Name :- Manthan Joshi

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

To load a CSV file into pandas dataframe, we have to use "read_csv".

```
import pandas as pd
df = pd.read_csv("file path")
print(df.head())
```

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

To find the datatype of a column we have to use "dtypes".

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape","Orange"], "Cost_per_kg":[100, 80, 60, 90]}
df = pd.DataFrame(x)
print(df.dtypes)
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape","Orange"], "Cost_per_kg":[100, 80, 60, 90]}
df = pd.DataFrame(x)
result = df[df["Cost_per_kg"] > 60]
print(result)
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape","Orange"], "Cost_per_kg":[100, 80, 60, 90]}
df = pd.DataFrame(x)

df.rename(columns = {'Cost_per_kg':'Cost_per_kilogram'}, inplace = True)
print(df)
```

Name :- Manthan Joshi

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape","Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df.rename(columns = {'Cost_per_kg':'Cost_per_kilogram'}, inplace = True)

df.drop(["Fruits_From", "Demand"], axis = 1, inplace=True)
print(df)
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape","Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df.rename(columns = {'Cost_per_kg':'Cost_per_kilogram'}, inplace = True)

df2 = df['Demand'].unique()
print(df2)
```

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

```
from numpy import NaN
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", NaN], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
```

Name :- Manthan Joshi

```
df = pd.DataFrame(x)
df.rename(columns = {'Cost_per_kg':'Cost_per_kilogram'}, inplace = True)

print(df["Fruits"].isnull().sum())
print(df["Cost_per_kilogram"].isnull().sum())
print(df["Fruits_From"].isnull().sum())
print(df["Demand"].isnull().sum())
```

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

```
from numpy import NaN
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", NaN], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df.rename(columns = {'Cost_per_kg':'Cost_per_kilogram'}, inplace = True)

df["Fruits"].fillna("fruit is not identified", inplace = True)
print(df)
```

Q9. How do you concatenate two Pandas DataFrames?

```
import pandas as pd
x = {"City":["Hyderabad","Pune","Bhuwaneshwar","Ahmedabad"],
"State":["Telangana","Maharastra","Orissa","Gujarat"]}
df_1 = pd.DataFrame(x)

y = {"City":["Trivendrum","Amaravati", "Chennai","Kolkata"], "State":["Kerala","Andhra
Pradesh","Tamil Nadu","West Bengal"]}
df_2 = pd.DataFrame(y)

df = pd.concat([df_1,df_2])
```

Name :- Manthan Joshi

```
print(df)
```

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

```
import pandas as pd
x = {"City":["Hyderabad","Pune","Bhuwaneshwar","Ahmedabad"],
    "State":["Telangana","Maharastra","Orissa","Gujarat"]}
df_1 = pd.DataFrame(x)

y =
{"Name":["Akash","Ravinder","Geetha","Lavanya"],"City":["Hyderabad","Pune","Bhuwaneshwar","A
hmedabad"],

"State":["Telangana","Maharastra","Orissa","Gujarat"],"Vehicle":["Car","Bike","Car","Scooty"]}
df_2 = pd.DataFrame(y)

df = df_2.merge(df_1[["City","State"]])
print(df)
```

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

```
import pandas as pd
x = {"Day":
["1/1/2017","1/1/2017","1/1/2017","1/1/2017","1/1/2017","1/1/2017","1/1/2017","1/1/2017","1/1
/2017","1/1/2017","1/1/2017","1/1/2017"],

"City":["Newyork","Newyork","Newyork","Newyork","Mumbai","Mumbai","Mumbai","Mumbai","Pa
ris","Paris","Paris","Paris"],
    "Temperature":[32,36,28,33,90,85,87,92,45,50,54,42],"Wind
Speed":[6,7,12,7,5,12,15,5,20,13,8,10],

"Event":["Rain","Sunny","Snow","Cloudy","Cloudy","Cloudy","Cloudy","Rain","Rain","Fog","Fog","Sn
ow"]}
df = pd.DataFrame(x)
```

Name :- Manthan Joshi

```
for i,j in df.groupby("City"):
  print(i,j["Temperature"].max())
```

Q12. How do you pivot a Pandas DataFrame?

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
print(df.pivot(index="Fruits_From",columns = "Demand"))
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df["Cost_per_kg"] = df["Cost_per_kg"].apply(pd.to_numeric)
print(df['Cost_per_kg'].dtypes)
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df.sort_values("Fruits")
```

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

Name :- Manthan Joshi

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
df_2 = df.copy()
print(df_2)
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
display(df.loc[(df['Cost_per_kg']>=80) & (df['Fruits_From']=="Kadapa")])
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
print(df["Cost_per_kg"].mean())
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
```

Name :- Manthan Joshi

```
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }

    df = pd.DataFrame(x)

    print(df["Cost_per_kg"].std())
```

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

```
    import pandas as pd

    x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
"Last_Month_Price":[120,60,40,100],

    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }

    df = pd.DataFrame(x)

    print(df["Cost_per_kg"].corr(df["Last_Month_Price"]))
```

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

```
    import pandas as pd

    x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
"Last_Month_Price":[120,60,40,100],

    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }

    df = pd.DataFrame(x)

    print(df["Fruits"])
```

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

```
    import pandas as pd

    x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90],
"Last_Month_Price":[120,60,40,100],

    "Fruits_From": ["Kashmir", "Kadapa", "Shimla",
"Banglore"],"Demand":["High","Medium","Medium","Low"] }

    df = pd.DataFrame(x)
```

Name :- Manthan Joshi

```
print(df.iloc[0])
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90], "Last_Month_Price":[120,60,40,100],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla", "Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
print(df.sort_values(by="Fruits"))
```

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

```
import pandas as pd
x = {"Fruits":["Apple", "Banana","Grape", "Orange"], "Cost_per_kg":[100, 80, 60, 90], "Last_Month_Price":[120,60,40,100],
    "Fruits_From": ["Kashmir", "Kadapa", "Shimla", "Banglore"],"Demand":["High","Medium","Medium","Low"] }
df = pd.DataFrame(x)
print(df.loc[:,"Fruits"])
```

Q24. How do you remove duplicates from a DataFrame?

```
import pandas as pd

boxes = {'Fruits': ['Apple','Apple','Apple','Banana','Banana','Orange','Orange','Orange'],
      'Demand': ['Medium','Low','Medium','High','Low','Low','Medium','High']
     }
df = pd.DataFrame(boxes, columns = ['Fruits', 'Demand'])

print(df)
```

Name :- Manthan Joshi

```
print()
print(df.drop_duplicates())
```

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

loc is label-based, which means that we have to specify rows and columns based on their row and column labels.

iloc is integer position-based, so we have to specify rows and columns by their integer position values.