

Home Assignment 2 - Data Engineering

Group 2

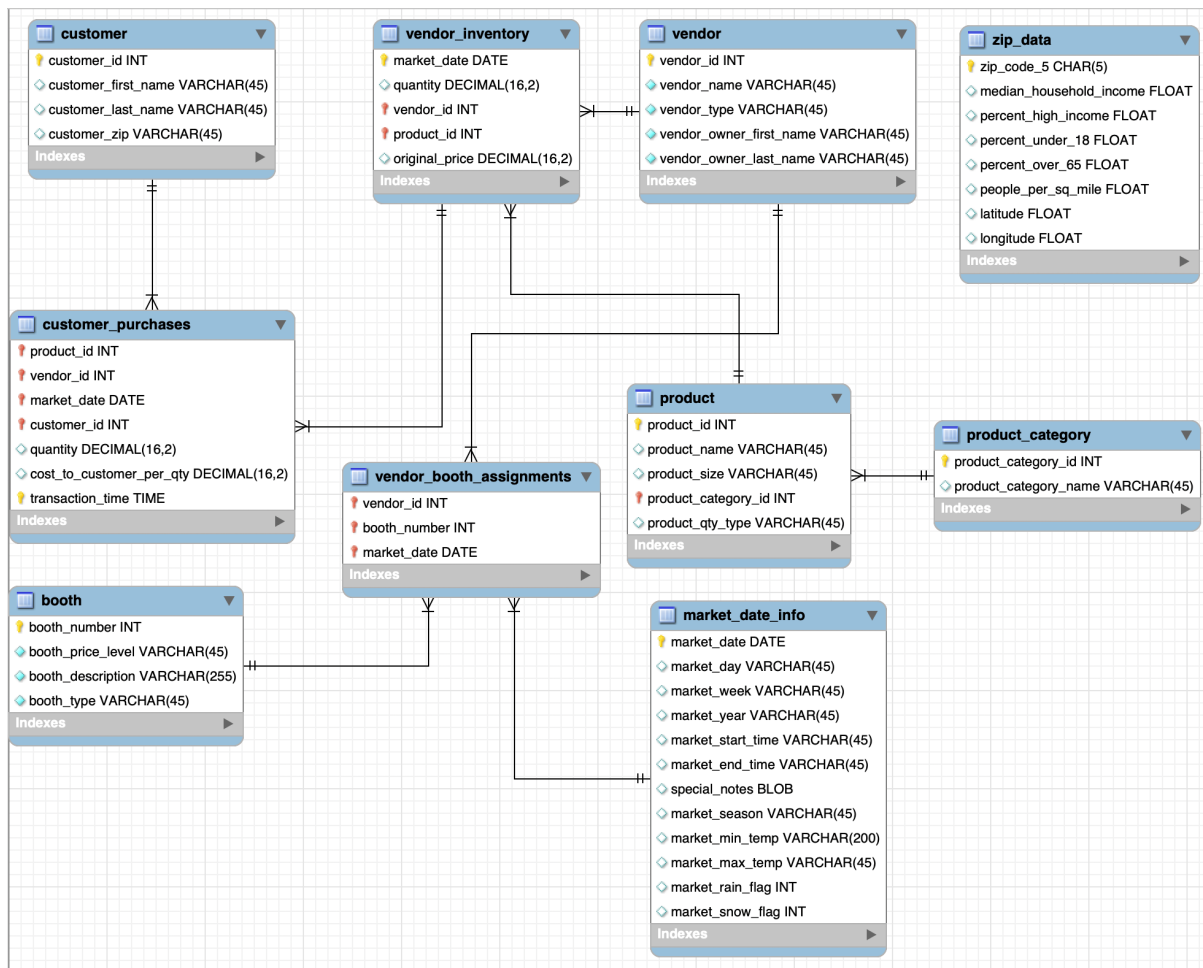
Geraldine Soto: 274032

Nutan Gupta: 274037

1. If you have not installed the Farmers Market database, do so.

Done.

2. Can you make an ERD of the tables in the database so we can get an overview of what we have stored and how it is connected?



3. Why is the “vendor” table not directly linked to the “booth” table?

- The reason why the vendor_table is not directly linked to the booth_table is because there is an indirect link between them, through “vendor_booth_assignments”. Using this link makes it easier to handle the complex many-to-many relationship between these two entities. The intermediate table enables the system to efficiently track which vendors are assigned to which booth and on which dates these assignments take place.

With a direct link between vendor_table and booth_table, it would be difficult to handle this complexity in an appropriate way and could lead to data duplication and possibly inconsistent data. By rather structuring it with the additional link, it provides a more flexible and scalable way of handling the relationship between vendors and booth.

4. Why doesn't the "vendor_booth_assignments" table have a primary key?

- The reason the «vendor_booth_asignments» table does not have a primary key is because it acts as a join table to handle the many-to-many relationship between vendor_table and booth_table with additional market date information. This means that each combination of a vendor, a booth and a date is unique, and together these fields act as a composite key and there is less need for a primary key.

5. What is the earliest and latest date from which we have market data?

```
3 • SELECT
4     MIN(market_date) AS earliest_date,
5     MAX(market_date) AS latest_date
6 FROM market_date_info;
```

The result shows that the earliest date is 02 March, 2019 and the latest date is 10 October, 2020.

Result Grid		Filter Rows:
earliest_date	latest_date	
2019-03-02	2020-10-10	

6. Which packaged product categories do we sell on the farmers market? Make sure the code only returns product categories which contain "packaged" in the name.

```
8 • SELECT product_category_name
9 FROM product_category
10 WHERE product_category_name LIKE '%packaged%';
11
```

The result to find product categories containing "packaged" in their name shows that the Farmers Market sells two packaged product categories: Packaged Pantry Goods and Packaged Prepared Food.

	product_category...	product_category_name
	2	Packaged Pantry Goods
	3	Packaged Prepared Food

7. The farmers market organizers must be notified if any customers are entered with missing data. Make code to return rows with missing data in any column from the "customer" table

```
SELECT*
FROM customer
WHERE customer_id IS NULL
or customer_first_name IS NULL
or customer_last_name IS NULL
or customer_zip IS NULL;
```

customer_id	customer_first_na...	customer_last_na...	customer_zip
NULL	NULL	NULL	NULL

8. The market organizers are required to report the maximum and minimum temperature at any one market day. Add a column to the table "market_date_info". This column should be called "missing_data". The value in this column should be 0 if no minimum or maximum temperature were recorded that day.

```

1 • USE farmers_market;
2 • ALTER TABLE market_date_info;
3 • SHOW COLUMNS FROM market_date_info;
4 • ALTER TABLE market_date_info;
5 • UPDATE market_date_info
6
7 SET missing_data = CASE
8     WHEN market_min_temp IS NULL OR market_max_temp IS NULL THEN 1
9     ELSE 0
10 END;
11
12 • SELECT * FROM market_date_info;

```

✓	4	14:21:19	USE farmers_market	0 row(s) affected	0.00
✓	5	14:21:21	UPDATE market_date_info SET missing_data = CASE WHEN market_min_temp I...	150 row(s) affected Rows matched: 150 Changed: 150 Warnings: 0	0.01
✓	6	14:21:39	UPDATE market_date_info SET missing_data = CASE WHEN market_min_temp I...	0 row(s) affected Rows matched: 150 Changed: 0 Warnings: 0	0.00
✓	7	14:22:44	SHOW COLUMNS FROM market_date_info	13 row(s) returned	0.00
✓	8	14:25:38	SELECT * FROM market_date_info LIMIT 0, 1000	150 row(s) returned	0.00
✓	9	14:26:53	USE farmers_market	0 row(s) affected	0.00
✓	10	14:27:37	SELECT * FROM market_date_info LIMIT 0, 1000	150 row(s) returned	0.00

9. Make a query which outputs the product name and the product category name. Sort on product category name and then on product name.

```

SELECT product_name, product_category_id
FROM product
ORDER BY product_category_id, product_name;

