Artificial Intelligence (Machine Learning & Deep Learning) [Course]

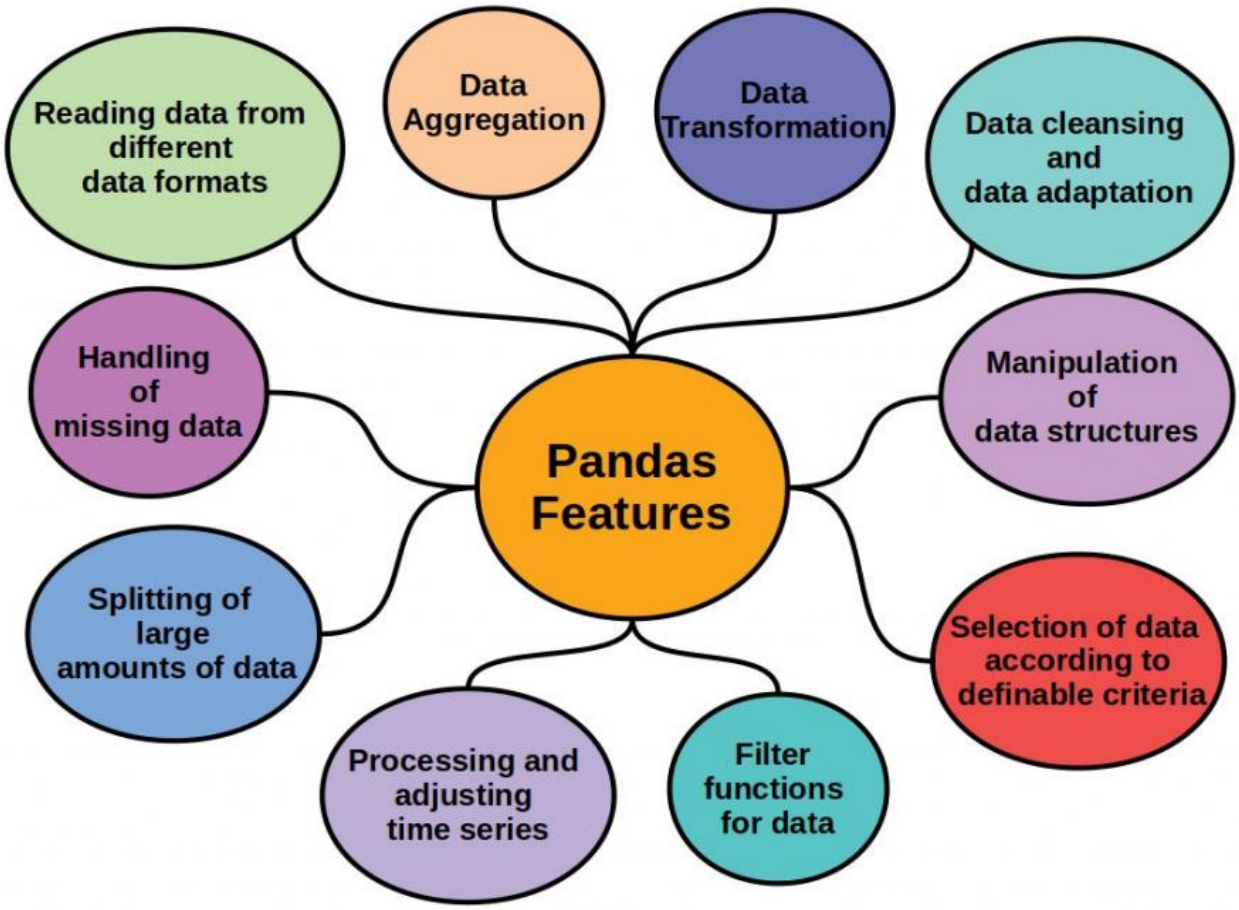
Week 4 - Day 1 n Day 2 - Pandas

[See examples / code in GitHub code repository]

