



Data Exploration

Stages of Data Exploration



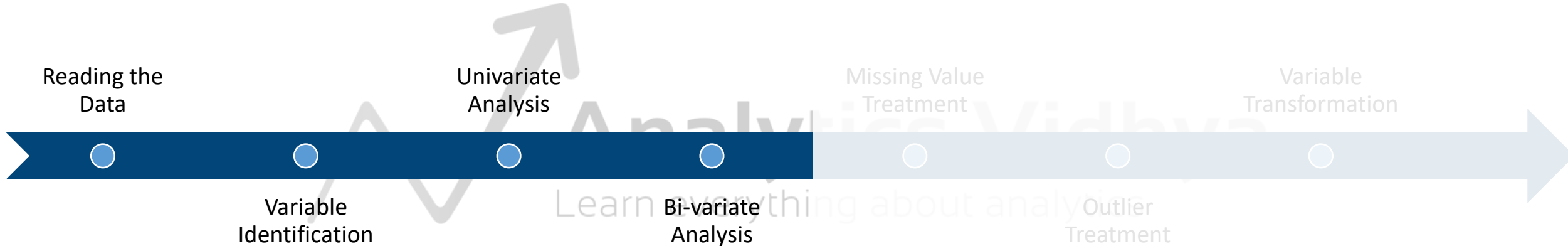
Stages of Data Exploration



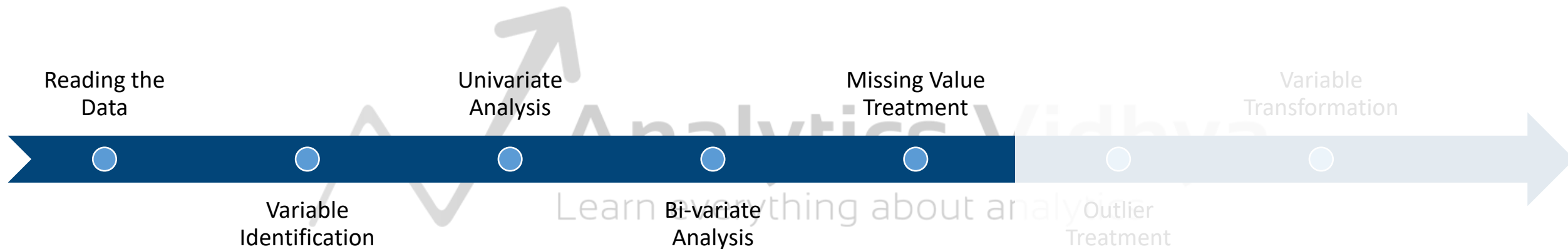
Stages of Data Exploration



Stages of Data Exploration



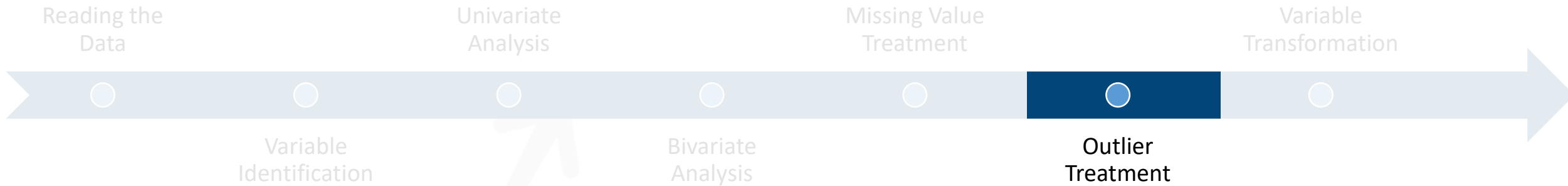
Stages of Data Exploration



Stages of Data Exploration



Stages of Data Exploration



Topics to be covered

1. Reasons for Outliers
2. Types of Outliers
3. Identifying Outliers
4. Different methods to deal with Outliers

Analytics Vidhya
Learn everything about analytics

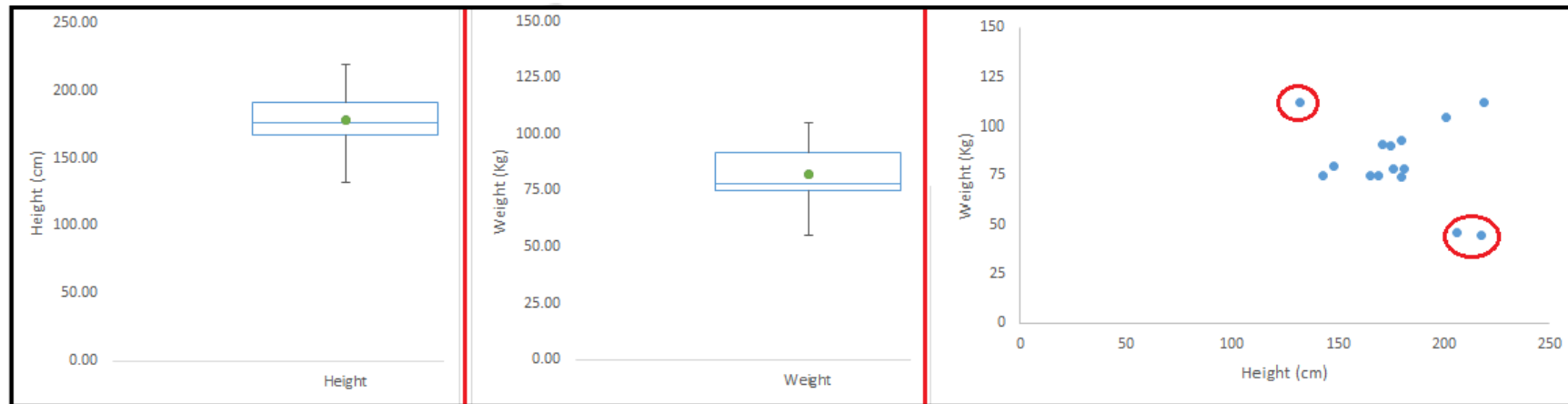
Outlier Treatment

Reasons for Outliers

1. **Data Entry Errors** : Entering salary of a professional as \$1000,000 instead of \$100,000
