# index

Rahmath

Name : Barath Kumar S J Std : III rd Sec : AI & DS

Roll No : 2 3 1 8 0 1 0 2 0 Sub :

| S.No. | Date | Title | Page No. | Teacher's Sign / Remarks |
|---|---|---|---|---|
| 1. | 2.8.25 | Load & Convert dataset into dataframe | | |
| 2. | 2.8.25 | Data discovery & preparation | | |
| 3. | 5.8.25 | Customer segmentation | 10 | |
| 4. | 8.8.25 | Text preprocessing | | |
| 5. | 9.9.25 | NLP Task & Info Retrieval | 10 | |
| 6. | 9.9.25 | EDA with python | | |
| 7. | 16.9.25 | Clustering | | |
| 8. | 4.10.25 | Interactive Dashboard with Tableau | | |

Output :

| | Survived | pclass | age | sibsp | fare | Class |
|---|---|---|---|---|---|---|
| 0 | 0 | 3 | 22.0 | 1 | 7.2500 | Third |
| 1 | 1 | 1 | 38.0 | 1 | 71.233 | First |
| 2 | 1 | 3 | 26.0 | 0 | 7.9250 | Third |
| 3 | 1 | 1 | 35.0 | 1 | 53.100 | First |
| 4 | 0 | 3 | 35.0 | 0 | 8.0500 | Third |

Missing values per column :

Survived : 0

pclass : 0

Sex : 0

Age : 177

sibsp : 0

parch : 0

Duplicate removed : 56.

EXP. NO : 1

Load the titanic dataset and convert it into dataframe

AIM :

To perform basic program and exploring data analysis.

PROGRAM CODE :

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import
                    Labeler code, standard
                                    scale

tit = sns.load_dataset('titanic')
print('tit.head())
tit.isnull().sum()
tit['age'] = tit['age'].fill na[method.
                        fill na(method ='bull'

tit['deck'] = tit['deck'].cat.add.categorical(
                        ['unknown'])

miss_ = tit[tit['deck'].isna()].index[:5]

tit.loc[miss_,'deck'] = unknown

le = labelEncoder()

tit['Sex_encoded'] = le.fit_transform.
                    (titanic['sex'])

Scaler = StandardScaler()

tit['fbase_scaled'] = Scaler.fit_transform(
                    tit['fare'])
```

Correlation Heatmap :

| | | | | | |
|---|---|---|---|---|---|
| Pclass | 1.00 | -0.29 | 0.10 | 0.03 | -0.55 |
| age | 0.29 | 1.00 | -0.21 | -0.17 | 0.08 |
| sibsp | 0.10 | -0.21 | 1.00 | 0.41 | 0.15 |
| parch | 0.03 | -0.17 | 0.41 | 1.00 | 0.21 |
| fare | -0.55 | 0.8 | 0.15 | 0.2 | 1.00 |
| | Pclass | age | Sibsp | Parch | fare |

```
sns.pairplot(tit, vars=['pclass', 'sex', 'survived',
                        'age', 'sibsp'])
plt.suptitle('pairplot of selected feature',
             y=1.02)
cor_features = ['pclass', 'age', 'sibsp',
                'pair', 'false']
corr = tit[cor_features].corr()
plt.figure(figsize=(8, 6))
sns.heatmap(corr, annot=True, cmap='cool',
            fmt='.2f')
plt.title('correlation Heatmap')
plt.show()
```

RESULT:

Thus, the data processing and cleaning using titanic dataset has been executed successfully.