# onhrbunbj

### April 29, 2024

```
[]: import numpy as np
     import pandas as pd
     import seaborn as sns
     import matplotlib.pyplot as plt
     from matplotlib import style
[ ]: test_df = pd.read_csv("test.csv")
     train_df = pd.read_csv("train.csv")
[]: test_df.info
[]: <bound method DataFrame.info of
                                             PassengerId Pclass
     Name
     0
                   892
                              3
                                                               Kelly, Mr. James
                   893
                              3
     1
                                              Wilkes, Mrs. James (Ellen Needs)
     2
                   894
                              2
                                                     Myles, Mr. Thomas Francis
     3
                   895
                              3
                                                               Wirz, Mr. Albert
     4
                   896
                                 Hirvonen, Mrs. Alexander (Helga E Lindqvist)
                              3
     413
                  1305
                                                             Spector, Mr. Woolf
     414
                  1306
                              1
                                                  Oliva y Ocana, Dona. Fermina
     415
                              3
                                                  Saether, Mr. Simon Sivertsen
                  1307
     416
                  1308
                              3
                                                            Ware, Mr. Frederick
     417
                  1309
                              3
                                                       Peter, Master. Michael J
                         SibSp
             Sex
                    Age
                                 Parch
                                                     Ticket
                                                                  Fare Cabin Embarked
     0
            male
                   34.5
                              0
                                     0
                                                     330911
                                                                7.8292
                                                                          NaN
          female
                   47.0
                                                                7.0000
                                                                                      S
     1
                              1
                                     0
                                                     363272
                                                                          NaN
     2
            male 62.0
                              0
                                     0
                                                     240276
                                                                9.6875
                                                                          NaN
                                                                                      Q
                                                                                      S
     3
            male
                   27.0
                              0
                                     0
                                                     315154
                                                                8.6625
                                                                          NaN
     4
                   22.0
                              1
                                                               12.2875
                                                                                      S
          female
                                     1
                                                    3101298
                                                                          NaN
                                                                                      S
                              0
                                     0
                                                  A.5. 3236
                                                                8.0500
     413
            male
                    NaN
                                                                          NaN
                                                                                      С
     414
          female
                   39.0
                              0
                                     0
                                                   PC 17758
                                                              108.9000
                                                                         C105
                              0
                                     0
                                        SOTON/O.Q. 3101262
                                                                                      S
     415
            male
                   38.5
                                                                7.2500
                                                                          NaN
                                                                                      S
     416
            male
                    NaN
                              0
                                     0
                                                     359309
                                                                8.0500
                                                                          NaN
     417
            male
                    NaN
                              1
                                     1
                                                        2668
                                                               22.3583
                                                                          NaN
                                                                                      С
```

#### []: train\_df.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 891 entries, 0 to 890 Data columns (total 12 columns): Column Non-Null Count Dtype 0 PassengerId 891 non-null int64 Survived 1 891 non-null int64 2 Pclass 891 non-null int64 3 Name 891 non-null object 4 Sex 891 non-null object 5 714 non-null float64 Age 6 891 non-null int64 SibSp 7 Parch 891 non-null int64 Ticket 891 non-null object float64 9 Fare 891 non-null 10 Cabin 204 non-null object 11 Embarked 889 non-null object dtypes: float64(2), int64(5), object(5) memory usage: 83.7+ KB []: total = train\_df.isnull().sum() total []: 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 []: total = train\_df.isnull().sum().sort\_values(ascending=False) percent\_1 = train\_df.isnull().sum()/train\_df.isnull().count()\*100 percent\_2 = (round(percent\_1, 1)).sort\_values(ascending=False) #rounds the\_ →values in percent\_1 to one decimal place and sorts them in descending order.

missing\_data = pd.concat([total, percent\_2], axis=1, keys=['Total', '%'])

#### missing\_data []: Total % 687 77.1 Cabin 19.9 Age 177 Embarked 0.2 2 PassengerId 0 0.0 Survived 0 0.0 Pclass 0.0 0 Name 0 0.0 Sex 0.0 0 0.0 SibSp Parch 0.0 0 Ticket 0 0.0 Fare 0 0.0 []: train\_df.duplicated() []: 0 False 1 False 2 False False 3 4 False 886 False 887 False 888 False 889 False 890 False Length: 891, dtype: bool []: train\_df.drop\_duplicates(inplace=True) train\_df PassengerId Survived Pclass \ []: 0 0 3 1 1 2 1 1 3 2 1 3 3 4 1 1 4 5 0 3 2 886 887 0 887 888 1 1 3 888 889 0

