

---

---

---

---

---



# Topic 5

## Data types:

Structured data: Highly organized information in data base or can be search easily.

unstructured data: Opposite of structured data

## Data acquisition:

Direct approach: converted from real world data into digital data.

Indirect approach: can be found or bought.

Handling missing data: Global constant,

Ignoring, Imputing or Inference approach.

Filtering noisy data: Identify outlier and removing.

Munging or Wrangling: Involves reformatting data into usable format.

## Reduction

- Data Cube Aggregation.
- Dimensionality Reduction.

## Discretization

- Nominal
- ordinal
- Continuous

## Conversions

- Onehot Encoding
- Feature Construction

## Normalization

- min-max
- Z-Score
- Scaling

Above techniques do not need domain experts

but, Smoothing, Generalization, Feature Construction need domain experts

- info()

gives info about columns and null values.

Label encoder is for ordinal and categories  
is meaningful

One Hot encoding is for nominal  
and categories are not meaningful