

Untitled1

April 12, 2023

```
[1]: import numpy as np
import pandas as pd
```

```
[2]: df = pd.read_csv("titanic_train.csv")
df.head()
```

```
[2]:
```

	passenger_id	pclass	name \
0	1216	3	Smyth, Miss. Julia
1	699	3	Cacic, Mr. Luka
2	1267	3	Van Impe, Mrs. Jean Baptiste (Rosalie Paula Go...
3	449	2	Hocking, Mrs. Elizabeth (Eliza Needs)
4	576	2	Veal, Mr. James

	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body \
0	female	NaN	0	0	335432	7.7333	NaN	Q	13	NaN
1	male	38.0	0	0	315089	8.6625	NaN	S	NaN	NaN
2	female	30.0	1	1	345773	24.1500	NaN	S	NaN	NaN
3	female	54.0	1	3	29105	23.0000	NaN	S	4	NaN
4	male	40.0	0	0	28221	13.0000	NaN	S	NaN	NaN

	home.dest	survived
0	NaN	1
1	Croatia	0
2	NaN	0
3	Cornwall / Akron, OH	1
4	Barre, Co Washington, VT	0

```
[8]: df.isnull().sum()
```

```
[8]:
```

passenger_id	0
pclass	0
name	0
sex	0
age	174
sibsp	0
parch	0
ticket	0

```

fare          1
cabin         659
embarked      1
boat          542
body          777
home.dest     386
survived      0
dtype: int64

```

```
[15]: df["age"] = df["age"].fillna(0)
```

```
[16]: df["fare"] = df["fare"].fillna(0)
```

```
[17]: df["boat"] = df["boat"].fillna(0)
```

```
[18]: df["body"] = df["body"].fillna(0)
```

```
[19]: df.isnull().sum()
```

```

[19]: passenger_id    0
      pclass          0
      name           0
      sex            0
      age            0
      sibsp          0
      parch          0
      ticket          0
      fare           0
      cabin         659
      embarked       1
      boat           0
      body           0
      home.dest      386
      survived       0
      dtype: int64

```

```
[20]: df = df.drop_duplicates()
```

```
[22]: df
```

```

[22]:   passenger_id  pclass                                name \
0          1216        3                                Smyth, Miss. Julia
1           699        3                                Cacic, Mr. Luka
2          1267        3  Van Impe, Mrs. Jean Baptiste (Rosalie Paula Go...
3           449        2          Hocking, Mrs. Elizabeth (Eliza Needs)
4           576        2                                Veal, Mr. James
..          ...        ...                                ...

```

845	158	1	Hipkins, Mr. William Edward
846	174	1	Kent, Mr. Edward Austin
847	467	2	Kantor, Mrs. Sinai (Miriam Sternin)
848	1112	3	Peacock, Miss. Treasteall
849	425	2	Greenberg, Mr. Samuel

	sex	age	sibsp	parch		ticket	fare	cabin	embarked	\
0	female	0.0	0	0		335432	7.7333	NaN	Q	
1	male	38.0	0	0		315089	8.6625	NaN	S	
2	female	30.0	1	1		345773	24.1500	NaN	S	
3	female	54.0	1	3		29105	23.0000	NaN	S	
4	male	40.0	0	0		28221	13.0000	NaN	S	
..
845	male	55.0	0	0		680	50.0000	C39	S	
846	male	58.0	0	0		11771	29.7000	B37	C	
847	female	24.0	1	0		244367	26.0000	NaN	S	
848	female	3.0	1	1	SOTON/O.Q.	3101315	13.7750	NaN	S	
849	male	52.0	0	0		250647	13.0000	NaN	S	

	boat	body	home.dest	survived
0	13	0.0	NaN	1
1	0	0.0	Croatia	0
2	0	0.0	NaN	0
3	4	0.0	Cornwall / Akron, OH	1
4	0	0.0	Barre, Co Washington, VT	0
..
845	0	0.0	London / Birmingham	0
846	0	258.0	Buffalo, NY	0
847	12	0.0	Moscow / Bronx, NY	1
848	0	0.0	NaN	0
849	0	19.0	Bronx, NY	0

