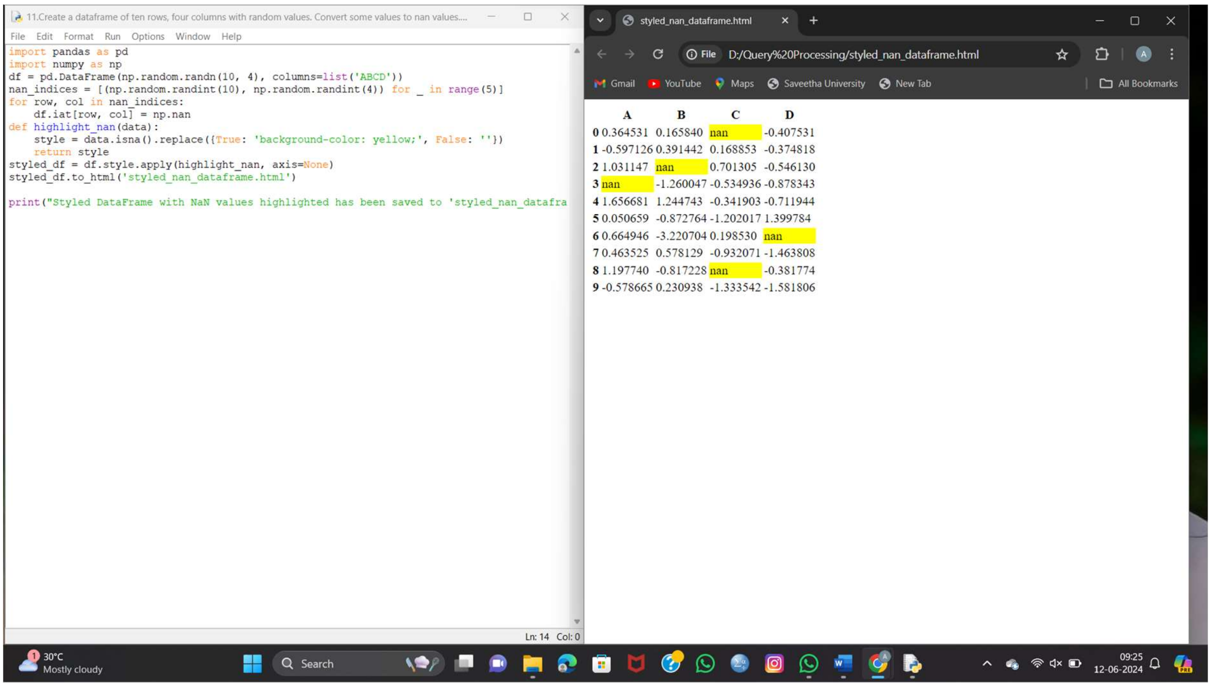
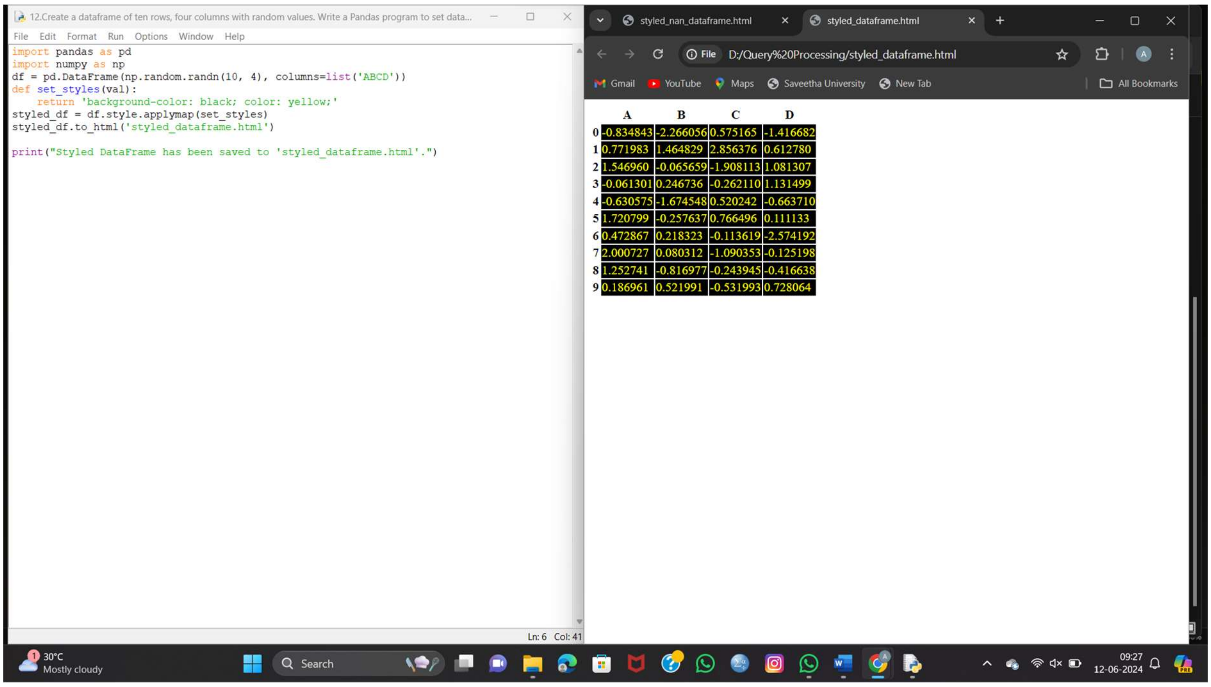
11.Create a dataframe of ten rows, four columns with random values. Convert some values to nan values. Write a Pandas program which will highlight the nan values.


