import pandas as pd In [52]: In [53]: import numpy as np pip install pandas_summary In [54]: Requirement already satisfied: pandas_summary in c:\users\hp\anaconda3\lib\site-packages (0.2.0) Requirement already satisfied: datatile in c:\users\hp\anaconda3\lib\site-packages (from pandas_summary) (1.0.3) Requirement already satisfied: pandas in c:\users\hp\anaconda3\lib\site-packages (from pandas_summary) (2.2.2) Requirement already satisfied: numpy in c:\users\hp\anaconda3\lib\site-packages (from pandas_summary) (1.23.5) Requirement already satisfied: traceml<1.1 in c:\users\hp\anaconda3\lib\site-packages (from datatile->pandas_summary) (1.0.8) Requirement already satisfied: python-dateutil>=2.8.2 in c:\users\hp\anaconda3\lib\site-packages (from pandas->pandas_summary) (2.8.2) Requirement already satisfied: tzdata>=2022.7 in c:\users\hp\anaconda3\lib\site-packages (from pandas->pandas_summary) (2024.1) Requirement already satisfied: pytz>=2020.1 in c:\users\hp\anaconda3\lib\site-packages (from pandas->pandas_summary) (2022.7) Requirement already satisfied: six>=1.5 in c:\users\hp\anaconda3\lib\site-packages (from python-dateutil>=2.8.2->pandas->pandas_summary) (1.16.0) Note: you may need to restart the kernel to use updated packages. from pandas_summary import DataFrameSummary In [55]: # Data Reading: In [56]: df=pd.read_csv('C:\\Users\\HP\Downloads\\train.csv') df In [57]: Passengerld Survived Pclass Sex Age SibSp Parch **Ticket** Cabin Embarked Out[57]: Name Fare 0 1 0 3 A/5 21171 S Braund, Mr. Owen Harris male 22.0 0 7.2500 NaN 2 1 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 C85 С 3 1 3 2 STON/O2. 3101282 S Heikkinen, Miss. Laina female 26.0 0 0 7.9250 NaN Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 3 4 1 1 0 113803 53.1000 C123 S 1 0 3 4 5 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 NaN S 0 2 0 211536 13.0000 S 886 887 Montvila, Rev. Juozas male 27.0 0 NaN 112053 30.0000 887 888 1 1 Graham, Miss. Margaret Edith female 19.0 0 B42 S 0 889 3 Johnston, Miss. Catherine Helen "Carrie" 2 W./C. 6607 23.4500 S 888 female NaN NaN С 889 890 1 1 Behr, Mr. Karl Howell male 26.0 111369 30.0000 C148 0 3 Dooley, Mr. Patrick 0 Q 890 891 male 32.0 0 370376 7.7500 NaN 891 rows × 12 columns df.shape In [58]: (891, 12)Out[58]: df.head() In [59] Out[59]: Passengerld Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked 0 1 0 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 NaN S 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 C85 2 3 3 0 S 1 0 STON/O2. 3101282 7.9250 NaN Heikkinen, Miss. Laina female 26.0 Futrelle, Mrs. Jacques Heath (Lily May Peel) 0 113803 53.1000 female 35.0 4 5 0 3 0 0 S 8.0500 Allen, Mr. William Henry male 35.0 373450 NaN In [60]: # Data Summary: df.describe() Out[60]: Survived **Parch** Passengerld **Pclass** Age SibSp Fare 891.000000 891.000000 891.000000 714.000000 891.000000 891.000000 891.000000 count mean 446.000000 0.383838 2.308642 29.699118 0.523008 0.381594 32.204208 0.836071 1.102743 std 257.353842 