Data Exploration

Stages of Data Exploration

Stages of Data Exploration

Stages of Data Exploration

Stages of Data Exploration

Reading the
Data

Univariate
Analysis

Missing Value
Treatment

Variable
Transformation

Variable
Identification

Bi-variate
Analysis

Outlier
Treatment

Topics to be covered:

1. What is Univariate Analysis?

2. Why do we need Univariate Analysis?

3. What we look for in Univariate Analysis?

4. How to perform Univariate Analysis?

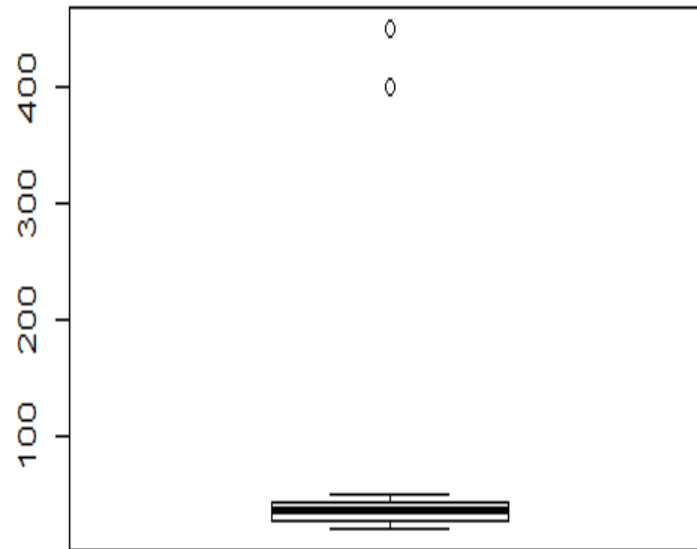# Stages of Data Exploration

What is Univariate Analysis?

