

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

df = pd.read_csv('/content/train.csv')
df1 = pd.read_csv('/content/test.csv')

df

{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 891,\n  \"fields\": [\n    {\n      \"column\": \"PassengerId\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 257,\n        \"min\": 1,\n        \"max\": 891,\n        \"num_unique_values\": 891,\n        \"samples\": [\n          710,\n          440,\n          841\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"Survived\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 0,\n          \"max\": 1,\n          \"num_unique_values\": 2,\n          \"samples\": [\n            1,\n            0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\": \"Pclass\",\n          \"properties\": {\n            \"dtype\": \"number\",\n            \"std\": 0,\n            \"min\": 1,\n            \"max\": 3,\n            \"num_unique_values\": 3,\n            \"samples\": [\n              3,\n              1\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n          },\n          {\n            \"column\": \"Name\",\n            \"properties\": {\n              \"dtype\": \"string\",\n              \"num_unique_values\": 891,\n              \"samples\": [\n                \"Moubarek, Master. Halim Gonios (\\\"William George\\\")\",\n                \"Kvillner, Mr. Johan Henrik Johannesson\"\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\"\n            },\n            {\n              \"column\": \"Sex\",\n              \"properties\": {\n                \"dtype\": \"category\",\n                \"num_unique_values\": 2,\n                \"samples\": [\n                  \"female\",\n                  \"male\"\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\"\n              },\n              {\n                \"column\": \"Age\",\n                \"properties\": {\n                  \"dtype\": \"number\",\n                  \"std\": 14.526497332334042,\n                  \"min\": 0.42,\n                  \"max\": 80.0,\n                  \"num_unique_values\": 88,\n                  \"samples\": [\n                    0.75,\n                    22.0\n                  ],\n                  \"semantic_type\": \"\",\n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"SibSp\",\n                  \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 1,\n                    \"min\": 0,\n                    \"max\": 8,\n                    \"num_unique_values\": 7,\n                    \"samples\": [\n                      1,\n                      0\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"Parch\",\n                    \"properties\": {\n                      \"dtype\": \"number\",\n                    }

```

```

\"std\": 0,\n      \"min\": 0,\n      \"max\": 6,\n\"num_unique_values\": 7,\n      \"samples\": [\n      0,\n1\n      ],\n      \"semantic_type\": \"\",\n\"description\": \"\"\n      }\n      },\n      {\n      \"column\":\n\"Ticket\",\n      \"properties\": {\n      \"dtype\": \"string\",\n\"num_unique_values\": 681,\n      \"samples\": [\n\"11774\",\n      \"248740\"\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\": \"Fare\",\n      \"properties\": {\n      \"dtype\": \"number\",\n      \"std\": 49.6934285971809,\n      \"min\": 0.0,\n      \"max\": 512.3292,\n\"num_unique_values\": 248,\n      \"samples\": [\n11.2417,\n      51.8625\n      ],\n      \"semantic_type\":\n\"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\": \"Cabin\",\n      \"properties\": {\n      \"dtype\":\n\"category\",\n      \"num_unique_values\": 147,\n      \"samples\": [\n      \"D45\",\n      \"B49\"\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\": \"Embarked\",\n      \"properties\":\n      {\n      \"dtype\": \"category\",\n      \"num_unique_values\":\n3,\n      \"samples\": [\n      \"S\",\n      \"C\"\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      }\n      ],\n      \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

df1

```

{"summary": "{\n  \"name\": \"df1\",\n  \"rows\": 418,\n  \"fields\":\n  [\n    {\n      \"column\": \"PassengerId\",\n      \"properties\": {\n      \"dtype\": \"number\",\n      \"std\": 120,\n      \"min\": 892,\n      \"max\": 1309,\n      \"num_unique_values\":\n418,\n      \"samples\": [\n      1213,\n      1216,\n1280\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\":\n\"Pclass\",\n      \"properties\": {\n      \"dtype\": \"number\",\n      \"std\": 0,\n      \"min\": 1,\n      \"max\": 3,\n\"num_unique_values\": 3,\n      \"samples\": [\n      3,\n2,\n      1\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\":\n\"Name\",\n      \"properties\": {\n      \"dtype\": \"string\",\n      \"num_unique_values\": 418,\n      \"samples\": [\n      \"Krekorian, Mr. Neshan\",\n      \"Kreuchen, Miss. Emilie\",\n      \"Canavan, Mr. Patrick\"\n      ],\n      \"semantic_type\":\n\"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\": \"Sex\",\n      \"properties\": {\n      \"dtype\":\n\"category\",\n      \"num_unique_values\": 2,\n      \"samples\":\n[\n      \"female\",\n      \"male\"\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n      }\n      },\n      {\n      \"column\": \"Age\",\n      \"properties\": {\n      \"dtype\": \"number\",\n      \"std\": 14.181209235624422,\n      \"min\": 0.17,\n      \"max\": 76.0,\n      \"num_unique_values\":\n

```

```

79,\n          \"samples\": [\n          10.0,\n          34.5\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"SibSp\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 0,\n          \"max\": 8,\n          \"num_unique_values\": 7,\n          \"samples\": [\n          0,\n          1\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Parch\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 0,\n          \"max\": 9,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          1,\n          6\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Ticket\",\n          \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 363,\n          \"samples\": [\n          \"2673\",\n          \"W./C. 6607\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Fare\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 55.907576179973844,\n          \"min\": 0.0,\n          \"max\": 512.3292,\n          \"num_unique_values\": 169,\n          \"samples\": [\n          41.5792,\n          57.75\n          ],\n          \"semantic_type\":\n          \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Cabin\",\n          \"properties\": {\n          \"dtype\":\n          \"category\",\n          \"num_unique_values\": 76,\n          \"samples\": [\n          \"A21\",\n          \"E45\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Embarked\",\n          \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\":\n          3,\n          \"samples\": [\n          \"Q\",\n          \"S\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        }\n      }\n    ],\n    \"type\": \"dataframe\", \"variable_name\": \"df1\"}

```

```
df.head()
```

```

{"summary": "{\n  \"name\": \"df\",\n  \"rows\": 891,\n  \"fields\": [\n    {\n      \"column\": \"PassengerId\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 257,\n        \"min\": 1,\n        \"max\": 891,\n        \"num_unique_values\": 891,\n        \"samples\": [\n          710,\n          440,\n          841\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"Survived\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n          0,\n          \"min\": 0,\n          \"max\": 1,\n          \"num_unique_values\": 2,\n          \"samples\": [\n          1,\n          0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"Pclass\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 1,\n          \"max\": 3,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          3,\n          3,\n          3\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        }\n      }\n    ]\n  ]\n}

```

```

1\n      ],\n      \"semantic_type\": \"\",\n      \"description\": \"\"\n    },\n    {\n      \"column\":\n      \"Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 891,\n        \"samples\": [\n          \"Moubarek, Master. Halim Gonios (\\\"William George\\\")\",\n          \"Kvillner, Mr. Johan Henrik Johannesson\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      {\n        \"column\": \"Sex\",\n        \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 2,\n          \"samples\": [\n            \"female\",\n            \"male\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        }\n      },\n      {\n        \"column\": \"Age\",\n        \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 14.526497332334042,\n          \"min\": 0.42,\n          \"max\": 80.0,\n          \"num_unique_values\": 88,\n          \"samples\": [\n            0.75,\n            22.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\":\n          \"SibSp\",\n          \"properties\": {\n            \"dtype\": \"number\",\n            \"std\": 1,\n            \"min\": 0,\n            \"max\": 8,\n            \"num_unique_values\": 7,\n            \"samples\": [\n              1,\n              0\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n          },\n          {\n            \"column\":\n            \"Parch\",\n            \"properties\": {\n              \"dtype\": \"number\",\n              \"std\": 0,\n              \"min\": 0,\n              \"max\": 6,\n              \"num_unique_values\": 7,\n              \"samples\": [\n                0,\n                1\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\"\n            },\n            {\n              \"column\":\n              \"Ticket\",\n              \"properties\": {\n                \"dtype\": \"string\",\n                \"num_unique_values\": 681,\n                \"samples\": [\n                  \"11774\",\n                  \"248740\"\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\"\n              },\n              {\n                \"column\": \"Fare\",\n                \"properties\": {\n                  \"dtype\": \"number\",\n                  \"std\": 49.6934285971809,\n                  \"min\": 0.0,\n                  \"max\": 512.3292,\n                  \"num_unique_values\": 248,\n                  \"samples\": [\n                    11.2417,\n                    51.8625\n                  ],\n                  \"semantic_type\":\n                  \"\",\n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"Cabin\",\n                  \"properties\": {\n                    \"dtype\":\n                    \"category\",\n                    \"num_unique_values\": 147,\n                    \"samples\": [\n                      \"D45\",\n                      \"B49\"\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"Embarked\",\n                    \"properties\": {\n                      \"dtype\": \"category\",\n                      \"num_unique_values\":\n                      3,\n                      \"samples\": [\n                        \"S\",\n                        \"C\"\n                      ],\n                      \"semantic_type\": \"\",\n                      \"description\": \"\"\n                    }\n                  }\n                }\n              }\n            }\n          }\n        }\n      ],\n      \"type\": \"dataframe\", \"variable_name\": \"df\"

