

TASK 2

```
In [3]: import pandas as pd
import numpy as np
#READ THE DATSET: Here I ve used Titanic dataset
data = pd.read_csv("uses.csv")
```

```
In [4]: data.head(10)
```

Out[4]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1
5	897	0
6	898	1
7	899	0
8	900	1
9	901	0

```
In [6]: data.tail(10)
```

Out[6]:

	PassengerId	Survived
408	1300	1
409	1301	1
410	1302	1
411	1303	1
412	1304	1
413	1305	0
414	1306	1
415	1307	0
416	1308	0
417	1309	0

```
In [7]: data.dtypes
```

Out[7]:

```
PassengerId    int64
Survived        int64
dtype: object
```

```
In [8]: categorical_columns = data.select_dtypes(include=['object', 'category']).columns.tolist()

# Identify numerical columns
numerical_columns = data.select_dtypes(include=['int64', 'float64']).columns.tolist()
```

```
In [9]: print("Categorical columns:")
print(categorical_columns)

print("Numerical columns:")
print(numerical_columns)
```

Categorical columns:
[]
Numerical columns:
['PassengerId', 'Survived']

```
In [11]: columns_used=['PassengerId', 'Survived'
                        ]
selected_columns = data.loc[:, columns_used]
```

```
In [13]: selected_columns.head(10)
```

Out[13]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1
5	897	0
6	898	1
7	899	0
8	900	1
9	901	0

```
In [14]: column_means = selected_columns.mean()
print("Mean for each column:",column_means )
```

Mean for each column: PassengerId 1100.500000
Survived 0.363636
dtype: float64

```
In [15]: column_modes = selected_columns.mode().iloc[0]
print("Modes for each column:" ,column_modes )
```

Modes for each column: PassengerId 892.0
Survived 0.0
Name: 0, dtype: float64

```
In [16]: column_median = selected_columns.median()
print("Median for each column:" ,column_median )
```

Median for each column: PassengerId 1100.5
Survived 0.0
dtype: float64

```
In [17]: column_std_deviation = selected_columns.std()
print("Standard Deviation for each column:" ,column_std_deviation )
```

Standard Deviation for each column: PassengerId 120.810458
Survived 0.481622
dtype: float64

```
In [18]: selected_columns.describe()
```

Out[18]:

	PassengerId	Survived
count	418.000000	418.000000
mean	1100.500000	0.363636
std	120.810458	0.481622
min	892.000000	0.000000
25%	996.250000	0.000000
50%	1100.500000	0.000000
75%	1204.750000	1.000000
max	1309.000000	1.000000

```
In [22]: data['PassengerId'].unique()
```

Out[22]:

```
array([ 892,  893,  894,  895,  896,  897,  898,  899,  900,  901,  902,
        903,  904,  905,  906,  907,  908,  909,  910,  911,  912,  913,
        914,  915,  916,  917,  918,  919,  920,  921,  922,  923,  924,
        925,  926,  927,  928,  929,  930,  931,  932,  933,  934,  935,
        936,  937,  938,  939,  940,  941,  942,  943,  944,  945,  946,
        947,  948,  949,  950,  951,  952,  953,  954,  955,  956,  957,
        958,  959,  960,  961,  962,  963,  964,  965,  966,  967,  968,
        969,  970,  971,  972,  973,  974,  975,  976,  977,  978,  979,
        980,  981,  982,  983,  984,  985,  986,  987,  988,  989,  990,
        991,  992,  993,  994,  995,  996,  997,  998,  999, 1000, 1001,
       1002, 1003, 1004, 1005, 1006, 1007, 1008, 1009, 1010, 1011, 1012,
       1013, 1014, 1015, 1016, 1017, 1018, 1019, 1020, 1021, 1022, 1023,
       1024, 1025, 1026, 1027, 1028, 1029, 1030, 1031, 1032, 1033, 1034,
       1035, 1036, 1037, 1038, 1039, 1040, 1041, 1042, 1043, 1044, 1045,
       1046, 1047, 1048, 1049, 1050, 1051, 1052, 1053, 1054, 1055, 1056,
       1057, 1058, 1059, 1060, 1061, 1062, 1063, 1064, 1065, 1066, 1067,
       1068, 1069, 1070, 1071, 1072, 1073, 1074, 1075, 1076, 1077, 1078,
       1079, 1080, 1081, 1082, 1083, 1084, 1085, 1086, 1087, 1088, 1089,
       1090, 1091, 1092, 1093, 1094, 1095, 1096, 1097, 1098, 1099, 1100,
       1101, 1102, 1103, 1104, 1105, 1106, 1107, 1108, 1109, 1110, 1111,
       1112, 1113, 1114, 1115, 1116, 1117, 1118, 1119, 1120, 1121, 1122,
       1123, 1124, 1125, 1126, 1127, 1128, 1129, 1130, 1131, 1132, 1133,
       1134, 1135, 1136, 1137, 1138, 1139, 1140, 1141, 1142, 1143, 1144,
       1145, 1146, 1147, 1148, 1149, 1150, 1151, 1152, 1153, 1154, 1155,
       1156, 1157, 1158, 1159, 1160, 1161, 1162, 1163, 1164, 1165, 1166,
       1167, 1168, 1169, 1170, 1171, 1172, 1173, 1174, 1175, 1176, 1177,
       1178, 1179, 1180, 1181, 1182, 1183, 1184, 1185, 1186, 1187, 1188,
       1189, 1190, 1191, 1192, 1193, 1194, 1195, 1196, 1197, 1198, 1199,
       1200, 1201, 1202, 1203, 1204, 1205, 1206, 1207, 1208, 1209, 1210,
       1211, 1212, 1213, 1214, 1215, 1216, 1217, 1218, 1219, 1220, 1221,
       1222, 1223, 1224, 1225, 1226, 1227, 1228, 1229, 1230, 1231, 1232,
       1233, 1234, 1235, 1236, 1237, 1238, 1239, 1240, 1241, 1242, 1243,
       1244, 1245, 1246, 1247, 1248, 1249, 1250, 1251, 1252, 1253, 1254,
       1255, 1256, 1257, 1258, 1259, 1260, 1261, 1262, 1263, 1264, 1265,
       1266, 1267, 1268, 1269, 1270, 1271, 1272, 1273, 1274, 1275, 1276,
       1277, 1278, 1279, 1280, 1281, 1282, 1283, 1284, 1285, 1286, 1287,
       1288, 1289, 1290, 1291, 1292, 1293, 1294, 1295, 1296, 1297, 1298,
       1299, 1300, 1301, 1302, 1303, 1304, 1305, 1306, 1307, 1308, 1309],
      dtype=int64)
```

```
In [19]: data.head(10)
```

Out[19]:

	PassengerId	Survived
0	892	0
1	893	1
2	894	0
3	895	0
4	896	1
5	897	0
6	898	1
7	899	0
8	900	1
9	901	0

```
In [21]: data['Survived'].unique()
```

Out[21]:

```
array([0, 1], dtype=int64)
```