

**Domanda 1**

Risposta non data

Punteggio max.:

1,50

In MongoDB, given a replica set with 7 members, the fault tolerance (number of nodes that can fail) is equal to:

---

- (a) 5
- (b) 3
- (c) 4
- (d) 2
- (e) 1
- (f) none of the other answers is correct

Risposta errata.

La risposta corretta è: 3

**Domanda 2**

Risposta non data

Punteggio max.:

1,50

The NoSQL design pattern named “document versioning” has the disadvantage of:

---

- (a) being difficult to be implemented on existing systems
- (b) always requiring to execute twice each write operation
- (c) none of the other answers is correct
- (d) often requiring two indexes for the same field during data migration
- (e) negatively impacting the performance of the queries

Risposta errata.

La risposta corretta è: always requiring to execute twice each write operation

**Domanda 3**

Risposta non data

Punteggio max.:

1,50

In a dataset about the weather by date and location you find the date expressed with hours and minutes. In the context of data quality, it is an issue of...

---

- (a) Completeness
- (b) Credibility
- (c) Precision
- (d) Accuracy
- (e) Consistency

Risposta errata.

La risposta corretta è: Precision

**Domanda 4**

Risposta non data

Punteggio max.: 0,50

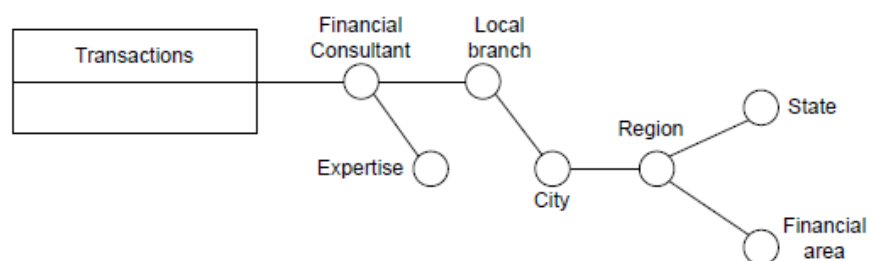
Data analysts of an international financial group are interested in analysing the behaviour of their local branches and their customer transactions, in terms of average amount (euro) per transaction.

They would like the analysis to address the following features:

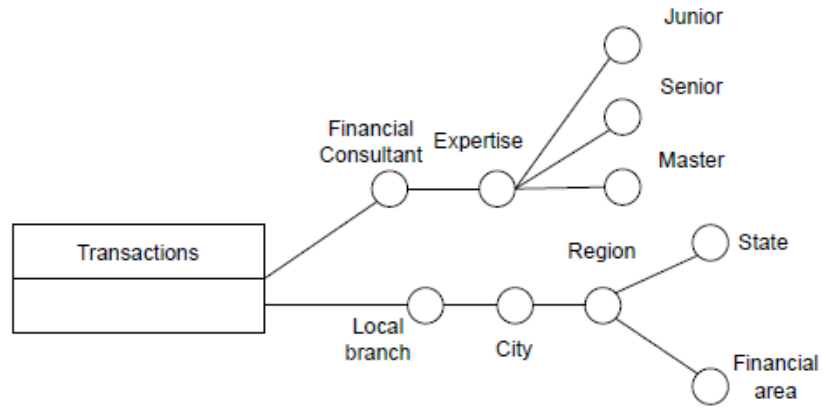
- Local branches, financial consultants, and their expertise level
  - Each consultant can execute financial transactions.
  - Each financial consultant can work only in one local branch.
  - Each financial consultant has a level of expertise (i.e., Junior, Senior, or Master).
  - The data warehouse is required to track the financial consultant, their expertise level, their local branch, also in terms of their city, region, state, and financial area.
  - A financial area is intended to cover contiguous regions, which might be in different states.
  - A state can be associated with more than one financial area.
- Financial services
  - Each transaction is associated with a specific financial service, e.g., insurance, international markets, or bank accounts.
  - The data warehouse is required to track the financial services of the transactions.
- Customers
  - The customer city, region and state are analyzed.
  - The level of risk of the customers, which can be low, moderate, or high.
  - Customers can be interested in different investments, e.g., stocks, bonds, retirement plans, etc. The list of investments can be extended over time and can be potentially quite long.
  - The data warehouse is required to track the customers, their investment interests, their risk levels, their city, region, and state.
- The analysis of the transactions must be performed for each day of the week, holiday (yes/no), date, month, 3-month period, semester, and year.

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

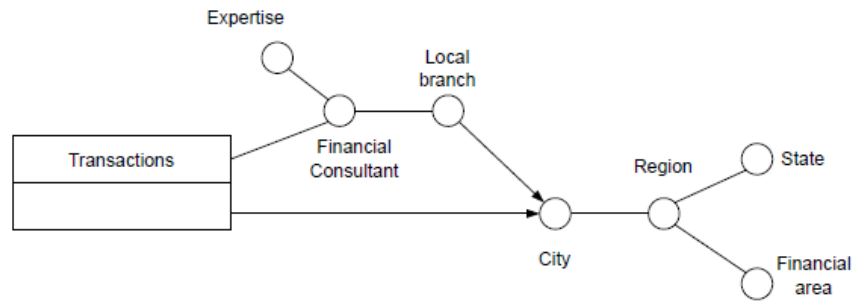
(a)



(b)

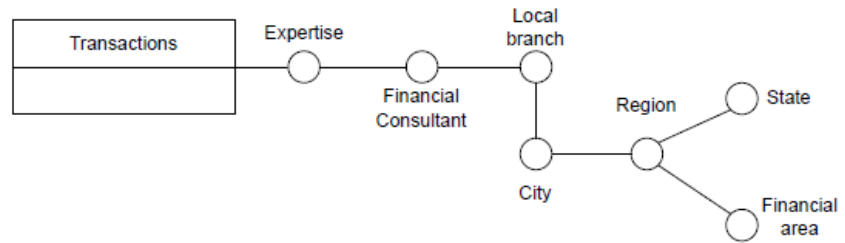


(c)

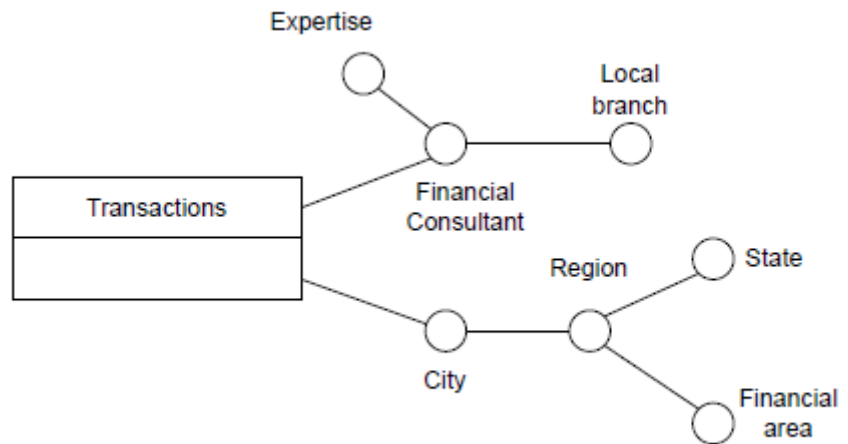


(d) None of the others is correct.

