

## PANDAS

### Sample Viva questions:

- 1) What is stack and its types?
- 2) Dataframe Vs series.
- 3) Methods to reshape the dataframe? (Or) how to add an Index, row, or column to a Pandas DataFrame?
- 4) what is the Flatten method in pandas?
- 5) Difference between sort values and sort index.
- 6) Name some methods of pandas that are useful to handle the missing values.
- 7) What is the difference between loc and iloc in Pandas?
- 8) Mention the different Types of Data structures in pandas?
- 9) Difference between merge,join,concat?
- 10) How to iterate over a Pandas DataFrame?
- 11) How to add an Index, row, or column to a Pandas DataFrame?
- 12) How will you slice rows in a panda DataFrame?
- 13) How can we get the statistical summary of data in a panda DataFrame?
- 14) How can you check if a DataFrame is empty in pandas?
- 15) How will you get the number of rows and columns of a DataFrame in pandas?
- 16) What is use of Reindexing in pandas?
- 17) What is use of GroupBy objects in Pandas?
- 18) What is Vectorization in pandas?
- 19) How to convert a DataFrame to an array in Pandas?
- 20) What is multiple Indexing?
- 21) Difference between Pandas Series and NumPy Arrays
- 22) Difference between Groupby and pivot.

### List of some Pandas functions:

Sl.No.	Attribute name	Purpose
1	df.index	to display row labels
2	df.columns	to display column labels
3	df.dtypes	to display data type of each column in the DataFrame
4	df.shape	to display a tuple representing the dimensionality of the DataFrame (rows,columns)
5	df.size	Returns number of elements
6	df.head(n)	to display the first n rows in the DataFrame
7	df.tail(n)	to display the last n rows in the DataFrame
8	df.index	Describe Index
9	df.info()	Info on Dataframe
10	df.count()	Number of non-NA values
11	df.sum()	Sum of values
12	df.cumsum()	Cumulative sum of values
13	df.min() / Df.max()	Minimum/maximum values
14	df.idxmin() / Df.idxmax()	Minimum/maximum index values
15	df.describe()	Summary statistics
16	df.mean()/df.median()	Mean /median of values
17	df.apply(function)	Apply function
18	df.applymap(function)	Apply function element wise
19	df.drop()	
20	df.sort_index(axis=0/1)	Sort by labels along an axis / Sorting by an axis
21	Df.sort_values(by=coln_name,axis=0/1)	Sorting by values along an axis
22	df.T	Transposing the data
23	df.rank()	Assign ranks to entries
24	pd.melt(dataframe,id_vars=[coln_name],value_vars=[coln_name],value_name=coln_name)	Gather columns into rows
25	df.iteritems()	(column index,series) pairs
26	df.iterrows()	(row index,series)Pairs
27	df.dropna()	Drop NaN values
28	df.fillna()	Fill NaN values with a predetermined value
29	df.replace('a','b')	Replace values with others

30	<code>df.groupby(by='coln_name').mean()</code>	Group the dataframe using specific column
31	<code>df.set_index()</code>	Set the index
32	<code>df.reset_index()</code>	Reset the index
33	<code>df.reindex(new_index)</code>	Conform Series/DataFrame to new index
34	<code>df.duplicated()</code>	Check row duplicates
35	<code>df.duplicated([coln name])</code>	Check for specific column duplications
36	<code>df.drop_duplicates()</code>	Drop duplicates
37	<code>pd.merge(df1,df2,how='left/right/inner/outer',on='coln name')</code>	Combining 2 dataframes
38	<code>Df1.join(df2,how='left/right/inner/outer')</code>	Join columns with other DataFrame either on index or on a key column
39	<code>pd.concat([df1, df3], axis=1,join="inner/outer")</code>	Concatenate pandas objects along a particular axis with optional set logic along the other axes.
40	<code>pd.read_csv(filename)</code>	Read a csv file / import csv data
41	<code>df.to_csv(filename)</code>	Write to a CSV file
42		
43	<code>pd.isnull()</code>	Checks for null Values
44	<code>df.rename(columns={'old_name': 'new_name'})</code>	Renaming the column name
45	<code>df1.append(df2)</code>	Add the rows in df1 to the end of df2 (columns should be identical)
46	<code>df.corr()</code>	Returns the correlation between columns in a DataFrame
47	<code>df.std()</code>	Returns the standard deviation of each column
48	<code>df['coln name'].value_counts()</code>	Count number of rows with each unique value of variable
49	<code>len(df)</code>	Number of rows in DataFrame
50	<code>df['coln name'].nunique()</code>	Number of distinct values in a column
51	<code>df.stack()</code>	Pivot column level to index, i.e. "stacking the columns"
52	<code>df.unstack()</code>	Pivot index level to columns, "unstack the columns"
53	<code>pd.get_dummies(df, prefix=['col1', 'col2'])</code>	Convert categorical variable into dummy/indicator variables.
54	<code>pd.isna(df)</code>	Detect missing values for an array-like object.
55	<code>pd.to_numeric(arg)</code>	Convert argument to a numeric type.
56	<code>df.select_dtypes()</code>	Return a subset of the DataFrame's columns based on the column dtypes.