0.486592 14.526497 0.806057 49.693429 0.000000 1.000000 0.000000 1.000000 0.420000 0.000000 0.000000 min 223.500000 0.000000 20.125000 25% 2.000000 0.000000 0.000000 7.910400 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 891.000000 1.000000 3.000000 80.000000 8.000000 6.000000 512.329200 max df.describe(include='all') In [61]: Out[61]: Passengerld Survived **Pclass** Name Sex Age SibSp Parch Ticket Fare Cabin Embarked 891.000000 891.000000 891.000000 891 714.000000 891.000000 891.000000 891 891.000000 204 889 count 891 unique NaN 891 NaN NaN NaN 681 NaN 147 NaN NaN B96 B98 S NaN NaN Braund, Mr. Owen Harris male NaN NaN 347082 top NaN NaN NaN NaN 577 NaN NaN NaN NaN 644 freq NaN NaN 446.000000 0.383838 2.308642 29.699118 0.523008 0.381594 32.204208 mean NaN NaN NaN NaN NaN std 257.353842 0.486592 0.836071 NaN NaN 14.526497 1.102743 0.806057 NaN 49.693429 NaN NaN 1.000000 0.000000 0.420000 0.000000 0.000000 0.000000 1.000000 NaN NaN NaN min NaN NaN 25% 223.500000 0.000000 2.000000 20.125000 0.000000 0.000000 7.910400 NaN NaN NaN NaN NaN 446.000000 0.000000 3.000000 28.000000 0.000000 0.000000 14.454200 NaN 50% NaN NaN NaN NaN 668.500000 1.000000 3.000000 38.000000 1.000000 0.000000 31.000000 NaN **75**% NaN NaN NaN NaN 891.000000 1.000000 3.000000 80.000000 8.000000 6.000000 NaN 512.329200 max NaN NaN NaN NaN df.describe(include=['object']) Ticket Cabin Embarked Out[62]: Sex Name count 891 891 891 204 889 unique 891 2 681 147 3 S Braund, Mr. Owen Harris male 347082 B96 B98 1 577 644 freq df.describe(include=['object', 'float']) Out[66]: Age Ticket Fare Cabin Embarked 891 714.000000 891.000000 204 889 891 891 count 681 NaN 3 unique NaN Braund, Mr. Owen Harris male 347082 B96 B98 S NaN NaN top 577 NaN 7 NaN 644 freq 29.699118 NaN 32.204208 NaN NaN mean NaN NaN 14.526497 49.693429 NaN std NaN NaN NaN 0.420000 0.000000 min NaN NaN NaN NaN NaN 20.125000 25% 7.910400 NaN NaN NaN NaN NaN 28.000000 **50**% NaN NaN NaN 14.454200 NaN NaN **75**% NaN NaN 38.000000 NaN 31.000000 NaN 80.000000 NaN 512.329200 NaN NaN NaN NaN max #Handling Missing Values: df.isnull() Out[75]: Passengerld Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked 0 False True False 1 False 2 False True False 3 False 4 False False False False False False False False False True False 886 False True False 887 False 888 False False False False False True False False False False True False 889 False 890 False False False False False False False True False 891 rows × 12 columns In [76]: df.isnull().sum() 0 PassengerId Out[76]: 0 Survived 0 Pclass Name 0 Sex 0 177 Age SibSp Parch 0 Ticket 0 Fare 0 Cabin 687 Embarked 2 dtype: int64 df.isnull().sum().sum() In [77]: Out[77]: #filling Null Values df2=df.fillna(value=0) Passengerld Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked Out[78]: 0 0 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 0 S C85 С 1 Cumings, Mrs. John Bradley (Florence Briggs Th... 0 PC 17599 71.2833 female 38.0 1 3 1 S 2 3 Heikkinen, Miss. Laina 26.0 0 0 STON/O2. 