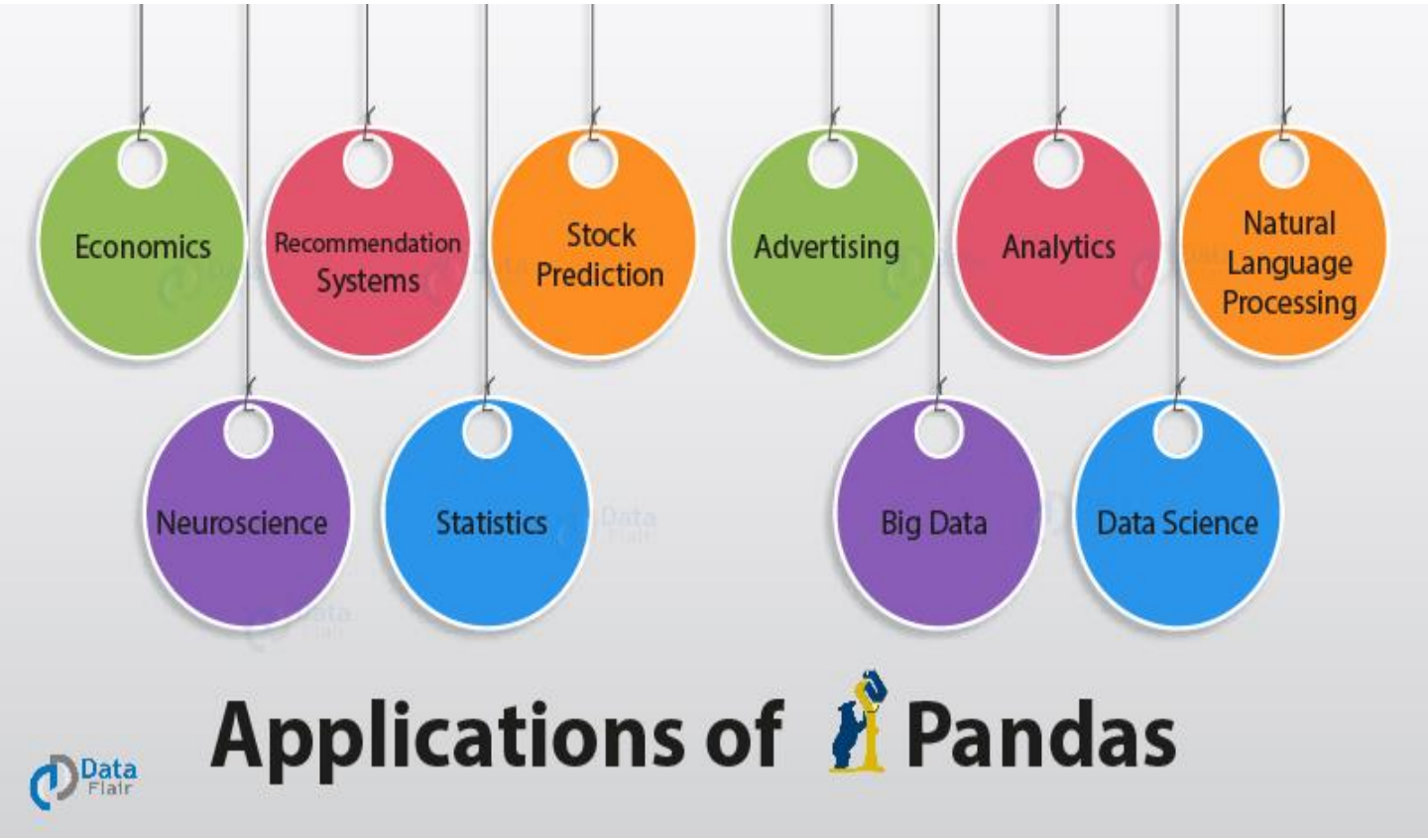
It is not about Theory, it is 20% Theory and 80% Practical – Technical/Development/Programming [Mostly Python based]

Pandas - Overview

- Pandas is a Python library used for working with data sets.
- It has functions for analyzing, cleaning, exploring, and manipulating data.
- The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.



