Fatima, Sayeda

05/05/2022

MODULE 1: WHAT IS DATA SCIENCE?

PEER GRADED ASSIGNMENT-FINAL ASSIGNMENT

# Overview

In this Assignment, you will demonstrate your understanding of the videos and the readings by answering open-ended questions,

1. defining data science
2. data scientist, and
3. describing the different sections comprising a final deliverable of a data science project.

Please note that this assignment is worth 10% of your final grade.

**Table of Contents**

[Overview 0](https://pgone-my.sharepoint.com/personal/fatima_s_pg_com/Documents/1-DATA%20SCIENCE/COURSEA/MODULE1-What%20is%20Data%20Science-FINAL%20ASSIGNMENT.docx#_Toc102624707)

[Review Criteria 2](#_Toc102624708)

[1. Defining Data Science [Problem 1: 3 points] 3](#_Toc102624709)

[2. Defining Data Scientist [Problem 2: 2 points] 3](#_Toc102624710)

[3. Describing the different sections comprising a final deliverable of a data science project [Problem 3: 5 points] 3](#_Toc102624711)

[References 5](#_Toc102624712)

# Review Criteria

This assignment will be graded by your peers who are also completing this course during the same session.

There are a total of **10 points** possible for this final project.

Below is the breakdown of possible points:

1. **Defining Data Science [Problem 1: 3 points]**
2. **Defining Data Scientist [Problem 2: 2 points]**
3. **Describing the different sections comprising a final deliverable of a data science project [Problem 3: 5 points]**

## **Defining Data Science [Problem 1: 3 points]**

The rise of data science is marked by inevitability and so, its progress has extreme urgency and significance. From a sociological perspective, data science is a set of knowledge systems comprising of new concepts, methods, models, technologies, tools and applications which seeks to understand new challenges, opportunities, thought patterns and approaches in an age of big data (Li and Cheng, 2012). To Xu, et al. (2021), data science is the fundamental approach and procedure of utilizing modelling, analysing, computing and learning to facilitate data conversion from information to knowledge and subsequently, to decision-making for the ultimate realization of data value chain. In summary, while data science may be described as exploration of data which may enhance understandings for making strategic decisions and choices, it may simply be regarded as whatever any data scientist does in order to gain this insight. To put it even more simply, data science is not only about using a particular tool or approach like machine learning to reveal patterns or predictions but instead, it is the art of using any available tool or approach which can best resolve a query in hand. Whether the data is big or small is irrelevant, for as long as its exploration can lead to meaningful insights, it falls within the field of data science. For example, my KNIME model investigating merely a few thousand lines of Written-off invoices at P&G helped identify risky customers and bad debt patterns among different legal entities and business areas which, in turn, identified key areas of improvement within recovery and credit granting policies. So, it follows that intelligent exploration and analysis of even small data may fall within the remit of data science as long as it provides meaningful insights for business strategic making.

## **Defining Data Scientist [Problem 2: 2 points]**

The data scientist profile was recognized to describe individuals working on data applications investigating voluminous and varied data sets in order to find solutions to business queries and create immediate and massive organizational impacts (Costa and Santos, 2017). To Patil (2011), a data scientist is really an evolution of data analyst role exhibiting robust business acumen and abilities to effectively communicate analytical findings in the form of sound solutions carrying maximum value to organizations. Hence, one may describe a data scientist as someone who explores data, whether big or small, and arrives at answers which can then be communicated to appropriate stakeholders in an easy narrative to facilitate strategic decision making. To elaborate a bit more, a data scientist is someone who, driven by his/her natural inquisitiveness, may competently use any suitable tool at his/her disposable to exhaustively analyse data, make sounds judgements and argumentative conclusions which can then be effectively communicated in a simple, easy to understand format without complicated jargon.

## **Describing the different sections comprising a final deliverable of a data science project [Problem 3: 5 points]**

The eventual aim of a data science exploration and analysis is effective communication of key findings to targeted audience in the form of a narrative so that it may be utilized to influence policies and strategies. This narrative or final deliverable often takes the form of an essays or reports, normally, around 1,000 to 7,000 words long and may contain illustrations like graphs, tables and plots, or it may even be a more comprehensive document offering the reader with critical reviews, in-depth data analysis and exhaustive explanations. Whether it is brief or comprehensive, it is advisable that this final deliverable should observe a prescribed format containing some key elements.

To begin with, the document must start with an informative cover page containing a title which will help its readers in understanding the topic of preceding analysis. It should also have names of authors, their contacts and affiliations which will help readers communicate with them in future should they require any explanation. Then, a very important element to insert in the cover page is the full publication date and details of the institutional publisher (if applicable) to facilitate future citations and tracing.

For longer documents say, those exceeding five pages, it is crucial that the cover page be followed by a table of contents (ToC) which will serve like a guiding map to readers and will furnish visual representation of actual narrative. The ToC should, therefore, contain lists of tables and figures along with key headings to furnish a glimpse of what lies ahead in the document.

What follows next is an executive summary of not more than three paragraphs, (which may be longer for larger documents), which serves to furnish readers with an overview and clarify their expectations. The subsequent introductory section will further convey document’s purpose and basic information which should then followed by a cited literature review discoursing relevant research on the topic, existing gaps, research questions and hypothesis.

While research methods, data collection approaches, tools and variables are best described in the methodology section, outcomes can be carefully summarized in the results section upon which key arguments are constructed in the subsequent discussion section. It is here where significance of research endeavours in terms of closing identified research gaps will be highlighted. The document can then conclude its findings in conclusion section, which is normally followed by general housekeeping segments like referencing, acknowledgments and any appendices.

# References

Costa, C. & Santos, M. Y., 2017. The data scientist profile and its representativeness in the European e-Competence framework and the skills framework for the information age. International Journal of Information Management, 37(6), pp. Pages 726-734.

Li, G.-J. & Cheng, X., 2012. Big Data Research: A Major Strategic Field of Future Science and Technology and Economic and Social Development-Research Status and Scientific Thinking of Big Data.. Journal of Chinese Academy of Sciences, 27(6), pp. 647-657.

Patil, D., 2011. Building data science teams − O’Reilly radar. [Online]   
Available at: http://radar.oreilly.com/2011/09/building-data-science-teams.html  
[Accessed 01 05 2022].

Xu, Z., Tang, N., Xu, C. & Cheng, X., 2021. Data science: connotation, methods, technologies, and development. Data Science and Management, 1(1), pp. 32-37.