

PRE-PROCESSING AND VISUALIZATON

BINDU K R



WHICH OF THE STATEMENTS ARE TRUE ?

A Dataframe is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns).

B.A data frame is a two-dimensional data structure, Data is aligned in a tabular fashion in rows and columns.

C.Dataframe consists of three principal components, the data, rows, and columns.

WHICH OF THE STATEMENTS ARE TRUE ?

A Dataframe is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns).

B.A data frame is a two-dimensional data structure, Data is aligned in a tabular fashion in rows and columns.

C.Dataframe consists of three principal components,

The diagram illustrates a DataFrame as a table with labeled axes. The word 'Columns' is at the top with arrows pointing to the column headers: 'Name', 'Team', 'Number', 'Position', and 'Age'. The word 'Rows' is on the left with arrows pointing to the row indices: 0, 1, 2, 3, 4, 5, and 6. A purple box labeled 'Data' encompasses the data cells of the table.

	<i>Name</i>	<i>Team</i>	<i>Number</i>	<i>Position</i>	<i>Age</i>
0	Avery Bradley	Boston Celtics	0.0	PG	25.0
1	John Holland	Boston Celtics	30.0	SG	27.0
2	Jonas Jerebko	Boston Celtics	8.0	PF	29.0
3	Jordan Mickey	Boston Celtics	NaN	PF	21.0
4	Terry Rozier	Boston Celtics	12.0	PG	22.0
5	Jared Sullinger	Boston Celtics	7.0	C	NaN
6	Evan Turner	Boston Celtics	11.0	SG	27.0

WHICH OF THE STATEMENTS ARE TRUE ?

- A. To create DataFrame from dict of ndarray/list, all the ndarray must be of same length.
- B. If index is passed then the length index should be equal to the length of arrays.
- C. If no index is passed, then by default, index will be `range(n)` where `n` is the array length.

WHICH OF THE
STATEMENTS
ARE TRUE ?

- A. To create DataFrame from dict of ndarray/list, all the ndarray must be of same length.
- B. If index is passed then the length index should be equal to the length of arrays.
- C. If no index is passed, then by default, index will be `range(n)` where `n` is the array length.

WHICH OF THE STATEMENTS ARE TRUE ?

- A. In Order to select a column in Pandas DataFrame, we can either access the columns by calling them by their columns name.
- B. # select two columns

```
print(df[['Name', 'Qualification']])
```
- C. Rows can also be selected by passing integer location to an `iloc[]` function.
- D. Retrieve rows from a Data frame. `DataFrame.loc[]` method is used

WHICH OF THE STATEMENTS ARE TRUE ?

- A. In Order to select a column in Pandas DataFrame, we can either access the columns by calling them by their columns name.
- B. # select two columns

```
print(df[['Name', 'Qualification']])
```
- C. Rows can also be selected by passing integer location to an iloc[] function.
- D. Retrieve rows from a Data frame. DataFrame.loc[] method is used

SELECTING A SINGLE ROW

A. In Order to select a column in Pandas DataFrame, we can either access the columns by calling them by their columns name.

B. # select two columns

```
print(df[['Name', 'Qualification']])
```

C. Rows can also be selected by passing integer location to an iloc[] function.

D. Retrieve rows from a Data frame. DataFrame.loc[] method is used

SELECTING A SINGLE ROW

- # retrieving rows by iloc method
- `row2 = data.iloc[3]`

SELECT THE TRUE STATEMENTS OF MISSING DATA

- A. Missing Data can occur when no information is provided for one or more items or for a whole unit.
- B. Missing Data is a very big problem in real life scenario.
- C. Missing Data can also refer to as NA(Not Available) values in pandas.

SELECT THE TRUE STATEMENTS OF MISSING DATA

- A. Missing Data can occur when no information is provided for one or more items or for a whole unit.
- B. Missing Data is a very big problem in real life scenario.
- C. Missing Data can also refer to as NA(Not Available) values in pandas.

Check missing
values in
pandas
dataframe

Check missing values in pandas dataframe

- # using isnull() function
- `df.isnull()`

Fill null values in a datasets

- `fillna()`
- `replace()`
- `interpolate()` function

Dropping missing values

Dropping missing values

- # using dropna() function
- df.dropna()

Iterating over
rows

Iterating over rows

```
# iterating over rows using  
iterrows() function  
for i, j in df.iterrows():  
    print(i, j)  
    print()
```

Iterating over Columns

Iterating over Columns

```
print(df)
```

```
for i in columns:
```

```
# printing the third element of the column
```

```
    print (df[i][2])
```

Retrieve an
rows of
dataframe

Retrieve an
rows of
dataframe

```
# using iteritems() function to  
retrieve rows  
for key, value in df.iteritems():  
    print(key, value)  
    print()
```

DataFrame Methods

FUNCTION	DESCRIPTION
<code>index()</code>	Method returns index (row labels) of the DataFrame
<code><u>insert()</u></code>	Method inserts a column into a DataFrame
<code><u>add()</u></code>	Method returns addition of dataframe and other, element-wise (binary operator add)
<code><u>sub()</u></code>	Method returns subtraction of dataframe and other, element-wise (binary operator sub)
<code><u>mul()</u></code>	Method returns multiplication of dataframe and other, element-wise (binary operator mul)
<code><u>div()</u></code>	Method returns floating division of dataframe and other, element-wise (binary operator truediv)

DataFrame Methods

FUNCTION	DESCRIPTION
unique()	Method extracts the unique values in the dataframe
<u>nunique()</u>	Method returns count of the unique values in the dataframe
value_counts()	Method counts the number of times each unique value occurs within the Series
columns()	Method returns the column labels of the DataFrame
axes()	Method returns a list representing the axes of the DataFrame
<u>isnull()</u>	Method creates a Boolean Series for extracting rows with null values
<u>notnull()</u>	Method creates a Boolean Series for extracting rows with non-null values

Sort the rows in
a DataFrame by
the values of
one or more
columns

Sort the rows in
a DataFrame by
the values of
one or more
columns

```
df.sort_values("attribute name")
```

Sort the rows in
a DataFrame by
the values of
one or more
columns in
descending
order

Sort the rows in
a DataFrame by
the values of
one or more
columns

```
df.sort_values(  
    by="attribute name",  
    ascending=False  
)
```

Sort the rows in
a DataFrame by
the values of
one or more
columns and
select sorting
algorithm

Sort the rows in a DataFrame by the values of one or more columns and select sorting algorithm

```
df.sort_values(  
    by="attribute name",  
    ascending=False,  
    kind="mergesort"  
)
```

Split the data
into groups
based on some
criteria

Split the data
into groups
based on some
criteria

```
Team = df.groupby('Teamname')
```


Form groups
based on more
than one
category

```
Team = df.groupby('Teamname')
```

Form groups
based on more
than one
category

```
TeamPosition =  
df.groupby(['Team', 'Position'])
```

Unpivots a
DataFrame from
wide format to
long format

Heatmap

A heatmap is a two-dimensional graphical representation of data where the individual values that are contained in a matrix are represented as colors.

The seaborn python package allows the creation of annotated heatmaps which can be tweaked using Matplotlib tools as per the creator's requirement.

Unpivots a
DataFrame from
wide format to
long format

Pandas.melt()

melt() function is useful to manage a DataFrame into a format where one or more columns are identifier variables, while all other columns, considered measured variables, are unpivoted to the row axis, leaving just two non-identifier columns, variable and value.

```
pd.melt(df, id_vars = ['Name'],  
value_vars = ['Course'])
```

WHAT IS TRUE ABOUT DATA VISUALIZATION?

A. DATA VISUALIZATION IS USED TO COMMUNICATE INFORMATION CLEARLY AND EFFICIENTLY TO USERS BY THE USAGE OF INFORMATION GRAPHICS SUCH AS TABLES AND CHARTS.

B. DATA VISUALIZATION HELPS USERS IN ANALYZING A LARGE AMOUNT OF DATA IN A SIMPLER WAY.

C. DATA VISUALIZATION MAKES COMPLEX DATA MORE ACCESSIBLE, UNDERSTANDABLE, AND USABLE.

D. ALL OF THE ABOVE

I. WHAT IS TRUE ABOUT DATA VISUALIZATION?

A. DATA VISUALIZATION IS USED TO COMMUNICATE INFORMATION CLEARLY AND EFFICIENTLY TO USERS BY THE USAGE OF INFORMATION GRAPHICS SUCH AS TABLES AND CHARTS.

B. DATA VISUALIZATION HELPS USERS IN ANALYZING A LARGE AMOUNT OF DATA IN A SIMPLER WAY.

C. DATA VISUALIZATION MAKES COMPLEX DATA MORE ACCESSIBLE, UNDERSTANDABLE, AND USABLE.

D. ALL OF THE ABOVE

- DATA VISUALIZATION IS USED TO COMMUNICATE INFORMATION CLEARLY AND EFFICIENTLY TO USERS BY THE USAGE OF INFORMATION GRAPHICS SUCH AS TABLES AND CHARTS. IT HELPS USERS IN ANALYZING A LARGE AMOUNT OF DATA IN A SIMPLER WAY. IT MAKES COMPLEX DATA MORE ACCESSIBLE, UNDERSTANDABLE, AND USABLE.

DATA CAN BE VISUALIZED USING?

