

Python - Analiza danych z modulem PANDAS

www.udemy.com (<http://www.udemy.com>) (R)

LAB - S08-L004 - pie plot

1. Zaimportuj **moduł pandas, numpy i matplotlib** i nadaj im standardowe aliasy. Dodaj instrukcję powodującą wyświetlenie wykresu generowanego przez matplotlib w jupyter notebook.
2. Uruchom poniższy fragment kodu, aby przygotować dane do rysowania wykresu (wszystkie zastosowane polecenia powinny Ci już być na tym etapie znane):

```
import datetime

#Import data
nasa = pd.read_csv("nasa_facebook_statuses.csv",
                  usecols=["status_published", "num_reactions",
                           "num_comments", "num_shares", "num_likes"])

#Convert column type to date time
nasa["status_published"] = pd.to_datetime(nasa["status_published"])

#Select observations between two datetimes - May 2015
filter =(nasa['status_published'] >= '2016-3-1') & (nasa['status_published'] < '2016-6-1')
nasa = nasa[filter]

#Add column with day only
nasa['month'] = nasa.apply(lambda row: row["status_published"].month, axis=1)

#Group all columns by sum
nasa_by_month = nasa.groupby(by='month').sum()

nasa_by_month.head()
```

3. Wyświetl wykres kołowy prezentujący serię **num_likes** znajdującą się w **nasa_by_month**
4. Dodaj do poprzedniego polecenia parametr zmieniający rozmiar wykresu na 5 x 5
5. Zmień polecenie dodając tytuł na **NASA Facebook activity** oraz zmieniający etykiety zawierające numer miesiąca na jego nazwę
6. Zmień polecenie tak, aby wygenerowane zostały 4 wykresy prezentujące wszystkie serie numeryczne z **nasa_by_month**
7. Zmień rozmiar na 10 x 10, ułożenie na "2 wykresiki u góry i 2 na dole", usuń legendę
8. W poprzednim wykresie ponownie włącz legendę, dodaj etykietę z informacją o udziale procentowym wycinka w całości, zmień kolor czcionki tej etykiety na biały

Rozwiązania:

Poniżej znajdują się propozycje rozwiązań zadań. Prawdopodobnie istnieje wiele dobrych rozwiązań, dlatego jeżeli rozwiążesz zadania samodzielnie, to najprawdopodobniej zrobisz to inaczej, może nawet lepiej :) Możesz pochwalić się swoimi rozwiązaniami w sekcji Q&A

```
In [1]: import pandas as pd
import numpy as np
import matplotlib as plt
%matplotlib inline
```

```
In [2]: import datetime

#Import data
nasa = pd.read_csv("nasa_facebook_statuses.csv",
                  usecols=["status_published", "num_reactions",
                           "num_comments", "num_shares", "num_likes"])

#Convert column type to date time
nasa["status_published"] = pd.to_datetime(nasa["status_published"])

#Select observations between two datetimes - May 2015
filter = (nasa['status_published'] >= '2016-3-1') & (nasa['status_published'] < '2016-6-1')
nasa = nasa[filter]

#Add column with day only
nasa['month'] = nasa.apply(lambda row: row["status_published"].month, axis=1)

#Group all columns by sum
nasa_by_month = nasa.groupby(by='month').sum()

nasa_by_month.head()
```

Out[2]:

| | num_reactions | num_comments | num_shares | num_likes |
|-------|---------------|--------------|------------|-----------|
| month | | | | |
| 3 | 3003815 | 49108 | 392585 | 2846361 |
| 4 | 2927616 | 120118 | 370803 | 2760385 |
| 5 | 2708977 | 46122 | 393576 | 2548960 |

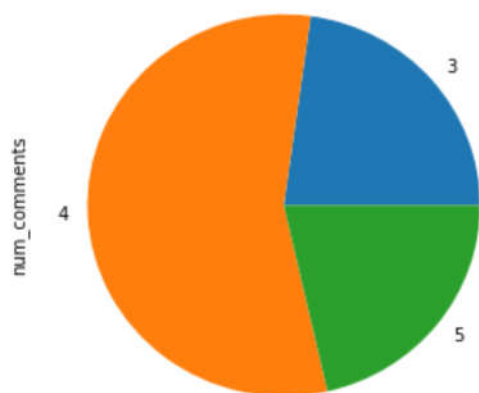
```
In [3]: nasa_by_month["num_comments"].plot(kind="pie")
```

```
Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x172116b2d30>
```



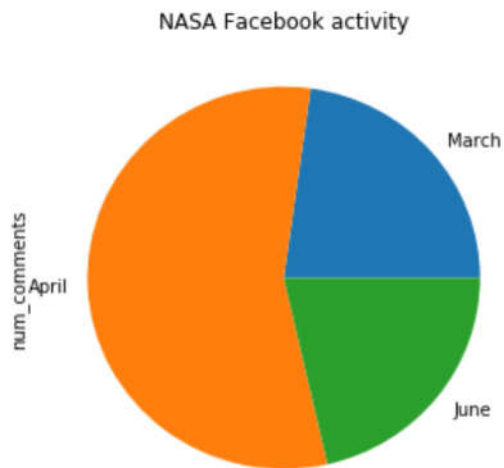
```
In [4]: nasa_by_month["num_comments"].plot(kind="pie", figsize=(5,5))
```

```
Out[4]: <matplotlib.axes._subplots.AxesSubplot at 0x172117e74e0>
```



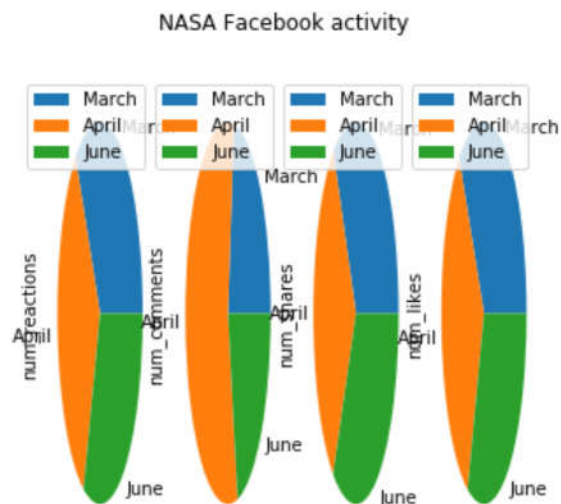
```
In [5]: nasa_by_month["num_comments"].plot(kind="pie", figsize=(5,5),
                                             labels=("March", "April", "June"),
                                             title="NASA Facebook activity")
```

```
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x17211822be0>
```



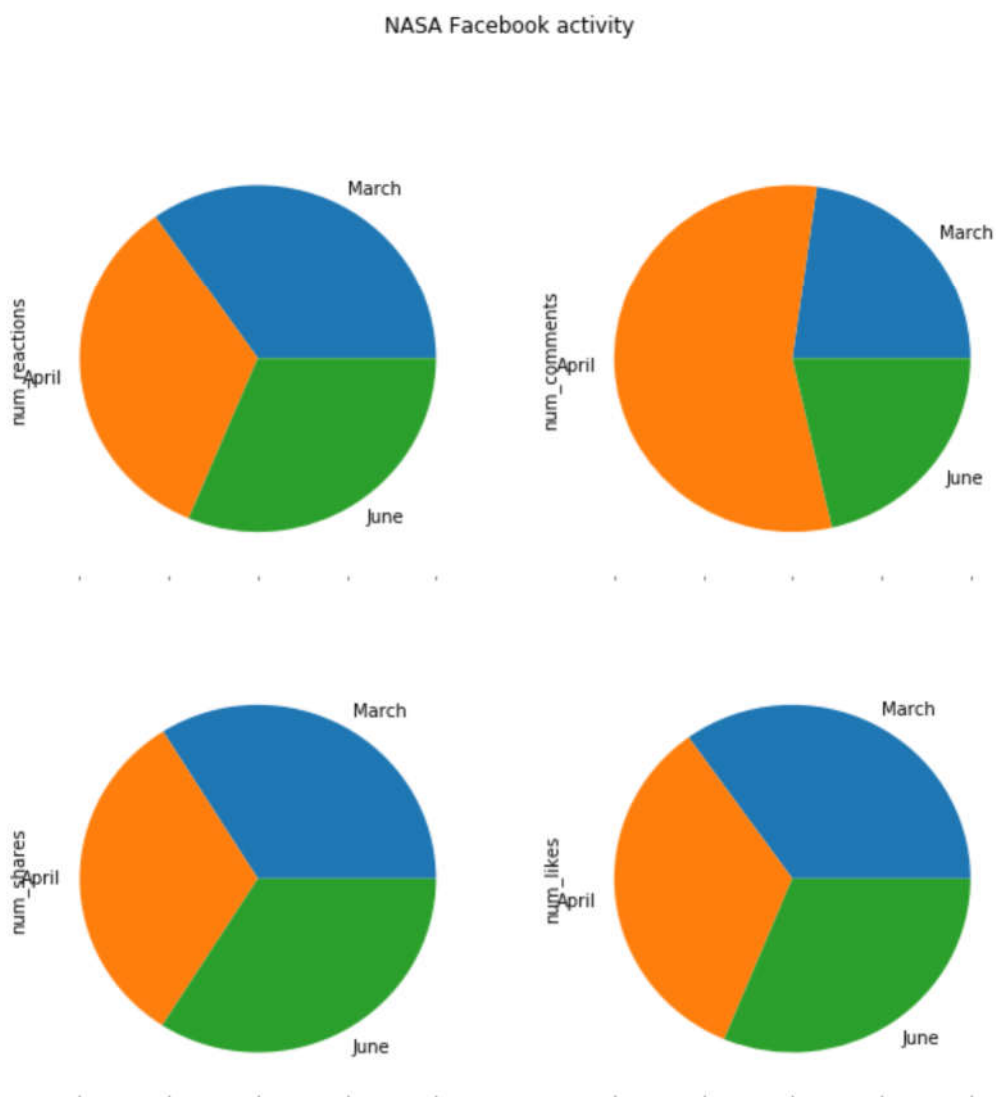
```
In [6]: nasa_by_month.plot(kind="pie", figsize=(5,5),
                             labels=("March", "April", "June"),
                             title="NASA Facebook activity",
                             subplots=True)
```

```
Out[6]: array([<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211859F60>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x0000017211ACF400>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x0000017211AF8940>,
                <matplotlib.axes._subplots.AxesSubplot object at 0x0000017211B20EB8>],
              dtype=object)
```



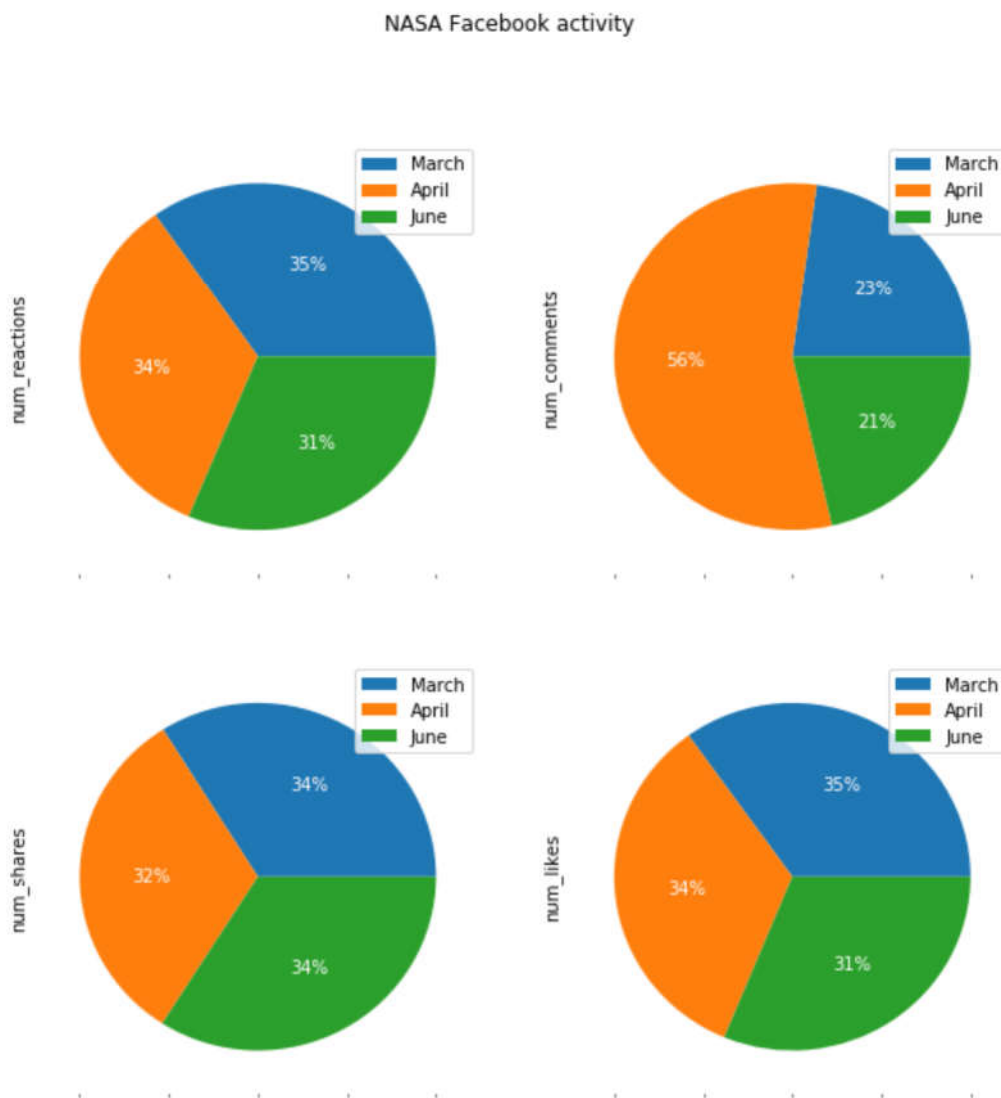
```
In [7]: nasa_by_month.plot(kind="pie", figsize=(10,10),
                        labels=("March", "April", "June"),
                        title="NASA Facebook activity",
                        subplots=True,
                        layout=(2,2), legend=False)
```

```
Out[7]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211BA0588>,
<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211BF1630>],
[<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211C19B70>,
<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211C4C128>]],
dtype=object)
```



```
In [8]: nasa_by_month.plot(kind="pie", figsize=(10,10),
                             labels=("March", "April", "June"),
                             title="NASA Facebook activity",
                             subplots=True,
                             layout=(2,2), legend=True,
                             autopct="%.0f%%", textprops=dict(color='white'))
```

```
Out[8]: array([[<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211BBBF98>,
                 <matplotlib.axes._subplots.AxesSubplot object at 0x0000017211D25F28>],
               [<matplotlib.axes._subplots.AxesSubplot object at 0x0000017211D554A8>,
                 <matplotlib.axes._subplots.AxesSubplot object at 0x0000017211D7AA20>]],
          dtype=object)
```



```
In [ ]:
```