```

```
df.tail()
```

```

{"summary":{"\n  \"name\": \"df\", \n  \"rows\": 5, \n  \"fields\": [\n    {\n      \"column\": \"PassengerId\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 1, \n        \"min\": 887, \n        \"max\": 891, \n        \"num_unique_values\": 5, \n        \"samples\": [\n          888, \n          891, \n          889\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Survived\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 0, \n        \"min\": 0, \n        \"max\": 1, \n        \"num_unique_values\": 2, \n        \"samples\": [\n          1, \n          0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Pclass\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 1, \n        \"min\": 1, \n        \"max\": 3, \n        \"num_unique_values\": 3, \n        \"samples\": [\n          2, \n          1\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Name\", \n      \"properties\": {\n        \"dtype\": \"string\", \n        \"num_unique_values\": 5, \n        \"samples\": [\n          \"Graham, Miss. Margaret Edith\", \n          \"Doooley, Mr. Patrick\"\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Sex\", \n      \"properties\": {\n        \"dtype\": \"category\", \n        \"num_unique_values\": 2, \n        \"samples\": [\n          \"female\", \n          \"male\"\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Age\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 5.354126134736337, \n        \"min\": 19.0, \n        \"max\": 32.0, \n        \"num_unique_values\": 4, \n        \"samples\": [\n          19.0, \n          32.0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"SibSp\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 0, \n        \"min\": 0, \n        \"max\": 1, \n        \"num_unique_values\": 2, \n        \"samples\": [\n          1, \n          0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Parch\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 0, \n        \"min\": 0, \n        \"max\": 2, \n        \"num_unique_values\": 2, \n        \"samples\": [\n          2, \n          0\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Ticket\", \n      \"properties\": {\n        \"dtype\": \"string\", \n        \"num_unique_values\": 5, \n        \"samples\": [\n          \"112053\", \n          \"370376\"\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }, \n    {\n      \"column\": \"Fare\", \n      \"properties\": {\n        \"dtype\": \"number\", \n        \"std\": 10.09253436952285, \n        \"min\": 7.75, \n        \"max\": 30.0, \n        \"num_unique_values\": 4, \n        \"samples\": [\n          30.0, \n          7.75\n        ], \n        \"semantic_type\": \"\", \n        \"description\": \"\"\n      }\n    }\n  ]\n}

```

```
\\"description\\": \\'\\'\\n      }\\n    },\\n    {\\n      \\\"column\\\":  
\\\"Cabin\\\",\\n      \\\"properties\\\": {\\n        \\\"dtype\\\": \\\"category\\\",\\n  
        \\\"num_unique_values\\\": 2,\\n        \\\"samples\\\": [\\n  
        \\\"C148\\\",\\n          \\\"B42\\\"\\n        ],\\n        \\\"semantic_type\\\":  
        \\\"\\\",\\n        \\\"description\\\": \\'\\'\\'\\n      }\\n    },\\n    {\\n      \\\"column\\\": \\\"Embarked\\\",\\n      \\\"properties\\\": {\\n        \\\"dtype\\\":  
        \\\"string\\\",\\n        \\\"num_unique_values\\\": 3,\\n        \\\"samples\\\":  
        [\\n          \\\"S\\\",\\n            \\\"C\\\"\\n        ],\\n        \\\"semantic_type\\\": \\\"\\\",\\n        \\\"description\\\": \\'\\'\\'\\n      }\\n  
    }\\n  ]\\n}\\", \"type\":\"dataframe\"}
```

```
df.shape
```

(891, 12)

```
df.isnull().sum()
```

PassengerId	0
Survived	0
Pclass	0
Name	0
Sex	0
Age	0
SibSp	0
Parch	0
Ticket	0
Fare	0
Embarked	0
dtype:	int64

```
df.describe()
```

```
{
  "summary": {
    "name": "df",
    "rows": 8,
    "fields": [
      {
        "column": "PassengerId",
        "properties": {
          "dtype": "number",
          "std": null,
          "min": 0.0,
          "max": 0.0,
          "num_unique_values": 1,
          "samples": [
            0.0
          ],
          "semantic_type": "\"\"",
          "description": ""
        },
        "column": "Survived",
        "properties": {
          "dtype": "number",
          "std": null,
          "min": 0.0,
          "max": 0.0,
          "num_unique_values": 1,
          "samples": [
            0.0
          ],
          "semantic_type": "\"\"",
          "description": ""
        },
        "column": "Pclass",
        "properties": {
          "dtype": "number",
          "std": null,
          "min": 0.0,
          "max": 0.0,
          "num_unique_values": 1,
          "samples": [
            0.0
          ],
          "semantic_type": "\"\"",
          "description": ""
        },
        "column": "Sex",
        "properties": {
          "dtype": "number",
          "std": null,
          "min": 0.0,
          "max": 0.0,
          "num_unique_values": 1,
          "samples": [
            0.0
          ],
          "semantic_type": "\"\"",
          "description": ""
        }
      ]
    }
  }
}
```

```

{"semantic_type": "",\n
  },\n
  {\n
    "column": "Age",\n
    "properties": {\n
      "dtype": "number",\n
      "std": null,\n
      "min": 0.0,\n
      "max": 0.0,\n
      "num_unique_values": 1,\n
      "samples": [\n
        0.0\n
      ],\n
      "semantic_type": ""\n
    },\n
    {\n
      "column": "SibSp",\n
      "properties": {\n
        "dtype": "number",\n
        "std": null,\n
        "min": 0.0,\n
        "max": 0.0,\n
        "num_unique_values": 1,\n
        "samples": [\n
        0.0\n
      ],\n
      "semantic_type": ""\n
    },\n
    {\n
      "column": "Parch",\n
      "properties": {\n
        "dtype": "number",\n
        "std": null,\n
        "min": 0.0,\n
        "max": 0.0,\n
        "num_unique_values": 1,\n
        "samples": [\n
        0.0\n
      ],\n
      "semantic_type": ""\n
    },\n
    {\n
      "column": "Fare",\n
      "properties": {\n
        "dtype": "number",\n
        "std": null,\n
        "min": 0.0,\n
        "max": 0.0,\n
        "num_unique_values": 1,\n
        "samples": [\n
        0.0\n
      ],\n
      "semantic_type": ""\n
    },\n
    {\n
      "column": "Embarked",\n
      "properties": {\n
        "dtype": "number",\n
        "std": null,\n
        "min": 0.0,\n
        "max": 0.0,\n
        "num_unique_values": 1,\n
        "samples": [\n
        0.0\n
      ],\n
      "semantic_type": ""\n
    }\n
  ]\n}, "type": "dataframe"}