3101282 7.9250 0 S 3 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) 0 113803 53.1000 C123 female 35.0 4 5 0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 0 S 886 887 0 2 Montvila, Rev. Juozas male 27.0 0 0 211536 13.0000 0 S 888 S 887 0 112053 30.0000 B42 1 Graham, Miss. Margaret Edith female 19.0 0 S 888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" 2 W./C. 6607 23.4500 0 female С 889 890 1 Behr, Mr. Karl Howell male 26.0 0 111369 30.0000 C148 890 891 0 3 Dooley, Mr. Patrick male 32.0 0 0 370376 7.7500 0 Q 891 rows × 12 columns df2.isnull().sum().sum() Out[79]: df3=df.fillna(value=5) In [80]: df3 Sex Age SibSp Passengerld Survived Pclass **Ticket** Cabin Embarked Out[80]: Name Parch Fare 0 0 3 S 1 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 5 С 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 C85 1 2 3 1 3 5 S Heikkinen, Miss. Laina female 26.0 0 0 STON/O2. 3101282 7.9250 Futrelle, Mrs. Jacques Heath (Lily May Peel) 3 4 1 1 0 113803 53.1000 C123 S female 35.0 1 4 5 0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 5 S 2 0 5 S 886 887 0 Montvila, Rev. Juozas male 27.0 0 211536 13.0000 887 888 1 0 112053 30.0000 B42 S 1 Graham, Miss. Margaret Edith female 19.0 0 889 0 3 2 5 S 888 Johnston, Miss. Catherine Helen "Carrie" female 5.0 1 W./C. 6607 23.4500 С 889 890 1 1 Behr, Mr. Karl Howell male 26.0 0 111369 30.0000 C148 0 3 0 5 Q 890 891 Dooley, Mr. Patrick male 32.0 0 370376 7.7500 891 rows × 12 columns df3.isnull().sum() PassengerId Out[81]: Survived 0 Pclass 0 Name 0 Sex 0 0 Age SibSp 0 Parch 0 Ticket 0 Fare 0 Cabin 0 0 Embarked dtype: int64 #filling Null Values with a previous value df4=df.fillna(method='pad') df4 C:\Users\HP\AppData\Local\Temp\ipykernel_8268\2637400133.py:2: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead. df4=df.fillna(method='pad') Out[82]: Passengerld Survived Pclass Sex Age SibSp Parch Ticket Fare Cabin Embarked Name 0 1 0 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 NaN S 1 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... PC 17599 С 38.0 1 0 71.2833 C85 2 3 female 26.0 S 3 1 Heikkinen, Miss. Laina 0 0 STON/O2. 3101282 7.9250 C85 3 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 0 113803 53.1000 C123 S 4 5 0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 C123 S ... S 886 887 0 2 Montvila, Rev. Juozas male 27.0 0 0 211536 13.0000 C50 Graham, Miss. Margaret Edith female 19.0 S 887 888 1 0 0 112053 30.0000 B42 Johnston, Miss. Catherine Helen "Carrie" 888 889 0 3 female 19.0 1 2 W./C. 6607 23.4500 B42 S С 890 1 0 889 1 Behr, Mr. Karl Howell male 26.0 0 111369 30.0000 C148 0 3 Q 890 891 Dooley, Mr. Patrick male 32.0 0 0 370376 7.7500 C148 891 rows × 12 columns In [83]: df4.isnull().sum() PassengerId Survived 0 0 **Pclass** 0 Name 0 Sex 0 Age 0 SibSp 0 Parch 0 Ticket Fare 0 Cabin Embarked dtype: int64 #filling Null Values with a nextvalue In [84]: df5=df.fillna(method='bfill') C:\Users\HP\AppData\Local\Temp\ipykernel_8268\3242335653.py:2: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead. df5=df.fillna(method='bfill') Passengerld Survived Pclass Sex Age SibSp Parch **Ticket** Fare Cabin Embarked Out[84]: Name 0 1 0 3 Braund, Mr. Owen Harris 22.0 0 A/5 21171 7.2500 C85 S male 1 С 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 C85 1 2 3 1 3 Heikkinen, Miss. Laina female 26.0 0 STON/O2. 