889

890

890

891

1

0

1

3

```
Name
                                                                Sex
                                                                       Age
                                                                            SibSp \
0
                                  Braund, Mr. Owen Harris
                                                               male
                                                                      22.0
                                                                                 1
1
     Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0
                                                                               1
2
                                   Heikkinen, Miss. Laina
                                                             female
                                                                      26.0
                                                                                 0
3
           Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                                             female
                                                                      35.0
                                                                                 1
4
                                Allen, Mr. William Henry
                                                               male
                                                                      35.0
                                                                                 0
                                                                        ...
                                    Montvila, Rev. Juozas
                                                                                 0
886
                                                               male
                                                                      27.0
887
                            Graham, Miss. Margaret Edith
                                                                      19.0
                                                             female
                                                                                 0
888
               Johnston, Miss. Catherine Helen "Carrie"
                                                             female
                                                                                 1
                                                                       NaN
889
                                    Behr, Mr. Karl Howell
                                                                      26.0
                                                                                 0
                                                               male
890
                                      Dooley, Mr. Patrick
                                                               male
                                                                      32.0
                                                                                 0
     Parch
                        Ticket
                                    Fare Cabin Embarked
0
         0
                    A/5 21171
                                  7.2500
                                           NaN
                                                       S
         0
                                           C85
                                                       С
1
                     PC 17599
                                71.2833
2
                                                       S
         0
             STON/02. 3101282
                                  7.9250
                                           NaN
3
         0
                        113803
                                53.1000
                                          C123
                                                       S
                                                       S
4
         0
                                  8.0500
                        373450
                                           NaN
                                            •••
. .
886
         0
                        211536
                                13.0000
                                           {\tt NaN}
                                                       S
887
         0
                        112053
                                30.0000
                                           B42
                                                       S
888
         2
                   W./C. 6607
                                23.4500
                                           NaN
                                                       S
                                                       С
889
                                30.0000
                                          C148
         0
                        111369
890
         0
                        370376
                                  7.7500
                                           NaN
                                                       Q
```

[891 rows x 12 columns]

### []: train\_df.isnull()

[]:		Passen	gerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	\
	0		False	False	False	False	False	False	False	False	False	
	1		False	False	False	False	False	False	False	False	False	
	2		False	False	False	False	False	False	False	False	False	
	3		False	False	False	False	False	False	False	False	False	
	4		False	False	False	False	False	False	False	False	False	
				•••		•••		•••	•••			
	886		False	False	False	False	False	False	False	False	False	
	887		False	False	False	False	False	False	False	False	False	
	888		False	False	False	False	False	True	False	False	False	
	889		False	False	False	False	False	False	False	False	False	
	890		False	False	False	False	False	False	False	False	False	
		Fare	Cabin	Embarked	_							
	0	False	True	False								
	1	False	False	False								
	2	False	True	False								

```
False False
3
                     False
4
    False
            True
                     False
     •••
886 False
                     False
            True
                     False
887 False False
888 False
                     False
            True
889 False False
                     False
890 False
            True
                     False
```

[891 rows x 12 columns]

# []: train\_df.dropna()

[]:		PassengerId	Survived	Pclass	١
	1	2	1	1	
	3	4	1	1	
	6	7	0	1	
	10	11	1	3	
	11	12	1	1	
		***	•••	•••	
	871	872	1	1	
	872	873	0	1	
	879	880	1	1	
	887	888	1	1	
	889	890	1	1	

	Name	Sex	Age	SibSp	\
1	Cumings, Mrs. John Bradley (Florence Briggs Th f	emale 3	8.0	1	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
6	McCarthy, Mr. Timothy J	male	54.0	0	
10	Sandstrom, Miss. Marguerite Rut	female	4.0	1	
11	Bonnell, Miss. Elizabeth	female	58.0	0	
			•••		
871	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	
872	Carlsson, Mr. Frans Olof	male	33.0	0	
879	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	
887	Graham, Miss. Margaret Edith	female	19.0	0	
889	Behr, Mr. Karl Howell	male	26.0	0	

	Parch	Ticket	Fare		Cabin	Embarked
1	0	PC 17599	71.2833		C85	C
3	0	113803	53.1000		C123	S
6	0	17463	51.8625		E46	S
10	1	PP 9549	16.7000		G6	S
11	0	113783	26.5500		C103	S
	•••	•••		•••		
871	1	11751	52.5542		D35	S

872	0	695	5.0000	B51 B53 B55	S
879	1	11767	83.1583	C50	C
887	0	112053	30.0000	B42	S
889	0	111369	30.0000	C148	C

[183 rows x 12 columns]