2. **Measurement Errors** : Measuring some particular data in km instead of m.
3. **Processing Errors**
4. **Change in the underlying population**

Outlier Treatment

Types of Outliers



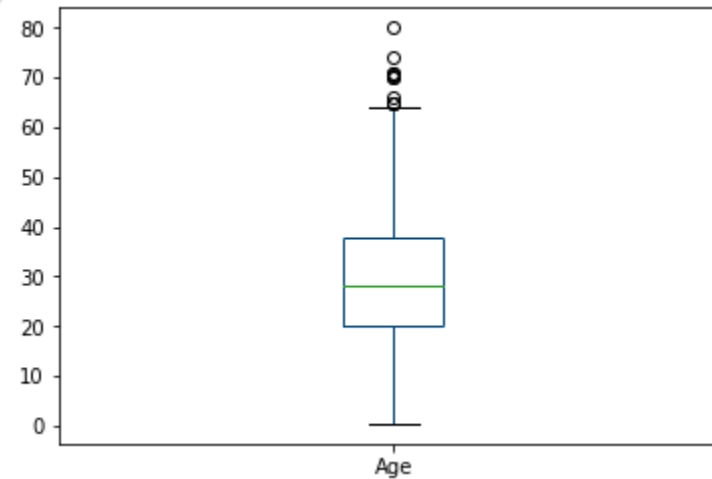
Univariate Outliers

Bivariate Outliers

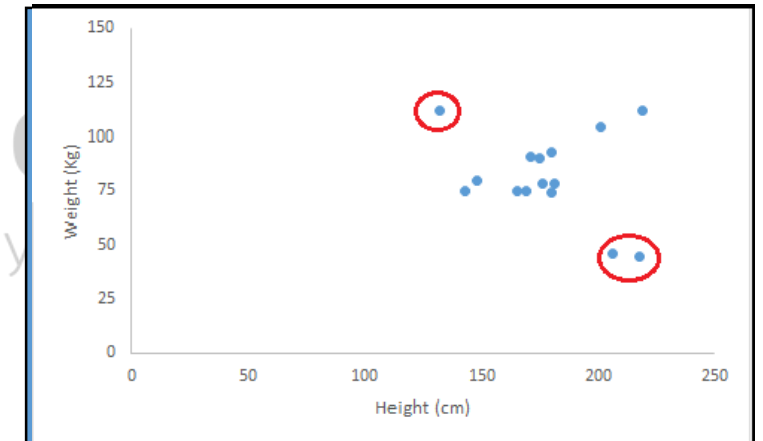
Outlier Treatment

Identifying Outliers

Graphical Methods



Boxplot



Scatter plot

Outlier Treatment

Identifying Outliers

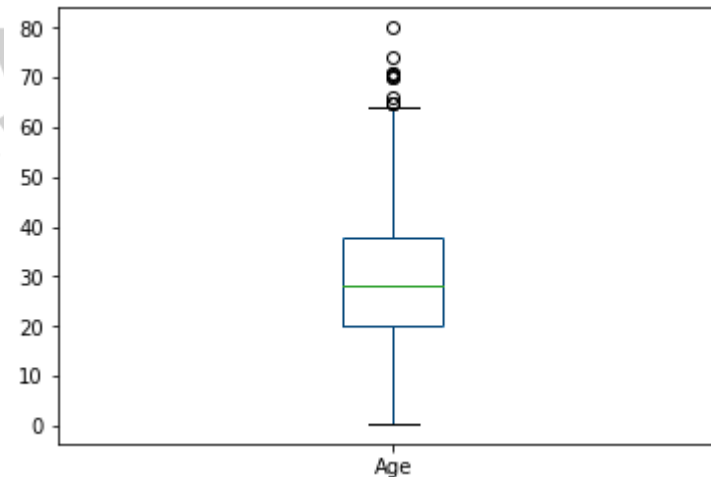
Formula Methods

$< Q1 - 1.5 * IQR$ or $> Q3 + 1.5 * IQR$

Where $IQR = Q3 - Q1$,

$Q3$ = Value at 3rd quartile

$Q1$ = Value at 2nd Quartile



Outlier Treatment

Treating Outliers

1. Deleting Observations
2. Transforming and Binning Values
3. Imputing Outliers like Missing Values
4. Treat them as separately

