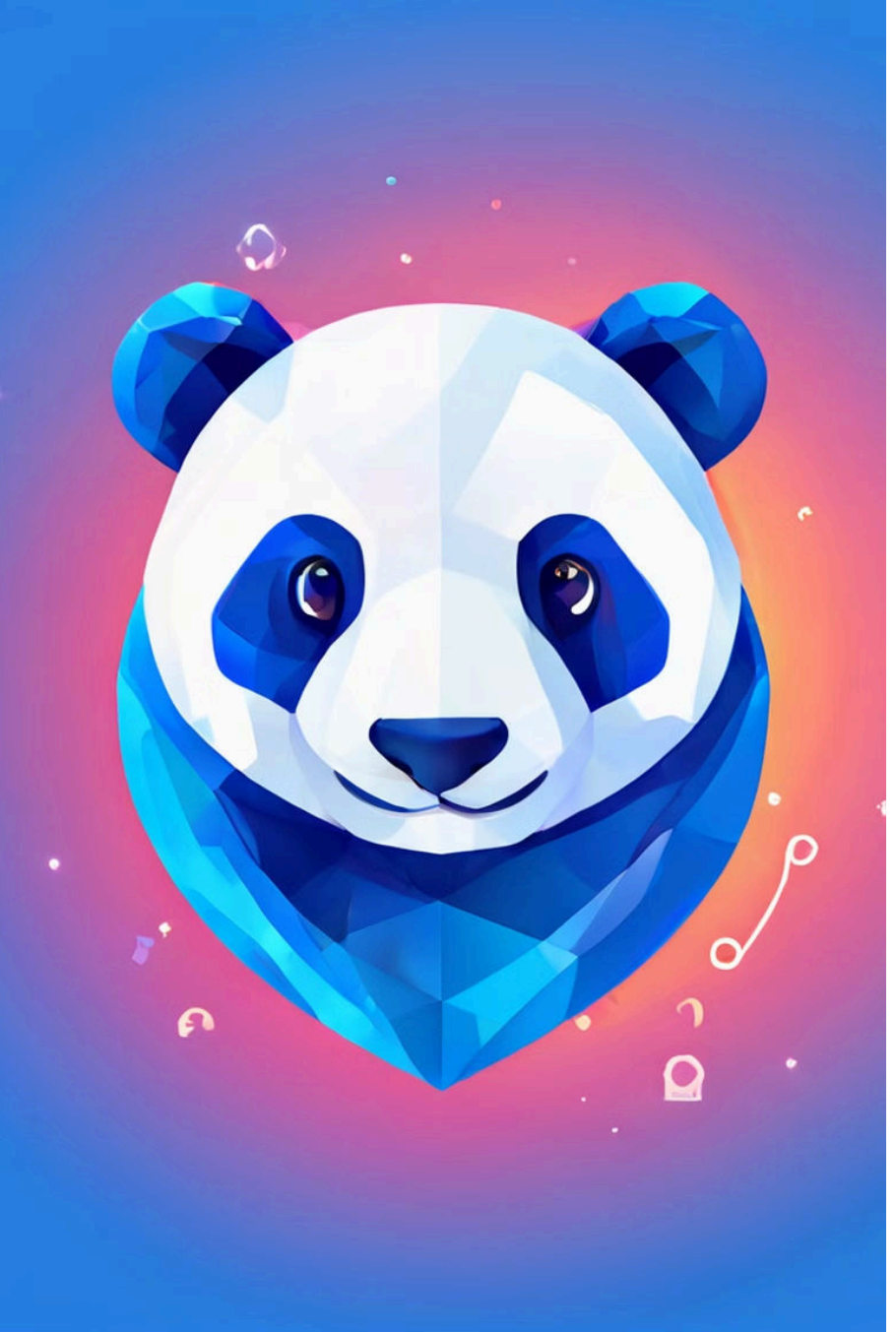

Introduction to Pandas

Pandas is a data analysis and data manipulation library built on top of NumPy in python. It provides data structures and functions needed to work with structured data seamlessly and efficiently. It is particularly useful for handling data in a tabular form, similar to SQL tables or Excel spreadsheets or CSV spreadsheets.

Key Features of Pandas

1 **Series**

One-dimensional labeled array capable of holding any data type.

2 **DataFrame**

Two-dimensional labeled data structure with columns of potentially different types. Kinda like a dictionary of series or a 2-dimensional array.

3 **Data Cleaning**

Handling missing data, filtering data, and replacing values.

4 **Data Transformation**

Applying functions to data, pivoting tables, and merging/joining datasets.

5 **Data Aggregation**

Grouping data and performing operations like sum, mean, or count.

Creating Data Structures

1

Install Pandas

Open your terminal and type: pip install pandas.

2

Create Series

```
s = pd.Series([1, 3, 5, np.nan, 6, 8])
print(s)
```

3

Create DataFrame

```
data = {'Name': ['Hamid', 'Bashir', 'Ghilas', 'Tiziri'], 'Age': [24, 27, 22, 32], 'City': ['Alger',
'Annaba', 'Oran', 'Tizi ouzou']}
df = pd.DataFrame(data)
print(df)
```

4

Read/Write CSV

```
df = pd.read_csv('path_to_file/data.csv')
print(df.head())
df.to_csv('output.csv', index=False)
```

Inspecting and Manipulating Data

Inspecting Data

- print(df.head())
- print(df.tail())
- print(df.info())
- print(df.describe())

Indexing & Selecting

- df['Age']
- df.loc[1:3, ['Name', 'Age']]

Filtering & Modifying

- filtered_df = df[df['Age'] > 25]
- df['Salary'] = [50000, 60000, 45000, 70000]
- df.fillna(0, inplace=True)
- df.dropna(inplace=True)

Why Use Pandas?

Ease of Use

Simple and intuitive syntax for data manipulation.

Performance

Optimized for performance, built on top of NumPy.

Versatility

Handles a wide variety of data formats and integrates well with other data analysis libraries.

Rich Functionality

Extensive functionality for data cleaning, manipulation, analysis, and visualization.