## []: train\_df.fillna(0)

[]:		Passen	gerId	Survived		\						
	0		1	0	3							
	1		2	1	1							
	2		3	1	3							
	3		4	1	1							
	4		5	0	3							
	• •		•••		•••							
	886		887	0	2							
	887		888	1	1							
	888		889	0	3							
	889		890	1	1							
	890		891	0	3							
									_			
	_				_	_		Name	Sex	Age	SibSp	\
	0							Owen Harris	male	22.0	. 1	
	1	Cuming	s, Mrs	. John Bra	•			riggs Th			1	
	2							Miss. Laina		26.0	0	
	3	F	utrell	e, Mrs. Ja	_			y May Peel)	female	35.0	1	
	4				Alle	n, M	r. Wil	lliam Henry	male	35.0	0	
	• •							•••				
	886							Rev. Juozas	male	27.0	0	
	887							garet Edith	female	19.0	0	
	888		Johi	nston, Mis				en "Carrie"	female	0.0	1	
	889				В			Karl Howell	male	26.0	0	
	890					Doo	ley, N	Mr. Patrick	male	32.0	0	
		ъ.,		m· 1			a 1 ·					
	^	Parch		Ticke				Embarked				
	0	0		A/5 2117			0	S				
	1	0	amos: //	PC 1759			C85	C				
	2	0	STUN/	02. 310128			0	S				
	3	0		11380			C123	S				
	4	0		37345	0 8.0	500	0	S				
				 01150				G				
	886	0		21153			0	S				
	887	0		11205			B42	S				
	888	2		W./C. 660			0	S				
	889	0		11136			C148	C				
	890	0		37037	6 7.7	500	0	Q				

## []: train\_df.dropna()

[]:		Passen	gerId	Survived	Pclass	\					
	1		2	1	1						
	3		4	1	1						
	6		7	0	1						
	10		11	1	3						
	11		12	1	1						
			•••	•••							
	871		872	1	1						
	872		873	0	1						
	879		880	1	1						
	887		888	1	1						
	889		890	1	1						
								_		~	,
		a .		7.1 D	(5		Name	Sex	Age	SibSp	\
	1								8.0	1	
	3	F	utrelle	, Mrs. Ja	_		ly May Peel)	female	35.0	1	
	6			~		•	r. Timothy J	male	54.0	0	
	10			Sand			rguerite Rut	female	4.0	1	
	11				Bonn	ell, Mis	s. Elizabeth	female	58.0	0	
		ъ.		D: 1		1 (0.11					
	871	Beckw	ith, Mr	s. Richar			ie Monypeny)	female	47.0	1	
	872	_					. Frans Olof	male	33.0	0	
	879	Ро	tter, M			•	enia Wilson)	female	56.0	0	
	887			(			rgaret Edith	female	19.0	0	
	889				В	enr, Mr.	Karl Howell	male	26.0	0	
		Parch	Tick	et Fa	ire	Cabin	Embarked				
	1	0	PC 175			C85	С				
	3	0	1138			C123	S				
	6	0	174			E46	S				
	10	1	PP 95			G6	S				
	11	0	1137	83 26.55	500	C103	S				
		•••	•••	•••		•••					
	871	1	117	51 52.55	542	D35	S				
	872	0	6	95 5.00	000 B51	B53 B55	S				
	879	1	117			C50	С				
	887	0	1120	53 30.00	000	B42	S				
	889	0	1113	69 30.00	000	C148	C				

