Data Science, lesson 5

Exploratory Data Analysis

GENERAL ASSEMBLY

Learning Objectives

After this lesson, you'll be able to:

- Explain what is Exploratory Data Analysis (EDA).
- Describe the differences between NumPy Arrays and Pandas DataFrames.
- Manipulate Pandas DataFrames and Series.
- Handle null and missing values.

What is EDA and why do I need to waste my time on it?

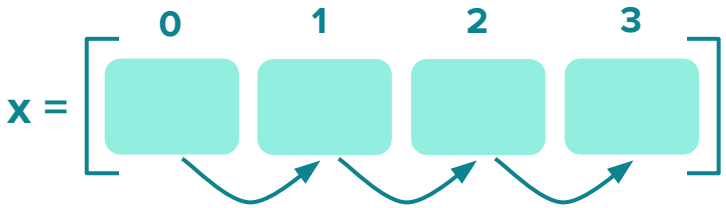
Answer questions
Test business assumptions
Generate hypotheses

Rearrange and combine datasets
Clean the data
Prepare the data for modelling

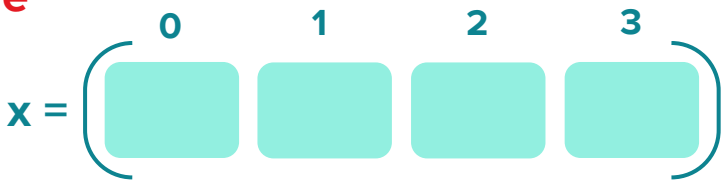
Good knowledge of your data!

Complex Variables

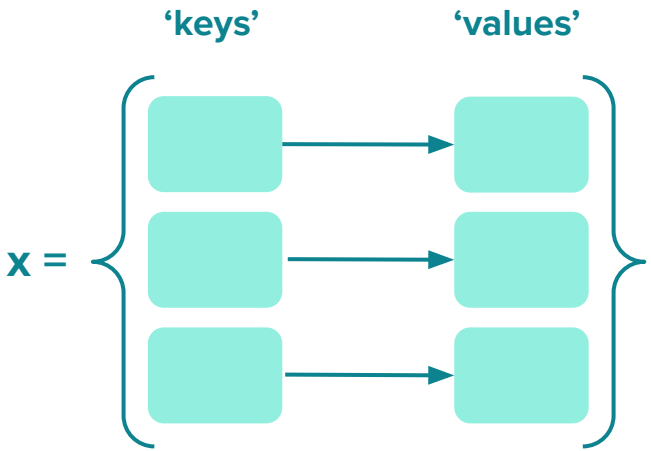
List



Tuple

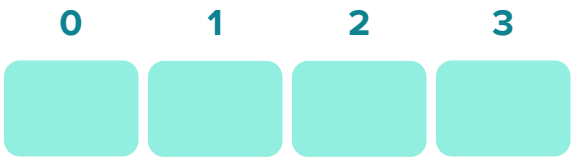


Dictionary



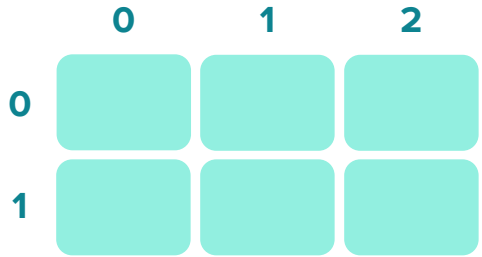
NumPy Arrays

1D



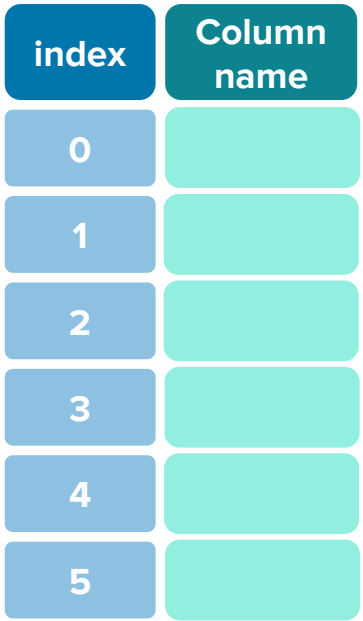
3D

2D

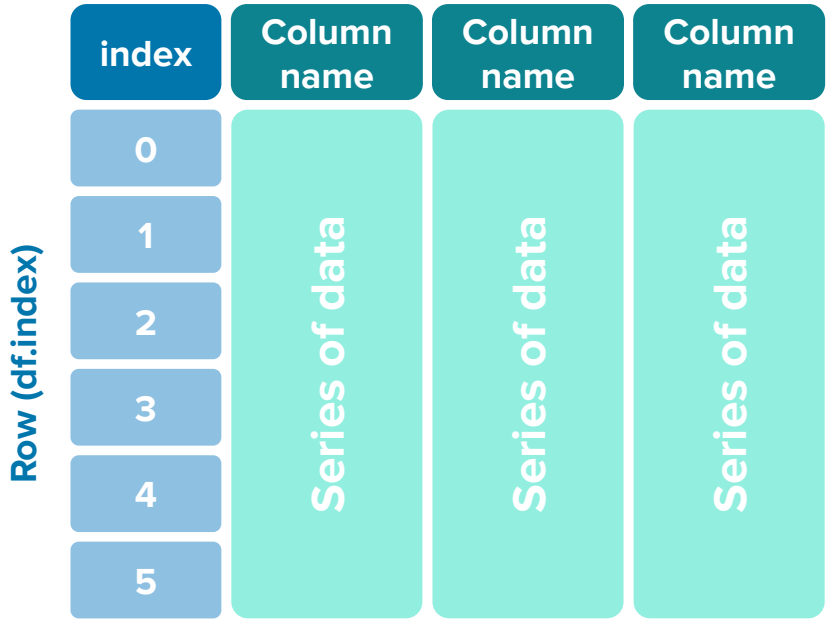


Pandas DataFrames

Series



Dataframe



Attributes and Methods

Each object of a class has:

Attributes are the features of the objects or the variables used in a class

Methods are the operations or activities performed by/on that object defined as functions in the class.

```
# Pandas DataFrames are Pandas class objects and
therefore come with attributes and methods. To access
these, follow the variable name with a dot.

# Access the 'index' attribute -- note there are no
parentheses. Attributes are not callable:
users.index

# Call the 'head' method (since there are open/closed
parentheses):
users.head()

# Call the 'head' method with parameter '10',
indicating the first 10 rows:
users.head(10)

# Alternatively, call the 'head' method, explicitly
setting the parameter 'n' to have a value of '10'.
users.head(n=10)
```


Data Science

Conclusion

Ask Me Anything!

Before Next Class:

A. Get into habit of using Git and GitHub
B. Practice your Python!

See you next time!

Thank you!

GENERAL ASSEMBLY


GA