal with this, we classify orders into two categories—delivery orders and offline pick-up orders and create related entities to show different traits of them. More specifically, both the delivery orders and offline pick-up orders have online and offline ordering channels, while only the delivery orders have a relationship with

the delivery person.

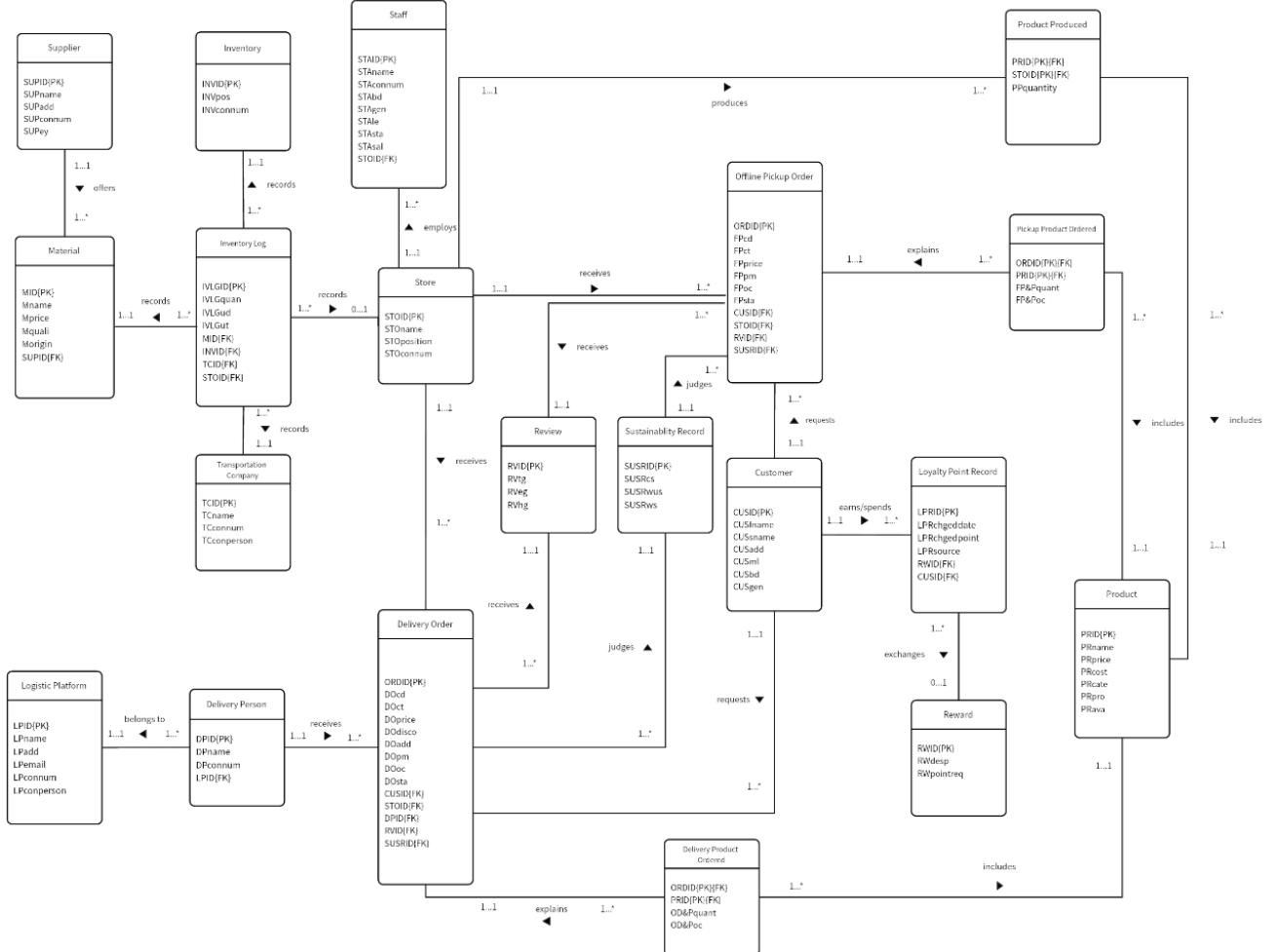


Figure 2.3

## 2.4 Data schema

<b>Supplier</b> (SUPID, SUPname, SUPadd, SUPconnum, SUPey) <b>Primary Key</b> SUPID	<b>Transportation company</b> (TCID, TCname, TCconnum, TCconperson) <b>Primary Key</b> TCID
<b>Material</b> (MID, Mname, Mprice, Mquali, Morigin, SUPID) <b>Primary Key</b> MID <b>Foreign Key</b> SUPID	<b>Inventory Log</b> (IVLID, MID, IVLGquan, IVLGud, IVLGut, INVID, TCID, STOID) <b>Primary Key</b> IVLID <b>Foreign Key</b> MID, INVID, STOID, TCID
<b>Inventory</b> (INVID, INVpos, INVconnum) <b>Primary Key</b> INVID	<b>Product</b> (PRID, PRname, PRprice, PRcost, PRcate, PRpro, PRava) <b>Primary Key</b> PRID
<b>Store</b> (STOID, STOname, STOpos, STOconnum) <b>Primary Key</b> STOID	<b>Review</b> (RVID, RVtg, RVeg, RVhg) <b>Primary Key</b> RVID
<b>Sustainability Record</b> (SUSRID, SUSRcs, SUSRwus, SUSRws) <b>Primary Key</b> SUSRID	<b>Staff</b> (STAID, STAfname, STAsname, STAconnum, STAbd, STAgens, STAlte, STAsa, STAsal, STOID) <b>Primary Key</b> STAID <b>Foreign Key</b> STOID
<b>Delivery Order</b> (ORDID, DOcd, DOct, DOpice, DOdisco, DOadd, DOpm, DOoc, DOsta, CUSID, STOID, DPID, RVID, SUSRID) <b>Primary Key</b> ORDID <b>Foreign Key</b> CUSID, STOID, DPID, RVID, SUSRID	<b>Offline Pickup Order</b> (ORDID, FPcd, FPct, FPprice, FPpm, FPoc, FPsta, CUSID, STOID, RVID, SUSRID) <b>Primary Key</b> ORD_ID <b>Foreign key</b> CUSID, STOID, RVID, SUSRID
<b>Delivery Product Ordered</b> (ORDID, PRID, OD&Pquant, OD&Poc) <b>Primary Key</b> ORDID, PRID <b>Foreign Key</b> ORDID, PRID	<b>Pickup Product Ordered</b> (ORDID, PRID, FP&Pquant, FP&Poc) <b>Primary Key</b> ORDID, PRID <b>Foreign Key</b> ORDID, PRID
<b>Customer</b> (CUSID, CUSfname, CUSsname, CUSadd, CUSml, CUSbd, CUSgen) <b>Primary Key</b> CUSID	<b>Loyalty Point Record</b> (LPRID, LPRchgeddate, LPRchgedpoint, LPRsource, RWID, CUSID) <b>Primary Key</b> LPRID <b>Foreign Key</b> CUSID, RWID
<b>Reward</b> (RWID, RWdesp, RWpointreq) <b>Primary Key</b> RWID	<b>Logistic Platform</b> (LPID, LPname, LPadd, LPemail, LPconnum, LPconperson) <b>Primary Key</b> LPID
<b>Delivery Person</b> (DPID, DPname, DPconnum, LPID) <b>Primary Key</b> DPID <b>Foreign Key</b> LPID	<b>Product Produced</b> (PRID, STOID, PPquantity) <b>Primary Key</b> PRID, STOID <b>Foreign Key</b> PRID, STOID