[183 rows x 12 columns]

```
train_df
[]:
                         Survived
                                    Pclass
                                            \
           PassengerId
     0
                      1
                                 0
                                         3
     1
                      2
                                 1
                                          1
     2
                      3
                                 1
                                          3
     3
                      4
                                 1
                                          1
     4
                      5
                                 0
                                          3
     . .
                                          2
                                 0
     886
                   887
     887
                   888
                                 1
                                          1
                                          3
     888
                   889
                                 0
     889
                   890
                                 1
                                         1
     890
                   891
                                          3
                                                            Name
                                                                      Sex
                                                                                 SibSp
                                                                             Age
                                                                                         \
     0
                                       Braund, Mr. Owen Harris
                                                                            22.0
                                                                     male
                                                                                       1
     1
           Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0
                                                                                     1
     2
                                        Heikkinen, Miss. Laina
                                                                                       0
                                                                   female
                                                                            26.0
     3
                Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                                                   female
                                                                            35.0
                                                                                       1
     4
                                      Allen, Mr. William Henry
                                                                     male
                                                                            35.0
                                                                                       0
                                         Montvila, Rev. Juozas
                                                                                       0
     886
                                                                     male
                                                                            27.0
     887
                                  Graham, Miss. Margaret Edith
                                                                            19.0
                                                                                       0
                                                                   female
     888
                     Johnston, Miss. Catherine Helen "Carrie"
                                                                   female
                                                                             0.0
                                                                                       1
                                         Behr, Mr. Karl Howell
     889
                                                                            26.0
                                                                                       0
                                                                     male
     890
                                            Dooley, Mr. Patrick
                                                                     male
                                                                            32.0
                                                                                       0
          Parch
                             Ticket
                                         Fare Cabin Embarked
     0
                          A/5 21171
                                       7.2500
                                                    0
               0
                                                             S
     1
               0
                           PC 17599
                                      71.2833
                                                 C85
                                                             C
     2
                  STON/02. 3101282
                                       7.9250
                                                    0
                                                             S
     3
               0
                                      53.1000
                                                             S
                             113803
                                                C123
               0
                                                             S
     4
                             373450
                                       8.0500
                                                    0
     . .
                              •••
                                           •••
     886
               0
                             211536
                                      13.0000
                                                   0
                                                             S
     887
                             112053
                                      30.0000
                                                 B42
                                                             S
               0
               2
                                                             S
     888
                         W./C. 6607
                                      23.4500
                                                    0
                                                             С
     889
               0
                             111369
                                      30.0000
                                                C148
```

[891 rows x 12 columns]

0

370376

7.7500

[]: train\_df.fillna(0,inplace=True)

```
[]: train_df.head(8)
```

890

0

Q

```
[]:
        PassengerId Survived Pclass
     0
                   1
                              0
                                       3
                   2
     1
                              1
                                       1
     2
                   3
                              1
                                       3
                   4
     3
                              1
                                       1
     4
                   5
                              0
                                       3
     5
                   6
                              0
                                       3
                   7
     6
                              0
                                       1
     7
                   8
                              0
                                       3
                                                         Name
                                                                   Sex
                                                                         Age
                                                                               SibSp \
     0
                                    Braund, Mr. Owen Harris
                                                                  male
                                                                        22.0
                                                                                   1
        Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0
                                                                                 1
     1
     2
                                      Heikkinen, Miss. Laina
                                                                female
                                                                        26.0
                                                                                   0
     3
             Futrelle, Mrs. Jacques Heath (Lily May Peel)
                                                                        35.0
                                                                female
                                                                                   1
     4
                                   Allen, Mr. William Henry
                                                                  male
                                                                        35.0
                                                                                   0
     5
                                            Moran, Mr. James
                                                                  male
                                                                         {\tt NaN}
                                                                                   0
     6
                                    McCarthy, Mr. Timothy J
                                                                  male 54.0
                                                                                   0
     7
                             Palsson, Master. Gosta Leonard
                                                                  male
                                                                         2.0
                                                                                   3
        Parch
                           Ticket
                                       Fare Cabin Embarked
     0
            0
                       A/5 21171
                                    7.2500
                                              NaN
                                                          S
                                                          C
     1
            0
                        PC 17599
                                   71.2833
                                              C85
     2
            0
                STON/02. 3101282
                                    7.9250
                                                          S
                                              NaN
     3
            0
                           113803 53.1000
                                             C123
                                                          S
                                                          S
     4
            0
                           373450
                                    8.0500
                                              {\tt NaN}
     5
            0
                                                          Q
                           330877
                                    8.4583
                                              NaN
                                                          S
     6
                                              E46
            0
                            17463
                                   51.8625
     7
                                                          S
             1
                           349909
                                   21.0750
                                              NaN
    Adressing Missing values
[]: total = train_df.isnull().sum().sort_values(ascending=False)
     percent_1 = train_df.isnull().sum()/train_df.isnull().count()*100
```

```
percent_2 = (round(percent_1, 1)).sort_values(ascending=False)
missing_data = pd.concat([total, percent_2], axis=1, keys=['Total', '%'])
missing_data
```

```
[]:
                   Total
                              %
                           77.1
     Cabin
                      687
                      177
                           19.9
     Age
                            0.2
     Embarked
                        2
     PassengerId
                            0.0
                        0
     Survived
                            0.0
                        0
     Pclass
                        0
                            0.0
     Name
                            0.0
                        0
     Sex
                            0.0
```

SibSp	0	0.0
Parch	0	0.0
Ticket	0	0.0
Fare	0	0.0