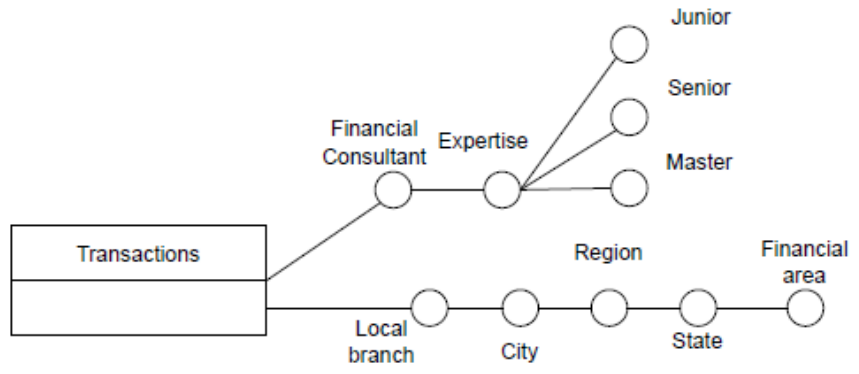
(e)



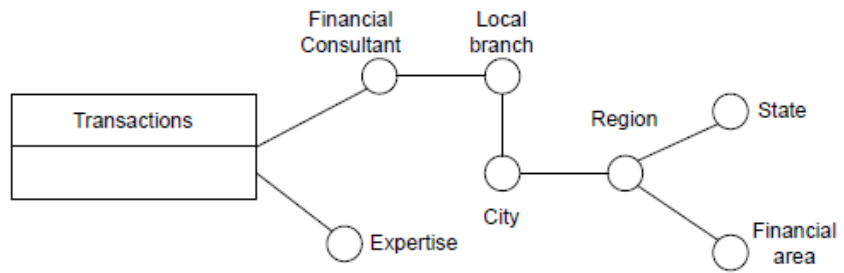
(f)



○ (g)

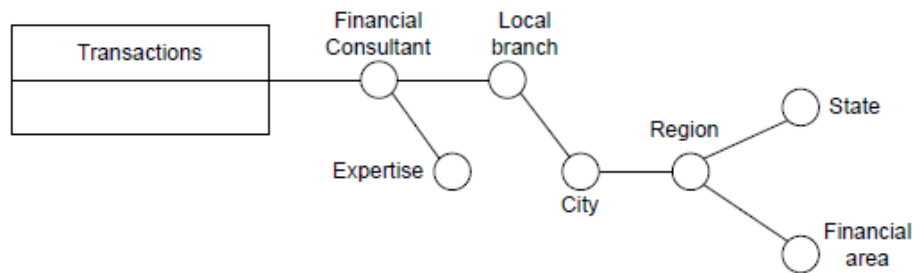


○ (h)



Risposta errata.

La risposta corretta è:



**Domanda 5**

Risposta non data

Punteggio max.:

0,50

Data analysts of an international financial group are interested in analysing the behaviour of their local branches and their customer transactions, in terms of average amount (euro) per transaction.

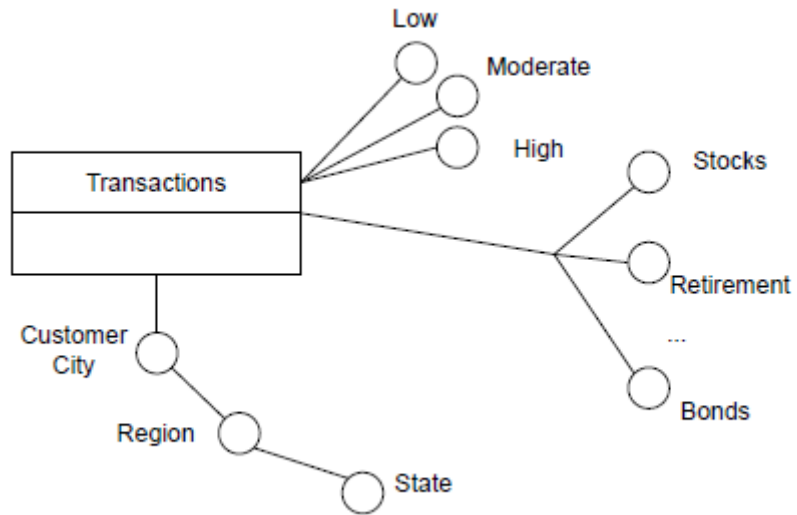
They would like the analysis to address the following features:

- Local branches, financial consultants, and their expertise level
  - Each consultant can execute financial transactions.
  - Each financial consultant can work only in one local branch.
  - Each financial consultant has a level of expertise (i.e., Junior, Senior, or Master).
  - The data warehouse is required to track the financial consultant, their expertise level, their local branch, also in terms of their city, region, state, and financial area.
  - A financial area is intended to cover contiguous regions, which might be in different states.
  - A state can be associated with more than one financial area.
- Financial services
  - Each transaction is associated with a specific financial service, e.g., insurance, international markets, or bank accounts.
  - The data warehouse is required to track the financial services of the transactions.
- Customers
  - The customer city, region and state are analyzed.
  - The level of risk of the customers, which can be low, moderate, or high.
  - Customers can be interested in different investments, e.g., stocks, bonds, retirement plans, etc. The list of investments can be extended over time and can be potentially quite long.
  - The data warehouse is required to track the customers, their investment interests, their risk levels, their city, region, and state.
- The analysis of the transactions must be performed for each day of the week, holiday (yes/no), date, month, 3-month period, semester, and year.

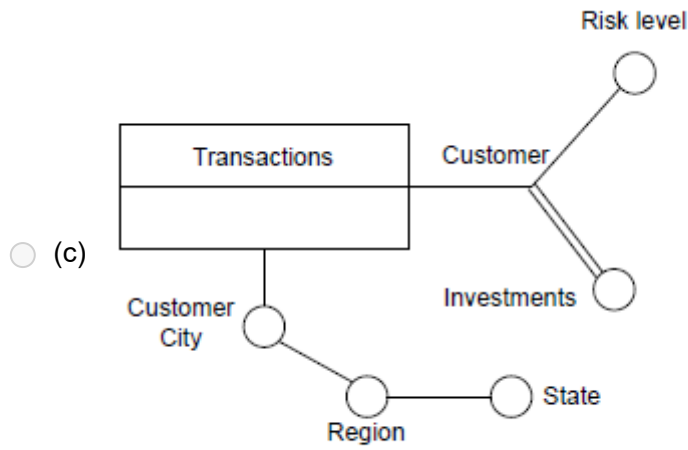
Select, among the following dimensions, those that meet the requirements described in the problem specification (at most one answer is correct).

---

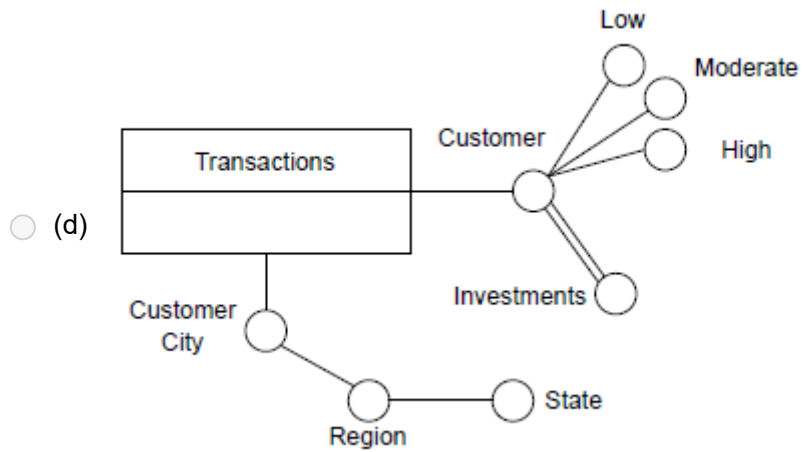
(a)



(b) None of the others is correct.

