```
In [1]:  import requests
         from bs4 import BeautifulSoup
         import csv
         from IPython.display import display
```

```
In [2]:  #url = "https://www.avito.ru/sevastopol/kvartiry/prodam/vtorichka-ASgBAQICAUSSA8YQAUDmBxSMUg?cd=1&p=2&proprofile=1"
         def get_html(url):
             r = requests.get(url) # запрос к серверу
             return r.text # ответ в виде текста
```

Количество страниц на авито

```
In [3]:  def get_total_pages(html):
             soup = BeautifulSoup(html, 'lxml')
             divs = soup.find('div', class_='pagination-root-2oCjZ')
             pages = divs.find_all('span', class_='pagination-item-1WyVp')[-2].get('data-marker')
             total_pages = pages.split('(')[1].split(')')[0]
             return int(total_pages)
```

```
In [4]:  url = "https://www.avito.ru/sevastopol/kvartiry/prodam/vtorichka-ASgBAQICAUSSA8YQAUDmBxSMUg?cd=1&p=1&proprofile=1"
         r = requests.get(url)
         soup = BeautifulSoup(r.text, 'lxml')
```

```
In [5]:  def get_page_data(html):
             soup = BeautifulSoup(html, 'lxml')
         #    divs = soup.find('div', class_='snippet-list js-catalog_serp')
             ads = soup.find_all('div', class_='description item_table-description')
             for ad in ads:
                 try:
                     room = ad.find('div', class_='snippet-title-row').find('span').text.strip() # метод strip убирает пропуски слева и справа

                     rooms = room.split(',')[0].split(' ')[0].split('-')[0] # метод slip Разбиение строки по разделителю и выбор нужного элемента
                 except:
                     rooms = ''
                 try:
                     are = ad.find('div', class_='snippet-title-row').find('span').text.strip()
                     area = are.split(',')[1].split(' ')[1]
                 except:
                     area = ''
                 try:
                     floo = ad.find('div', class_='snippet-title-row').find('span').text.strip()
                     floor = floo.split(',')[2].split('/')[0].split(' ')[1]
                 except:
                     floor = ''
                 try:
                     floo_t = ad.find('div', class_='snippet-title-row').find('span').text.strip()
                     floor_total = floo_t.split(',')[2].split('/')[1].split(' ')[0]
                 except:
                     floor_total = ''
         #        try:
         #            div = ad.find('div', class_='snippet-title-row').find('h3')
         #            url = "https://avito.ru" + div.find('a').get('href')
         #        except:
         #            url = ''
                 try:
                     pric = ad.find('div', class_="snippet-price-row").text.strip()
                     price = pric.strip('P').rstrip().replace(' ', '')
                 except:
                     price = ''
                 try:
                     adr = ad.find('div', itemprop="address").text.strip()
                     adress = adr.split(',')[0]
                 except:
                     adress = ''
         #        try:
         #            adr = ad.find('div', itemprop="address").text.strip()
         #            adress2 = adr.split(',')[1]
         #        except:
         #            adress2 = ''
                 try:
                     do = ad.find('div', itemprop="address").text.strip()
                     dom = do.split(',')[-1]
                 except:
                     dom = ''
                 try:
                     rielty = ad.find('div', class_="snippet-line-row").text.strip()
                 except:
                     rielty = ''
                 try:
                     day = ad.find('div', class_="snippet-date-info").get('data-tooltip') # достаем данные из 'div' методом get
                 except:
                     day = ''
                 data = {'day':day,
                         'adress':adress,
                         #'adress2':adress2,
                         'dom':dom,
                         'price':price,
                         'rooms':rooms,
                         'area':area,
                         'floor':floor,
                         'floor_total':floor_total,
                         'rielty':rielty}
         #                'url':url}
                 write_csv(data)
```

```
In [6]:  # divs = soup.find('div', class_='snippet-list js-catalog_serp')
         # ads = divs.find_all('div', class_='description item_table-description')
```

```
In [7]:  # di/vs
```

Разобъем url на части.

```
In [8]:  def write_csv(data):
             with open('avito_10_10_2020.csv', 'a', encoding='utf-8') as f:
                 writer = csv.writer(f)
                 writer.writerow( (data['day'],
                                   data['adress'],
                                   data['dom'],
                                   data['price'],
                                   data['rooms'],
                                   data['area'],
                                   data['floor'],
                                   data['floor_total'],
                                   data['rielty']))
                 #writer.writeheader(['date', 'adress', 'dom', 'price', 'room', 'total_area', 'floor', 'floor_total', 'realty'])
```

```
In [9]:  # https://www.avito.ru/sevastopol/kvartiry/prodam/vtorichka-ASgBAQICAUSSA8YQAUDmBxSMUg?cd=1&p=2&proprofile=1
```

```
In [10]: def main():
             url = "https://www.avito.ru/sevastopol/kvartiry/prodam/vtorichka-ASgBAQICAUSSA8YQAUDmBxSMUg?cd=1&p=1&proprofile=1"
             base_url = "https://www.avito.ru/sevastopol/kvartiry/prodam/vtorichka-ASgBAQICAUSSA8YQAUDmBxSMUg?cd=1&"
             page_part = "p="
             page_part_2 = "&proprofile=1"

             total_pages = get_total_pages(get_html(url))

             for i in range(1, total_pages):
                 url_gen = base_url + page_part + str(i) + page_part_2
                 html = get_html(url_gen)
                 get_page_data(html)

         if __name__ == '__main__':
             main()
```