Layth Zubaidi

BDA 594

Professor Ming

Graduate Paper – Literature Review in their Specialty Areas with Big Data Applications

1. Find out two websites which focus on your own special areas and big data analytics and write a 500-word paragraph to introduce each web site with their URLs.

1st Website: <https://towardsdatascience.com/analytics-essentials-for-data-science-25da994f4bed>

* So to preface I come from a background of computer science in my undergraduate program. Going into the field of data science I felt I had a slight edge in understanding the concepts and tools that the field of data science likes to utilize. As such when learning Python I found many useful tools that this programming language was able to utilize in order to create code that can take in and analyze datasets and make some actual sense with them, and even use them for any additional insight. This website is extremely useful for giving insight of some of the most important concepts to learn with Python in order to utilize data analytics. The website has specifically mentioned on a focus on programming skills that are important to manipulating and analyzing data in an efficient way. These skills should prove useful in a multitude of situations in the data engineering field. Some of the concepts that this website discusses about are: Dataframes, Arrays, Visualizations, Descriptive statistics, Working with dates and time, and also Machine learning. Starting off they dive into a topic that I find to be important in fundamental in programming usage of data analytics, just as they mention how it is a core concept of data science analytics. To start off with a basic definition of them, they’re essentially tables with rows and columns to store data. The rows are seen as a record and the columns being attributes of said records. For a simple example that they used, is a dataframe of employees, where each row is a person and the columns hold information on them such as name, home address, salary, and job position. It is then explained how due to their nature they are quite important to understand their functionality as one can easily manipulate the data within as well as create insightful data visualization through them such as graphs and maps. It is also important to understand how to properly clean this data to make it more readable and avoid any issues that the tables can create when developing visualization through them. One should make sure to remove rows with missing values and erroneous values as well as reformatting the table to be more understandable.

1. Write an additional 1000-word essay about the impact of big data science on an area of your interest and identify the potential connections of big data science with your job or your major field.