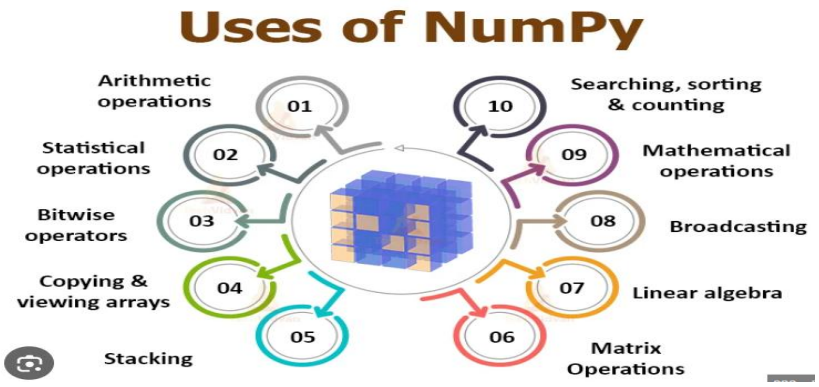
# Python Data Challenge : Class Activity & Homework

This activity is designed to reinforce your understanding of Python basics, NumPy, Pandas, and Matplotlib. The challenge is split into three main parts, each containing interesting and challenging tasks. You are expected to apply Python fundamentals such as variables, control flow, functions, and file handling while working with data.

## Part A : NumPy Challenges



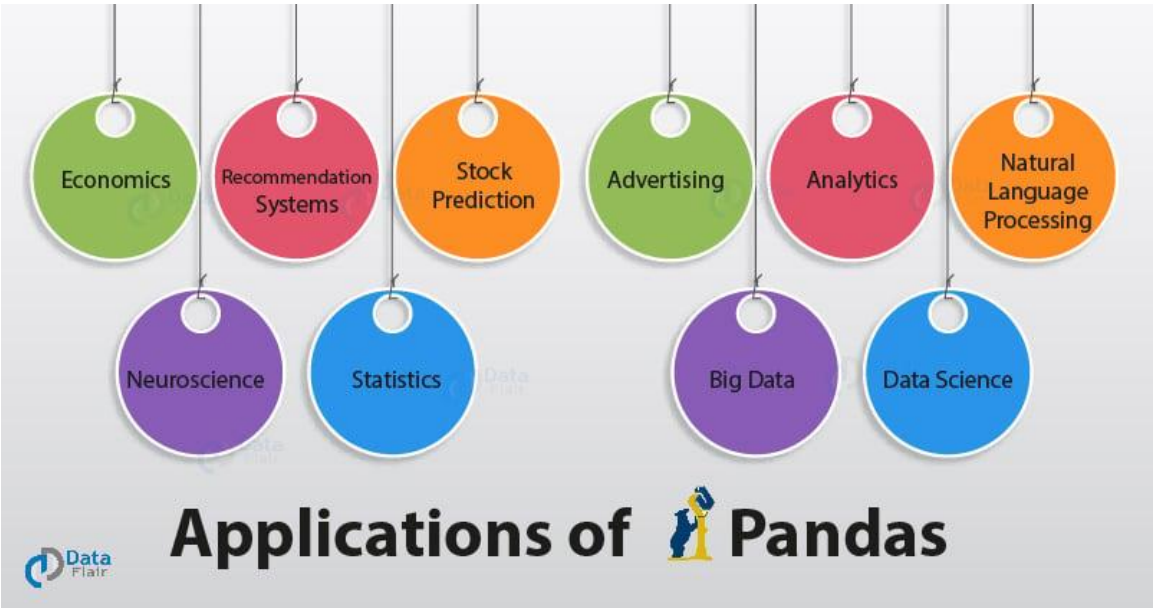

In this section, you will use NumPy arrays to practice numerical operations and array manipulations.

1. Create a NumPy array of 100 random integers between 1 and 100.
   - Find the mean, median, and standard deviation.
   - Sort the array and print the top 5 largest values.

2. Using np.arange, create a 5x5 matrix.
   - Replace the diagonal values with 0.
   - Multiply this matrix by its transpose.

3. Generate 50 random values between -5 and 5.
   - Find how many are positive, negative, and between -1 and 1.

## Part B : Pandas Challenges




Here you will practice working with data using Pandas, including selection, filtering, grouping, and handling missing values.

