Guided by Professor:
Sreeram Chavali

Submitted By Group 1:
Huixin Li
Alex Soek
Mengzhen (Jane) Wang
Zhaoyi Zhang

Course: MSIS 2621 Business Intelligence and Data Warehousing

Phase 1:
Business Analysis

Business Analysis Project

Business Scenario:

Boba Time is a tea drink franchise founded in 2019 in Bay Area, California. Boba time expanded to northern California by the end of 2019, and by 2020 had further opened more locations in southern California. Since the company has expanded to the majority of cities in California and the district manager wants to have a central data warehouse to generate revenue reports and other key operational metrics to track the business. With accountable reports, the district manager can figure out strategies to increase regional revenue and profitability.

Purpose of the Project:

- Expand business and increase revenue
- Know how the business is doing in each store
- Check whether certain strategies are working
- What we can improve on our products, i.e. boba drinks
- Know our members

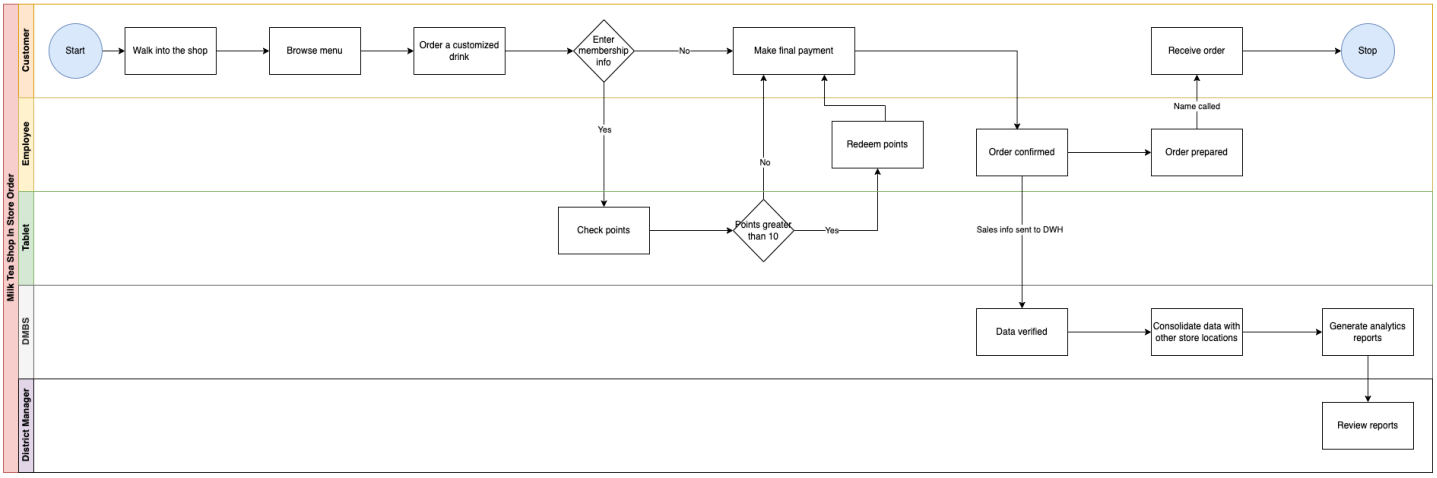
Business Process and Swim Lane Diagram:

Customers can order drinks through either employers or self-service kiosks. The process of ordering between these two methods has subtle discrepancies.

- In-store order:
  The customer walks into a boba shop and walks to the front of the store. After that, the customer browses through the menu and decides what he wants. Then the customer selects the toppings, sugar level, ice level, and milk preference. Before making the payment, if the customer has enrolled in membership, he could type his information on the machine (tablet on the desk) to earn points. If there are 10 points, the customer can pick one drink for free. Otherwise, the customer will continue to check out with either card or cash. Then the customer just needs to wait until his name is called and gets his drink. The order and membership information will be sent straight to the data warehouse. That information will be verified and transformed before entering the data warehouse. Those data will be summarized and analyzed every day to send to each store. The district management team will also get a copy of the report. After seeing how different metrics perform in each store, some business processes might be changed to achieve business requirements.

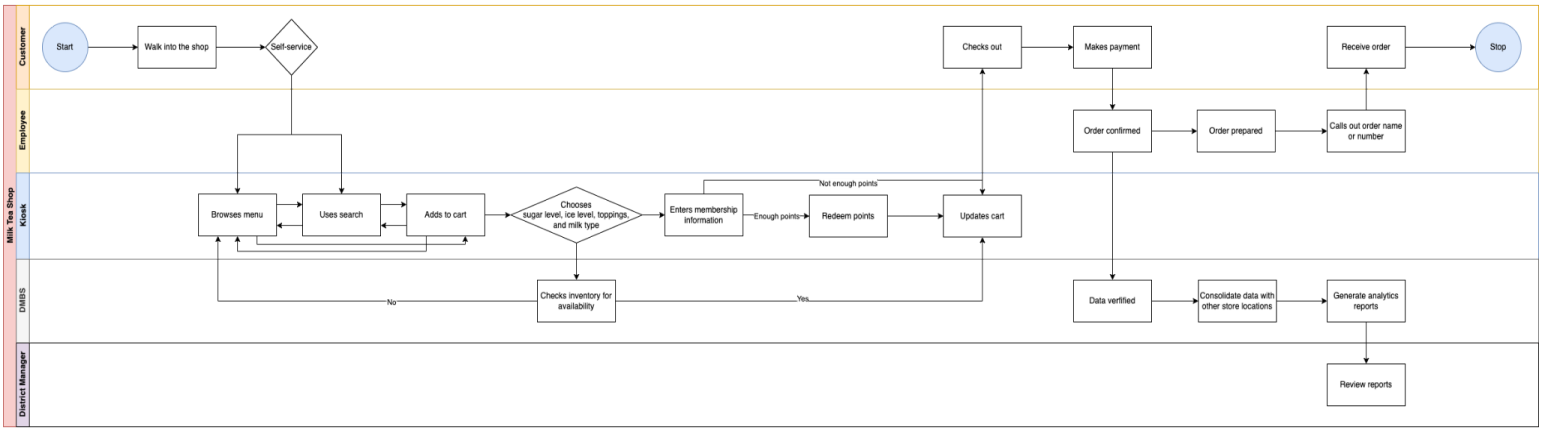
It's always a great time for boba!

Swim Lane Diagram for in-store order



- Self-service order:
  The customer decides to order at the kiosk. Then browses through the menu and uses the search function if the client wants anything in particular. Picks the items, and adjusts sugar, ice level, toppings, and milk alternatives. Before paying, the customer can log into the membership. If there are 10 points, he can pick one drink for free. Otherwise, proceed to check out. Pick a way to pay, either by card or apple pay. The customer then waits for his name to be called. The order and membership information will be sent straight to the data warehouse. That information will be verified and transformed before entering the data warehouse. Those data will be summarized and analyzed every day to send to each store. The district management team will also get a copy of the report. After seeing how different metrics perform in each store, some business processes might be changed to achieve business requirements.

Swim Lane Diagram for self-service order



Stakeholders:
Store Managers, District Managers, Analytics team, Customers

Lead Measures:
The number of orders per day/week/month
The number of orders per store location
Average price per order
The number of orders made by members
Total revenue by age groups

Lag Measure:
Revenue growth
Membership enrollment improvement
Text opt-in promotion effectiveness

Descriptive analytics:
1. What is the daily/monthly/annual revenue?
2. What is the daily order number?
3. What are the top 5 most popular drinks?
4. What are the least popular drinks?
5. What is the busiest hour of a day?
6. What does the demographic look like for the membership(e.g. Sex, enrollment time)
7. Which store location has the highest number of orders?
8. Which store location has the highest number of members?
9. What day of the week has the most revenue?
10. Average price per order?
11. What are the different payment types? And the fees associated with it.
12. What are the most and least popular drinks in the region?
13. What is the avg visit per member per month?
14. What is the total membership enrollment per year?
15. How many members accept the text option? And the total sales associated with it.
16. What are the total sales contributed by customers of two genders?
17. What are the total sales contributed by customers of different age groups?