[850 rows x 15 columns]

```
[23]: # There were 853 rows first
```

```
[24]: df.head()
```

```
[24]:
```

	passenger_id	pclass		name	\
0	1216	3		Smyth, Miss. Julia	
1	699	3		Cacic, Mr. Luka	
2	1267	3	Van Impe, Mrs. Jean Baptiste (Rosalie Paula Go...		
3	449	2	Hocking, Mrs. Elizabeth (Eliza Needs)		
4	576	2		Veal, Mr. James	

	sex	age	sibsp	parch	ticket	fare	cabin	embarked	boat	body	\
0	female	0.0	0	0	335432	7.7333	NaN	Q	13	0.0	

1	male	38.0	0	0	315089	8.6625	NaN	S	0	0.0
2	female	30.0	1	1	345773	24.1500	NaN	S	0	0.0
3	female	54.0	1	3	29105	23.0000	NaN	S	4	0.0
4	male	40.0	0	0	28221	13.0000	NaN	S	0	0.0

	home.dest	survived
0	NaN	1
1	Croatia	0
2	NaN	0
3	Cornwall / Akron, OH	1
4	Barre, Co Washington, VT	0

```
[25]: # Removing outliers
```

```
[26]: df = df.drop("passenger_id",axis=1)
```

```
[27]: df = df.drop("cabin",axis=1)
```

```
[30]: df = df.drop("home.dest",axis=1)
```

```
-----
KeyError                                Traceback (most recent call last)
/tmp/ipykernel_14947/3063006230.py in <module>
----> 1 df = df.drop("home.dest",axis=1)

~/anaconda3/lib/python3.9/site-packages/pandas/util/_decorators.py in _
↳ wrapper(*args, **kwargs)
    309         stacklevel=stacklevel,
    310     )
--> 311     return func(*args, **kwargs)
    312
    313     return wrapper

~/anaconda3/lib/python3.9/site-packages/pandas/core/frame.py in drop(self,
↳ labels, axis, index, columns, level, inplace, errors)
    4955         weight 1.0      0.8
    4956         """
-> 4957         return super().drop(

    4958             labels=labels,
    4959             axis=axis,

~/anaconda3/lib/python3.9/site-packages/pandas/core/generic.py in drop(self,
↳ labels, axis, index, columns, level, inplace, errors)
    4265         for axis, labels in axes.items():
    4266             if labels is not None:
-> 4267                 obj = obj._drop_axis(labels, axis, level=level,
↳ errors=errors)
```

```

4268
4269         if inplace:

~/anaconda3/lib/python3.9/site-packages/pandas/core/generic.py in
↳ _drop_axis(self, labels, axis, level, errors, consolidate, only_slice)
4309             new_axis = axis.drop(labels, level=level, errors=errors)
4310         else:
-> 4311             new_axis = axis.drop(labels, errors=errors)
4312             indexer = axis.get_indexer(new_axis)
4313

~/anaconda3/lib/python3.9/site-packages/pandas/core/indexes/base.py in
↳ drop(self, labels, errors)
6659         if mask.any():
6660             if errors != "ignore":
-> 6661                 raise KeyError(f"{list(labels[mask])} not found in axis ")
6662             indexer = indexer[~mask]
6663             return self.delete(indexer)

KeyError: "['home.dest'] not found in axis"

```

```
[31]: df = df.drop("name",axis = 1)
```

```
[32]: df.head()
```

```
[32]:
```

	pclass	sex	age	sibsp	parch	ticket	fare	embarked	boat	body	\
0	3	female	0.0	0	0	335432	7.7333	Q	13	0.0	
1	3	male	38.0	0	0	315089	8.6625	S	0	0.0	
2	3	female	30.0	1	1	345773	24.1500	S	0	0.0	
3	2	female	54.0	1	3	29105	23.0000	S	4	0.0	
4	2	male	40.0	0	0	28221	13.0000	S	0	0.0	

	survived
0	1
1	0
2	0
3	1
4	0

```
[ ]:
```