**Ordine fasi progetto:**

**Data understanding**

* Print dei dati v
* Descrizione dei dati v
* Attributi e cambio attributi v
* Prima scrematura di data quality problem: • Wrong data • Duplicate data • Noise and outliers • Missing value
* Statistics: (Compare statistics with the expected behaviour • Exploit visualization techniques that can help in detecting • general patterns and trends • outliers and unusual patterns)
* **Scatter Matrix e scatterplot**
* Correlation Analysis

**Data preparetion**

- Aggregation

- Data Reduction: Sampling

- Dimensionality Reduction

- Feature subset selection

- Feature creation

- Discretization and Binarization

- Attribute Transformation