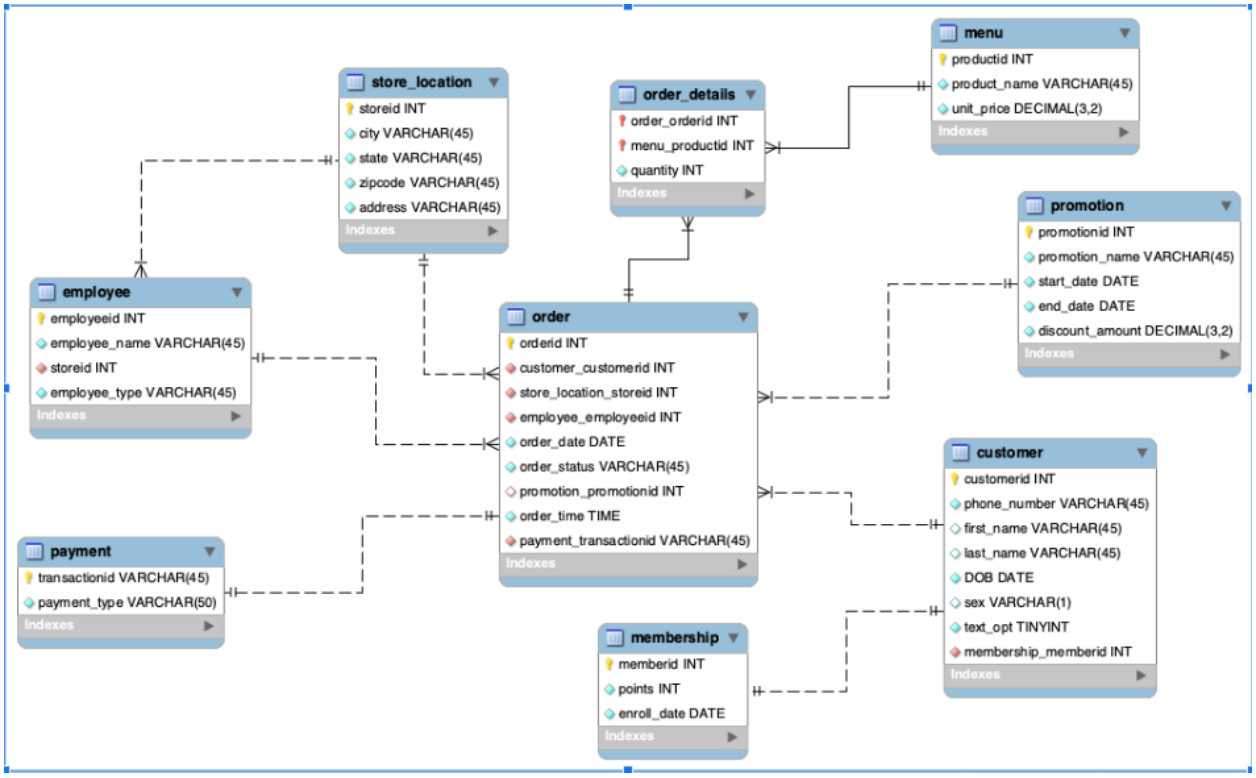
Predictive analytics:
1. Forecast which month will have the most revenue.
2. Forecast whether removing some items from the menu will increase revenue.
3. Forecast where to open new stores or shut down existing stores can increase revenue.
4. Forecast new members enrolled in the region

Key Takeaway:
During the project brainstorming, we opted to visit several boba shops rather than merely visualizing the business process mentally. We observed customers placing in-store orders with staff and even tried out a self-service kiosk, as well as attempting to sign up for a membership to understand the enrollment process. By immersing ourselves in these experiences and

observations, we gained a comprehensive grasp of the entire business operation, which made it much easier for us to create a swim lane diagram.
We conducted research on sales reports relating to food franchises for the purpose of developing descriptive and predictive analytics. Generally, such reports consist of elements such as monthly revenue, product category-based sales, best-selling items, customer demographic data, and marketing and promotional efforts. Utilizing this information, we successfully devised both descriptive and predictive analytics for our project.

Phase 2:
Data Modeling

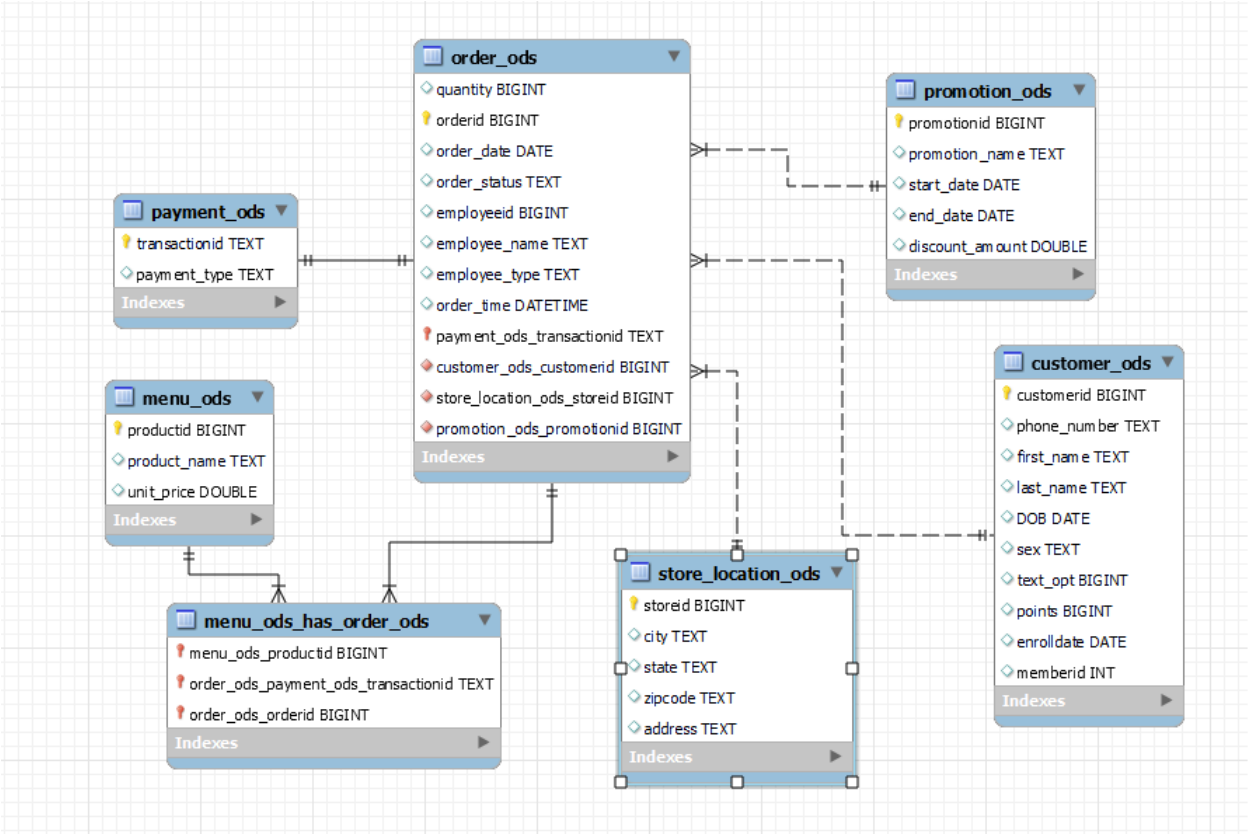
OLTP Schema:
The OLTP schema is designed to have 9 tables: payment, employee, store_location, order_details, order, membership, customer, promotion and menu.
Screenshot of OLTP Data Model:



ODS Schema:

In ODS Schema, there are 7 tables: payment, menu, menu_order, order, store_location, promotion and customer. Customer and Membership from OLTP schema are merged into one table, and merging order, oder_detail, payment, menu and employee to make order_ods.