3101282 7.9250 C123 S 4 S 3 1 1 0 Futrelle, Mrs. Jacques Heath (Lily May Peel) 35.0 1 113803 53.1000 C123 female 4 5 0 3 Allen, Mr. William Henry 0 0 8.0500 E46 S male 35.0 373450 Montvila, Rev. Juozas 886 887 0 2 male 27.0 0 0 211536 13.0000 B42 S 887 888 1 female 19.0 0 B42 S 1 Graham, Miss. Margaret Edith 0 112053 30.0000 888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" 26.0 1 2 W./C. 6607 23.4500 C148 S female С 889 890 male 26.0 0 1 1 Behr, Mr. Karl Howell 111369 30.0000 C148 0 890 891 0 3 Dooley, Mr. Patrick 0 370376 7.7500 Q male 32.0 NaN 891 rows × 12 columns df6=df.fillna(method='pad',axis=1) In [85]: C:\Users\HP\AppData\Local\Temp\ipykernel_8268\2061083641.py:1: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead. df6=df.fillna(method='pad',axis=1) Passengerld Survived Pclass Age SibSp Parch **Ticket** Cabin Embarked Out[85]: Name Sex Fare 3 0 0 Braund, Mr. Owen Harris A/5 21171 S 1 male 22.0 1 0 7.25 7.25 1 Cumings, Mrs. John Bradley (Florence Briggs Th... 1 2 1 38.0 1 0 PC 17599 71.2833 C85 С 2 3 1 3 0 STON/O2. 3101282 S Heikkinen, Miss. Laina female 26.0 0 7.925 7.925 3 4 1 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) 35.0 0 113803 C123 S female 1 53.1 5 0 3 Allen, Mr. William Henry 0 S 4 male 35.0 0 373450 8.05 8.05 886 887 0 2 Montvila, Rev. Juozas 27.0 0 0 211536 13.0 13.0 S male 887 888 Graham, Miss. Margaret Edith 19.0 0 112053 B42 S 1 1 female 0 30.0 Johnston, Miss. Catherine Helen "Carrie" 2 888 889 0 3 female female 1 W./C. 6607 23.45 23.45 S 889 890 0 С 1 1 Behr, Mr. Karl Howell 26.0 0 111369 C148 male 30.0 0 3 0 Q 890 891 Dooley, Mr. Patrick male 32.0 0 370376 7.75 7.75 891 rows × 12 columns df7=df.fillna(method='bfill',axis=1) df7 C:\Users\HP\AppData\Local\Temp\ipykernel_8268\1055879924.py:1: FutureWarning: DataFrame.fillna with 'method' is deprecated and will raise in a future version. Use obj.ffill() or obj.bfill() instead. df7=df.fillna(method='bfill',axis=1) Passengerld Survived Pclass Sex Age SibSp Parch Fare Cabin Embarked Out[86]: Name Ticket 0 1 0 3 Braund, Mr. Owen Harris male 22.0 1 0 A/5 21171 7.25 S S 1 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 C85 С 2 3 1 3 S Heikkinen, Miss. Laina female 26.0 0 STON/O2. 