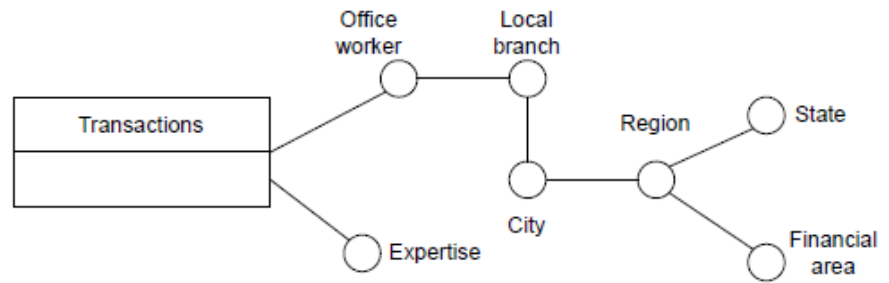


(c)

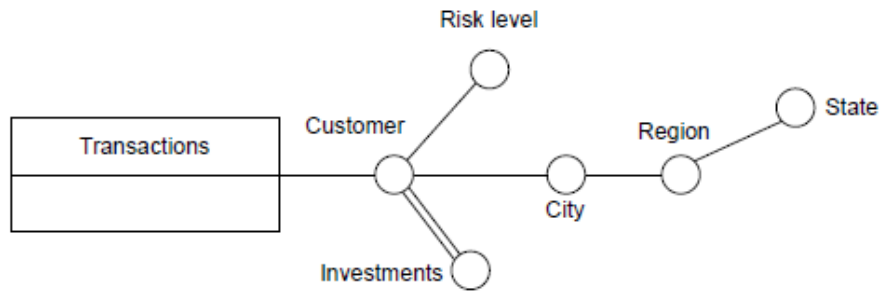


(d)

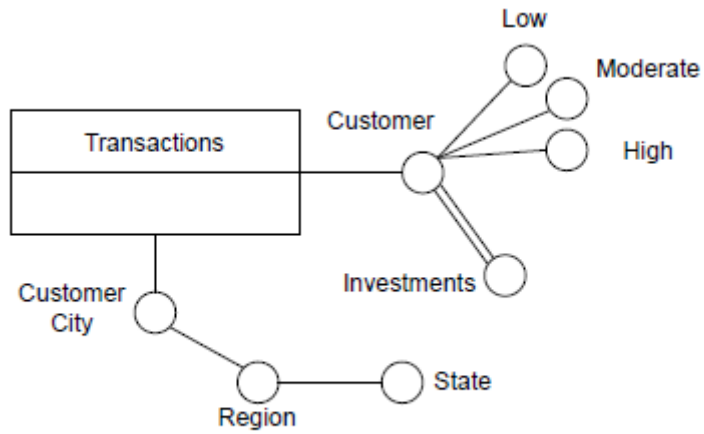
(e)



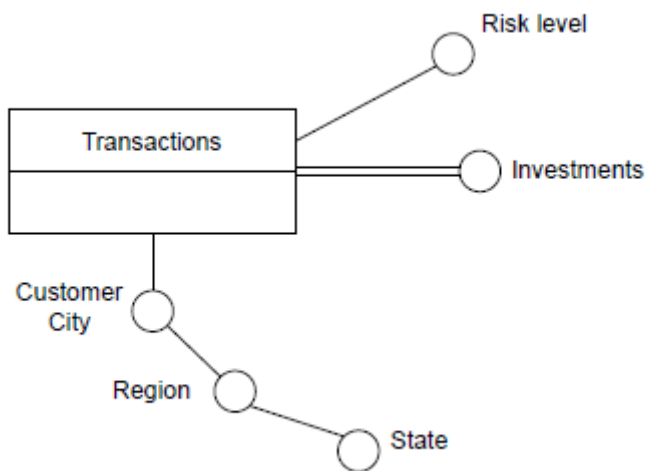
(f)



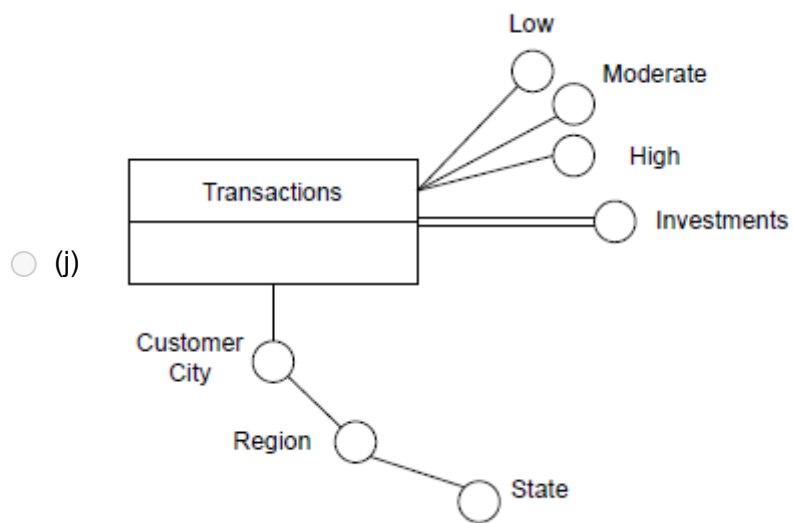
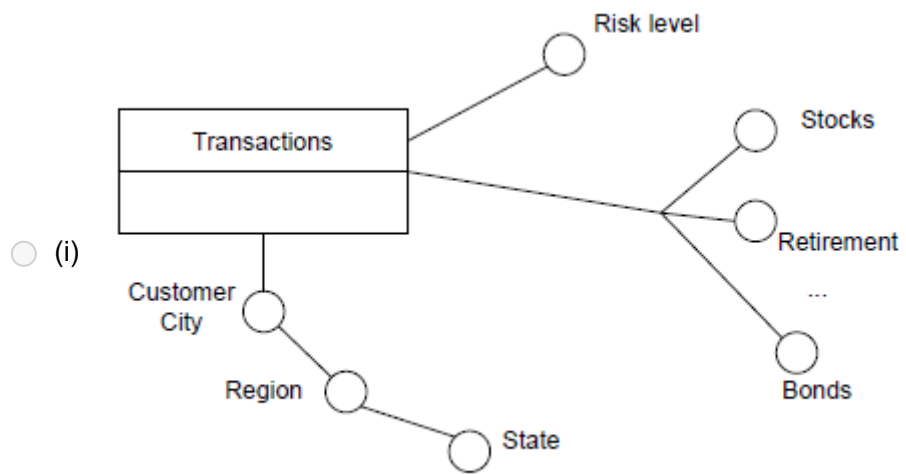
(g)



(h)

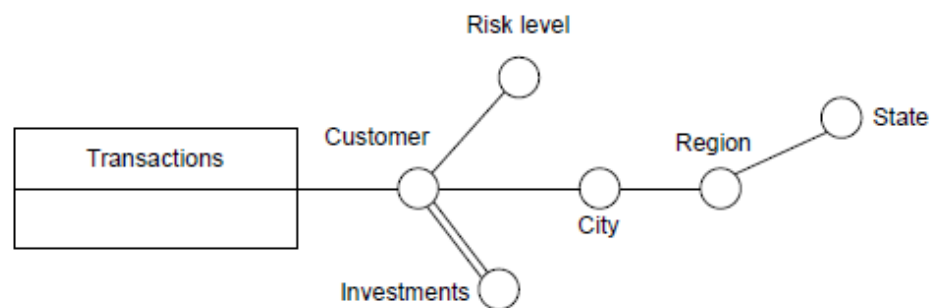






Risposta errata.

La risposta corretta è:



**Domanda 6**

Risposta non data

Punteggio max.:

0,50

Data analysts of an international financial group are interested in analysing the behaviour of their local branches and their customer transactions, in terms of average amount (euro) per transaction.

