Dealing with Outliers

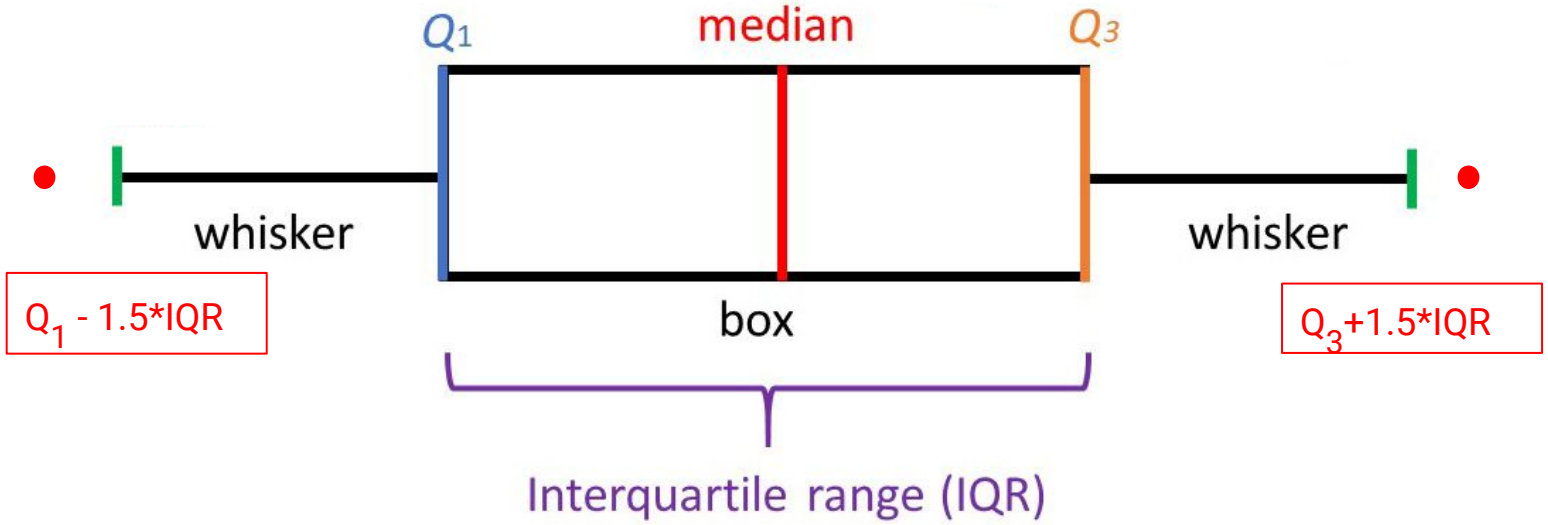
Outliers

Extreme(large or small) values relative to other observations

Reasons for Outliers Values

- Human Error
- Measurement Error
- Experimental Error
- Intentional Outlier
- Sampling Error
- Actual Outliers

Identifying Outliers



Analytics Vidhya
Learn everything about analytics

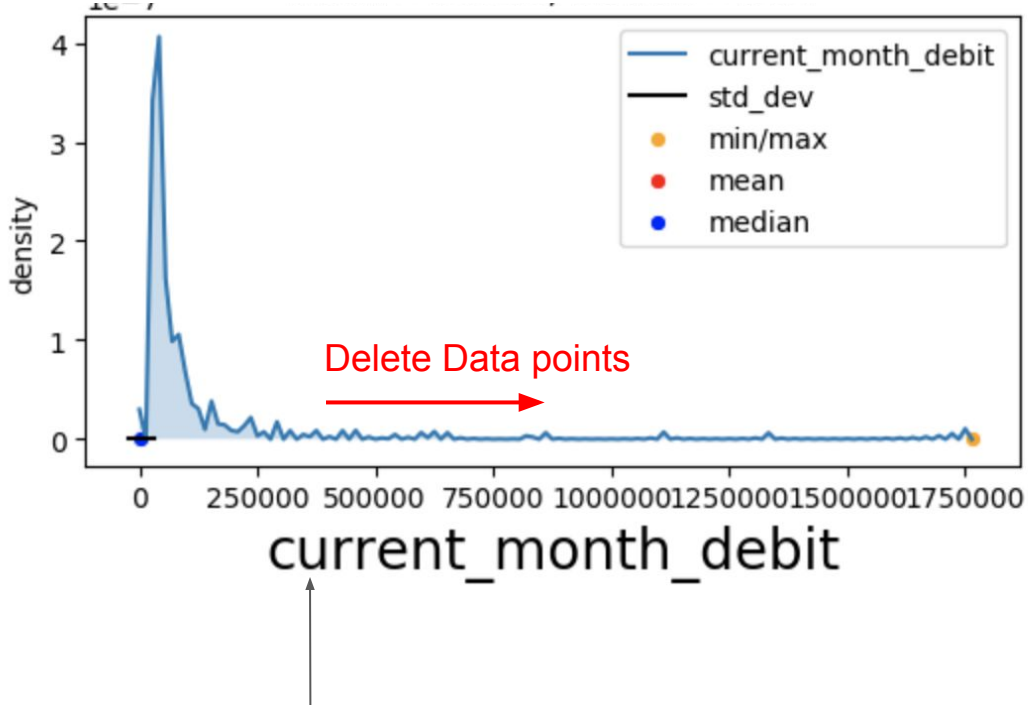
Dealing with Outliers

- Delete the data points
- Replace Outliers
  - Using central tendency
  - Relationship with other Variables
  - Replace with whisker values
  - Using an ML model
- Transform the values
  - Variable transformation
  - Binning

Dealing with Outliers

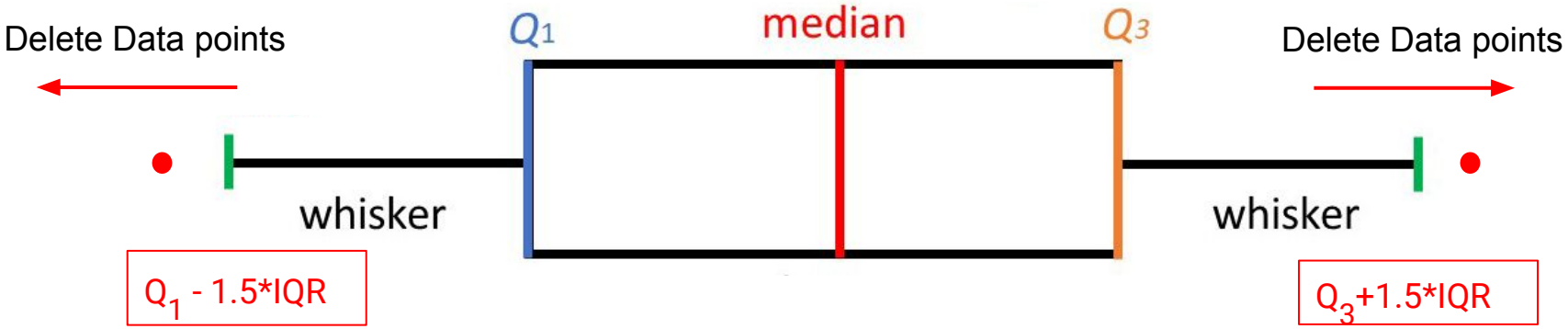
- Delete the data points
- Replace Outliers
  - Using central tendency
  - Replace with whisker values
  - Using an ML model
- Transform the values
  - Variable transformation
  - Binning

