Thank you. Your test submitted.

You have cleared this assessment.

Obtained Percentage	Obtained Marks
78.57 %	11 / 14

Best Attempt Score:78.57 % on 22-03-2025

Which of the following is/are correct way(s) of detecting outliers in a dataset?

1. Values that are 1.5 IQR times lesser than the 1st quartile
2. Values that are 1.5 IQR times greater than the 3rd quartile
3. Values that are 2 times greater or le

- (●) 1 and 2
- ( ) Only 1
- ( ) Only 2
- ( ) Only 3
- ( ) 1, 2 and 3

Warning

This operation is disabled.

Ok

Which of the following statement is true?

- ○ Outliers should always be replaced by the m
- ◉ For a right skewed data mean is greater thar
- ○ A normalized vertical box plot has no outlier
- ○ Outliers should always be replaced by the m

Warning

This operation is disabled.

To visualise growth in population of a country over time, which of the following graph will be the most suitable?

- (●) Line chart
- ( ) Bar graph
- ( ) Histogram
- ( ) Pie chart

Warning

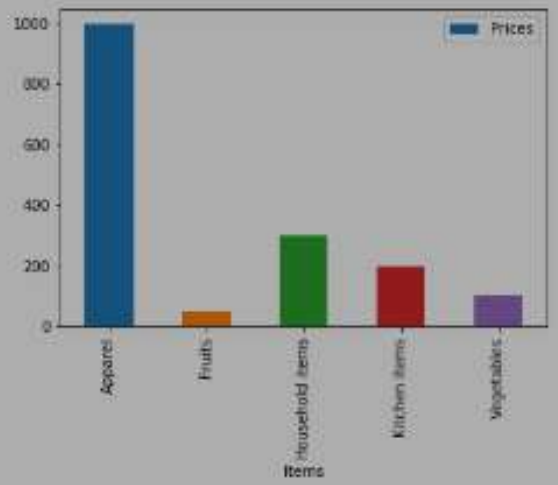
This operation is disabled.

Consider the below dataframes shopping_items_df and shopping_prices_df respectively.

	Id	Items
0	5001	Apparel
1	5002	Fruits
2	5003	Household items
3	5004	Kitchen items
4	5005	Vegetables

	Id	Prices
0	5001	1000
1	5002	50
2	5003	300
3	5004	200
4	5005	100

Which of the following options give the below shown output between Prices and Items?



1. `now_df = pd.merge(shopping_items_df,shopping_prices_df, on="Id" )`
   `now_df.plot.bar(x='Items',y="Prices")`

2. `now_df = shopping_items_df.merge(shopping_prices_df, on="Id" )`
   `now_df.barplot(x='Items',y="Prices")`

3. `now_df = shopping_prices_df.merge(shopping_items_df, on="Id" )`
   `now_df.plot (x='Items',y="Prices",kind="bar")`

4. `now_df = shopping_items_df.merge(shopping_prices_df, on="Id" )`
   `now_df[["Id","Prices"]].plot()`

- (●) 1 and 3
- ( ) 2 and 3
- ( ) 3 and 4
- ( ) 1 and 4

Consider the following iris_df dataframe from iris dataset:

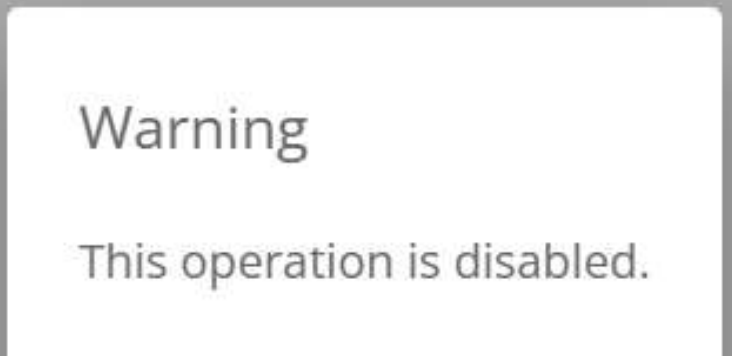
Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
6.3	3.3	6.0	2.5	virginica
5.8	2.7	5.1	1.9	virginica
7.1	3.0	5.9	2.1	virginica
6.3	2.9	5.6	1.8	virginica
6.5	3.0	5.8	2.2	virginica
7.0	3.2	4.7	1.4	versicolor
6.4	3.2	4.5	1.5	versicolor
6.9	3.1	4.9	1.5	versicolor
5.5	2.3	4.0	1.3	versicolor
6.5	2.8	4.6	1.5	versicolor

Which of the following commands will return the mean of all the numerical columns with resp

- ○ iris_df["Species"].describe()
- ○ iris_df["Species"].mean()
- ○ map(mean,iris_df["Species"])
- ◉ iris_df.groupby("Species").mean()
- ○ iris_df.iloc[:,:4].groupby("Species").mean()

Which of the following is shown when the describe method of the dataframe is invoked?

- (●) Mode
- ( ) Variance
- ( ) Count of unique values
- ( ) Median



Consider the below code. What will be dimensions of df?

```
CustomerID=np.arange(1,7)
Product=np.array(['Toaster','Toaster','Toaster','Radio','Radio','Radio'])
df1 = pd.DataFrame({'CustomerID':CustomerID, 'Product':Product})
CustomerID=np.array([2,4,6])
State=np.array(['Alabama','Alabama','Ohio'])
df2 = pd.DataFrame({'CustomerID':CustomerID, 'Sta
df = pd.merge(df1,df2)
df.shape
```

Warning

This operation is disa

- ○ 3 Rows and 2 Columns
- ◉ 3 Rows and 3 Columns
- ○ 6 Rows and 2 Columns
- ○ 6 Rows and 3 Columns

Which of the following options best describes Exploratory Data Analysis?

- ○ Increase the number of features
- ○ Creating new data
- ◉ Finding patterns, Detect and remove anamo
- ○ Increase the number of datapoints

Warning

This operation is disabled.

Which of the following represents the relation between a quantitative and a categorical attribute?

- ○ Two dimensional box plot
- ◉ Scatter plot
- ○ Histogram
- ○ Pie chart

Warning

This operation is disabled.

Below is the data frame containing the data points about the applicants for a job. The data frame is named application_df. Which of the following option helps to sort the application_df by Marks_Secured?

	ApplicationId	Gender	Domicile	Mark_Secured	Percentage
0	A1001	M	Yes	167.25	
1	A1002	F	No	201.67	
2	A1003	M	No	365.85	
3	A1004	M	Yes	307.50	
4	A1005	F	Yes	223.75	

- ○ sort(application_df).by("Mark_Secured")
- ○ application_df.sort(by="Mark_Secured")
- ◉ application_df.sort_values(by="Mark_Secured")
- ○ sorted(application_df)

Which of the following is/are performed as part of data cleaning?

- ○ Removing outliers
- ○ Filling missing values
- ○ Removing missing values
- ○ Changing column names
- ◉ All of the given options

Warning

This operation is disabled.

Study the below employee_df dataframe.

Id	name	Salary
101	Swetha	3000.0
102	Sekhar	NaN
103	NaN	3400.0
104	Shadab	NaN
105	Aishwarya	500.0

Which of the following statements gives the out

Id	name	Salary
False	False	False
False	False	True
False	True	False
False	False	True
False	False	False

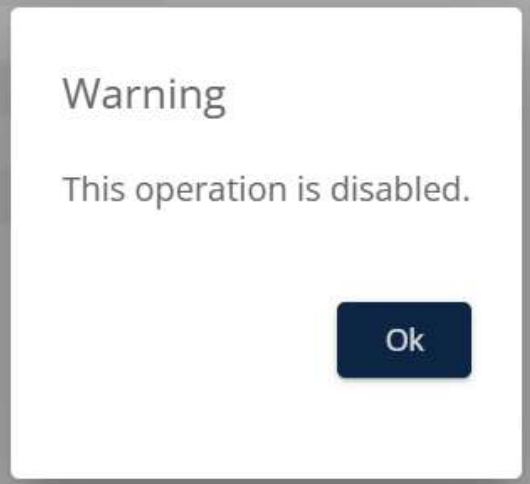
Warn

This op

- (●) employee_df.isna()
- ( ) employee_df.isNAN()
- ( ) employee_df[employee_df != np.Null]
- ( ) employee_df[employee_df!=None]

Which of the following options helps to convert categorical data to numerical data on the application_df dataframe given below?

	ApplicationId	Gender	Domicile	Mark_Secured	Percentage
0	A1001	M	Yes	167.25	p<=50
1	A1002	F	No	201.67	
2	A1003	M	No	365.85	
3	A1004	M	Yes	307.50	
4	A1005	F	Yes	223.75	



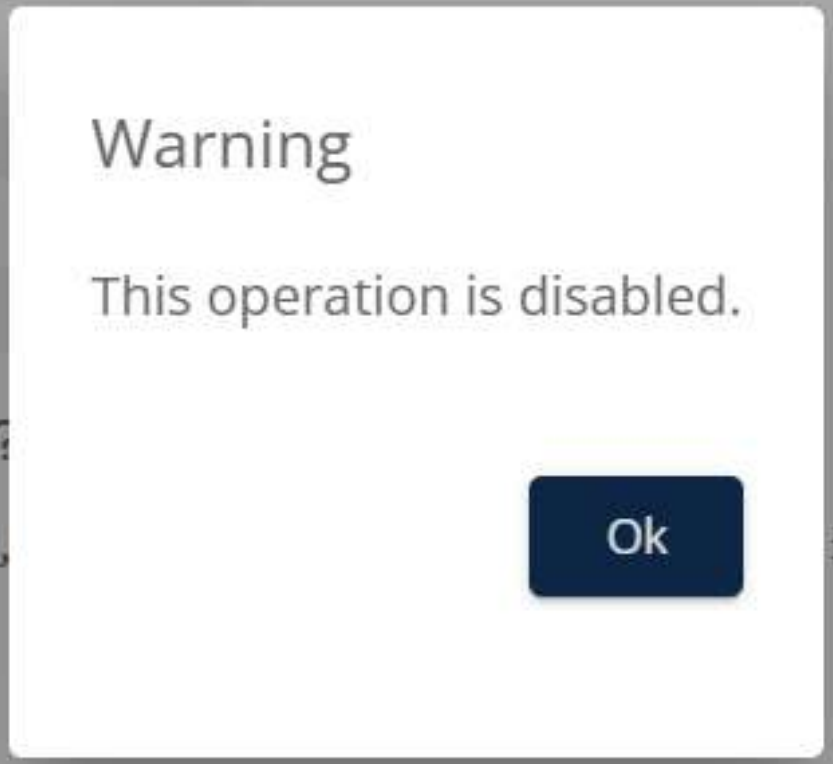
- (●) pd.get_dummies(application_df)
- ( ) application_df.one_hot_encoding()
- ( ) application_df.get_dummies()
- ( ) application_df.to_numerical()

Consider the below dataframe application_df:

	ApplicationId	Gender	Domicile	Mark_Secured	Percentage
0	A1001	M	Yes	167.25	p<=50
1	A1002	F	No	201.67	
2	A1003	M	No	365.85	
3	A1004	M	Yes	307.50	
4	A1005	F	Yes	223.75	

What is the output of the following code?

```
pd.crosstab(index = application_df["Percentage"],                    iloc[0,0]
```



- ○ 1
- ○ 2
- ○ 3
- ◉ Error