11.Create a dataframe of ten rows, four columns with random values. Convert some values to nan values....
File Edit Format Run Options Window Help
styled_nan_dataframe.html
File D:/Query%20Processing/styled_nan_dataframe.html
Gmail YouTube Maps Saveetha University New Tab All Bookmarks
Ln: 14 Col: 0
30°C Mostly cloudy
Search
09:25 12-06-2024

```python
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.randn(10, 4), columns=list('ABCD'))
nan_indices = [(np.random.randint(10), np.random.randint(4)) for _ in range(5)]
for row, col in nan_indices:
    df.iat[row, col] = np.nan
def highlight_nan(data):
    style = data.isna().replace({True: 'background-color: yellow;', False: ''})
    return style
styled_df = df.style.apply(highlight_nan, axis=None)
styled_df.to_html('styled_nan_dataframe.html')

print("Styled DataFrame with NaN values highlighted has been saved to 'styled_nan_datafra
```

	A	B	C	D
0	0.364531	0.165840	nan	-0.407531
1	-0.597126	0.391442	0.168853	-0.374818
2	1.031147	nan	0.701305	-0.546130
3	nan	-1.260047	-0.534936	-0.878343
4	1.656681	1.244743	-0.341903	-0.711944
5	0.050659	-0.872764	-1.202017	1.399784
6	0.664946	-3.220704	0.198530	nan
7	0.463525	0.578129	-0.932071	-1.463808
8	1.197740	-0.817228	nan	-0.381774
9	-0.578665	0.230938	-1.333542	-1.581806

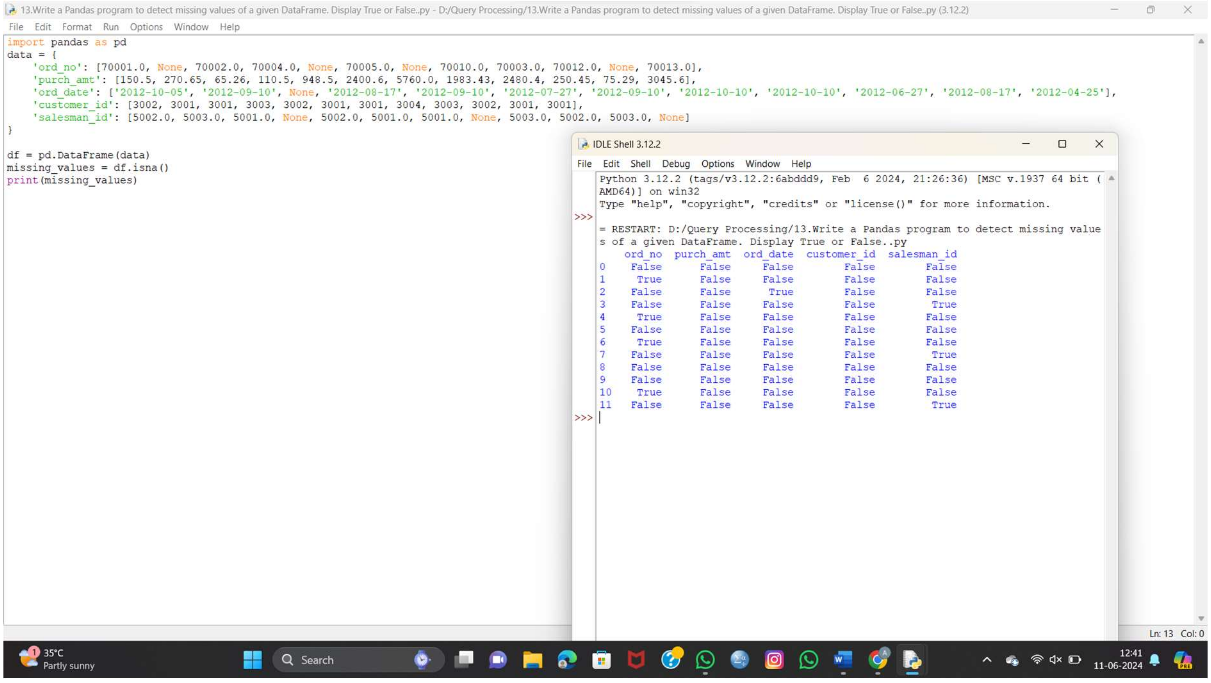
12.Create a dataframe of ten rows, four columns with random values. Write a Pandas program to set dataframe background Color black and font color yellow.


12.Create a dataframe of ten rows, four columns with random values. Write a Pandas program to set data...
File Edit Format Run Options Window Help
styled_nan_dataframe.html
styled_dataframe.html
File D:/Query%20Processing/styled_dataframe.html
Gmail YouTube Maps Saveetha University New Tab All Bookmarks
Ln: 6 Col: 41
30°C Mostly cloudy
Search
09:27 12-06-2024

```python
import pandas as pd
import numpy as np
df = pd.DataFrame(np.random.randn(10, 4), columns=list('ABCD'))
def set_styles(val):
    return 'background-color: black; color: yellow;'
styled_df = df.style.applymap(set_styles)
styled_df.to_html('styled_dataframe.html')

print("Styled DataFrame has been saved to 'styled_dataframe.html'.")
```

	A	B	C	D
0	-0.834843	-2.266056	0.575165	-1.416682
1	0.771983	1.464829	2.856376	0.612780
2	1.546960	-0.065659	-1.908113	1.081307
3	-0.061301	0.246736	-0.262110	1.131499
4	-0.630575	-1.674548	0.520242	-0.663710
5	1.720799	-0.257637	0.766496	0.111133
6	0.472867	0.218323	-0.113619	-2.574192
7	2.000727	0.080312	-1.090353	-0.125198
8	1.252741	-0.816977	-0.243945	-0.416638
9	0.186961	0.521991	-0.531993	0.728064

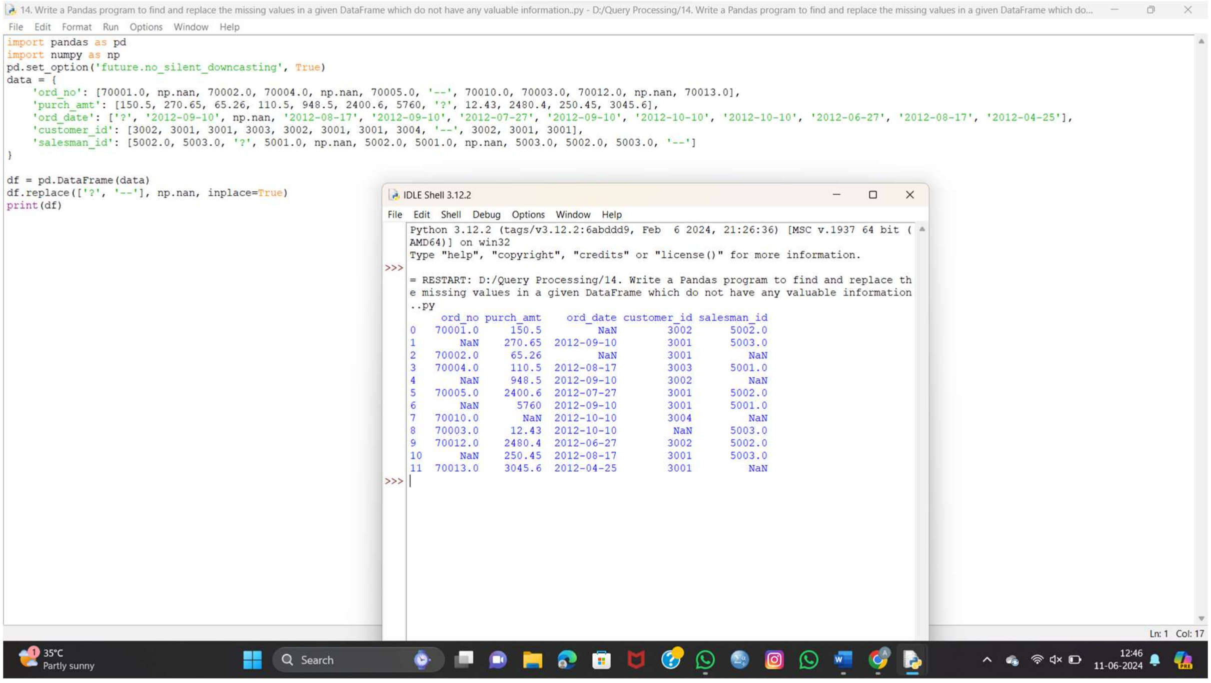
13.Write a Pandas program to detect missing values of a given DataFrame. Display True or False.


13.Write a Pandas program to detect missing values of a given DataFrame. Display True or False..py - D:/Query Processing/13.Write a Pandas program to detect missing values of a given DataFrame. Display True or False..py (3.12.2)
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 13 Col: 0
35°C
Partly sunny
Search
12:41
11-06-2024

```python
import pandas as pd
data = {
    'ord_no': [70001.0, None, 70002.0, 70004.0, None, 70005.0, None, 70010.0, 70003.0, 70012.0, None, 70013.0],
    'purch_amt': [150.5, 270.65, 65.26, 110.5, 948.5, 2400.6, 5760.0, 1983.43, 2480.4, 250.45, 75.29, 3045.6],
    'ord_date': ['2012-10-05', '2012-09-10', None, '2012-08-17', '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', '2012-04-25'],
    'customer_id': [3002, 3001, 3001, 3003, 3002, 3001, 3001, 3004, 3003, 3002, 3001, 3001],
    'salesman_id': [5002.0, 5003.0, 5001.0, None, 5002.0, 5001.0, 5001.0, None, 5003.0, 5002.0, 5003.0, None]
}

df = pd.DataFrame(data)
missing_values = df.isna()
print(missing_values)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/13.Write a Pandas program to detect missing value
s of a given DataFrame. Display True or False..py
    ord_no  purch_amt  ord_date  customer_id  salesman_id
0    False      False     False        False        False
1     True      False     False        False        False
2    False      False      True        False        False
3    False      False     False        False         True
4     True      False     False        False        False
5    False      False     False        False        False
6     True      False     False        False        False
7    False      False     False        False         True
8    False      False     False        False        False
9    False      False     False        False        False
10    True      False     False        False        False
11   False      False     False        False         True
>>> 
```

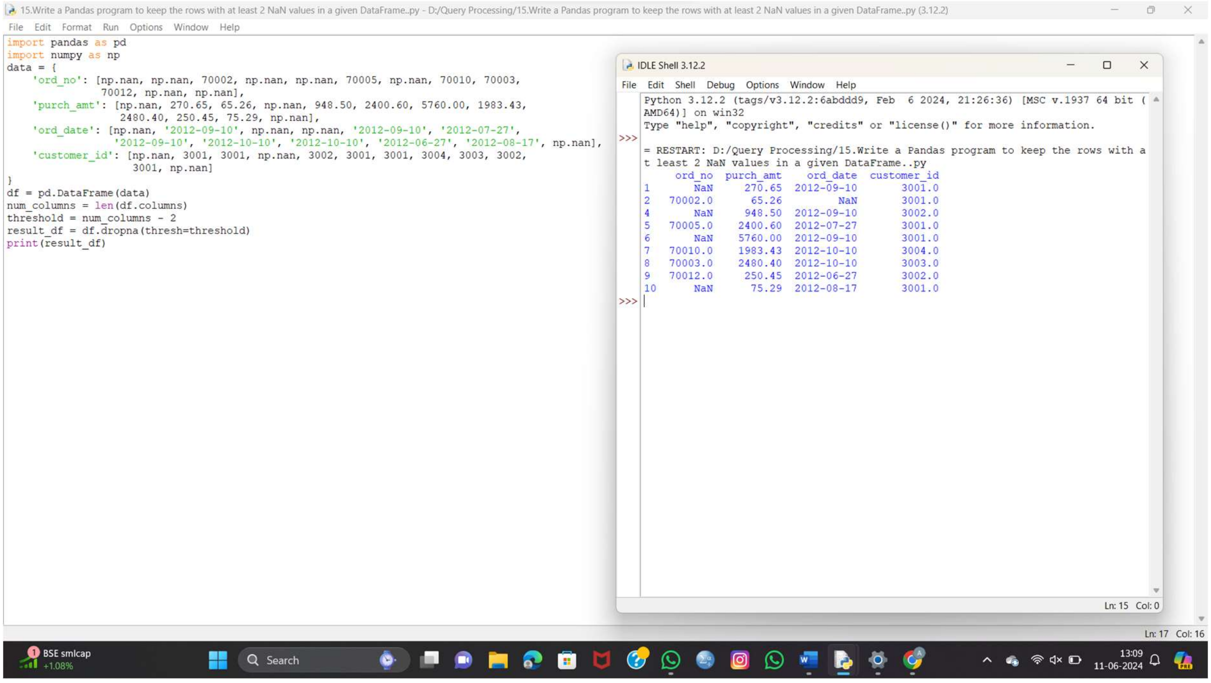
14. Write a Pandas program to find and replace the missing values in a given DataFrame which do not have any valuable information.


14. Write a Pandas program to find and replace the missing values in a given DataFrame which do not have any valuable information..py - D:/Query Processing/14. Write a Pandas program to find and replace the missing values in a given DataFrame which do...
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 1 Col: 17
35°C
Partly sunny
Search
12:46
11-06-2024

```python
import pandas as pd
import numpy as np
pd.set_option('future.no_silent_downcasting', True)
data = {
    'ord_no': [70001.0, np.nan, 70002.0, 70004.0, np.nan, 70005.0, '--', 70010.0, 70003.0, 70012.0, np.nan, 70013.0],
    'purch_amt': [150.5, 270.65, 65.26, 110.5, 948.5, 2400.6, 5760, '?', 12.43, 2480.4, 250.45, 3045.6],
    'ord_date': ['?', '2012-09-10', np.nan, '2012-08-17', '2012-09-10', '2012-07-27', '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', '2012-04-25'],
    'customer_id': [3002, 3001, 3001, 3003, 3002, 3001, 3001, 3004, '--', 3002, 3001, 3001],
    'salesman_id': [5002.0, 5003.0, '?', 5001.0, np.nan, 5002.0, 5001.0, np.nan, 5003.0, 5002.0, 5003.0, '--']
}

df = pd.DataFrame(data)
df.replace(['?', '--'], np.nan, inplace=True)
print(df)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/14. Write a Pandas program to find and replace th
e missing values in a given DataFrame which do not have any valuable information
..py
     ord_no purch_amt    ord_date customer_id salesman_id
0   70001.0     150.5         NaN        3002      5002.0
1       NaN    270.65  2012-09-10        3001      5003.0
2   70002.0     65.26         NaN        3001         NaN
3   70004.0     110.5  2012-08-17        3003      5001.0
4       NaN     948.5  2012-09-10        3002         NaN
5   70005.0    2400.6  2012-07-27        3001      5002.0
6       NaN      5760  2012-09-10        3001      5001.0
7   70010.0       NaN  2012-10-10        3004         NaN
8   70003.0     12.43  2012-10-10         NaN      5003.0
9   70012.0    2480.4  2012-06-27        3002      5002.0
10      NaN    250.45  2012-08-17        3001      5003.0
11  70013.0    3045.6  2012-04-25        3001         NaN
>>> 
```

15.Write a Pandas program to keep the rows with at least 2 NaN values in a given DataFrame.

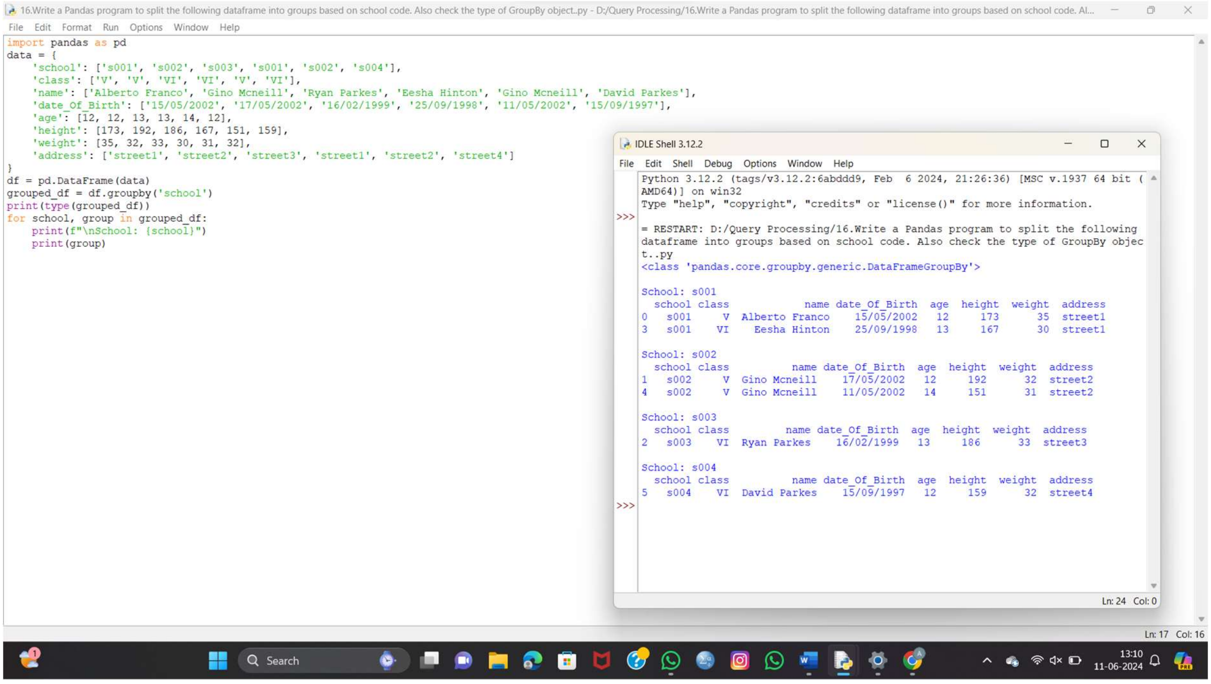


15.Write a Pandas program to keep the rows with at least 2 NaN values in a given DataFrame..py - D:/Query Processing/15.Write a Pandas program to keep the rows with at least 2 NaN values in a given DataFrame..py (3.12.2)
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 15 Col: 0
Ln: 17 Col: 16
BSE smlcap +1.08%
Search
13:09 11-06-2024

```
import pandas as pd
import numpy as np
data = {
    'ord_no': [np.nan, np.nan, 70002, np.nan, np.nan, 70005, np.nan, 70010, 70003,
               70012, np.nan, np.nan],
    'purch_amt': [np.nan, 270.65, 65.26, np.nan, 948.50, 2400.60, 5760.00, 1983.43,
                  2480.40, 250.45, 75.29, np.nan],
    'ord_date': [np.nan, '2012-09-10', np.nan, np.nan, '2012-09-10', '2012-07-27',
                 '2012-09-10', '2012-10-10', '2012-10-10', '2012-06-27', '2012-08-17', np.nan],
    'customer_id': [np.nan, 3001, 3001, np.nan, 3002, 3001, 3001, 3004, 3003, 3002,
                    3001, np.nan]
}
df = pd.DataFrame(data)
num_columns = len(df.columns)
threshold = num_columns - 2
result_df = df.dropna(thresh=threshold)
print(result_df)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/15.Write a Pandas program to keep the rows with a
t least 2 NaN values in a given DataFrame..py
     ord_no  purch_amt    ord_date  customer_id
1       NaN     270.65  2012-09-10       3001.0
2   70002.0      65.26         NaN       3001.0
4       NaN     948.50  2012-09-10       3002.0
5   70005.0    2400.60  2012-07-27       3001.0
6       NaN    5760.00  2012-09-10       3001.0
7   70010.0    1983.43  2012-10-10       3004.0
8   70003.0    2480.40  2012-10-10       3003.0
9   70012.0     250.45  2012-06-27       3002.0
10      NaN      75.29  2012-08-17       3001.0
>>>
```

16.Write a Pandas program to split the following dataframe into groups based on school code. Also check the type of GroupBy object.

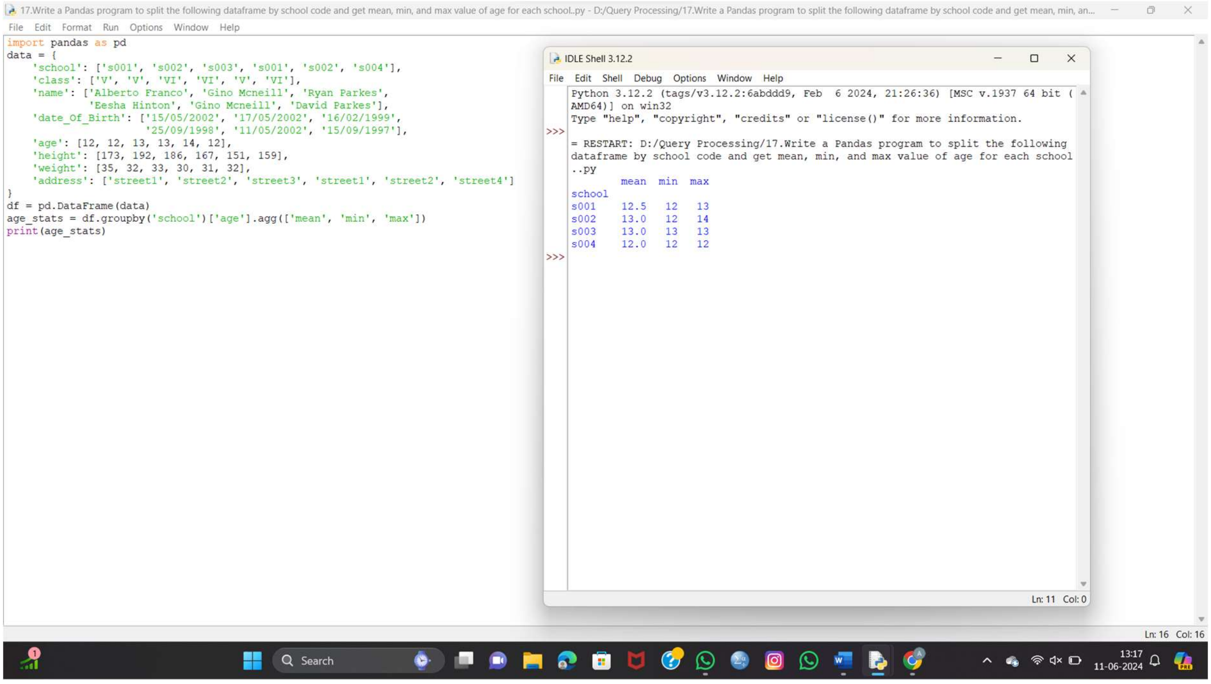


16.Write a Pandas program to split the following dataframe into groups based on school code. Also check the type of GroupBy object..py - D:/Query Processing/16.Write a Pandas program to split the following dataframe into groups based on school code. Al...
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 24 Col: 0
Ln: 17 Col: 16
Search
13:10 11-06-2024

```
import pandas as pd
data = {
    'school': ['s001', 's002', 's003', 's001', 's002', 's004'],
    'class': ['V', 'V', 'VI', 'VI', 'V', 'VI'],
    'name': ['Alberto Franco', 'Gino Mcneill', 'Ryan Parkes', 'Eesha Hinton', 'Gino Mcneill', 'David Parkes'],
    'date_Of_Birth': ['15/05/2002', '17/05/2002', '16/02/1999', '25/09/1998', '11/05/2002', '15/09/1997'],
    'age': [12, 12, 13, 13, 14, 12],
    'height': [173, 192, 186, 167, 151, 159],
    'weight': [35, 32, 33, 30, 31, 32],
    'address': ['street1', 'street2', 'street3', 'street1', 'street2', 'street4']
}
df = pd.DataFrame(data)
grouped_df = df.groupby('school')
print(type(grouped_df))
for school, group in grouped_df:
    print(f"\nSchool: {school}")
    print(group)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/16.Write a Pandas program to split the following
dataframe into groups based on school code. Also check the type of GroupBy objec
t..py
<class 'pandas.core.groupby.generic.DataFrameGroupBy'>

School: s001
  school class            name date_Of_Birth  age  height  weight  address
0   s001     V  Alberto Franco    15/05/2002   12     173      35  street1
3   s001    VI    Eesha Hinton    25/09/1998   13     167      30  street1

School: s002
  school class          name date_Of_Birth  age  height  weight  address
1   s002     V  Gino Mcneill    17/05/2002   12     192      32  street2
4   s002     V  Gino Mcneill    11/05/2002   14     151      31  street2

School: s003
  school class         name date_Of_Birth  age  height  weight  address
2   s003    VI  Ryan Parkes    16/02/1999   13     186      33  street3

School: s004
  school class          name date_Of_Birth  age  height  weight  address
5   s004    VI  David Parkes    15/09/1997   12     159      32  street4
>>>
```

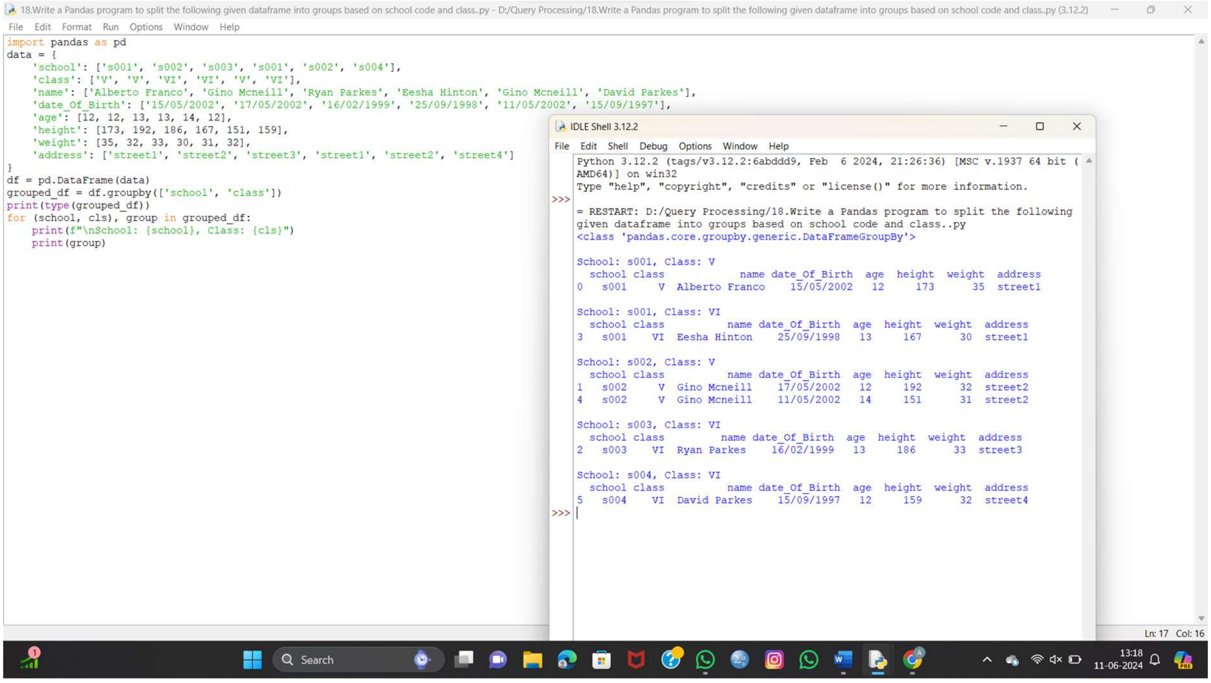
17.Write a Pandas program to split the following dataframe by school code and get mean, min, and max value of age for each school.


