

2018

WEDNESDAY

DAY (136-229)

20th Week

16

PANDAS

for

DATA ANALYSIS

~~10~~

>> Wes McKinney (2008)

F S S M T W T F S S M T W T F S S M T W T F S S M T W T F S S JUN
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 • '18

2018

FRIDAY
DAY (138-227)
20th Week

18

Appointment	Notes	Work to do
<h2>→ Pandas : Reading and Writing Files</h2>		
Category		Topics Covered
Introduction :-		<ul style="list-style-type: none"> - overview of Pandas file handling. - supported formats : CSV, Excel, etc.
Reading Files :-		<ul style="list-style-type: none"> - Reading CSV file using <code>pd.read_csv("file.csv")</code>. - Excel : <code>pd.read_excel("file.xlsx", sheet_name = "sheet 1")</code>. - JSON : <code>pd.read_json("file.json")</code>.
Writing Files :-		<ul style="list-style-type: none"> - Saving to CSV using <code>df.to_csv("output.csv", index = False)</code>. - Excel : <code>df.to_excel("output.xlsx", index = False)</code>.
Real-world Applications :-		<ul style="list-style-type: none"> > Handling e-commerce data (product, user, and transaction records).
Key Considerations		<ul style="list-style-type: none"> > Renaming encoding issues with <code>encoding = "utf-8"</code> or <code>encoding = "latin1"</code>. > Processing large datasets in smaller chunks using <code>chunksize</code>.
Exploring Data		<ul style="list-style-type: none"> > data using : <code>head()</code> <code>tail()</code>. > data structure : <code>info()</code> <code>describe()</code>. > dataset : <code>shape</code> and <code>columns</code>.

F	S	S	M	T	W	T	F	S	S	M	T	W	T	F	S	S	JUN														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	•	'18

19

SATURDAY
DAY (139-226)
20th Week

Filtering and Selecting

- > Selecting specific columns with `df[["column"]]` or `df[[col1, col2]]`
- > Filtering rows using condition with `col == (AND)` and `"`
- > Applying filtering techniques for real world Data Manipulation.

Filtering and Selecting

- > Selecting specific columns with `If ["column"] of [col1, col2, ...]`
 - > Filtering rows using conditions with `col, (AND) and "`
 - > Applying filtering techniques for real world Data selection.

Some of the parts will be used
for other vehicles.

What is the relationship between the two? How can we use this?

20 SUNDAY

Ma

2018

MONDAY
DAY (141-224)
21st Week

21

Roles and Uses :-

Field/Job Role	Use Cases
Data Scientist	> processing data, feature selection, and exploratory data analysis.
Data Analyst	> cleaning and summarizing data, Generating Insights.
Machine Learning Engineer	> preparing datasets for training machine learning models.
Artificial Intelligence Engineer	> handling large datasets for AI model development.
Business Analyst	>> analyzing business data to improve decision-making.
Financial Analyst	> analysing financial trends, managing reports.
Research Scientist	> handling research datasets, statistical analysis.
Marketing Analyst	>> Analyzing customer behavior and campaign data.
Operations Analyst	> Optimizing operational processes through data insights.
Software Developer	>> Creating backend systems that process and analyze data.

TUESDAY
DAY (142-223)
21st Week

22

Note

Work to do

May

2018

Data Manipulation :-

>> Definition : Changing, organizing, or preparing data to make it useful and easier to understand.

>> Goal : To clean, transform, and structure raw data for better usability.

>> Example :-

(i) Organizing a grocery list i.e., sorting random items into categories like "Fruits" or "Dairy".

(ii) Fixing errors in student record e.g., correcting missing or wrong grades.

Data Analysis :-

→ Creator :-

>> Wes McKinney, a data scientist and software developer, created Pandas in 2008.

* While working at AQR Capital Management, Wes faced challenges analyzing large financial data sets.

* Existing tools like Excel were inefficient for large-scale data cleaning and analysis.

i) Tracking fitness progress

May	T	W	F	S	S	M	T	W	F	S	S	M	T	W	F	S
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

May	T	W	F	S	S	M	T	W	F	S	S	M	T	W	F	S
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16

WEDNESDAY
DAY (143-222)
21st Week

23

Notes

Work to do

May

2018

Data Manipulation

Data Analysis

Organize and structure raw data. Extracting insights from prepared data.

Fixing errors in a student's grade sheet.

Analyzing which student scored marks the highest.

Find patterns, trends, and outliers.

24 THURSDAY
DAY (144-221)
21st Week

May

2018

Appointment

Note

Work to do

Appointment

Note

Work to do

Pandas :-

- » Pandas is a powerful and popular Python library designed for data manipulation (cleaning, transforming, and structuring data) and data analysis (finding patterns, trends and insights).

»

- » It simplifies working with structured data such as CSV files, Excel, databases, etc.

→ tables

→ spreadsheets

→ time-series data

* Key Features :-

- work seamlessly with structured data formats like CSV and Excel.
- Handles missing values easily.
- Built on Numpy for fast computations.

* Data Frame :-

- Data Frame :- a two-dimensional labeled data structure in Pandas, similar to a table in a database, an Excel spreadsheet or a SQL table.

* Consists of Rows / Column where

→ Row have indices

» Columns have names

Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

FRIDAY
DAY (145-220)
21st Week

25