Introduction to Data Analytics

Module – 1 Syllabus

- Data-Information
- characteristics of data
- data munging
- Scraping
- Sampling
- Cleaning
- importance of data analytics

Data Analysis

- Data analysis is a process of obtaining raw data and converting it into information useful for decision-making by users.

Data Analytics

- Data Analytics the science of examining raw data with the purpose of drawing conclusions about that information

Applications of Analytics

- In commercial industries, to enable organizations to make more-informed business decisions and by scientists

- By researchers, to verify or disprove scientific models, theories and hypotheses.

Analysis vs Analytics

- Data analytics is a broader term and includes data analysis as necessary subcomponent.

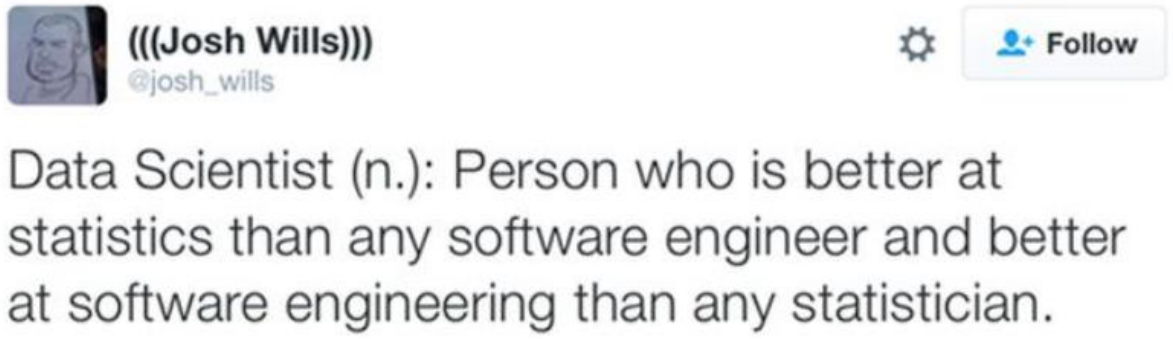
- Analytics defines the science behind the analysis.

Analysis vs Analytics

- The science means understanding the cognitive processes an analyst uses to understand problems and explore data in meaningful ways.

- Analytics also include data extract, transform, and load; specific tools, techniques, and methods; and how to successfully communicate results.

Data Science???



Data Science

- Dealing with unstructured and structured data, Data Science is a field that comprises of everything that related to data cleansing, preparation, and analysis.

- Involves in creation of new algorithms

Data vs Information

- Data are the facts or details from which information is derived.
- Individual pieces of data are rarely useful alone.
- For data to become information, data needs to be put into context.

Data vs Information

Comparison Chart

BASIS FOR COMPARISON	DATA	INFORMATION
Meaning	Data means raw facts gathered about someone or something, which is bare and random.	Facts, concerning a particular event or subject, which are refined by processing is called information.
What is it?	It is just text and numbers.	It is refined data.
Based on	Records and Observations	Analysis
Form	Unorganized	Organized
Useful	May or may not be useful.	Always
Specific	No	Yes
Dependency	Does not depend on information.	Without data, information cannot be processed.

Characteristics of data

The seven characteristics that define data quality are:

1. Accuracy and Precision
2. Legitimacy and Validity
3. Reliability and Consistency
4. Timeliness and Relevance
5. Completeness and Comprehensiveness
6. Availability and Accessibility
7. Granularity and Uniqueness

Characteristics of data

Accuracy and Precision: This characteristic refers to the exactness of the data.

Legitimacy and Validity: Requirements governing data set the boundaries of this characteristic.

Characteristics of data

Reliability and Consistency: Regardless of what source collected the data or where it resides, it cannot contradict a value residing in a different source or collected by a different system.

Timeliness and Relevance: Data collected too soon or too late could misrepresent a situation and drive inaccurate decisions.

Characteristics of data

Completeness and Comprehensiveness: Incomplete data is as dangerous as inaccurate data.

Availability and Accessibility: This presumes that the data exists and is available for access to be granted.

Characteristics of data

Completeness and Comprehensiveness: Incomplete data is as dangerous as inaccurate data.

Availability and Accessibility: This presumes that the data exists and is available for access to be granted.

Granularity and Uniqueness: The level of detail at which data is collected is important, because confusion and inaccurate decisions can otherwise occur.