```
from google.colab import files


uploaded = files.upload()
```

Choose Files train.csv

- **train.csv**(application/vnd.ms-excel) - 61194 bytes, last modified: 6/24/2021 - 100% done

```
Saving train.csv to train.csv


for fn in uploaded.keys():
 print('User uploaded file "{name}" with length {length} bytes'.format(
name=fn, length=len(uploaded[fn])))


     User uploaded file "train.csv" with length 61194 bytes


import pandas as pd
train=pd.read_csv("train.csv")
print(train)


          PassengerId  Survived  Pclass  ...     Fare Cabin  Embarked
     0              1         0       3  ...   7.2500   NaN         S
     1              2         1       1  ...  71.2833   C85         C
     2              3         1       3  ...   7.9250   NaN         S
     3              4         1       1  ...  53.1000  C123         S
     4              5         0       3  ...   8.0500   NaN         S
     ..           ...       ...     ...  ...      ...   ...       ...
     886          887         0       2  ...  13.0000   NaN         S
     887          888         1       1  ...  30.0000   B42         S
     888          889         0       3  ...  23.4500   NaN         S
     889          890         1       1  ...  30.0000  C148         C
     890          891         0       3  ...   7.7500   NaN         Q

     [891 rows x 12 columns]


df=pd.DataFrame(pd.read_csv("train.csv"))
print(df)

          PassengerId  Survived  Pclass  ...     Fare Cabin  Embarked
     0              1         0       3  ...   7.2500   NaN         S
     1              2         1       1  ...  71.2833   C85         C
     2              3         1       3  ...   7.9250   NaN         S
     3              4         1       1  ...  53.1000  C123         S
     4              5         0       3  ...   8.0500   NaN         S
     ..           ...       ...     ...  ...      ...   ...       ...
     886          887         0       2  ...  13.0000   NaN         S
     887          888         1       1  ...  30.0000   B42         S
     888          889         0       3  ...  23.4500   NaN         S
     889          890         1       1  ...  30.0000  C148         C
     890          891         0       3  ...   7.7500   NaN         Q

     [891 rows x 12 columns]


df.head()
```

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599
2	3	1	3	Heikkinen, Miss	female	26.0	0	0	STON/O2.

```
df.index
```

```
RangeIndex(start=0, stop=891, step=1)
```

```
df.columns
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
      dtype='object')
```

```
df.iloc[[0,1,2,3]]
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599

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

```
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
```

```
drop=df.isnull().sum()[df.isnull().sum()>(35/100*df.shape[0])]
drop
```

```
Cabin    687
dtype: int64
```

```
x=df.isnull().sum()
drop=x[x>(35/100*df.shape[0])]
drop
```

```
Cabin    687
dtype: int64
```

```
drop.index
```

```
Index(['Cabin'], dtype='object')
```

```
df.drop(drop.index,axis=1,inplace=True)
df.isnull().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         2
dtype: int64
```

```
df['Embarked'].describe()
```

```
count     889
unique      3
top         S
freq      644
Name: Embarked, dtype: object
```

```
df['Embarked'].fillna('S',inplace=True)
df.isnull().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Embarked         0
dtype: int64
```

```
df.corr()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
PassengerId	1.000000	-0.005007	-0.035144	0.036847	-0.057527	-0.001652	0.01
Survived	-0.005007	1.000000	-0.338481	-0.077221	-0.035322	0.081629	0.25
Pclass	-0.035144	-0.338481	1.000000	-0.369226	0.083081	0.018443	-0.54
Age	0.036847	-0.077221	-0.369226	1.000000	-0.308247	-0.189119	0.09
SibSp	-0.057527	-0.035322	0.083081	-0.308247	1.000000	0.414838	0.15
Parch	-0.001652	0.081629	0.018443	-0.189119	0.414838	1.000000	0.21
Fare	0.012658	0.257307	-0.549500	0.096067	0.159651	0.216225	1.00

```
df['FamilySize']=df['SibSp']+df['Parch']
df.drop(['SibSp','Parch'],axis=1,inplace=True)
df.corr()
```

	PassengerId	Survived	Pclass	Age	Fare	FamilySize
PassengerId	1.000000	-0.005007	-0.035144	0.036847	0.012658	-0.040143
Survived	-0.005007	1.000000	-0.338481	-0.077221	0.257307	0.016639
Pclass	-0.035144	-0.338481	1.000000	-0.369226	-0.549500	0.065997
Age	0.036847	-0.077221	-0.369226	1.000000	0.096067	-0.301914
Fare	0.012658	0.257307	-0.549500	0.096067	1.000000	0.217138
FamilySize	-0.040143	0.016639	0.065997	-0.301914	0.217138	1.000000

```
df['Alone']=[0 if df['FamilySize'][i]>0 else 1 for i in df.index]
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	Ticket	Fare	Embar
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	A/5 21171	7.2500	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	PC 17599	71.2833	

```
df.groupby(['Alone'])['Survived'].mean()
```

```
Alone
0    0.505650
1    0.303538
Name: Survived, dtype: float64
```

```
df[['Alone','Fare']].corr()
```

	Alone	Fare
Alone	1.000000	-0.271832
Fare	-0.271832	1.000000

```
df['Sex']=[0 if df['Sex'][i]=='male' else 1 for i in df.index]
df.groupby(['Sex'])['Survived'].mean()
```

```
Sex
0    0.188908
1    0.742038
Name: Survived, dtype: float64
```

```
df.groupby(['Sex'])['Survived'].mean()
```

```
Sex
0    0.188908
1    0.742038
Name: Survived, dtype: float64
```

```
df.groupby(['Embarked'])['Survived'].mean()
```

```
Embarked
C    0.553571
Q    0.389610
S    0.339009
Name: Survived, dtype: float64
```

CONCLUSION

- Female passengers were prioritized over men.
- People with high class or rich people have higher survival rate than others.The hierarichy might saving the passengers.
- Passengers travelling with their family have higher survivsl rate.
- Passengers who boarded the ship at cherbourg, survived more in proportion than the others

✓ 0s completed at 8:17 PM