Risposta non data

Punteggio max.: 2,00 The following document structure represents a document of a sale.

Each document collects some information about the customer, the list of sold items and the shop.

```
{
 "_id": ObjectId(),
 "saleDate": <Date>,
 "items": [
  {
    "name": <string>,
    "tags": [ <string>, ... ],
    "price": <float>,
    "quantity": <integer>
  }, ...
 ],
 "storeLocation": <string>,
 "customer": {
  "gender": <string>,
  "age": <integer>,
  "email": <string>,
  "satisfaction": <float>
 },
 "couponUsed": <boolean>,
 "purchaseMethod": <string>
}
```

Write a MongoDB query to find all sales documents where male customers over 50 years old used a coupon and purchased at least one item with the "school" tag. Show only the purchase date and the store location.

```
db.sales.find(
  { items.tags: "school",
    customer.gender: "M",
    customer.age: {$gt: 50},
    couponUsed: true },
    {storelocation: 1, salesDate: 1, _id0}
)
```

Risposta non data

Punteggio max.: 4,00

```
Fact(AutID, BuyerID, TimeID, TotalIncome,
TotalArtworksSold)
Author(AutID, ArtisticMovement, Alive, Nation, Region)
Buyer(BuyerID, Type, ReliabilityIndex, Nation, Region)
Time(TimeID, Month, 4m, 6m, year)
```

- Buyers can be either private consumers, Art houses, or museum
- Reliability index is an integer number between 1 and 5
- Alive can be either 0 or 1

Separately for each alive author, buyer type and semester (6m), select:

A. the ratio between total income and the total income for that author

B. the monthly average income (note that some months might be missing, you must consider only existing months)

C. assign an increasing rank to each author for each semester, according to its total income (rank first the highest).

Risposta non data

Punteggio max.: 0,50

An energy company provides its service in various Italian cities on numerous buildings, where smart metering systems are installed to measure the energy consumed and produced every 5 minutes. You are prompted to design a data warehouse to analise consumed and produced energy amounts (kWh) based on the following features.

- Weather conditions in terms of presence or absence of rain, fog, snow, cloud, and sun
  - Multiple weather conditions can be present at the same time
- Additional contextual measurements
  - Wind speed (high-medium-low), external temperature (very high-high-medium-low-very low), and solar radiation (high-medium-low).
- Considering the building
  - Year of construction, the number of rooms, the number of people usually living inside, the type of building (residential, office, or public service), the city, the city district and the region.
- The analysis must be performed for each time slot of the day (morning 6-12, afternoon 12-18, evening 18-22, night 22-6), date, day of the week, holiday or nonholiday, month, 4-month period, 6-month period, year, month of the year

Select, among the following dimensions, those that meet the requirements described in the problem specification (at most one answer is correct).









Risposta non data Punteggio max.: <b>and</b> 1,50	An energy company provides its service in various Italian cities on numerous buildings, where smart metering systems are installed to measure the energy consumed produced every 5 minutes. You are prompted to design a data warehouse to analise consumed and produced energy amounts (kWh) based on the following features.
	<ul> <li>Weather conditions in terms of presence or absence of rain, fog, snow, cloud, and sun         <ul> <li>Multiple weather conditions can be present at the same time</li> </ul> </li> </ul>
	<ul> <li>Additional contextual measurements Wind speed         <ul> <li>(high-medium-low), external temperature (very high-high-medium-low-very low), and solar radiation (high-medium-low).</li> </ul> </li> <li>Considering the building</li> </ul>
	<ul> <li>Year of construction, the number of rooms, the number of people usually living inside, the type of building (residential, office, or public service), the city, the city district and the region.</li> </ul>
	<ul> <li>The analysis must be performed for each time slot of the day (morning 6-12, afternoon 12-18, evening 1822, night 22-6), date, day of the week, holiday or nonholiday, month, 2-month period, quarter, year, month of the year</li> </ul>
	Select all and only the required measures of the fact table in the conceptual schema design among the following (multiple-choice question). <b>Hint:</b> do consider the dimensions defined by the previous answers.
	Scegli una o più alternative:
	(a) Average number buildings per city (count)
	(b) Average produced energy (kWh)
	(c) Total monthly consumed energy (kWh)
	(d) Total consumed energy cost (euros)
	(e) Average produced energy (kWh)
	(I) Iotal consumed energy (KVVN)
	(y) rotal energy cost per building (euros)
	(i) Total produced energy (kWh)
	(j) Total buildings (count)

