Assi 1 - Data Weangling - I

csv file | Dataset - titanic_train.csv

Required Libraries -

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Functions used -

df = pd.read_csv - To load files
df.head() - It's showing top 5 results.
df.tail() - It's showing bottom 5 results
df.isnull().sum() - calculating the null values
sns.heatmap
df.describe() - Get some initial statistics.
df.info() - Getting some info about dataset.
df.dtypes - finding data types.
df.shape - finding Dimensions of the data frame. (no. of rows & no of columns)
df.drop_duplicates()
df.columns

Real Questions with Answers.

① What is mean by Data Preprocessing?

→ ① checking of missing values using pandas isnull() function.
② By using describe() function to get some initial statistics.
③ Provide variable description.

④ Types of Variables
⑤ checking the dimensions of data frame.

② What is mean by Data Weangling ?
→ ① Data exploration - The data is studies, analyzed & understood by visualizing representations of data.
② Dealing with missing values
③ Reshaping data - Data is manipulated according to the requirement. Where new data can be added or pre-existing data canbe modified.
④ filtering data - Some data sets exist with unwanted rows & columns which are required to be removed or filtered.

③ Importance of Data Weangling.
→ - Improve data usability.
- Converts data into compatible format

④ What is mean by data Weangling process?
→ Cleaning, organising & enriching raw data so that it can be used for decision making process

⑤ Uses of pandas libeary.
→ Panda is open source libeary.
Uses -
① Data cleasing ② Data fill
② Data normalization ④ Merges & joins
⑤ Data visualization ⑥ statistical analysis
⑦ Data inspection ⑧ loading & saving data

(6) Uses of numpy libeary.
→ numpy is the numerical Python libeary.
Uses -
(1) Working with arrays
(2) Working in domain of linear algebra.
(3) Fourier transform
(4) Working with matrices.
It is also open source libeary.
(5) Working in vector - vector multiplication.

(7) Uses of matplotlib libeary.
→ matplotlib is used for data visualization & graphical plotting libeary (histograms, scatter plots, bar charts)

(8) Uses of seaborn libeary.
→ seaborn libeary used for making statistical graphics in Python.

(9) Difference between matplotlib & seaborn libeary

matplotlib -
- matplotlib is a Python libeary used to plot various graphs with the help of additional libraries like Numpy & Pandas.
- matplotlib creates simple graphs, including histograms, bargraphs, pie charts, scatter plots, lines & other visual representation of data.
- Mainly used to plot 2D graphs of arrays.
- It uses syntax that relatively complicated & extensive.

eg. matplotlib.pyplot.bar(x-axis, y-axis) is the syntax for bar graph.

Seaborn -

- Seaborn is also a pandas library that utilises Matplotlib, Pandas & Numpy to plot graphs.
- It is a superset of matplotlib library.
- It has relatively simple syntax
  eg. seaborn.barplot (x-axis, y-axis) syntax for a bar graph.
- Seaborn is more comfortable with Pandas data frames.
- Seaborn prevents overlapping with the help of default themes.

(10) How to install any library in python program

pip install package_name

eg pip install seaborn.