

DATA ANALYSIS PROJECT

CASE STUDY OF A CAKE DELIVERY SERVICE

Szimonetta Farkas | Data Analyst

INTRODUCTION

Umer, a cafe owner, came up with an idea. He observed that his customers rather stay at home when the weather is cold in winter. He decided to start a cake delivery service for people who would like to enjoy a piece of cake at their warm and cosy home.

The 10 menu items customers can get delivered :

**Butter Cake, Chocolate Cake, Red Velvet Cake, Pineapple Cake, Coffee Cake,
Apple Cake, Cheesecake, Strawberry Cake, Lemon Cake, Honey Cake.**

Umer wants to gain insights about his service idea, so he requested me to help analyzing a collection of data about his customers.

PROBLEM STATEMENT

The cake delivery service is active from 12:00 pm until 4:00 pm, 3 days (Monday, Tuesday, Wednesday) a week.

Umer needs information about the favorite choices of his customers, ordering patterns, and about how much money the customers spent.

He requested me to take a look at the data about the delivery system and analyze it as well.



DETAILS ABOUT THE DATASET

I have 3 tables. The first one includes the type of cakes, cake IDs and the prices(Euro)

CAKES.ID = ORDERS.CAKE_ID

CAKES

ID	CAKE_NAME	PRICE
1	Butter Cake	4.5
2	Chocolate Cake	5.5
3	Red Velvet Cake	7.5
4	Pineapple Cake	6.5
5	Coffee Cake	5.5
6	Apple Cake	5
7	Cheesecake	5.5
8	Strawberry Cake	6.5
9	Lemon Cake	5.5
10	Honey Cake	6.5

The second table includes the order IDs, customer IDs, cake IDs and the date of order

ORDERS.CAKE_ID = CAKE.ID

ORDERS.ORDER_ID =
DELIVERY.ORDER_ID

ORDERS

ORDER_ID	CUST_ID	CAKE_ID	ORDER_DATE
1	101	6	2023-01-02 13:20:00
1	101	5	2023-01-02 13:20:00
1	101	10	2023-01-02 13:20:00
1	101	4	2023-01-02 13:20:00
2	102	6	2023-01-02 14:09:00
3	103	5	2023-01-02 15:35:00
4	104	2	2023-01-03 14:11:00
5	105	1	2023-01-03 14:50:00
5	105	9	2023-01-03 14:50:00
6	101	8	2023-01-03 15:26:00
6	101	10	2023-01-03 15:26:00
7	106	4	2023-01-03 15:17:00
8	103	7	2023-01-04 13:56:00
9	107	6	2023-01-04 15:17:00
10	101	8	2023-01-09 14:53:00
10	101	10	2023-01-09 14:53:00
11	108	4	2023-01-09 15:18:00
12	109	9	2023-01-10 14:30:00
12	109	4	2023-01-10 14:30:00
13	110	2	2023-01-10 14:51:00
14	101	10	2023-01-11 13:54:00
14	101	6	2023-01-11 13:54:00
15	107	8	2023-01-11 14:39:00
16	104	5	2023-01-16 14:57:00
17	101	4	2023-01-16 15:35:00
18	102	7	2023-01-17 14:05:00
19	110	10	2023-01-17 14:31:00
20	101	4	2023-01-17 14:59:00
21	109	7	2023-01-18 14:03:00
22	106	6	2023-01-18 14:48:00
23	101	5	2023-01-23 14:47:00
23	101	9	2023-01-23 14:47:00
24	102	1	2023-01-23 15:01:00
25	109	2	2023-01-24 14:57:00
26	101	10	2023-01-25 14:06:00
27	104	3	2023-01-25 14:41:00
28	110	8	2023-01-30 14:18:00
29	107	1	2023-01-30 14:50:00
30	103	6	2023-01-31 14:55:00

The third table includes the information about delivery: order IDs, employer IDs, pick up time, the delivery distance (km) and the duration of the delivery (minutes)