Figure 2.4

## 2.5 Data dictionary

**Data Dictionary**

Entity	Attribute names	Data Type	Description	Allowable values (Nulls)
Supplier	SUPID	Text	Supplier ID, PK of Supplier table	SUP001- SUP050 Can't be null
	SUPname	Text	Supplier name, Supplier (as a company)'s full name	Variable characters Can't be null
	SUPadd	Text	Supplier address, Supplier's registration address, where the company locates eg. abc Street, m City	Variable characters Can't be null
	SUPconnum	Number	Supplier contact number, Supplier's contact number To keep in touch with main person in charge of the business	11 Variable numbers Can't be null
	SUPey	Date	Supplier established year, The year when the supplier is established, eg. When they began their business	Variable Date Can't be null
Material	MID	Text	Material ID, PK of material table	M100001- M999999 Can't be null
	Mname	Text	Material name, Material's full name	Variable characters Can't be null
	Mprice	Currency	Material price, Price of raw materials per 1kg (RMB)	Variable currency Can't be null
	Mquali	Text	Material quantity, Quality of material	A+, A-, A B+, B-, B C+, C-, C Can't be null
	Morigin	Text	Material origin, Country/province of origin of raw materials, e.g. UK	Variable characters Can't be null
	SUPID	Text	Supply ID, PK of Supply table, FK of Material table	SUP001- SUP050 Can't be null
Transportation Company	TCID	Text	Transportation Company ID, PK of Transportation company table	TC001-TC999 Can't be null
	TCname	Text	Transportation Company name, Full name of the transportation company	Variable characters, no more than 20 characters Can't be null
	TCconnum	Number	Transportation Company contact number, Contact number of the transportation company To keep in touch with person in charge	12 numbers Can't be null
	TCconperson	Text	Transportation Company contact person, Full name of the contact person in transportation company responsible for docking	Variable characters, no more than 20 characters Can't be null
Inventory	INVID	Text	Inventory ID, PK of Inventory table	I100001-I1999999 Can't be null
	INVpos	Text	Inventory position, Location of the warehouse e.g. ABC street, m City	Variable characters Can't be null
	INVconnum	Number	Inventory contact number, Warehouse's contact number, to keep in touch with inventories	11 Variable numbers Can't be null
Inventory Log	IVLGIN	Text	Inventory Log ID, PK of Inventory Log table	IL100001-IL999999 Can't be null
	MID	Text	Material ID, PK of Material table, FK of Inventory Log table	M100001- M999999 Can't be null
	IVLGquan	Number	Inventory Log quantity, Inventory logging, positive for incoming shipments, negative for outgoing shipments	Variable number Can't be null

	IVLGud	Date/Time	Inventory Log update, Update date of the Inventory log	Variable date Can't be null
	ILVGut	Date/Time	Inventory Log update time, Update time of the Inventory log	Variable date Can't be null
	INVID	Text	Inventory ID, PK of Inventory table, FK of Inventory Log table	I100001-I199999 Can't be null
	TCID	Text	Transportation Company ID, PK of Transportation Company table, FK of Inventory Log table	TC001-TC999 Can't be null
	STOID	Text	Store ID, PK of Store table, FK of Inventory Log table	STO001-STO999 Can be null if incoming shipments
Store	STOID	Text	Store ID, PK of Store table	STO001-STO999 Can't be null
	STOname	Text	Store full name, eg. Nottingham on-campus store	Variable characters, no more than 50 characters Can't be null
	STOpos	Text	Store position, Current position title, eg. abc Street	Variable characters, no more than 100 characters Can't be null
	STOconnum	Number	Store Contact Number, to keep in touch with the store	11 Variable numbers Can't be null
Product	PRID	Text	Product ID, PK of Product table	P100001-P 1000030 Can't be null
	PRname	Text	Product name, eg. Americano	Variable characters, no more than 30 characters Can't be null
	PRprice	Currency	Product price, Unit price of the product (RMB)	0-99 Can't be null

	PRcost	Currency	Product cost, Unit cost of the product (RMB)	0-99 Can't be null
	PRcate	Text	Product category, Category for the product, eg. Juice/milktea/coffee	Variable characters, no more than 20 characters Can't be null
	PRpro	Text	Property of the product, Distinguish the property, whether the product is a long-term supply product or a limited supply product, regular product = RP, limited product = LP	RP/LP Can't be null
	PRava	Number	Product availability, Availability of the product, Whether the product is being sold available, being sold = 1, not being sold = 0	0/1 Can't be null
Product Produced	PRID	Text	Prduct ID, PK of the Product table, PK and FK of the Product Produced table	P100001-P100030 Can't be null
	STOID	Text	Store ID, PK of the Store table, PK and FK of the Product Produced table	STO001-STO999 Can't be null
	PPquan	Number	Product Produced quantity, number of the product produced	Variable number Can't be null
Review	RVID	Text	Review ID, PK of the Review table	R100001-R100125 Can't be null
	RVtg	Number	Review taste grade, Taste Grade can be rated ranging from 1-5 (5 for best and 1 for worst)	Interger beween 1-5 Can't be null
	RVeg	Number	Review experience grade, Experience Grade can be rated ranging from 1-5 (5 for best and 1 for worst)	Interger beween 1-5 Can't be null
	RVhg	Number	Review health grade, Health Grade can be rated ranging from 1-5 (5 for best and 1 for worst)	Interger beween 1-5 Can't be null

Staff	STAID	Text	Staff ID, PK of Staff table	STA001-STA999 Can't be null
	STAfirstname	Text	Staff's first name	Variable characters Can't be null
	STAsurname	Text	Staff's surname	Variable characters Can't be null
	STAconnum	Number	Staff's contact number To keep in touch with the staff	Variable numbers Can't be null
	STAbd	Text	Staff's birthday, Personal information of staff	Variable date Can't be null
	STAgend	Text	Staff's gender, M=male, F=Female	F/M Can't be null
	STAle	Text	Staff Position Level: Intern =L0, Regular staff =L1, Supervisor=L2	L0/L1/L2 Can't be null
	STAsta	Text	Staff's status, Active=A, Resigned=F	A/F Can't be null
	STAsal	Currency	Staff's salary. Different levels of employees have different pay levels	Variable currency Can't be null
	STOID	Text	Store ID, PK of Store table, FK of Staff table	STO001-STO999 Can't be null
Delivery Order	ORDID	Text	Delivery Order ID, PK of Delivery Order table	ORD101001- ORD101250 Can't be null
	DOcd	Date/Time	Delivery Order create date, date for the delivery order created	2024/4/1-2024/4/7 Can't be null
	DOct	Date/Time	Delivery Order create time, time (hour, minute, second) for the delivery order created	Set between 9:00-21:00 Can't be null
Offline Pickup Order	DOprice	Currency	Delivery Order price, total price for the delivery order	Range from ¥12.00-¥228.00 Can't be null
	DOdisco	number	Delivery Order discount, discount rate	0.95/null Can be null
	DOadd	Text	Delivery Order address, the address of the delivery order	Can't be null
	DOpm	Text	Delivery Order Payment Method, payment way of delivery order	Card/Alipay/Wechatpay Can't be null
	DOsta	Text	Delivery order's status In production=P, Completed=C Abnormal=A	P/C/A Can't be null
	DOoc	Text	Delivery order's ordering channel online/offline	online/offline Can't be null
	CUSID	Text	Customer's ID, PK of Customer table, FK of Delivery Order table	CUS001-CUS999 Can't be nul
	STOID	Text	Store's ID, PK of Store table, FK of Delivery Order table	STO001-STO999 Can't be null
	DPID	Text	Delivery Person's ID, PK of Delivery Person, FK of Delivery Order table	DP001-DP050 Can't be null
	RVID	Text	Review ID, PK of Review table, FK of Delivery Order table	R100001-R100125 Can't be null
	SUSRID	Text	Sustainability Record, PK of Sustainability Record table, FK of Delivery Order table	STO001-STO999 Can't be null
	ORDID	Text	Offline Pickup Order ID, PK of Offline Pickup Order table	ORD101001-ORD102250 Can't be null
	FPcd	Date/Time	Offline Pickup Order create date, date for the Offline Pickup Order creation	2024/4/1-2024/4/7 Can't be null