Screenshot of ODS Data Model:

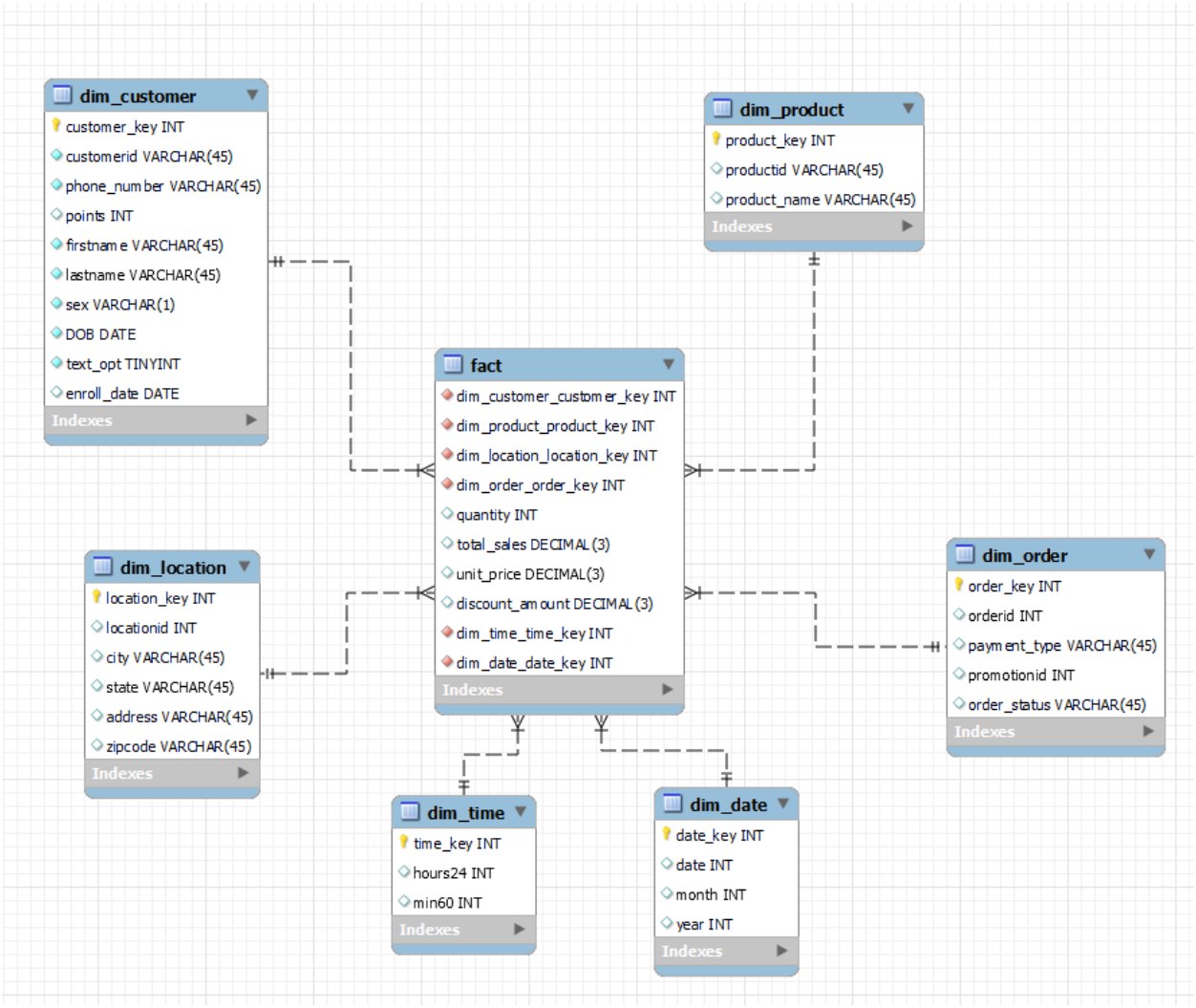


Data Warehouse Schema:

There are 6 dimension tables and one fact table. The dimension tables are:

1. Customer
2. Product
3. Order
4. Location
5. Date
6. Time

Screenshot of OLAP Data Model:



Key Takeaways:

When we designed the STAR schema and OLTP schema at first, we missed a few columns of information that we needed. But as we started generating data and looking at the business questions, we realized that there are more attributes needed. So we think It is important to always

go back and check whether our architecture and data model is complied with the business requirements

Phase 3:
ETL Implementation

Data Source used for implementation:

1. MySQL Database
2. CSV
3. Excel

Data Sets: All tables are auto generated from generatedata.com and Excel

Website: Generate.com

Part1: Extracting data from different data sources
We imported the Excel files into MySQL workbench

Customer OLTP table:

Result Grid | Filter Rows: | Edit: | Export/Import: | Wrap Cell Content:

customerid	phone_number	first_name	last_name	DOB	sex	text_opt
1	1234567890	Jane	Wang	2001-01-01	F	1
2	4472066910	Naida	Wagner	1989-05-15	F	1
3	2713733514	Kaseem	Bennett	1994-03-26	M	1
4	5602533437	Rama	Knapp	1982-07-11	F	0
5	2564864540	Anthony	Anthony	1978-09-26	M	1
6	4386181909	Rahim	Daniel	2000-12-29	M	0
7	4126071215	Cruz	Glover	2010-01-10	M	0
8	9019953333	Maile	Bowman	2011-06-29	M	1
9	4105619197	Stephen	Rasmussen	1991-03-12	M	1
10	5727538472	Ryan	Jensen	1999-01-30	M	1
11	3568487205	Stacey	Bennett	1986-07-15	M	0
12	9276503237	Jordan	Owen	2012-08-26	F	0
13	3506014592	Peter	Fleming	1984-01-29	M	1
14	5091693592	Megan	Rich	1993-11-28	M	0
15	5945067255	Kibo	Ryan	2011-09-20	F	0
16	7918078197	Brynn	Berg	1998-06-09	M	0

Employee OLTP table:

employeeid	employee_name	storeid	employee_type
1	kiosk	20232	self-service
2	kiosk	20233	self-service
3	kiosk	20234	self-service
4	kiosk	20235	self-service
5	kiosk	20236	self-service
6	kiosk	20237	self-service
7	kiosk	20238	self-service
8	kiosk	20232	self-service
9	Jena	20233	manager
10	Donovan	20234	manager
11	Macon	20235	manager
12	Althea	20236	manager
13	Anjolie	20237	manager
14	Rooney	20238	manager
15	Kermit	20232	manager
16	Donna	20233	manager

Membership OLTP table:

customerid	points	enrolldate	memberid
1	4	2020-11-18	15
2	13	2022-08-17	74
3	6	2020-08-11	11
4	8	2021-10-24	70
5	12	2022-06-28	NULL
6	5	2020-07-18	2
7	12	2020-04-02	54
8	10	2021-08-20	NULL
9	10	2022-08-20	66
10	10	2020-07-19	27
11	15	2020-07-11	38
12	4	2022-10-28	14
13	12	2023-01-14	NULL
14	14	2023-01-21	80
15	0	2022-11-28	53
16	13	2021-02-02	NULL

Menu OLTP table:

productid	product_name	unit_price
1001	Boba Milk Tea	5.95
1002	Almond Milk Tea	5.95
1003	Black Tea	5.55
1004	Chocolate Milk Tea	6.50
1005	Coconut Milk Tea	5.95
1006	Coconut Pineapple Tea	6.65
1007	Fruit Magic Tea	5.55
1008	Ginger Longan Tea	6.50
1009	Ginger Milk Tea	5.95
1010	Ginger Tea	5.55
1011	Hawaiian Fruit Tea	5.95
1012	Honey Green Milk Tea	6.65
1013	Honey Lemon Tea	6.65
1014	Honey Milk Tea	5.55
1015	Honeydew Milk Tea	5.50
1016	Green Apple Tea	5.75

Order OLTP table:

orderid	customer_customerid	store_location_storeid	employee_employeeid	order_date	order_status	promotion_promotionid	order_time
1	35	20239	2	2019-09-24	1	3	16:39:00
2	30	20237	58	2019-09-24	1	1	11:38:00
3	69	20238	50	2019-09-24	1	6	13:16:00
4	34	20239	35	2020-05-15	1	5	12:19:00
5	61	20236	27	2019-08-28	0	NULL	15:41:00
6	8	20233	40	2020-03-01	1	1	11:30:00
7	55	20236	10	2020-03-01	1	6	13:15:00
8	60	20235	40	2020-11-09	1	3	11:39:00
9	63	20235	55	2020-10-19	1	1	13:31:00
10	74	20235	29	2020-10-19	1	6	16:00:00
11	1	20232	9	2020-10-19	1	3	19:05:00
12	72	20237	40	2020-06-06	0	NULL	12:02:00
13	48	20237	21	2020-06-06	1	1	14:34:00
14	39	20236	56	2020-10-02	1	5	15:08:00
15	10	20235	41	2020-10-02	1	4	16:51:00
16	31	20237	34	2019-06-22	1	4	13:34:00

Order_details OLTP table:

order_orderid	menu_productid	quantity
1	1006	1
2	1016	3
3	1022	4
4	1009	3
5	1019	4
6	1013	2
7	1021	4
7	1022	2
8	1015	3
9	1017	4
10	1006	1
11	1002	4
12	1015	4
13	1020	2
13	1026	3
14	1026	3

Payment OLTP table:

transactionid	order_orderid	payment_type
142776	68	credit card
147384	10	cash
160997	57	credit card
164299	55	credit card
207330	11	credit card
208146	14	debit card
214197	29	cash
217077	4	credit card
217601	36	credit card
222271	69	credit card
226482	95	credit card
270645	78	credit card
286573	13	cash
310217	81	debit card
315260	49	credit card
326797	5	credit card

Promotion OLTP table:

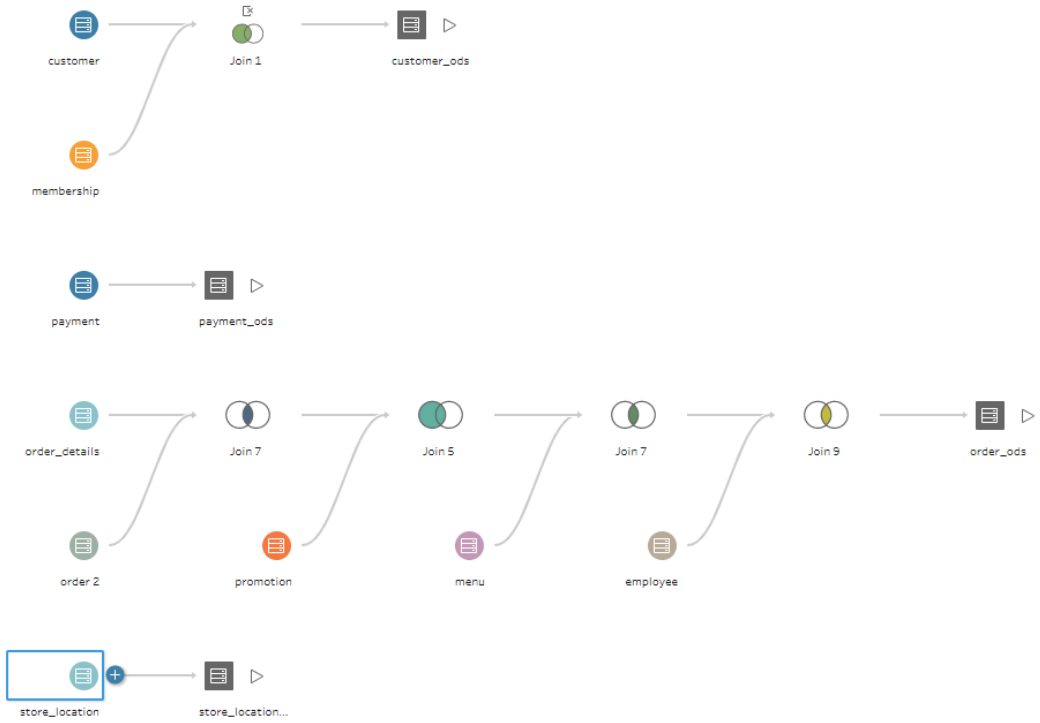
promotionid	promotion_name	start_date	end_date	discount_amount
1	Chinese New Year	2022-12-30	2023-01-03	0.12
2	Valentine's Day	2023-02-11	2023-02-14	0.50
3	Thanksgiving Day	2023-11-15	2023-11-30	0.10
4	Christmas	2023-12-20	2023-12-27	0.15
5	Boba Tea Day	2023-04-29	2023-04-30	0.12
6	Halloween	2023-10-15	2023-10-31	0.10

Store_Location OLTP table:

storeid	city	state	zipcode	address
20232	San Jose	CA	95054	2712 Augustine Dr #110
20233	Santa Clara	CA	95050	738 Franklin St
20234	Los Angeles	CA	90006	1001 Vermont Ave
20235	Santa Barbara	CA	93101	651 Paseo Nuevo #213
20236	San Diego	CA	92109	4343 Mission Blvd #104
20237	Monterey	CA	93940	398 Fremont St
20238	San Francisco	CA	94109	2125 Polk St
20239	San Mateo	CA	94401	128 E 3rd Ave

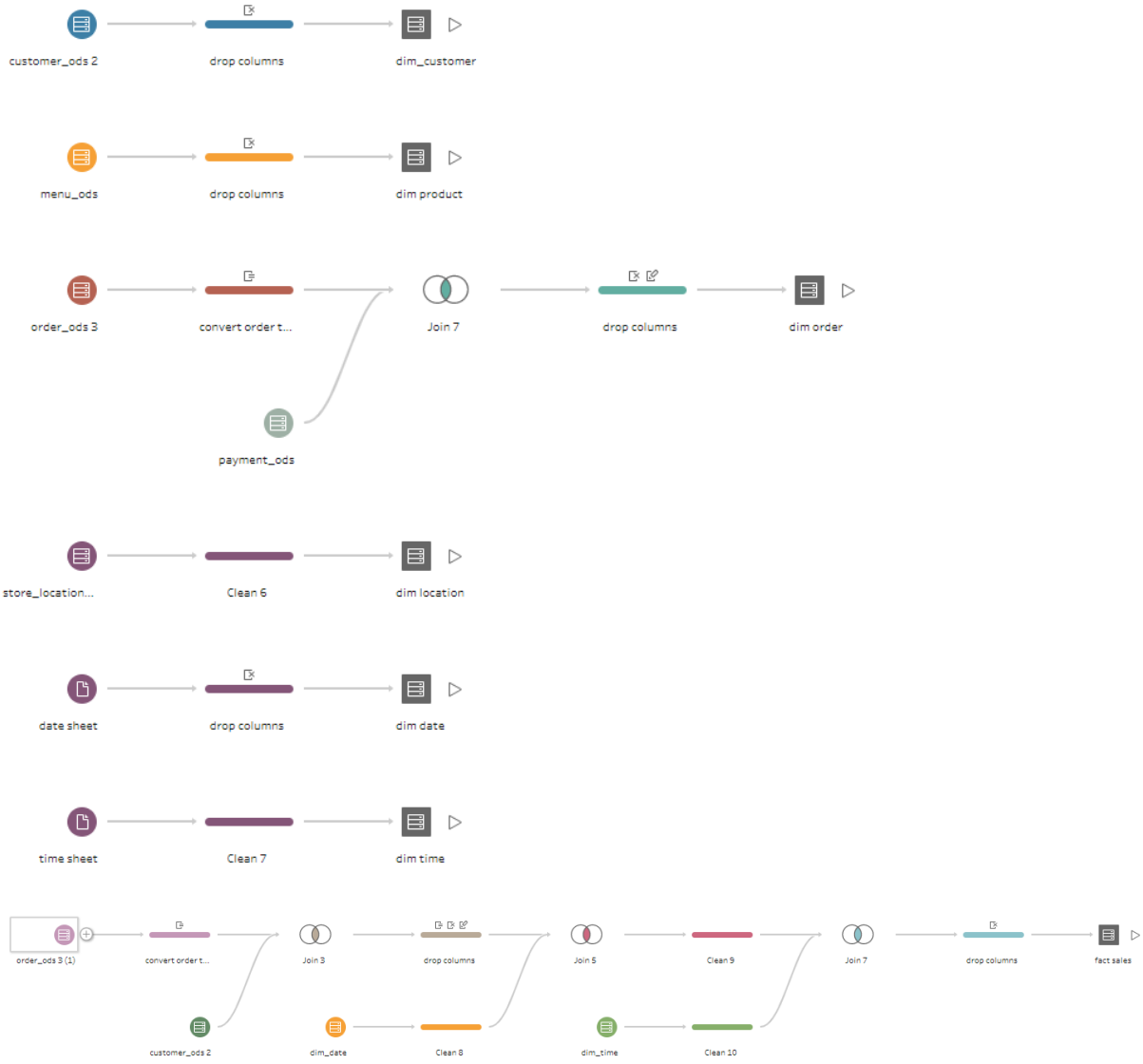
Part 2: Load data into ODS tables

We loaded data into ODS tables by joining from customer and member to make customer_ods, and merging order, oder_detail, payment, menu and employee to make order_ods.

