DATA SCIENCE – UNIVARIATE ANALYSIS

Standard Normal Distribution

```
def stdNBgraph(dataset):
        # Coverted to standard Normal Distribution
    import seaborn as sns
    mean=dataset.mean()
    std=dataset.std()

    values=[i for i in dataset]

    z_score=[((j-mean)/std) for j in values]

    sns.distplot(z_score,kde=True)

    sum(z_score)/len(z_score)
    #z_score.std()
```

stdNBgraph(dataset):

Functions Code Explanation:-

1. **def stdNBgraph(dataset):**
   - **This line defines a function called stdNBgraph. It's like a mini-program that takes a dataset (a collection of data) as input.**

2. import seaborn as sns
    - This line imports the seaborn library, which is a tool for creating statistical graphs.
3. mean = dataset.mean()
    - This calculates the average (mean) of all the numbers in the dataset and stores it in a variable named mean.
4. std = dataset.std()
    - This calculates the standard deviation of the dataset. Standard deviation tells you how spread out the data is. A high standard deviation means the data is very spread out, while a low standard deviation means the data is clustered close to the mean.
5. values = [i for i in dataset]
    - This line creates a list called values that contains all the data points from

the dataset. It's essentially just copying the data into a new list.

6. z_score = [((j - mean) / std) for j in values]
   - This is the important part! It calculates the "z-score" for each value in the dataset.
   - For each value j, it subtracts the mean and then divides by the std. This converts the data to a standard normal distribution (mean=0, standard deviation=1).
   - The z-score tells you how many standard deviations a data point is away from the mean.
7. sns.distplot(z_score, kde=True)
   - This line creates a graph (specifically a histogram with a Kernel Density Estimate) of the calculated z_scores using seaborn. This will visually show

you the distribution of your data after it's been standardized.

8. sum(z_score) / len(z_score)
    - This calculates the average of the z-scores. (It should be very close to zero).

Understanding Z-Scores and Standard Normal Distribution

What this code does:

1. Calculates Mean and Standard Deviation: It finds the average and spread of your data.
2. Calculates Z-Scores: It transforms your data into z-scores, which tell you how many standard deviations each data point is from the mean.

3. **Visualizes Data: It creates a graph of the z-scores to show the distribution of your data after standardization.**
4. **Standard Normal Distribution: The z-scores follow a standard normal distribution, which is a bell-shaped curve with a mean of 0 and a standard deviation of 1.**

Key Concepts:

- **Z-score: A way to standardize data by measuring how many standard deviations a data point is from the mean.**
- **Standard Normal Distribution: A special bell-shaped curve with a mean of 0 and a standard deviation of 1.**

- **Data Transformation: Changing the data to a new scale to make it easier to analyze.**
- **Visualization: Using graphs to understand patterns in data.**
- **Statistics: Calculating mean and standard deviation to summarize data.**

- **This process is useful because it allows you to compare data from different datasets that might have different scales. Z-scores also make it easier to calculate probabilities and identify outliers.**