Pandas – Use cases

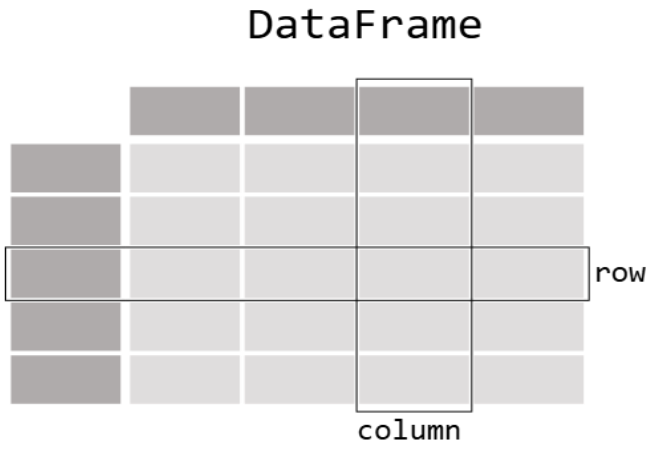


Reference:

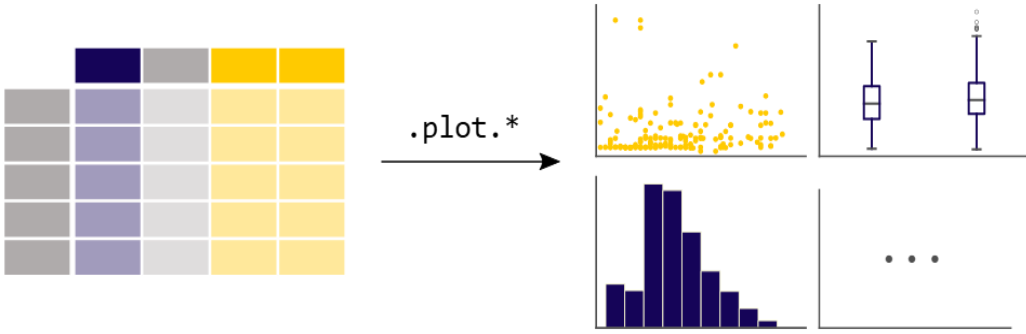
https://data-flair.training/blogs/applications-of-pandas/

Pandas – Key Concepts

- A data table is called a DataFrame.



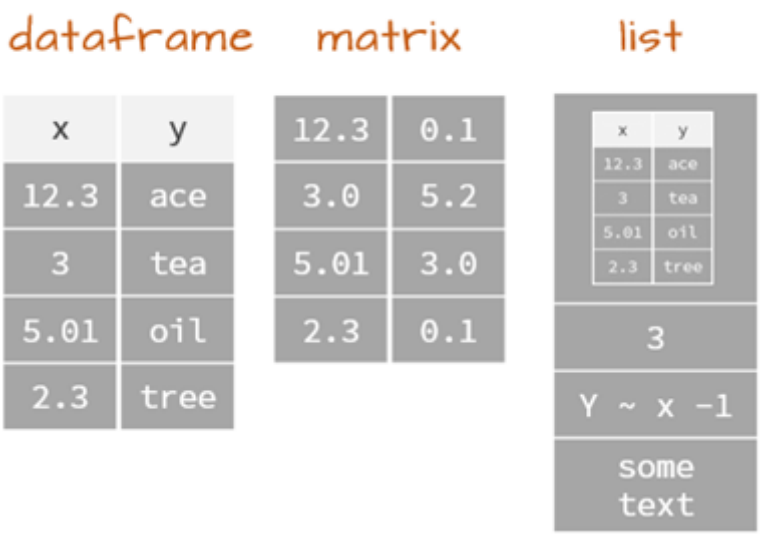
- Pandas provides plotting your data out of the box, using the power of Matplotlib.



