Q1. How do you load a CSV file into a Pandas DataFrame?

In jupyter notebook declare the pandas as pd and by using keyword read_csv command we can load a csv file.

```
[1]: import pandas as kh

[5]: kh1 = kh.read_csv('https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv')
     kh1
```

[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

891 rows × 12 columns

Q2. How do you check the data type of a column in a Pandas DataFrame?

In jupyter notebook by using the dtypes we can the data type of column

```
[9]: kh1.dtypes

[9]: PassengerId      int64
     Survived         int64
     Pclass           int64
     Name            object
     Sex             object
     Age            float64
     SibSp            int64
     Parch            int64
     Ticket          object
     Fare           float64
     Cabin           object
     Embarked        object
     dtype: object
```

Q3. How do you select rows from a Pandas DataFrame based on a condition?

By using loc and iloc function we can access the rows

Q4. How do you rename columns in a Pandas DataFrame?

The columns can be renamed by using the rename () function in the pandas dataframe

Q5. How do you drop columns in a Pandas DataFrame?

By using the drop () function we can delete the columns in pandas dataframe

```
[30]: pk1.drop('Parch' , axis = 1)
```

[30]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	373450	8.0500	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	211536	13.0000	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	112053	30.0000	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	W./C. 6607	23.4500	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	111369	30.0000	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	370376	7.7500	NaN	Q

891 rows × 11 columns

Q6. How do you find the unique values in a column of a Pandas DataFrame?

By using the unique values we can find the unique values in pandas dataframe

```
[39]: pk1['Pclass'].unique()

[39]: array([3, 1, 2])
```

Q7. How do you find the number of missing values in each column of a Pandas DataFrame?

By using the isnull() and sum() function we can find the number of missing values

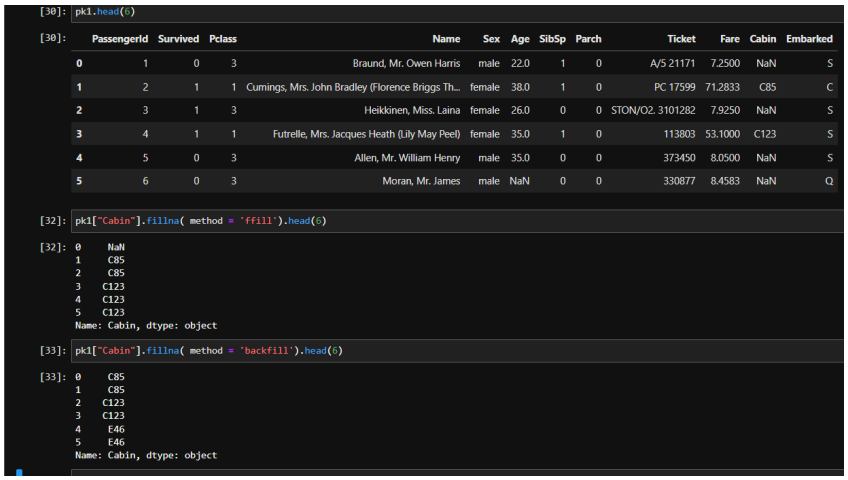
```
[41]: pk1.isnull().sum()

[41]: PassengerId      0
      Survived         0
      Pclass           0
      Name             0
      Sex              0
      Age            177
      SibSp            0
      Parch            0
      Ticket           0
      Fare             0
      Cabin          687
      Embarked         2
      dtype: int64
```

Q8. How do you fill missing values in a Pandas DataFrame with a specific value?

We can fill the missing values with fillna() function and the syntax for fillna()

Dataframe.fillna(value, method,axis,limit)



[30]:
```
pk1.head(6)
```

[30]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q

[32]:
```
pk1["Cabin"].fillna( method = 'ffill').head(6)
```

[32]:
```
0     NaN
1     C85
2     C85
3    C123
4    C123
5    C123
Name: Cabin, dtype: object
```

[33]:
```
pk1["Cabin"].fillna( method = 'backfill').head(6)
```

[33]:
```
0     C85
1     C85
2    C123
3    C123
4     E46
5     E46
Name: Cabin, dtype: object
```

Q9. How do you concatenate two Pandas DataFrames?

In pandas dataframes by using the concat() function we can concatenate two different dataframe

Q10. How do you merge two Pandas DataFrames on a specific column?

We can merge two pandas dataframes by using the join function , there always default join will be outer to make intersection inner is used

Q11. How do you group data in a Pandas DataFrame by a specific column and apply an aggregation function?

We group data in a pandas dataframe by passing the dictionary with key containing columns and value containing list of aggregate function for these specified columns and the column is not specified it will be replaced as nan in the result.

Q12. How do you pivot a Pandas DataFrame?

The pivot() function is used to reshape the pandas dataframe by using the index or the columns valuues

Q13. How do you change the data type of a column in a Pandas DataFrame?

By using the astype() function we can change the datatype of a column

Q14. How do you sort a Pandas DataFrame by a specific column?

By using the sort_values() we can sort the column by mentioning the specific columns name

Q15. How do you create a copy of a Pandas DataFrame?

We can copy the pandas dataframe by using the dataframe.copy(deep = true/false) . If deep = true any changes done on the original file it will not reflect to the copied file , if deep = false any changes done on the original file it will reflect to the copied file.

Q16. How do you filter rows of a Pandas DataFrame by multiple conditions?

We can filter the row by using the columns values and loc indexers

Q17. How do you calculate the mean of a column in a Pandas DataFrame?

To calculator the mean of a column in a pandas dataframe by using the pandas dataframe.mean() functions if don't specify the axis value it will be zero

Q18. How do you calculate the standard deviation of a column in a Pandas DataFrame?

By using the dataframe.std() function we can calculate the standard deviation of a column if axis value is not specified it will remain as zero and if want to skip the nan value we are using the skipna = true

Q19. How do you calculate the correlation between two columns in a Pandas DataFrame?

To calculate the correlation between two columns in a pandas dataframe by using the corr() function

Q20. How do you select specific columns in a DataFrame using their labels?

We can select the specific columns using the [ ] operators

Q21. How do you select specific rows in a DataFrame using their indexes?

By using loc function we can select specific rows in a DataFrame

Q22. How do you sort a DataFrame by a specific column?

By using the sort_values() we can sort the column by mentioning the specific columns name

Q23. How do you create a new column in a DataFrame based on the values of another column?

By using the dataframe.apply() and dataframe.map() we can create a new column in a dataframe

Q24. How do you remove duplicates from a DataFrame?

We can remove duplicates from a dataframe by using the drop_duplicates function

Q25. What is the difference between .loc and .iloc in Pandas?

Iloc is the system generated index and loc is the user generated index . If the user change the any columns as index the index value will change accordings .