DELIERY.ORDER_ID =
ORDERS.ORDER_ID

DELIVERY

ORDER_ID	EMP_ID	PICK_UP_TIME	DISTANCE	DURATION
1	1	2023-01-02 13:40:00	1	11
2	3	2023-01-02 14:14:00	1	15
3	1	2023-01-02 15:40:00	2	18
4	1	2023-01-03 14:16:00	1	10
5	3	2023-01-03 15:01:00	1	11
6	1	2023-01-03 15:36:00	2	20
7	3	2023-01-03 15:22:00	1	9
8	1	2023-01-04 14:01:00	1	10
9	2	2023-01-04 15:22:00	2	18
10	1	2023-01-09 15:03:00	1	10
11	2	2023-01-09 15:23:00	1	12
12	1	2023-01-10 14:40:00	1	9
13	2	2023-01-10 14:56:00	1	8
14	1	2023-01-11 14:04:00	2	18
15	2	2023-01-11 14:44:00	1	10
16	1	2023-01-16 15:02:00	2	19
17	2	2023-01-16 15:40:00	1	10
18	1	2023-01-17 14:10:00	2	15
19	2	2023-01-17 14:36:00	1	10
20	3	2023-01-17 15:04:00	1	8
21	1	2023-01-18 14:08:00	1	11
22	2	2023-01-18 14:53:00	3	28
23	1	2023-01-23 14:57:00	1	12
24	2	2023-01-23 15:06:00	2	17
25	1	2023-01-24 15:02:00	3	30
26	1	2023-01-25 14:11:00	2	21
27	2	2023-01-25 14:46:00	1	9
28	1	2023-01-30 14:23:00	1	10
29	2	2023-01-30 14:55:00	1	10
30	1	2023-01-31 15:00:00	2	15

1, HOW MANY CAKES WERE ORDERED?

```
SELECT COUNT(CAKE_ID) AS NUMBER_OF_ORDERED_CAKES FROM ORDERS;
```

	NUMBER_OF_ORDERED_CAKES
▶	39

2, HOW MANY ITEMS WERE ORDERED BY EACH CUSTOMER?

```
SELECT CUST_ID, COUNT(CAKE_ID) AS NUMBER_OF_PURCHASED_ITEMS
FROM ORDERS
GROUP BY CUST_ID
ORDER BY COUNT(CAKE_ID) DESC;
```

CUST_ID	NUMBER_OF_PURCHASED_ITEMS
101	15
109	4
102	3
103	3
104	3
107	3
110	3
105	2
106	2
108	1

3, HOW MUCH MONEY WAS MADE BY CAKE DELIVERY?

```
SELECT SUM(PRICE) AS INCOME  
FROM ORDERS  
INNER JOIN CAKES  
ON CAKES.ID = ORDERS.CAKE_ID;
```

INCOME

226.5

4, HOW MUCH MONEY WAS MADE BY EACH TYPE OF CAKE?

```
SELECT CAKE_NAME, SUM(PRICE) AS MONEY_MADE_BY_EACH_TYPE
FROM ORDERS
INNER JOIN CAKES ON CAKES.ID = ORDERS.CAKE_ID
GROUP BY CAKE_NAME
ORDER BY SUM(PRICE) DESC;
```

CAKE_NAME	MONEY_MADE_BY_EACH_TYPE
Honey Cake	39
Pineapple Cake	39
Apple Cake	30
Strawberry Cake	26
Coffee Cake	22
Chocolate Cake	16.5
Lemon Cake	16.5
Cheesecake	16.5
Butter Cake	13.5
Red Velvet Cake	7.5

5, HOW MUCH MONEY WAS SPENT BY EACH CUSTOMER?

```
SELECT CUST_ID, SUM(PRICE) AS MONEY_SPENT_BY_CUSTOMER
FROM ORDERS
INNER JOIN CAKES
ON CAKES.ID = ORDERS.CAKE_ID
GROUP BY CUST_ID
ORDER BY SUM(PRICE) DESC;
```

CUST_ID	MONEY_SPENT_BY_CUSTOMER
101	91.5
109	23
104	18.5
110	18.5
103	16
107	16
102	15
106	11.5
105	10
108	6.5

6, WHAT WAS THE MOST PURCHASED ITEM BY EACH CUSTOMER?

```
WITH FAVORITE AS (  
  SELECT CUST_ID, CAKE_NAME, COUNT(CAKE_ID) AS PURCHASES,  
         DENSE_RANK() OVER(PARTITION BY CUST_ID ORDER BY COUNT(CAKE_ID) DESC) AS NUM  
  FROM ORDERS  
  INNER JOIN CAKES ON CAKES.ID = ORDERS.CAKE_ID  
  GROUP BY CUST_ID, CAKE_NAME)  
  
SELECT CUST_ID, CAKE_NAME, PURCHASES  
FROM FAVORITE WHERE NUM = 1;
```