	pandas	NumPy
DATA TYPE	Tabular	Numerical
STRENGTHS	Data frame, Series	Arrays
AMOUNT OF DATA BEING ANALYZED	>500K rows	<50K rows
DATA TYPES	Can contain dissimilar data types	Homogenous data types
MEMORY USAGE	More	Less
SPEED	Slower	Faster

python

Pandas – Key Concepts

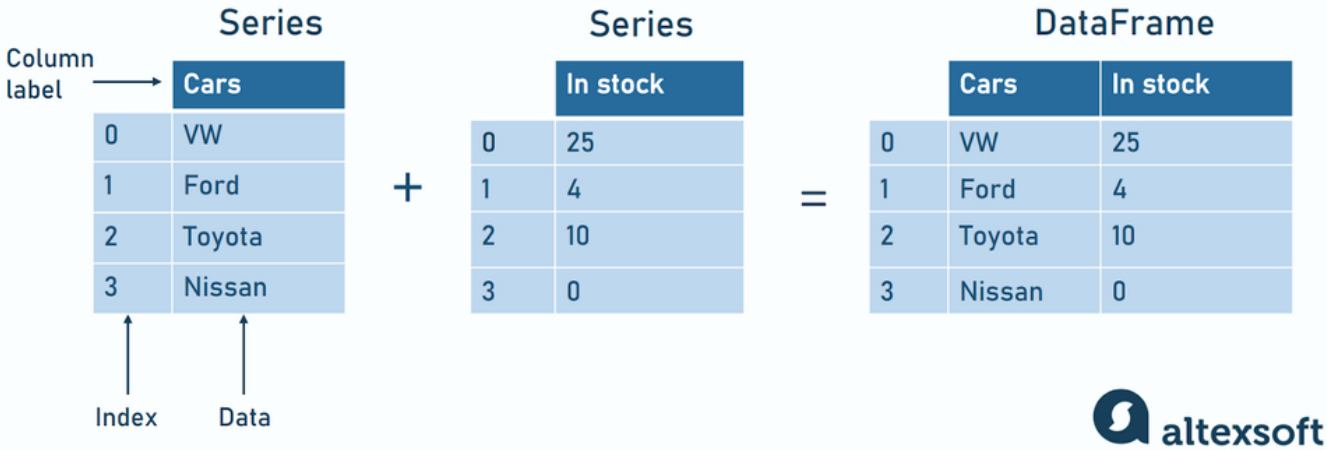


Characteristics	NumPy Array	Pandas Dataframe
Homogeneity	Arrays consist of only homogeneous elements (elements of same data type)	Dataframes have heterogeneous elements.
Mutability	Arrays are mutable	Dataframes are mutable
Access	Array elements can be accessed using integer positions.	Dataframes can be accessed using both integer position as well as index.
Flexibility	Arrays do not have flexibility to deal with dynamic data sequence and mixed data types.	Dataframes have that flexibility.
Data type	Array deals with numerical data.	Dataframes deal with tabular data.

Pandas – Key Concepts

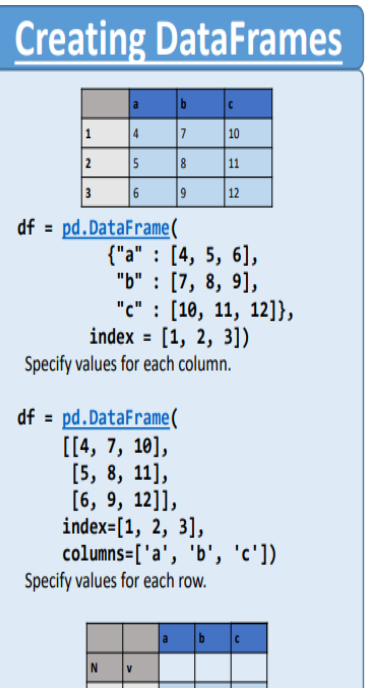
Pandas dtype	Python type	NumPy type	Usage
object	str or mixed	string_, unicode_, mixed types	Text or mixed numeric and non-numeric values
int64	int	int_, int8, int16, int32, int64, uint8, uint16, uint32, uint64	Integer numbers
float64	float	float_, float16, float32, float64	Floating point numbers
bool	bool	bool_	True/False values
datetime64	NA	datetime64[ns]	Date and time values
timedelta[ns]	NA	NA	Differences between two datetimes
category	NA	NA	Finite list of text values

