Graph Link:

https://www.datylon.com/blog/types-of-charts-graphs-examples-data-visualization

QUIZs:

Tile 1:

Question 1: Which of the following statements best describe a data-driven application?

Select one:

a. An application that has access to lots and lots of data!
b. Businesses that produce goods and services based primarily on the collection, analysis and synthesis of information.
c. An application that uses data to understand and create value.
d. Any system that uses a tool like a relational database.

Question 2: In the generic pipeline discussed in this module, finding errors in a dataset and correcting mistakes is part of which phase?

Select one:

a. Gathering
b. Processing
c. Analysing
d. Presenting
e. Preserving

Question 3: In the generic pipeline discussed in this module, creating a chart to explain your business proposal to your client is part of which phase?

Select one:

a. Gathering
b. Processing
c. Analysing
d. Presenting

e. Preserving
f.

Question 4: In the generic pipeline discussed in this module, exploratory visualisations are generally part of which phase?

Select one:

a. Gathering
b. Processing
c. Analysing
d. Presenting
e. Preserving

Question 5: Is this statement True or False? "By collecting more data, a firm has more scope to improve its products, which attracts more users, generating even more data, and so on."

Select one:

a. True
b. False

Tile 2:

Question 1: Complete the following sentence: "A survey of the eye colour of this class is an example of... "

Select one:

a.Nominal data.

b.Ordinal data.

c.Interval data.

d.Ratio data.

Question 2: Complete the following sentence: "How many cups of coffee (or tea!) you drink in a day is an example of... "

Select one:

a.Nominal data.

b.Ordinal data.

c.Interval data.

d.Ratio data.

Question 3Complete the following sentence: "A nominal variable... "

Select one:

a. has a natural order.

B. can be described by the average of all values.

c.is a category.

Question 4: Complete the following sentence: "A qualitative variable is... "

Select one:

a. sometimes measured on a continuous scale
b. a categorical variable
c. a numeric variable

Question 5: Complete the following sentence: "A rating (1 to 5) of how much you agree (1) or disagree (5) with the statement “Data is the new oil” is... "

Select one:

a. a categorical variable.
b. an ordinal variable.
c. an interval variable.

Question 6: Complete the following sentence: "Biological species (taxonomic classification) is a type of... "

Select one:

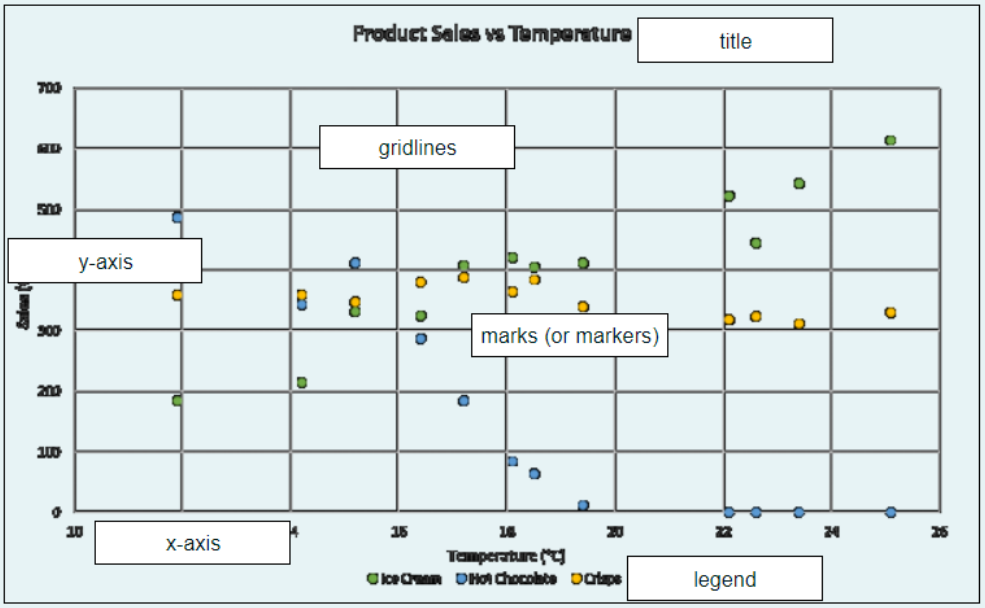
a. Nominal data.
b. Ordinal data.
c. Interval data.
d. Ratio data.

Question 7: Complete the following sentence: "The temperature in Dublin over the last seven days is an example of... "

Select one:

a. Nominal data.
b. Ordinal data.
c. Interval data.
d. Ratio data.

Correctly label each of the marked elements in the graph



Which mark is used to encode the data in the graph?

Question 2

Select one:

a.

e. Point

b.

Line

c.

Area

d.

Form

Feedback

Your answer is correct.

Question text

Select all attributes used to encode data in the graph.

Question 3

Select one or more:

a.

Quantitative: position x & y

b.

Qualitative: position x

c.

Qualitative: colour (hue)

d.

Quantitative: colour (saturation)

e.

Quantitative: area

Feedback

Your answer is correct.

How much data is being encoded for each marker in this graph?

Question 4

Select one:

a.

2: product sales & temperature

b.

3: product sales, product type & temperature

c.

4: location, product sales, product type & temperature

d.

2: product sales and product type

Feedback

Your answer is incorrect.

Question text

What type of graph is this?

Question 5

Select one:

a.

Bar chart

b.

Line chart

c.

Scatterplot

d.

Pie chart

Q.1 You are asked to plan a data analytics project to *analyse student feedback to DCU in relation to online teaching in 2020 and 2021*. Using the Generic Data Analytics Pipeline discussed in CA682, assign each of the following activities to one of the 5 main categories: Gathering, Processing, Analysing, Presenting and Preserving.

Liaising with DCU Registry to get datasets from the student registration and results systems. → Gathering,
Removing incorrect entries from the student datasets. → Processing,
Converting student words into sentiment ratings and correlating with the field of study. → Analysing,
Documenting the data formats used in the study and saving all of the created datasets. → Preserving,
Creating a document to share with senior university management summarising the findings. → Presenting,
Anonymising student comments that include identifying details. → Processing,
Conducting student surveys to answer the key questions about their experience. → Gathering,
Calculating the average satisfaction levels based on the sentiment ratings. → Analysing

According to the three classical definitions of big data, which of the following datasets is *most likely* to be classified as "big data"?

Question 2

Answer

a.

Sales records from the DCU merchandise store.

b.

The "Titanic" dataset showing passenger details from the final voyage of the ship.

c.

Records from Spotify of the tracks listened to by each user (est. 232M users).

For data that records the "Type of pet (e.g., cat, dog, bird, fish)" choose *all* of the following descriptions that can apply. Marks will be deducted for including wrong choices.

Qualitative

Quantitative

Discrete

Continuous

Nominal

Ordinal

Interval

Ratio

For data that records the "Number of pets currently owned" choose *all* of the following descriptions that can apply. Marks will be deducted for including wrong choices.

Qualitative

Quantitative

Discrete

Continuous
Nominal
Ordinal
Interval

Ratio

For data that records the "Weight of pets (in grams)" choose *all* of the following descriptions that can apply. Marks will be deducted for including wrong choices.

Qualitative

Quantitative

Discrete

Continuous

Nominal
Ordinal
Interval

Ratio

For data that records the "Happiness of pet owners (self-rated from 1 to 5)" choose *all* of the following descriptions that can apply. Marks will be deducted for including wrong choices.

Qualitative

Quantitative
Discrete
Continuous
Nominal

Ordinal

Interval
Ratio

Which of the following is *not* a valid description of metadata?

Question 7

Answer

a.

Metadata is created by humans and is often incorrect.

b.

Metadata is information on the organisation of the data, data domains and the relationship between them.

c.

Metadata is data about data.

d.

Metadata is an inferior form of cataloguing.

Which of the following statements are correct in relation to open data?

Question 8

Answer

a.

Open data can be used commercially.

b.

Open data is allowed to contain personal information.

c.

Open data is only provided by governments.

d.

Open data may help make governments and corporations more transparent.

Which of the following statements are correct in relation to data cleaning and data quality metrics?

Question 9

Answer

a.

A good measure of data quality is accuracy and completeness.

b.

Data quality metrics can be used for contracts for service delivery.

c.

It is possible to perfectly and absolutely measure quality of a dataset to compare performance.

d.

Many data quality metrics (accuracy, completeness) are unmeasurable

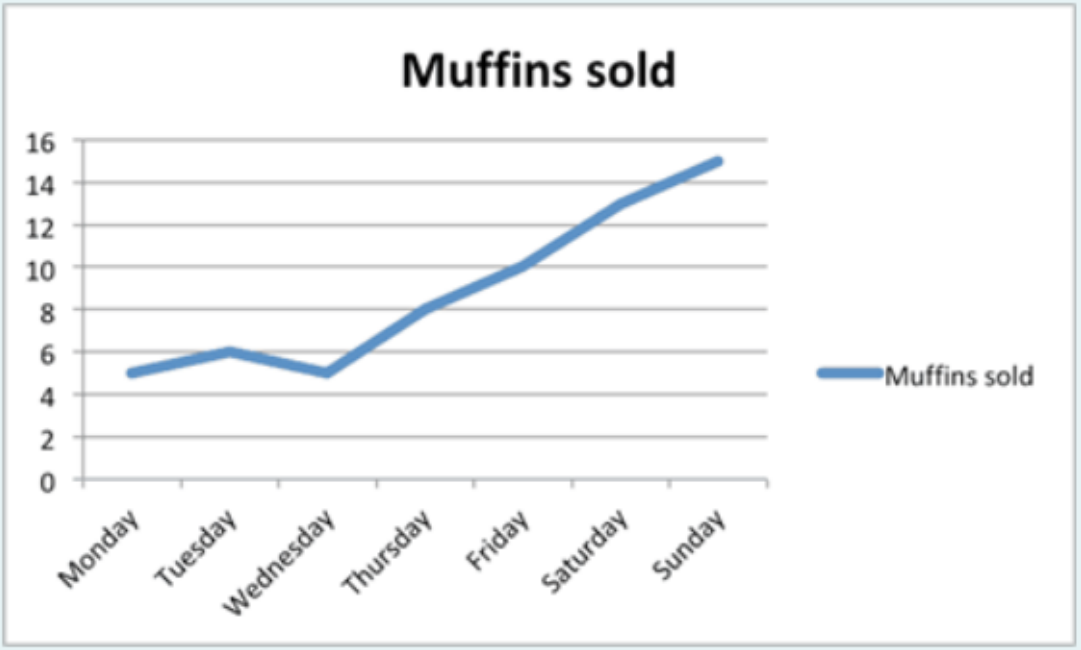
Match the error with the *most likely* phase of the generic data analytics pipeline where it was introduced