- (k) Average consumed energy per building (kWh)
- (I) Total buildings per region (count)
- (m) Total produced energy cost (euros)
- (n) Average consumed energy cost (euros)

Risposta errata.

La risposta corretta è: Total consumed energy (kWh), Total produced energy (kWh)

Domanda **5** 

The approximation pattern has the advantage of:

Risposta non data Punteggio max.:

1,50

- (a) fewer writes to the database
- (b) reduction in the overall number of documents in a collection
- (c) none of the answers is correct
- (d) improvement of performance when there are a lot of join operations
- (e) reduction in the overall size of the working set

Risposta errata.

La risposta corretta è: fewer writes to the database





# **Visual Proportionality**

Are the values encoded in a uniformly proportional way?

Domanda **7** Risposta non data Punteggio max.: 1,25



## Data

Is the data quality appropriate? Select true answers only.

Scegli una o più alternative:

- (a) Data is updated to the first half of 2022.
- (b) Data is not consistent as different periods of time are considered to measure the net worth.
- (c) Data is understandable as the meaning the values is clear (billions of dollars).
- (d) Some values are reported without decimal digits while others with one decimal digit.
- (e) Data is not credible because the source available is not a trusted one.
- (f) The net worth should have been measured on 1st January 2023, not on 15th December 2022.
- (g) Data is complete because several billionaries are considered.
- (h) Data is not understandable because different unit of measures are used.
- (i) Data is not complete because only 10 billionaries have been considered.
  - (j) Data is not accurate as the values look too rounded to be real.

La risposta corretta è: Data is not complete because only 10 billionaries have been considered., Data is understandable as the meaning the values is clear (billions of dollars)., Some values are reported without decimal digits while others with one decimal digit., The net worth should have been measured on 1st January 2023, not on 15th December 2022.

Risposta non data

Punteggio max.: 2,00

The following document structure represents a post uploaded on a blog.

```
{
 " id": ObjectId(),
 "body": "Lorem ipsum ...",
 "permalink": <string>,
 "author": "machine",
 "title": "Gettysburg Address",
 "tags": [
  "avenue",
  "distance"
   ],
 "comments": [
  {
    "body": "Lorem ipsum ...",
    "email": "name@domain.com",
    "author": "Linnie Weigel"
  }, ...
 ],
 "date": Date("2023-01-18")
}
```

You are required to find the 3 most popular tags, i.e., those associated with the largest number of posts. Which of the following queries satisfies the request?

) (a)

db.posts.aggregate( [ { \$group: { \_\_id: '\$tags', n: { \$sum: 1 } }}, {\$ limit: 3} ])

(b)

db.posts.aggregate( [{\$unwind: '\$tags'}, { \$group: { \_\_id: '\$tags', n: { \$sum: 1 } }}, { \$sort: { n: -1 } }, {\$ limit: 3} ])



La risposta corretta è:

```
db.posts.aggregate( [{$unwind: '$tags'},
{ $group: {
__id: '$tags',
n: { $sum: 1 } }},
{ $sort: { n: -1 } },
{$ limit: 3} ])
```

Risposta non data

Punteggio max.: 0,25



## Design data

Design the visualization based on the following data structure.

BILLIONARIE	Scegli	~
WORTH_JANUARY	Scegli	~
WORTH_DECEMBER	Scegli	~

#### Risposta errata.

La risposta corretta è: BILLIONARIE – Dimension, WORTH\_JANUARY – Measure, WORTH\_DECEMBER – Measure

Risposta non data

Punteggio max.: 1,50 Which one of the following sentences related to discretization (or quantization) is true?

- (a) Discretization could be used to encode an ordinal-friendly visual attribute
- (b) Discretization can be applied without creating intervals of values
- (c) Discretization may increase the precision of the measures
- (d) Discretization is a technique that maps ordinal measures into quantitative ones
- (e) Discretization must be used to encode values with a color scale

Risposta errata.

La risposta corretta è: Discretization could be used to encode an ordinalfriendly visual attribute



## **Visual Utility**

All the elements in the graph convey useful information?

Risposta non data

Punteggio max.: 4,00 Design a MongoDB database to store orders from shops for a shipping company.

The application must efficiently retrieve all shipping orders for each month and for each VAT number of the stores.

Each shipping order is characterized by a timestamp, the number of packages, the total weight of the packages, the issuing store, and the shipping address (full name of the recipient, street name, street number, city, state).

Stores from which orders are placed are characterized by a name, address, VAT number and contact information such as telephone, website (if available) and social profiles.

For the address you are required to keep track of the street name, street number, city and state, as well as the geographical coordinates (latitude and longitude).

For the social profiles, the name of the social service (e.g., "Twitter") and the profile url must be saved. Multiple social profiles can be present, also on the same social service (e.g., multiple Twitter accounts).

Given a specific month, in addition to the list of all orders placed by a store, some statistics must be readily presented. Specifically, the total number of orders, the average weight of packages, and the total number of packages must be returned.

Moreover, for each order, the name of the store should be retrieved without additional lookups.

Write a sample document for each collection of the database.

Besides the sample documents, explicitly indicate the design patterns used.

ORDER

```
{
_id: ObjectId(),
store: {
_id: ObjectId(),
VAT: <string>
}
month: <date>,
statistics: {
     tot_count: <number>,
     tot_weight: <number>,
     tot_packages: <number>
  },
orders: [
{ ts: <datetime>,
 n_packages: <number>,
 weight: <string>,
 store: {
     _id: ObjectId(),
     name: <string>
   },
 address: {
     full_name: <string>,
     street: <string>,
     number: <number>,
     city: <string>,
     state: <string>
  }
}]
}
```

## STORE

```
{_id: ObjectId(),
name: <string>,
VAT: <string>,
address: {
street: <string>,
number: <number>,
city: <string>,
state: <string>
},
          type:"Point",
loc: {
     coordinates":[ <number>, <number>]
}
contacts: {
     tel: <string>,
     website: <url>,
     social: [{k:<string>, v:<url>}]
```

#### Pattern used:

- Bucket pattern to track orders
- **Computed** pattern for the shipping stats in the order collection
- **Extended reference** for the order collection to show the store information
- **Polymorphic** for contact information
- Attribute pattern for social profile [optional]

Risposta non data

Punteggio max.: 0,50

An energy company provides its service in various Italian cities on numerous buildings, where smart metering systems are installed to measure the energy consumed and produced every 5 minutes. You are prompted to design a data warehouse to analise consumed and produced energy amounts (kWh) based on the following features.

- Weather conditions in terms of presence or absence of rain, fog, snow, cloud, and sun
  - Multiple weather conditions can be present at the same time
- Additional contextual measurements
  - Wind speed (high-medium-low), external temperature (very high-high-medium-low-very low), and solar radiation (high-medium-low).
- Considering the building
  - Year of construction, the number of rooms, the number of people usually living inside, the type of building (residential, office, or public service), the city, the city district and the region.
- The analysis must be performed for each time slot of the day (morning 6-12, afternoon 12-18, evening 18-22, night 22-6), date, day of the week, holiday or nonholiday, month, 4-month period, 6-month period, year, month of the year

Select, among the following dimensions, those that meet the requirements described in the problem specification (at most one answer is correct).









Risposta non data

This is a blank question to be used as your personal notepad during the exam.

Non valutata

Anything written here will NOT be evaluated.

Domanda 15 Risposta non data Punteggio max.: 0,25



## Question

Which one of the following questions represents the purpose of this visualization?

- (a) Which is the correlation between billionaries worth and their businesses?
- (b) What is the trend of net worth of the richest billionaries over several years?
- (c) What is the rank of the richest billiories?
- (d) What is the change of net worth of the richest billionaries in 2022?
- (e) Which one is the richest billionarie?

#### Risposta errata.

La risposta corretta è: What is the change of net worth of the richest billionaries in 2022?

Risposta non data

Punteggio max.: 0,50

An energy company provides its service in various Italian cities on numerous buildings, where smart metering systems are installed to measure the energy consumed and produced every 5 minutes. You are prompted to design a data warehouse to analise consumed and produced energy amounts (kWh) based on the following features.

- Weather conditions in terms of presence or absence of rain, fog, snow, cloud, and sun
  - Multiple weather conditions can be present at the same time
- Additional contextual measurements
  - Wind speed (high-medium-low), external temperature (very high-high-medium-low-very low), and solar radiation (high-medium-low).
- Considering the building
  - Year of construction, the number of rooms, the number of people usually living inside, the type of building (residential, office, or public service), the city, the city district and the region.
- The analysis must be performed for each time slot of the day (morning 6-12, afternoon 12-18, evening 18-22, night 22-6), date, day of the week, holiday or nonholiday, month, 4-month period, 6-month period, year, month of the year

Select, among the following dimensions, those that meet the requirements described in the problem specification (at most one answer is correct).













```
Tickets(ConsID, SellID, EventID, BenID, TimeID,
TotalIncome, TotalTicketsSold)
Consumer(ConsID, Province, Region, State)
Seller(SellID, SellerName, IsOnline)
Event (EventID, EventName, Private, Typology,
Region, State)
Benefits(BenID, Benefit)
Time(TimeID, Month, 3m, 4m, 6m, year)
```

1,00

Risposta non data



## Design schema & Sketch

Fill in the required schema elements; formulas can be used if required. Then describe in words the design proposal.

1,25

Risposta non data

Punteggio max.:

The following document structure represents a document of a sale.

Each document collects some information about the customer, the list of sold items and the shop.

3,00

```
{
 "_id": ObjectId(),
 "saleDate": <Date>,
 "items": [
  {
    "name": <string>,
    "tags": [ <string>, ... ],
    "price": <float>,
    "quantity": <integer>
  }, ...
 ],
 "storeLocation": <string>,
 "customer": {
  "gender": <string>,
  "age": <integer>,
  "email": <string>,
  "satisfaction": <float>
 },
 "couponUsed": <boolean>,
 "purchaseMethod": <string>
}
```

Considering only items with a selling price above 100, display for each month and year the total number of products sold and the total revenue.

Note: to extract the month and the year from a date field, you can use respectively the operators {\$month: '\$dateFieldName'} and {\$year: '\$dateFieldName'}.

db.sales.a	ggregate( [{ \$unwind: '\$items'},
	{ \$match: { 'items.price': { \$gt: 100 } }},
	{ \$group: {
	_id: {
	month: { \$month: '\$saleDate'},
	year: { \$year: '\$saleDate'}
	},
	n: { \$sum: 1 },
	income: { \$sum: '\$items.price'}
	}
	}])

Domanda 20 Risposta non data Punteggio max.: 4,00	Fact( <u>AutID</u> , <u>BuyerID</u> , <u>TimeID</u> , TotalIncome, TotalArtworksSold) Author( <u>AutID</u> , ArtisticMovement, Alive, Nation, Region) Buyer( <u>BuyerID</u> , Type, ReliabilityIndex, Nation, Region) Time( <u>TimeID</u> , Month, 4m, 6m, year)
	<ul> <li>Buyers can be either private consumers, Art houses, or museum</li> <li>Reliability index is an integer number between 1 and 5</li> <li>Alive can be either 0 or 1</li> </ul>
	<ul> <li>Separately for each artistic movement and month, select:</li> <li>A. the average income from all the artworks sold</li> <li>B. the percentage of artworks sold for each month with respect to the total number of artworks sold of that year</li> <li>C. the cumulative monthly number of artworks sold since the beginning of the year</li> </ul>





## **Visual Clarity**

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