KodeKlubben 2.0

Øvelsesgang 1

Kristian Urup Olesen Larsen, Jakob Jul Elben

October 30, 2018

Økonomisk Institut, KU

Velkommen!

Hvem er vi?

- Økonomistuderende
- RA's på Økonomisk Institut
- Arbejder typisk i Python, R, STATA eller SAS

Hvem er i?

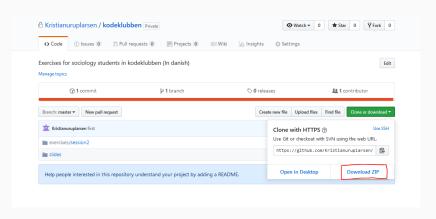
• Det er sådan set ligemeget, vi skal bare have det sjovt og lære noget data science.

1

Setup

Alt materiale ligger på GitHub

- Åben https://github.com/Kristianuruplarsen/kodeklubben
- Klik på knappen Clone or download og på Download ZIP



• Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

• Vi har slides der viser hvordan i kan løse alle spørgsmålene,

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

 Vi har slides der viser hvordan i kan løse alle spørgsmålene, men den bedste vej frem er learning by doing

- Hvad mener (folketings)partierne egentligt? Gid vi kunne spørge dem...
- DR's kandidattest fra KV17 er stadig online!
- Alle kandidaternes svar er "frit" tilgængelige.

Kort note om formattet:

- Vi har slides der viser hvordan i kan løse alle spørgsmålene, men den bedste vej frem er learning by doing
- 1) Læs dokumentation 2) spørg en sidemakker 3) google det (stackoverflow har som regel løsningen) 4) spørg os.

Projektet

Op til KV17 lavede DR en kandidattest. Alle kandidaterne har selvfølgelig svaret på spørgsmålene. De svar vil vi gerne have.

Denne gang:

- Download data fra DR
 - internet-hacks
 - Interagere med nettet gennem python
- Rens datasættet og gør klar til alt det sjove
 - jonglere med data og formatter

Næste gang:

- Alt det sjove.
 - Dimensionality reduction
 - Interaktive plot

Øvelser

Det link i leder efter er: https://www.dr.dk/tjenester/kv17-candidateapi/api/constituency/all Prøv at åbne linket i en browser!

- Hvis ikke i har installeret requests er det en god ide at gøre det nu
- Google:
 - "python requests get link"
 - "python requests json"

```
import requests

url = #[FILL IN]

response = requests.get(url)
kommuner = response.json()
```

- Kig i den json fil i har downloadet hvordan er den struktureret?
- I kan hente alle slugs med et loop:

```
1 slugs = list()
2 for x in kommuner['constituencies']:
3    if not x['slug'] == 'ikke-oplyst':
4         slugs.append(x['slug'])
5
```

• Eller med en *list comprehension*:

```
slugs = [x['slug'] for x in kommuner['constituencies']
if not x['slug'] == 'ikke-oplyst']
```

 I skal ind i jeres developer console igen - denne gang leder i en json fil med samme navn som den kommune i har søgt på



```
def hent_kommune_data(kommuneslug):
      base = 'https://www.dr.dk/tjenester/kv17-candidateapi/'
2
      url = base + f'api/ballot/{kommuneslug}'
3
4
      response = requests.get(url)
5
6
      if response.ok:
7
          return response.json()
8
      else:
9
          return None
10
11
```

```
import pandas as pd
2
3
     def rens_kommune_data(kommunejson):
4
       ballot = kommunejson['ballot']['constituency']
       parties = kommunejson['ballot']['ballotParties']
5
6
7
       _name = ballot['slug']
8
       _id = ballot['id']
9
       for party in parties:
            party_candidates = party['candidates']
12
13
           for candidate in party_candidates:
                _fname = candidate['firstName']
14
15
                _Iname = candidate['lastName']
16
17
                data = { 'id ': [_id],
18
                        'kommune': [_name],
                        'letter': [candidate['partyLetter']],
19
20
                        'party': [candidate['partyName']],
21
                        'name': ['{} {}'.format(_fname, _Iname)],
22
                        'answers': [candidate['answerString']]
24
25
                    df = pd.concat([df, pd.DataFrame(data)])
26
                except NameError:
                    df = pd.DataFrame(data)
27
28
29
       return df.reset_index(drop = True)
30
```

```
1 from time import sleep
2
3 for kom in slugs:
      sleep(2)
4
      print(kom)
5
      data = hent_kommune_data(kom)
6
      clean = rens_kommune_data(data)
7
8
      try:
9
           df = pd.concat([df, pd.DataFrame(clean)])
10
      except NameError:
11
           df = pd.DataFrame(clean)
12
```

Næste gang/tak for i dag

Næste gang/tak for i dag

- Vi lægger en rettevejledning ud til dagens øvelser på github
- Vi lægger datasættet op før næste øvelsesgang ikke nødvendigt med hjemmearbejde!

Næste gang/tak for i dag

- Vi lægger en rettevejledning ud til dagens øvelser på github
- Vi lægger datasættet op før næste øvelsesgang ikke nødvendigt med hjemmearbejde!

Næste gang:

- Samme tid, samme sted, næste fredag.
- Vi skal lave noget analyse på de data i har scrapet i dag.