CUST_ID	CAKE_NAME	PURCHASES
101	Honey Cake	5
102	Apple Cake	1
102	Cheesecake	1
102	Butter Cake	1
103	Coffee Cake	1
103	Cheesecake	1
103	Apple Cake	1
104	Chocolate Cake	1
104	Coffee Cake	1
104	Red Velvet Cake	1
105	Butter Cake	1
105	Lemon Cake	1
106	Pineapple Cake	1
106	Apple Cake	1
107	Apple Cake	1
107	Strawberry Cake	1
107	Butter Cake	1
108	Pineapple Cake	1
109	Lemon Cake	1
109	Pineapple Cake	1
109	Cheesecake	1
109	Chocolate Cake	1
110	Chocolate Cake	1
110	Honey Cake	1
110	Strawberry Cake	1

7, HOW MANY UNIQUE CUSTOMER ORDER WERE MADE?

```
SELECT COUNT(DISTINCT(CUST_ID)) FROM ORDERS;
```

COUNT(DISTINCT(CUST_ID))
10

8, HOW MANY OF EACH TYPE OF CAKE WERE DELIVERED?

```
SELECT CAKE_NAME , COUNT(CAKE_ID) AS NUMBER_OF_SOLD_CAKES
FROM ORDERS
INNER JOIN CAKES ON CAKES.ID = ORDERS.CAKE_ID
GROUP BY CAKE_NAME;
```

CAKE_NAME	NUMBER_OF_SOLD_CAKES
Apple Cake	6
Coffee Cake	4
Honey Cake	6
Pineapple Cake	6
Chocolate Cake	3
Butter Cake	3
Lemon Cake	3
Strawberry Cake	4
Cheesecake	3
Red Velvet Cake	1

9, WHAT WAS THE MAXIMUM NUMBER OF ITEMS DELIVERED IN A SINGLE ORDER?

```
SELECT ORDER_ID, COUNT(CAKE_ID) NUMBER_OF_ITEMS FROM ORDERS  
GROUP BY ORDER_ID  
ORDER BY COUNT(CAKE_ID) DESC  
LIMIT 1;
```

ORDER_ID	NUMBER_OF_ITEMS
1	4

10, WHAT WAS THE TOTAL VOLUME OF CAKES ORDERED FOR EACH HOUR OF THE DAY?

```
SELECT EXTRACT(HOUR FROM ORDER_DATE) AS HOUR_DATA , COUNT(DISTINCT(ORDER_ID)) AS NUMBER_OF_ORDERS  
FROM ORDERS  
GROUP BY HOUR_DATA;
```

HOUR_DATA	NUMBER_OF_ORDERS
13	3
14	20
15	7

11, WHAT WAS THE VOLUME OF ORDERS FOR EACH DAY OF THE WEEK?

```
SELECT DAYNAME(ORDER_DATE) AS NAME_OF_THE_DAY, COUNT(DISTINCT(ORDER_ID)) AS NUMBER_OF_ORDERS
FROM ORDERS
GROUP BY NAME_OF_THE_DAY;
```

NAME_OF_THE_DAY	NUMBER_OF_ORDERS
Monday	11
Tuesday	11
Wednesday	8

12, IS THERE ANY RELATIONSHIP BETWEEN THE NUMBER OF CAKES AND HOW LONG THE ORDER TAKES TO PREPARE?

```
SELECT O.ORDER_ID, COUNT(O.ORDER_ID) AS NUMBER_OF_CAKES,  
TIMESTAMPDIFF(MINUTE, ORDER_DATE, PICK_UP_TIME) AS PREPARING_TIME  
FROM ORDERS AS O  
INNER JOIN DELIVERY AS D ON O.ORDER_ID = D.ORDER_ID  
GROUP BY O.ORDER_ID, PREPARING_TIME  
ORDER BY PREPARING_TIME DESC;
```

ORDER_ID	NUMBER_OF_CAKES	PREPARING_TIME
1	4	20
5	2	11
6	2	10
10	2	10
12	2	10
14	2	10
23	2	10
2	1	5
3	1	5
4	1	5
7	1	5
8	1	5
9	1	5
11	1	5
13	1	5
15	1	5
16	1	5
17	1	5
18	1	5
19	1	5
20	1	5
21	1	5
22	1	5
24	1	5
25	1	5
26	1	5
27	1	5
28	1	5
29	1	5
30	1	5