```

*# replacing missing values with the average in the columns(Data cleaning)*

```

df['Age'].fillna(df['Age'].median(), inplace = True)
df['Fare'].fillna(df['Fare'].mean(), inplace = True)
df['Embarked'].fillna(df['Embarked'].mode(), inplace = True)
df.dropna(inplace = True)
df['Sex'] = df['Sex'].map({'male': 0, 'female': 1})
df['Embarked'] = df['Embarked'].map({'C': 0, 'Q': 1, 'S': 2})

```

*#summary*

```
df.describe()
```

```

{"summary": "{\n  "name": "df",\n  "rows": 8,\n  "fields": [\n    {\n      "column": "PassengerId",\n      "properties": {\n        "dtype": "number",\n        "std": 284.008380046864,\n        "min": 2.0,\n        "max": 890.0,\n        "num_unique_values": 8,\n        "samples": [\n          455.4950495049505,\n          457.5,\n          202.0\n        ],\n        "semantic_type": ""\n      },\n      {\n        "column": "Survived",\n        "properties": {\n          "dtype": "number",\n          "std": 71.21005300204565,\n          "min": 0.0,\n          "max": 202.0,\n          "num_unique_values": 5,\n
```



```

\"samples\": [\n                0.6633663366336634,\n                1.0,\n                0.47373245850820106\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\",\n            \"column\": \"Pclass\",\n            \"properties\": {\n                \"dtype\": \"number\",\n                \"std\": 70.9809043732544,\n                \"min\": 0.5282054850451682,\n                \"max\": 202.0,\n                \"num_unique_values\": 5,\n                \"samples\": [\n                    1.198019801980198,\n                    3.0,\n                    0.5282054850451682\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\",\n                \"column\": \"Sex\",\n                \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 71.26895778854988,\n                    \"min\": 0.0,\n                    \"max\": 202.0,\n                    \"num_unique_values\": 5,\n                    \"samples\": [\n                        0.47029702970297027,\n                        1.0,\n                        0.5003569978173238\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\",\n                    \"column\": \"Age\",\n                    \"properties\": {\n                        \"dtype\": \"number\",\n                        \"std\": 63.90403415867557,\n                        \"min\": 0.92,\n                        \"max\": 202.0,\n                        \"num_unique_values\": 8,\n                        \"samples\": [\n                            34.952574257425745,\n                            33.5,\n                            202.0\n                        ],\n                        \"semantic_type\": \"\",\n                        \"description\": \"\",\n                        \"column\": \"SibSp\",\n                        \"properties\": {\n                            \"dtype\": \"number\",\n                            \"std\": 71.16834722564758,\n                            \"min\": 0.0,\n                            \"max\": 202.0,\n                            \"num_unique_values\": 6,\n                            \"samples\": [\n                                202.0,\n                                0.44554455445544555,\n                                3.0\n                            ],\n                            \"semantic_type\": \"\",\n                            \"description\": \"\",\n                            \"column\": \"Parch\",\n                            \"properties\": {\n                                \"dtype\": \"number\",\n                                \"std\": 71.11834851325345,\n                                \"min\": 0.0,\n                                \"max\": 202.0,\n                                \"num_unique_values\": 6,\n                                \"samples\": [\n                                    202.0,\n                                    0.4405940594059406,\n                                    4.0\n                                ],\n                                \"semantic_type\": \"\",\n                                \"description\": \"\",\n                                \"column\": \"Fare\",\n                                \"properties\": {\n                                    \"dtype\": \"number\",\n                                    \"std\": 165.4724689492812,\n                                    \"min\": 0.0,\n                                    \"max\": 512.3292,\n                                    \"num_unique_values\": 8,\n                                    \"samples\": [\n                                        76.10330099009902,\n                                        55.0,\n                                        202.0\n                                    ],\n                                    \"semantic_type\": \"\",\n                                    \"description\": \"\",\n                                    \"column\": \"Embarked\",\n                                    \"properties\": {\n                                        \"dtype\": \"number\",\n                                        \"std\": 71.00630836403613,\n                                        \"min\": 0.0,\n                                        \"max\": 202.0,\n                                        \"num_unique_values\": 5,\n                                        \"samples\": [\n                                            1.297029702970297,\n                                            2.0,\n                                            0.9467888063815835\n                                        ]\n                                    },\n                                }\n                            }\n                        }\n                    }\n                }\n            ],\n            \"type\": \"dataframe\"}

