📊 Day 11 – Pandas Data Analysis: Used Bikes Dataset

Today, we explored the Pandas library using a real-world dataset: **Used_Bikes.csv**.

🐼 What is Pandas?

Pandas is a Python library used for working with structured data.
It provides two primary data structures:

- `DataFrame` : Table-like structure
- `Series` : Single column

📥 1. Importing Libraries and Reading Data

```
import pandas as pd
df = pd.read_csv("Used_Bikes.csv")
```

ℹ️ 2. Overview of the Dataset

`df.info()`
Output:

Column	Non-Null Count	Dtype
bike_name	7324	object
price	7324	float64
city	7324	object
kms_driven	7324	float64
owner	7324	object
age	7324	float64
power	7324	float64
brand	7324	object

🔍 3. Removing Duplicates

```
df.duplicated().sum()
# Output: 25324
```

```
df.drop_duplicates(inplace=True)
df.duplicated().sum()
# Output: 0
```

📈 4. Descriptive Statistics

```
df.describe()
```

Metric	price	kms_driven	age	power
count	7324	7324	7324	7324
mean	84883.9	23910.5	6.66	228.13
std	120966.2	27317.6	3.61	158.32
min	4400	1	1	100
max	1900000	750000	63	1800

🗒️ 5. Viewing Data

```
df.head(5)
df.tail(5)
```

📊 6. Columns, Data Types & Shape

```
df.columns  # output- no. of columns in the table
df.dtypes   # output- data types of the columns
df.shape    # output- dimensions mentioning number of rows x number of columns
```

Shape: (7324, 8)

Data types:

- object (categorical): bike_name, city, owner, brand
- float64 (numerical): price, kms_driven, age, power

🔢 7. Value Counts and Uniques

```
df['owner'].value_counts()
```

Owner Type	Count
First Owner	6642
Second Owner	588
Third Owner	84
Fourth Owner Or More	10

```
df['brand'].value_counts()
```

Top brands include Bajaj, Royal Enfield, Hero, Honda, Yamaha, etc.

🔍 8. Filtering Data

🎯 Royal Enfield with specific filters

```
royal_df = df[(df['brand']=="Royal Enfield") &
              (df['kms_driven'] <= 30000) &
              (df['owner'] == "First Owner") &
              (df['price'] <= 100000)]

# output--- list of bikes satisfying above conditions
```

🧪 Using `.query()` method:

```
df.query("brand=='Royal Enfield' and kms_driven==30000 and
owner=='First Owner' and price<=100000")
```

🚀 9. Filter by Brand and Condition

🔸 TVS bikes with age < 1 and price ≤ 40000

```
tvs_df = df[df['brand'] == "TVS"]
tvs_age_df = tvs_df[tvs_df['age'] < 1]
tvs_price_df = tvs_df[tvs_df['price'] <= 40000]
```

🛠️ 10. Feature Selection

```
df['brand']
df['owner']
```

You can use square brackets to select a single or multiple columns.

✅ Summary of Commands Used:

Command	Description
`pd.read_csv()`	Load data from CSV
`df.info()`	Basic info about DataFrame
`df.describe()`	Summary stats
`df.head()/df.tail()`	Preview first/last rows
`df.columns / df.shape / df.dtypes`	Metadata and structure
`df['col'].value_counts()`	Unique counts of a column

Command	Description
`df.duplicated().sum()`	Count duplicate rows
`df.drop_duplicates(inplace=True)`	Remove duplicates
`df.query("condition")`	Filter with readable syntax
`df[(condition1) & (condition2)]`	Boolean indexing for custom filtering

📌 Final Thoughts

Mastering **Pandas** is the first big step in becoming a data scientist.
You now have the skills to load, inspect, clean, filter, and explore datasets — which are critical in **real-world analytics and ML projects**.