Data management and visualization

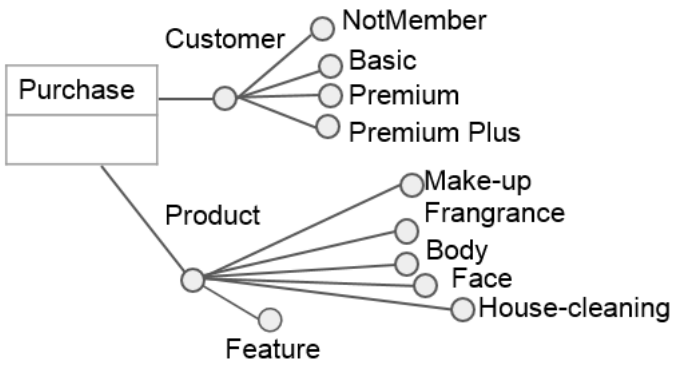
Question 1

Text of the question

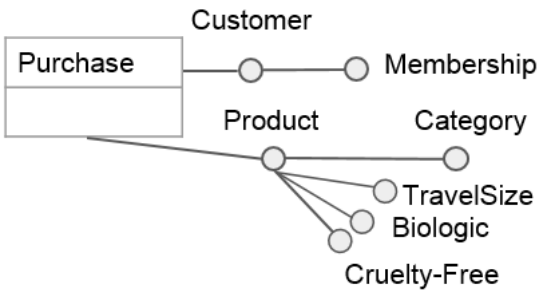
Data analysts of an italian company selling products for the personal care and house cleaning are interested in analyzing statistics about the purchases of their products. In particular, they would like the analyses to address the following features.

- **Customer** and the information of its **membership** to the company fidelity program; the membership can have one of the following values:
  - NotMember
  - Basic
  - Premium
  - Premium plus
- Each **product** belongs to one of the following **categories**:
  - Make-up
  - Fragrance
  - Body
  - Face
  - House Cleaning
- Each **product** may have one or more of the following **features**:
  - Travel size
  - Biologic
  - Cruelty-free
  - For instance, the analysts would like to compute KPIs on all biologic and travel-size products, or on all cruelty-free and biologic products, or on all travel-size non-biologic products.

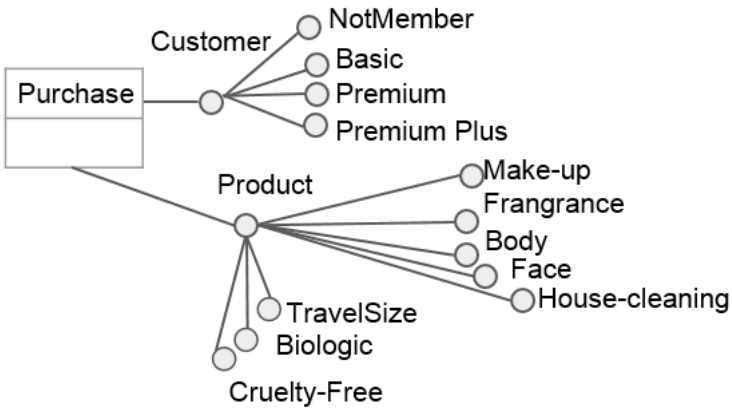
Choose the correct conceptual schema from the proposed ones to properly define the characteristics of customers and products according to the given specifications (at most one answer is allowed).



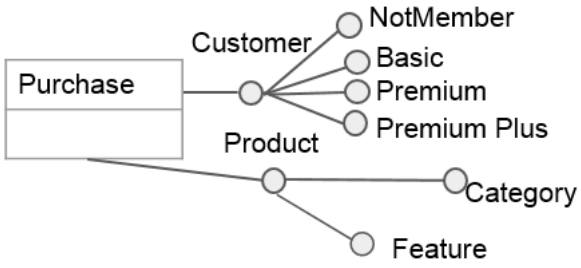
(a)



