

## PANDAS:

1- Create a Pandas Data frame from the given data and create a new column “Voter” based on voter age, i.e., if age >18 then voter column should be “Yes” otherwise if age <18 then voter column should be “No”

```
raw_Data = {'Voter_name': ['Geek1', 'Geek2', 'Geek3', 'Geek4',  
                          'Geek5', 'Geek6', 'Geek7', 'Geek8'],  
            'Voter_age': [15, 23, 25, 9, 67, 54, 42, np.NaN]}
```

2 – Create a Pandas Data frame from the given data and collapse First and Last column into one column as Full Name, so the output contains Full Name and Age, then convert column age to index

```
raw_Data = {'First': ['Manan ', 'Raghav ', 'Sunny '],  
            'Last': ['Goel', 'Sharma', 'Chawla'],  
            'Age': [12, 24, 56]}
```

3- Create a Pandas Data frame from the given data -

```
raw_Data = {'Date': ['10/2/2011', '11/2/2011', '12/2/2011', '13/2/2011'],  
            'Product': ['Umbrella', 'Matress', 'Badminton', 'Shuttle'],  
            'Price': [1250, 1450, 1550, 400],  
            'Expense': [21525220.653, 31125840.875, 23135428.768, 56245263.942]}
```

a- Add Index as Item1, Item2, Item3, Item4

b- Find the index labels of all items whose ‘Price’ is greater than 1000.

c- Replace products using Map() with respective codes- Umbrella : ‘U’, Matress : ‘M’, Badminton : ‘B’, Shuttle: ‘S’

d- Round off the Expense column values to two decimal places.

e- Create a new column called ‘Discounted\_Price’ after applying a 10% discount on the existing ‘price’ column.(try using lambda function)

f- Convert the column type of “Date” to datetime format

g- Create a column rank which ranks the products based on the price (one with the highest price will be rank 1).

### Assignment: Exploring NBA Player Data

Download the **nba.csv** file containing NBA player data Complete the following tasks using Python, Pandas, and data visualization libraries:

#### 1. Load Data:

- Load the **nba.csv** data into a Pandas DataFrame.
- Display basic information about the DataFrame.

#### 2. Data Cleaning:

- Handle missing values by either removing or imputing them.
- Remove duplicate rows.

#### 3. Data Transformation:

- Create a new column 'BMI' (Body Mass Index) using the formula:  $BMI = (\text{weight in pounds} / (\text{height in inches})^2) * 703$ . (Assuming a fixed height value of 70 inches (5 feet 10 inches))

#### 4. Exploratory Data Analysis (EDA):

- Display summary statistics of the 'age', 'weight', and 'salary' columns.
- Calculate the average age, weight, and salary of players in each 'position' category.

#### 5. Data Visualization:

- Create a histogram of player ages.
- Create a boxplot of player salaries for each 'position'.
- Plot a scatter plot of 'age' vs. 'salary' with a different color for each 'position'.

#### 6. Top Players:

- Display the top 10 players with the highest salaries.

#### 7. College Analysis:

- Determine the top 5 colleges with the most represented players.

#### 8. Position Distribution:

- Plot a pie chart to show the distribution of players across different 'positions'.

#### 9. Team Analysis:

- Display the average salary of players for each 'team'.
- Plot a bar chart to visualize the average salary of players for each 'team'.

#### 10. Extras

- Get the index at which the minimum weight value is present.
- Sort values based on name in alphabetical order for the rows (the original Dataframe sorting should not change)
- Create a series from given dataframe on “name” column and display top and last 10

#### Guidelines:

1. Write Python code to complete each task.
2. Provide comments explaining your code.
3. Use meaningful variable names.
4. Include necessary library imports.
5. Present your findings in a clear and organized manner.
6. Feel free to use additional code cells for each task.