Dealing with Outliers



Analytics Vidhya
Learn everything about analytics

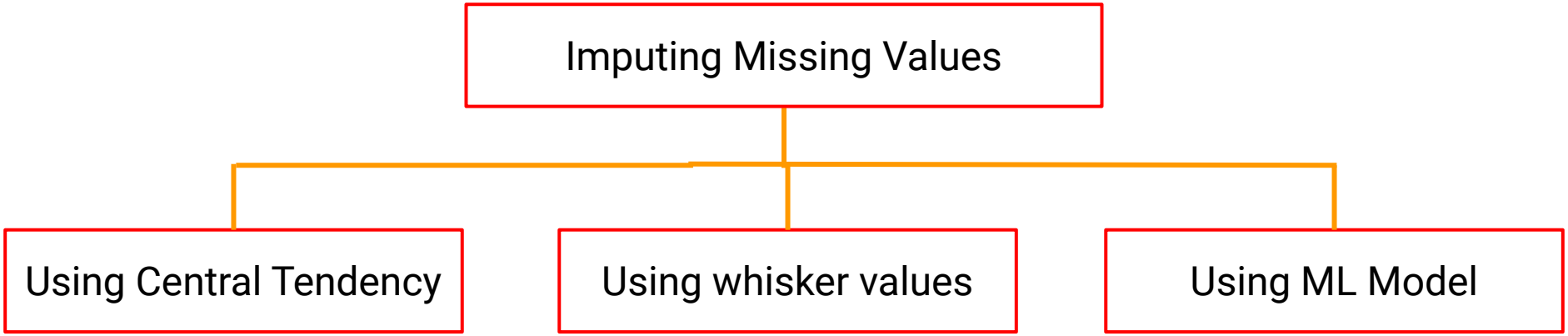
Dealing with Outliers



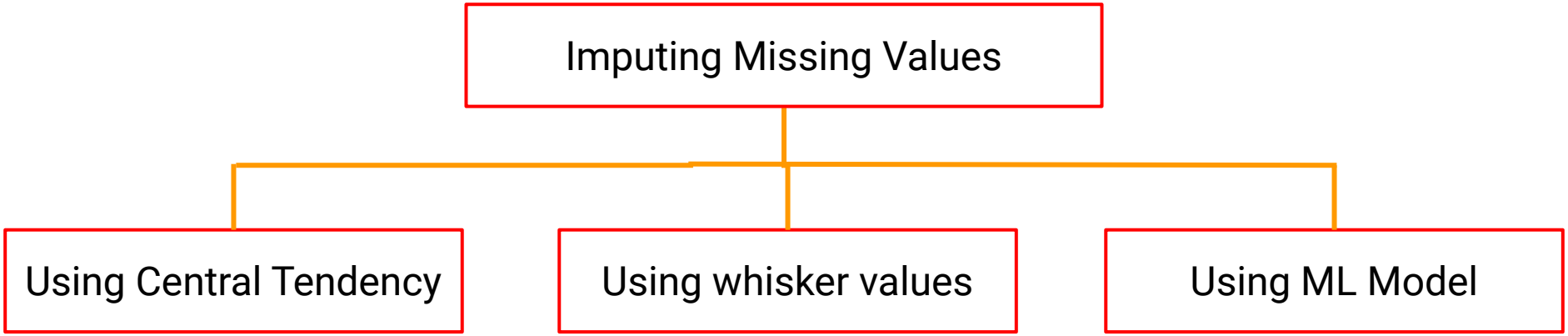
Dealing with Outliers

- Delete the data points
- Replace Outliers
  - Using central tendency
  - Relationship with other Variables
  - Replace with whisker value
  - Using an ML model
- Transform the values
  - Variable transformation
  - Binning