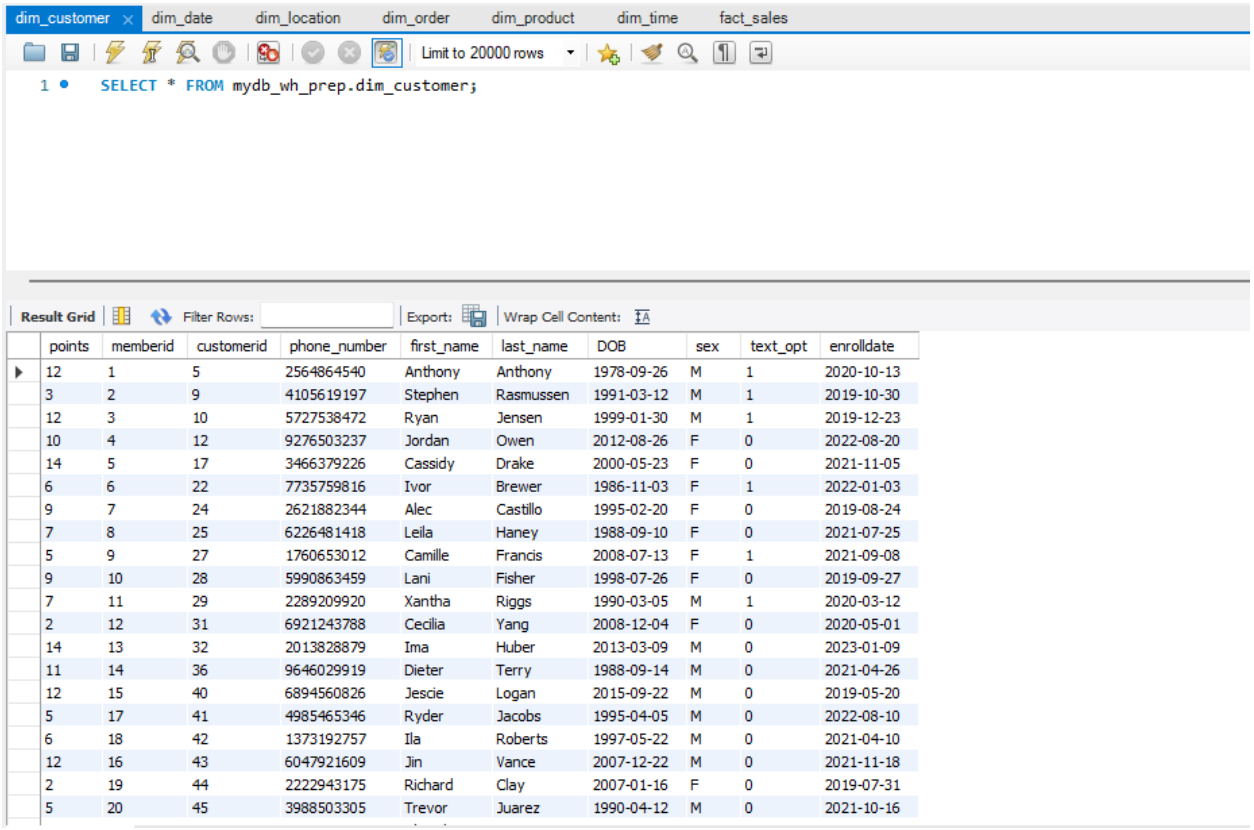


Part 3: Load data from ODS tables to Dimension tables

We joined order_ods and payment_ods to generate dim_order, and merged order_ods, customer_ods, dim_date and dim_time to generate the fact_sales table.



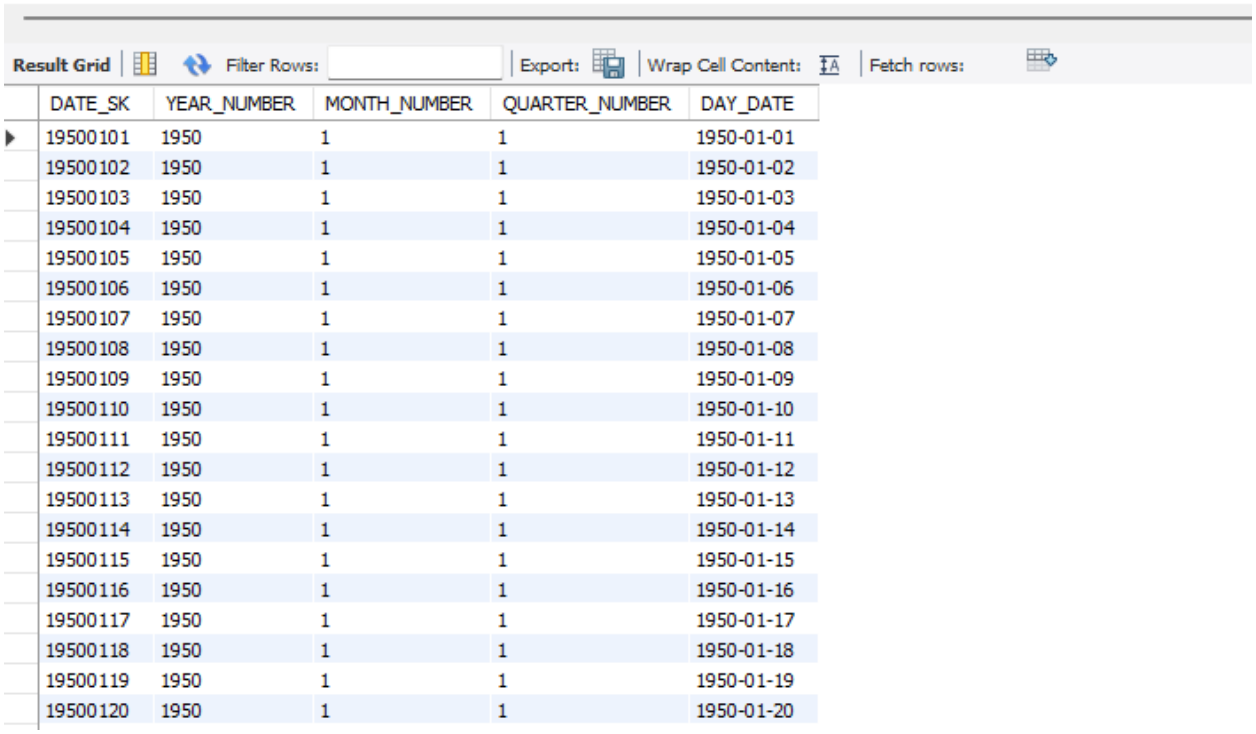
Customer dimension table:



points	memberid	customerid	phone_number	first_name	last_name	DOB	sex	text_opt	enrolldate
12	1	5	2564864540	Anthony	Anthony	1978-09-26	M	1	2020-10-13
3	2	9	4105619197	Stephen	Rasmussen	1991-03-12	M	1	2019-10-30
12	3	10	5727538472	Ryan	Jensen	1999-01-30	M	1	2019-12-23
10	4	12	9276503237	Jordan	Owen	2012-08-26	F	0	2022-08-20
14	5	17	3466379226	Cassidy	Drake	2000-05-23	F	0	2021-11-05
6	6	22	7735759816	Ivor	Brewer	1986-11-03	F	1	2022-01-03
9	7	24	2621882344	Alec	Castillo	1995-02-20	F	0	2019-08-24
7	8	25	6226481418	Leila	Haney	1988-09-10	F	0	2021-07-25
5	9	27	1760653012	Camille	Francis	2008-07-13	F	1	2021-09-08
9	10	28	5990863459	Lani	Fisher	1998-07-26	F	0	2019-09-27
7	11	29	2289209920	Xantha	Riggs	1990-03-05	M	1	2020-03-12
2	12	31	6921243788	Cecilia	Yang	2008-12-04	F	0	2020-05-01
14	13	32	2013828879	Ima	Huber	2013-03-09	M	0	2023-01-09
11	14	36	9646029919	Dieter	Terry	1988-09-14	M	0	2021-04-26
12	15	40	6894560826	Jescie	Logan	2015-09-22	M	0	2019-05-20
5	17	41	4985465346	Ryder	Jacobs	1995-04-05	M	0	2022-08-10
6	18	42	1373192757	Ila	Roberts	1997-05-22	M	0	2021-04-10
12	16	43	6047921609	Jin	Vance	2007-12-22	M	0	2021-11-18
2	19	44	2222943175	Richard	Clay	2007-01-16	F	0	2019-07-31
5	20	45	3988503305	Trevor	Juarez	1990-04-12	M	0	2021-10-16

