

```
import pandas as pd

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

```
df.dtypes

PassengerId      int64
Survived          int64
Pclass            int64
Name              object
Sex              object
Age              float64
SibSp             int64
Parch            int64
Ticket            object
Fare             float64
Cabin            object
Embarked          object
dtype: object
```

df.describe()

	PassengerId	Survived	Pclass	Age	SibSp	Parch	Fare
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204200
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693422
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910460
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329000

```
df.isna().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
```

```
age_mean_value=df['Age'].mean()
df['Age']=df['Age'].fillna(age_mean_value)
```

```
df.drop('Cabin',axis=1,inplace=True)
```

df.head()

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833

```
filtered_age = df[df.Age>40]
filtered_age
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5
15	16	1	2	Hewlett, Mrs. (Mary D Kingcome)	female	55.0	0	0	248706	16.0
33	34	0	2	Wheadon, Mr. Edward H	male	66.0	0	0	C.A. 24579	10.5
35	36	0	1	Holverson, Mr. Alexander Oskar	male	42.0	1	0	113789	52.0

Next steps:

Generate code with filtered_age

View recommended plots

```
sorted_passengers = df.sort_values('Name',ascending=True,kind='heapsort')
```

```
sorted_passengers.head(10)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare
845	846	0	3	Abbing, Mr. Anthony	male	42.0	0	0	C.A. 5547	7.55
746	747	0	3	Abbott, Mr. Rossmore Edward	male	16.0	1	1	C.A. 2673	20.25
279	280	1	3	Abbott, Mrs. Stanton (Rosa Hunt)	female	35.0	1	1	C.A. 2673	20.25
308	309	0	2	Abelson, Mr. Samuel	male	30.0	1	0	P/PP 3381	24.00

Next steps:

Generate code with sorted_passengers

View recommended plots

```
merged_df=pd.merge(df.head(2),df.tail(2),how='outer',indicator=True)
merged_df
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Embarked	_merge
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	S	left_only
1	2	1	1	Cummings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C	left_only
2	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C	right_only
3	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	Q	right_only

Next steps:

Generate code with merged_df

View recommended plots

```
group_df=df.groupby("Name")
```

```
group_df
```

<pandas.core.groupby.generic.DataFrameGroupBy object at 0x7bdfd253b970>

Start coding or generate with AI.

