COLLEGE CODE : 8100

REGISTER NO :810021205053

WEBSITE TRAFFIC ANALYSIS

DATA ANALYTICS WITH COGNOS : GROUP 5

PHASE : 3

This phase involves in designing of the steps that defining in each phase of the previous documentation this involves importing necessary functions, data processing and so on in this phase we have to begin our project by loading and preprocessing the dataset.

The IBM suggests using the jupyter notebook for loading and preprocess the dataset:

Here for this project title we need to define the loading the libraries, understand the data and visualize the missing values.

For this certain inputs are defined for this project.in this phase each of the input

Codes of project is given below:

untitled7

October 18, 2023

1. 1Sunday1 9/14/20142,1461,582
2. 2Monday2 9/15/20143,6212,528
3. 3Tuesday3 9/16/20143,6982,630
4. 4 Wednesday4 9/17/20143,6672,614
5. 5Thursday5 9/18/20143,3162,366

First.Time.Visits Returning.Visits

1. 1,430152
2. 2,297231
3. 2,352278
4. 2,327287
5. 2,130236

[4]:RowDay Day.Of.WeekDate Page.Loads Unique.Visits \

2162 2163Saturday7 8/15/20202,211,696 2163 2164Sunday1 8/16/20202,7242,037

1. 2165Monday2 8/17/20203,4562,638

2. 2166Tuesday3 8/18/20203,5812,683
3. 2167 Wednesday4 8/19/20202,0641,564

First.Time.Visits Returning.Visits

1. 1,373323
2. 1,686351
3. 2,181457

21652,184 499

21661,297 267

[5]: (2167, 8)

<class 'pandas.core.frame.DataFrame'> RangeIndex: 2167 entries, 0 to 2166 Data columns (total 8 columns):

#ColumnNon-Null Count Dtype

--- ---------- -----

0Row2167 non-null int64

1Day2167 non-null object

2Day.Of.Week2167 non-null int64

3Date2167 non-null object

4Page.Loads2167 non-null object

5Unique.Visits2167 non-null object

6First.Time.Visits 2167 non-null object

7 Returning.Visits 2167 non-null dtypes: int64(2), object(6) memory usage: 135.6+ KB object

[7]: array(['Row', 'Day', 'Day.Of.Week', 'Date', 'Page.Loads', 'Unique.Visits',

'First.Time.Visits', 'Returning.Visits'], dtype=object)

[8]: Row int64

Day object

Day.Of.Week int64

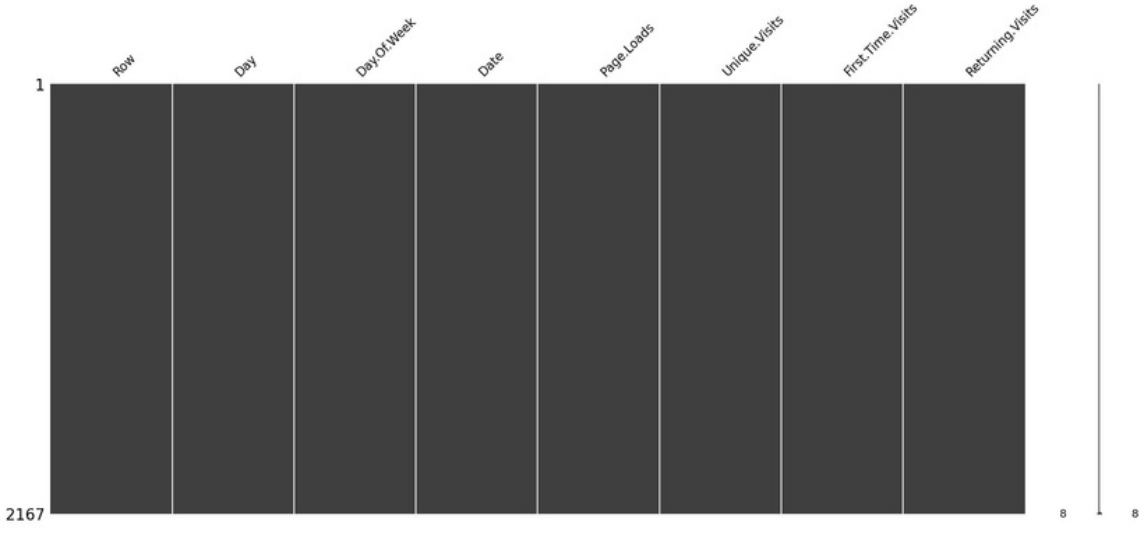
Date object

Page.Loads object

Unique.Visits object

First.Time.Visits object

Returning.Visits dtype: object object



[10]:RowDay Day.Of.WeekDate Page.Loads First.Time.Visits \

1. 1Sunday1 9/14/20142,1461,430
2. 2Monday2 9/15/20143,6212,297
3. 3Tuesday3 9/16/20143,6982,352
4. 4 Wednesday4 9/17/20143,6672,327
5. 5Thursday5 9/18/20143,3162,130

Returning.Visits

1. 152
2. 231

3. 278
4. 287
5. 236

[11]:RowDay Day.Of.WeekDate Page.Loads First.Time.Visits \

0False False False False False False

1False False False False False False

2False False False False False False

3False False False False False False

4False False False False False False

....

2162 False False False False False False

2163 False False False False False False

2164 False False False False False False

2165 False False False False False False

2166 False False False False False False

Returning.Visits

1. False
2. False
3. False
4. False
5. False

.....

1. False
2. False
3. False
4. False
5. False

[2167 rows x 7 columns]

[12]: Row 0

Day 0

Day.Of.Week 0

Date 0

Page.Loads 0

First.Time.Visits 0

Returning.Visits dtype: int64 0

[13]: Row 0

Day 0

Day.Of.Week 0

Date 0

Page.Loads 0

First.Time.Visits 0

Returning.Visits dtype: int64 0

[14]: Empty DataFrame

Columns: [Row, Day, Day.Of.Week, Date, Page.Loads, First.Time.Visits,

Returning.Visits]

Index: []

[15]:RowDay Day.Of.WeekDate Page.Loads First.Time.Visits \

1. 1Sunday1 9/14/20142,1461,430
2. 2Monday2 9/15/20143,6212,297
3. 3Tuesday3 9/16/20143,6982,352

4. 4 Wednesday4 9/17/20143,6672,327
5. 5Thursday5 9/18/20143,3162,130

.....

1. 2163Saturday7 8/15/20202,2211,373
2. 2164Sunday1 8/16/20202,7241,686
3. 2165Monday2 8/17/20203,4562,181
4. 2166Tuesday3 8/18/20203,5812,184
5. 2167 Wednesday4 8/19/20202,0641,297

Returning.Visits

1. 152
2. 231
3. 278
4. 287
5. 236

.....

1. 323
2. 351
3. 457
4. 499
5. 267

[2167 rows x 7 columns]

<class 'pandas.core.frame.DataFrame'> RangeIndex: 2167 entries, 0 to 2166 Data columns (total 7 columns):

#ColumnNon-Null Count Dtype

--- ---------- -----

1. Row2167 non-nullint64
2. Day2167 non-nullobject
3. Day.Of.Week2167 non-nullint64
4. Date2167 non-nulldatetime64[ns]
5. Page.Loads2167 non-nullobject 5First.Time.Visits 2167 non-nullobject 6Returning.Visits2167 non-nullobject dtypes: datetime64[ns](1), int64(2), object(4) memory usage: 118.6+ KB

None

[17]: Row 0

Day 0

Day.Of.Week 0

Date 0

Page.Loads 0

First.Time.Visits 0

Returning.Visits dtype: int64 0

[18]: df["Returning.Visits"]=df['Returning.Visits'].map({0:"no", 1: "yes"}) df.head()

[18]:RowDay Day.Of.WeekDate Page.Loads First.Time.Visits \

1. 1Sunday1 2014-09-142,1461,430
2. 2Monday2 2014-09-153,6212,297
3. 3Tuesday3 2014-09-163,6982,352
4. 4 Wednesday4 2014-09-173,6672,327
5. 5Thursday5 2014-09-183,3162,130

Returning.Visits

1. NaN
2. NaN
3. NaN
4. NaN
5. NaN

[19]: df["Returning.Visits"].describe(include=['object','bool'])

[19]: count 0

unique 0

top NaN

freq NaN

Name: Returning.Visits, dtype: object

[20]: Int64Index([], dtype='int64')

[21]: numerical_cols = ['Row','First.Time.Visits','Returning.Visits'] df[numerical_cols].describe()

[21]:Row count 2167.000000

mean	1084.000000
std	625.703338
min	1.000000
25%	542.500000
50%	1084.000000
75%	1625.500000
max	2167.000000

[ ]:

Accuracy :0.1726268

Consistency:23.28393

[ ]:PHASE3[1]:**import**pandas**as**pd**import**numpy**as**np**import**missingno**as**msno[2]:df=pd.read_csv('daily-website-visitors.csv')[3]:df.head()[3]:RowDay Day.Of.WeekDate Page.Loads Unique.Visits \

[4]:df.tail()

[5]:df.shape

[6]:df.info()

[7]:df.columns.values

[8]:df.dtypes

[9]:msno.matrix(df);

[10]:df=df.drop(['Unique.Visits'],axis=1)df.head()

[11]:df.isnull()

[12]:df.isnull().sum()

```
[13]:df['Row']=pd.to_numeric(df.Row,errors='coerce')df.isnull().sum()

[14]:df[np.isnan(df['Row'])]

[15]:df.fillna(df['Row'].mean())

[16]:df["Date"]=pd.to_datetime(df["Date"],format="%m/%d/%Y")print(df.info())

[17]:df.isnull().sum()

[20]:df[df['Row']==0].index
```