

## Week 4 Quiz

This quiz **does not** contribute towards your overall course marks. It is a checkpoint and also serves as a sample of the Multiple-Choice-Question (MCQ) section of the final exams. The final exams will have MCQ questions and short answers. The sample of questions for the short answers will be released at the end of the teaching weeks.

**No time limit**

**No due date**

Q1 [1pt] Choose the best definition for data science:

- Data Science is what a Data Scientist does
- Data Science and Data Analysis are the same
- Data Science is the technology of handling and extracting value from data that is always inter-disciplinary in nature.
- Data Science is the process of deriving insights from data, using scientific methods, processes, and system tools.

Q2 [1pt] Which of the following are Types of Data Scientist as mentioned in the article "Analyzing the Analyzers"?

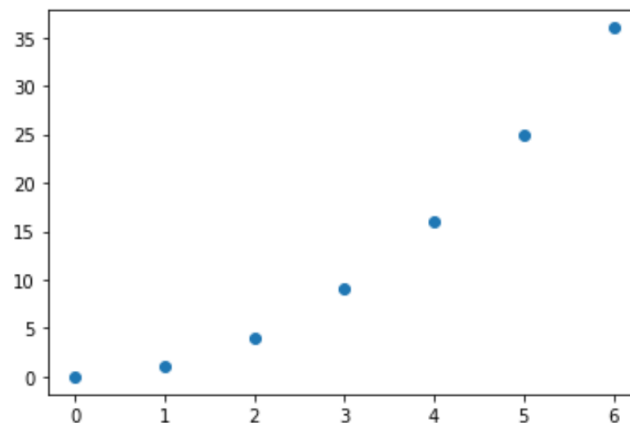
- Data Researcher
- Data Analyst
- Data Businessperson
- Data Officer

Q3 [1pt] What are the 5 number summary in statistics?

- Minimum, 1st Quartile, Median, 3rd Quartile, Maximum
- Minimum, 1st Quartile, Mean, 3rd Quartile, Maximum
- Mean, 1st Quartile, Median, 3rd Quartile, Mode
- Mean, Median, Mode, Minimum, Maximum

Q4 [1pt] Which of the following will produce a plot that looks like below:

*(If you are wondering about whether there will be programming questions, we do not expect you to provide code but be able to at least read or fill in the blanks. For this question, you may have to complete Week 5 labs first. We will repeat this in the sample exam questions at the end of the teaching week for reinforcement on what you need to know.)*



For reference, the code prior to the plot (code) is

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
df = pd.DataFrame({'X' : [0,1,2,3,4,5,6], 'Y' : [0,1,4,9,16,25,36]})
```

- `plt.boxplot(df['Y'])`
- `plt.plot(df['Y'])`
- `plt.scatter(df['X'], df['Y'])`
- `plt.hist(df['Y'])`

### Feedback

It requires data from both axes and is a scatter plot. It is also possible to use `plt.plot()` but the answer is using only the data from the Y-Axis.

Q5 [1pt] Which of these tasks might a data scientist typically perform?

- Criticise the lack of tools to deliver the project
- Develop and present a project idea
- Conduct industrial espionage to obtain project ideas.
- Financially sponsor the data science project

## Feedback

This is the same as pitching project ideas.

Q6 [1pt] Which option is the Mean, Median and Mode of the following set of values respectively? 1,2,2,3,4,7,9

- 4,2,3
- 4,3,3
- 4,3,2
- 5,3,2