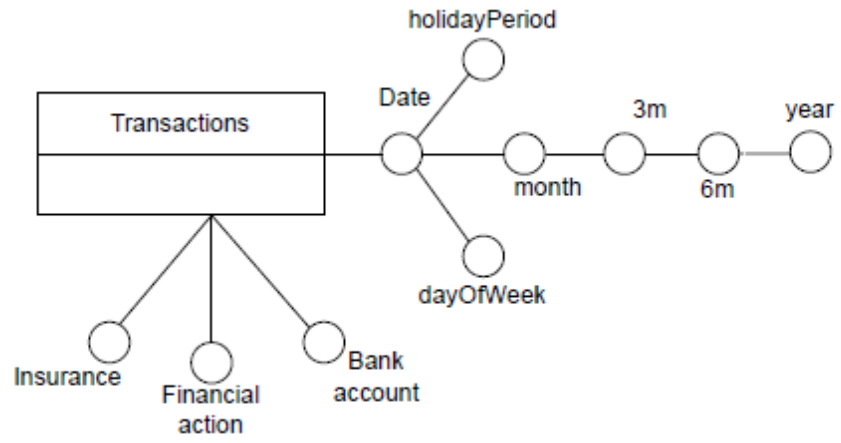
They would like the analysis to address the following features:

- Local branches, financial consultants, and their expertise level
  - Each consultant can execute financial transactions.
  - Each financial consultant can work only in one local branch.
  - Each financial consultant has a level of expertise (i.e., Junior, Senior, or Master).
  - The data warehouse is required to track the financial consultant, their expertise level, their local branch, also in terms of their city, region, state, and financial area.
  - A financial area is intended to cover contiguous regions, which might be in different states.
  - A state can be associated with more than one financial area.
- Financial services
  - Each transaction is associated with a specific financial service, e.g., insurance, international markets, or bank accounts.
  - The data warehouse is required to track the financial services of the transactions.
- Customers
  - The customer city, region and state are analyzed.
  - The level of risk of the customers, which can be low, moderate, or high.
  - Customers can be interested in different investments, e.g., stocks, bonds, retirement plans, etc. The list of investments can be extended over time and can be potentially quite long.
  - The data warehouse is required to track the customers, their investment interests, their risk levels, their city, region, and state.
- The analysis of the transactions must be performed for each day of the week, holiday (yes/no), date, month, 3-month period, semester, and year.

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

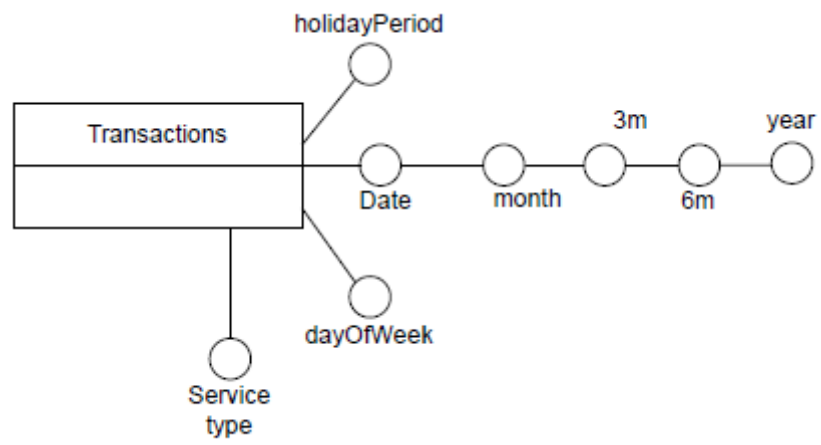
---

(a)

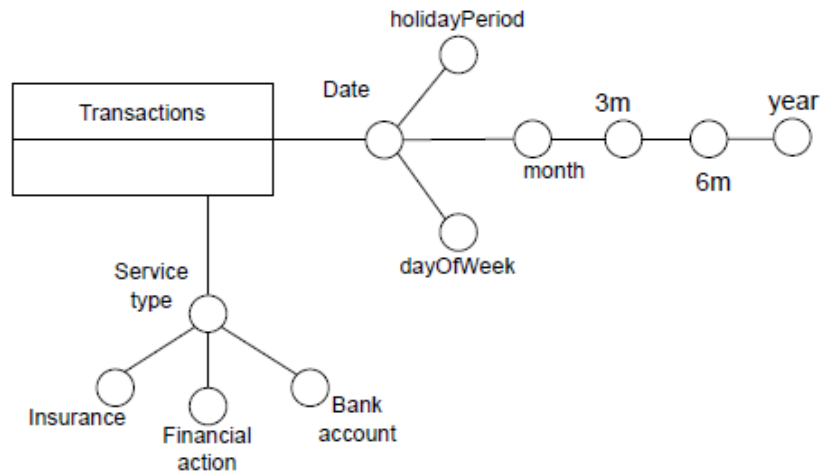


(b) None of the others is correct.

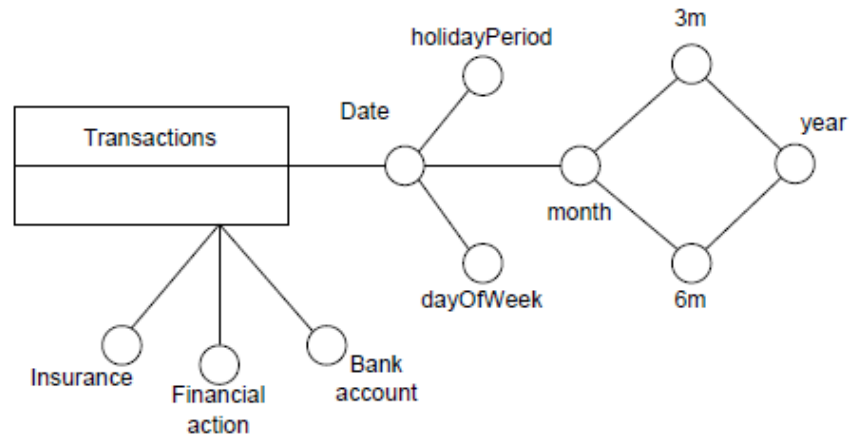
(c)



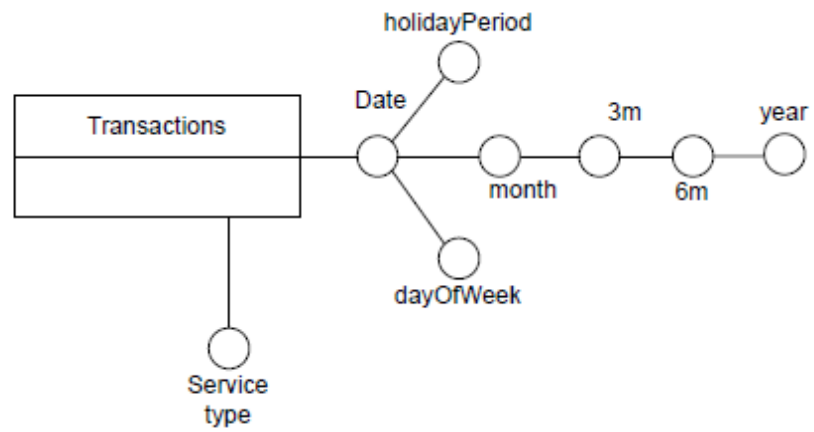
(d)



(e)

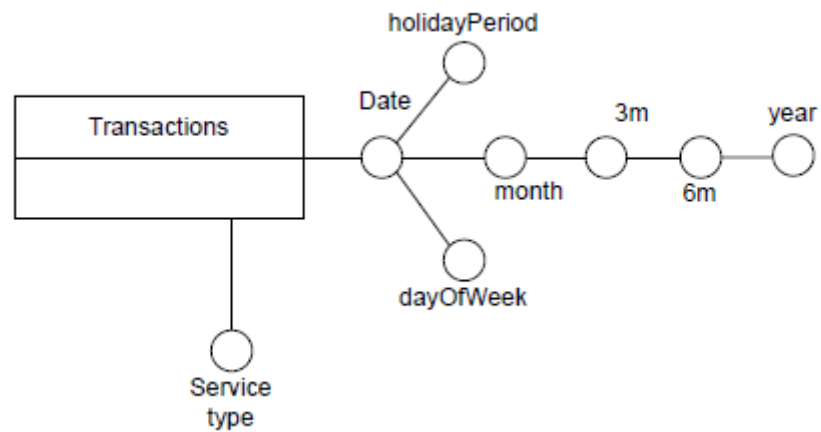


(f)



Risposta errata.