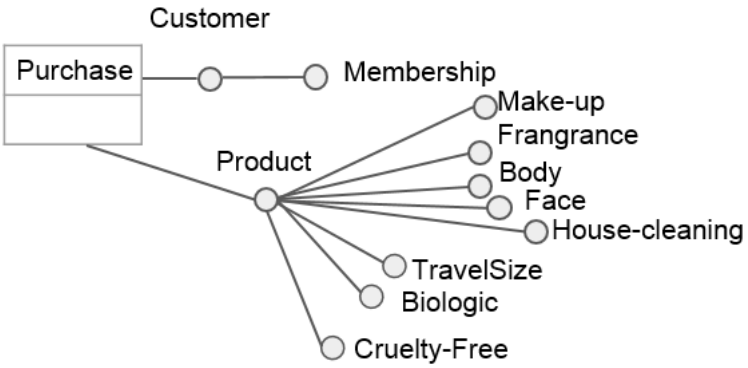
(b) ✓



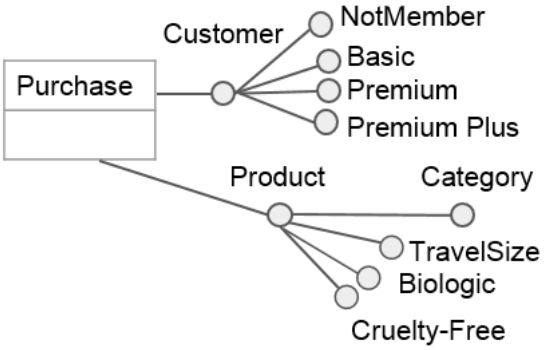
(c)


Purchase
Customer
NotMember
Basic
Premium
Premium Plus
Product
Category
Feature

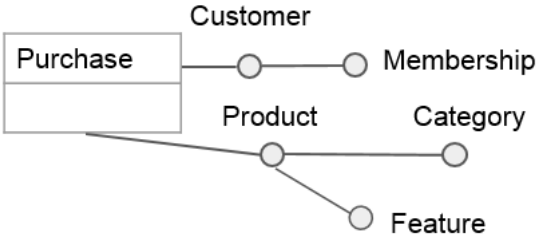
(d)


Purchase
Customer
Membership
Product
Make-up
Frangrance
Body
Face
House-cleaning
TravelSize
Biologic
Cruelty-Free

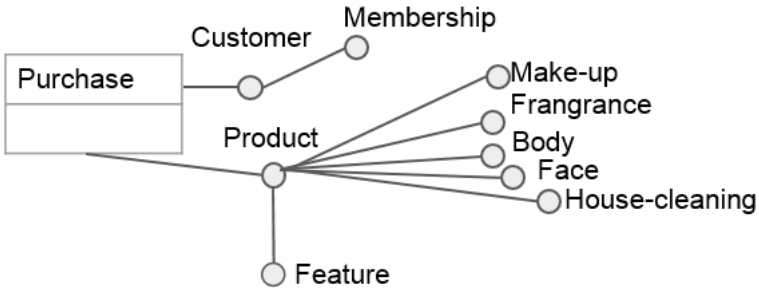
(e)


Purchase
Customer
NotMember
Basic
Premium
Premium Plus
Product
Category
TravelSize
Biologic
Cruelty-Free

(f)

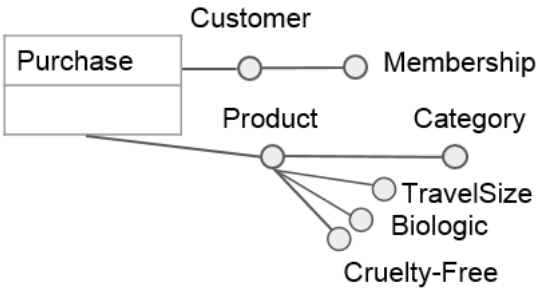

Purchase
Customer
Membership
Product
Category
Feature

(g)


Purchase
Customer
Membership
Product
Make-up
Frangrance
Body
Face
House-cleaning
Feature

(h)

Feedback



The correct answer is:

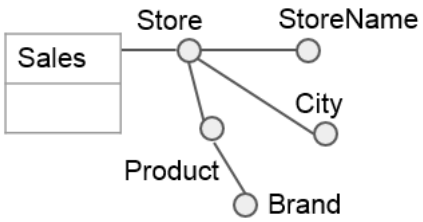
Question 2

Text of the question

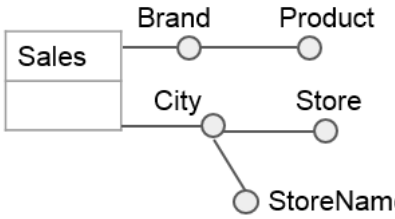
Data analysts of an italian company selling products for the personal care and house cleaning are interested in analyzing statistics about the sales of products in their stores. Consider the following description of the data to design a proper datawarehouse.

- For each **product**, the **brand** of the product is known.
- Each product is sold by multiple stores and each store sells different products.
- A **store** is identified by a specific unique number.
- Stores are characterized by their **name** (not unique) and their full **address**.
- Different stores can have the same same.
- At the same address, only one store is present. Each store has only one address.
- The analyses must be performed for each **store, store name, store city**, **product**, and **brand**.

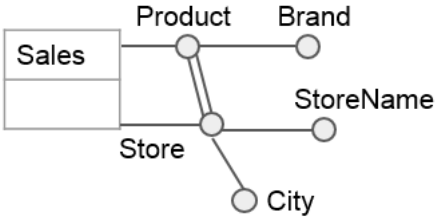
Choose the correct conceptual schema from the proposed ones to properly define the characteristics of products and stores according to the given specifications (at most one answer is allowed).



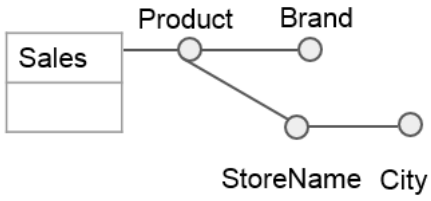
(a)



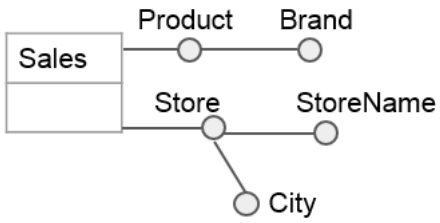
(b)



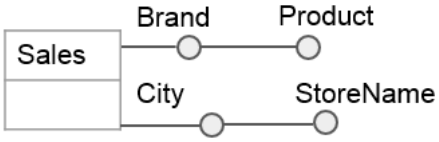
(c)



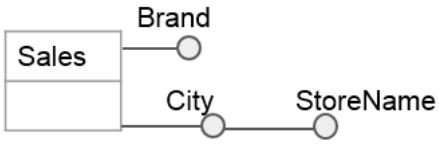
(d)



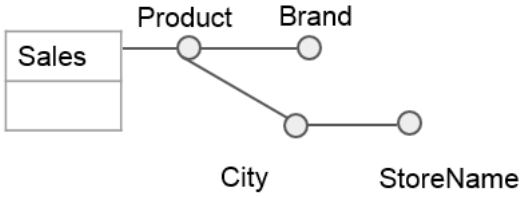
(e) ✓



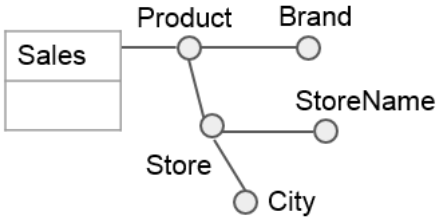
(f)



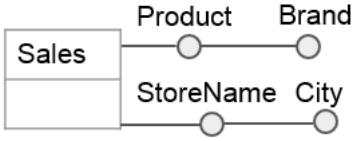
(g)



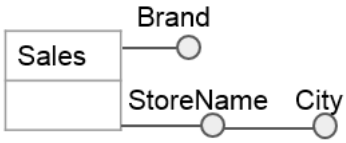
(h)



(i)

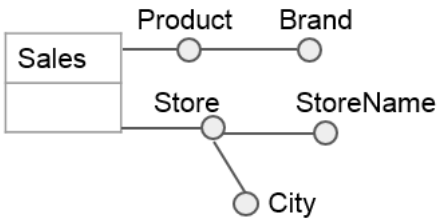


(j)



(k)

Feedback



The correct answer is:

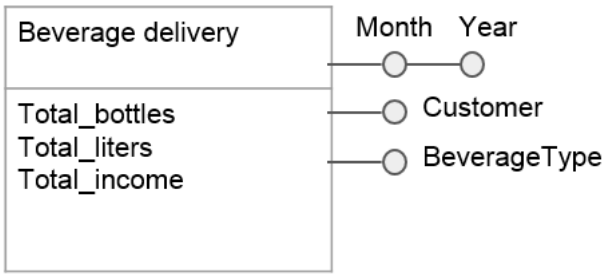
Question 3

Text of the question

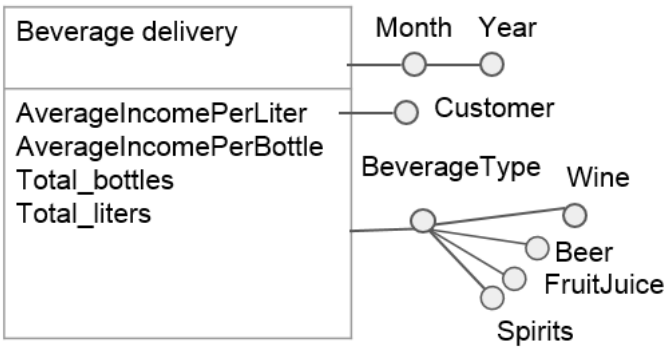
Data analysts are interested in analyzing statistics about beverage deliveries.

- The delivery company is interest in analyzing sales of 5 categories of beverages: wine, beer, fruit juice and spirits. Each beverage belongs to one category.
- Beverages are sold in bottles. Bottles are of different formats, in liters: 0.33, 0.5, 1.0, 1.5.
- Beverages are sold to customers having a personal account on the company platform.
- They want to analyze the statistics based on the customer, the month of the delivery (e.g. may 2020, june 2020, etc.) and year, for each category.
- They want to analyze the average income per bottle, the average income per liter, the total number of liters, and the total number of bottles sold.

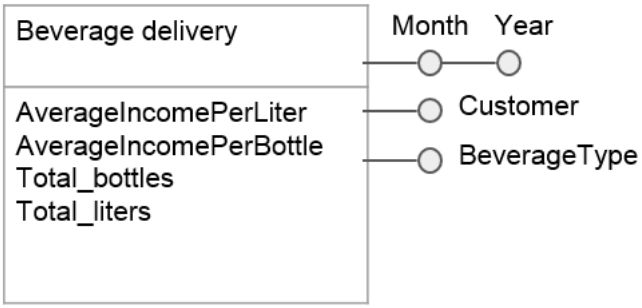
Choose the best solution for the conceptual schema design among those proposed (at most one answer is allowed).



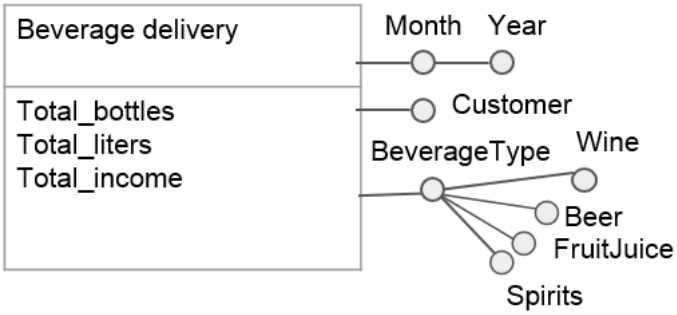
(a) ✓



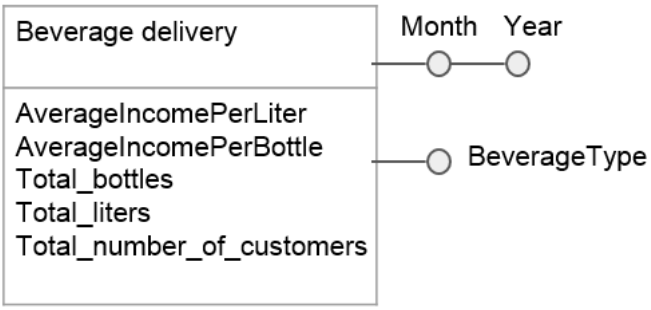
(b)



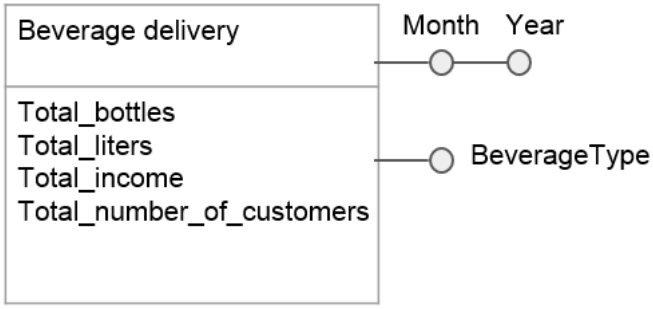
(c)



(d)

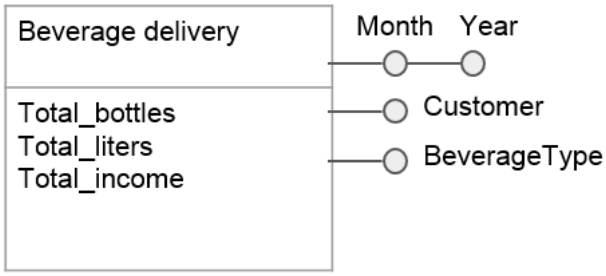


(e)



(f)

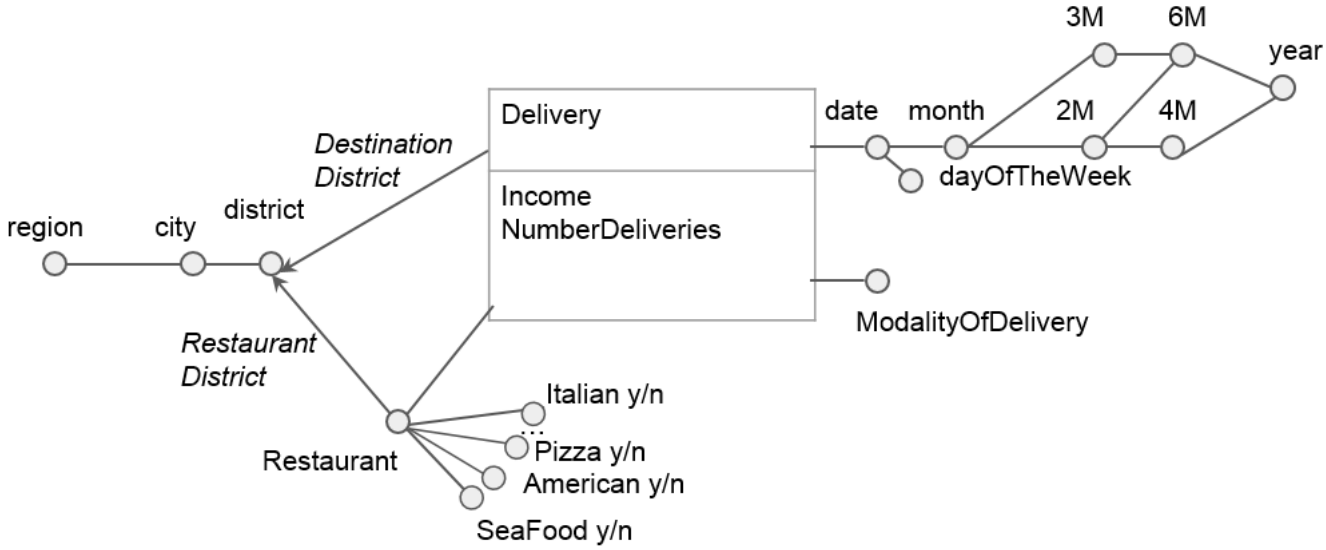
Feedback



The correct answer is:

Question 4

Text of the question



- Each restaurant has a unique name and it can belong to more than one category.
- There are 10 categories of restaurants: Italian, Pizza, .., American, SeaFood.
- The cardinality of “ModalityOfDelivery” is 3, and it can be “B” for bike, “C” for car, or “O” for other means.

Write the logical design of the conceptual DW schema indicated in the picture.
Write each table on a new line.
Use the **bold** or the underline for identifying primary-key attributes.

Feedback

Delivery(TimeId, ModalityOfDelivery, RestaurantId, DestinationDistrict, Income, NumberDeliveries)
Time(TimeId, date, month, 2M, 3M, 4M, 6M, year, dayOfTheWeek)
Restaurant(RestaurantId, Restaurant, Italian, Pizza, America, …, SeaFood, RestaurantDistrictId)
District(DiscritId, District, City, Region)

Question 5

Text of the question

```
Purchase(TimeId, StoreId, income, number_of_items_sold)
Store(StoreId, Name, Brand, City, Region)
Date (TimeId, date, month, year)
```

Separately for each city and month, compute:

- the total income
- the percentage of income in each month, with respect to the total of the year
- inside each region, assign a rank to the cities based on the income (rank 1st the highest income city), separately for each month

Write the requested SQL query.

Feedback

```
SELECT city, month,
     SUM(Income),
     100*SUM(Income)/SUM(SUM(Income)) OVER (PARTITION BY city, year),
      RANK() OVER (PARTITION BY region, month ORDER BY SUM(Income) DESC),

FROM Store S, Date D, Purchase P

WHERE P.StoreId=S.StoreId AND P.Timeid=D.Timeid

GROUP BY city, month, year, region
```

Question 6

Text of the question

```
Purchase(TimeId, StoreId, income, number_of_items_sold)
Store(StoreId, Name, Brand, City, Region)
Date (TimeId, date, month, year)
```

Separately for each month and brand, compute:

- the average income per item
- the daily average number of items
- the cumulative total number of items since the beginning of the year

Write the requested SQL query.

Feedback

```
SELECT Brand, month, year,
     SUM(income)/Sum(number_of_items_sold),
     SUM(number_of_items_sold)/COUNT(DISTINCT Date),
     SUM(SUM(number_of_items_sold)) OVER (PARTITION BY Brand, year
                    ORDER BY month ROWS UNBOUNDED PRECEDING)

FROM Store S, Date D, Purchase P

WHERE P.StoreId=S.StoreId AND P.Timeid=D.Timeid

GROUP BY Brand, month, year
```

Question 6

Text of the question

Design a document-based NoSQL database for storing the following data describing restaurant dishes and deliveries, and specifically optimizing the described access patterns.

A **restaurant** is described by the following data.

- A unique identification string, starting with “R” and followed by a unique number.
- A name, as indicated on its public sign.
- An address, consisting of street name (e.g., “Via Po”), house number (e.g., “1A”), city name, postal code (e.g., 10100), and country name.

A **dish** is described by the following data.

- A unique identification string, starting with “D” and followed by a unique number.
- A name, as indicated on the menu (e.g., “Super-size hot dog”).
- Different prices, such the regular price, the promotional price, the discounted price, etc.
- An approximate weight value in kg.
- A list of the main ingredients, with each ingredient being a text string, such as “bread”, “wurstel”, etc.
- A list of filenames of the pictures of the dish.

Each restaurant belongs to one or more categories, such as pizza, sea food, or street food. The number of categories a restaurant belongs to is low and limited. It is often required to filter restaurants based on their categories.

Each restaurant has a list of dishes in its menu. Dishes are always accessed by their identification number. Each dish might be in the menu of more than one restaurant. The number of dishes for each restaurant can be large, but it is finite.

The access pattern to the above-described data typically consists of (i) a restaurant search, then, after selecting a restaurant, (ii) its dishes are accessed in a second interaction with a menu preview of the dish name and its regular price. Finally, (iii) selecting a dish from the preview menu presents the full data of the selected dish to the customer.

Customers of the restaurants can ask to deliver some dishes at their address. The database must record all deliveries of all dishes for all restaurants. The number of deliveries can grow indefinitely.

A **delivery** is described by the following data and is typically accessed by the delivery courier who needs to know the following information in their mobile app for managing the deliveries.

- A unique identification timestamp automatically generated by the database.
- The customer id (e.g., “c123456”), the customer full name, the customer phone number, and the customer address (in the same format as the restaurant address).
- The list of dishes, which are rarely accessed by the delivery courier.
- The restaurant id, name, and address.

Provide below a relevant sample document for each collection you design to address the described context.

- E.g., if you design the database with four collections, a sample document for each collection must be provided, writing explicitly the collection name before each sample document.
- Sample documents must provide values for all the attributes you expect to store inside them.
- When lists are used, provide at least 2 sample items in each list.
- Please use a new line for each new attribute of the document, also for sub-documents.
- Please use pretty indentation to make the document human readable.

Sample document of collection ABC

```
{ "attributeA": "ValueA",
"attributeB": [
{ "ABC": [1, 2, 3] },
```

```
                                { "XYZ":  [3, 4, 5] },
                                ],
                                "attributeC": [
                                { "X": "ABC",
                                "Y": "XYZ" },
                                { "X":  "A123",
                                "Z": "Z123" },
                                ],
                                ...
                        }
```

}

Feedback

Restaurant collection

```
{
“id”: “r123456”,
“name”: “Da Antonio”,
“address”:
       {
       “streetName”: “Via Po”,
       “houseNumber”: “1A”,
       “city”: “Torino”,
       “postalCode”: 10100,
       “countryName”: “Italy”
       },
“categories”: [“pizza”, “sea food”, “street food”, …],

“dishes”: [
       {“id”: “d123456”,
        “name”: “Super-size hot dog”,
        “regular_price”: 5.0,
       },
       {“id”: “d345678”,
        “name”: “Another dish”,
        “regular_price”: 123.45,
       }, …]

“better_dishes_solution”: {
       “d123456”: {
              “name”: “Super-size hot dog”,
              “regular_price”: 5.0,
             },
       “d345678”: {
              “name”: “Another dish”,
              “regular_price”: 123.45,
             }
       }

}
```

Dishes collection


```
{
“id”: “d123456”,
“name”: “Super-size hot dog”,
“prices”:
       {
       “regular”: 5.0,
       “promo”: 4.0,
       “discounted”: 3.5,
       },
“weight”: 0.3,
“ingredients”: [“bread”, “wurstel”, …]
“pictureFiles”: [“d123456_A.jpg”, “d123456_B.jpg”, …]
}
```

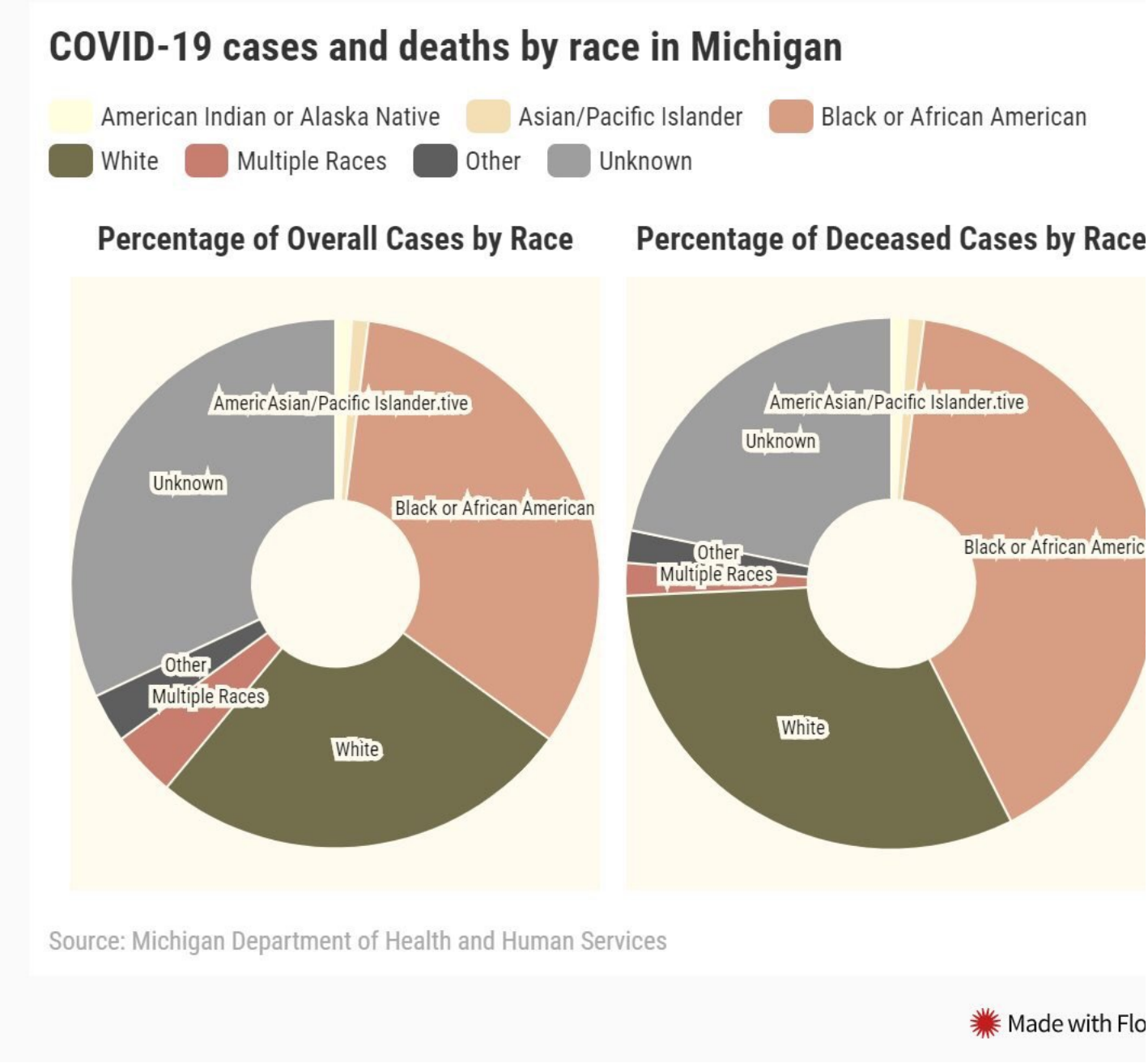

Deliveries collection


```
{
“timestamp”: “1234567890123”,
“customer”:
       {
       “ID”: “c123456”,
       “name”: “John Doe”,
       “phone”: “+393939393939”
       “address”:
              {
              “streetName”: “Via Po”,
              “streetNumber”: “99B”,
              “city”: “Torino”,
              “postalCode”: 10100,
              “country”: “IT”
              }
       },
“dishes”: [d123456, d234567, …],
“restaurant”:
       {
       “id”: “r123456”,
       “name”: “Da Antonio”,
       “address”:
              {
              “streetName”: “Via Po”,
              “streetNumber”: “1A”,
              “city”: “Torino”,
              “postalCode”: 10100,
              “country”: “IT”
```

```
            },
        }
}
```

Informazione

Testo informativo



Analyze the above graph comparing the number of COVID-19 cases and deaths by race in Michigan.

Feedback

The suggested redesigns of the plot are available at this link: https://public.tableau.com/profile/mtorchiano#!/vizhome/shared/86F2P2JJH

Question 8

Text of the question

Question

Is there a clearly defined question addressed by the visualization? Write it down.

Question 9

Text of the question

Data

Is the data quality appropriate? Identify the inadequate characteristics and explain.

Question 10

Text of the question

Visual Proportionality

Are the values encoded in a uniformly proportional way?

Question 11

Text of the question

Visual Utility

All the elements in the graph convey useful information?

Question 12

Text of the question

Visual Clarity

Are the data in the graph clearly identifiable and understandable (properly described)?

Question 13

Text of the question

Design data

Design the visualization based on the following data structure (to be completed).

Question 14

Text of the question

Design schema

Fill in the required schema elements; formulas can be used if required.

Question 15

Text of the question

Sketch of the resulting graph

Describe in words the design proposal.