ET2-65 SAEE PANDIT (Grocery analysis)

```
import pandas as pd

import numpy as np


df = pd.read_csv('groceriesdata.csv')


#1

print("Display first 5 rows of the data",df.head())

#Display first 5 rows of the data    Member_number       Date   itemDescription

#0          1808  21-07-2015    tropical fruit

#1          2552  05-01-2015       whole milk

#2          2300  19-09-2015        pip fruit

#3          1187  12-12-2015  other vegetables

#4          3037  01-02-2015       whole milk


#2

print("Display last 5 rows of the data",df.tail())

#Display last 5 rows of the data       Member_number       Date       itemDescription

#38760          4471  08-10-2014         sliced cheese

#38761          2022  23-02-2014               candy

#38762          1097  16-04-2014            cake bar

#38763          1510  03-12-2014  fruit/vegetable juice

#38764          1521  26-12-2014            cat food


#3

print("Display the number of rows and columns of data",df.shape)

#Display the number of rows and columns of data (38765, 3)


#4

print("Display coloumn names",df.columns)

#Display coloumn names Index(['Member_number', 'Date', 'itemDescription'], dtype='object')


#5

print("display 1 coloumn",df['itemDescription'])
```

ET2-65 SAEE PANDIT (Grocery analysis)

```
#display 1 coloumn 0              tropical fruit

#1                 whole milk

#2                  pip fruit

#3           other vegetables

#4                 whole milk

#                ...

#38760           sliced cheese

#38761                 candy

#38762               cake bar

#38763    fruit/vegetable juice

#38764               cat food

#Name: itemDescription, Length: 38765, dtype: object


#6

print("Most brought items(5)",df['itemDescription'].value_counts().head(5))

# itemDescription

# whole milk         2502

# other vegetables    1898

# rolls/buns         1716

# soda               1514

# yogurt             1334

# Name: count, dtype: int64


#7

print("Least freuquently brought item",print(df['itemDescription'].value_counts().idxmin()))

# Least freuquently brought item None


#8


print("First date:", df['Date'].min())

# First date: 01-01-2014


#9
```

ET2-65 SAEE PANDIT (Grocery analysis)

```
print("Last date:", df['Date'].max())

# Lastt date: 31-10-2015


#-------------------------#NUMPY ANALYSIS-----------------------------------------------------------#


#10

print("List first 5 unique items:", np.unique(df['itemDescription'])[:5])

# List first 5 unique items: ['Instant food products' 'UHT-milk' 'abrasive cleaner' 'artif. sweetener'

#  'baby cosmetics']


#11

print("Find number of unique items bought:", np.unique(df['itemDescription']).size)

# Find number of unique items bought: 167


#12

print("Find the sum of all Member_numbers:", np.sum(df['Member_number']))

# Find the sum of all Member_numbers: 116436177


#13

print("Find the unique member numbers count:", np.unique(df['Member_number']).size)

# Find the unique member numbers count: 3898


#14

print("Find indexes where Date is '20-06-2014':", np.where(df['Date'] == '20-06-2014'))

# Find indexes where Date is '20-06-2014': (array([ 8713,  9858, 10079, 10271, 11021, 11361, 11571, 12807, 13394,

   #    13623, 13702, 13806, 13903, 13907, 14170, 14434, 14643, 14749,

   #    14812, 15007, 15378, 15401, 24862, 26007, 26228, 26420, 27170,

   #    27510, 27720, 28956, 29543, 29772, 29851, 29955, 30052, 30056,

   #    30319, 30583, 30792, 30898, 30961, 31156, 31527, 31550, 32961,

   #    33161, 33577, 33936, 34033, 34037, 34487, 35309, 36789, 37149,

   #    37253, 37350, 37354, 37512, 37712, 38128, 38604]),)


#15
```

ET2-65 SAEE PANDIT (Grocery analysis)

```
print("Find indexes where item bought is 'whole milk':", np.where(df['itemDescription'] == 'whole milk'))

# Find indexes where item bought is 'whole milk': (array([    1,     4,     8, ..., 38688, 38689, 38745], shape=(2502,)),)


#16

print("Mean of Member_numbers:", np.mean(df['Member_number']))

# Mean of Member_numbers: 3003.64186766413


#17

maxindex = np.argmax(df['Member_number'])

print("Index where Member_number is maximum:", maxindex)

# Index where Member_number is maximum: 3578


#18

minindex=np.argmin(df["Member_number"])

print("Index where Member Number is minimum ",minindex)

# Index where Member Number is minimum  1629


#19

print("Most popular item",df['itemDescription'].value_counts().idxmax())

# Most popular item whole milk


#20

soda_count = np.sum(df['itemDescription'].values == 'soda')

print("Number of times 'soda' was bought:", soda_count)

# Number of times 'soda' was bought: 1514
```

