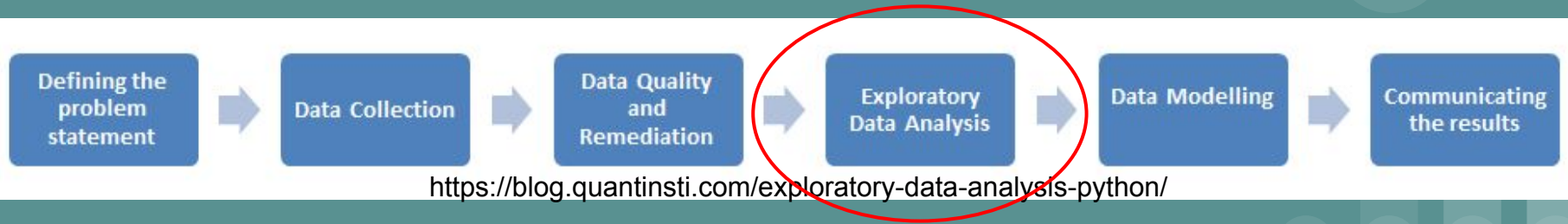


# EDA 4 Learning Data

*Exploratory Data Analysis,  
a gentle introduction*



# Data Collection

<https://www.kaggle.com/prasy46/innercity>



- Download Dataset

- innercity.csv.zip

- Dataset

- Innercity
  - Code
  - Discussion
  - Activity
  - Metadata

**innercity**

Prasanna Venkatesh • updated 2 years ago (Version 1)

Data Tasks Code (5) Discussion Activity Metadata

Download (858 KB) New Notebook

Usability 4.7 Tags No tags yet

**Description**

**Problem statement:**

A house value is simply more than location and square footage. Like the features that make up a person, an educated party would want to know all aspects that give a house its value. For example, you want to sell a house and you don't know the price which you usually try to find similar properties in your neighbourhood and based on gathered data you

**Objective:**

```
dfCity.describe()
```

	cid	price	room_bed	room_bath	living_measure	lot_me
count	2.161300e+04	2.161300e+04	21613.000000	21613.000000	21613.000000	2.161300
mean	4.580302e+09	5.401822e+05	3.370842	2.114757	2079.899736	1.510697
std	2.876566e+09	3.673622e+05	0.930062	0.770163	918.440897	4.142051
min	1.000102e+06	7.500000e+04	0.000000	0.000000	290.000000	5.200000
25%	2.123049e+09	3.219500e+05	3.000000	1.750000	1427.000000	5.040000
50%	3.904930e+09	4.500000e+05	3.000000	2.250000	1910.000000	7.618000
75%	7.308900e+09	6.450000e+05	4.000000	2.500000	2550.000000	1.068800
max	9.900000e+09	7.700000e+06	33.000000	8.000000	13540.000000	1.651359

10 of 23 columns

Opening innercity.csv.zip

You have chosen to open:  
innercity.csv.zip  
which is: Archivio zip (858 KB)  
from: https://storage.googleapis.com

What should Firefox do with this file?

☐ Open with Gestore di archivi Engrampa (default)

☒ Save File

☐ Do this automatically for files like this from now on.

Cancel OK

# Statistica Descrittiva

## ● Misure di posizione

- Moda
- Mediana
- Quantili, percentili, ...

