I. Introduction

Our client, Dacheng's Bistro, is a renowned culinary influencer in Taiwan. Dacheng's Bistro's career is centered around creating cooking and cuisine content. With a growing reputation, Dacheng's Bistro ventured into the business realm by establishing Dacheng's Bistro. This innovative venture initially focused on providing affordable instant food options to consumers.

Dacheng's Bistro specializes in a diverse range of instant food products, including items such as jams, noodles, and hotpot ingredients. The core of their business model revolves around a collaborative effort with specialized factories. Dacheng's Bistro, in close partnership with these facilities, engages in the development and refinement of product flavors. Once the flavor profiles are perfected and finalized, the factories commence production based on the specified quantities and specifications.

After product production begins, the products are promptly made available on their official website, allowing customers to place orders seamlessly. Customers are offered the flexibility of choosing between paying upon delivery or making an upfront payment. The company diligently monitors payment statuses and ensures that orders are dispatched.

In cases where customers fail to complete their payments, the order is canceled, and products returned to the warehouse. Additionally, Dacheng's Bistro places a strong emphasis on customer satisfaction, permitting customers to return products if they have any concerns or issues.

II. The Business Objectives

The business in question is a small-scale start-up and therefore had no formal system of organizing data. Instead, they have procured services from an ordering platform that enables them to capture essential order-related data, including details such as the ordered products and quantities, order timestamps, order values, payment statuses, delivery statuses, as well as information pertaining to cancellations or returns. We intend to provide them with:

1. Comprehensive Database Design: Develop a robust and tailored database structure that encompasses all aspects of Dacheng's Bistro's operations, including order data, product details, customer information, and transaction records.

2. Data-Driven Business Analysis: Utilize the available order data to conduct a series of analyses. This includes identifying the best-selling products based on both total quantities sold, and revenue generated, pinpointing peak selling times for these products, calculating return rates and order cancellation rates, and identifying products with the highest return or cancellation rates.

3. Customer Insights: Extract valuable insights from the data to better understand the target audience. This includes identifying and quantifying loyal customers who place orders more than three times a year and assessing whether the average order amount differs significantly between loyal customers and other customer segments.

By achieving these objectives, we aim to empower Dacheng's Bistro with actionable insights that will inform strategic decisions, enhance customer experiences, and contribute to the company's ongoing success and future product development efforts.

The business insights that will be provided are:

1. What is the best-selling category/item by amount sold?
2. What is the worst-selling category/item by amount?
3. What is the peak month for sales?
4. What percentage of payments are actually completed?
5. What is the average amount of money spent per customer?
6. What is the average order number of items bought per customer?
7. Who are the most loyal customers? (loyalty calculated based on RFM: recency, frequency and monetary)
8. Which promotions result in higher orders?
9. Which city has the most customers?
10. What is the percentage share of each delivery method?

III. Entity Relationship Diagram

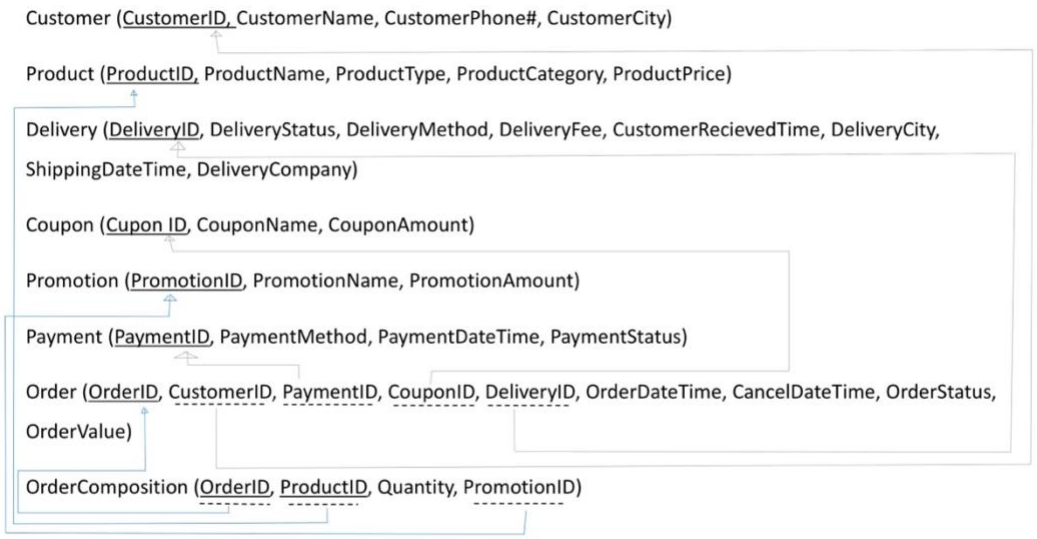
There are 8 entities in the ERD presented in rectangular shapes. Each entity will become a table in the final database.

