⌄ Exercise 13: Mining Frequent Patterns from Air Pollution Data (Google Collab)

s1292001

First, the data is loaded into a dataframe

```
import pandas as pd

dataset = pd.read_csv("airPollutionData.csv")
dataset.head()
```

	Timestamp	POINT(128.1779451 26.7456488)	POINT(128.8409418 32.6954809)	POINT(128.8460344 32.6989579)	POINT(129.285… 34.2426…
0	2024-06-01 01:00:00	NaN	NaN	8.0	
1	2024-06-01 02:00:00	NaN	NaN	10.0	
2	2024-06-01 03:00:00	NaN	NaN	8.0	
3	2024-06-01 04:00:00	NaN	NaN	9.0	
4	2024-06-01 05:00:00	NaN	NaN	9.0	

5 rows × 1623 columns

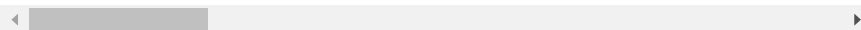## ⌄ Data pre-processing

Step 1:

Removing the timestamp column

```
dataset.drop(columns=["Timestamp"], inplace=True)
dataset.head()
```

	POINT(128.1779451 26.7456488)	POINT(128.8409418 32.6954809)	POINT(128.8460344 32.6989579)	POINT(129.2857639 34.2426737)	POINT… …
0	NaN	NaN	8.0	11.0	
1	NaN	NaN	10.0	10.0	
2	NaN	NaN	8.0	11.0	
3	NaN	NaN	9.0	10.0	
4	NaN	NaN	9.0	9.0	

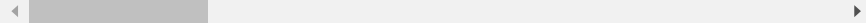5 rows × 1622 columns

⌄ Step 2:

Removing Sensors with no Point information

```
sensorsWithPointInformation = [col for col in dataset if 'Unnamed' in col]
dataset.drop(columns=sensorsWithPointInformation, inplace=True, axis=1)
dataset.head()
```

	POINT(128.1779451 26.7456488)	POINT(128.8409418 32.6954809)	POINT(128.8460344 32.6989579)	POINT(129.2857639 34.2426737)	POINT
0	NaN	NaN	8.0	11.0	
1	NaN	NaN	10.0	10.0	
2	NaN	NaN	8.0	11.0	
3	NaN	NaN	9.0	10.0	
4	NaN	NaN	9.0	9.0	

5 rows × 1622 columns

◄ |████████| ► ▶

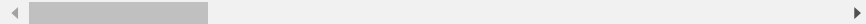## Step 3:

Replacing NaN values with 0

```
dataset.fillna(0, inplace=True)
dataset.head()
```

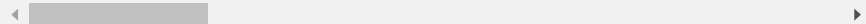|	POINT(128.1779451 26.7456488)	POINT(128.8409418 32.6954809)	POINT(128.8460344 32.6989579)	POINT(129.2857639 34.2426737)	POINT
0	0.0	0.0	8.0	11.0	
1	0.0	0.0	10.0	10.0	
2	0.0	0.0	8.0	11.0	
3	0.0	0.0	9.0	10.0	
4	0.0	0.0	9.0	9.0	

5 rows × 1622 columns

◄ |████████| ► ▶

Step 4:

Changing the columns to first character capitalized, lowercase format

```
dataset.columns = [x.capitalize() for x in dataset.columns]
dataset.head()
```

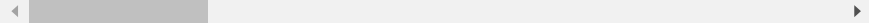|	Point(128.1779451 26.7456488)	Point(128.8409418 32.6954809)	Point(128.8460344 32.6989579)	Point(129.2857639 34.2426737)	Point
0	0.0	0.0	8.0	11.0	
1	0.0	0.0	10.0	10.0	
2	0.0	0.0	8.0	11.0	
3	0.0	0.0	9.0	10.0	
4	0.0	0.0	9.0	9.0	

5 rows × 1622 columns

◄ |████████| ► ▶

Step 5:

Removing abnormal values in the data

```
dataset.where(dataset < 100, 0, inplace=True) # Get rid of values greater or equal to 100
dataset.where(dataset > 0, 0, inplace=True) # Get rid of negative values
dataset.head()
```

	Point(128.1779451 26.7456488)	Point(128.8409418 32.6954809)	Point(128.8460344 32.6989579)	Point(129.2857639 34.2426737)	Point
0	0.0	0.0	8.0	11.0	
1	0.0	0.0	10.0	10.0	
2	0.0	0.0	8.0	11.0	
3	0.0	0.0	9.0	10.0	
4	0.0	0.0	9.0	9.0	

5 rows × 1622 columns

Installing PAMI

```
!pip install -U PAMI
```

```
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from PAMI) (3.7.1)
Requirement already satisfied: resource in /usr/local/lib/python3.10/dist-packages (from PAMI) (0.2.1)
Requirement already satisfied: validators in /usr/local/lib/python3.10/dist-packages (from PAMI) (0.33.0)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from PAMI) (2.0.7)
Requirement already satisfied: Pillow in /usr/local/lib/python3.10/dist-packages (from PAMI) (9.4.0)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages (from PAMI) (1.25.2)
Requirement already satisfied: sphinx in /usr/local/lib/python3.10/dist-packages (from PAMI) (5.0.2)
Requirement already satisfied: sphinx-rtd-theme in /usr/local/lib/python3.10/dist-packages (from PAMI) (2.0.0)
Requirement already satisfied: discord.py in /usr/local/lib/python3.10/dist-packages (from PAMI) (2.4.0)
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages (from PAMI) (3.3)
Requirement already satisfied: deprecated in /usr/local/lib/python3.10/dist-packages (from PAMI) (1.2.14)
Requirement already satisfied: wrapt<2,>=1.10 in /usr/local/lib/python3.10/dist-packages (from deprecated->PAMI) (1.14.1)
Requirement already satisfied: aiohttp<4,>=3.7.4 in /usr/local/lib/python3.10/dist-packages (from discord.py->PAMI) (3.9.5)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (1.2.1)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (1.4.5)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (24.1)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (3.1.2)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->PAMI) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->PAMI) (2024.1)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->PAMI) (2024.1)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.10/dist-packages (from plotly->PAMI) (8.5.0)
Requirement already satisfied: JsonForm>=0.0.2 in /usr/local/lib/python3.10/dist-packages (from resource->PAMI) (0.0.2)
Requirement already satisfied: JsonSir>=0.0.2 in /usr/local/lib/python3.10/dist-packages (from resource->PAMI) (0.0.2)
Requirement already satisfied: python-easyconfig>=0.1.0 in /usr/local/lib/python3.10/dist-packages (from resource->PAMI) (0.1.7)
Requirement already satisfied: sphinxcontrib-applehelp in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (1.0.8)
Requirement already satisfied: sphinxcontrib-devhelp in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (1.0.6)
Requirement already satisfied: sphinxcontrib-jsmath in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (1.0.1)
Requirement already satisfied: sphinxcontrib-htmlhelp>=2.0.0 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (2.0
Requirement already satisfied: sphinxcontrib-serializinghtml>=1.1.5 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAM
Requirement already satisfied: sphinxcontrib-qthelp in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (1.0.8)
Requirement already satisfied: Jinja2>=2.3 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (3.1.4)
Requirement already satisfied: Pygments>=2.0 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (2.16.1)
Requirement already satisfied: docutils<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (0.18.1)
Requirement already satisfied: snowballstemmer>=1.1 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (2.2.0)
Requirement already satisfied: babel>=1.3 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (2.15.0)
Requirement already satisfied: alabaster<0.8,>=0.7 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (0.7.16)
Requirement already satisfied: imagesize in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (1.4.1)
Requirement already satisfied: requests>=2.5.0 in /usr/local/lib/python3.10/dist-packages (from sphinx->PAMI) (2.31.0)
Requirement already satisfied: sphinxcontrib-jquery<5,>=4 in /usr/local/lib/python3.10/dist-packages (from sphinx-rtd-theme->PAM
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->discord.py->
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->discord.py->PAM
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->discord.py-
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->discord.p
Requirement already satisfied: yarl<2.0,>=1.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->discord.py->PA
Requirement already satisfied: async-timeout<5.0,>=4.0 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.7.4->disco
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.3->sphinx->PAMI) (2.1.
Requirement already satisfied: jsonschema in /usr/local/lib/python3.10/dist-packages (from JsonForm>=0.0.2->resource->PAMI) (4.2
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->PAMI)
Requirement already satisfied: PyYAML in /usr/local/lib/python3.10/dist-packages (from python-easyconfig>=0.1.0->resource->PAMI)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.0->sphinx
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.0->sphinx->PAMI) (3.7
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests>=2.5.0->sphinx->PAMI
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema-
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema->JsonForm>=0.0.2-
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema->JsonForm>=0.0.2->reso
```