17.Write a Pandas program to split the following dataframe by school code and get mean, min, and max value of age for each school..py - D:/Query Processing/17.Write a Pandas program to split the following dataframe by school code and get mean, min, an...
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 11 Col: 0
Ln: 16 Col: 16
Search
13:17
11-06-2024

```
import pandas as pd
data = {
    'school': ['s001', 's002', 's003', 's001', 's002', 's004'],
    'class': ['V', 'V', 'VI', 'VI', 'V', 'VI'],
    'name': ['Alberto Franco', 'Gino Mcneill', 'Ryan Parkes',
             'Eesha Hinton', 'Gino Mcneill', 'David Parkes'],
    'date_Of_Birth': ['15/05/2002', '17/05/2002', '16/02/1999',
                      '25/09/1998', '11/05/2002', '15/09/1997'],
    'age': [12, 12, 13, 13, 14, 12],
    'height': [173, 192, 186, 167, 151, 159],
    'weight': [35, 32, 33, 30, 31, 32],
    'address': ['street1', 'street2', 'street3', 'street1', 'street2', 'street4']
}
df = pd.DataFrame(data)
age_stats = df.groupby('school')['age'].agg(['mean', 'min', 'max'])
print(age_stats)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> 
= RESTART: D:/Query Processing/17.Write a Pandas program to split the following
dataframe by school code and get mean, min, and max value of age for each school
..py
        mean  min  max
school
s001    12.5   12   13
s002    13.0   12   14
s003    13.0   13   13
s004    12.0   12   12
>>> 
```

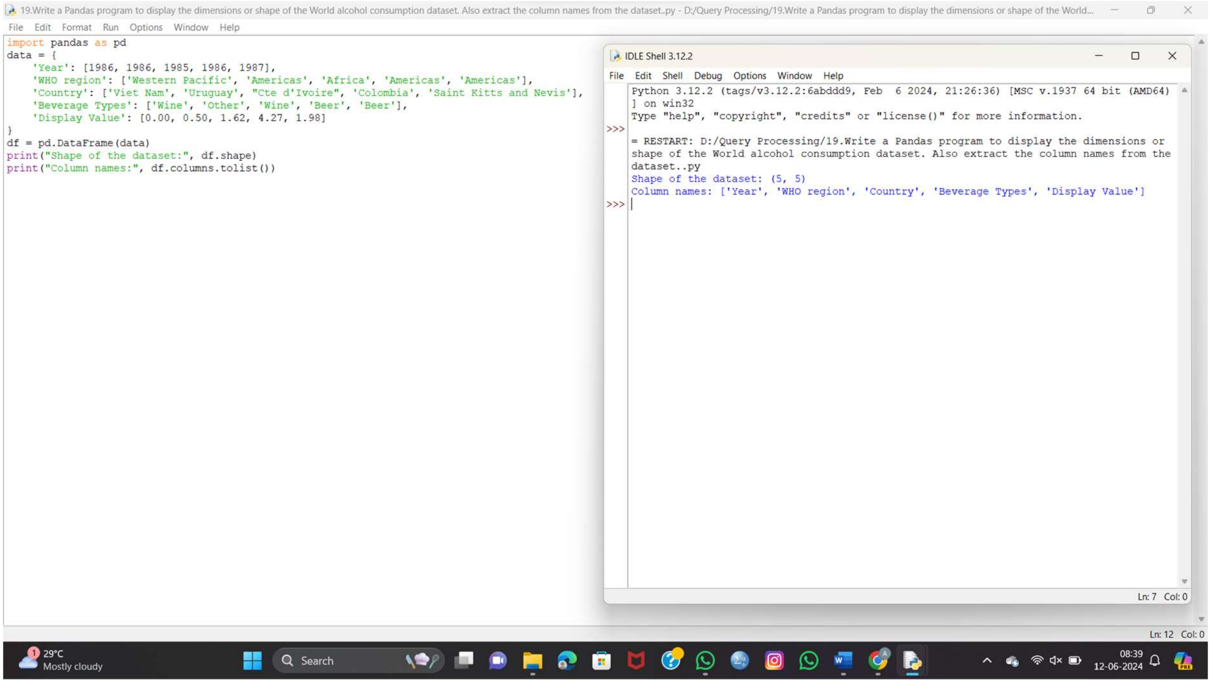
18.Write a Pandas program to split the following given dataframe into groups based on school code and class.


18.Write a Pandas program to split the following given dataframe into groups based on school code and class..py - D:/Query Processing/18.Write a Pandas program to split the following given dataframe into groups based on school code and class..py (3.12.2)
File Edit Format Run Options Window Help
IDLE Shell 3.12.2
File Edit Shell Debug Options Window Help
Ln: 17 Col: 16
Search
13:18
11-06-2024

```
import pandas as pd
data = {
    'school': ['s001', 's002', 's003', 's001', 's002', 's004'],
    'class': ['V', 'V', 'VI', 'VI', 'V', 'VI'],
    'name': ['Alberto Franco', 'Gino Mcneill', 'Ryan Parkes', 'Eesha Hinton', 'Gino Mcneill', 'David Parkes'],
    'date_Of_Birth': ['15/05/2002', '17/05/2002', '16/02/1999', '25/09/1998', '11/05/2002', '15/09/1997'],
    'age': [12, 12, 13, 13, 14, 12],
    'height': [173, 192, 186, 167, 151, 159],
    'weight': [35, 32, 33, 30, 31, 32],
    'address': ['street1', 'street2', 'street3', 'street1', 'street2', 'street4']
}
df = pd.DataFrame(data)
grouped_df = df.groupby(['school', 'class'])
print(type(grouped_df))
for (school, cls), group in grouped_df:
    print(f"\nSchool: {school}, Class: {cls}")
    print(group)
```

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>> 
= RESTART: D:/Query Processing/18.Write a Pandas program to split the following
given dataframe into groups based on school code and class..py
<class 'pandas.core.groupby.generic.DataFrameGroupBy'>

School: s001, Class: V
  school class            name date_Of_Birth  age  height  weight  address
0   s001     V  Alberto Franco    15/05/2002   12     173      35  street1

School: s001, Class: VI
  school class          name date_Of_Birth  age  height  weight  address
3   s001    VI  Eesha Hinton    25/09/1998   13     167      30  street1