The correct answer is: Data delivery issues such as transmission problems that may result in loss of network connectivity, buffer overflows or corruption. → Gathering,

Time synchronization errors resulting in missing values. → Processing,

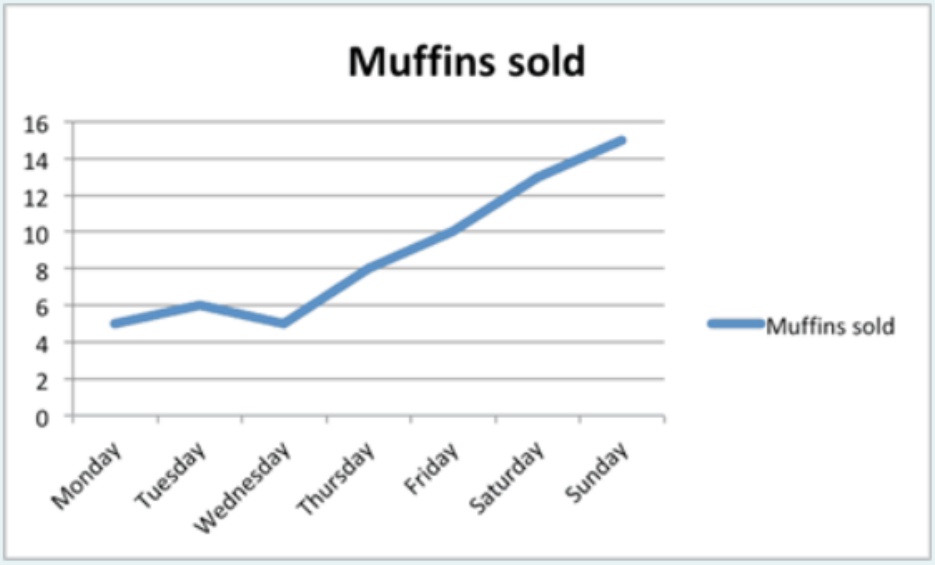
No documentation provided on format for missing values (“”, Nan, -999) → Preserving,

Unnecessary precision of generated numerical data. → Analysing



line chart

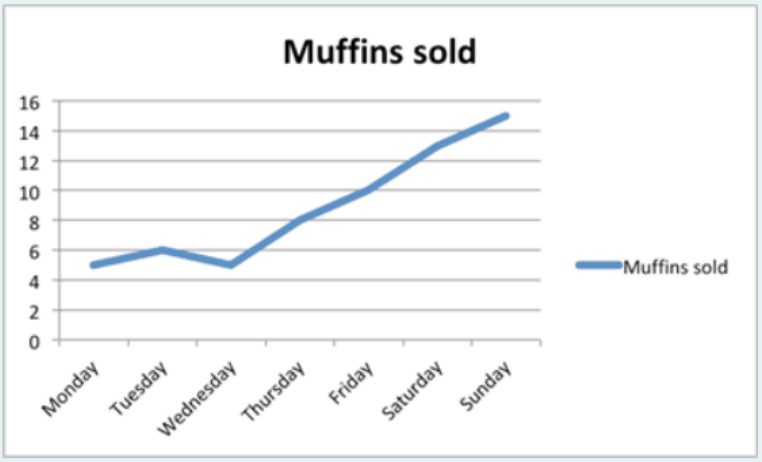
In the graph below, how many muffins were sold on Thursday?



- 2
- 4
- 6
- 8
- 10 ✖
- 12

Answer is 8

Identify all of the marks and attributes used to encode data in the graph below.



- [ ] point
- [ ] line
- [x] area ✖
- [ ] position
- [ ] colour (hue)
- [ ] colour (saturation)
- [ ] slope
- [ ] size (length)
- [ ] size (area)
- [ ] quantity

The correct answers are: line, position, slope

Match the scenario with the *most appropriate* choice of chart to visualise the given data.

The correct answer is: The most popular method of travel to DCU during 2019. → Bar chart,

Show the improvement in sales (total profit in €) over each of the past 5 years for your product compared to your competitors. → Line chart,

Distribution of grades in CA682 over the past 5 years. → Box plot,

Understand the relationship between maximum daily temperature (^{0}C) and average daily personal water consumption (Litres) in Ireland. → Scatterplot