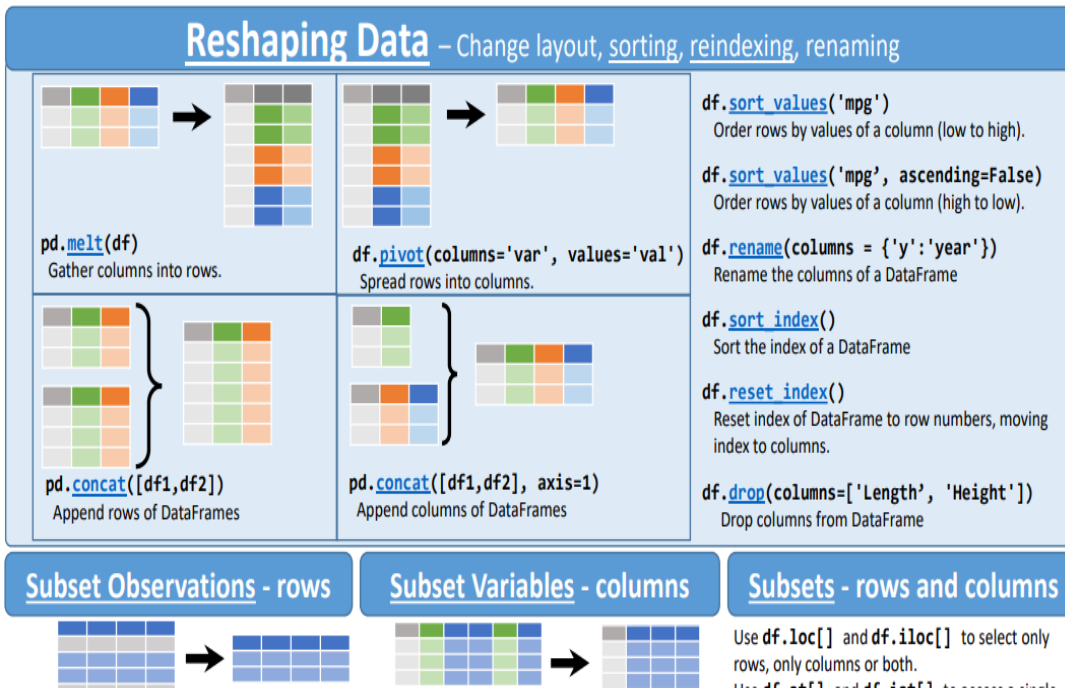
DATA STRUCTURES IN PANDAS



altexsoft

Pandas – cheat sheets





Python Pandas Cheat Sheet

https://www.datasciencecentral.com/data-science-in-python-pandas-cheat-sheet/

https://pandas.pydata.org/Pandas_Cheat_Sheet.pdf

https://www.datacamp.com/cheat-sheet/pandas-cheat-sheet-data-wrangling-in-python

Pandas - Exercises

See code here: https://github.com/ShahzadSarwar10/AI-ML-Explorer
You should be able to analyze – each code statement, you should be able to see trace information – at each step of debugging. "DEBUGGING IS BEST STRATEGY TO LEARN A LANAGUAGE." So debug code files, line by line, analyze the values of variable – changing at each code statement. BEST STRATEGY TO LEARN DEEP.

Let's put best efforts.
Thanks.
Shahzad – Your AI – ML Instructor

Thank you - for listening and participating

- ❑**Questions / Queries**
- ❑**Suggestions/Recommendation**
- ❑**Ideas…..?**

Shahzad Sarwar
Cognitive Convergence
https://cognitiveconvergence.com
shahzad@cognitiveconvergence.com
voice: +1 4242530744 (USA) +92-3004762901 (Pak)