```

```
# Visualization distributions
```

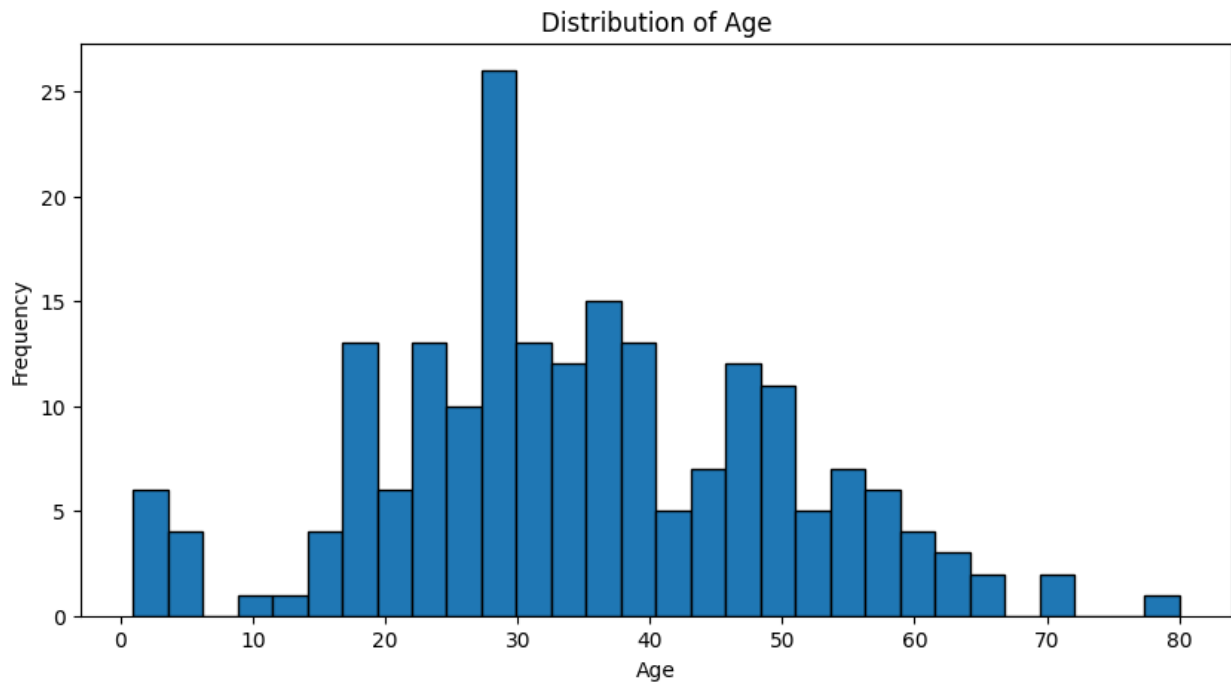
```

plt.figure(figsize=(10,5))
plt.hist(df['Age'] , bins =30, edgecolor = 'black')
plt.xlabel('Age')

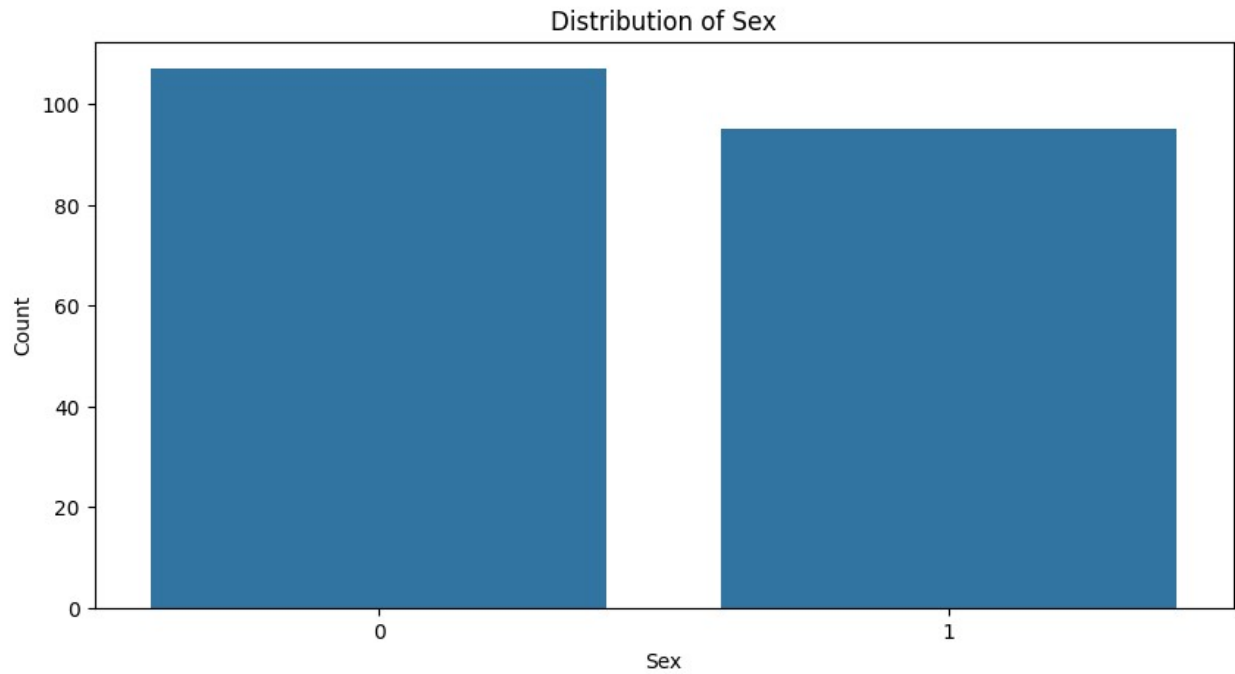
```