1. Order
2. Product
3. Payment
4. Delivery
5. Customer
6. Coupon
7. Promotion
8. Composition (associative entity)

The attributes associated with each of these are presented in oval shapes and the relationships are presented in diamond shapes.
The main entities for Dacheng's Bistro are Order, Customer, and Product. The ERD is created based on saving all order related data in structural way. For the promotion is applied based on the product and the order quantity, while promotion is applied to entire order; the system of Dacheng's Bistro allows only one promotion for one order. To avoid sharing personal information confidentiality, we exclude the personal information in our dataset, but is registered in the actual order detail.

IV. 3 Normal Form Relational Schema

This relational schema has been adjusted to include only atomic values, have only full functional dependencies and no transitive dependencies. This will ensure that there are no anomalies in the final database.



The primary keys for each table are underlined in a continuous line and the foreign keys are underlined with dotted lines.

V. Analysis and Insights

The SQL file to create tables is submitted in SQL format. The queries that were run to get answers to the business questions are given below.

Q.1 What is the best-selling category/item by amount sold?

```sql
##Analysis
#1. What is the best-selling category/item by amount sold?
select distinct(a.product_category) as "Product Category", sum(c.TotalAmount) as "Total Sales"
from sc.products as a
join sc.order_composition as b
join sc.orders as c
on a.product_ID = b.product_id and b.order_id = c.OrderID
where (c.OrderStatus = "Closed") and (a.product_category is not null or a.product_category <> '')
group by a.product_category
order by sum(c.TotalAmount) desc
limit 2
offset 1; #There are blanks in product category since the business is new, so excluding the blank row
```

100% 1:16

Result Grid | Filter Rows: | Export: | Fetch rows:

Product Category	Total Sales
dried noodles	12037914
hot pot food	6044378

Considering the 2022 is at the same time where COVID19 still exist while the severity of it is declining, we can see people still store instant food, even hot pot which is usually eaten outside of the house for friend gatherings. The result definitely shows the change of dietary habits due to COVID19.

Q.2 What is the worst-selling category/item by amount?

```sql
##Analysis
#2. What is the worst-selling category/item by amount?
select distinct(a.product_category) as "Product Category", sum(c.TotalAmount) as "Total Sales"
from sc.products as a
join sc.order_composition as b
join sc.orders as c
on a.product_ID = b.product_id and b.order_id = c.OrderID
where (c.OrderStatus = "Closed") and (a.product_category is not null or a.product_category <> '')
group by a.product_category
order by sum(c.TotalAmount) asc
limit 2;
```

100% 1:32

Result Grid | Filter Rows: | Export: | Fetch rows:

Product Category	Total Sales
one-flavor jam	576106
mixed flavor jam	1095307

Jam is the first product developed and produced by Dacheng's Bistro which was always considered to be successful, but from the analysis we can see that Jam is worst selling Product category by Total Sales. Promotion campaigns can be conducted for Jam to boost up sales as it is the longest running product for Dacheng's Bistro

Q.3 What is the peak month for sales?

```
##Analysis
#3. What is the peak month for sales?
select distinct monthname(a.OrderDateTime) as "Month", year(a.OrderDateTime) as "Year", sum(a.TotalAmount) as "Total Sales"
from sc.orders as a
where a.OrderStatus = "Closed"
group by monthname(a.OrderDateTime), year(a.OrderDateTime)
order by sum(a.TotalAmount) desc
limit 5;
```

Result Grid | Filter Rows: | Export: | Fetch rows:

Month	Year	Total Sales
January	2022	3056367
April	2022	2836177
December	2022	1586188
November	2022	1575620
February	2022	1354856

For January the Bistro is usually preparing for the Lunar New Year and continues to do so until end of the Lunar New Year (Usually till February), where people tend to pile up on instant food, which helps boost business. Considering the best-selling products are hotpot items, it is not surprising that the winter season (from November to February) has the highest total sales.

For April, it is the start of the Spring and also time for Tomb Sweeping Festival, which usually prepared instant food such as noodle to worship the ancestors.

Q.4 What percentage of payments are actually completed?

```
SET @TotalOrderCount = (SELECT COUNT(DISTINCT OrderID) FROM orders);

SELECT OrderStatus, (COUNT(DISTINCT OrderID))*100/@TotalOrderCount AS OrderPercentage
FROM orders
GROUP BY OrderStatus;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

OrderStatus	OrderPercentage
Canceled	11.5944
Closed	88.4056

There is only ~12% orders that get cancelled or are left incomplete. There's high percentage of completion for orders placed with Dacheng's Bistro.

Q.5 What is the order amount (monetary) per customer?

```sql
SELECT SUM(TotalAmount)/COUNT(DISTINCT CustomerID) AS AverageRevenuePerCustomer
FROM orders
WHERE OrderStatus = "Closed";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

AverageRevenuePerCustomer
2491.7651398708886

Considering the bestselling product is the instant noodles which are low price items, it is not surprising that the annual average monetary value per customer is low. It is suggested that we can consider to provide higher value products to boost the business.

Q.6 What is the average order amount (items) per customer?

```sql
SELECT COUNT(B.ProductID)/COUNT(DISTINCT A.CustomerID) AS AverageProductsPerCustomer
FROM orders AS A
LEFT JOIN order_composition AS B ON A.OrderID = B.OrderID
WHERE OrderStatus = "Closed";
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

AverageProductsPerCustomer
3.7456

Q.7 Get the most loyal customer list

```sql
INNER JOIN TopRevCustomers AS B ON A.CustomerID = B.CustomerID
INNER JOIN TopOrderCustomers AS C ON A.CustomerID = C.CustomerID)

# Loyal Customer Info
SELECT A.CustomerID,
IF(CustomerName = '', "Unknown", CustomerName) AS CustomerName,
IF(CustomerCity = '', "Unknown", CustomerCity) AS CustomerCity,
IF(CustomerPhoneNumber = '', "Unknown", CustomerPhoneNumber) AS CustomerPhoneNumber
FROM LoyalCustomers AS A
INNER JOIN Customer AS B ON A.CustomerID = B.CustomerID;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

CustomerID	CustomerName	CustomerCity	CustomerPhoneNumber
10536079	Unknown	Unknown	Unknown
11056222	Unknown	nantou county	Unknown
11561489	Unknown	Unknown	Unknown
12094756	Unknown	miaoli county	Unknown
12387105	Unknown	new taipei city	Unknown
12484779	Unknown	new taipei city	Unknown
6892204	Unknown	hsinchu city	Unknown

Result 4

With this data, we can decide which customers to provide loyalty benefits to.

Q.8 Which promotions result in higher orders?

```sql
#8
select sum(quantity) as order_amount,PromotionName from order_composition as t1
join promotion as t2
on t1.PromotionID = t2.PromotionID
group by PromotionName
order by order_amount desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

order_amount	PromotionName
49647	Choose 20 packs of dried noodles for $850 or 3...
28499	Anniversary Event-Choose 27 packs of dried no...
7388	1111-Choose 27 packs of noodles for $1111
2148	1111-Choose 8 boxes of hot pot food for $1111
469	Anniversary Event-Choose 4 jars of jam for $999
271	1111-Choose 5 jars of jam for $1111
248	Grouponly-Choose 27 packages of noodles for ...
72	Grouponly-Choose 8 boxes of hot pot food for ...
15	Grouponly-Choose 5 jars of jam for$1111

Q.9 Which city has the most customers?

```sql
#9
select count(CustomerID) as customer_amount,member_city from orders as t1
left join customer as t2
on t1.CustomerID = t2.member_id
group by member_city
order by customer_amount desc
limit 1
offset 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

customer_amount	member_city
1882	new taipei city

Dacheng is from Taipei City, which is close to New Taipei City. Meanwhile, New Taipei City has a larger population than Taipei, which indicates where he has more social media influence, his product sells better.

Q.10 What is the percentage share of each delivery method?

```
18    #10
19 •  select DeliveryMethod,count(DeliveryMethod) as total,
20    (count(DeliveryMethod)*100)/(select count(*) from delivery) as percentage
21    from delivery
22    group by DeliveryMethod
23    order by percentage desc;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

DeliveryMethod	total	percentage
7-11cash on delivery	4301	47.1188
home delivery(main island)	4005	43.8760
7-11 pick up	777	8.5123
frozen home delivery(main island)	28	0.3067
home delivery(outer islands)	15	0.1643
t-cat delivery	2	0.0219

Considering the density of convenience store (7-11) in Taiwan it is not surprising that people tend to select to pick up their order in the store. However, from the results we can see that if the customers are paying up-front, they tend to have the order delivered directly to their home.

Combining with question 6, if we want to avoid cancel rate, we can take out the 7-11 cash on delivery to make sure all customers paid up-front thus avoid the cancel from not paying

VI. Conclusion

In this project, we have created a database for our client Dacheng's Bistro to be able to store his business data in an organized fashion. We have also come up with relevant business questions that are answered through this analysis of this database. The insights generated will allow Dacheng to run his business in a data-informed way and thereby his business will have a better chance at growth and success.