**Scraping Law firm Address, Name, Phone numbers**

* Class Names are not properly used in the websites, So it was difficult to scrape the details completely
* So some of the sites are leftover

1. Total : 5033
2. Unique ones: 5024
3. Inactive or unsubscribed: 300 Remaining: 4724
4. Sites with no proper classnames:

No geographic location: 154

No law firm name: 7

No phone numbers: 14

No state: 203

No city: 203

* Also the geographic location or addresses are not in a proper format, means all the addresses are in same format so we were not able to scrape other states and cities outside US
* Some states are in full-form and some are in short form ( e.g. California-CA )

**Scraping Practice areas**

* Same as scraping dump, we faced problem with extracting the practice areas, As the classnames are not used properly. So we tried to extract it from the nav bar.
* So we couldn’t scrape
  + Main practice areas for 591 sites out of 4724
  + Sub practice areas for 434 sites out of 4724
* We got the practice\_areas from the superlawyers site
  + Practice areas matching with Super lawyers site – ( 4724-554 ) out of 4724

**Uploading all the datas into a postgresql table**

Combining both the dump file which has lawfirm name, address, phone number, state and city with the