- A. GRAPHS
- B. CHARTS
- C. MAPS
- D. ALL OF THE ABOVE

•

- DATA VISUALIZATION IS USED TO COMMUNICATE INFORMATION CLEARLY AND EFFICIENTLY TO USERS BY THE USAGE OF INFORMATION GRAPHICS SUCH AS TABLES AND CHARTS. IT HELPS USERS IN ANALYZING A LARGE AMOUNT OF DATA IN A SIMPLER WAY. IT MAKES COMPLEX DATA MORE ACCESSIBLE, UNDERSTANDABLE, AND USABLE.

DATA CAN BE VISUALIZED USING?

- A. GRAPHS
- B. CHARTS
- C. MAPS
- D. ALL OF THE ABOVE

•

- DATA VISUALIZATION IS USED TO COMMUNICATE INFORMATION CLEARLY AND EFFICIENTLY TO USERS BY THE USAGE OF INFORMATION GRAPHICS SUCH AS TABLES AND CHARTS. IT HELPS USERS IN ANALYZING A LARGE AMOUNT OF DATA IN A SIMPLER WAY. IT MAKES COMPLEX DATA MORE ACCESSIBLE, UNDERSTANDABLE, AND USABLE.

DATA VISUALIZATION
IS ALSO AN ELEMENT
OF THE
BROADER _____
_____.

- A. DELIVER PRESENTATION ARCHITECTURE
- B. DATA PRESENTATION ARCHITECTURE
- C. DATASET PRESENTATION ARCHITECTURE
- D. DATA PROCESS ARCHITECTURE

DATA VISUALIZATION IS ALSO AN ELEMENT OF THE BROADER _____.

- A. DELIVER PRESENTATION ARCHITECTURE
- B. DATA PRESENTATION ARCHITECTURE
- C. DATASET PRESENTATION ARCHITECTURE
- D. DATA PROCESS ARCHITECTURE

Data visualization is also an element of the broader data presentation architecture (DPA) discipline, which aims to identify, locate, manipulate, format and deliver data in the most efficient way possible.

WHICH METHOD
SHOWS
HIERARCHICAL
DATA IN A
NESTED FORMAT?

A. TREEMAPS

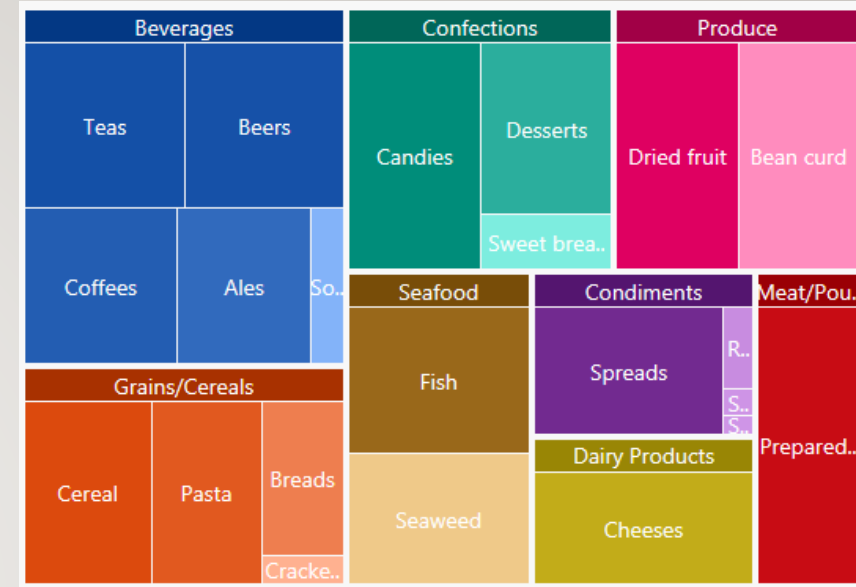
B. SCATTER PLOTS

C. POPULATION PYRAMIDS

D. AREA CHARTS

WHICH METHOD SHOWS HIERARCHICAL DATA IN A NESTED FORMAT?

- A. TREEMAPS
- B. SCATTER PLOTS
- C. POPULATION PYRAMIDS
- D. AREA CHARTS



EXPLANATION:

Treemaps are best used when multiple categories are present, and the goal is to compare different parts of a whole.

WHICH OF THE
FOLLOWING IS
FALSE?

- A. Data visualization include the ability to absorb information quickly
- B. Data visualization is another form of visual art
- C. Data visualization decrease the insights and take solwer decisions
- D. None of the above

WHICH OF THE
FOLLOWING IS
FALSE?

- A. DATA VISUALIZATION INCLUDE THE ABILITY TO ABSORB INFORMATION QUICKLY
- B. DATA VISUALIZATION IS ANOTHER FORM OF VISUAL ART
- C. DATA VISUALIZATION DECREASE THE INSIGHTS AND TAKE SOLWER DECISIONS**
- D. NONE OF THE ABOVE

EXPLANATION: DATA VISUALIZATION DECREASE THE INSIGHTS AND TAKE SOLWER DECISIONS IS FALSE STATEMENT.