PANDAS

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