ET2-65 SAEE PANDIT (Grocery analysis)

edsdataset.py > ...

```python
import pandas as pd
import numpy as np


df = pd.read_csv('groceriesdata.csv')

#1
print("Display first 5 rows of the data",df.head())
#Display first 5 rows of the data    Member_number        Date   itemDescription
#0           1808  21-07-2015    tropical fruit
#1           2552  05-01-2015        whole milk
#2           2300  19-09-2015         pip fruit
#3           1187  12-12-2015  other vegetables
#4           3037  01-02-2015        whole milk

#2
print("Display last 5 rows of the data",df.tail())
#Display last 5 rows of the data        Member_number        Date        itemDescription
#38760           4471  08-10-2014          sliced cheese
#38761           2022  23-02-2014                 candy
#38762           1097  16-04-2014              cake bar
#38763           1510  03-12-2014  fruit/vegetable juice
#38764           1521  26-12-2014              cat food

#3
print("Display the number of rows and columns of data",df.shape)
#Display the number of rows and columns of data (38765, 3)

#4
print("Display coloumn names",df.columns)
#Display coloumn names Index(['Member_number', 'Date', 'itemDescription'], dtype='object')

#5
print("display 1 coloumn",df['itemDescription'])
#display 1 coloumn 0              tropical fruit
#1                    whole milk
#2                     pip fruit
#3                   other vegetables
#4                          whole milk
#                        ...
#38760            sliced cheese
#38761                   candy
#38762                cake bar
#38763    fruit/vegetable juice
#38764                cat food
#Name: itemDescription, Length: 38765, dtype: object


#6
print("Most brought items(5)",df['itemDescription'].value_counts().head(5))
# itemDescription
# whole milk          2502
# other vegetables    1898
# rolls/buns          1716
# soda                1514
# yogurt              1334
# Name: count, dtype: int64

#7
print("Least freuquently brought item",print(df['itemDescription'].value_counts().idxmin()))
# Least freuquently brought item None

#8

print("First date:", df['Date'].min())
# First date: 01-01-2014

#9
print("Last date:", df['Date'].max())
# Lastt date: 31-10-2015
```

ET2-65 SAEE PANDIT (Grocery analysis)

```python
#-------------------------#NUMPY ANALYSIS-----------------------------------------------------------------#


#10
print("List first 5 unique items:", np.unique(df['itemDescription'])[:5])
# List first 5 unique items: ['Instant food products' 'UHT-milk' 'abrasive cleaner' 'artif. sweetener'
#  'baby cosmetics']

#11
print("Find number of unique items bought:", np.unique(df['itemDescription']).size)
# Find number of unique items bought: 167

#12
print("Find the sum of all Member_numbers:", np.sum(df['Member_number']))
# Find the sum of all Member_numbers: 116436177

#13
print("Find the unique member numbers count:", np.unique(df['Member_number']).size)
# Find the unique member numbers count: 3898

#14
print("Find indexes where Date is '20-06-2014':", np.where(df['Date'] == '20-06-2014'))
# Find indexes where Date is '20-06-2014': (array([ 8713,  9858, 10079, 10271, 11021, 11361, 11571, 12807, 13394,
    #    13623, 13702, 13806, 13903, 13907, 14170, 14434, 14643, 14749,
    #    14812, 15007, 15378, 15401, 24862, 26007, 26228, 26420, 27170,
    #    27510, 27720, 28956, 29543, 29772, 29851, 29955, 30052, 30056,
    #    30319, 30583, 30792, 30898, 30961, 31156, 31527, 31550, 32961,
    #    33161, 33577, 33936, 34033, 34037, 34487, 35309, 36789, 37149,
    #    37253, 37350, 37354, 37512, 37712, 38128, 38604]),)


#15
print("Find indexes where item bought is 'whole milk':", np.where(df['itemDescription'] == 'whole milk'))
# Find indexes where item bought is 'whole milk': (array([    1,     4,     8, ..., 38688, 38689, 38745], shape=(2502,)),)

#16
print("Mean of Member_numbers:", np.mean(df['Member_number']))
```

```python
#15
print("Find indexes where item bought is 'whole milk':", np.where(df['itemDescription'] == 'whole milk'))
# Find indexes where item bought is 'whole milk': (array([    1,     4,     8, ..., 38688, 38689, 38745], shape=(2502,)),)

#16
print("Mean of Member_numbers:", np.mean(df['Member_number']))
# Mean of Member_numbers: 3003.64186766413

#17
maxindex = np.argmax(df['Member_number'])
print("Index where Member_number is maximum:", maxindex)
# Index where Member_number is maximum: 3578

#18
minindex=np.argmin(df["Member_number"])
print("Index where Member Number is minimum ",minindex)
# Index where Member Number is minimum  1629

#19
print("Most popular item",df['itemDescription'].value_counts().idxmax())
# Most popular item whole milk

#20
soda_count = np.sum(df['itemDescription'].values == 'soda')
print("Number of times 'soda' was bought:", soda_count)
# Number of times 'soda' was bought: 1514
```

```
PROBLEMS  OUTPUT  DEBUG CONSOLE  TERMINAL  PORTS

       33161, 33577, 33936, 34033, 34037, 34487, 35309, 36789, 37149,
       37253, 37350, 37354, 37512, 37712, 38128, 38604]),)
Find indexes where item bought is 'whole milk': (array([    1,     4,     8, ..., 38688, 38689, 38745], shape=(2502,)),)
Mean of Member_numbers: 3003.64186766413
Index where Member_number is maximum: 3578
Index where Member Number is minimum  1629
Most popular item whole milk
Number of times 'soda' was bought: 1514
```

powershell
Python
Python

Ln 125, Col 1 Spaces: 4 UTF-8 CRLF {} Python 3.12.4 Go Live