	FPct	Date/Time	Offline Pickup Order create time, time (hour, minute, second) to the Offline Pickup Order creation Set between 9:00-21:00	Set between 9:00-21:00 Can't be null
	FPprice	Currency	Offline Pickup Order price, the total price for the whole Offline Pickup Order	Range from ¥12.00-¥228.00 Can't be null
	FPPm	Text	Offline Pickup Order's Payment way Eg. Cash/Card/Alipay/Wechatpay/Card	Card/Alipay/Wechatpay/Card Can't be null
	FPOC	Text	Offline Pickup Order's ordering channel online/offline	online/offline Can't be null
	FPsta	Text	Offline Pickup Order status, Status of Offline Pickup Order In production=P, completed=C	P/C
	CUSID	Text	Customer ID, PK of Customer table, FK of Offline Pickup Order	CUS001-CUS999 Can't be null
	STOID	Text	Store ID, PK of Store table, FK of Offline Pickup Order	STO001-STO999 Can't be null
	RVID	Text	Review ID, PK of Review table, FK of Offline Pickup Order	RV100001- R100125 Can't be null
	SUSRID	Text	Sustainability Record ID, PK of Sustainability Record table, FK of Offline Pickup Order	SUS001- SUS125 Can't be null
Delivery Product Ordered	ORDID	Text	Delivery Product Ordered ID, PK of the Delivery Product Ordered table, FK of the Delivery Order table	Delivery Product Ordered ORD103001- ORD104050 Can't be null
	PRID	Text	Product ID, PK of the Product Table, FK of the Delivery Product Ordered	P100001-P100030 Can't be null
Pickup Product Ordered	OD&Pquant	Number	Delivery Product Ordered quantity, Quantity of each product in the order ID (in pieces)	From 1-20 Can't be null
	OD&Poc	Text	Delivery Product Ordered ordering channel, Pickup Product Ordered Distribution channel: Online/offline	online/offline Can't be null
	ORDID	Text	Pickup Product Ordered ID, PK of the Pickup Product Ordered table, FK of the Offline Pickup Order table	ORD101001- ORD102069 Can't be null
	PRID	Text	Product ID, PK of the Product table, FK of the Pickup Product Ordered table	P100001-P100030 Can't be null
	FP&Pquant	Number	Pickup Product Ordered quantity, Quantity of each product in the order ID (in pieces)	From 1-6 Can't be null
	FP&Poc	Text	Pickup Product Ordered ordering channel, Distribution channel: online/offline	online/offline Can't be null
	SUSRID	Text	Sustainability Record ID, PK of the Sustainability Record Table	SUS001-SUS125 Can't be null
	SUSRCS	Number	Sustainability Record carbon score, CO2 emission index rating (1-5), the higher the more environmentally friendly	Integer between 1-5 Can't be null
Sustainability Record	SUSRWS	Number	Sustainability Record water use score, Water Use Index rating (1-5), higher is more environmentally friendly	Integer between 1-5 Can't be null
	SUSRWS	Number	Sustainability Record waste score, Waste index rating (1-5), the higher the less wasteful	Integer between 1-5 Can't be null
	Customer	CUSID	Customer ID, PK of the customer table	CUS001-CUS999 Can't be null

	CUSfname	Text	Customer First name	Variable characters Can't be null
	CUSsname	Text	Customer Surname	Variable characters Can't be null
	CUSadd	Text	Customer Address, for delivering the order	Variable characters Can't be null
	CUSml	number	Customer Membership Level, points rating based on the amount of money spent, free registration is available, member level (1-5), non-member = 0	0-5 Can't be null
	CUSbd	Date/time	Customer Birthday, the customer's birthday (yy-mm-dd)	Variable date Can't be null
	CUSgen	Text	Customer gender, M=male, F=Female	F/M Can't be null
Loyalty Point Record	LPRID	Text	Loyalty Point Record ID, PK of the Loyalty Point Record table	LPR001-LPR309 Can't be null
	LPRchgeddate	Date/time	Loyalty Point Record changed date	24.4.1-24.4.30 Can't be null
	LPRchgedpoint	Number	Loyalty Point Record changed point, points quantity	Integer (can be negative and positive) Can't be null
	LPRsource	Text	Loyalty Point Record source, point sources, e.g. purchase, activity	Variable characters Can be null if LPRchgedpoint<0
	RWID	Text	RewardID, PK of the Reward table, FK of the Loyalty Point Record table	R001-R009 Can be null
	CUSID	Text	Customer ID, PK of Customer table, FK of Royalty Point Record table	CUS001-CUS400 Can't be null if LPRchgedpoint>0
Reward	RWID	Text	Reward ID, PK of the Reward table	R001-R009 Can't be null
Delivery Person	RWdesp	Text	Reward Description, Names of redemption prizes, e.g. stickers, mugs	Variable characters Can't be null
	Rwpointreq	Number	Reward Point Required, Points needed to convert this prize	10-90 Can't be null
	DPID	Text	Delivery Person ID, PK of the Delivery table	DP001-DP050 Can't be null
	DPname	Text	Delivery Person name, the delivery person's name	Variable characters Can't be null
	DPconnum	Number	Delivery Person contact number, the delivery person's contact number	Variable numbers Can't be null
	LPID	Text	Logistic Platform ID, PK of the Logistic Platform table, FK of the Delivery Person table	LP001-LP002 Can't be null
Logistic Platform	LPID	Text	Logistic Platform ID, PK of the Logistic Platform table which logistic platform the delivery person belongs to	LP001-LP002 Can't be null
	LPname	Text	Logistic Platform Name	Variable characters Can't be null
	LPadd	Text	Logistic Platform Address, the company's address, e.g. ABC Street	Variable characters Can't be null
	LPemail	Text	Logistic Platform's Email, to contact with the company	Variable characters Can't be null
	LPconnum	Number	Logistic Platform's contact number (person in charge)	11 numbers Can't be null
	LPconperson	Text	Logistic Platform's contact person, Name of the person in charge of the company	Variable characters Can't be null

**Figure2.5**

### 3.SQL queries

#### 3.1 Customer

These two SQL queries (Figure 3.1) are designed to enhance the customer experience. The first one allows customers to check the quantity of ongoing orders in front of them at any time and decide whether to place an order, which reduces the waiting time of customers. The second can help customers view the top 5 highest-selling products, helping decisions of making choices.

<b>Customer1: Find number of orders currently processing in store</b>																									
<b>Code</b>																									
<pre>customer1: Find number of orders currently processing in store</pre> <pre>SELECT Count(*) AS OngoingOrderCount FROM (SELECT D0sta FROM [Delivery Order] WHERE D0sta = "P" UNION ALL SELECT FPsta FROM [Offline Pickup Order] WHERE FPsta = "P") AS CombinedOrders;</pre>																									
<b>Result</b>																									
<pre>customer1: Find number of orders currently processing in store</pre> <table border="1"> <thead> <tr> <th>OngoingOrderCount</th> </tr> </thead> <tbody> <tr> <td>8</td> </tr> </tbody> </table>		OngoingOrderCount	8																						
OngoingOrderCount																									
8																									
<b>Customer2: Find the top 5 products</b>																									
<b>Code</b>																									
<pre>customer2: Find the top 5 products</pre> <pre>SELECT TOP 5 Product.PRID, Product.PRname, Product.PRname, Sum([Delivery Product Ordered].[OD&amp;Pquant]) AS QuantityOfSales FROM [Delivery Order] INNER JOIN (Product INNER JOIN [Delivery Product Ordered] ON Product.PRID = [Delivery Product Ordered].PRID ) ON [Delivery Order].ORDID = [Delivery Product Ordered].ORDID GROUP BY Product.PRID, Product.PRname, Product.PRname, [Delivery Product Ordered].PRID , [Delivery Product Ordered].PRID ORDER BY Sum([Delivery Product Ordered].[OD&amp;Pquant]) DESC;</pre>																									
<b>Result</b>																									
<pre>customer2: Find the top 5 products</pre> <table border="1"> <thead> <tr> <th>PRID</th> <th>Expr1001</th> <th>PRname</th> <th>QuantityOfSales</th> </tr> </thead> <tbody> <tr> <td>P100018</td> <td>Blueberry Muffin</td> <td>Blueberry Muffin</td> <td>98</td> </tr> <tr> <td>P100001</td> <td>Tapioca Milk Tea</td> <td>Tapioca Milk Tea</td> <td>94</td> </tr> <tr> <td>P100019</td> <td>Chocolate Croissant</td> <td>Chocolate Croissant</td> <td>85</td> </tr> <tr> <td>P100011</td> <td>Cappuccino</td> <td>Cappuccino</td> <td>80</td> </tr> <tr> <td>P100021</td> <td>Cranberry Scone</td> <td>Cranberry Scone</td> <td>70</td> </tr> </tbody> </table>		PRID	Expr1001	PRname	QuantityOfSales	P100018	Blueberry Muffin	Blueberry Muffin	98	P100001	Tapioca Milk Tea	Tapioca Milk Tea	94	P100019	Chocolate Croissant	Chocolate Croissant	85	P100011	Cappuccino	Cappuccino	80	P100021	Cranberry Scone	Cranberry Scone	70
PRID	Expr1001	PRname	QuantityOfSales																						
P100018	Blueberry Muffin	Blueberry Muffin	98																						
P100001	Tapioca Milk Tea	Tapioca Milk Tea	94																						
P100019	Chocolate Croissant	Chocolate Croissant	85																						
P100011	Cappuccino	Cappuccino	80																						
P100021	Cranberry Scone	Cranberry Scone	70																						

