In [ ]:
```
#Handling rare type of categorical variables
```

In [24]:
```
categorical_features=[feature for feature in dataset.columns if dataset[feature].dtype=='O']
```

In [26]:
```
categorical_features
```

Out[26]:
```
['Name', 'Sex', 'Ticket', 'Cabin', 'Embarked']
```

In [31]:
```
dataset.head(100)
```

Out[31]:

	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	C
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
...	...	...	...	...	...	...	...	...	...	...	...	...
95	96	0	3	Shorney, Mr. Charles Joseph	male	NaN	0	0	374910	8.0500	NaN	S
96	97	0	1	Goldschmidt, Mr. George B	male	71.0	0	0	PC 17754	34.6542	A5	C
97	98	1	1	Greenfield, Mr. William Bertram	male	23.0	0	1	PC 17759	63.3583	D10 D12	C
98	99	1	2	Doling, Mrs. John T (Ada Julia Bone)	female	34.0	0	1	231919	23.0000	NaN	S
99	100	0	2	Kantor, Mr. Sinai	male	34.0	1	0	244367	26.0000	NaN	S

100 rows × 12 columns

In [36]:
```
for feature in categorical_features:
    labels_ordered=dataset.groupby([feature])['Fare'].mean().sort_values().index
    labels_ordered={k:i for i,k in enumerate(labels_ordered,0)}
    dataset[feature]=dataset[feature].map(labels_ordered)
```

In [38]:
```
dataset.head(10)
```

Out[38]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	74	0	22.0	1	0	66	7.2500	NaN	1.0
1	2	1	1	788	1	38.0	1	0	636	71.2833	89.0	2.0
2	3	1	3	223	1	26.0	0	0	235	7.9250	NaN	1.0
3	4	1	1	749	1	35.0	1	0	617	53.1000	73.0	1.0
4	5	0	3	245	0	35.0	0	0	250	8.0500	NaN	1.0
5	6	0	3	292	0	NaN	0	0	287	8.4583	NaN	0.0
6	7	0	1	734	0	54.0	0	0	609	51.8625	66.0	1.0
7	8	0	3	530	0	2.0	3	1	480	21.0750	NaN	1.0
8	9	1	3	364	1	27.0	0	2	356	11.1333	NaN	1.0
9	10	1	2	657	1	14.0	1	0	564	30.0708	NaN	2.0

In [37]:
```
scaling_feature=[feature for feature in dataset.columns if feature not in ['PassengerId','Embarked'] ]
len(scaling_feature)
```

Out[37]: 10

In [20]:
```
scaling_feature
```

Out[20]:
```
['PassengerId',
 'Survived',
 'Pclass',
 'Name',
 'Sex',
 'Age',
 'SibSp',
 'Parch',
 'Ticket',
 'Fare',
 'Cabin',
 'Agenan']
```

In [33]:
```
dataset.head()
```

Out[33]:

	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	C
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 [ ]:
```

```