

- Week2 Quiz Solution

Consider the definition given for data science, is the boundary between data science, data engineering and data analysis fixed?

Select one: True or False

Answer: False

What does the code below return (from slide 40 Week 2 lecture)?

```
>>> df.loc [(df['Sex'] == 'female') & (df['Survived'] == 1)]
```

Answer: Data for all female passengers that survived

What is a dataframe?

- a. An array
- b. A structure that stores tabular data
- c. A theory about data
- d. A list

Answer: structure that stores tabular data

What is .ipynb?

- a. Intelligent Python Nota Bene
- b. Typo, it should be 'pinyin'
- c. Interactive Python NoteBook
- d. An illegal file extension

Answer: Interactive Python NoteBook

- **Week 3 Quiz answers**

- 1) Give an example of a categorical-ordinal data.

Answer) Student grade (HD, D, ..., F), etc..

- 2) Calculate the Mean, Median, and Mode of the following set of values respectively.

1,2,2,2,3

Answer) the mean is 2, the median is 2, the mode is 2

- 3) Name a measure of spread introduced as a "Descriptive statistic" in week 3 lecture

Answer) range or standard deviation

- 4) According to Week 3 lecture slides, Histograms are a special type of bar charts

Answer) TRUE

Week 4 Quiz Answers

Q1) Open data is a private data that is machine readable.

Select one: True/False

Answer: **False**

Q2) Name a popular data/information API.

Example APIs: Google maps API, Twitter API, Facebook API.

Refer in below links for more examples

<https://rapidapi.com/blog/most-popular-api/>

<https://www.programmableweb.com/news/10-most-popular-products-apis/brief/2021/06/27>

Q3) When we have missing data in a dataset, how should we deal with it?

- a. Replace with a special “unknown” value
- b. Replace with an average value
- c. Remove the row or column in which missing data resides in
- d. All of the above options can be correct

Answer: **All of the above options can be correct**

- **Week 5 quiz answers**

Q) In general, the more training data we have, the better the test performance is

A) True

Q) According to lecture 5: "Error" measures the distance between the ... and the actual value.

A) Prediction

Q) Predictive models are learnt from

A) Training data

- Week 7 quiz answers

1. Please fill in the blank in the following sentence:

Clustering is machine learning task.

Answer: An unsupervised

2. Explain "sensitivity" as one of the classification metrics

Answer: When the actual value is positive, how often is the prediction correct?

$TP/(TP+FN)$

Example: Fraudulent transaction detector: Optimise sensitivity

3. Explain "specificity" as one of the classification metrics.

Answer: When the actual value is negative, how often is the prediction correct?

$TN/(TN+FP)$

Example: Spam filter: Optimise precision or specificity

- Week09

1. Unix shell commands like “less” and “grep”:

- a) are used to fit regression tree models
- b) can be used to manipulate large data files easily
- c) are poorly documented
- d) are examples of technology that is too old to be useful to a modern data scientists

Answer) b

2. Give an example of Descriptive metadata.

Answer) e.g. title, author of a book

3. Give an example of Structural metadata.

Answer) e.g. chapters in a book, elements in XML, containers in MPEG

4. Give an example of Administrative metadata.

Answer) e.g. version number, archiving date, Digital Rights Management (DRM)

Week 10 Quiz

1.) Give an example of a data parallel task.

Processing a large document collection to extract specific terms.

A classification task in a large dataset.

Summary statistics calculations in a large data set.

2.) Which one of the following is suitable for real-time data processing?

a. Hadoop

b. Spark

Answer: b.

3. Use SQL database when data is unstructured and unchanging.

Select one:

TRUE

FALSE

Answer: FALSE

Week 12 Quiz

Q1. What compliance and audit are?

Compliance refers to an organisations' adherence to policies, procedures, and regulations, particularly in the context of information confidentiality, data security and privacy.

Audit refers to the process of validating that an organisation is compliant with applicable regulations, policies, and procedures.

Q2. What is Confidentiality?

Confidentiality focuses on information privacy and relates to how information about individuals are treated and shared.