Python Pandas

1. Name two data structures in Pandas.
2. Name the two libraries that we generally import while working with Pandas.
3. What is the other name for library?
4. What is a Series?
5. Name the two components of Series.
6. What is the other name for index?
7. Name the function used to create a Series type object.
8. Define DataFrame.
9. Name the data structure in Pandas which can store data of different types.
10. Can a Series store data of different types?
11. Write the major characteristics of DataFrame.
12. How can we call the two axes of a DataFrame?/What does axis=0 and axis=1 means?
13. Is DataFrame both value mutable and size mutable?
14. Give the syntax for creating a DataFrame.
15. What are the different ways of passing data while creating a DataFrame?

1. Write the use of the following attributes of DataFrame.

   i. index ii. columns iii. axes iv. dtypes v. size
   vi. shape vii.valuesviii.empty ix.ndim x. T

   (or) Name the attribute of the DataFrame which returns the row names with datatype. Etc.,
2. What are the two ways to access the column of a DataFrame.Give syntax.
3. Give syntax to access multiple columns of a DataFrame.
4. Look at the following DataFrame and find the errors in the code given:

   df:

	Population	**Avg_Income**	**Per_Capita_Income**
Delhi	10927986	789456123	6.12354
Mumbai	12691836	652314987	5.23486
Kolkata	4631392	258963147	8.62495

   Code: df.'Population'
   df[Population]
   df['Population','Avg_Income']

5.Write code to create the above DataFrame.

1.Give syntax to access a row of a DataFrame.
2.Give syntax to access a subset of a DataFrame.
3.Write the difference in the output of the following code:

```
df.loc['Delhi']
df.loc['Delhi',:]
```

4.Find the output:

```
df.loc['Delhi':'Mumbai']
df.loc['Delhi':'Mumbai','Population':'Avg_Income']
df.loc['Delhi':'Mumbai',:]
```

1. What is the use of iloc? Write its expansion.
2. Find the output:
   df.iloc[0:2,1:2]
3. How to access an individual value from a DataFrame?
4. Find the output: df.Population[0]
5. Find the output: df.Population[0]=11875546
   print(df)
6. Write the difference between the following code:
   df.at['Bangalore']=[2342343,4545758,4.34556]
   df.at['Bangalore',:]=[2342343,4545758,4.34556]
   df.loc['Bangalore']=[2342343,4545758,4.34556]
7. Find the output:
   df['Density']=[1400,3445,3876]
   print(df)

1. Find the output:
   (i) del df['Density']
   print(df)
   (ii) df1= df.pop('Density')
   print(df)
   print(df1)
   (iii) df.drop([0,1])
   print(df)
   (iv) df.drop(['Delhi','Mumbai'])
   print(df)
   (v) df.drop(['Population'],axis=1)
   print(df)
2. Write the use of iteritems() and iterrows()
3. Find the output:
   (i) for x,y in df.iteritems():
   print(x,y)
   (ii) for x,y in df.iterrows():
   print(x,y)

1. What is the use of min() and max() functions.Write the syntax and explain its arguments.
2. Find the output:
   print(df.min())
   print(df.min(axis=1))
   print(df.max())
   print(df.max(axis=1))
3. What will happen if we apply min()/max() function to the name column of a DataFrame.

1. df

	Eng	Maths	Tamil
2018	75	95	75
2019	99	99	NaN
2020	98	95	95

Find the output:

```
print(df.mean())
print(df.mean(axis=1))
print(df.mean(axis=1,skipna=False))
print(df.median())
print(df.median(axis=1))
print(df.mode())
print(df.mode(axis=1))
```

1. Find the Output:

```
print(df.count())
print(df.count(axis=1))
print(df.sum())
print(df.sum(axis=1))
```

2. What is the default value of the argument skipna in aggregate functions.
3. What is the default value of the argument numeric_only in aggregate functions.
4. What is the difference in giving True/False to numeric _only argument in aggregate functions.
5. What is the use of min_count() argument in sum().

1. What is quantile?
2. What is the use of quantile() function?
3. How quantile is related to percentile?
4. Expand IQR.How can we calculate IQR?
5. How the 2-quantile, 3-quantile and 4-quantile are called?
6. What is the other name for tercile?
7. What are the other names for IQR?
8. What do T and Q denote in quantile?
9. Write the code to calculate the quartile for the above DataFrame of both row-wise and column-wise.
10. What is the default value of q in quantile() function?
11. Write code to calculate the variance for the above DataFrame df both row-wise and column-wise.
12. Find the output:

(i)
```
import pandas as pd
s=pd.Series([1,2,4,5,6,8,10,12,16,20])
r=s.quantile(0.3)
print(r )
```

(ii)
```
import pandas as pd
df=DataFrame(np.array([[11,1],[12,10],[13,100],[14,100],[15,1000]]),columns=['a','b'])
r=df.quantile(0.2)
print( r)
```

(iii)
```
import pandas as pd
d={'Age':[2,5],'Score':[8,7]}
print(df)
print(df.var())
```

1. Find the output:
```
print(df['Eng'].min())
print(df[['Eng','Tam']].count())
print(df.loc[2018].max())
print(df.loc[2019:2020,'Maths':'Tam'].max())
```
2. What is pivoting/data pivoting?

1. What pivoting does?

1. Give syntax for pivot() function and explain its arguments.
2. What will happen in the pivoted table if it do not have a matching entry for the given row and column in original table?
3. What does the pivot() function returns?

1. What does the pivot() function does?
2. What will happen if we skip the 'values' argument in pivot() function?

1. When does the pivot() function raises error even when the syntax is correct? What is the solution for that?
2. When should we use pivot_table function?
3. How does pivot-table differ() from pivot()?
4. Give syntax for pivot_table() and explain its arguments.
5. What will happen if we skip the 'argfunc' argument in pivot_table() function?

1. Can we skip 'columns' argument in pivot_table()?
2. Write program to create and display the following dataframe(df):

Item	**Company**	**Rupees**	**USD**
TV	LG	12000	700
TV	VIDEOCON	10000	650
TV	LG	15000	800
AC	SONY	14000	750

3. Find the output:
```
df.pivot_table(index='Item','Columns='Company',Values='Rupees')
print(df)
```
4. Write code

(i) To compute total rupees per item.
(ii) To compute number of company per item
(iii) To compute total USD per company
(iv) To compute total Rupees on two fields Item,Company.