School: s002, Class: V
  school class          name date_Of_Birth  age  height  weight  address
1   s002     V  Gino Mcneill    17/05/2002   12     192      32  street2
4   s002     V  Gino Mcneill    11/05/2002   14     151      31  street2

School: s003, Class: VI
  school class         name date_Of_Birth  age  height  weight  address
2   s003    VI  Ryan Parkes    16/02/1999   13     186      33  street3

School: s004, Class: VI
  school class          name date_Of_Birth  age  height  weight  address
5   s004    VI  David Parkes    15/09/1997   12     159      32  street4
>>> 
```

19.Write a Pandas program to display the dimensions or shape of the World alcohol consumption dataset. Also extract the column names from the dataset.



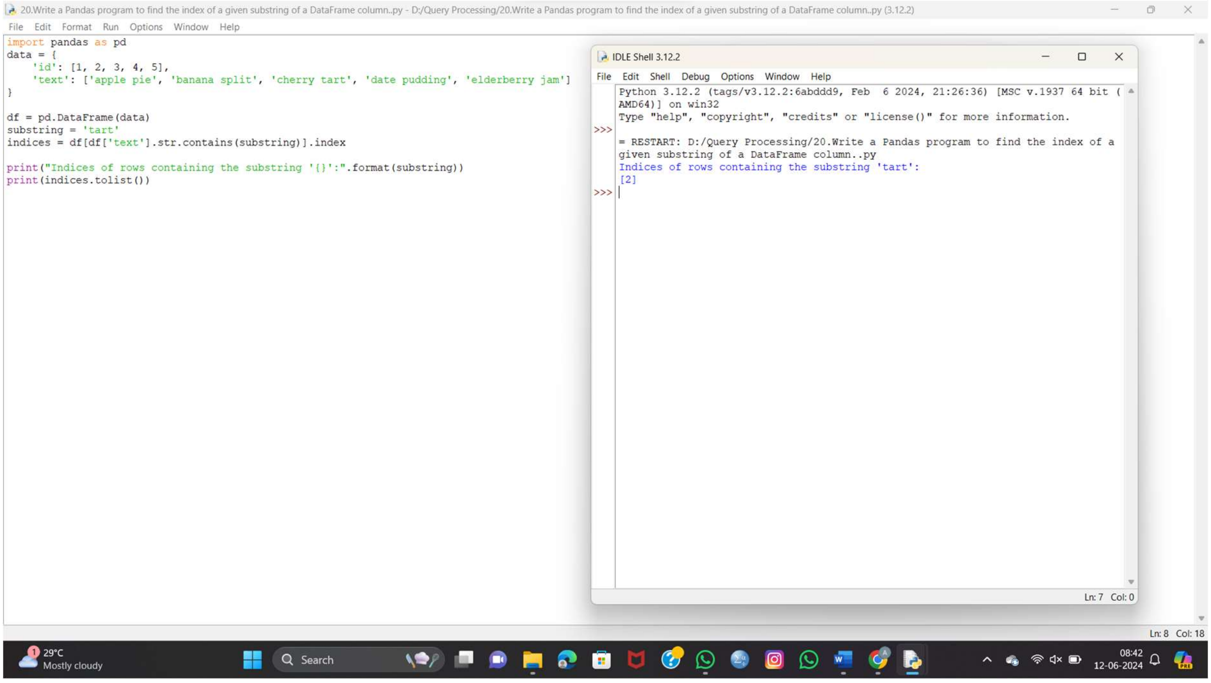
19.Write a Pandas program to display the dimensions or shape of the World alcohol consumption dataset. Also extract the column names from the dataset..py - D:/Query Processing/19.Write a Pandas program to display the dimensions or shape of the World...

```python
import pandas as pd
data = {
    'Year': [1986, 1986, 1985, 1986, 1987],
    'WHO region': ['Western Pacific', 'Americas', 'Africa', 'Americas', 'Americas'],
    'Country': ['Viet Nam', 'Uruguay', "Cte d'Ivoire", 'Colombia', 'Saint Kitts and Nevis'],
    'Beverage Types': ['Wine', 'Other', 'Wine', 'Beer', 'Beer'],
    'Display Value': [0.00, 0.50, 1.62, 4.27, 1.98]
}
df = pd.DataFrame(data)
print("Shape of the dataset:", df.shape)
print("Column names:", df.columns.tolist())
```

IDLE Shell 3.12.2

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (AMD64)
] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/19.Write a Pandas program to display the dimensions or
shape of the World alcohol consumption dataset. Also extract the column names from the
dataset..py
Shape of the dataset: (5, 5)
Column names: ['Year', 'WHO region', 'Country', 'Beverage Types', 'Display Value']
>>>
```

20.Write a Pandas program to find the index of a given substring of a DataFrame column.



20.Write a Pandas program to find the index of a given substring of a DataFrame column..py - D:/Query Processing/20.Write a Pandas program to find the index of a given substring of a DataFrame column..py (3.12.2)

```python
import pandas as pd
data = {
    'id': [1, 2, 3, 4, 5],
    'text': ['apple pie', 'banana split', 'cherry tart', 'date pudding', 'elderberry jam']
}

df = pd.DataFrame(data)
substring = 'tart'
indices = df[df['text'].str.contains(substring)].index

print("Indices of rows containing the substring '{}':".format(substring))
print(indices.tolist())
```

IDLE Shell 3.12.2

```
Python 3.12.2 (tags/v3.12.2:6abddd9, Feb  6 2024, 21:26:36) [MSC v.1937 64 bit (
AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
= RESTART: D:/Query Processing/20.Write a Pandas program to find the index of a
given substring of a DataFrame column..py
Indices of rows containing the substring 'tart':
[2]
>>>
```