**Figure 3.1**

### 3.2 Staff

Targeting the needs of the staff group to operate the business, three SQL queries are designed. Firstly, it helps staff to count the omnichannel revenue of a particular store on a particular day, playing the role of reconciliation. The second one is used to count the number of returning customers, which contributes to judging customer retention ability and loyalty at a certain period. The last one could provide inventory warning by listing material whose quantity is below 20kg, which is crucial to inventory operations.

<b>Staff1: Calculate certain store revenue on certain day</b>	
<b>Code</b>	

**staff1: Calculate certain store's revenue on certain day**

```

SELECT STOID, Transaction_Date, SUM(Sale_Amount) AS Total_Revenue
FROM (SELECT STOID, D0cd AS Transaction_Date, D0price AS Sale_Amount
      FROM [Delivery Order]
     WHERE STOID = "ST0001" AND D0cd = #2024/4/1#
UNION ALL
      SELECT STOID, FPcd AS Transaction_Date, FPprice AS Sale_Amount
      FROM [Offline Pickup Order]
     WHERE STOID = "ST0001" AND FPcd = #2024/4/1#
) AS CombinedSales
GROUP BY STOID, Transaction_Date;

```

**Result**

STOID	Transaction_Date	Total_Revenue
ST0001	2024/4/1	400

**Staff2: Calculate the number of repeat customers**

**Code**

```

staff2: Calculate the number of repeat customers
SELECT COUNT(CUSID) AS Returning_Customers
FROM (SELECT CUSID
      FROM [Delivery Order]
     GROUP BY CUSID
    HAVING COUNT(*) > 1
UNION ALL

      SELECT CUSID
      FROM [Offline Pickup Order]
     GROUP BY CUSID
    HAVING COUNT(*) > 1 ) AS ReturningCustomers;

```

**Result**

Returning_Customers
223

**Staff3: Find out Materials whose quantity < 20kg**

**Code**

```

staff3: Find out products whose quantity < 20 kg
SELECT [MID], SUM([IVLGquan]) AS TotalQuantity
FROM [Inventory Log]
GROUP BY [MID]
HAVING Sum([IVLGquan])<20;

```

**Result**

MID	TotalQuant
M100012	14
M100016	7
M100018	7
M100020	0
M100041	5
M100044	17
M100049	10
M100051	18
M100055	3
M100056	6
M100078	15
M100088	4

Figure 3.2

### 3.3 Suppliers

The supplier community plays an important role in the supply of raw materials to the beverage store. Suppliers can query which inventory warehouse the materials are shipped into with the first SQL. Besides, the second SQL allows suppliers to calculate their revenues for providing certain raw materials.

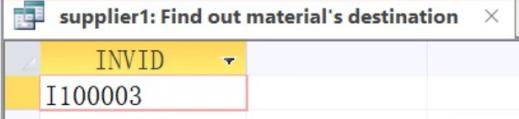
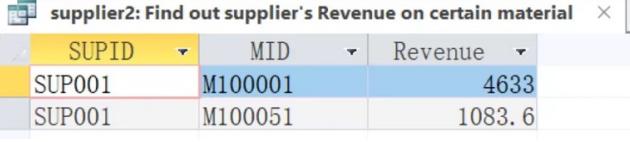
Supplier1: Find out material's destination										
<b>Code</b>										
	 <pre>supplier1: Find out material's destination SELECT [Inventory Log].INVID FROM [Inventory Log] WHERE ((([Inventory Log].MID) = "M100001")) GROUP BY INVID</pre>									
<b>Result</b>										
	 <table border="1"><thead><tr><th>INVID</th></tr></thead><tbody><tr><td>I100003</td></tr></tbody></table>	INVID	I100003							
INVID										
I100003										
Supplier2: Find out supplier's revenue on certain material										
<b>Code</b>										
	 <pre>supplier2: Find out supplier's Revenue on certain material SELECT m.SUPID, m.MID, m.Mprice*(SELECT SUM(IVLGquan)     FROM [Inventory Log]     WHERE MID = m.MID AND IVLGquan &gt; 0) AS Revenue     FROM Material AS m     WHERE m.SUPID = "SUP001";</pre>									
<b>Result</b>										
	 <table border="1"><thead><tr><th>SUPID</th><th>MID</th><th>Revenue</th></tr></thead><tbody><tr><td>SUP001</td><td>M100001</td><td>4633</td></tr><tr><td>SUP001</td><td>M100051</td><td>1083.6</td></tr></tbody></table>	SUPID	MID	Revenue	SUP001	M100001	4633	SUP001	M100051	1083.6
SUPID	MID	Revenue								
SUP001	M100001	4633								
SUP001	M100051	1083.6								

Figure 3.3

### 3.4 Third-party logistics provider

Two SQLs were designed for the third-party platform checking and working with the beverage store delivery service. The first SQL can be applied to the platform's performance review of delivery person by listing delivery person with more than two orders that were not delivered on time (which is shown as an abnormal order in the database). The second SQL performs the task of counting the total number of orders delivered by a certain platform's total delivery person.

<b>Logistic1: Find the delivery person &gt;=2 abnormal order record</b>							
<b>Code</b>							
	<pre>logistic1: Find the delivery person &gt;= 2 abnormal order record SELECT DPID FROM [Delivery Order] WHERE [Dosta] = "A" GROUP BY DPID HAVING COUNT(*) &gt;= 2;</pre>						
<b>Result</b>							
	<table border="1"> <thead> <tr> <th>DPID</th> </tr> </thead> <tbody> <tr> <td>DP013</td> </tr> </tbody> </table>	DPID	DP013				
DPID							
DP013							
<b>Logistic2: Find the order quantity for delivery platform</b>							
<b>Code</b>							
	<pre>Logistic2: Find the order quantity for delivery platform SELECT     dp.LPID AS LPID,     COUNT(do.DPID) AS Total_Orders FROM     [Delivery Order] do INNER JOIN     [Delivery Person] dp ON do.DPID = dp.DPID GROUP BY     dp.LPID</pre>						
<b>Result</b>							
	<table border="1"> <thead> <tr> <th>LPID</th> <th>Total_Orde...</th> </tr> </thead> <tbody> <tr> <td>LP001</td> <td>112</td> </tr> <tr> <td>LP002</td> <td>188</td> </tr> </tbody> </table>	LPID	Total_Orde...	LP001	112	LP002	188
LPID	Total_Orde...						
LP001	112						
LP002	188						

Figure 3.4

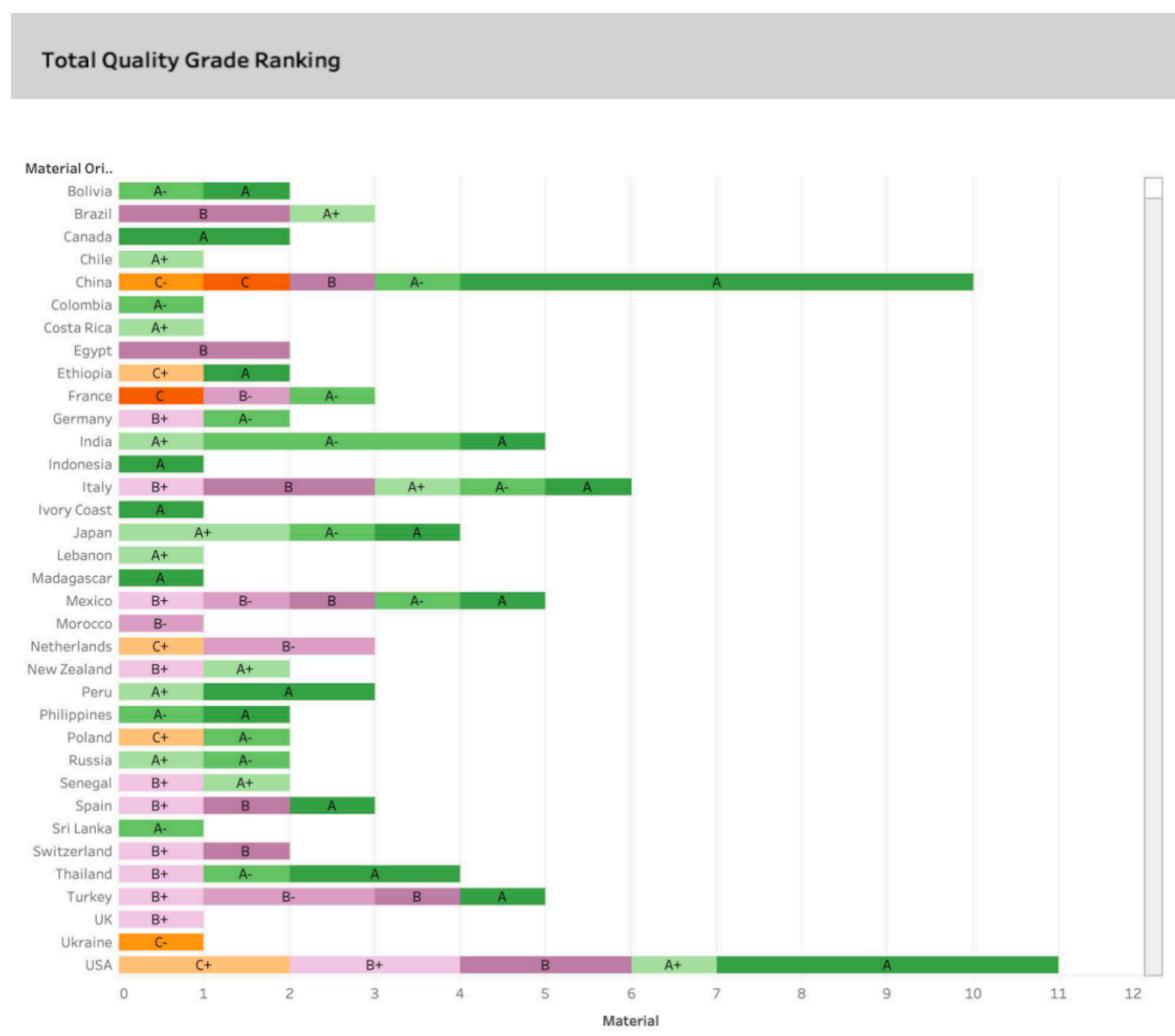
## 4. Visualization

This report presents a data visualization of the material supply situation, user distribution, and sales analysis of the beverage shop chain from 1 April 2024 to 7 April 2024 by using Tableau. It aims to help staffs and managers understand the current state of the company's business more intuitively, identify potential problems, and provide data support for future decision-making.

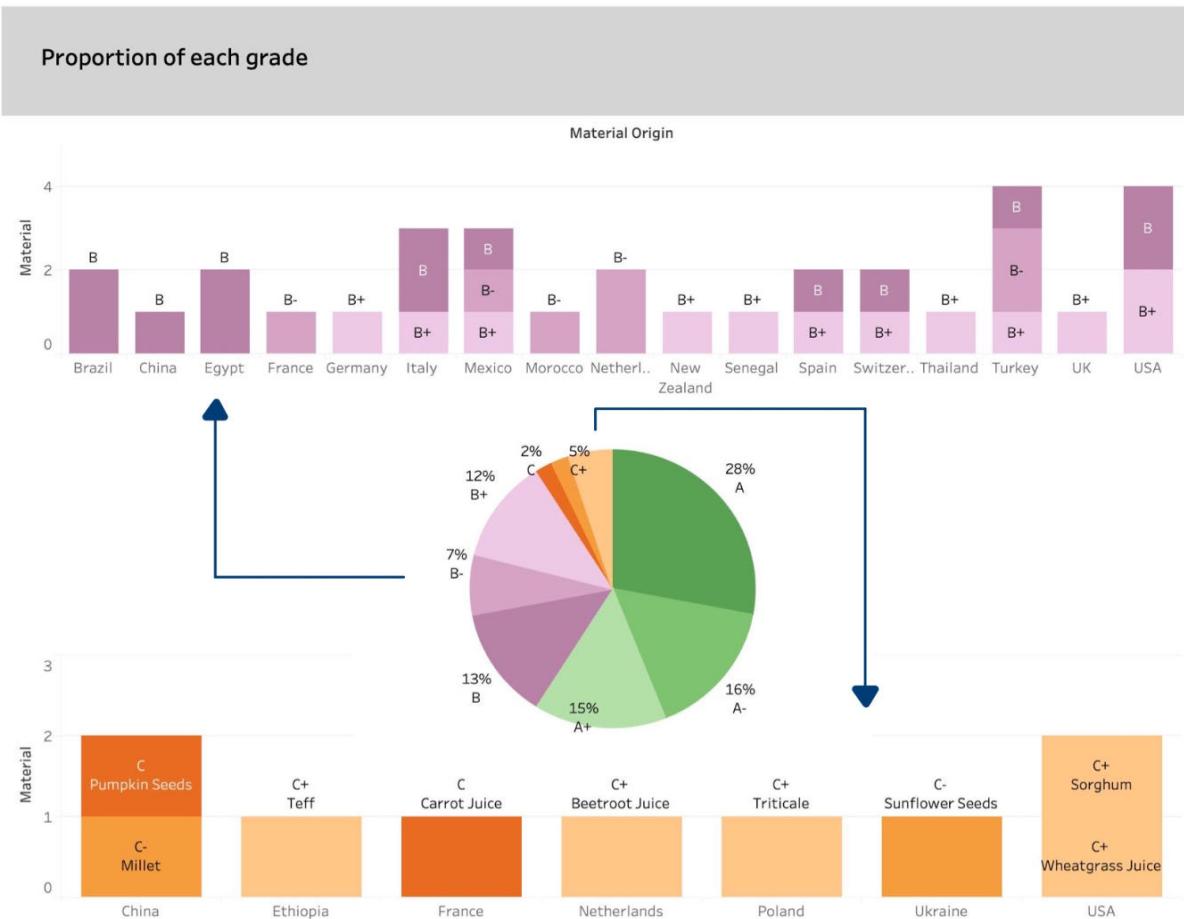
### 4.1 Material supply situation

In Figure 4.1, the beverage chain's materials are sourced from a wide range of locations, with China and the United States standing out in terms of volume. In Figure 4.2, 59% of the materials are in the A rating, 32% in the B rating and 9% in the C rating. Therefore, managers should focus on further improving and

upgrading the materials in the B rating. The materials in the C rating should be screened out or changed purchasing place for rectification and monitoring, aiming to reduce the number of C-rated materials as much as possible to less than 3%.



