IBM Data Science

There are 9 courses in this certificate training –

1. What is Data Science?
2. Open Source tools for Data Science
3. Data Science Methodology
4. Python for Data Science and AI
5. Databases and SQL for Data Science
6. Data Analysis with Python
7. Data Visualization with Python
8. Machine Learning with Python
9. Applied Data Science Capstone

1. What is Data Science –

It is the process of using data to understand different things, understand the world. For me is when you have a model or hypothesis of a problem, and you try to validate that hypothesis or model with your data. Data science is the art of uncovering the insights and trends that are hiding behind data. It's when you translate data into a story. So use storytelling to generate insight. And with these insights, you can make strategic choices for a company or an institution.
Data science is a field about processes and systems to extract data from various forms of whether it is unstructured or structured form. Data science is the study of data. Like biological sciences is a study of biology, physical sciences, it's the study of physical reactions. Data is real, data has real properties, and we need to study them if we're going to work on them. Data Science involves data and some signs. The definition or the name came up in the 80s and 90s when some professors were looking into the statistics curriculum, and they thought it would be better to call it data science.

But what is Data Science?
I'd see data science as one's attempt to work with data, to find answers to questions that they are exploring. In a nutshell, it's more about data than it is about science. If you have data, and you have curiosity, and you working with data, and you're manipulating it, you're exploring it, the very exercise of going through analyzing data, trying to get some answers from it is data science. Data science is relevant today because we have tons of data available. We used to worry about lack of data. Now we have a data deluge. In the past, we didn't have algorithms, now we have algorithms. In the past, the software was expensive, now it's open source and free. In the past, we couldn't store large amounts of data, now for a fraction of a cost, we can have gazillions of datasets for a very low cost.
So, the tools to work with data, the variability of data, and the ability to store and analyze data, it's all cheap, it's all available, it's all ubiquitous, it's here. There's never been a better time to be a data scientist.

I have also learned about regression analysis in this course.

2. Open Source tools for Data Science

In this course, I have learned about various open source tools for Data Science.

- Skill Network Labs
- Jupyter Notebooks
- Apache Zeppelin Notebooks
- Rstudio IDE
- IBM Watson studio

3. Data Science Methodology

In this course I have learned about the major steps involved in tackling a data science problem. - The major steps involved in practicing data science, from forming a concrete business or research problem, to collecting and analyzing data, to building a model, and understanding the feedback after model deployment. - How data scientists think!

4. Python for Data Science and AI

In this course I have learned about Python Basics like types, expressions, variables, string operations, lists, tuples, sets, dictionaries, Loops, objects and classes, file handling, pandas and numpy.

5. Databases and SQL for Data Science

In this course, I have learned about relational database concepts that helps to apply foundational knowledge of the SQL language, performing SQL access in a data science environment. The emphasis in this course is on hands-on and practical learning. I have also created some database instances in the cloud. I have done series of hands-on labs, practice building and running SQL queries in this lab. I have also learned how we can access databases from Jupyter notebooks using SQL and Python.

6. Data Analysis with Python

In this course I have learned about Importing Datasets, Cleaning the Data , Data frame manipulation, Summarizing the Data. It includes following parts: Data Analysis libraries, use of

Pandas, Numpy and Scipy libraries to work with a sample dataset. I have used this library to load, manipulate, analyze, and visualize cool datasets.

7. Data Visualization with Python

This course was all about several data visualization libraries in Python like Matplotlib, Seaborn, and Folium and how we can tell a compelling story by visualizing the data and findings from the data.

8. Machine Learning with Python

In this course I have learned about some of machine learning topics like supervised and unsupervised learning, classification, clustering and some Python libraries like Sci-kit learn and Scipy.

9. Applied Data Science Capstone

In this course I have learned about FourSquare API (It is a restful API to retrieve the data about venues in different neighborhoods around the world and I have applied this learnings to complete my Capstone Project