

# Assignment 1

```
In [ ]: import pandas as pd
import numpy as np
# from sklearn.impute import SimpleImputer
```

```
In [ ]: df_train = pd.read_csv('./Dataset/train.csv')
df_test = pd.read_csv('./Dataset/test.csv')
```

```
In [ ]: df_train.head(5)
```

```
Out[ ]:
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85	S
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S



```
In [ ]: df_train.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 891 entries, 0 to 890
```

```
Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	PassengerId	891 non-null	int64
1	Survived	891 non-null	int64
2	Pclass	891 non-null	int64
3	Name	891 non-null	object
4	Sex	891 non-null	object
5	Age	714 non-null	float64
6	SibSp	891 non-null	int64
7	Parch	891 non-null	int64
8	Ticket	891 non-null	object
9	Fare	891 non-null	float64
10	Cabin	204 non-null	object
11	Embarked	889 non-null	object

```
dtypes: float64(2), int64(5), object(5)
```

```
memory usage: 83.7+ KB
```

```
In [ ]: df_train.isna().sum()
```

```
Out[ ]: PassengerId      0
Survived              0
Pclass               0
Name                 0
Sex                  0
Age                 177
SibSp                0
Parch                0
Ticket              0
Fare                 0
Cabin               687
Embarked             2
dtype: int64
```

---

## Handling Missing Values

---

### ■ Filling Embraked

```
In [ ]: df_train['Embarked'].unique()
```

```
Out[ ]: array(['S', 'C', 'Q', nan], dtype=object)
```

```
In [ ]: df_train['Embarked'].nunique()
```

```
Out[ ]: 3
```

```
In [ ]: df_train['Embarked'].mode()
```

```
Out[ ]: 0    S
Name: Embarked, dtype: object
```

```
In [ ]: df_train['Embarked'].fillna(df_train['Embarked'].mode(), inplace=True)
```

---

## Visulaization - Heatmap

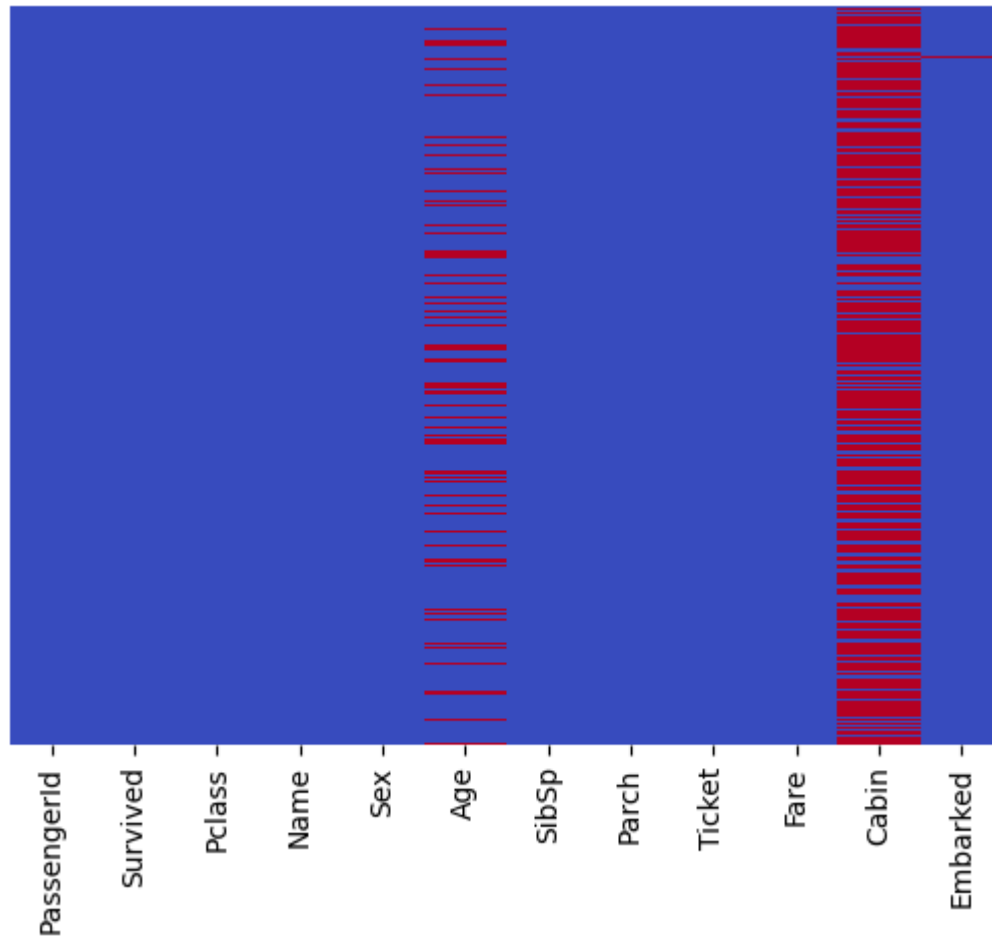
---

```
In [ ]: %pip install seaborn
```

```
In [ ]: import seaborn as sns
```

```
In [ ]: sns.heatmap(df_train.isnull(),yticklabels=False,cbar=False,cmap='coolwarm')
```

```
Out[ ]: <Axes: >
```