3101282 7.925 S 0 3 C123 S Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 0 113803 53.1 5 0 3 0 373450 S S 4 Allen, Mr. William Henry 0 8.05 male 35.0 887 0 2 0 0 211536 S S 886 Montvila, Rev. Juozas male 27.0 13.0 888 S 887 1 Graham, Miss. Margaret Edith female 19.0 0 0 112053 30.0 B42 S 888 889 0 3 2 S Johnston, Miss. Catherine Helen "Carrie" female 1 1 W./C. 6607 23.45 С 889 890 1 Behr, Mr. Karl Howell male 26.0 0 111369 30.0 C148 0 3 0 Q 890 891 Dooley, Mr. Patrick male 32.0 0 370376 7.75 Q 891 rows × 12 columns #filling different values is null in different columns df8=df.fillna({'Age':'abcd','Cabin':'defg'}) Passengerld Survived Pclass Age SibSp Cabin Embarked Out[87]: Name Sex Parch **Ticket** Fare 3 0 1 0 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 defg S 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 0 PC 17599 71.2833 C85 С 1 38.0 3 1 3 2 Heikkinen, Miss. Laina 0 STON/O2. 3101282 7.9250 defg S Futrelle, Mrs. Jacques Heath (Lily May Peel) female 3 4 1 1 35.0 0 113803 53.1000 C123 S 4 5 0 3 Allen, Mr. William Henry male 35.0 0 0 373450 8.0500 defg S 886 887 0 2 Montvila, Rev. Juozas male 27.0 0 0 211536 13.0000 defg S 887 888 Graham, Miss. Margaret Edith female 0 0 112053 30.0000 B42 S 1 1 19.0 888 889 0 3 Johnston, Miss. Catherine Helen "Carrie" female 2 W./C. 6607 23.4500 defg S 0 889 890 1 1 Behr, Mr. Karl Howell 111369 30.0000 C148 С male 26.0 Q 890 891 0 3 Dooley, Mr. Patrick male 32.0 0 0 370376 7.7500 891 rows × 12 columns #filling null values with the mean of a columns df9=df.fillna(value=df['Age'].mean()) df9 Cabin Embarked Passengerld Survived Pclass Out[88]: Name Sex Age SibSp Parch Ticket Fare 0 S 1 0 3 Braund, Mr. Owen Harris male 22.000000 0 A/5 21171 7.2500 29.699118 1 2 1 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.000000 PC 17599 71.2833 С 2 3 1 3 0 STON/O2. 3101282 S Heikkinen, Miss. Laina female 26.000000 0 7.9250 29.699118 3 4 Futrelle, Mrs. Jacques Heath (Lily May Peel) 35.000000 113803 53.1000 S 4 5 0 3 Allen, Mr. William Henry 0 0 S male 35.000000 373450 8.0500 29.699118 886 887 0 2 male 27.000000 0 0 211536 13.0000 29.699118 S Montvila, Rev. Juozas 887 888 Graham, Miss. Margaret Edith female 19.000000 112053 30.0000 889 890 Behr, Mr. Karl Howell male 26.000000 111369 30.0000 C148 890 0 3 Q 891 Dooley, Mr. Patrick male 32.000000 0 370376 7.7500 29.699118 891 rows × 12 columns df9=df.fillna(value=df['Age'].max()) In [89]: Out[89]: Passengerld Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked 0 0 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 80.0 S 1 1 71.2833 С 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 C85 2 3 1 3 S Heikkinen, Miss. Laina female 26.0 0 0 STON/O2. 3101282 7.9250 80.0 3 S Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 113803 53.1000 C123 0 3 5 0 0 S 4 Allen, Mr. William Henry male 35.0 373450 8.0500 80.0 887 2 male 27.0 S 886 0 0 0 211536 13.0000 80.0 Montvila, Rev. Juozas 888 S 887 1 Graham, Miss. Margaret Edith female 19.0 112053 30.0000 B42 888 889 0 3 2 S Johnston, Miss. Catherine Helen "Carrie" female 80.0 W./C. 6607 23.4500 80.0 890 С 889 Behr, Mr. Karl Howell male 26.0 0 111369 30.0000 C148 0 3 890 891 Dooley, Mr. Patrick male 32.0 0 0 370376 7.7500 80.0 Q 891 rows × 12 columns df9=df.fillna(value=df['Age'].min()) In [90]: df9 Out[90]: Passengerld Survived Pclass Name Sex Age SibSp Parch **Ticket** Fare Cabin Embarked 0 1 0 3 Braund, Mr. Owen Harris male 22.00 0 A/5 21171 7.2500 0.42 S 1 Cumings, Mrs. John Bradley (Florence Briggs Th... 1 2 0 PC 17599 71.2833 C85 С 1 38.00 female 1 2 3 1 3 S Heikkinen, Miss. Laina 26.00 0 STON/O2. 