Date dimension table:

```
1 •   SELECT * FROM mydb_wh_prep.dim_date;
```

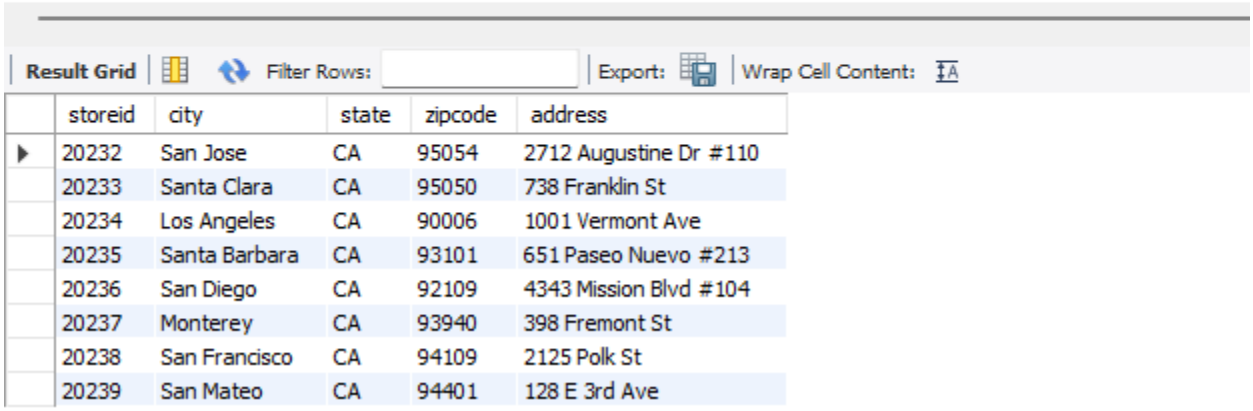


Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

DATE_SK	YEAR_NUMBER	MONTH_NUMBER	QUARTER_NUMBER	DAY_DATE
19500101	1950	1	1	1950-01-01
19500102	1950	1	1	1950-01-02
19500103	1950	1	1	1950-01-03
19500104	1950	1	1	1950-01-04
19500105	1950	1	1	1950-01-05
19500106	1950	1	1	1950-01-06
19500107	1950	1	1	1950-01-07
19500108	1950	1	1	1950-01-08
19500109	1950	1	1	1950-01-09
19500110	1950	1	1	1950-01-10
19500111	1950	1	1	1950-01-11
19500112	1950	1	1	1950-01-12
19500113	1950	1	1	1950-01-13
19500114	1950	1	1	1950-01-14
19500115	1950	1	1	1950-01-15
19500116	1950	1	1	1950-01-16
19500117	1950	1	1	1950-01-17
19500118	1950	1	1	1950-01-18
19500119	1950	1	1	1950-01-19
19500120	1950	1	1	1950-01-20

Location dimension table:

```
1 •   SELECT * FROM mydb_wh_prep.dim_location;
```



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

storeid	city	state	zipcode	address
20232	San Jose	CA	95054	2712 Augustine Dr #110
20233	Santa Clara	CA	95050	738 Franklin St
20234	Los Angeles	CA	90006	1001 Vermont Ave
20235	Santa Barbara	CA	93101	651 Paseo Nuevo #213
20236	San Diego	CA	92109	4343 Mission Blvd #104
20237	Monterey	CA	93940	398 Fremont St
20238	San Francisco	CA	94109	2125 Polk St
20239	San Mateo	CA	94401	128 E 3rd Ave

Order dimension table:

```
1 •  SELECT * FROM mydb_wh_prep.dim_order;
```

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

orderid	promotionid	order_status	payment_type
1	3	1	credit card
2	1	1	cash
3	6	1	debit card
4	5	1	credit card
6	1	1	credit card
7	6	1	credit card
7	6	1	credit card
8	3	1	credit card
9	1	1	credit card
10	6	1	cash
11	3	1	credit card
13	1	1	cash
13	1	1	cash
14	5	1	debit card
15	4	1	credit card
16	4	1	credit card
16	4	1	credit card
16	4	1	credit card
17	6	1	credit card
18	4	1	credit card

Product dimension table:

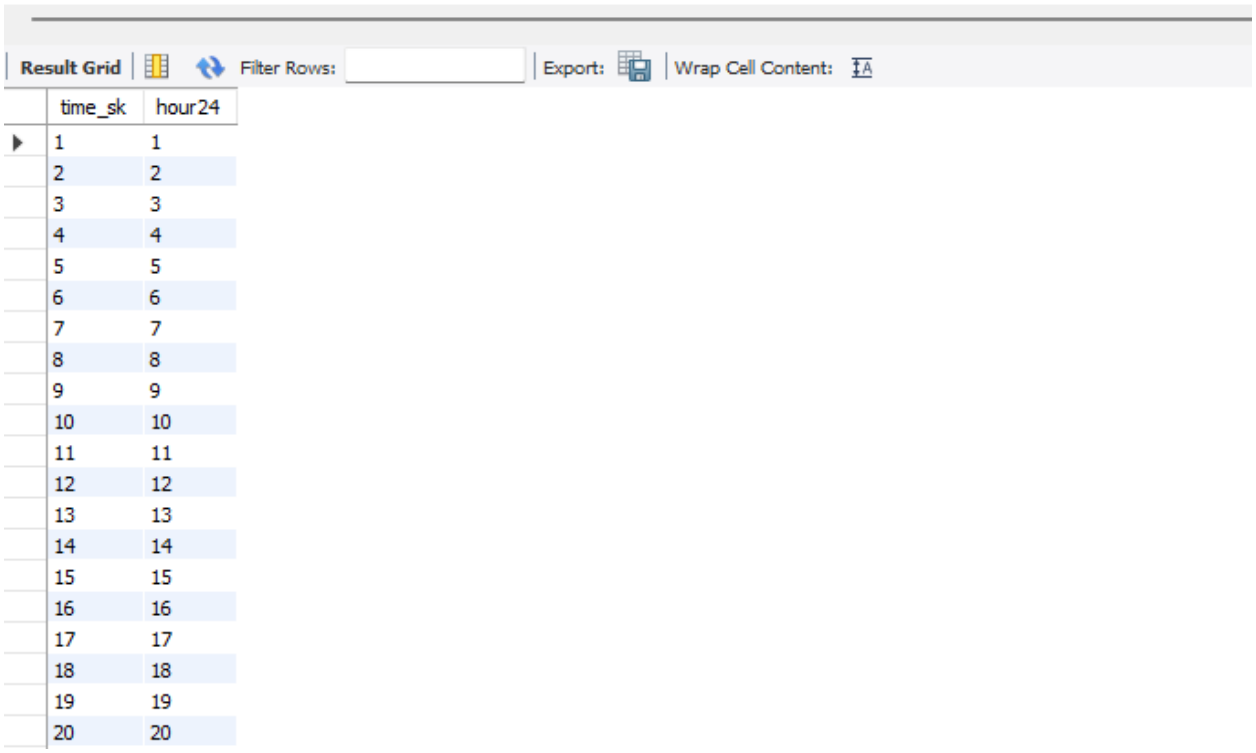
```
1 •  SELECT * FROM mydb_wh_prep.dim_product;
```

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

productid	product_name
1001	Boba Milk Tea
1002	Almond Milk Tea
1003	Black Tea
1004	Chocolate Milk Tea
1005	Coconut Milk Tea
1006	Coconut Pineapple Tea
1007	Fruit Magic Tea
1008	Ginger Longan Tea
1009	Ginger Milk Tea
1010	Ginger Tea
1011	Hawaiian Fruit Tea
1012	Honey Green Milk Tea
1013	Honey Lemon Tea
1014	Honey Milk Tea
1015	Honeydew Milk Tea
1016	Green Apple Tea
1017	Guava Tea
1018	Honey Tea
1019	Iced Tea
1020	Kiwi Green Tea

Time dimension table:

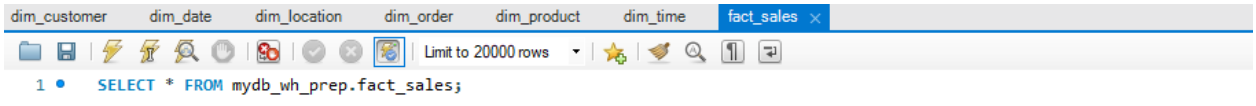
```
1 •  SELECT * FROM mydb_wh_prep.dim_time;
```



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

time_sk	hour24
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20

Fact_sales dimension table:



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

total_sales	customerid	orderid	prioductid	quantity	discount_amount	unit_price	storeid	DATE_SK	time_sk
20.24	5	39	1020	4	0.12	5.75	20233	20200313	11
19.36	5	39	1015	4	0.12	5.5	20233	20200313	11
14.025	9	44	1021	3	0.15	5.5	20234	20190816	18
4.675	10	15	1015	1	0.15	5.5	20237	20201002	16
22.61	17	28	1013	4	0.15	6.65	20237	20190717	19
5.06	17	47	1028	1	0.12	5.75	20238	20190803	12
14.025	24	36	1022	3	0.15	5.5	20234	20200701	19
19.36	25	53	1018	4	0.12	5.5	20237	20190803	13
5	27	97	1019	2	0.5	5	20234	20190606	12
10.472	28	81	1005	2	0.12	5.95	20233	20200114	19
19.98	29	65	1003	4	0.1	5.55	20238	20190923	13
15.18	29	43	1028	3	0.12	5.75	20237	20191104	19
13.2	29	43	1027	3	0.12	5	20237	20191104	19
10.12	29	43	1016	2	0.12	5.75	20237	20191104	19
14.025	31	16	1021	3	0.15	5.5	20237	20190622	13
11.305	31	16	1012	2	0.15	6.65	20237	20190622	13
11.305	31	16	1006	2	0.15	6.65	20237	20190622	13
11	31	27	1022	4	0.5	5.5	20237	20190717	18
18.7	31	62	1025	4	0.15	5.5	20234	20200822	17
14.6625	31	62	1020	3	0.15	5.75	20234	20200822	17

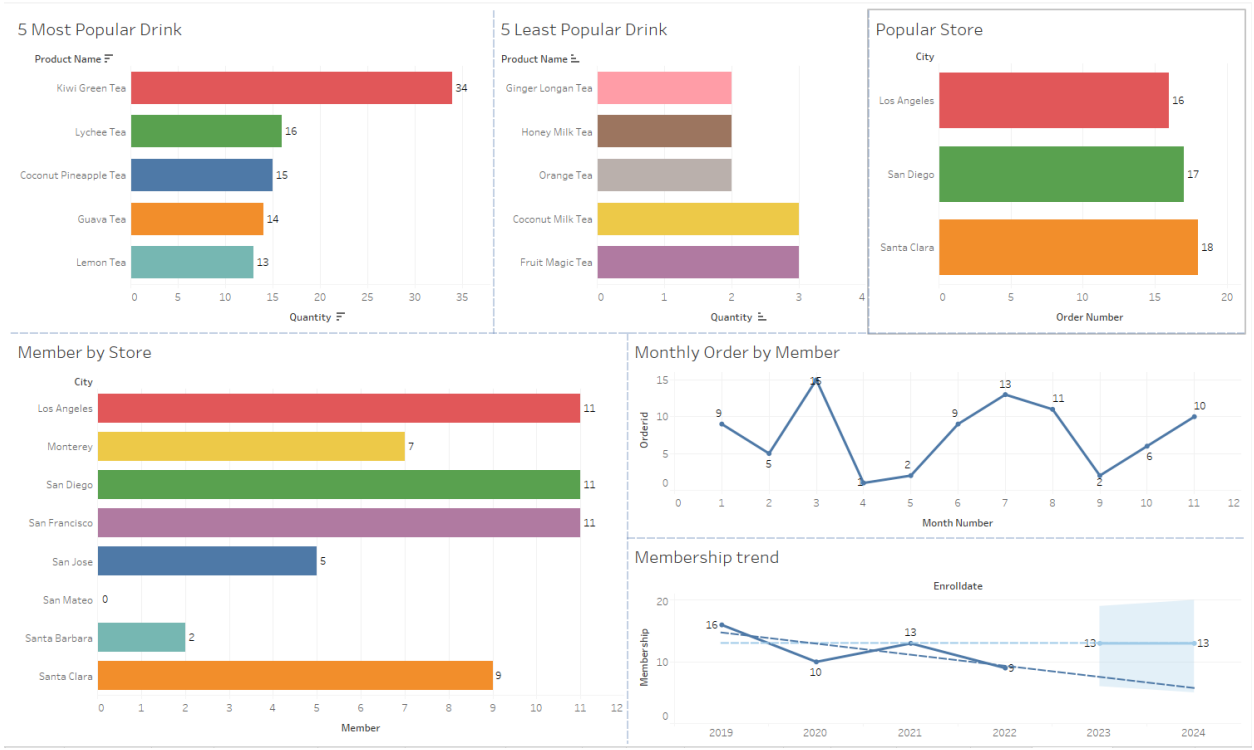
Key Takeaways:

While implementing ETL, the extraction part was good, but we encountered a bit of an issue at the beginning of the transform state. Although we did not have that many tables–while joining them, there were many extra columns that were not needed. We cleaned them up using Tableau Prep, but it was a bit of a mess. When loading the data in Workbench, we were able to run them successfully and pull out the data that we needed to answer our business questions and do analysis.

Phase 4:
Tableau Implementation

Descriptive Analysis1:

1. What are the top 5 most popular drinks?
2. What are the least 5 popular drinks?
3. The number of orders by store location
4. The number of members by store location
5. Monthly orders by members
6. Member trend by year



In this dashboard, we implemented descriptive analysis about products and membership. On the top, we analyzed top and bot 5 drinks by calculating the most and least order count among all drinks. We also listed the most popular stores.

And for bottom charts, we analyzed number of membership by stores. Since San Mateo is a newly open location, it hasn't recruit any member yet. Given the line chart of monthly order by member, we found members like order drinks in summer and spring season. Last but not least, membership trend by dot line shows stores are losing new members and need more actions to boost enrollment.

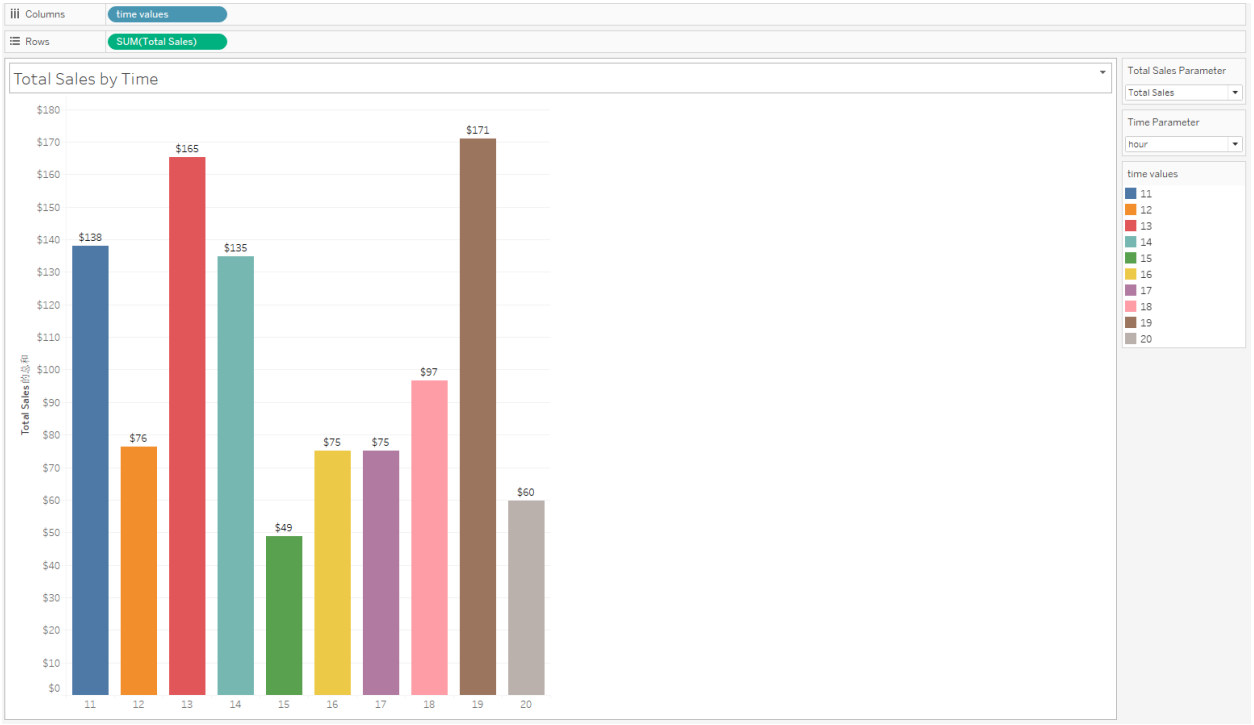
Descriptive Analysis2:

1.What are the different payment types? And the total sales amount associated with it
2.How many members accept the text option? And the total sales associated with it.
3.What are the total sales contributed by customers of two genders?
4.What are the total sales contributed by customers of different age groups?

