Unit 3.1Graded Assignment

Instructions:
Implement a label encoder for categorical data using pure Python, Pandas and NumPy

Submitted by:

1. **Ali Nasir (2303.KHI.DEG.012)**
2. **Saif ur Rehman (2303.KHI.DEG.007)**

Solution:

1. First we have import our dataset.

Assignment

```python
import pandas as pd
import numpy as np

# DataSets
data1 = pd.read_csv('/home/saifrehman/Documents/DataScience_course/Datasets/ipl-matches.csv')
data1.sample(5)
```

[84] ✓ 0.0s Python

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players
241	1175368	Chandigarh	2019-04-01	2019	13	Kings XI Punjab	Delhi Capitals	Punjab Cricket Association IS Bindra Stadium	Delhi Capitals	field	N	Kings XI Punjab	Runs	14.0	NaN	SM Curran	['KL Rahul', 'SM Curran', 'MA Agarwal', 'SN Kh...
147	1216524	NaN	2020-10-27	2020/21	47	Sunrisers Hyderabad	Delhi Capitals	Dubai International Cricket Stadium	Delhi Capitals	field	N	Sunrisers Hyderabad	Runs	88.0	NaN	WP Saha	['DA Warner', 'WP Saha', 'MK Pandey', 'KS Will...
869	392206	Johannesburg	2009-05-02	2009	26	Chennai Super Kings	Delhi Daredevils	New Wanderers Stadium	Delhi Daredevils	field	N	Chennai Super Kings	Runs	18.0	NaN	SB Jakati	['M Vijay', 'ML Hayden', 'SK Raina', 'S Badrin...
22	1304098	Mumbai	2022-05-07	2022	52	Punjab Kings	Rajasthan Royals	Wankhede Stadium, Mumbai	Punjab Kings	bat	N	Rajasthan Royals	Wickets	6.0	NaN	YBK Jaiswal	['JM Bairstow', 'S Dhawan', 'PBB Rajapaksa', '...
496	734041	Mumbai	2014-05-25	2014	56	Mumbai Indians	Rajasthan Royals	Wankhede Stadium	Mumbai Indians	field	N	Mumbai Indians	Wickets	5.0	NaN	CJ Anderson	['LMP Simmons', 'MEK Hussey', 'CJ Anderson', '...

2. We have implemented function in two use cases if the run by default it will labelize all the object based column. In second case we will labialize in the desired column. We have saved all the str type data frame a list alled var, which will labelize all the categorical data. The def labelencoder function states that make copy of our data frame in encoded_dataframe variable → used for loop pn dataframe columns → using if condition to check the data type = " O" → using for loop to encode the columns → categories to set the unique values separate →using for loop to emurate the the values assign unique value to a label → The map() function is used to apply the label encoding dictionary (encode) to each value in the column. The astype() method is used to convert the resulting column to the int16 data type for memory efficiency → finally return encoded dataframe

Code

```python
var = [i for i in data1 if data1[i].dtype == 'O']
def labelencoder(dataFrame, index=var):
    encoded_dataFrame = dataFrame.copy()
    for i in dataFrame.columns:
        if dataFrame[i].dtype == 'O':
            for j in dataFrame[index]:
                categories = dataFrame[j].unique()
                # This line creates a dictionary that maps each unique category to a unique integer value.
                encode = {cat: x for x, cat in enumerate(categories)}
                encoded_dataFrame[j] = dataFrame[j].map(encode).astype(np.int16)

    return encoded_dataFrame
```

[80] ✓ 0.0s

3. When we run the function we can see that it has labialize all the categorical data

```python
a = labelencoder(data1)
a.sample(5)
```

[82] ✓ 0.2s Python

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2Players	Umpire1
9	1304111	2	9	0	9	2	4	2	4	1	0	6	1	3.0	0	9	9	9	0
602	598022	1	465	9	48	7	9	19	7	0	0	9	0	4.0	0	67	592	592	41
536	729309	10	420	8	58	11	4	7	4	0	0	12	0	6.0	0	165	527	526	39
10	1304110	3	10	0	10	3	3	4	6	1	0	7	1	17.0	0	10	10	10	7
69	1304051	4	58	0	69	0	7	5	3	1	0	1	1	61.0	0	53	67	67	2

[ ] Python

```python
a = labelencoder(data1, ['City','SuperOver'])
a.sample(5)
```

[81] ✓ 0.0s Python

	ID	City	Date	Season	MatchNumber	Team1	Team2	Venue	TossWinner	TossDecision	SuperOver	WinningTeam	WonBy	Margin	method	Player_of_Match	Team1Players	Team2F
794	419146	13	2010-04-09	2009/10	41	Kings XI Punjab	Mumbai Indians	Punjab Cricket Association Stadium, Mohali	Mumbai Indians	bat	0	Kings XI Punjab	Wickets	6.0	NaN	KC Sangakkara	['AB Barath', 'DPMD Jayawardene', 'KC Sangakka...	['S Dh 'SR Tend 'AT R
409	980947	2	2016-04-28	2016	24	Mumbai Indians	Kolkata Knight Riders	Wankhede Stadium	Mumbai Indians	field	0	Mumbai Indians	Wickets	6.0	NaN	RG Sharma	['RG Sharma', 'PA Patel', 'AT Rayudu', 'KH Pan...	['RV Utl 'G Ga 'Sh H
921	336009	8	2008-05-08	2007/08	28	Delhi Daredevils	Chennai Super Kings	Feroz Shah Kotla	Chennai Super Kings	field	0	Chennai Super Kings	Wickets	4.0	NaN	MS Dhoni	['G Gambhir', 'V Sehwag', 'AB de Villiers', 'S...	['S Vidy Flemir Dhc
4	1304116	2	2022-05-22	2022	70	Sunrisers Hyderabad	Punjab Kings	Wankhede Stadium, Mumbai	Sunrisers Hyderabad	bat	0	Punjab Kings	Wickets	5.0	NaN	Harpreet Brar	['PK Garg', 'Abhishek Sharma', 'RA Tripathi', ...	['JM Bai 'S Dhaw Sh *
373	981019	17	2016-05-29	2016	Final	Royal Challengers Bangalore	Sunrisers Hyderabad	M Chinnaswamy Stadium	Sunrisers Hyderabad	bat	0	Sunrisers Hyderabad	Runs	8.0	NaN	BCJ Cutting	['CH Gayle', 'V Kohli', 'AB de Villiers', 'KL ...	['DA W 'S Dh Henr