3101282 7.9250 0.42 3 Futrelle, Mrs. Jacques Heath (Lily May Peel) 113803 53.1000 S 4 1 1 female 35.00 1 0 C123 4 5 0 3 Allen, Mr. William Henry male 35.00 0 0 373450 8.0500 0.42 S 2 886 887 0 Montvila, Rev. Juozas male 27.00 0 0 211536 13.0000 0.42 S 887 888 112053 30.0000 B42 S 1 1 Graham, Miss. Margaret Edith female 19.00 0 0 0 S 888 889 3 Johnston, Miss. Catherine Helen "Carrie" 0.42 2 W./C. 6607 23.4500 0.42 889 890 Behr, Mr. Karl Howell 26.00 0 0 111369 30.0000 C148 С male Q 890 891 0 3 Dooley, Mr. Patrick male 32.00 0 0 370376 7.7500 0.42 891 rows × 12 columns df10=df.dropna() In [91]: df10 Out[91]: Passengerld Survived Pclass Name Sex Age SibSp Parch Ticket Fare Cabin Embarked 1 2 C85 С 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 0 PC 17599 71.2833 1 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 3 4 1 1 113803 53.1000 C123 S 0 7 0 6 1 McCarthy, Mr. Timothy J male 0 0 17463 51.8625 E46 S PP 9549 10 11 1 3 16.7000 G6 Sandstrom, Miss. Marguerite Rut female 11 12 113783 26.5500 C103 S 1 1 Bonnell, Miss. Elizabeth female 0 871 872 D35 S 1 1 Beckwith, Mrs. Richard Leonard (Sallie Monypeny) female 47.0 11751 52.5542 872 873 0 5.0000 B51 B53 B55 Carlsson, Mr. Frans Olof male 33.0 695 879 880 1 1 Potter, Mrs. Thomas Jr (Lily Alexenia Wilson) female 56.0 0 1 11767 83.1583 C50 С 888 1 112053 30.0000 S 887 Graham, Miss. Margaret Edith female 19.0 0 B42 111369 30.0000 889 890 1 C148 С 1 Behr, Mr. Karl Howell male 26.0 183 rows × 12 columns #replace In [92]: df12=df.replace(to_replace=np.nan, value=60) Passengerld Survived Pclass Fare Cabin Embarked Out[92]: Name Sex Age SibSp Parch Ticket 0 1 0 3 Braund, Mr. Owen Harris male 22.0 0 A/5 21171 7.2500 60 S 1 1 PC 17599 71.2833 C85 С 1 Cumings, Mrs. John Bradley (Florence Briggs Th... female 38.0 2 3 1 3 0 STON/O2. 3101282 7.9250 S Heikkinen, Miss. Laina female 26.0 0 60 Futrelle, Mrs. Jacques Heath (Lily May Peel) female 35.0 113803 53.1000 3 1 S C123 5 3 S 4 0 0 0 Allen, Mr. William Henry male 35.0 373450 8.0500 60 887 0 2 male 27.0 0 0 211536 13.0000 S 886 Montvila, Rev. Juozas 60 887 888 1 112053 30.0000 S Graham, Miss. Margaret Edith female 19.0 B42 Johnston, Miss. Catherine Helen "Carrie" S 888 889 0 3 2 W./C. 6607 23.4500 60 female 60.0 1 889 890 С Behr, Mr. Karl Howell male 26.0 0 111369 30.0000 C148 0 3 890 891 Dooley, Mr. Patrick male 32.0 0 0 370376 7.7500 60 Q 891 rows × 12 columns In [97]: **#Categorical Data Encoding** from sklearn.preprocessing import LabelEncoder label_encoder=LabelEncoder() In [98]: df['Sex'] In [99] male 0 Out[99] female 2 female 3 female male . . . 886 male 887 female 888 female 889 male 890 male Name: Sex, Length: 891, dtype: object label_encoder.fit_transform(df['Sex']) array([1, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, Out[100]: 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 1]) In []: