

```
import pandas as pd

df = pd.read_csv('/titanic.csv')

df.head()
```

	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 Mav Peel)	female	35.0	1	0	113803	53.1000	C123	S

```
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
```

```
df.shape

(891, 12)
```

```
def mean_imputation(dataset, column, mean):
    dataset[column+'_mean'] = dataset[column].fillna(mean)
```

```
mean = df.Age.mean()
```

```
mean

29.69911764705882
```

```
mean_imputation(df, 'Age', mean)
```

```
df.head()
```

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

```
df[['Age', 'Age_mean']].isna()
```

	Age	Age_mean
0	False	False
1	False	False
2	False	False
3	False	False
4	False	False
...
886	False	False
887	False	False
888	True	False
889	False	False
890	False	False

891 rows × 2 columns

```
df.loc[888]

PassengerId      889
Survived          0
Pclass           3
Name      Johnston, Miss. Catherine Helen "Carrie"
Sex              female
Age              NaN
SibSp            1
Parch           2
Ticket      W./C. 6607
Fare           23.45
Cabin           NaN
Embarked         S
Age_mean      29.699118
Name: 888, dtype: object
```

```
def median_imputation(dataset, column, median):
    dataset[column+'_median'] = dataset[column].fillna(median)
```

```
median = df.Age.median()
```

```
median

28.0
```

```
median_imputation(df, 'Age', median)
```

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Age_mean	Age_median
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	22.0	22.0
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C	38.0	38.0
2	3	1	3	Heikkinen, Miss. Laina Futrelle, Mrs.	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	26.0	26.0

df.loc[888]

```
PassengerId      889
Survived          0
Pclass            3
Name      Johnston, Miss. Catherine Helen "Carrie"
Sex              female
Age              NaN
SibSp             1
Parch             2
Ticket      W./C. 6607
Fare           23.45
Cabin           NaN
Embarked         S
Age_mean      29.699118
Age_median     28.0
Name: 888, dtype: object
```

df2 = pd.read_csv('/titanic.csv')

df2.head()

	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

from sklearn.impute import SimpleImputer

impute_mean = SimpleImputer(strategy = 'mean')

impute_mean.fit(df2[['Age']])

▼ SimpleImputer

SimpleImputer()

df2['Age_mean'] = impute_mean.transform(df2[['Age']])

df2.head()

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

df2.loc[888]

```
PassengerId      889
Survived          0
Pclass            3
Name      Johnston, Miss. Catherine Helen "Carrie"
Sex              female
Age              NaN
SibSp             1
```

```

Parch                2
Ticket              W./C. 6607
Fare                23.45
Cabin              NaN
Embarked            S
Age_mean            29.699118
Name: 888, dtype: object

```

```
impute_median = SimpleImputer(strategy = 'median')
```

```
impute_median.fit(df2[['Age']])
```

```

▼ SimpleImputer
SimpleImputer(strategy='median')

```

```
df2[['Age_median']] = impute_median.transform(df2[['Age']])
```

```
df2.loc[888]
```

```

PassengerId          889
Survived              0
Pclass               3
Name      Johnston, Miss. Catherine Helen "Carrie"
Sex                female
Age                NaN
SibSp               1
Parch               2
Ticket              W./C. 6607
Fare                23.45
Cabin              NaN
Embarked            S
Age_mean            29.699118
Age_median          28.0
Name: 888, dtype: object

```

```
from sklearn.model_selection import train_test_split
```

```
X = df2[['Age_mean', 'Pclass']]
```

```
y = df2['Survived']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=100)
```

```
from sklearn.linear_model import LogisticRegression
```

```
model = LogisticRegression()
```

```
model.fit(X_train, y_train)
```

```

▼ LogisticRegression
LogisticRegression()

```

```
model.score(X_train, y_train)
```

```
0.7008426966292135
```

```
X_data = df2[['Age_median', 'Pclass']]
```

```
X_train, X_test, y_train, y_test = train_test_split(X_data, y, test_size=0.2, random_state=100)
```

```
model.fit(X_train, y_train)
```

```

▼ LogisticRegression
LogisticRegression()

```

```
model.score(X_train, y_train)
```

```
0.6980337078651685
```

