Assi 3 - Descriptive Statistics.

CSV file | Dataset - ~~loan_data_set.csv~~ Iris_Dataset

Required Libraries:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.datasets import load_iris
```

functions used -

```
df = pd.read_csv("Iris data set.csv")
df.head()
df.shape
df.info()
df.describe()
df.isnull().sum()
plt.show()
mean = grouped_df.mean()
median = grouped_df.median()
min = grouped_df.min()
max = grouped_df.max()
std = grouped_df.std()
df.skew()
```

Do all operations on each column

Also Draw boxplot for each column.

Oral Questions With Answers.

① Which terms include under Measures of central tendency?
→ Mean, Median, Mode.

② Which terms include under measures of variability?
→ Standard deviation, Variance & interquartile range.

③ Describe Iris dataset.
→ 1) What is mean by Iris
2) How many species exist of Iris flower
3) For creating dataset How many species used.
4) List 3 species used for creating dataset.
5) Which ~~fff~~ columns used in Iris dataset.
6) Total no. of rows in dataset.

④ Describe Measures of Central Tendency.
→ It describe the centre of the data.
& often represented by mean, median & mode.

⑤ Explain mean.
→ Mean represents the arithmetic average of the data.

⑥ Explain Median.
→ Median represents the 50th percentile. or the middle value of the data, that separates the distribution into two halves.

(7) Explain Mode-

→ Mode represents the ~~data~~ most frequent value of a variable in the data.

(8) Explain standard deviation.

→ - Standard deviation is a measure that is used to quantify the amount of variation of a set of data values from its mean.
- A low standard deviation for a variable indicates that the data points tend to be close to its mean. & vice versa.

(9) Explain variance.

→ - It is square of the standard deviation & the covariance of the random variable with itself.

(10) Explain Interquartile Range (IQR)-

→ - Measure of statistical dispersion.
- It is calculated as the difference between the upper quartile (75th percentile) & the lower quartile (25th percentile)
- IQR is a very important measure for identifying outliers & could be visualize using boxplot.

(12) Explain term outliers.

→ - outliers are always give wrong direction for your expected results.
- Outliers always talk about extremities.
- Too small or too large.

(13) Explain skewness-

→ - It is used to measure of symmetry or lack of symmetry.
- The skewness value can be positive, negative or undefined.
- In a perfectly symmetrical distribution, the mean, the median & the mode will all have the same value.