**Figure 4.1**



**Figure 4.2**

## 4.2 Customer portrait

Observing the distribution of customers' age and gender according to Figure 4.3, the main customer group is young females, concentrated in 14-29 years old. Besides, most of the customers have membership levels 1-4, which indicates that customers tend to become members and upgrade. Observing the returned customers of the cross-channel orders according to Figure 4.4, online sales channels have more returned customers and the number of returned females are more than the number of returned males. The results assist the refinement of customer profiling. The chain's marketing strategy can be skewed towards a younger female demographic and the online ordering service can become a business focus to support the increased business.

Membership age and gender distribution

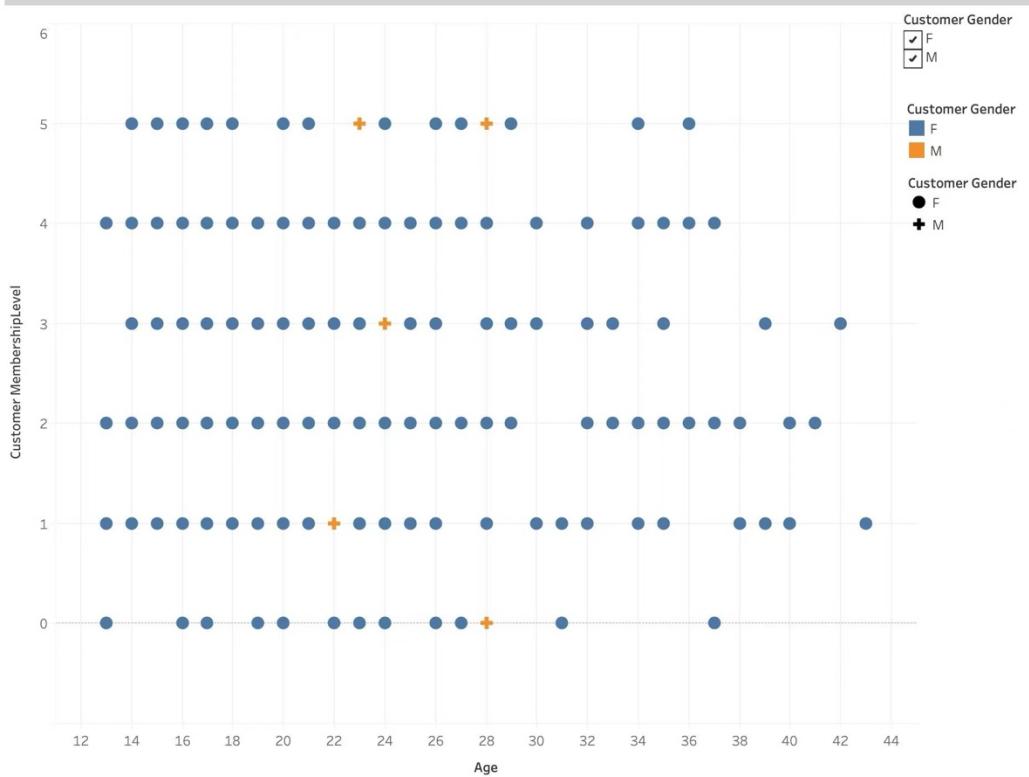


Figure 4.3

Returned customer

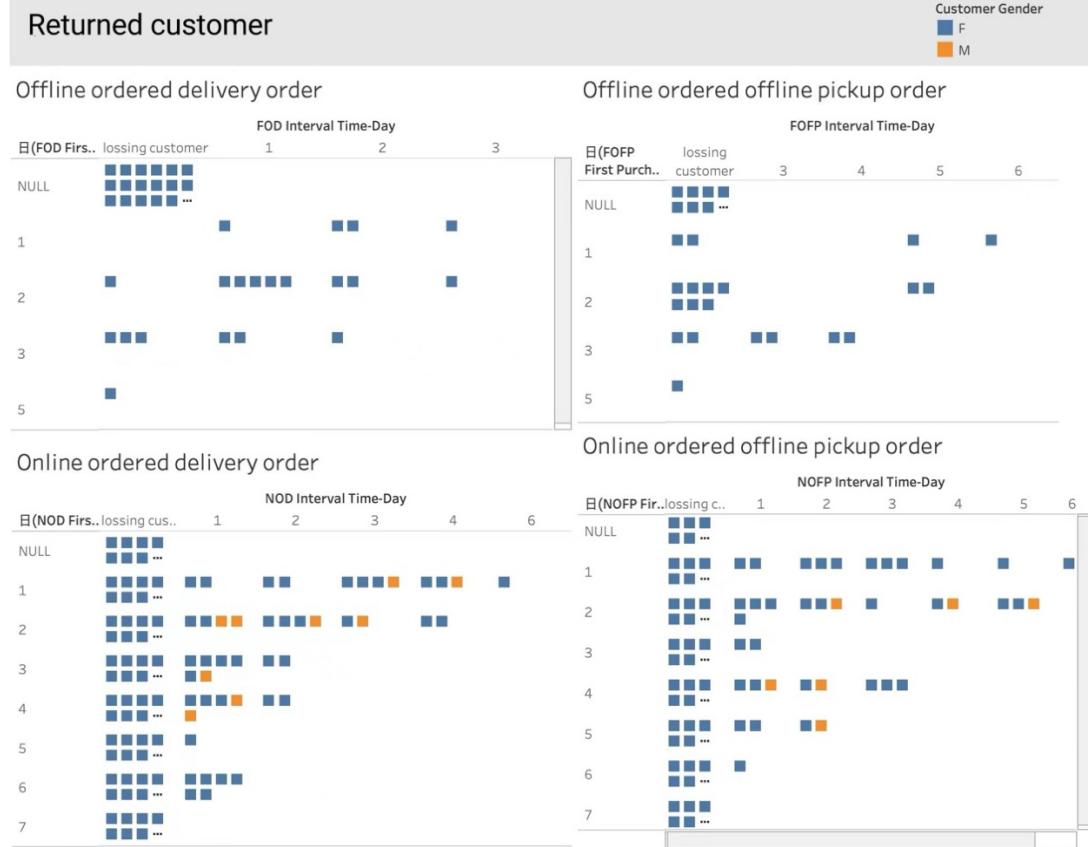


Figure 4.4

## **4.3 Sales analysis**

### **4.3.1 Store performance**

Figure 4.5 presents a comprehensive overview of total sales performance and each store's performance from 4.1 to 4.7, specifically during the hours of 9 am to 20 pm. A cursory glance at the TOTAL SALES reveals a distinct pattern: weekday sales generally surpass weekend sales. Stores' performance also proves this opinion, which is shown in the peaks and weekdays over weekends in the store profit area graph. Delving deeper into the temporal distribution, sales peaks appear at 9 a.m. and 18 p.m., which could be defined as commuting hours. Based on these manifestations, it could be assumed the stores' location plays a significant role in sales peak, given that most of our stores are situated in proximity to office buildings across Yinzhou District, Haishu District, and Beilun District.

Stepping into analyzing the sales performance of each store, the sales figures across stores have clear differences. The timing of peak sales varies from store to store. Overall, with Store001 standing out, Store002 and Store003 register the lowest sales, amounting to only half of Store001's sales. If this trend persists, we may need to reevaluate our store locations. One potential strategy could be to consider closing Store003 and exploring the possibility of opening a new store in a location proximate to Store001, given its impressive sales performance. This could potentially help us capitalize on the strong sales momentum in that area and further optimize our store network.