Data transformation

Converting the data frame into a transactional database

```
from PAMI.extras.DF2DB import DenseFormatDF as db

filename = 'PM24HeavyPollutionRecordingSensors.csv'

obj = db.DenseFormatDF(dataset)
obj.convert2TransactionalDatabase(filename, '>=', 15)
```
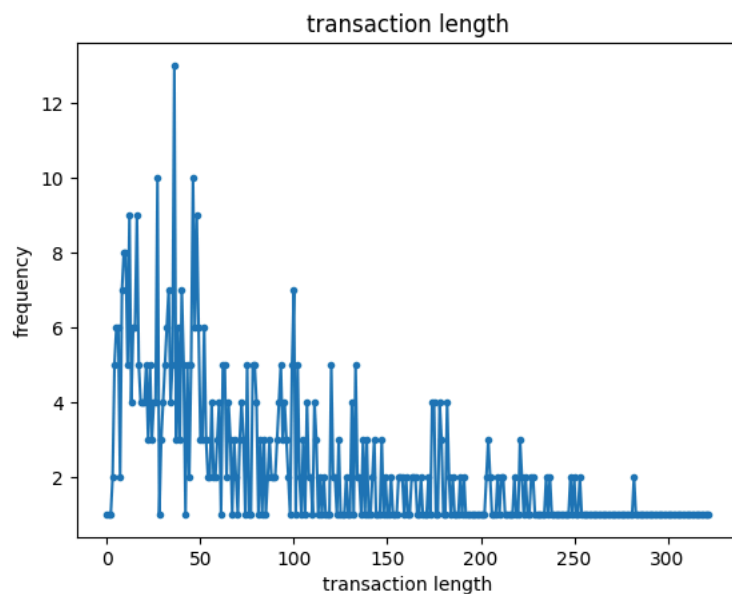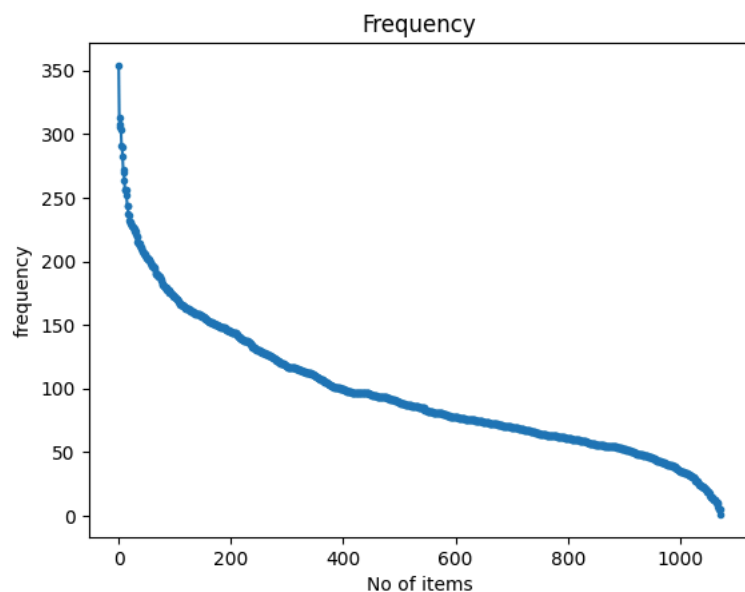
⌄ Data Mining

```
from PAMI.extras.dbStats import TransactionalDatabase as tds

obj = tds.TransactionalDatabase(filename)
obj.run()
obj.printStats()
obj.plotGraphs()
```

⇥

```
Average Transaction Size : 144.9791666666666
Maximum Transaction Size : 706
Standard Deviation Transaction Size : 151.67913267221184
Variance in Transaction Sizes : 23038.557284422808
Sparsity : 0.8648842808325566
```

```
from PAMI.frequentPattern.basic import FPGrowth as ab

minSup=150
obj = ab.FPGrowth(filename, minSup)
obj.mine()
obj.printResults()
obj.save('frequentPatterns.txt')
```

```
Frequent patterns were generated successfully using frequentPatternGrowth algorithm
Total number of Frequent Patterns: 779
Total Memory in USS: 353562624
Total Memory in RSS 370946048
Total ExecutionTime in ms: 1.5116190910339355
```

∨ Visualizing the results

```
from PAMI.extras.graph import visualizePatterns as fig

obj = fig.visualizePatterns('frequentPatterns.txt', 20)
obj.visualize(width=1000,height=900)
```

```
from PAMI.frequentPattern.basic import FPGrowth as ab

minSup=150
obj = ab.FPGrowth(filename, minSup)
obj.mine()
obj.printResults()
obj.save('frequentPatterns.txt')
```

```
Frequent patterns were generated successfully using frequentPatternGrowth algorithm
```