Data Analytics
An Introduction

Outcomes

Students would learn.

1. **Basic definition of Data, Information, and Data analytics**
2. **Different types of variables**
3. **Types of analytics**
4. **Analytics Life Cycle**

Basic Definition

- Data: Data is a set of values of qualitative or quantitative variables. It is information in raw or unorganized form. It may be a fact, figure, characters, symbols etc.
- Information: Meaningful or organised data is information.
- Analytics: Analytics is the discovery , interpretation, and communication of meaningful patterns or summery in data.
- Data Analytics (DA) is the process of examining data sets in order to draw conclusion about the information it contains.
- Analytics is not a tool or technology, rather it is the way of thinking and acting on data.

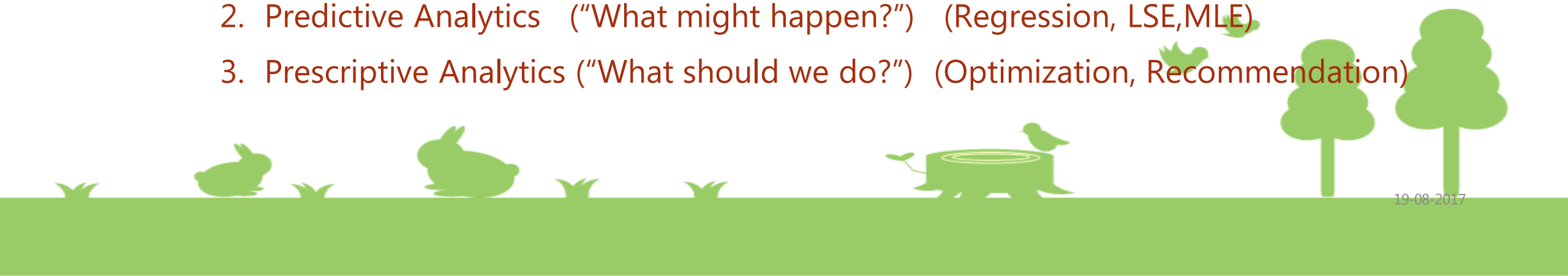
Data Analytics (Cont..)

- Examples

1. Business analytics
2. Risk ,,
3. Fraud ,,
4. Health ,,
5. Web ,,

- Types of analytics

1. Descriptive Analytics (“What has happened?”) (Data aggregation, summary, data mining)
2. Predictive Analytics (“What might happen?”) (Regression, LSE,MLE)
3. Prescriptive Analytics (“What should we do?”) (Optimization, Recommendation)



Variable Types

1. Numerical

Continuous (ex: height, weight, profit.)

Discrete (# items, population, count of students etc) 2 categorical

Nominal categorical (location, caste, gender)

Ordinal categorical (Grade, hotness, coldness etc.)

Answer the question

- 1. Which type of variable is “Street Number” ?
- 2. Which type of variable is “Phone Number” ?
- 3. Which type of variable is “Annual Income” ?

Analytics Life Cycle

1. Problem Identification
2. Hypothesis formulation
3. Data Collection
4. Data Exploration/preparation
5. Model Building
6. Model Validation and Evaluation

Analytics Life Cycle(Cont..)

1. Problem Identification

- The problem is a situation which is judged to be corrected or solved
- Problem can be identified through

1. Comparative/benchmarking studies
2. Performance Reporting
3. Asking some basic questions
    a) Who are affected by the problem?
    b) What will happen if problem is not solved?
    c) When and where does the problem occur?
    d) Why is the problem occurring
    e) How are the people currently handling the problem?

Analytics Life Cycle(Cont..)

2. Hypothesis formulation

1. Frame the questions which need to be answered.
2. Develop a comprehensive list of all possible issues related to the problem.
3. Reduce the list by eliminating duplicates and combining overlapping issues.
4. Using consensus building get down to a major issue list.

3. Data Collection

Data collection techniques are

1. Using data that is already collected by others
2. Systematically selecting and watching characteristics of people, objects, and events.
3. Oral questioning respondents either individually or as a group
4. Collecting data based on answers provided by the respondents in written format.

Analytics Life Cycle(Cont..)

4. Data Exploration

1. Importing data
2. Variable Identification
3. Data Cleaning
4. Summarizing data
5. Selecting subset of data

5. Model Building

- Building a Model is a very iterative process because there is no such thing as final and perfect solution.
- Many of the machine learning and statistical techniques are available in traditional technology platform

6. Model validation and Evaluation

- Like model building the process of validating model is also a iterative process.

- There are so many ways …
- **Confusion Matrix**.
- **Confidence Interval**.
- **ROC curve**
- **Chi Square**.
- **Root Mean Square Error**
- **Gain and Lift Chart**.