📓 Day 3 – AI-ML, Pandas Dataset Exploration & Kaggle

📅 Date: **June 26, 2025**

◆ Overview

On Day 3, we explored the use of Python for real-world data analysis using the **Pandas** library. We also had an introduction to **Artificial Intelligence and Machine Learning**, explored **Kaggle.com**, and worked on a real dataset: **Happiness Report 2018–2019**.

□ Introduction to AI-ML

✨ What is AI?

Artificial Intelligence (AI) is the ability of machines to mimic human intelligence. It powers systems like voice assistants, recommendation engines, and self-driving cars.

□ What is Machine Learning?

ML is a subfield of AI that enables machines to learn from data without being explicitly programmed.

✅ Real-World Applications:

- YouTube & Netflix recommendations
- Face recognition systems
- Predictive typing
- Self-driving vehicles

📊 Dataset Operations using Pandas in Python

We used the **Happiness Report 2018–2019** dataset and performed the following operations using the **Pandas** library in **Google Colab**.

📌 Dataset Columns:

- Overall rank
- Country or region
- Year
- Score
- GDP per capita
- Social support
- Healthy life expectancy
- Freedom to make life choices

- Generosity
- Perceptions of corruption

□ Code and Outputs

✅ 1. Importing Dataset

```
import pandas as pd
df = pd.read_csv("report_2018-2019.csv")
```

✅ 2. View First 5 Rows

```
df.head()
```

🔍 Output (sample):

Overall rank	Country	Year	Score	GDP	Social support	Life exp.	Freedom	Generosity	Corruption
154	Afghanistan	2019	3.203	0.35	0.517	0.361	0.000	0.158	0.025
145	Afghanistan	2018	3.632	...	...	...	...	...	...

✅ 3. View Last 5 Rows

```
df.tail()
```

🔍 Output:

Last few countries like Zimbabwe, Yemen, Zambia and their scores

---

✅ 4. Get Dataset Info

```
df.info()
```

🔍 Output Summary:

- Total entries: 312
- 10 columns: including int, float, and object types
- No null values present

✅ 5. Drop Missing Values

```
df.dropna()
```

- Shows the dataset again as there were **no missing values** to drop.

✅ 6. Check for Missing Values

```
df.isnull()
```

- Output shows `False` for all entries, meaning no null values exist.

✅ 7. Get Statistical Summary

```
df.describe()
```

🔍 Output Summary:

Metric	Score	GDP	Social Support	Life Expectancy
Count	312	312	312	312
Mean	5.39	0.89	1.21	0.66
Std Dev	1.11	...	...	...
Min–Max	2.85–7.76	...	...	...

✅ 8. Null Value Count Column-Wise

```
df.isnull().sum()
```

- All columns show **0 missing values**.

✅ 9. Select Specific Rows by Index

```
df.iloc[[0, 1, 2]]
```

- Displays the first 3 rows from the dataset (Afghanistan & Albania records)

✅ 10. Select Data Based on a Condition

```
df.loc[(df.Score > 5)]
```

- Filters and displays all countries with a **happiness score > 5**
- Total **194 countries** met this condition.

🌐 Exploring Kaggle

We also explored Kaggle.com, a powerful platform for data science and machine learning enthusiasts

✨ **What We Explored:**

- Public Datasets (CSV, Excel, JSON formats)
- Competitions on real-world problems
- Shared Notebooks from global users
- Learning resources and courses
- Discussions and Q&A forums

📁 Notebook Summary

We wrote and executed our code in a Jupyter Notebook using **Google Colab**, making it easier to share and store our daily progress.

Notebook operations included:

- Importing libraries
- Loading datasets
- Exploring data using `.head()`, `.tail()`, `.info()`, `.describe()`
- Filtering and selecting data using `.iloc[]`, `.loc[]`