

Introduction to Data Science Course by IBM- Answers

By Jaswanth Reddy Vulchi

Module 1 – Defining Data Science Answers

Q1- In the report by the McKinsey Global Institute, by 2018, it is projected that there will be a shortage of people with deep analytical skills in the United States. What is the size of this shortage?

- ◆ **140 000 – 190 000 people**
- ◆ 120 000
- ◆ 20 000 – 50 000 people
- ◆ 800 000 – 900 000 people
- ◆ 3 – 6 million people

Q2- How is Walmart reported to have addressed its analytical needs?

- ◆ **Crowdsourcing**
- ◆ Code sharing
- ◆ Social media
- ◆ Outsourcing
- ◆ None of the options is correct

Q3- In the reading, the New York Times reported the base salary for data scientists as:

- ◆ \$150 000
- ◆ \$85 000 + Bonus
- ◆ **\$112 000**
- ◆ \$16 per hour
- ◆ \$100 000

Module 2 – What do data science people do? Answers

Q1- In the reading, what was the real added value of the research?

- ◆ **Quantifying the magnitude of relationships**
- ◆ Analyzing consumer behavior
- ◆ Proximity to transport and infrastructure resulted in higher housing prices
- ◆ Shopping centers had a nonlinear impact on housing prices
- ◆ all else being equal’ is a powerful assumption

Q2- In the reading, what is an example of a question that can be put to a regression analysis?

- ◆ Do homes with brick exterior sell in rural areas?
- ◆ **What is the impact of lot size on housing price?**
- ◆ What are typical land taxes in a house sale?
- ◆ How much does a finished basement cost?
- ◆ How much should a house near a park cost?

Q3- Who developed the statistical technique known as Regression?

- ◆ Andrew Gelman
- ◆ **Sir Frances Galton**
- ◆ Anindya Ghose
- ◆ Saeed Aghabozorgi
- ◆ Dhanurjay “DJ” Patil

Module 3 – Data Science in Business Answers

Q1- In the reading, what is the ultimate purpose of analytics:

- ♦To evangelize data science
- ♦To facilitate meetings between sales and marketing
- ♦**To communicate findings to the concerned**
- ♦To build models
- ♦To generate reports

Q2-In the reading, the report successfully did the job of:

- ♦**Using data and analytics to generate the likely economic scenarios**
- ♦Calculating projections for the economy
- ♦Convincing the leadership team to act on an initiative
- ♦Using PowerPoint to deliver a message
- ♦Summarizing pages and pages of research

Q3- In this reading, what is the role of the data scientist?

- ♦Email the stakeholders about the analysis
- ♦Manage a team of analysts to create a model
- ♦Develop the strategy to fix the problems in the findings
- ♦**Use the insights to build the narrative to communicate the findings**
- ♦Use the data to tell the story the CEO wants to tell

Module 4 – Use Cases for Data Science Answers

Q1- An introductory section is always helpful in:

- ♦**Setting up the problem for the reader**
- ♦Presenting the statistical calculations
- ♦Summarizing the text
- ♦Introducing the research methods
- ♦Advertising the product

Q2- The results section is where you present:

- ♦**The empirical findings**
- ♦R Squared
- ♦The conclusion
- ♦The contributors
- ♦The methods used

Q3- In the reading, what is an example of housekeeping?

- ◆ Adding slide numbers
- ◆ **Including a list of references**
- ◆ Adding headings to charts
- ◆ Adding pictures to graphs
- ◆ Saving the report as a PDF

Module 5 – Data Science People Answers

Q1- In the reading, how does the author define 'data science'?

- ◆ Data science is way of understanding things, of understanding the world
- ◆ Data science is a physical science like physics or chemistry
- ◆ Data science is some data and more science
- ◆ **Data science is what data scientists do**
- ◆ Data science is the art of uncovering the hidden secrets in data

Q2- In the reading, what is admirable about Dr. Patil's definition of a 'data scientist'?

- ◆ His definition limits data science to activities involving machine learning
- ◆ His definition is only for people who program in Python
- ◆ His definition excludes statistics
- ◆ His definition is about weaving strong narratives into analytics
- ◆ **His definition is inclusive of individuals from various academic backgrounds and training**

Q3- In the reading, what characteristics are said to be exhibited by "The best" data scientists?

- ◆ Ask good questions, really curious people, engineers
- ◆ Really curious, ask good questions, at least 10 years of experience
- ◆ **Thinkers, ask good questions, O.K. dealing with unstructured situations**
- ◆ Thinkers, really curious, PhDs
- ◆ Really curious people, engineers, statisticians

Final Exam Answers: Introduction to Data Science Answers

Q1- In the reading, the output of a data mining exercise largely depends on:

- ◆ The engineer
- ◆ The programming language used
- ◆ **The quality of the data**
- ◆ The scope of the project
- ◆ The data scientist

Q2- In the reading, what are some of the steps down the data mine?

- Establish goals, store data, mine data, present data
- **Establish goals, select data, pre-process data, transform data**
- Establish goals, team meeting, select data, transform data
- Establish goals, mine data, evaluate data mining results, create database
- Establish goals, select data, pre-process data, present data

Q3- What should you do when data are missing in a systematic way?

- Extrapolate data
- Use Python to generate values
- Determine the average of the values around the missing data
- **Determine the impact of missing data on the results**
- Determine who was managing the database

Q4- What is an example of a data reduction algorithm?

- Conjoint Analysis
- Programmatization
- A/B Testing
- **Principal Component Analysis**
- Prior Variable Analysis

Q5- What should be a prime concern for storing data?

- **Data safety and privacy**
- Hiring the right database manager
- The size of the files
- The physical location of the servers
- Hadoop clusters

Q6- What is a good starting point for data mining?

- **Data Visualization**
- Writing a data dictionary
- Non-parametric methods
- Creating a relational database
- Machine learning

Q7- When evaluating mining results, data mining and evaluating becomes:

- A transformative process
- An intuitive process
- A data driven process
- A strategic process
- **An iterative process**

Q8- When establishing data mining goals, the accuracy expected from the results also influences the:

- The timelines for the project
- The scope of the project
- The costs**
- The presentation
- Data scientist

Q9- When processing data, what factor can lead to errors in data?

- Synchronizing the database
- Changing services
- providers Renaming
- variables **Human error**
- Overfitting

Q10- Formal evaluation could include testing the predictive capabilities of the models on observed data to see how effective and efficient the algorithms have been in reproducing data.” This is known as:

- Prototyping
 - Overfitting
 - In-sample forecast**
 - False positive
 - Reverse engineering
-
-
-