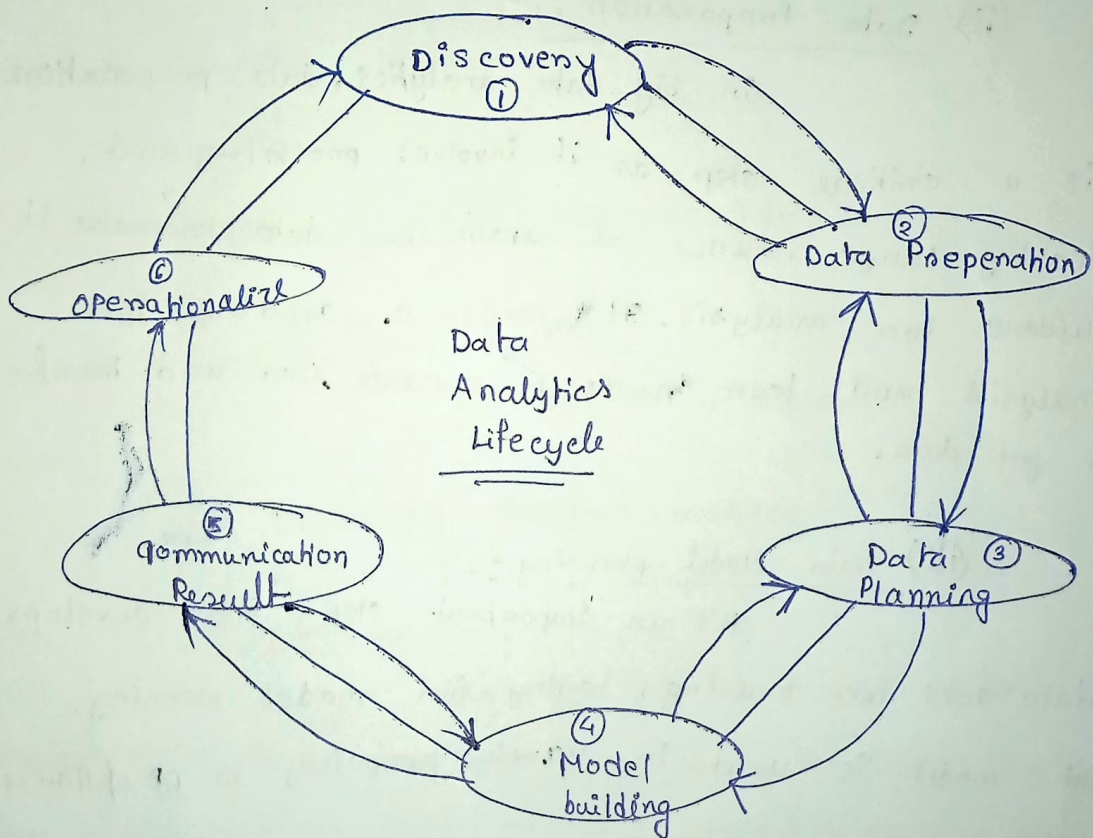


1) Data Analytics Lifecycle :-

It's the roadmap of how data is gathered, collected, processed, used and analyzed to achieve business goals. It's offers a systematic way to manage data for converting it into information that can be used to fulfill organized and project goals.

(b) Different state of Data Analytics Lifecycle :-



(c)

The phases of the data analytics lifecycle are also used in big data analytics projects with some modifications and additional considerations due to the large and complex nature of big data.

(i) Data Discovery :-

In big data analytics data is often collected from variety of source, including social media. Those data are unstructured. Some tools like Hadoop, MySQL are used to do this kind of work.

(ii) Data Preparation :-

On big data analytics, data preparation is a critical step, as it involves processing and cleaning large volume of unstructured data to make it suitable for analysis. It needs a sandbox for analysis and team needs to encode load and transform to get data.

(iii) Data Model planning :-

It's an important step. Team develops data sets for training, testing for model planning. This model is used in various purposes to optimize.

(iv) Model building :-

Predictive models are often built using machine learning algorithms. That are designed to work with large volume of data.

<v> Communication Results :-

Model accuracy is very important. Team needs to compare outcomes of modeling to predict for success and failure.

<vi> Operationalize :-

Model needs to continuously monitored and optimized to ensure their accuracy and later deploy in controlled way in the industry.

2) (a) Importance of Data visualization :-

<i> Easy to understand :-

It's easy to understand.

<ii> Improved decision-making.

<iii> Better agreement :-

Analyzing reports helps business stakeholders focus on the areas that require attention. The visual medium helps analysts understand the key points needed for their business.

<iv> Fasten Decision Making :-

Human process visual better than any tedious tabular forms or reports. If the data communication well, decision-makers can quickly

take action based on the new data insights, accelerating decision-making, and business growth simultaneously.

<v> Making sense of Complicated Data:-

It allows users to gain insight into their vast amounts of data. It benefits them to recognize new patterns and errors in the data. Making sense of these patterns helps the users pay attention to areas that indicate red flags on progress. This process, in turn drives the business ahead.

(b) Real time issues in data gathering and preparation:-

<i> Data volume and velocity:-

In realtime data gathering the volume and velocity of data can be very high, which can make it difficult to process and analyze the data in real-time.

<ii> Data quality:-

Real-time data may have quality issues such as missing values, incomplete data or incorrect data, which affect accuracy of data.

<iii> Data Integration:-

Real-time data may come from different sources and in different formats making it, challenging to integrate.

(iv) Data security :-

Real time data gathering may contain sensitive information, which requires careful handling to ensure data security and privacy.

(v) Time sensitivity :-

Real time data is time-sensitive which means that needs to be analyzed and acted upon quickly to be useful.

(c) "Data Cleaning" step used during data preprocessing :-

(i) Remove duplicate or irrelevant observation :-

In this process we remove all unwanted dataset, including duplicate dataset.

(ii) Fix Structural error :-

Structural error when you measure or transfer data and notice strange naming conventions, typos, or incorrect capitalization. These inconsistencies can cause mislabeled categories or classes.

(iii) Filter unwanted outliers :-

Outliers may affect the accuracy of a model for prediction so it needs to be removed by data cleaning we can do this.

(iv) Handle missing data :-

As no algorithm does not accept missing data, it needs to be fixed. We can drop that missing values or we can replace it by avg, median values of that columns or we can assign null values.

(v) Valid & QA :-

At the end of data cleaning we can able to answers those question -

- Does that data make sense?
- Does that data follow the appropriate rules for its field?
- Can you find trends in the data to help you from your next theory?

if can't get answers it may be dirty data and needs to be cleaned for preprocessing.