La risposta corretta è:



**Domanda 7**

Risposta non data

Punteggio max.:

1,50

Data analysts of an international financial group are interested in analysing the behaviour of their local branches and their customer transactions, in terms of average amount (euro) per transaction.

They would like the analysis to address the following features:

- Local branches, financial consultants, and their expertise level
  - Each consultant can execute financial transactions.
  - Each financial consultant can work only in one local branch.
  - Each financial consultant has a level of expertise (i.e., Junior, Senior, or Master).
  - The data warehouse is required to track the financial consultant, their expertise level, their local branch, also in terms of their city, region, state, and financial area.
  - A financial area is intended to cover contiguous regions, which might be in different states.
  - A state can be associated with more than one financial area.
- Financial services
  - Each transaction is associated with a specific financial service, e.g., insurance, international markets, or bank accounts.
  - The data warehouse is required to track the financial services of the transactions.
- Customers
  - The customer city, region, and state are analyzed.
  - Customers can be interested in different categories of investments, e.g., stocks, bonds, retirement plans, etc. Such a list can be extended over time and can be potentially quite long. The data warehouse is required to track the investment category of interest of the customers.
  - The level of risk for the customers can be low, moderate, or high.
- The analysis of the transactions must be performed for each day of the week, holiday (yes/no), date, month, 3-month period, semester, and year.

Select all and only the required measures of the fact table in the conceptual schema design among the following (multiple choice question).

**Hint:** do consider the dimensions defined by the previous answers.

---

Scegli una o più alternative:

- (a) Total amount of euros from companies (euros)
- (b) Total number of customers (count)
- (c) Total amount of transactions (euro)
- (d) Total number of customers (count)
- (e) Total number of transactions (count)
- (f) Average amount of euros per region (euros)
- (g) Average amount for state (euros)
- (h) Maximum amount per transaction (euros)
- (i) Average number of office workers per branch (count)
- (j) Average number of cities per financial area (count)

- (k) Average amount per category (euros)
- (l) Average number of customers per city (count)
- (m) Number of stock transactions (count)
- (n) Total number of local branches (count)

Risposta errata.

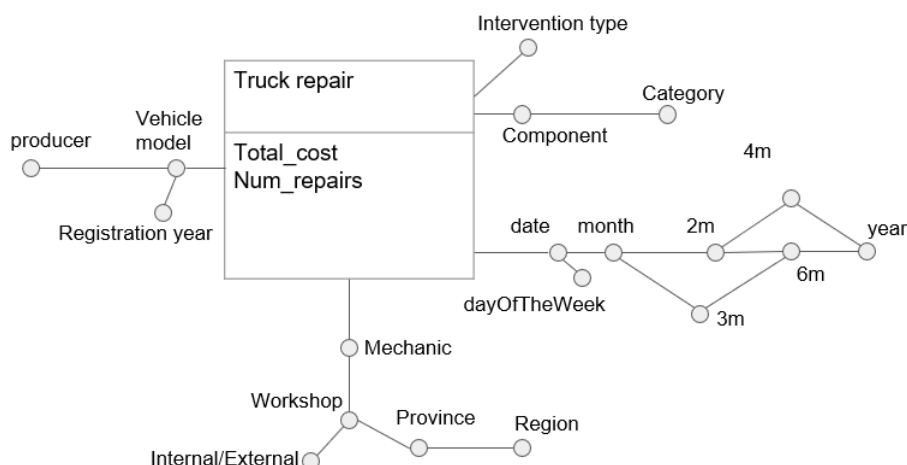
La risposta corretta è: Total number of transactions (count), Total amount of transactions (euro)

### Domanda 8

Risposta non data

Punteggio max.: 1,00

Given the following conceptual schema:



- Intervention type can be one among 3 possible values (substitution, repair, cleaning)
- Repairs are associated with a unique component (brake pads, tailer, etc.), which falls into a category
- Workshops are identified by a unique name. They can be considered as internal or external to the truck company.

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.

---

```
Repairs(InterventionType, CompID, TimeID, MechID,
VehicleID, Total_cost, Num_repairs)
Component(CompID, Component, Category)
Time(TimeID, Date, Month, DayOfTheWeek, 2m, 3m, 4m,
6m, year)
Mechanic(MechID, MechanicName, Workshop, Province,
Region, Internal)
Vehicle(VehicleID, Model, RegistrationYear,
Producer)
```

**Domanda 9**

Risposta non data

Punteggio max.:  
4,00

```
Fact(TimeId, AdvId, NumberOfClicks, ReadingTime)
Time(TimeId, date, month, 2M, 3M, 6M, year,
dayOfTheWeek)
Advertisement(AdvId, AdvTitle, Collection, Company,
referential, mythical, oblique, substantial)
```

- The data warehouse collects the total number of times each advertisement has been read (number of advertisements), and the total reading time of each advertisement, over time.
- More advertising can be part of one and only one collection.

Separately for each advertisement of category "referential" and bimester, select:

1. the average reading time per advertising
2. the total reading time of each advertisement with respect to the total number of clicks of all the advertisements of the same collection
3. assign a rank to each advertisement separately for collection and trimester, based on its number of streamings (rank 1st the advertisement with the highest number of clicks for each album)

Write the requested SQL query.

---

```
SELECT AdvId, 3M,
       SUM(ReadingTime) / SUM(NumberOfClicks) as A,
       SUM(ReadingTime) / SUM(SUM(NumberOfClicks)) OVER
(PARTITION BY Collection, T.3M) as B,
       RANK() OVER (PARTITION BY Collection, T.3M,
                   ORDER BY SUM(NumberOfClicks) DESC) as C
FROM Fact F, Advertisement A, Time T
WHERE A.AdvId=F.AdvId and T.TimeId=F.TimeId
AND referential = 1
GROUP BY AdvId, T.3M, Collection
```

**Domanda 10**

Risposta non data

Punteggio max.:

4,00

MusicStreaming(TimeId, PremiumUser, SongId,  
NumberOfStreamings, PlayingTime)  
Time(TimeId, date, month, 2M, 3M, 6M, year,  
dayOfTheWeek)  
Song(SongId, SongTitle, Album, Artist, indie, pop,  
classic, rock)

- The data warehouse collects the total number of times that each song has been listened to (number of streamings), and the total playing time of each song, over time.
- The cardinality of "PremiumUser" is 2, and it can be "0" for "no", "1" for "yes". The same is for indie, pop, classic, rock.

Separately for each album and for each month, select:

1. the cumulative number of streamings since the beginning of the year
2. the daily average playing time

Write the requested SQL query.

