[4:15 pm, 20/10/2023] Arulmohi: import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

data = pd.read\_csv("../input/market-basket-analysis/Assignment-1\_Data.csv", sep=';')

data.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 522064 entries, 0 to 522063

Data columns (total 7 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 BillNo 522064 non-null object

1 Itemname 520609 non-null object

2 Quantity 522064 non-null int64

3 Date 522064 non-null object

4 Price 522064 non-null object

5 CustomerID 388023 non-null float64

6 Country 522064 non-null object

dtypes: float64(1), int64(1), object(5)

memory usage: 27.9+ MB

/opt/conda/li…

[4:15 pm, 20/10/2023] Arulmohi: import numpy as np # linear algebra

import pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv)

data = pd.read\_csv("../input/market-basket-analysis/Assignment-1\_Data.csv", sep=';')

data.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 522064 entries, 0 to 522063

Data columns (total 7 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 BillNo 522064 non-null object

1 Itemname 520609 non-null object

2 Quantity 522064 non-null int64

3 Date 522064 non-null object

4 Price 522064 non-null object

5 CustomerID 388023 non-null float64

6 Country 522064 non-null object

dtypes: float64(1), int64(1), object(5)

memory usage: 27.9+ MB

/opt/conda/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3553: DtypeWarning: Columns (0) have mixed types.Specify dtype option on import or set low\_memory=False.

exec(code\_obj, self.user\_global\_ns, self.user\_ns)

data["Price"] = data["Price"].str.replace(",",".")

data["Price"] = data["Price"].astype("float64")

data.info()

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 522064 entries, 0 to 522063

Data columns (total 7 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 BillNo 522064 non-null object

1 Itemname 520609 non-null object

2 Quantity 522064 non-null int64

3 Date 522064 non-null object

4 Price 522064 non-null float64

5 CustomerID 388023 non-null float64

6 Country 522064 non-null object

dtypes: float64(2), int64(1), object(4)

memory usage: 27.9+ MB

dates = data['Date'].unique()

dates.sort()

dates[0]

'01.02.2011 08:23'

dates[-1]

'31.10.2011 17:19'

frequency = freq\_table = data.drop\_duplicates(subset = "BillNo").groupby(["CustomerID"])[["BillNo"]].count()

freq\_table.columns = ["Frequency"]

freq\_table.head()

Frequency

CustomerID

12346.0 1

12347.0 7

12349.0 1

12350.0 1

12352.0 8