**Data Science Assignment (IBM Data Science Coursera Certification)**

**Course 1: What is Data Science?**

**Define Data Science and Data Scientist:**

Data Science: Data science is the scientific study of data to gain knowledge. This field combines multiple disciplines to extract knowledge from both structured data and unstructured data for the purpose of making informed decisions and predictions. It involves multiple processes such as storing that data in right place, manipulating, exploring, and analyzing the data in order to answer questions to make strategic data-driven decisions or recommendations.

In today’s data-driven world, a vast quantity of data from many varied sources is getting generated every single moment. These sources, but not limited to, are log files, email, social media, sales data, patient information files, sports performance data, sensor data, security cameras, and many more besides. Data science can help organizations understand their environments, analyze existing issues, and reveal previously hidden opportunities, and predict the future outcomes based on the previously stored data and implementing algorithms on the data to extract insights and build predictive models. Many organizations use data science to focus on a specific problem, and so it's essential to clarify the question that the organization wants answered. The fact is the need for data science is growing rapidly as the amount of data increases exponentially and companies depend more heavily on analytics to drive revenue and innovation.

A data scientist is an analytics professional who is responsible for collecting, storing, manipulating, analyzing and interpreting data to help drive decision-making in an organization. The data scientist role combines elements of several traditional and technical jobs, including mathematician, scientist, statistician, and computer programmer. It involves the use of advanced analytics techniques, such as machine learning and predictive modeling, along with the application of scientific principles.

As part of data science initiatives, data scientists often work with big data that can be defined as datasets that have greater variety and velocity to develop model and test hypotheses, make inferences, and analyze things such as customer and market trends, financial risks, cybersecurity threats, stock trades, equipment maintenance needs and medical conditions etc.

Data scientists must possess the following skills:

1. Working with Big data which is a large amount of data such as;

**Structured data**, which is typically organized in rows and columns and includes words and numbers such as names, dates, and credit card information.

**Unstructured data**, which is unorganized and includes text in document files, social media and mobile data, website content, images, audio, and videos.

1. Have a strong understanding of databases
2. Computer Programming
3. Mathematics, Calculus, Algebra, Statistics, and Probability
4. Domain Knowledge
5. Communications, public speaking, Report Writing, Storytelling, and data visualization.

**Second Question: Data Science can be applied to various problems across various industries. Which industries you are passionate about to pursue the data science for and why?**

Being a self-motivated individual throughout my entire life and someone with a keen eye for data-driven results, I thrive on turning complex datasets into meaningful and actionable insights that help stakeholders' informed decision-making.

My passion for data analysis manifests with each passing year. Being a Data Analyst for a couple of years, I always thrive on working with data as my curiosity always shows me a new path to invest my time in learning new tools and technologies that help me bring innovative ideas and strategies.   
  
As data science applies to many problems across many industries, I have always wanted to work with financial institutions in order to help achieve a common goal and I also want to work in health industry where our common objective lies on the same path to make the world a safer place and help in advancing healthcare facilities and processes that will help boost productivity in diagnosis and treatment and enhances the workflow of healthcare systems.

According to a study, the data generated by every human body is 2 terabytes per day. This data includes activities of the brain, stress level, heart rate, sugar level, and many more. In order to handle such a large amount of data, Data Science plays a vital role in providing hidden insights and innovative ideas based on patients’ health recorded data. With the help of the application of Data Science in healthcare, it has now become possible to detect the symptoms of a disease at a very early stage. Also, with the advent of various innovative tools and technologies, doctors are able to monitor patients’ conditions from remote locations.

**Data Science Report Structure:**

Before doing the analysis, it is very important to understand the business problems that need to be answered and accordingly, the structure of report. A brief report is more to the points and presents a summary of key findings while detailed report is more comprehensive and contains critical details about other relevant works, research methodology, data sources, and intermediate findings along with the main results.

The followings are main components of the data science project report;

1. Cover Page : It must contain 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
3. Abstract or Executive Summary
4. Introduction: always helpful in setting up the problem statement
5. Data source and Dataset Description
6. Methodology: Research Methods and Algorithms used
7. Data Analysis and Final Results
8. Discussion
9. Conclusion and Future Scope
10. References
11. Appendixes