Based on the videos and the reading material, how would you define a data scientist and data science? (3 points)

A data scientist is someone who finds solutions to problems by analyzing big or small data using appropriate tools and then tells stories to communicate her findings to the relevant stakeholders. They are curious and analytical thinkers who use a variety of math skills not limited to mathematics, statistics, probability, and activities involving machine learning to solve a problem. They are someone who can easily process millions of rows of data in a couple of hours and distrust (statistical) models in order to unlock the insights data and tell a fantastic story via the data. They apply different available methods and algorithms to draw insights and conclusions from various kinds of data. After applying data science methodologies, they are effective communicators and story tellers who can present their findings often to present new findings or confirm what was initially suspected.

Data science is something that data scientists do. Data science is a way of extracting insight from large volumes of disparate data and is the art of uncovering the hidden secrets in data. Data science involves drawing patterns from seemingly random structured and unstructured types of data to gain insights and supports data-driven decision-making.

As discussed in the videos and the reading material, data science can be applied to problems across different industries. Give a brief explanation describing what industry you are passionate about and would like to pursue a data science career in? (2 points)

I am passionate about pursuing a data science career in the medical industry. I regard this as an opportunity to make a profound impact on people’s lives. By harnessing the power of data analytics and innovative methodologies, I can potentially improve patient outcomes and drive medical advancements. As a data scientist, I have then the chance to unravel intricate patterns in medical data, uncover novel insights, and develop innovative solutions to complex challenges.

Based on the videos and the reading material, what are the **ten** main components of a report that would be delivered at the end of a data science project? (5 points)

1. **Cover page** containing the title of the report, names of authors, their affiliations, and contacts, the name of the institutional publisher (if any), and the date of publication.
2. **Table of contents (ToC)** with main headings, list of tables and figures to map what is contained in the document.
3. **Abstract/executive summary** to give a concise explanation of the key elements and findings in the report to give the readers a clear understanding of the study’s purpose, methods, results, and conclusions without requiring them to read the entire report.
4. **Introduction** that outlines the context and importance of the subject matter and that explains the problem for the reader who might be new to the topic and who might need to be gently introduced to the subject matter before being immersed in intricate details.
5. **Literature Review** of available relevant research on the subject matter which should include citations of influential authors and data sources. A literature review can be used to highlight gaps in the existing knowledge, which the conducted analysis will try to fill. This is where the research questions and hypothesis are formally introduced.
6. **Methodology** section that introduces the research methods, research design, data sources and data collection techniques employed for the analysis. If new data is collected, then an explanation of the data collection exercise is given in some detail.
7. **Results** section is where the empirical findings and outcomes are presented. Descriptive statistics, illustrative graphics and a significant statistical analysis of the hypothesis tested is included in this section.
8. **Discussion** section is where the main arguments built on the results are presented. The power/significance, implications, and limitations of the results in relation to the research objectives are elucidated to the readers. The understanding of the results is demonstrated and critically analyzed to interpret the data. It helps the readers to contextualize and evaluate the findings.
9. **Conclusion** section which generalizes findings, summarizes the implication of the research conducted and identifies possible future development in research and applications that could result from conducted research.
10. **References**, **Acknowledgement** and **Appendices** (if needed).