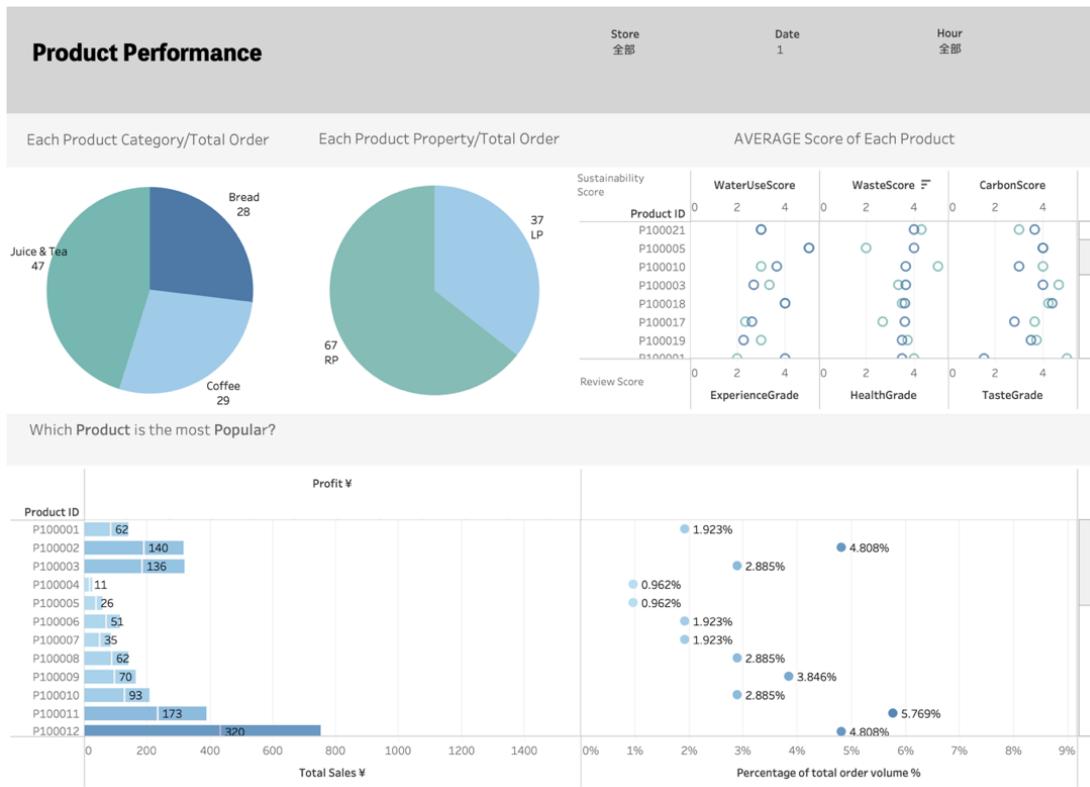
**Figure 4.5**

#### 4.3.2 Product performance

Figure 4.6 shows the product performance sold in our stores. Using the sales data from April 1<sup>st</sup> as an example, we can draw several key insights. Firstly, in terms of product categories, Juice & Tea emerges as the top-selling item. When it comes to product property, limited-edition products prove to be popular. Remarkably, when each store offers only one limited-edition product, it accounts for nearly 40% of the total orders, indicating a high demand for exclusive and limited-time offerings. Based on these observations, we could consider exploring collaborations with more popular IPs (Intellectual Properties) and introducing more limited-edition products to further boost sales.

On the other hand, the last four latitude specifically score between 4 and 5, stand out. This indicates that while our products are widely recognized and appreciated, while paying attention to sustainability. There are still areas for improvement in terms of customer satisfaction. For instance, we should focus on enhancing customer experience, such as service quality, store environment to ensure that we are meeting the expectations of our

customers and delivering a comprehensive experience.



**Figure 4.6**

## 5. Business intelligence

Database-based Business Intelligence (BI) is also an important part of the overall company strategy, which helps the organization to make better decisions. In this report, Regression and Classification, which are two typical BI tools, will be applied to the company for detailed analysis.

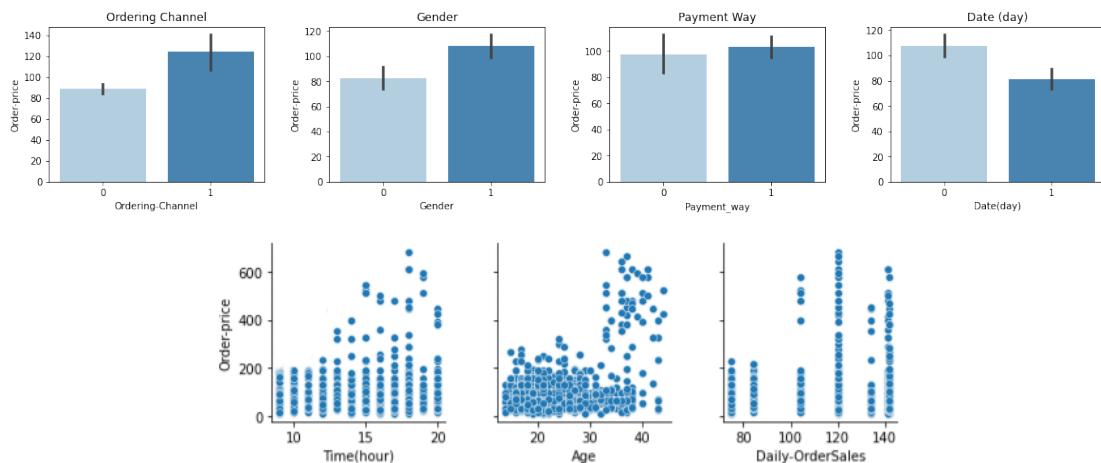
### 5.1 Regression

Regression analysis can help the company to explore the multiple factors that affect the price of each order. Firstly, all the elements in the database related to order price are selected as influential factors that may affect the order price (dependent variable Y). As shown in Figure 5.1, we select seven relevant entries. Then, view each of the seven entries individually in relation to the order price (Figure 5.2). '*Ordering-channel*', '*gender*', '*age*', and '*Date(day)*' can be seen in the scatterplot or box plot with a pronounced linear relationship with price. '*Time(hour)*' presents a seemly linear relationship with price. However, there seems to be a non-linear relationship between '*Payment\_way*' and '*Daily-*

*OrderSales*' with the price per order. Therefore, the variables that are relatively obviously linearly related in the graph were used as independent variables (X) for modeling regression analysis (Figure 5.3).

<b>Relevant Entity (Possible independent variables)</b>	<b>Description</b>
Ordering-Channel	Dummy variables: Online ordering = 0, offline ordering = 1;
Gender	Dummy variables: Male = 0, Female = 1;
Payment_way	Dummy variables: Online payments = 0, offline payments = 1;
Date(day)	Dummy variables: Working days = 0, non-working days = 1
Time(hour)	Time(hour) to create the order;
Age	Age of customers;
Daily-OrderSales	Number of orders per day.

**Figure 5.1**



**Figure 5.2**

<b>Dependent Variable (Y)</b>	<b>Independent variable (X)</b>
Order-price	Ordering-Channel
	Date(day)
	Time(hour)
	Age
	Gender

**Figure 5.3**

Since there are multiple independent variables and each variable does not have covariance (all VIF<5), Ordinary Least Squares (OLS) regression can be

used. From the results of the regression model (Figure 5.4), the p-values for 'const', 'Ordering-Channel', 'Age' and 'Gender' all have p-values less than 0.05, meaning that these coefficients are statistically significant. From the overall model assessment, the overall model is significant (the p-value corresponding to the F-statistic is small). Also, the residuals of our model are normally distributed (p-value of Omnibus near 1) and autocorrelated (Durbin-Watson near 2). Therefore, our OLS regression model and function (Figure 5.5) are relatively reliable.

OLS Regression Results						
Dep. Variable:	Order-price	R-squared:	0.793			
Model:	OLS	Adj. R-squared:	0.788			
Method:	Least Squares	F-statistic:	37.89			
Date:	Sun, 28 Apr 2024	Prob (F-statistic):	6.50e-35			
Time:	11:45:56	Log-Likelihood:	-4778.2			
No. Observations:	800	AIC:	9568.			
Df Residuals:	794	BIC:	9597.			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-67.2666	19.567	-3.438	0.001	-105.676	-28.857
Ordering-Channel	17.2511	7.103	2.429	0.015	3.309	31.193
Date(day)	-11.3101	8.525	-1.327	0.185	-28.044	5.423
Time(hour)	0.2847	0.973	0.293	0.770	-1.625	2.194
Age	6.1036	0.502	12.149	0.000	5.117	7.090
Gender	20.9581	8.220	2.550	0.011	4.823	37.093
Omnibus:	1.646	Durbin-Watson:	1.240			
Prob(Omnibus):	0.439	Jarque-Bera (JB):	1.017			
Skew:	0.726	Prob(JB):	0.601			
Kurtosis:	2.421	Cond. No.	329.			

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

**Figure 5.4**

$$\begin{aligned} \text{Order - price} = & -67.27 + 17.25 \times \text{Ordering - channel} - 11.31 \times \text{Date}(day) \\ & + 0.28 \times \text{Time}(hour) + 6.1 \times \text{Age} + 20.96 \times \text{Gender} \end{aligned}$$

**Figure 5.5**

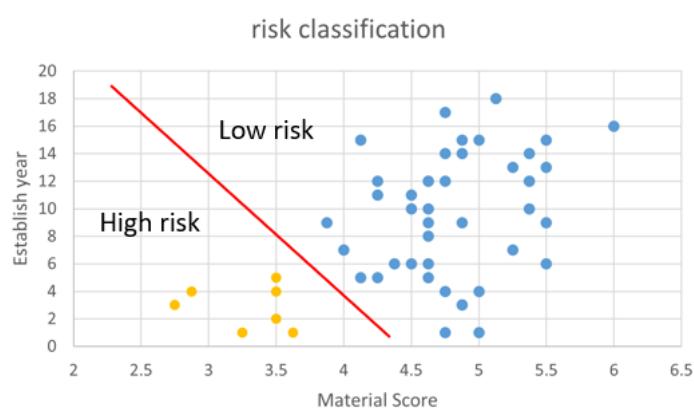
Finally, returning to the commercial application, despite the previously defined target group of young women, the company could tap into the market of middle-aged women by launching high-priced products with more expensive and high-quality materials. Secondly, as the order price of online ordering is lower than that of offline ordering channels, the company can set up special promotions such as "Drink with Friends" and "Group Purchase Set Meal" for online ordering channels. Moreover, for online ordering channels that require delivery services, the company can set up mechanisms such as full-value

reductions. Consequently, the price of each order on the online ordering channel can be increased, and higher revenues can be realized.

## 5.2 Classification

To evaluate the qualification of suppliers, we apply a classification methodology to verify the risk profile of suppliers on a regular basis. The year of establishment of the supplier in the database and the quality score of the material supplied can be used as a measure of the qualification of the supplier. Concerning the quality of materials, the higher the quality the higher the score. However, the variance is greater with lower scores. This is because lower quality materials are usually less expensive and do not have quality transportation. Their downgrading is more easily perceived by customers (e.g. spoilage problems). Therefore, downgrading may cause them more serious problems. As for the age of the supplier, it is reasonable to believe that the longer the age of the supplier, the more stable the supplier is in terms of reputation, business operation model and financial status, which means less risk of default.

The result (Figure 5.6) shows that suppliers with a shorter period of operation and supplying riskier materials are categorized as high default risk suppliers. Longer established suppliers are typically more experienced and stable, while suppliers that produce a higher percentage of poor-quality materials are more likely to be at higher risk of default due to the severe impact of degradation. The company can continue to use the classification model to work with low-risk companies on a long-term basis, while at the same time terminating contracts with high-risk companies in a timely manner.



**Figure 5.6**

## **6. Reflection: Advantage & Limitation**

Regarding the whole database solution, a relatively comprehensive framework has been built. The database involves the front, middle and back segments. The front side includes material supplements and inventory. Subsequently, it offers cross-channel sales and delivery service. Customer record system involving loyalty incentives, membership and order scoring mechanisms, allowing analyzing customer portrait, maintaining client relationships, and upgrading service reference. In general, the database includes a comprehensive supply chain, a robust sales system, and a stable pool of customers have been established.

However, there are also the following limitations. Currently, staff are recorded according to the store they belong to and their job level. As the company grows, there is a need to further refine the breakdown of staff by department and function. For example, more specialized positions such as administrative positions, inventory management positions, and marketing positions need to be identified to support the opening of more stores and increase revenue. Secondly, in terms of customer relationship maintenance, the current Reward mechanism is relatively simple, and a more complex REWARD mechanism combined with marketing may be needed subsequently. A feasible solution is to link membership level and reward, in addition to simple prize redemption, but also support different levels of discount. Additionally, the current order scoring is only limited to the given three latitude scores, which does not allow customers to comprehensively express their opinions about the consumption experience. To improve, it can be adjusted to allow customers to write literal type comments. Then more advanced technology, such as Python text analysis, could be used to capture the customer's consumption opinions, aiming to improve the quality of service.

## **7. Conclusion**

To conclude, after analyzing the beverage company's status and setting several business rules, a database system is built, taking staff, supplier, customer and logistic provider's views into consideration. It includes an ER diagram to show the relationship within entities and SQL queries are provided to satisfy users' needs. Visualization and BI analysis are then applied to make supplier situation, customer profile, store revenue, and product price factors analysis, which supports the determining of business focus and future strategic plan.

## Appendix

### Python code for BI regression

```
In [1]: import pandas as pd
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.api as sm
from statsmodels.stats.outliers_influence import variance_inflation_factor

In [2]: # Reading data from a CSV file
data = pd.read_csv('ddi_regression0428.csv')
print(data.head(5))
print(data.columns)

# Convert non-numeric columns to numeric types
data['Ordering-Channel'] = data['Ordering-Channel'].astype(int)
data['Payment_way'] = data['Payment_way'].astype(int)
data['Gender'] = data['Gender'].astype(int)
data['Date(day)'] = pd.to_numeric(data['Date(day)'], errors='coerce')
data['Time(hour)'] = pd.to_numeric(data['Time(hour)'], errors='coerce')
data['Daily-OrderSales'] = pd.to_numeric(data['Daily-OrderSales'], errors='coerce')
data['Order-price'] = pd.to_numeric(data['Order-price'], errors='coerce')

# Delete lines containing NaN
data.dropna(inplace=True)
...
```

```
In [3]: # Plotting the relationship between the independent and dependent variables

## scatterplot
sns.pairplot(data, x_vars=['Time(hour)', 'Age', 'Daily-OrderSales'], y_vars='Order-price', kind='scatter')
plt.subplots_adjust(wspace=0.3)
plt.show()

## Dummy variable box plot
blue_palette = sns.color_palette("Blues", n_colors=len(data['Ordering-Channel'].unique()))
fig, axes = plt.subplots(nrows=1, ncols=4, figsize=(20, 3))

sns.barplot(x='Ordering-Channel', y='Order-price', data=data, palette=blue_palette, ax=axes[0])
axes[0].set_title('Ordering Channel')

sns.barplot(x='Gender', y='Order-price', data=data, palette=blue_palette, ax=axes[1])
axes[1].set_title('Gender')

sns.barplot(x='Payment_way', y='Order-price', data=data, palette=blue_palette, ax=axes[2])
axes[2].set_title('Payment Way')

sns.barplot(x='Date(day)', y='Order-price', data=data, palette=blue_palette, ax=axes[3])
axes[3].set_title('Date (day)')

plt.subplots_adjust(wspace=0.2)
plt.show()
...
```

```
In [4]: # Calculate the correlation coefficient between the independent and dependent variables
correlation_matrix = data.corr()
print(correlation_matrix['Order-price'])

# Checking for multicollinearity
X = data[['Ordering-Channel', 'Date(day)', 'Time(hour)', 'Age', 'Gender']]
vif_data = pd.DataFrame()
vif_data["feature"] = X.columns
vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in range(len(X.columns))]
print(vif_data)

# Linear Regression Analysis
X = sm.add_constant(X)
y = data['Order-price']

model = sm.OLS(y, X).fit()
print(model.summary())
...
```