

Joanna Mielniczuk

Metody planowania i analizy eksperymentów

Zadanie domowe nr 1

Dane:

Analizie poddano dane na temat przebiegu pandemii COVID-19 w Polsce w okresie od maja do lipca 2020 roku. Dane pochodzą ze strony *gov.pl*, zawierają dzienne informacje o zakażeniach, zgonach, liczbie ozdrowieńców, aktywnych przypadkach choroby, liczbie osób przebywających na kwarantannie oraz znajdujących się pod nadzorem. Dodatkowo, zawarto w nich kumulatywne wskaźniki dla wybranych wartości.

	Data	Nowe przypadki	Przypadki (kumulatywnie)	Zgony	Zgony kumulatywnie	Ozdrowieńcy	Ozdrowieńcy kumulatywnie	Aktywne przypadki	Kwarantanna	Nadzór
0	01.05.2020	228	13105	7	649	271	3762	8694	95625	18383
1	02.05.2020	270	13375	12	661	183	3945	8769	96612	18306
2	03.05.2020	318	13693	15	676	150	4095	8922	96699	17785
3	04.05.2020	313	14006	19	695	185	4280	9031	100765	17291
4	05.05.2020	425	14431	19	714	375	4655	9062	101395	17081
...
87	27.07.2020	337	43402	5	1676	187	33043	8683	94920	7245
88	28.07.2020	502	43904	6	1682	147	33190	9032	95453	6222
89	29.07.2020	512	44416	12	1694	453	33643	9079	97561	8094
90	30.07.2020	615	45031	15	1709	344	33987	9335	97189	8069
91	31.07.2020	657	45688	7	1716	387	34374	9598	98282	8241

92 rows × 10 columns

Narzędzie analizy danych:

Do analizy danych wykorzystano język programowania Python oraz biblioteki *Pandas* i *Matplotlib*. Wyniki przedstawiono w pliku typu *Jupyter Notebook*, a do jego stworzenia wykorzystano *Google Colab*.

Podstawowa statystyka

Najpierw dla każdej kolumny przeprowadzono podstawową statystykę, tzn. wyliczono wartość minimalną, maksymalną, średnią oraz odchylenie standardowe.

```
Nowe przypadki
Minimum: 193 Maksimum: 657 Średnia: 356.64 Odchylenie standardowe: 99.21

Przypadki (kumulatywnie)
Minimum: 13105 Maksimum: 45688 Średnia: 29164.24 Odchylenie standardowe: 9298.24

Zgony
Minimum: 2 Maksimum: 30 Średnia: 11.67 Odchylenie standardowe: 6.23

Zgony kumulatywnie
Minimum: 649 Maksimum: 1716 Średnia: 1252.17 Odchylenie standardowe: 314.72

Ozdrowieńcy
Minimum: 82 Maksimum: 754 Średnia: 335.68 Odchylenie standardowe: 155.72

Ozdrowieńcy kumulatywnie
Minimum: 3762 Maksimum: 34374 Średnia: 17383.41 Odchylenie standardowe: 9418.23

Aktywne przypadki
Minimum: 7837 Maksimum: 14099 Średnia: 10528.65 Odchylenie standardowe: 1814.76

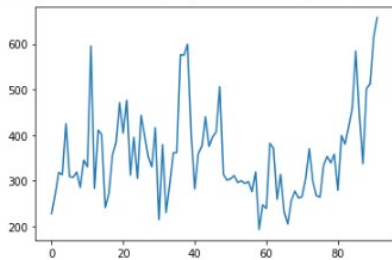
Kwarantanna
Minimum: 76645 Maksimum: 104024 Średnia: 88849.66 Odchylenie standardowe: 7244.12

Nadzór
Minimum: 6222 Maksimum: 20079 Średnia: 15324.05 Odchylenie standardowe: 4298.8
```

Dla każdej kolumny stworzono również wykres liniowy jej wartości od czasu. Jako że wskaźniki kumulatorywne okazały się być mniej więcej liniowo rosnące, poniżej przedstawiono wykresy dla pozostałych wartości.

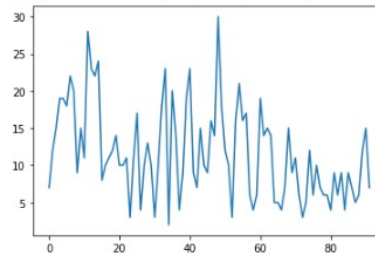
```
data['Nowe przypadki'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1cd92e10>
```



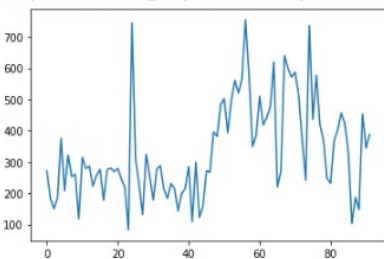
```
data['Zgony'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1cd64dd0>
```



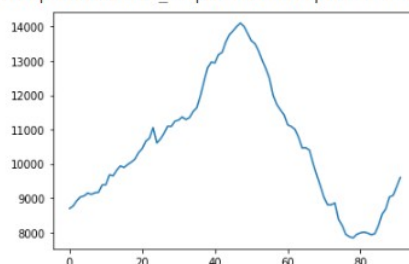
```
data['Ozdrowieńcy'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1cf05690>
```



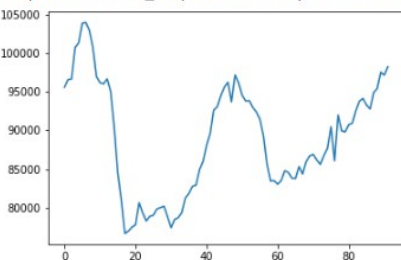
```
data['Aktywne przypadki'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1a875110>
```



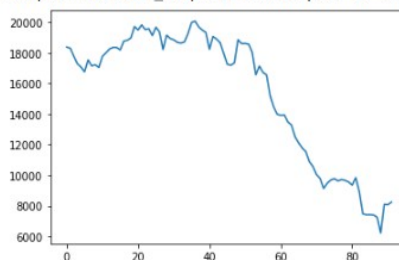
```
data['Kwarantanna'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1cebed50>
```



```
data['Nadzór'].plot.line()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1ccaa590>
```



Interesujące własności, które można zauważyć:

- zarówno liczba nowych przypadków, jak i zgonów notowały dosyć duże „skoki” w badanym okresie, jednak można zauważyć, że z czasem coraz mniej osób umierało na COVID-19, podczas gdy w lipcu nowych przypadków zaczęło przybywać
- liczba ozdrowieńców zmieniała się z dnia na dzień o wiele mniej gwałtownie, a w drugiej połowie czerwca i lipcu zaczęła oscylować wokół wyższych wartości
- liczba aktywnych przypadków rosła w miarę liniowo aż do 50. dnia badania (orientacyjnie – koniec czerwca), a następnie zaczęła gwałtownie spadać, około 80. dnia osiągnęła lokalne minimum
- liczba osób na kwarantannie największa była w maju, pod koniec miesiąca gwałtownie spadła – oprócz sporego „skoku” pod koniec czerwca od tego momentu rosła w miarę liniowo
- liczba osób pod nadzorem do połowy czerwca utrzymywała się na podobnym poziomie (ok. 20000), od tego momentu zaczęła liniowo maleć

Grupowanie danych:

Powyższe dane przydzielono na podstawie daty do każdego z 3 miesięcy, a następnie dla każdego z nich wyliczono sumaryczne i średnie wskaźniki dla wszystkich kolumn.

Sumaryczne wskaźniki:

```
[ ] monthSums = dataToGroup.groupby('Miesiąc').sum()
monthSums
```

	Nowe przypadki	Zgony	Ozdrowieńcy	Aktywne przypadki	Kwarantanna	Nadzór
Miesiąc						
5	10909	421	7958	309091	2740885	571684
6	10607	400	10342	380329	2657180	535877
7	11295	253	12583	279216	2776104	302252

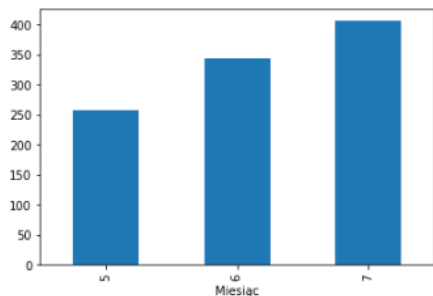
Średnie wskaźniki:

	Nowe przypadki	Zgony	Ozdrowieńcy	Aktywne przypadki	Kwarantanna	Nadzór
Miesiąc						
5	351.90	13.58	256.71	9970.68	88415.65	18441.42
6	353.57	13.33	344.73	12677.63	88572.67	17862.57
7	364.35	8.16	405.90	9006.97	89551.74	9750.06

Najbardziej interesujące okazały się wyniki dla *Ozdrowieńców*, *Aktywnych przypadków* oraz *Nadzoru*. Na wykresach przedstawiono wskaźniki średniej wymienionych kolumn.

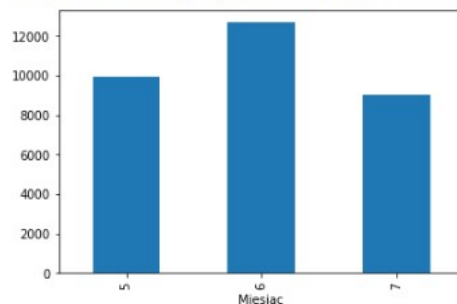
```
[ ] monthAvg['Ozdrowieńcy'].plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1a304110>



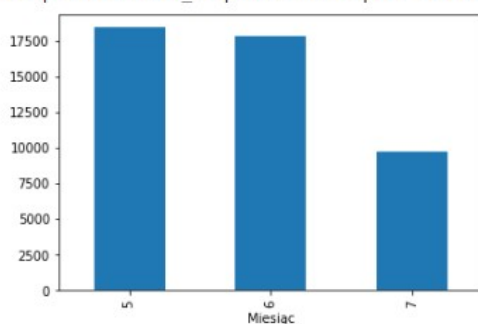
```
[ ] monthAvg['Aktywne przypadki'].plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1a2cbf10>



```
[ ] monthAvg['Nadzór'].plot.bar()
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e1a1e3bd0>



Wnioski:

W okresie przed wakacjami w 2020 roku liczba nowych przypadków wykrycia zakażenia koronawirusem wcale nie zmalała – na wykresie można zauważyć, że w samym lipcu wręcz ona wzrosła, a na przestrzeni miesięcy utrzymywała się na podobnym poziomie. Jednak stopniowo coraz więcej osób zaczęło zdrowieć, zanotowano mniej zgonów. Aktywnych przypadków choroby w lipcu było najmniej, jednak różnica wcale nie była aż tak duża. Wyraźnie zmniejszono nadzór, mimo że ludzi na kwarantannie wcale znacząco nie ubywało.

Link do Google Colabatory:

<https://colab.research.google.com/drive/1kYPmP4SGEFCNquY7qAntVNoWNB9UEvDZ?usp=sharing>