Dealing with Outliers



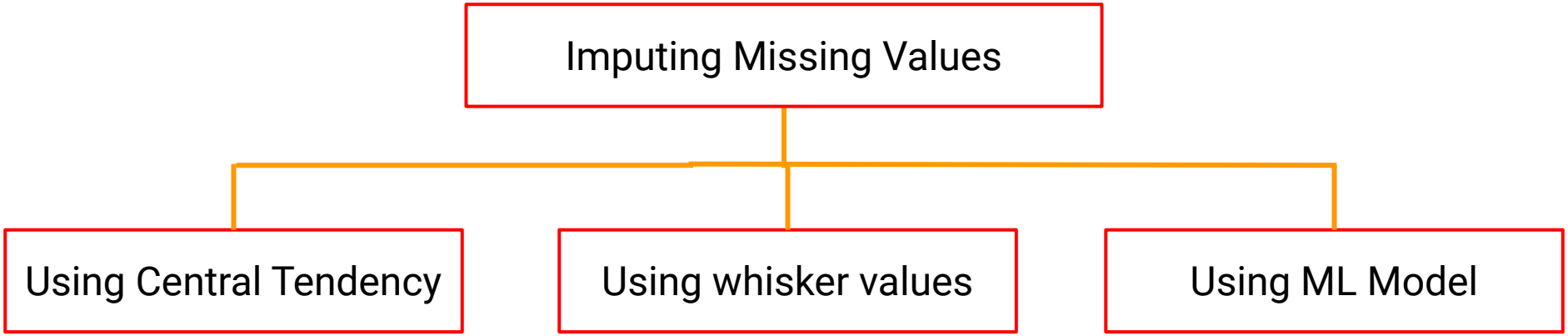
Dealing with Outliers



- Categorical - Mode
- Numerical - Median

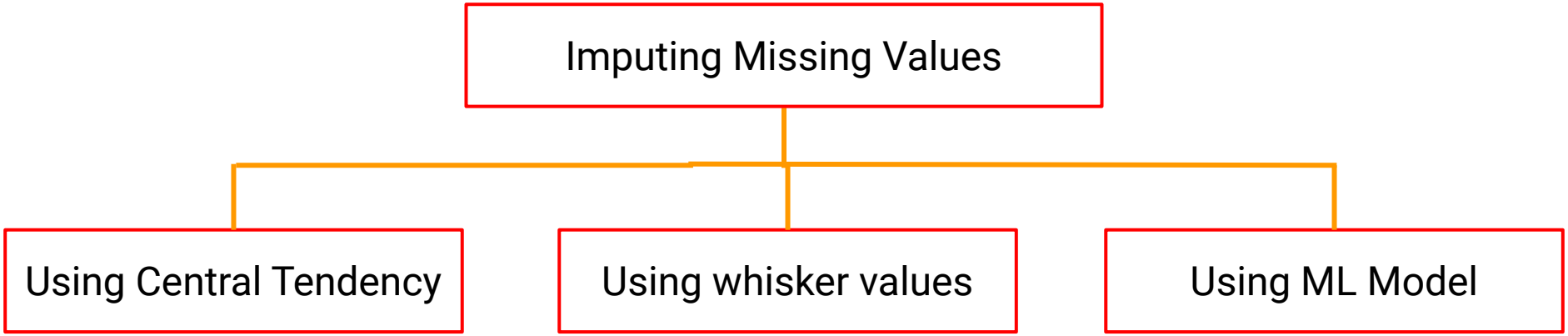
Analytics Vidhya
Learn everything about analytics