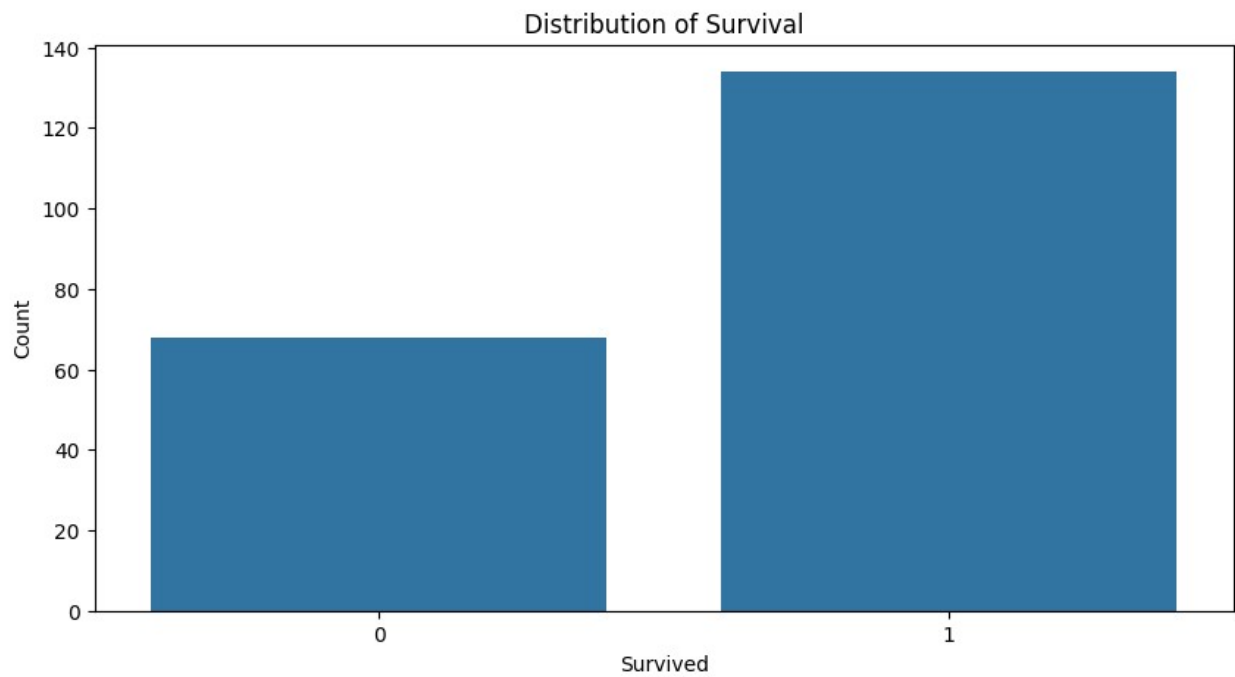
```
plt.ylabel('Frequency')
plt.title('Distribution of Age')
plt.show()
```



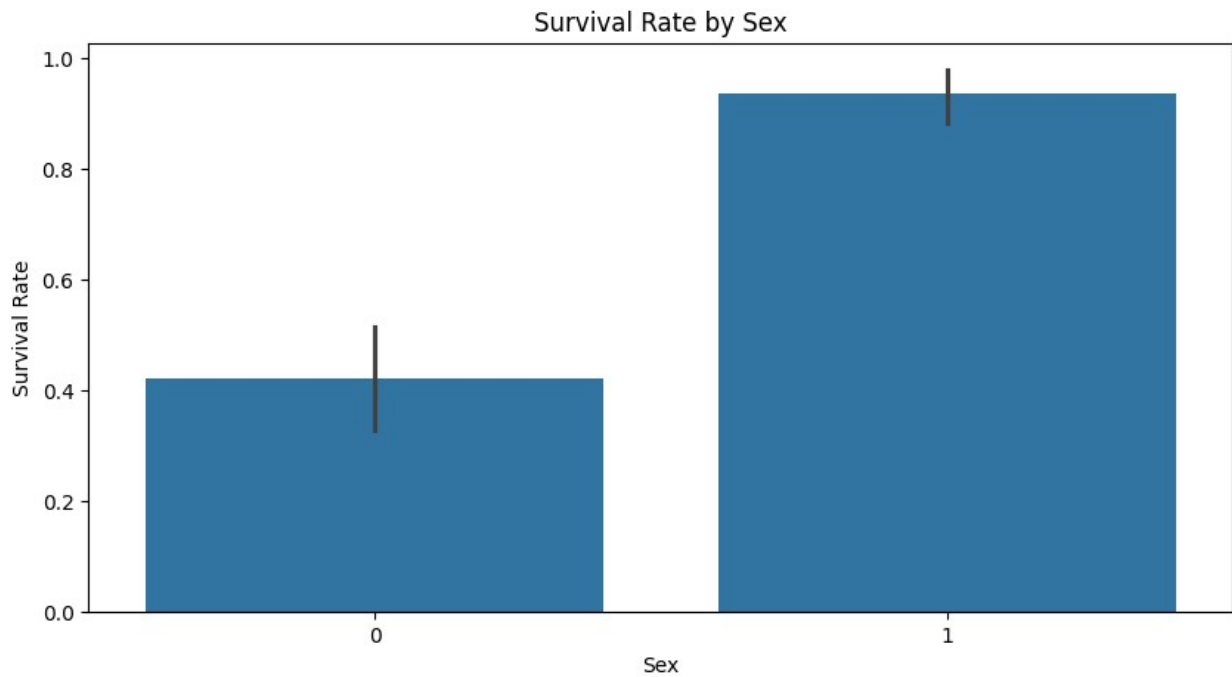
```
plt.figure(figsize=(10,5))
sns.countplot(x='Sex', data=df)
plt.xlabel('Sex')
plt.ylabel('Count')
plt.title('Distribution of Sex')
plt.show()
```



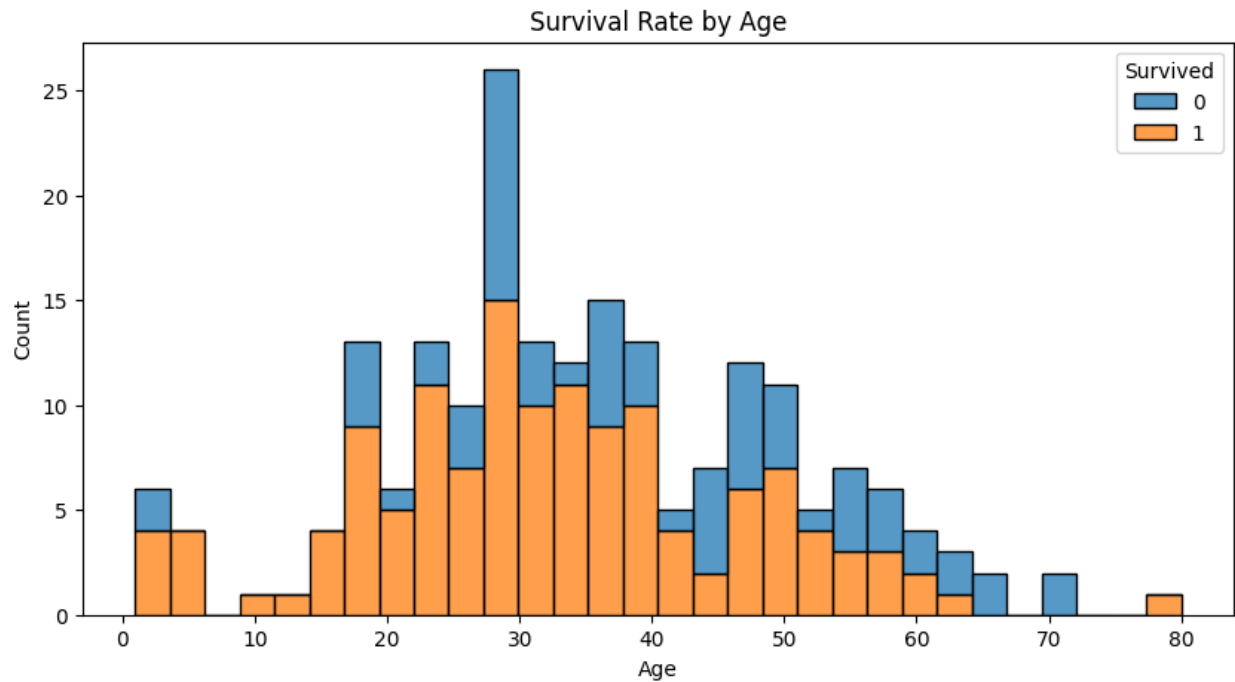
```
plt.figure(figsize=(10,5))  
sns.countplot(x='Survived', data = df)  
plt.xlabel('Survived')  
plt.ylabel('Count')  
plt.title('Distribution of Survival')  
plt.show()
```