---

### Filling Age

---

```
In [ ]: df_train['Age'].median()
```

```
Out[ ]: 28.0
```

```
In [ ]: df_train['Age'].fillna(df_train['Age'].median(),inplace=True)
```

```
In [ ]: df_train.isna().sum()
```

```
Out[ ]: PassengerId    0
Survived           0
Pclass             0
Name               0
Sex                0
Age                0
SibSp              0
Parch              0
Ticket             0
Fare               0
Cabin             687
Embarked           2
dtype: int64
```

```
In [ ]: df_train['Cabin'].unique()
```

```
Out[ ]: array([nan, 'C85', 'C123', 'E46', 'G6', 'C103', 'D56', 'A6',
               'C23 C25 C27', 'B78', 'D33', 'B30', 'C52', 'B28', 'C83', 'F33',
               'F G73', 'E31', 'A5', 'D10 D12', 'D26', 'C110', 'B58 B60', 'E101',
               'F E69', 'D47', 'B86', 'F2', 'C2', 'E33', 'B19', 'A7', 'C49', 'F4',
               'A32', 'B4', 'B80', 'A31', 'D36', 'D15', 'C93', 'C78', 'D35',
               'C87', 'B77', 'E67', 'B94', 'C125', 'C99', 'C118', 'D7', 'A19',
               'B49', 'D', 'C22 C26', 'C106', 'C65', 'E36', 'C54',
               'B57 B59 B63 B66', 'C7', 'E34', 'C32', 'B18', 'C124', 'C91', 'E40',
               'T', 'C128', 'D37', 'B35', 'E50', 'C82', 'B96 B98', 'E10', 'E44',
               'A34', 'C104', 'C111', 'C92', 'E38', 'D21', 'E12', 'E63', 'A14',
               'B37', 'C30', 'D20', 'B79', 'E25', 'D46', 'B73', 'C95', 'B38',
               'B39', 'B22', 'C86', 'C70', 'A16', 'C101', 'C68', 'A10', 'E68',
               'B41', 'A20', 'D19', 'D50', 'D9', 'A23', 'B50', 'A26', 'D48',
               'E58', 'C126', 'B71', 'B51 B53 B55', 'D49', 'B5', 'B20', 'F G63',
               'C62 C64', 'E24', 'C90', 'C45', 'E8', 'B101', 'D45', 'C46', 'D30',
               'E121', 'D11', 'E77', 'F38', 'B3', 'D6', 'B82 B84', 'D17', 'A36',
               'B102', 'B69', 'E49', 'C47', 'D28', 'E17', 'A24', 'C50', 'B42',
               'C148'], dtype=object)
```

---