1. Explore one variable at a time

2. Summarize the variable

3. Make sense out of that summary to discover insights, anomalies etc.

Stages of Data Exploration

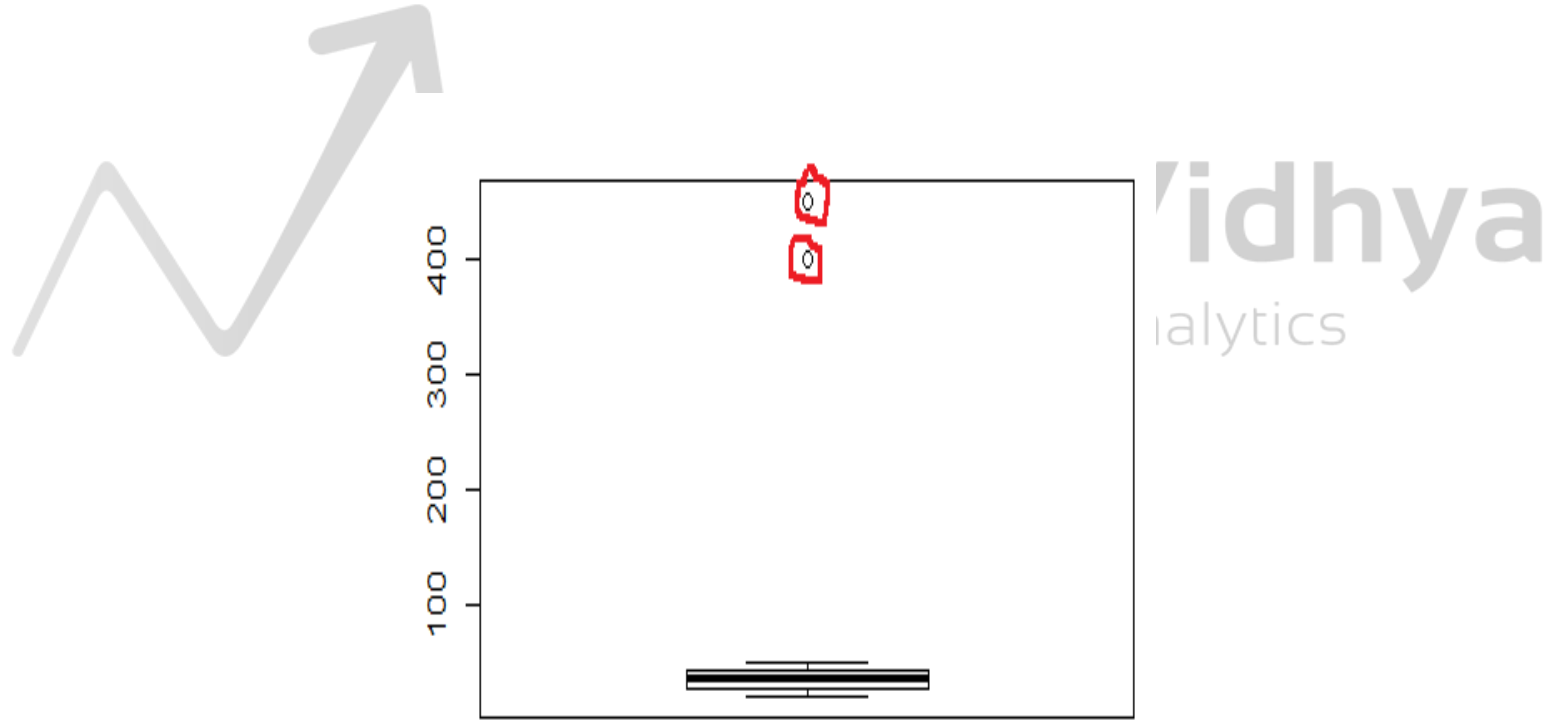**Why do we need Univariate Analysis?**

Stages of Data Exploration

Why do we need Univariate Analysis?

Stages of Data Exploration

Exploration methods depends on type of variable

Stages of Data Exploration

Continuous Variable

Categorical Variable

Stages of Data Exploration

What do we look for in Univariate Analysis?

Continuous Variable

- **Central Tendency and dispersion** – mean, median, standard deviation

- **Distribution of variable** – Symmetric / Right Skewed / Left Skewed

- **Presence of Missing Values**

- **Presence of Outliers**

Categorical Variable

Stages of Data Exploration

How to perform Univariate Analysis?

Continuous Variable

- **Tabular methods** – Mean, Median, Standard Deviation and missing values.

- **Graphical Methods** – Distribution of variable, Presence of Outliers

Categorical Variable

Analytics Vidhya
Learn everything about analytics

Stages of Data Exploration

How to perform Univariate Analysis?

Continuous
Variable

Categorical
Variable

Tabular methods

describe()

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

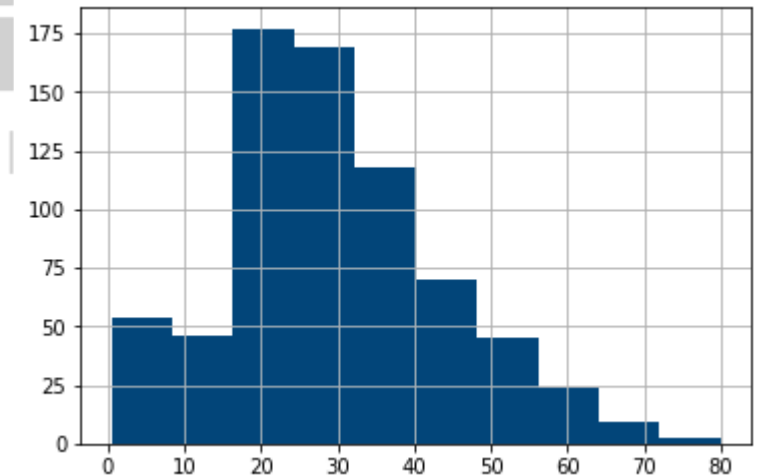# Stages of Data Exploration

How to perform Univariate Analysis?

Continuous
Variable

Categorical
Variable

Graphical Methods

Histogram – Distribution of a Variable

pd.DataFrame.hist()

Analytics Vidhya
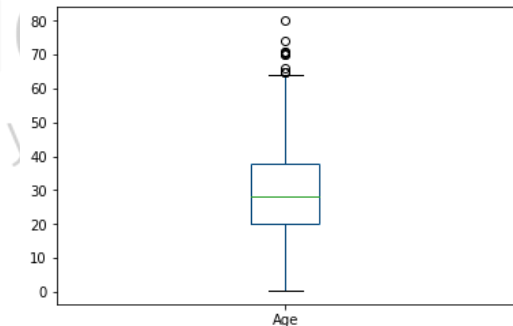Learn everything about analytics

Stages of Data Exploration

How to perform Univariate Analysis?

Continuous Variable

Categorical Variable

Graphical Methods

Boxplot – Detection of Outliers

pd.DataFrame.boxplot()

Stages of Data Exploration

Reading the
Data

**Univariate
Analysis**

Missing Value
Treatment

Variable
Transformation

Variable
Identification

Bi-variate
Analysis

Outlier
Treatment

Topics covered:

1.   What is Univariate Analysis?

2.   Why do we need Univariate Analysis?

3.   What we look for in Univariate Analysis?

4.   How to perform Univariate Analysis?