

## Лабораторная работа № 4

Перед началом выполнения каждой лабораторной работы, необходимо выполнить код в двух ячейках ниже. В случае если модуль `datetime` не установлен нужно установить. В последнем принте нужно вывести свою фамилию и инициалы.

In [39]:

```
!whoami
```

```
kaneva\ГЕ вГаЕ
```

In [2]:

```
from datetime import datetime

current_time = datetime.now()
print(current_time)
print("Kaneva E.P.") # написать здесь свою фамилию и инициалы
```

```
2022-12-07 14:15:12.740319
```

```
Kaneva E.P.
```

## Задания

In [1]:

```
import pandas as pd
df = pd.read_csv('data.csv', sep = ';')
```

**Задание 1.** В датафрейме `df` задана некоторая таблица. Написать функцию `get_rows_after_5(df, n)`, возвращающую датафрейм, в котором записано `n` строк, начиная с пятой сверху (включая 5-ю). Например, `get_row_after_5(df, 1)` должна вернуть только пятую строку, а `get_row_after_5(df, 2)` — 5-ю и 6-ю.

**Внимание!** Индексами (именами строк) могут быть не числа, а что угодно.

In [2]:

```
def get_rows_after_5(df, n):
    n += 4
    new_df = df.iloc[4:n]
    return new_df

print(get_rows_after_5(df, 2))
```

```
ID                                     Region  global_id \
4  302   город Москва (за исключением Троицкого и Новом...  19693301
5  313   город Москва (за исключением Троицкого и Новом...  19693302

TariffItem  UnitOfMeasure  StartDate  EndDate  MeasureDeviceAvail \
4  Электроэнергия      руб/кВт.ч  01.07.2012  31.07.2012             да
5  Отопление           руб/Гкал  01.09.2012  31.12.2012             да

StoveType                                     ConsumptionTime  ConsumptionTarget \
4  газовая с 17:00 по 21:00 ч (многотарифный учет) на бытовые нужды
5      NaN                                     NaN                NaN

Agency  TariffValue  Unnamed: 13
4      ОАО «Мосэнергосбыт»           4.03      NaN
5  ОАО «МОЭК» и иные, кроме ОАО «Мосэнерго» и ОАО...  1440.50      NaN
```

**Задание 2.** В датафрейме `df` задана некоторая таблица, её индексами являются целые числа, не обязательно идущие по порядку. Написать функцию `between(df, n, m)`, возвращающую все строки этой таблицы, расположенные между строками с индексами `n` и `m`, включая строки с индексами `n` и `m`. Гарантируется, что строка с индексом `n` встречается раньше строки с индексом `m`.

In [38]:

```
def between(df, n, m):
    m += 1
    new_df = df.iloc[n:m]
    return new_df

print(between(df, 2, 4))
```

ID	Region	global_id	\
2 285	город Москва (за исключением Троицкого и Новом...	19693299	
3 301	город Москва (за исключением Троицкого и Новом...	19693300	
4 302	город Москва (за исключением Троицкого и Новом...	19693301	

  

TariffItem	UnitOfMeasure	StartDate	EndDate	MeasureDeviceAvail	\
2 Электроэнергия	руб/кВт.ч	2012-01-01	2012-06-30	да	
3 Электроэнергия	руб/кВт.ч	2012-01-07	2012-07-31	да	
4 Электроэнергия	руб/кВт.ч	2012-01-07	2012-07-31	да	

  

StoveType	ConsumptionTime	ConsumptionTarget	\
2 газовая с 07:00 по 10:00 ч (многотарифный учет)	на бытовые нужды		
3 газовая с 07:00 по 10:00 ч (многотарифный учет)	на бытовые нужды		
4 газовая с 17:00 по 21:00 ч (многотарифный учет)	на бытовые нужды		

  

Agency	TariffValue	Unnamed: 13
2 ОАО «Мосэнергосбыт»	3.80	NaN
3 ОАО «Мосэнергосбыт»	4.03	NaN
4 ОАО «Мосэнергосбыт»	4.03	NaN

**Задание 3.** Загрузите файл `data.csv` (приложен к лабораторной работе).

1. Определите типы данных в каждом столбце и объем используемой памяти таблицей.
2. Столбцы с датами `StartDate` и `EndDate` приведите к типу `datetime64` (можно использовать функцию `pd.to_datetime()`).
3. Выберите из датафрейма агентство **ОАО «Мосэнергосбыт»** и подсчитайте среднее значение тарифа (`TariffValue`).
4. Подсчитайте для каждого агентства *сумму, максимальное, минимальное, среднее, медианное* значения тарифа (`TariffValue`).

In [15]:

```
# import pandas as pd
# df = pd.read_csv("data.csv", sep=';') - уже было сделано выше

# 3.1.

print(f'Data types:\n{df.dtypes}\n')
print(f'Memory used: {sum(df.memory_usage())} bytes')
```

```
Data types:
ID                int64
Region            object
global_id         int64
TariffItem        object
UnitOfMeasure     object
StartDate         object
EndDate          object
MeasureDeviceAvail object
StoveType         object
ConsumptionTime   object
ConsumptionTarget object
Agency           object
TariffValue       float64
Unnamed: 13       float64
dtype: object
```

```
Memory used: 8080 bytes
```

In [17]:

# 3.2.

```
df['StartDate'] = pd.to_datetime(df['StartDate'])
df['EndDate'] = pd.to_datetime(df['EndDate'])
print(f'Data types after the edit:\n{df.dtypes}\n')
```

```
Data types after the edit:
ID                int64
Region            object
global_id         int64
TariffItem        object
UnitOfMeasure     object
StartDate         datetime64[ns]
EndDate           datetime64[ns]
MeasureDeviceAvail object
StoveType         object
ConsumptionTime   object
ConsumptionTarget object
Agency           object
TariffValue       float64
Unnamed: 13       float64
dtype: object
```

In [22]:

# 3.3.

```
new_df = df[df['Agency'] == 'ОАО «Мосэнергосбыт»']
print(f'Mean value: {new_df["TariffValue"].mean()}')
```

Mean value: 2.5712499999999996

In [37]:

# 3.4.

```
agencies_list = list(set(df['Agency']))
# print(agencies_list) - просто осталось с проверки адекватности того, что пишу

new_df = pd.DataFrame({
    'Sum': 0.00,
    'MaxValue': 0.00,
    'MinValue': 0.00,
    'MeanValue': 0.00,
    'MedianValue': 0.00},
    index = agencies_list,
)
# print(new_df) - просто осталось с проверки адекватности того, что пишу

for agency in agencies_list:
    agency_split = df[df["Agency"] == agency]
    agency_tariff = agency_split["TariffValue"]
    new_df.Sum[agency] = sum(list(agency_tariff))
    new_df.MaxValue[agency] = max(list(agency_tariff))
    new_df.MinValue[agency] = min(list(agency_tariff))
    new_df.MeanValue[agency] = agency_tariff.mean()
    new_df.MedianValue[agency] = agency_tariff.median()

print(new_df)
```

	Sum	MaxValue	\
ООО «Мосрегионгаз»	7.02	4.09	
ОАО «Мосэнерго» и ОАО «МТК»	3366.70	1075.43	
ОАО «Мосэнергосбыт»	82.28	4.03	
ООО «Газпром межрегионгаз Москва»	13021.28	3370.00	
ОАО «МОЭК» и иные, кроме ОАО «Мосэнерго» и ОАО ...	4484.41	1440.50	
ОАО «Мосгаз»	1126.00	1126.00	
МГУП «Мосводоканал»	129.52	26.75	

  

	MinValue	MeanValue	\
ООО «Мосрегионгаз»	2.93	3.510000	
ОАО «Мосэнерго» и ОАО «МТК»	84.62	561.116667	
ОАО «Мосэнергосбыт»	0.67	2.571250	
ООО «Газпром межрегионгаз Москва»	4.09	723.404444	
ОАО «МОЭК» и иные, кроме ОАО «Мосэнерго» и ОАО ...	105.45	747.401667	
ОАО «Мосгаз»	1126.00	1126.000000	
МГУП «Мосводоканал»	16.65	21.586667	

  

	MedianValue
ООО «Мосрегионгаз»	3.510
ОАО «Мосэнерго» и ОАО «МТК»	541.455
ОАО «Мосэнергосбыт»	2.735
ООО «Газпром межрегионгаз Москва»	36.460
ОАО «МОЭК» и иные, кроме ОАО «Мосэнерго» и ОАО ...	720.850
ОАО «Мосгаз»	1126.000
МГУП «Мосводоканал»	21.155