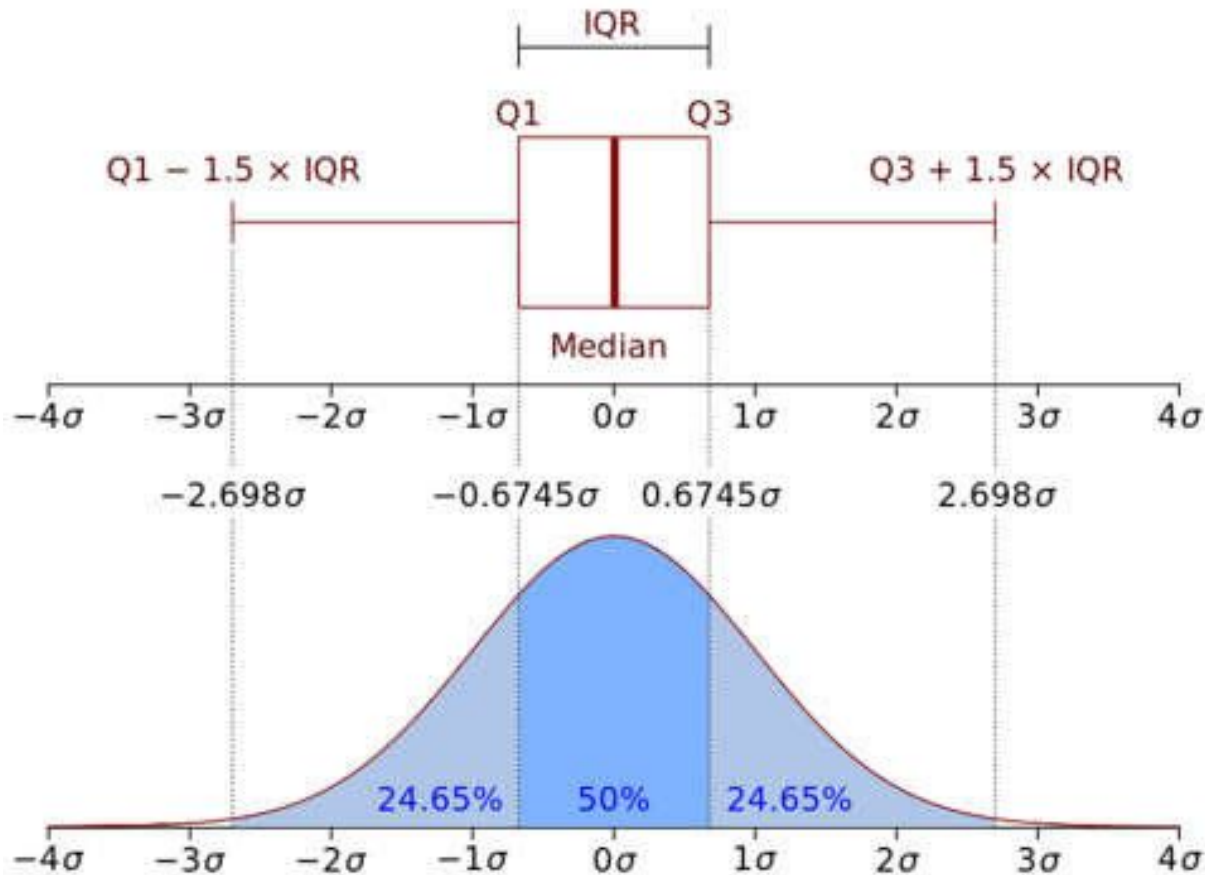
## ● Moda

- max absolute or relative frequency

## ● Mediana

- central position

## ● Quantili, Percentili



# Python Slicing

str\_object[start\_pos:end\_pos:step]

## ● Slice indexes

- rules
  - direct
  - reverse
  - steps

## ● Lambda Func

- one-line expression
  - keyword
  - argument
  - assignment

## ● Outliers

- Quantile & Boxplot
  - Q1 low limit
  - Q3 high limit

## ● Linear Regression

s = 'HelloWorld'

s[:i] + s[i:] == s

s1 = s[8:1:-1]

Start to End ->

H	e	l	l	o	W	o	r	l	D
0	1	2	3	4	5	6	7	8	9
9	8	7	6	5	4	3	2	1	0

JournalDev  
Output: lroWoll  
<- End to Start

Start to End ->

H	e	l	l	o	W	o	r	l	D
0	1	2	3	4	5	6	7	8	9
9	8	7	6	5	4	3	2	1	0

JournalDev  
Output: lool  
<- End to Start

Start to End ->

H	e	l	l	o	W	o	r	l	D
0	1	2	3	4	5	6	7	8	9
9	8	7	6	5	4	3	2	1	0
-9	-8	-7	-6	-5	-4	-3	-2	-1	0

JournalDev  
Output: or  
<- End to Start

Negative Index ->

# Python Programming

lambda argument (s): expression

## ● Slice indexes

- rules
  - direct
  - reverse
  - steps

## ● Lambda Func

- one-line expression
  - keyword
  - argument
  - assignment

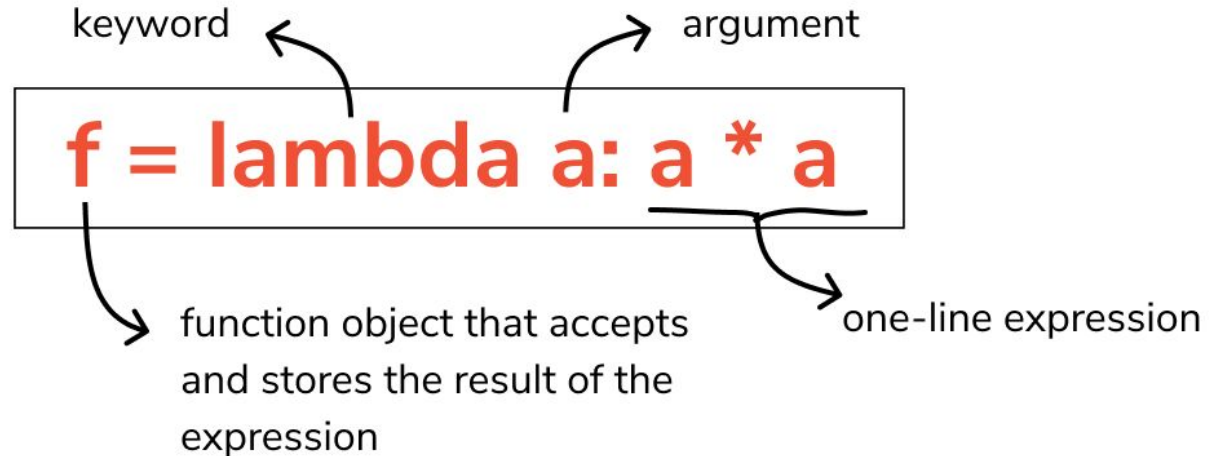
## ● Outliers

- Quantile & Boxplot
  - Q1 low limit
  - Q3 high limit

## ● Linear Regression

```
def square(a):
    return a * a
```

```
#main() function
res = square(6)
print(res)
```



# Linear Regression

## ● Slice indexes

### ○ rules

- direct
- reverse
- steps

## ● Lambda Func

### ○ one-line expression

- keyword
- argument
- assignment

## ● Outliers

### ○ Quantile & Boxplot

- Q1 low limit
- Q3 high limit

## ● Linear Regression