```

product_name	product_category_id
Baby Salad Lettuce Mix	1
Baby Salad Lettuce Mix - Bag	1
Carrots	1
Carrots - Organic	1
Habanero Peppers - Organic	1
Jalapeno Peppers - Organic	1
Organic Cherry Tomatoes	1
Poblano Peppers - Organic	1
Red Potatoes	1
Red Potatoes - Small	1
Roma Tomatoes	1
Sweet Corn	1
Sweet Potatoes	1
Maple Syrup - Jar	2
Apple Pie	3
Banana Peppers - Jar	3
Cherry Pie	3
Whole Wheat Bread	3
Cut Zinnias Bouquet	5
Eggs	6
Pork Chops	6
Farmer's Market Resuable Shopping Bag	7
Homemade Beeswax Candles	7

10. The market organizers would like to see how many transactions there have been per trading day. Information about the trades can be found in the “customer_purchases” table. Present the market_date and a column called “trades_per_day,” which contains the count of trades for each market_date.

```

SELECT
    market_date,
    COUNT(*) AS trades_per_day
FROM customer_purchases
GROUP BY market_date;|

```

market_date	trades_per_day
2019-07-20	32
2020-07-11	37
2020-07-22	33
2020-08-26	39
2020-09-05	35
2020-09-09	45
2019-07-13	34
2019-07-17	46
2019-08-07	34
2019-08-14	38
2019-09-11	41
2019-09-18	31
2020-07-01	35
2020-07-08	40
2020-07-15	38
2020-07-29	35
2020-08-01	30
2020-09-30	41
2019-07-03	40
2019-07-06	42
2019-07-27	43
2019-07-31	28
2019-08-17	31
2019-08-31	37
2019-09-07	36
2019-09-25	32
2020-07-18	35
2020-08-05	42

2020-08-12	42
2020-08-19	39
2020-09-02	37
2020-09-16	35
2020-09-19	40
2020-09-23	32
2020-09-26	39
2019-04-17	26
2019-04-24	26
2019-04-27	20
2019-05-01	26
2019-05-15	23
2019-05-18	21
2019-05-29	28
2019-06-12	34
2019-06-15	33
2019-08-10	40
2019-09-04	35
2019-10-02	23
2019-10-05	23
2019-10-19	23
2019-10-30	27
2019-11-06	25
2019-11-09	26
2019-11-30	31
2019-12-11	24
2019-12-18	33
2020-03-18	28
2020-03-21	21

2020-03-25	25
2020-04-08	21
2020-04-15	23
2020-04-25	18
2020-05-02	23
2020-05-06	27
2020-05-09	25
2020-05-20	26
2020-05-23	24
2020-05-30	26
2020-06-17	36
2020-06-27	31
2020-07-25	40
2020-08-08	48
2020-10-07	23
2020-10-10	25
2019-04-06	25
2019-04-13	21
2019-04-20	26
2019-05-04	29
2019-06-01	24
2019-06-22	33
2019-06-29	30
2019-07-10	32
2019-09-28	37
2019-10-16	28
2019-11-20	38
2019-12-14	25
2020-03-11	26

2020-04-29	20
2020-05-16	23
2020-06-06	30
2020-08-29	40
2019-05-08	26
2019-05-22	26
2019-08-28	33
2019-10-09	20
2019-10-12	22
2019-11-13	21
2019-12-28	33
2020-04-04	15
2020-06-20	30
2019-11-02	27
2019-12-21	30
2020-03-07	22
2019-06-19	29
2019-06-26	36
2019-08-24	38
2019-09-21	36
2020-07-04	44
2019-09-14	34
2020-09-12	26
2019-08-03	39
2019-08-21	32
2020-08-22	39
2019-05-25	19
2019-10-23	30
2019-10-26	26

2019-11-16	28
2019-12-04	23
2019-12-07	27
2019-12-25	27
2020-03-14	21
2020-03-28	24
2020-04-18	17
2020-04-22	24
2020-05-27	20
2020-06-13	28
2020-10-03	26
2019-04-10	23
2019-05-11	23
2019-06-08	36
2019-11-23	27
2020-05-13	18
2020-06-03	26
2020-06-10	31
2020-06-24	31
2020-08-15	38
2019-06-05	26
2019-07-24	30
2019-04-03	23
2020-04-11	17
2020-03-04	19
2020-04-01	16
2019-11-27	21