

Roll No - 130

Enrollment No - 24010101694

Lab - 5

Karan Sonagara

1) First, you need to read the titanic dataset from local disk and display Last five records

```
In [2]: import pandas as pd
import numpy as ny

In [4]: df = pd.read_csv("titanic.csv", encoding='cp1258')
df
```

]:		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. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
	891 rc	ws × 12 colur	mns								

In [8]: df.tail(5)

Out[8]:		Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Far
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7
	4										

2) Handle Missing Values in data set [use dropna(), fillna(), and interpolate]

In [10]: data_withdropna = df.dropna(how='all') #how = any # how = all # axis=1
data_withdropna

Out[10]:		Passengerld	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	7
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	(1)
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	(1)
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

In [20]: data_withfillna = df.fillna({'Age':30,'ICabin':"Not Avail"}) #({'Age':30,'ICabin
data_withfillna

Out[20]:		Passengerld	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
	0	1	0	3	Braund, Mr. Owen Harris	male		1	0	A/5 21171	
	1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	7
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
	•••			•••							
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	30.0	1	2	W./C. 6607	2
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	

891 rows × 12 columns

In [30]: datawithfillna = df.copy()
 meanAge = datawithfillna['Age'].mean()
 datawithfillna['Age'] = datawithfillna['Age'].fillna(meanAge)
 datawithfillna

Out[30]:		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	7
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	W./C. 6607	2
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
	004	40 '									

891 rows × 12 columns

In [34]: datawithfillna = df.copy()
 medianAge = datawithfillna['Age'].median()
 datawithfillna['Age'] = datawithfillna['Age'].fillna(medianAge)
 datawithfillna

34]:		Passengerld	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	7
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
	4	5	0		Allen, Mr. William Henry	male 	35.0	0	0	373450 	
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	28.0	1	2	W./C. 6607	2
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
	891 r	ows x 12 colur	nns								

891 rows × 12 columns

In [38]: data_interpolate = df.copy()
 data_interpolate['NewAge'] = data_interpolate['Age'].interpolate()
 data_interpolate

	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	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
•••										
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	3
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
891 r	ows × 13 colur	mns								

3) Apply Scaling to AGE attribute with min max, decimal scaling and z score.

```
In [40]: data_minmax = df.copy()
   age_main = data_minmax['Age']
   age_min = age_main.min()
   age_max = age_main.max()
   data_minmax['MinMaxAge'] = (age_main-age_min) / (age_max-age_main)
   data_minmax
```

	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	7
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
•••										
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	33
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	13
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
	1 2 3 4 886 888	0 1 1 2 2 3 4 5 886 887 887 888 888 889 889 890	0 1 0 1 2 1 2 3 1 3 4 1 4 5 0 886 887 0 887 888 1 888 889 0 889 890 1	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	1 0 3 Braund, Mr. Owen Harris 1 2 1 1 Equation of the property of th	1 0 3 Braund, Mr. Owen Harris male 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th female 2 3 1 3 Heikkinen, Mrs. Jacques Heath (Lijy May Peel) female 3 4 1 1 Futrelle, Mrs. Jacques Heath (Lijy May Peel) female 4 5 0 3 Allen, Mr. William Henry male 886 887 0 2 Montvila, Rev. male male 887 888 1 1 Graham, Miss. Margaret Edith female Edith 888 889 0 3 Catherine Helen "Carrie" female Helen "Carrie" 889 890 1 1 Behr, Mr. Karl Howell male Howell	0 1 0 3 Braund, Mr. Owen Harris male 22.0 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th female 38.0 2 3 1 3 Heikkinen, Mrs. Laina female 26.0 3 4 1 1 Futrelle, Mrs. Jacques Heath (Lijy May Peel) female 35.0 4 5 0 3 Allen, Mr. William Henry male 35.0 886 887 0 2 Montvila, Rev. Juozas male 27.0 887 888 1 1 Graham, Miss. Margaret Edith female Edith 19.0 888 889 0 3 Catherine Helen "Carrie" female Helen "Carrie" NaN 889 890 1 1 Behr, Mr. Karl Howell male 26.0 890 891 0 3 Behr, Mr. Mr. Moles male 26.0	1	1	0 1 0 3 Braund, Harris Mr. Owen Harris male 22.0 1 0 A/5 21/171 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th Female 38.0 1 0 PC 17599 2 3 1 3 Heikkinen, Mrs. Laina female 26.0 0 0 STON/O2. 3101282 3 4 1 1 Heikkinen, Mrs. Jacques Heath (Lily May Peel) female 35.0 1 0 113803 4 5 0 3 Allen, Mr. William Henry male 35.0 1 0 113803 886 887 0 2 Montvilliam Henry male 35.0 0 0 373450 887 888 1 1 Graham, Miss. Margaret Edith female 27.0 0 0 211536 888 889 0 3 Catherine Helen Carrie* Female Plane Female NaN NaN 1 2 W/C. G607

891 rows × 13 columns

```
In [64]: data_max = df.copy()
   age_main = data_max['Age']
   age_max = age_main.abs().max()
   temp = len(str(int(age_max)))
```

data_max["DecimalScaleAge"] = age_main/(10**temp)
data_max

Out[64]:		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	7
	2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	
	3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	5
	4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	
	•••										
	886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	1
	887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	3
	888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	2
	889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	(1)
	890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	
	891 rd	ows × 13 colur	nns								