This dashboard focuses on customer analysis. We use the data of our member to analyze their behavior based on what payment method they prefer, whether they accept text message, gender and age. By showing the total sales contributed by different group, we can see that most members choose to pay with credit card and our main customers are ranged from 20 to 40

Descriptive Analysis3:

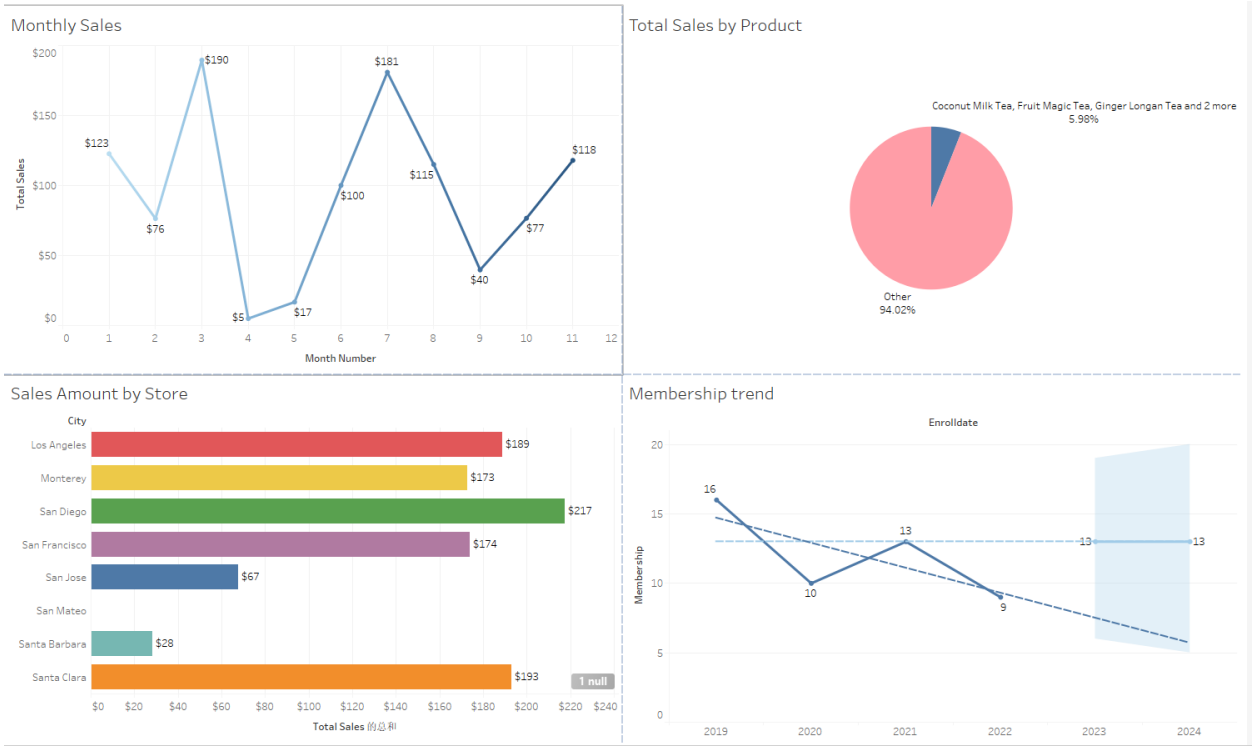
1. Which hour was the busiest in terms of total sales?
2. What are the total sales by hour, day, month, and year?



This dashboard uses the concept of parameters, we compare the total sales, average sales and quantity with different time periods: hour, day, month and year. Based on this, we can see what time the boba shop has the highest sales value.

Predictive Analysis1:

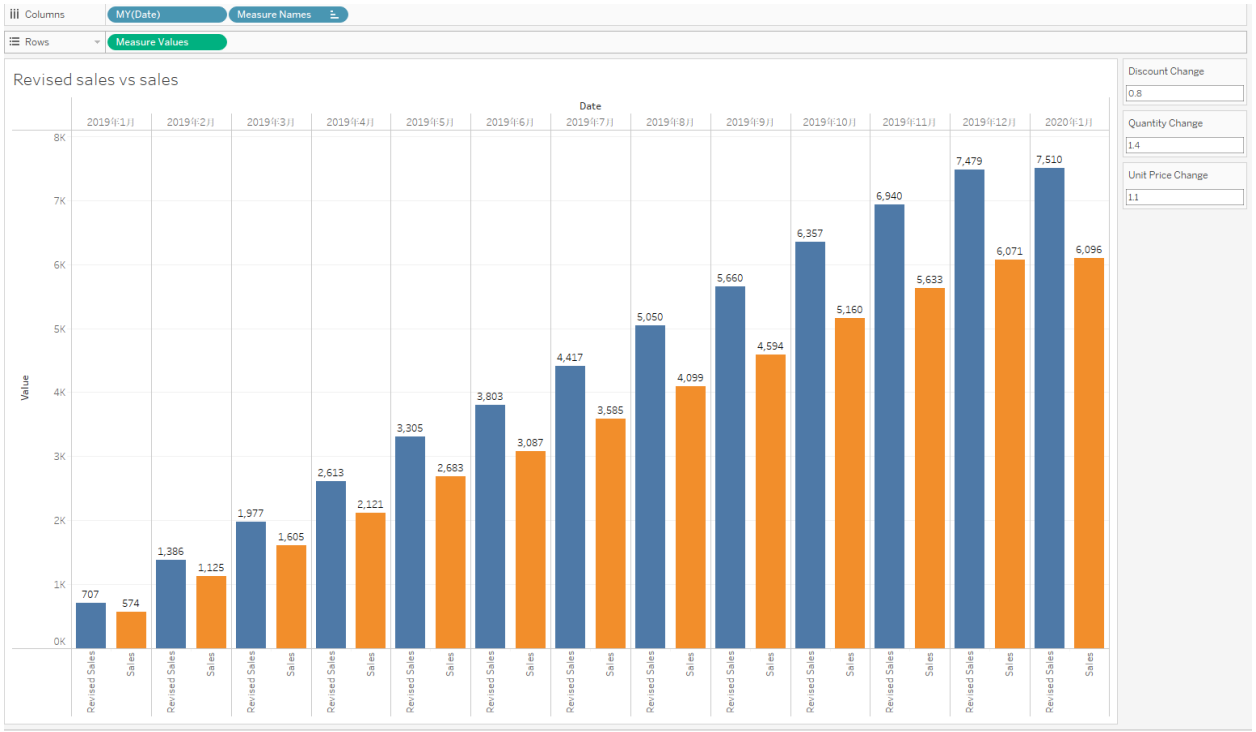
1. Which month of the year had the most sales? How should we allocate the resources?
2. Which menu items are the most popular?
3. Which 5 menu items are the least popular and what can we do to improve sales?
4. Which stores had the best total sales? What can we do to boost sales in that particular store?
5. Is our membership programming working? Should we continue to have promotions?



From this dashboard, we can conclude that our store will have the highest value in March. And since the sales amount of the least 5 popular products is less than 6% of the total sales, we can probably remove those products and try to introduce more new drinks. What's more, San Diego has the highest sales while San Mateo has the lowest . We think that is because the store in San Mateo is just open. Last, We can see from the trend line that the trend of Membership number is decreasing, maybe we need to offer more promotions to change that situation.

Predictive Analysis2:

We can see here by changing the parameters on the right, you can see what will the sales going to look like when we assume there is a certain percentage of growth or decline each month.



This uses the concept of using parameters as well as putting bar charts that represents the same period together. We did some calculations to represent growth or decline by changing the change parameters. In this graph, we forecasted that our quantity will grow 40%, promotion will discount the whole revenue by 20%, but we will increase the price by 10%. Combining everything, we can see what will be the predicted sales for next year.

Key Takeaways:

While implementing Tableau, the first thing is how to visualize the data according to our business question. After completing each sheet, we tried to find a good way to combine multiple sheets into one dashboard to make it reasonable and beautiful. Besides applying the knowledge about parameter we learned in the class, we also tried to create predictive analysis graph based on our data to help making business decisions.