13, HOW MANY ORDERS WERE DELIVERED BY EACH RUNNER?

```
SELECT EMP_ID, COUNT(EMP_ID) AS NUMBER_OF_DELIVERIES  
FROM DELIVERY  
GROUP BY EMP_ID;
```

EMP_ID	NUMBER_OF_DELIVERIES
1	16
3	4
2	10

14, WHAT WAS THE AVERAGE DISTANCE TRAVELED FOR EACH CUSTOMER?

```
SELECT O.CUST_ID , AVG(D.DISTANCE)
FROM ORDERS AS O
INNER JOIN DELIVERY AS D ON O.ORDER_ID = D.ORDER_ID
GROUP BY O.CUST_ID;
```

CUST_ID	AVG(D.DISTANCE)
101	1.3333
102	1.6667
103	1.6667
104	1.3333
105	1.0000
106	2.0000
107	1.3333
108	1.0000
109	1.5000
110	1.0000

15, WHAT WAS THE AVERAGE DISTANCE TRAVELED FOR ALL THE CUSTOMERS?

```
SELECT AVG(DISTANCE)  
FROM DELIVERY;
```

AVG(DISTANCE)
1.4333

16, WHAT WAS THE DIFFERENCE BETWEEN THE LONGEST AND SHORTEST DELIVERY TIMES FOR ALL ORDERS?

```
SELECT MAX(DURATION) - MIN(DURATION) AS DIFFERENCE_BETWEEN_LONGEST_AND_SHORTEST_DELIVERY  
FROM DELIVERY;
```

DIFFERENCE_BETWEEN_LONGEST_AND_SHORTE
22

17, WHAT WAS THE AVERAGE SPEED FOR EACH RUNNER FOR EACH DELIVERY AND DO YOU NOTICE ANY TREND FOR THESE VALUES?

```
SELECT DISTINCT(O.ORDER_ID),D.EMP_ID, D.DISTANCE*60/DURATION AS SPEED
FROM DELIVERY AS D
INNER JOIN ORDERS AS O ON O.ORDER_ID = D.ORDER_ID
ORDER BY SPEED DESC;
```

ORDER_ID	EMP_ID	SPEED
18	1	8.0000
30	1	8.0000
13	2	7.5000
20	3	7.5000
24	2	7.0588
3	1	6.6667
7	3	6.6667
9	2	6.6667
12	1	6.6667
14	1	6.6667
27	2	6.6667
22	2	6.4286
16	1	6.3158
4	1	6.0000
6	1	6.0000
8	1	6.0000
10	1	6.0000
15	2	6.0000
17	2	6.0000
19	2	6.0000
25	1	6.0000
28	1	6.0000
29	2	6.0000
26	1	5.7143
1	1	5.4545
5	3	5.4545
21	1	5.4545
11	2	5.0000
23	1	5.0000
2	3	4.0000

RESULTS

39 pieces of cakes were ordered during the given time period. Customer 101 purchased the most number of items (15). The maximum number of items delivered in a single order is 4. The most popular type of cakes: Apple Cake, Honey Cake and Pineapple Cake.

10 unique customer orders were made.

Most of the orders are between 2 pm and 3 pm. On Mondays and Tuesdays the number of orders are more compared to Wednesdays.

It takes approximately 5 minutes to prepare and pack 1 piece of cake.

Runner 1 delivered the most number of orders and he had the highest average speed as well. The average distance of the orders is 1.43 km.

226.5 euros were made by cake delivery. The most amount of money was made by Honey Cake and Pineapple Cake. The biggest amount of money was spent by Customer 101 (91.5 euros).

SUGGESTIONS

Customer 101 purchased the most number of items and he/she spent the most amount of money as well. To deeper the connection with this customer and appreciate his/her loyalty it is recommended to offer some special discount.

Apple Cake, Pineapple Cake and Honey Cake are the most popular items. It is recommended to increase their price slightly for a better income.

I suggest that Umer should make sure that between 2 pm and 3 pm there are always enough workers at the Cafe, especially on Mondays and Tuesdays.

Runner 1 delivered most of the orders and he was the fastest as well. He should be the worker of the month.

THANK YOU!