---

```
SELECT Album, month,  
       SUM(SUM(NumberOfStreamings)) OVER (  
         PARTITION BY Album, year  
         ORDER BY month  
         ROWS UNBOUNDED PRECEDING) as A,  
       SUM(PlayingTime)/COUNT(distinct date) as B  
FROM MusicStreaming M, Song S, Time T  
WHERE S.SongId=M.SongId and T.TimeId=M.TimeId  
GROUP BY T.month, S.Album, year
```



**Domanda 11**

Risposta non data

Punteggio max.:  
2,00

The following document structure represents a document of a book sold by an e-commerce website. Each document collects some information about the author and the last reviews. Each book is available in at least one language identified by the ISO code (e.g., "it" for Italian, "en" for English).

```
{
  "_id" : ObjectId("61fa5b8f6f631bb5339dc4b7"),
  "title" : "Lord of Rings ",
  "author" : {
    "id" : ObjectId("32af5b7a6f133ca5133dc4c8"),
    "name" : "J. R. R. Tolkien",
    "country" : "England"
  },
  "first_publication_year" : 1954,
  "languages" : [ "en", "it", "de", "fr", "es" ]
  "price_EU" : 20,
  "last_reviews" : [
    { "comment" : "Lorem ipsum ...", "rating" : 8, "date" : "2022-06-15" },
    { "comment" : "Lorem ipsum ...", "rating" : 9, "date" : "2022-06-13" },
    { "comment" : "Lorem ipsum ...", "rating" : 7.5, "date" : "2022-05-28" }
  ]
}
```

Write a MongoDB query to find all books written by English authors with a price greater than 20 euros and available in Italian language. Show only the title and price.

---

```
db.books.find (
{"author.country": "England", "price_EU":{$gt: 20}, languages:"it" } , {_id:0, title:1,
price:1}
)
```

**Domanda 12**

Risposta non data

Punteggio max.:

3,00

The following document structure represents a document of a book sold by an e-commerce website. Each document collects some information about the author and the last reviews. Each book is available in at least one language identified by the ISO code (e.g., "it" for Italian, "en" for English).

```
{
  "_id" : ObjectId("61fa5b8f6f631bb5339dc4b7"),
  "title" : "Lord of Rings ",
  "author" : {
    "id" : ObjectId("32af5b7a6f133ca5133dc4c8"),
    "name" : "J. R. R. Tolkien",
    "country" : "England"
  },
  "first_publication_year" : 1954,
  "languages" : [ "en", "it", "de", "fr", "es" ]
  "price_EU" : 20,
  "last_reviews" : [
    { "comment": "Lorem ipsum ...", "rating": 8, "date": "2022-06-15"},
    { "comment": "Lorem ipsum ...", "rating": 9, "date": "2022-06-13"},
    { "comment": "Lorem ipsum ...", "rating": 7.5, "date": "2022-05-28"}
  ]
}
```

Considering only books first published between the year 2000 and the year 2010, separately for each country of the author, select the average rating of the last reviews.

Show the countries with an average rating higher than 8 and sort the results in descending order according to the average rating.

```
db.collection_name.aggregate([
  { $match: {
    "year": { $gte: 2000, $lte: 2010 }
  } },
  { $unwind: '$last_reviews' },
  { $group: {
    '_id': '$author.country',
    'avg_rating': {
      '$avg': '$last_reviews.rating'
    }
  } },
  { $match: {
    "avg_rating": { $gt: 8 }
  } },
  { $sort: { 'avg_rating': -1 } }
])
```



**Domanda 13**

Risposta non data

Punteggio max.:

4,00

Design a MongoDB database to manage tv shows according to the following requirements.

The shows are characterized by a title, a description, a list of tags describing their topics, and the list of presenters. Each show has several editions. For each edition, the start date, the end date, and the number of viewers are known.

Presenters are characterized by their name, surname, email, date of birth and city of birth. Each presenter can have one or more online profiles on different platforms (e.g., Facebook, TikTok, Instagram, etc.). For each online profile, if available, the database tracks the corresponding URL of the profile and the number of followers. A presenter can present different tv shows.

Given a show, the database must be designed to efficiently provide the name and the surname of all its presenters.

Furthermore, given a show, the number of editions and the average number of viewers in each edition must be efficiently returned.

**Write a sample document for each collection of the database.**

**Explicitly indicate the design patterns used.**

---

**Presenter**

```
{
  _id: ObjectId(),
  name: <string>,
  surname: <string>,
  email: <string>,
  date_birth: <date>,
  city_birth: <string>,
  profiles: [
    { platform: <string>,
      url: <url>,
      followers: <number>},
    ...
  ]
}
```

**Show**

```
{
  _id: ObjectId(),
  title: <string>,
  description:<string>,
  tags: [ <string> ],
  presenters: [{
    _id: ObjectId(),
    name: <string>,
    surname: <string>},
    ...
  ],
  editions: [
    {start: <date>,
      end: <date>,
      viewers: <number>}
  ],
  n_editions: <number>,
  tot_viewers: <number>
}
```

**Pattern used:**

Attribute pattern to track the online profile information in the Presenter collection.

Extended reference pattern to track the presenter information associated with each show.

Bucket pattern to track when a show is provided.

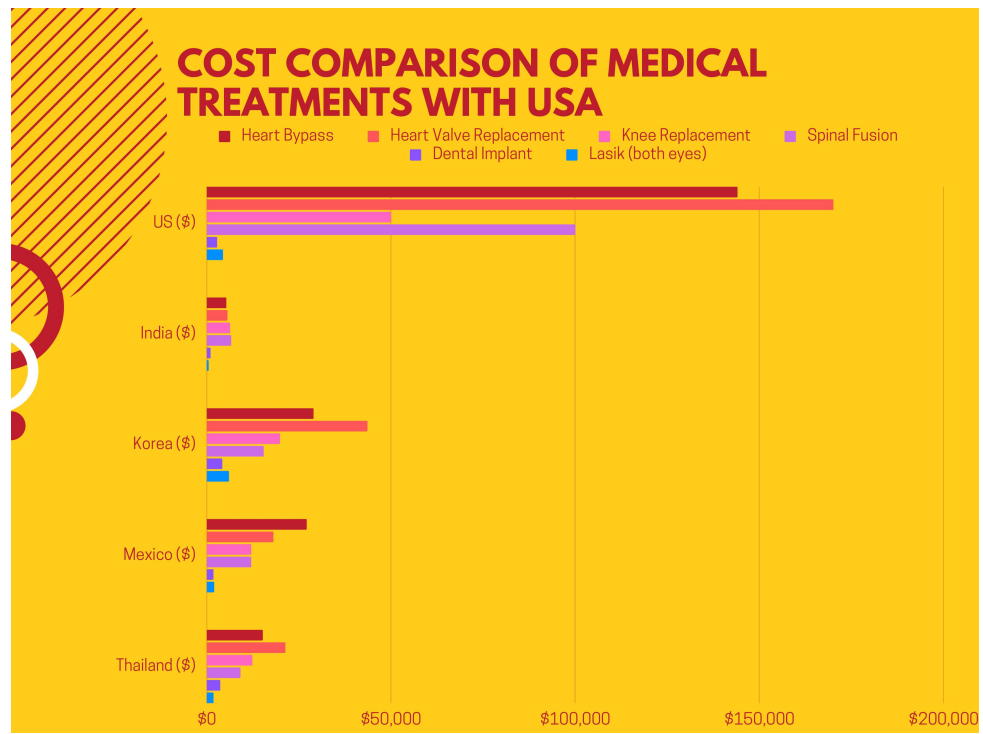
Computed pattern for average viewers on each edition

**Domanda 14**

Risposta non data

Punteggio max.:

0,25



**Question**

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

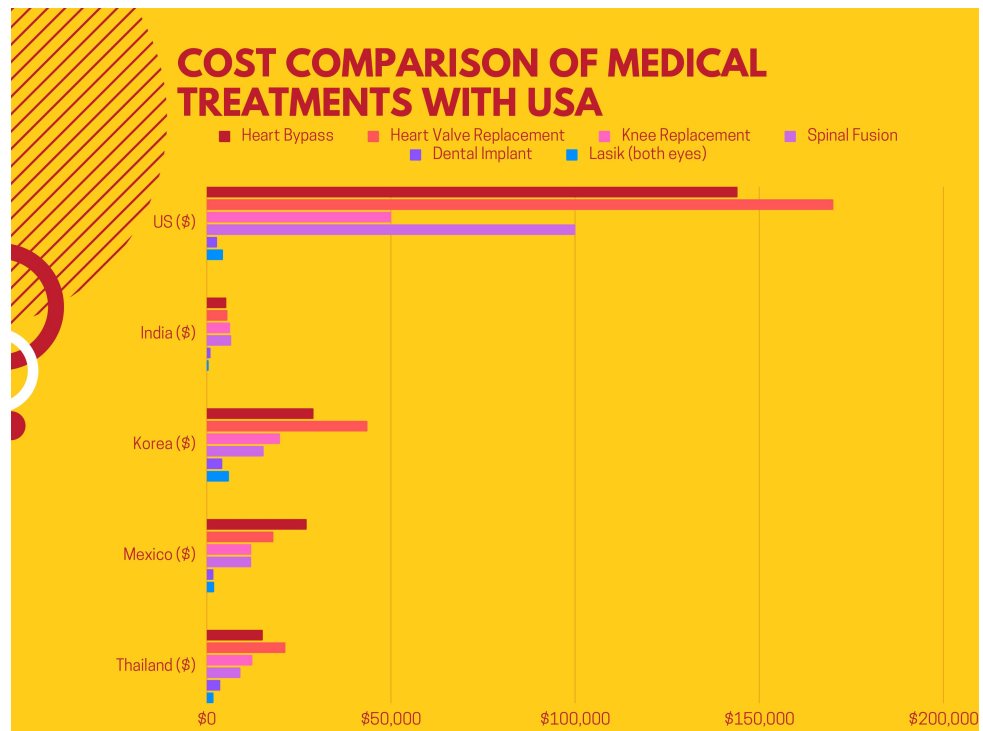
---

**Domanda 15**

Risposta non data

Punteggio max.:

1,25



**Data**

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

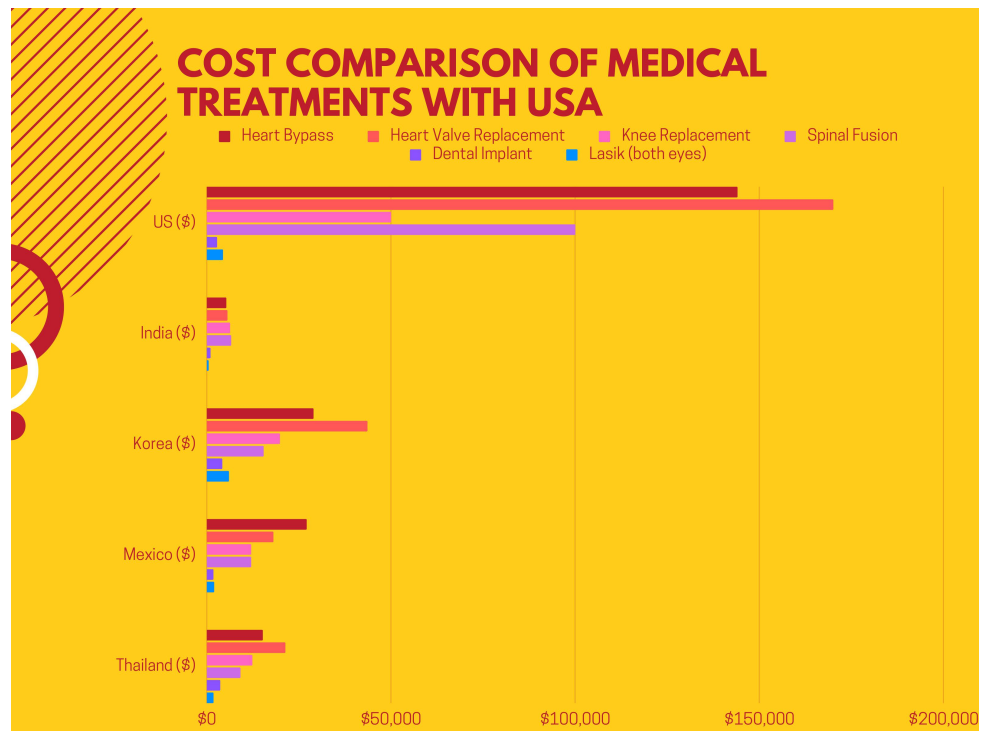
---

**Domanda 16**

Risposta non data

Punteggio max.:

0,75



**Visual Proportionality**

Are the values encoded in a uniformly proportional way?

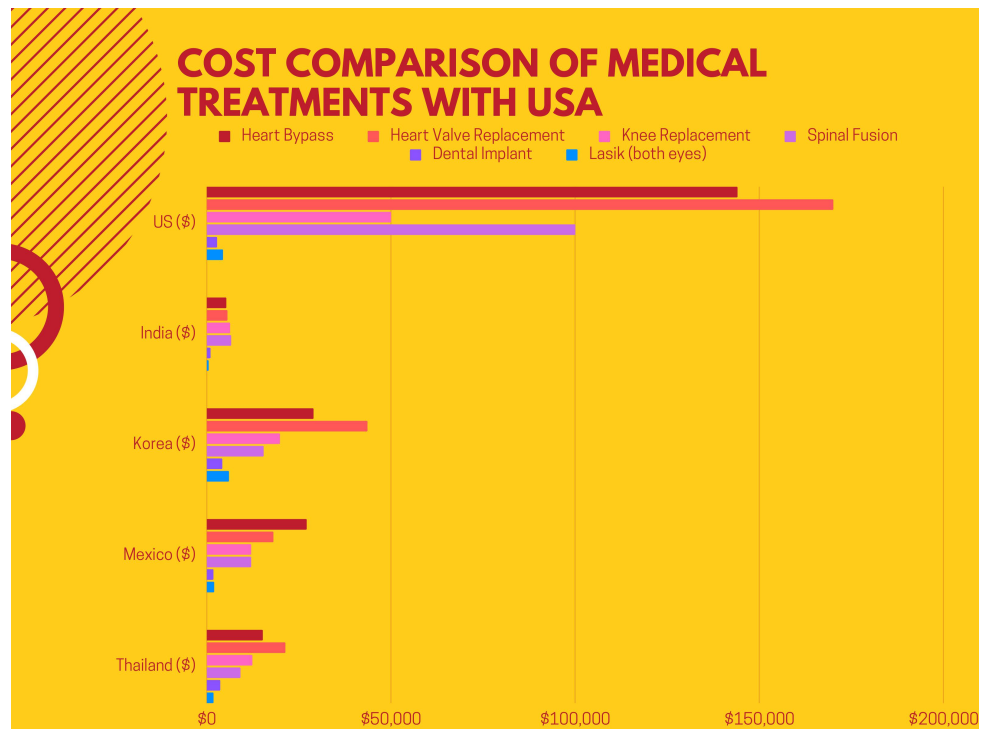
---

**Domanda 17**

Risposta non data

Punteggio max.:

0,75



**Visual Utility**

All the elements in the graph convey useful information?

---

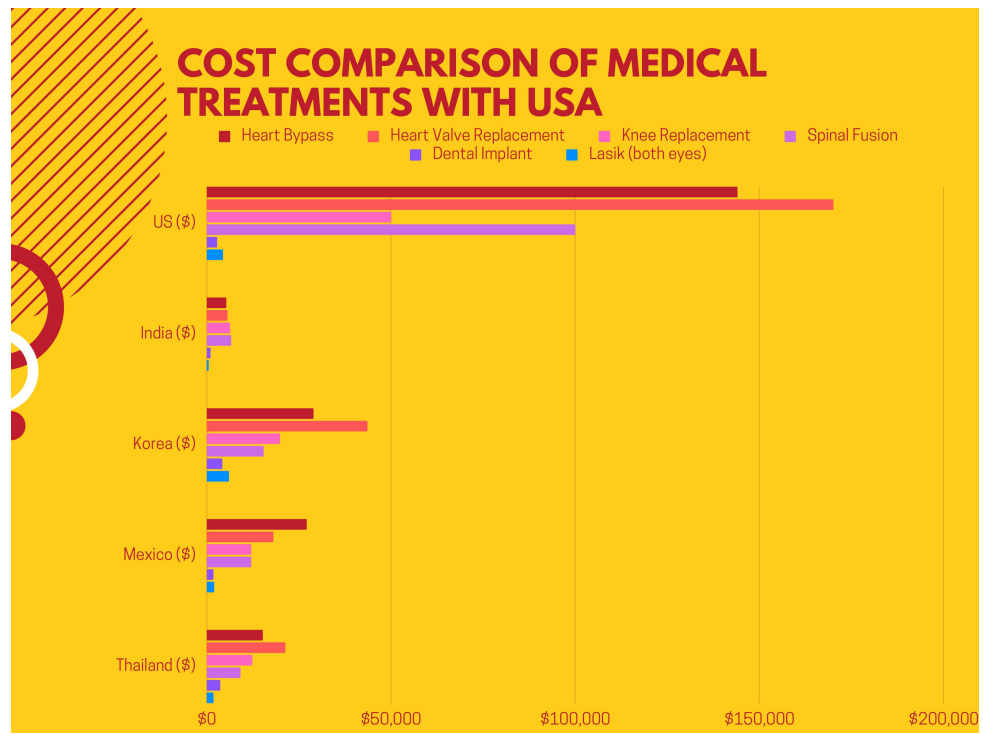


**Domanda 18**

Risposta non data

Punteggio max.:

0,50



**Visual Clarity**

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

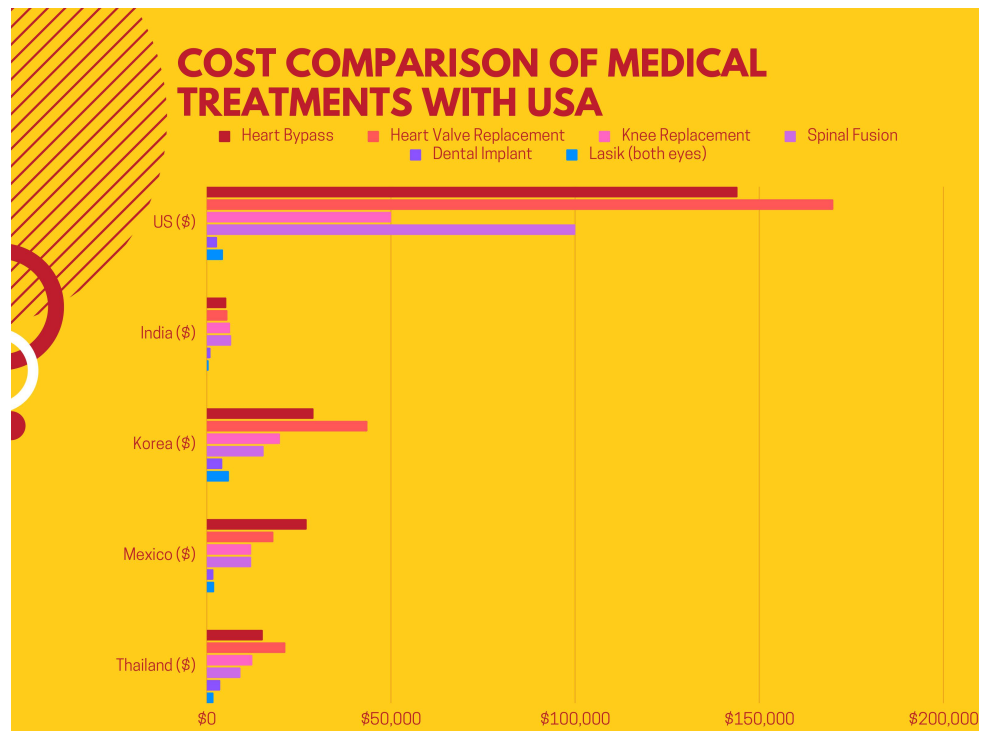
---

**Domanda 19**

Risposta non data

Punteggio max.:

0,25



**Design data**

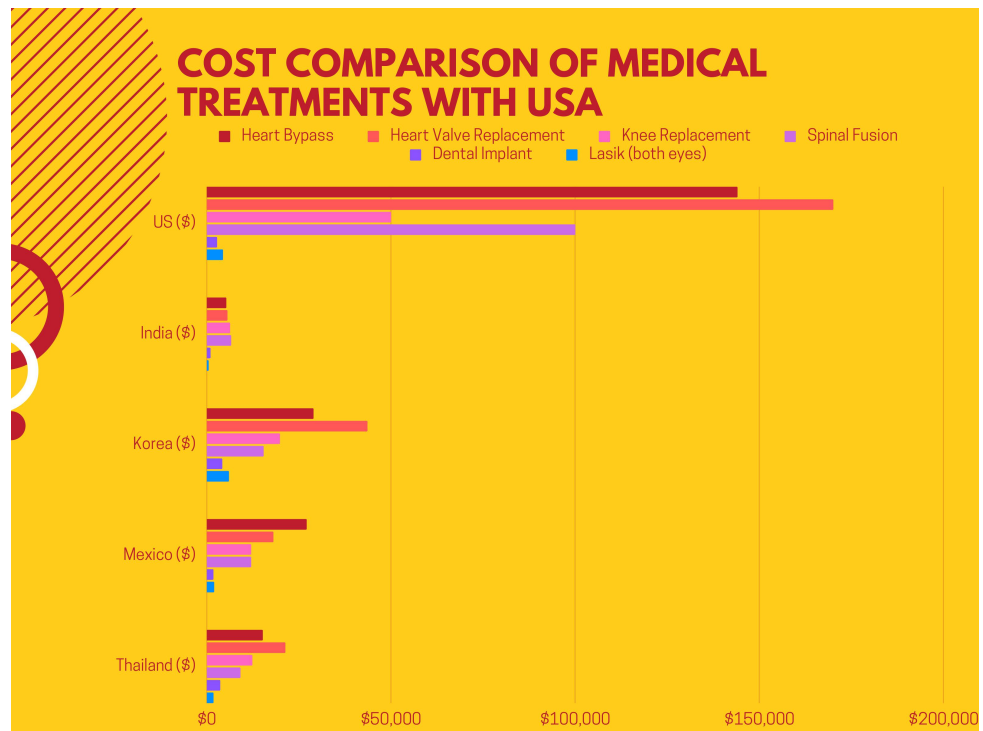
Design the visualization based on the following data structure (to be completed).

---

**Domanda 20**

Risposta non data

Punteggio max.: 1,25



**Design schema & Sketch**

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

---

**Domanda 21**

Risposta non data

Non valutata

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

Anything written here will NOT be evaluated.

---