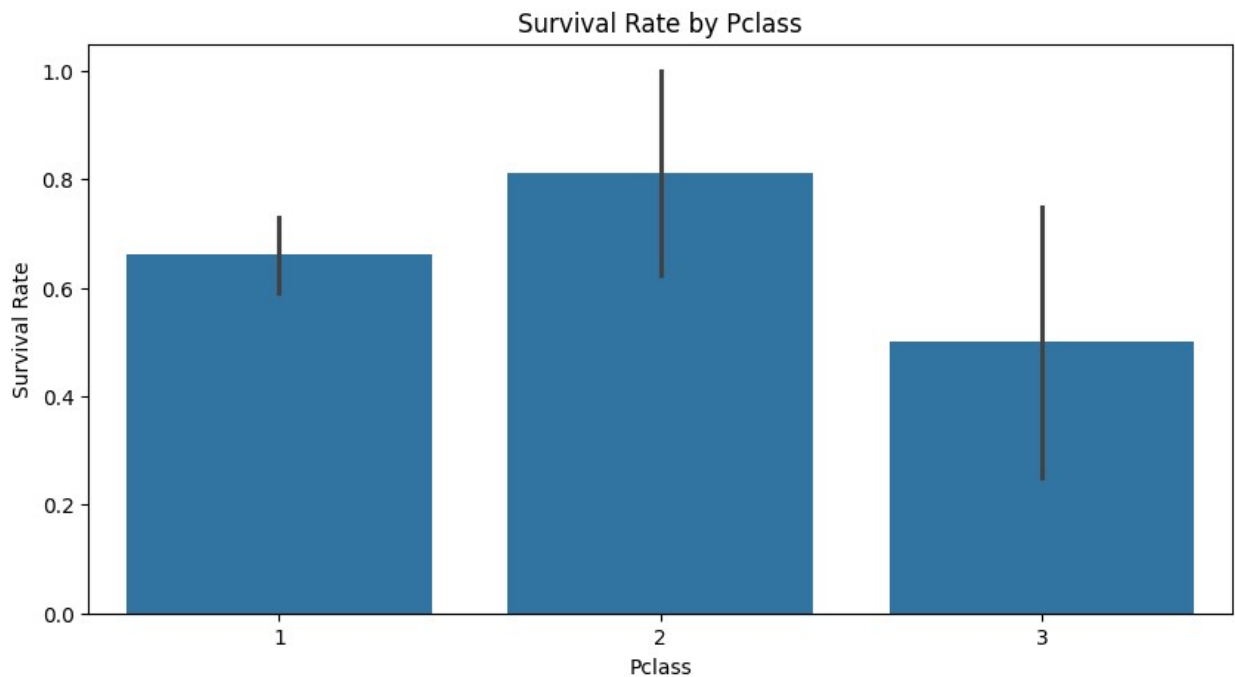
```
# Explore relationships between variables
plt.figure(figsize=(10,5))
sns.barplot(x='Sex', y='Survived', data=df)
plt.xlabel('Sex')
plt.ylabel('Survival Rate')
plt.title('Survival Rate by Sex')
plt.show()
```



```
plt.figure(figsize=(10,5))
sns.histplot(x='Age', hue='Survived', multiple='stack', bins=30,
data=df)
plt.xlabel('Age')
plt.ylabel('Count')
plt.title('Survival Rate by Age')
plt.show()
```



```
plt.figure(figsize=(10,5))
sns.barplot(x='Pclass', y='Survived', data=df)
plt.xlabel('Pclass')
plt.ylabel('Survival Rate')
plt.title('Survival Rate by Pclass')
plt.show()
```



```
#Correlation heatmap
correlation_matrix = df.select_dtypes(include=np.number).corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt='.2f')
plt.title('Correlation Matrix')
plt.show()
```