Dealing with Outliers



- Lower end - Q1 - 1.5*IQR
- Upper end - Q3 - 1.5*IQR

Dealing with Outliers



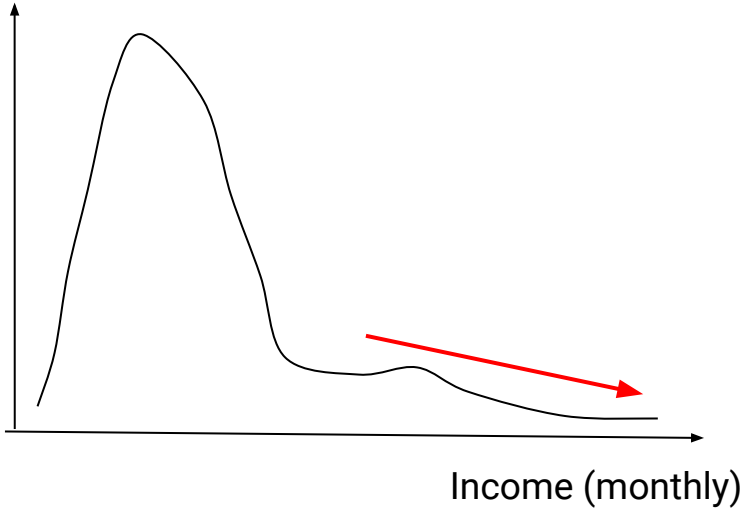
- Target -Column with Outlier
- Features - Other Attributes

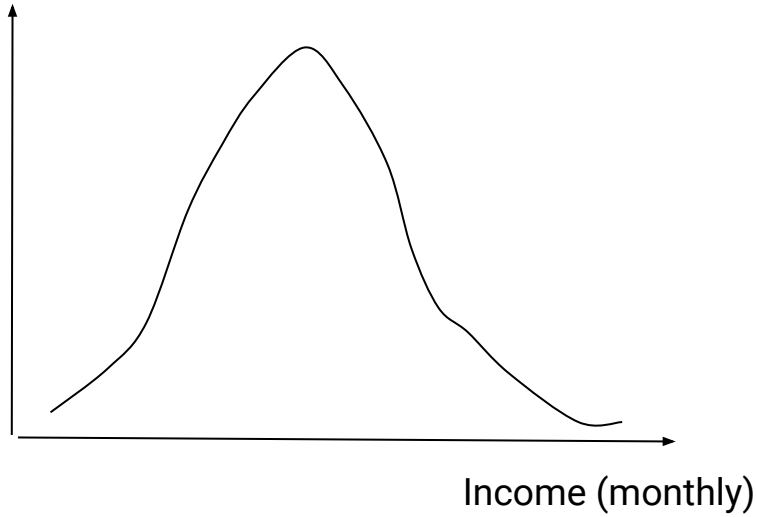
Analytics Vidhya
Learn everything about analytics

Dealing with Outliers

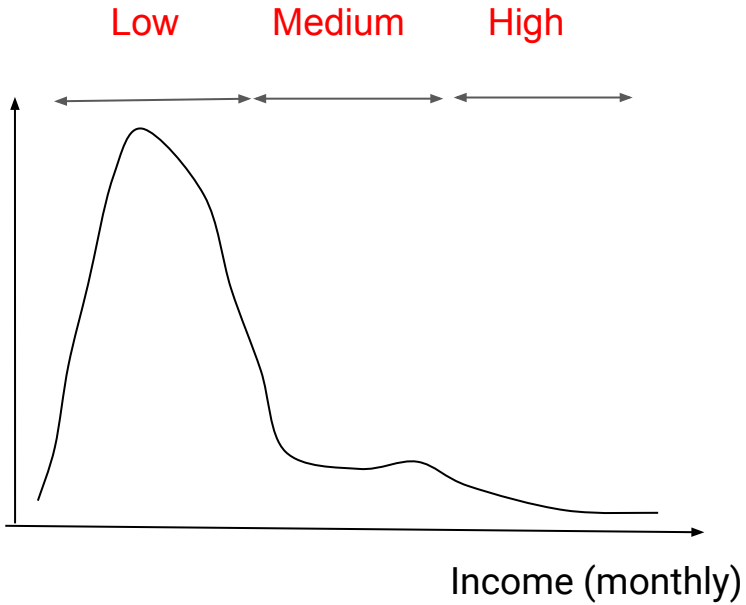
- Delete the data points
- Replace Outliers
  - Using central tendency
  - Replace with whisker values
  - Using an ML model
- Transform the values
  - Variable transformation
  - Binning

Variable Transformation





Binning Values



Thank You!

Delete Observations with Extreme Value

Notebook