

```
In [3]: import pandas as pd
import datetime
import numpy as np
import matplotlib.pyplot as plt
```

```
In [4]: csv_data = pd.read_csv("Res\\IBM.csv", delimiter=';') # Read csv and delimitate it by ";" char
csv_data
```

Out[4]:

	<DATE>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>
0	02.01.12	186.88	188.64	182.320	182.52	1002795
1	09.01.12	181.87	183.52	177.350	179.19	1528096
2	16.01.12	180.34	188.96	179.410	188.53	1633959
3	23.01.12	187.81	192.78	187.740	190.53	1286450
4	30.01.12	188.74	194.81	188.290	193.56	1224958
...
454	21.09.20	120.45	121.55	116.495	118.91	762538
455	28.09.20	120.54	123.22	118.850	120.53	530172
456	05.10.20	121.78	135.49	121.060	127.77	1186117
457	12.10.20	128.07	128.16	124.000	125.90	700642
458	19.10.20	126.87	127.29	105.920	111.66	2841388

459 rows × 6 columns

```
In [5]: csv_data.index = pd.to_datetime(csv_data['<DATE>'], format='%d.%m.%y') # set main index as date
csv_data
```

Out[5]:

	<DATE>	<OPEN>	<HIGH>	<LOW>	<CLOSE>	<VOL>
<DATE>						
2012-01-02	02.01.12	186.88	188.64	182.320	182.52	1002795
2012-01-09	09.01.12	181.87	183.52	177.350	179.19	1528096
2012-01-16	16.01.12	180.34	188.96	179.410	188.53	1633959
2012-01-23	23.01.12	187.81	192.78	187.740	190.53	1286450
2012-01-30	30.01.12	188.74	194.81	188.290	193.56	1224958
...
2020-09-21	21.09.20	120.45	121.55	116.495	118.91	762538
2020-09-28	28.09.20	120.54	123.22	118.850	120.53	530172
2020-10-05	05.10.20	121.78	135.49	121.060	127.77	1186117
2020-10-12	12.10.20	128.07	128.16	124.000	125.90	700642
2020-10-19	19.10.20	126.87	127.29	105.920	111.66	2841388

459 rows × 6 columns

```
In [6]: """1. Узнать в какие месяцы объём сделок был больше,
чем в остальные (для этого вывести суммарный объём сделок по месяцам в таблице,
а затем визуализировать в виде столбцовой диаграммы)."""
```

Out[6]:

```
'1. Узнать в какие месяцы объём сделок был больше, \nчем в остальные (для этого вывести суммарный объём сделок по месяцам в таблице, \na затем визуализировать в виде столбцовой диаграммы).'
```

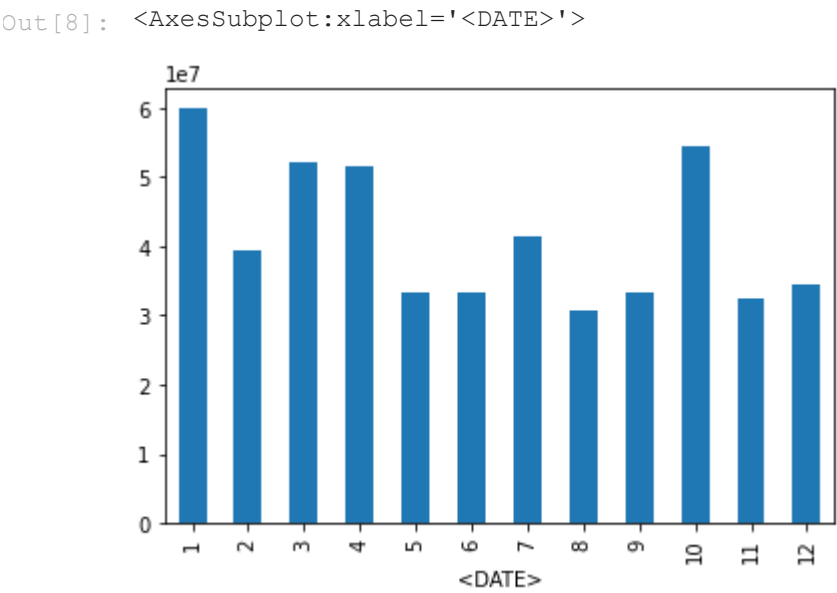
```
In [7]: # group the query by month 1-12 and sum 'VOLUME' in each group
val = csv_data.groupby(by=[csv_data.index.month])['<VOL>'].sum()
val
```

Out[7]:

<DATE>	
1	59985954
2	39448267
3	52047034
4	51531543
5	33380255
6	33444172
7	41393175
8	30809852
9	33203360
10	54529079
11	32516264
12	34368306

Name: <VOL>, dtype: int64

```
In [8]: val.plot(x='<DATE>', y='<VOL>', kind='bar')
```



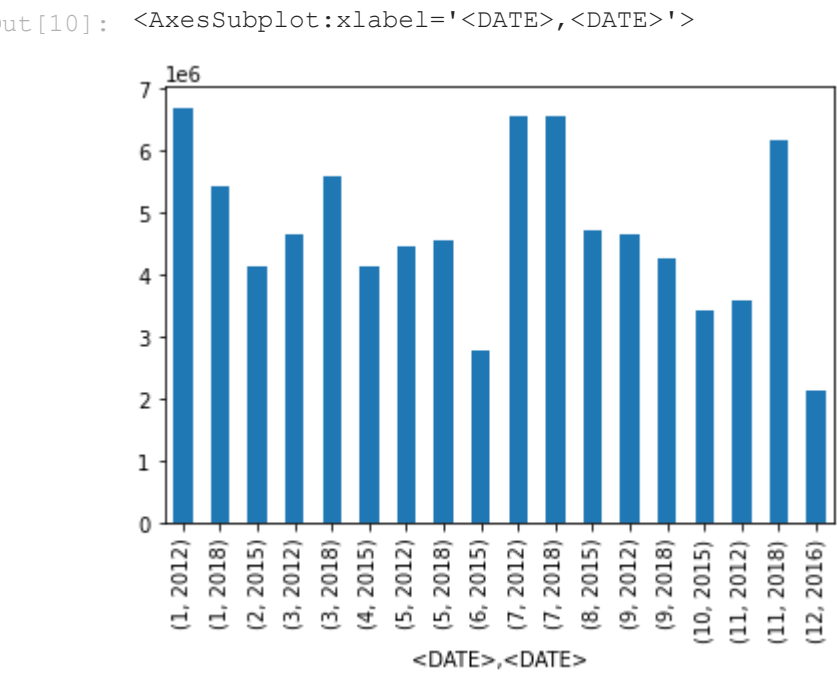
```
In [9]: # same but grouping by each year + each month
val = csv_data.groupby(by=[csv_data.index.month, csv_data.index.year])['<VOL>'].sum()
val
```

Out[9]:

<DATE>	<DATE>	
1	2012	6676258
	2013	5480431
	2014	3757794
	2015	16640622
	2016	4943424
	...	
12	2015	3306737
	2016	2117483
	2017	2636092
	2018	7291374
	2019	4062009

Name: <VOL>, Length: 106, dtype: int64

```
In [10]: val[:,6].plot(x='DATA', y='VOL', kind='bar')
```



```
In [11]: """2. Узнать среднюю цену акций на закрытии торгов по годам."""
```

Out[11]:

```
'2. Узнать среднюю цену акций на закрытии торгов по годам.'
```

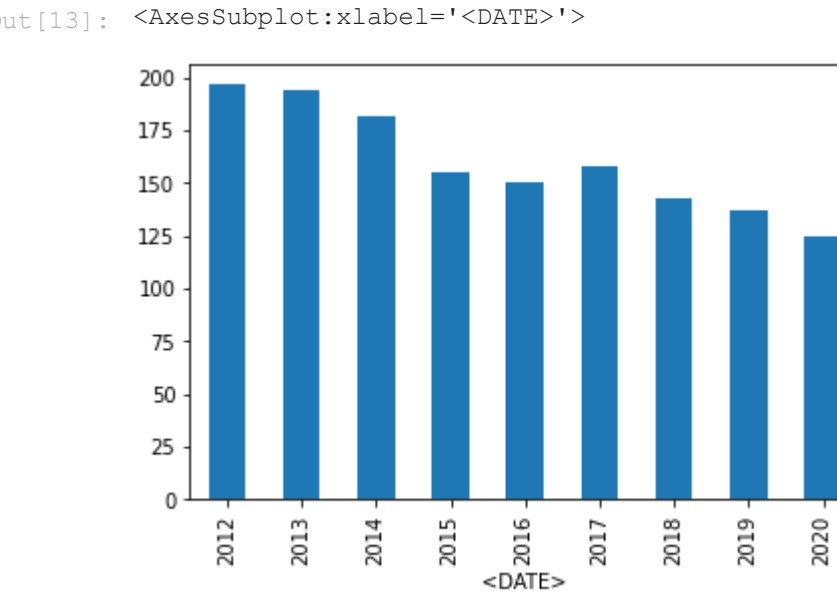
```
In [12]: # same but grouping by only each year
val = csv_data.groupby(by=[csv_data.index.year])['<CLOSE>'].mean()
val
```

Out[12]:

<DATE>	
2012	196.579245
2013	193.817885
2014	181.863750
2015	155.217451
2016	150.711731
2017	157.573654
2018	142.910000
2019	137.341346
2020	124.650357

Name: <CLOSE>, dtype: float64

```
In [13]: val.plot(x='DATA', y='VOL', kind='bar')
```



```
In [23]: %%html

Github:

https://github.com/KartonDCP/MathWorks/blob/master/SciPyProjects/FifthSemTasks/PyAndMl/MLLec4.ipynb
```

Github: <https://github.com/KartonDCP/MathWorks/blob/master/SciPyProjects/FifthSemTasks/PyAndMI/MLLec4.ipynb>