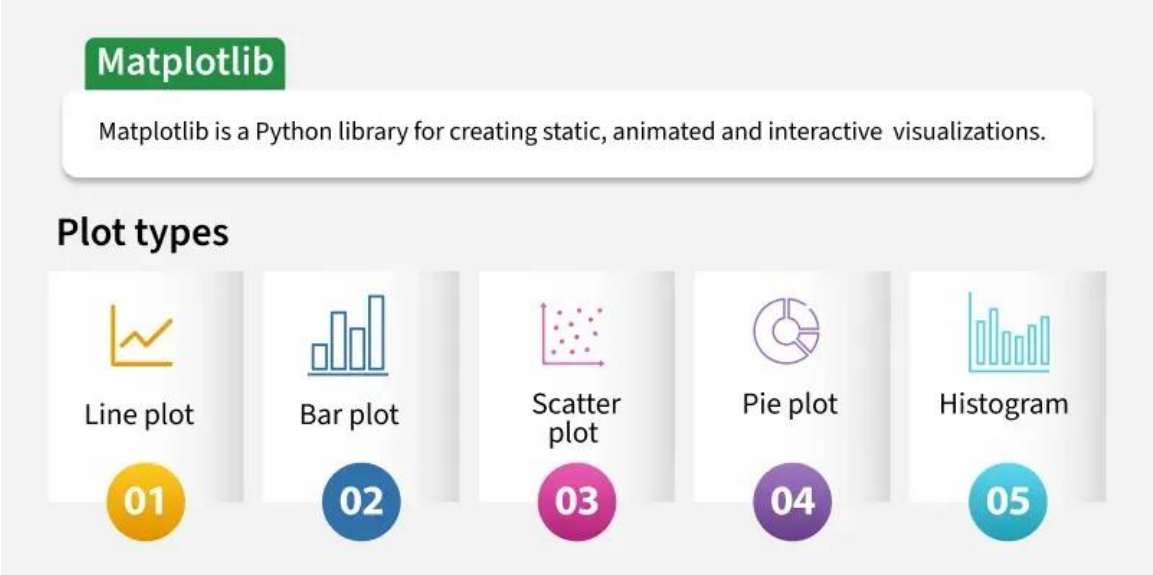
1. Load the Iris dataset as a DataFrame.
   - Show first 10 rows and dataset info.

2. Add a new column petal_area = petal length × petal width.
   - Find the average petal_area for each target class.

3. Filter rows where sepal length > 6 and petal width < 1.5.
   - Show the top 5 of this filtered set.

4. Introduce some missing values (set random 5 rows of sepal length to NaN).
   - Show how many are missing.
   - Fill missing with the column's median.

## Part C : Matplotlib Challenges




This section will test your ability to visualize data using Matplotlib. Make sure to label your plots properly.

1. Plot a histogram of petal length with 20 bins.

2. Create a scatter plot of sepal length vs. petal length.
   - Color points by target (class).

3. Plot a bar chart showing the mean sepal length per target class.

4. Create a boxplot comparing sepal width across target classes.

## Bonus Challenge :

Using Pandas + Matplotlib, create a correlation heatmap of all numeric columns in the dataset.
(Hint: df.corr() + plt.imshow() or use Seaborn if allowed.)

## Integration of Basic Python Concepts

While solving the above tasks, make sure to apply the following Python basics we covered:
- Use variables effectively to store results.
- Write at least one reusable function (e.g., a function that calculates mean, median, std of an array).
- Apply control structures (if/else) where needed (e.g., checking positive/negative values).
- Use loops for iterating when built-in functions are not allowed.
- Write clean code with comments and meaningful variable names.

## Deliverables

1. Submit a Jupyter Notebook with your solutions (code + outputs + explanations).
2. Clearly label each part (A, B, C).
3. Highlight at least one observation/insight from your analysis.
4. Bonus challenge is optional but encouraged.

2. Submit via email : memoona.saleem005@gmail.com

## Homework

For detailed practice of these concepts, you must watch these  tutorial playlist

Playlist Link: https://www.youtube.com/watch?v=ZyhVh-qRZPA&list=PL-osiE80TeTsWmV9i9c58mdDCSskIFdDS

https://www.youtube.com/watch?v=UO98lJQ3QGI&list=PL-osiE80TeTvipOqomVEeZ1HRrcEvtZB_