

# pandas-part1

September 12, 2024

```
[1]: import pandas as pd

# Create a DataFrame with some missing values and inconsistent data
data = {'Name': ['Alice', 'Bob', 'Charlie', 'David'],
        'Age': [25, None, 30, '40'],
        'Email': ['alice@example.com', 'bob@example', 'charlie@example.com',
        ↪None]}

df = pd.DataFrame(data)

# Fill missing values
df['Age'] = df['Age'].fillna(df['Age'].mode()[0])
df['Email'] = df['Email'].fillna('unknown@example.com')

# Convert 'Age' to a proper numeric format
df['Age'] = pd.to_numeric(df['Age'], errors='coerce')

print(df)
```

	Name	Age	Email
0	Alice	25	alice@example.com
1	Bob	25	bob@example
2	Charlie	30	charlie@example.com
3	David	40	unknown@example.com

C:\Users\Ravi Kumar Verma\AppData\Local\Temp\ipykernel\_10436\1206009490.py:11:  
UserWarning: Unable to sort modes: '<' not supported between instances of 'str'  
and 'int'

```
df['Age'] = df['Age'].fillna(df['Age'].mode()[0])
```

```
[2]: import pandas as pd

# Create a DataFrame
data = {'Name': ['Alice', 'Bob', 'Charlie', 'David'],
        'Age': [25, 30, 35, 40]}

df = pd.DataFrame(data)
```

```
# Filter rows where Age is greater than 30
filtered_df = df[df['Age'] > 30]

print(filtered_df)
```

```
      Name  Age
2  Charlie   35
3   David   40
```

```
[3]: import pandas as pd

# Create a DataFrame
data = {'Department': ['HR', 'IT', 'IT', 'HR'],
        'Employee': ['Alice', 'Bob', 'Charlie', 'David']}

df = pd.DataFrame(data)

# Group by department and count the number of employees in each department
department_counts = df.groupby('Department').size()

print(department_counts)
```

```
Department
HR      2
IT      2
dtype: int64
```

```
[4]: import pandas as pd

# Create a DataFrame with date strings
data = {'Dates': ['2024-01-01', 'not a date', '2024-02-01', '2024-03-01']}
df = pd.DataFrame(data)

# Convert 'Dates' column to datetime, coercing errors
df['Dates'] = pd.to_datetime(df['Dates'], errors='coerce')

print(df)
```

```
      Dates
0 2024-01-01
1         NaT
2 2024-02-01
3 2024-03-01
```

```
[5]: import pandas as pd

# Create a DataFrame
```

```
data = {'Name': ['Alice', 'Bob', 'Charlie'],
        'Location': ['New York', 'Los Angeles', 'Chicago']}

df = pd.DataFrame(data)

# Add a new column
df['Location_Upper'] = df['Location'].str.upper()

print(df)
```

	Name	Location	Location_Upper
0	Alice	New York	NEW YORK
1	Bob	Los Angeles	LOS ANGELES
2	Charlie	Chicago	CHICAGO

```
[1]: import pandas as pd

# Load the CSV file into a DataFrame
df = pd.read_csv('data.csv')

# Display the first few rows of the DataFrame
print("DataFrame preview:")
print(df.head())

# Filter books by genre
fiction_books = df[df['Genre'] == 'Fiction']
print("\nFiction Books:")
print(fiction_books)

# Sort books by publication year
sorted_books = df.sort_values(by='PublicationYear')
print("\nBooks sorted by Publication Year:")
print(sorted_books)

# Find books by a specific author
author_name = 'Harper Lee'
books_by_author = df[df['Author'] == author_name]
print(f"\nBooks by {author_name}:")
print(books_by_author)

# Get unique genres
unique_genres = df['Genre'].unique()
print("\nUnique Genres:")
print(unique_genres)
```

DataFrame preview:

	Title	Author	Genre \
0	The Catcher in the Rye	J.D. Salinger	Fiction

1	To Kill a Mockingbird	Harper Lee	Fiction
2	A Brief History of Time	Stephen Hawking	Science
3	The Great Gatsby	F. Scott Fitzgerald	Fiction
4	Sapiens: A Brief History of Humankind	Yuval Noah Harari	Non-Fiction

	PublicationYear
0	1951
1	1960
2	1988
3	1925
4	2011

Fiction Books:

	Title	Author	Genre	PublicationYear
0	The Catcher in the Rye	J.D. Salinger	Fiction	1951
1	To Kill a Mockingbird	Harper Lee	Fiction	1960
3	The Great Gatsby	F. Scott Fitzgerald	Fiction	1925

Books sorted by Publication Year:

	Title	Author	Genre	\
3	The Great Gatsby	F. Scott Fitzgerald	Fiction	
0	The Catcher in the Rye	J.D. Salinger	Fiction	
1	To Kill a Mockingbird	Harper Lee	Fiction	
2	A Brief History of Time	Stephen Hawking	Science	
4	Sapiens: A Brief History of Humankind	Yuval Noah Harari	Non-Fiction	

	PublicationYear
3	1925
0	1951
1	1960
2	1988
4	2011

Books by Harper Lee:

	Title	Author	Genre	PublicationYear
1	To Kill a Mockingbird	Harper Lee	Fiction	1960

Unique Genres:

['Fiction' 'Science' 'Non-Fiction']

[ ]: