

Usecase Statement:

This project is aimed for the general public to understand Practo website in the location Kakinada city. Here I scraped 3 webpages from this site using BeautifulSoup and loaded the data into Pandas dataframe. I answered queries like number of doctors listed on Practo, departments, max, min and average fees by department etc.

```
In [71]: ##Checkout: https://www.practo.com/kakinada/doctors
```

```
In [1]: import pandas as pd
from bs4 import BeautifulSoup
import requests
import seaborn as sns
import matplotlib.pyplot as plt

In [2]: n_pages=3
baseurl = "https://www.practo.com/kakinada/doctors?page="
total_names=[]
total_dept=[]
total_exp=[]
total_fees=[]
total_location=[]
total_hospital=[]
for page in range(1,n_pages+1):
    url=baseurl+str(page)
    #print(url)
    r=requests.get(url)
    #print(r)
    soup=BeautifulSoup(r.content, 'lxml')
    soup(soup)
    doc_details=soup.findall("div", attrs=("class", "u-border-general--bottom"))
    #print(len(doc_details))
    Names=[]
    dept=[]
    exp=[]
    fees=[]
    Location=[]
    hosp=[]
    for j in range(len(doc_details)):
        Names.append(doc_details[j].find("h2").get_text().strip())
        dept.append(doc_details[j].find("div", class="u-grey-3-text").find("span").get_text().strip())
        exp.append(doc_details[j].find("div", class="u-v2-spacer--xs-top").get_text().strip())
        Location.append(doc_details[j].find("div", class="u-bold u-d-inlineblock u-vv-fees.append(doc_details[j].find("span", attrs={'data-ga-id': 'consultation_fees_hosp.append(doc_details[j].find("span", attrs={'data-ga-id': 'doctor_clinic_n

total_names.extend(Names)
total_dept.extend(dept)
total_exp.extend(exp)
total_fees.extend(fees)
total_location.extend(Location)
total_hospital.extend(hosp)
#print(total_names)
#print(total_dept)
#print(total_exp)
#print(total_fees)
#print(total_location)
#print(total_hospital)

In [3]: df=pd.DataFrame()
df['Names'] = total_names
df['Dept'] = total_dept
df['Experience'] = total_exp
df['Fees'] = total_fees
df['Location'] = total_location
df['Hospital'] = total_hospital

In [4]: df

Out[4]:
```

	Names	Dept	Experience	Fees	Location	Hospital
0	Dr. B. Manohar Prasad	General Physician	23	400	Ramaraopet	Apollo Hospitals
1	Dr. S.Veerabhadra Raju	Ophthalmologist	58	150	Ramaraopet	Apollo Hospitals
2	Dr. P.P.Chatterjee	Orthopedist	20	300	Ramaraopet	Apollo Hospitals
3	Dr. D.V.S.Kiran Raju	Dentist	11	300	Temple Street	Srinivasa Multi Speciality Dental Hospital
4	Dr. M.V. Kiran Kumar	Neurosurgeon	17	300	Ramaraopet	Apollo Hospitals
5	Dr. G.V.Rao	General Surgeon	14	300	Ramaraopet	Apollo Hospitals
6	Dr. G.S.R. Murthy	Radiologist	20	300	Ramaraopet	Apollo Hospitals
7	Dr. D. Sudheer kumar	Cardiologist	17	500	Ramaraopet	Apollo Hospitals
8	Dr. K. K. Kiran Kumar Reddy	Cardiologist	10	500	Ramaraopet	Apollo Hospitals
9	Dr. T.Prem Kumar	Cardiologist	6	500	Ramaraopet	Apollo Hospitals
10	Dr. A. Neerja	Homeoeopath	14	300	Kakinada Bazar	Positive Homeoeopathy
11	Dr. R. D. Naik	Cardiothoracic Surgeon	22	300	Ramaraopet	Apollo Hospitals
12	Dr. P.Ravinder	Nephrologist	19	300	Ramaraopet	Apollo Hospitals
13	Dr. A.Ravi Kumar	Gastroenterologist	7	300	Ramaraopet	Apollo Hospitals
14	Dr. P. Raju	Pulmonologist	17	300	Ramaraopet	Apollo Hospitals
15	Dr. N. V. Srikanth	Ophthalmologist	12	150	Ramaraopet	Apollo Hospitals
16	Dr. Harshavardhan	Urologist	6	300	Ramaraopet	Sai Sudha Hospital
17	Dr. Manojna Battina	Pediatrician	Sp	300	APSP Camp	Dhanwin Child Neuro Care
18	Dr. Siva Rama Gandhi	Neurologist	14	300	APSP Camp	Dhanwin Child Neuro Care
19	Dr. Yasaswi Mamidiseti	Homeoeopath	3	300	Atchutapuratriam	Curewell Homeoeopathy
20	Dr. Amrutha Rao Thota	Gynecologist/Obstetrician	9	400	Ramaraopet	Sudha Urology and Andrology Hospital
21	Dr. K.V.Siva Rama Krishna	Homeoeopath	6	300	Suryaraopeta	Dr.Care Homeoeopathy
22	Dr. Vijaykumar Naranji	General Surgeon	13	400	Temple Street	Signia Precision Clinic
23	Dr. AM Reddy	Homeoeopath	24	999	Suryaraopeta	Dr.Care Homeoeopathy
24	Dr. T Kiran Kumar	Homeoeopath	27	300	Kakinada Bazar	Positive Homeoeopathy

```
In [5]: ##using requests module and Python's webscraping tool beautifulsoup we have gathered data
##loaded it into a pandas dataframe. Hereafter we perform data analysis of this data
```

Exploratory Data Analysis

```
In [6]: df.dtypes

Out[6]: Names          object
Dept              object
Experience         object
Fees              object
Location          object
Hospital          object
dtype: object

In [7]: df['Names'] = df['Names'].astype('string')
df['Dept'] = df['Dept'].astype('string')
df['Location'] = df['Location'].astype('string')
df['Hospital'] = df['Hospital'].astype('string')

In [8]: df.dtypes

Out[8]: Names          string
Dept              string
Experience         object
Fees              object
Location          string
Hospital          string
dtype: object

In [9]: df['Fees']=df['Fees'].astype(str).astype(int)

In [10]: df.dtypes

Out[10]: Names          string
Dept              string
Experience         object
Fees              int32
Location          string
Hospital          string
dtype: object

In [11]: df

Out[11]:
```

	Names	Dept	Experience	Fees	Location	Hospital
0	Dr. B. Manohar Prasad	General Physician	23	400	Ramaraopet	Apollo Hospitals
1	Dr. S.Veerabhadra Raju	Ophthalmologist	58	150	Ramaraopet	Apollo Hospitals
2	Dr. P.P.Chatterjee	Orthopedist	20	300	Ramaraopet	Apollo Hospitals
3	Dr. D.V.S.Kiran Raju	Dentist	11	300	Temple Street	Srinivasa Multi Speciality Dental Hospital
4	Dr. M.V. Kiran Kumar	Neurosurgeon	17	300	Ramaraopet	Apollo Hospitals
5	Dr. G.V.Rao	General Surgeon	14	300	Ramaraopet	Apollo Hospitals
6	Dr. G.S.R. Murthy	Radiologist	20	300	Ramaraopet	Apollo Hospitals
7	Dr. D. Sudheer kumar	Cardiologist	17	500	Ramaraopet	Apollo Hospitals
8	Dr. K. K. Kiran Kumar Reddy	Cardiologist	10	500	Ramaraopet	Apollo Hospitals
9	Dr. T.Prem Kumar	Cardiologist	6	500	Ramaraopet	Apollo Hospitals
10	Dr. A. Neerja	Homeoeopath	14	300	Kakinada Bazar	Positive Homeoeopathy
11	Dr. R. D. Naik	Cardiothoracic Surgeon	22	300	Ramaraopet	Apollo Hospitals
12	Dr. P.Ravinder	Nephrologist	19	300	Ramaraopet	Apollo Hospitals
13	Dr. A.Ravi Kumar	Gastroenterologist	7	300	Ramaraopet	Apollo Hospitals
14	Dr. P. Raju	Pulmonologist	17	300	Ramaraopet	Apollo Hospitals
15	Dr. N. V. Srikanth	Ophthalmologist	12	150	Ramaraopet	Apollo Hospitals
16	Dr. Harshavardhan	Urologist	6	300	Ramaraopet	Sai Sudha Hospital
17	Dr. Manojna Battina	Pediatrician	Sp	300	APSP Camp	Dhanwin Child Neuro Care
18	Dr. Siva Rama Gandhi	Neurologist	14	300	APSP Camp	Dhanwin Child Neuro Care
19	Dr. Yasaswi Mamidiseti	Homeoeopath	3	300	Atchutapuratriam	Curewell Homeoeopathy
20	Dr. Amrutha Rao Thota	Gynecologist/Obstetrician	9	400	Ramaraopet	Sudha Urology and Andrology Hospital
21	Dr. K.V.Siva Rama Krishna	Homeoeopath	6	300	Suryaraopeta	Dr.Care Homeoeopathy
22	Dr. Vijaykumar Naranji	General Surgeon	13	400	Temple Street	Signia Precision Clinic
23	Dr. AM Reddy	Homeoeopath	24	999	Suryaraopeta	Dr.Care Homeoeopathy
24	Dr. T Kiran Kumar	Homeoeopath	27	300	Kakinada Bazar	Positive Homeoeopathy

```
In [12]: df.loc[12]['Experience']

Out[12]: '19'
```

```
In [13]: df.iat[17, df.columns.get_loc('Experience')] = 11
```

```
In [14]: df

Out[14]:
```

	Names	Dept	Experience	Fees	Location	Hospital
0	Dr. B. Manohar Prasad	General Physician	23	400	Ramaraopet	Apollo Hospitals
1	Dr. S.Veerabhadra Raju	Ophthalmologist	58	150	Ramaraopet	Apollo Hospitals
2	Dr. P.P.Chatterjee	Orthopedist	20	300	Ramaraopet	Apollo Hospitals
3	Dr. D.V.S.Kiran Raju	Dentist	11	300	Temple Street	Srinivasa Multi Speciality Dental Hospital
4	Dr. M.V. Kiran Kumar	Neurosurgeon	17	300	Ramaraopet	Apollo Hospitals
5	Dr. G.V.Rao	General Surgeon	14	300	Ramaraopet	Apollo Hospitals
6	Dr. G.S.R. Murthy	Radiologist	20	300	Ramaraopet	Apollo Hospitals
7	Dr. D. Sudheer kumar	Cardiologist	17	500	Ramaraopet	Apollo Hospitals
8	Dr. K. K. Kiran Kumar Reddy	Cardiologist	10	500	Ramaraopet	Apollo Hospitals
9	Dr. T.Prem Kumar	Cardiologist	6	500	Ramaraopet	Apollo Hospitals
10	Dr. A. Neerja	Homeoeopath	14	300	Kakinada Bazar	Positive Homeoeopathy
11	Dr. R. D. Naik	Cardiothoracic Surgeon	22	300	Ramaraopet	Apollo Hospitals
12	Dr. P.Ravinder	Nephrologist	19	300	Ramaraopet	Apollo Hospitals
13	Dr. A.Ravi Kumar	Gastroenterologist	7	300	Ramaraopet	Apollo Hospitals
14	Dr. P. Raju	Pulmonologist	17	300	Ramaraopet	Apollo Hospitals
15	Dr. N. V. Srikanth	Ophthalmologist	12	150	Ramaraopet	Apollo Hospitals
16	Dr. Harshavardhan	Urologist	6	300	Ramaraopet	Sai Sudha Hospital
17	Dr. Manojna Battina	Pediatrician	Sp	300	APSP Camp	Dhanwin Child Neuro Care
18	Dr. Siva Rama Gandhi	Neurologist	14	300	APSP Camp	Dhanwin Child Neuro Care
19	Dr. Yasaswi Mamidiseti	Homeoeopath	3	300	Atchutapuratriam	Curewell Homeoeopathy
20	Dr. Amrutha Rao Thota	Gynecologist/Obstetrician	9	400	Ramaraopet	Sudha Urology and Andrology Hospital
21	Dr. K.V.Siva Rama Krishna	Homeoeopath	6	300	Suryaraopeta	Dr.Care Homeoeopathy
22	Dr. Vijaykumar Naranji	General Surgeon	13	400	Temple Street	Signia Precision Clinic
23	Dr. AM Reddy	Homeoeopath	24	999	Suryaraopeta	Dr.Care Homeoeopathy
24	Dr. T Kiran Kumar	Homeoeopath	27	300	Kakinada Bazar	Positive Homeoeopathy

```
In [15]: df.dtypes

Out[15]: Names          string
Dept              string
Experience         int32
Fees              int32
Location          string
Hospital          string
dtype: object

In [16]: df['Experience'] = df['Experience'].astype(str).astype(int)

In [17]: df.dtypes

Out[17]: Names          string
Dept              string
Experience         int32
Fees              int32
Location          string
Hospital          string
dtype: object
```

Questions Asked:

- Number of doctors listed on Practo
- Department having highest number of Doctors
- Hospital max listed on Practo
- Max, Min and Avg fee by department
- Practo listings by location
- Number of Doctors by Department

```
In [20]: ##Q1.
total doc listings=len(df)
print(f'There are {total_doc_listings} doctors listed in Practo Kakinada')

There are 25 doctors listed in Practo Kakinada
```

```
In [21]: ##Q2.
df['Dept'].value_counts()

Out[21]: Homeoeopath          5
Cardiologist              3
Ophthalmologist           2
General Surgeon           2
Neurosurgeon              1
Dentist                   1
Gynecologist/Obstetrician 1
Nephrologist              1
Pediatrician              1
Radiologist               1
Pulmonologist             1
Cardiothoracic Surgeon    1
Neurologist               1
Gastroenterologist        1
Urologist                 1
General Physician         1
Orthopedist               1
Name: Dept, dtype: Int64
```

```
In [ ]: ##Interestingly Homeoeopathy doctors are dominant in Practo listings Kakinada
```

```
In [22]: ##Q3.
df['Hospital'].value_counts()

Out[22]: Apollo Hospitals          14
Positive Homeoeopathy             2
Dhanwin Child Neuro Care          2
Dr.Care Homeoeopathy              2
Sai Sudha Hospital                1
Signia Precision Clinic            1
Sudha Urology and Andrology Hospital 1
Srinivasa Multi Speciality Dental Hospital 1
Curewell Homeoeopathy             1
Name: Hospital, dtype: Int64
```

```
In [23]: ##Interesting note: Apollo hospitals is max listed on Practo website while others are less
```

```
In [41]: ##Q5.
Locs=df['Location'].value_counts()
Locs

Out[41]: Ramaraopet          16
Temple Street              2
Suryaraopeta               2
Kakinada Bazar             2
APSP Camp                  2
Atchutapuratriam           1
Name: Location, dtype: Int64
```

```
In [42]: ##Ramaraopeta has highest number of Practo listings, 16.
```

```
In [50]: ##Q4.
Max_Fees=df['Fees'].max()
print("Max fees in Practo Kakinada is", Max_Fees)

Max fees in Practo Kakinada is 999
```

```
In [59]: Max_Fees_Dept=df.loc[df.Fees == 999, 'Dept']
Max_Fees_Dept=df.loc[df.Fees==999, 'Names']

In [57]: print("Department having max fees on Practo Kakinada is", Max_Fees_Dept)
print("Doctor having max fees on Practo Kakinada is", Max_Fees_Dept)

Department having max fees on Practo Kakinada is 23      Homeoeopath
Name: Dept, dtype: string
Doctor having max fees on Practo Kakinada is 23      Dr. AM Reddy
Name: Names, dtype: string
```

```
In [60]: ##Max fees is collected by a Homeoeopathy Doctor in Kakinada pertaining to Practo Kakinada
```

```
In [61]: Min_Fees=df['Fees'].min()
print("Min fees in Practo Kakinada is", Min_Fees)

Min fees in Practo Kakinada is 150
```

```
In [62]: Min_Fees_Dept=df.loc[df.Fees == 150, 'Dept']
Min_Fees_Dept=df.loc[df.Fees==150, 'Names']
print("Department having min fees on Practo Kakinada is", Min_Fees_Dept)
print("Doctor having min fees on Practo Kakinada is", Min_Fees_Dept)

Department having min fees on Practo Kakinada is 1      Ophthalmologist
Name: Dept, dtype: string
Doctor having min fees on Practo Kakinada is 1      Dr. S Veerabhadra Raju
Name: Names, dtype: string
```

```
In [63]: ##Min fees 150 is collected by an Ophthalmologist Dr. S Veerabhadra Raju
```

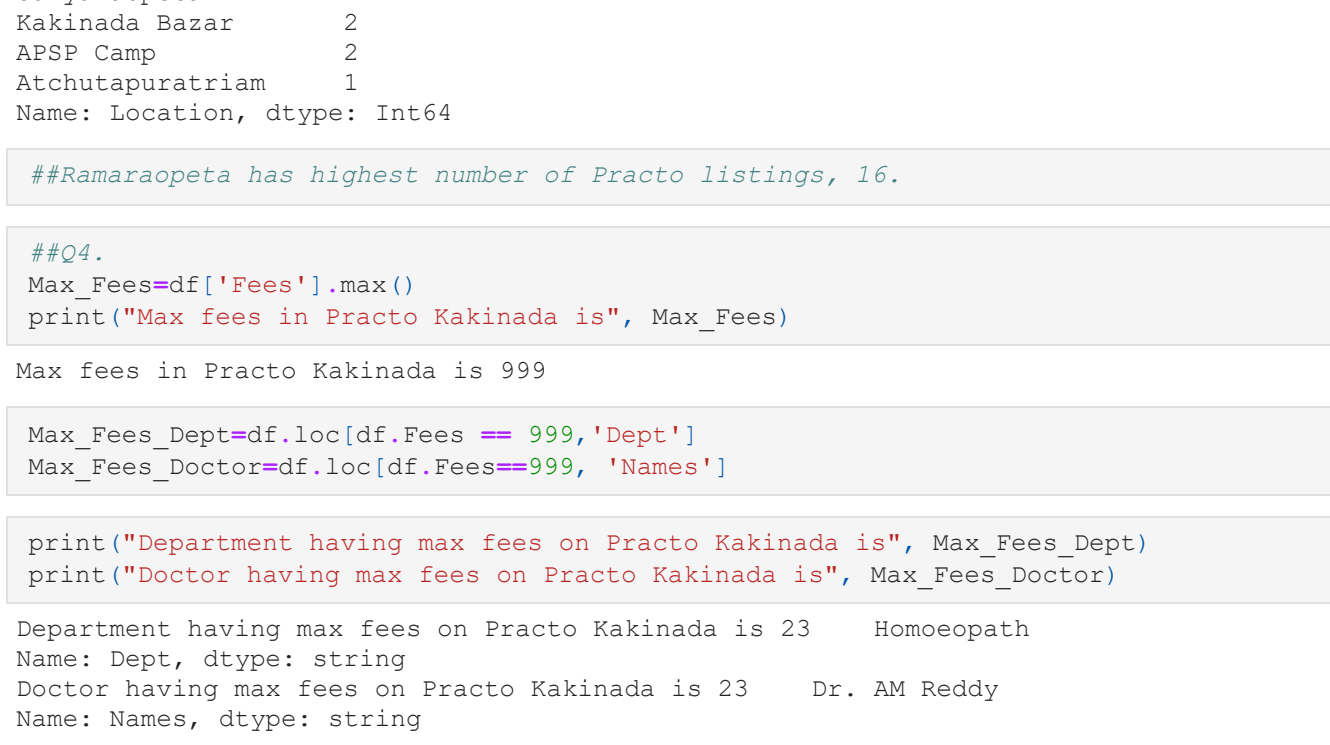
```
In [65]: Avg_Fees=df['Fees'].mean()
print("Avg fees in Practo Kakinada is", Avg_Fees)

Avg fees in Practo Kakinada is 351.96
```

Data Visualization

```
In [ ]: ##Plotting number of doctors by Department
```

```
In [66]: plt.figure(figsize=(15,10))
ax=sns.countplot(x='Dept', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.tight_layout()
plt.title('Number of doctors by Department')
plt.show()
```

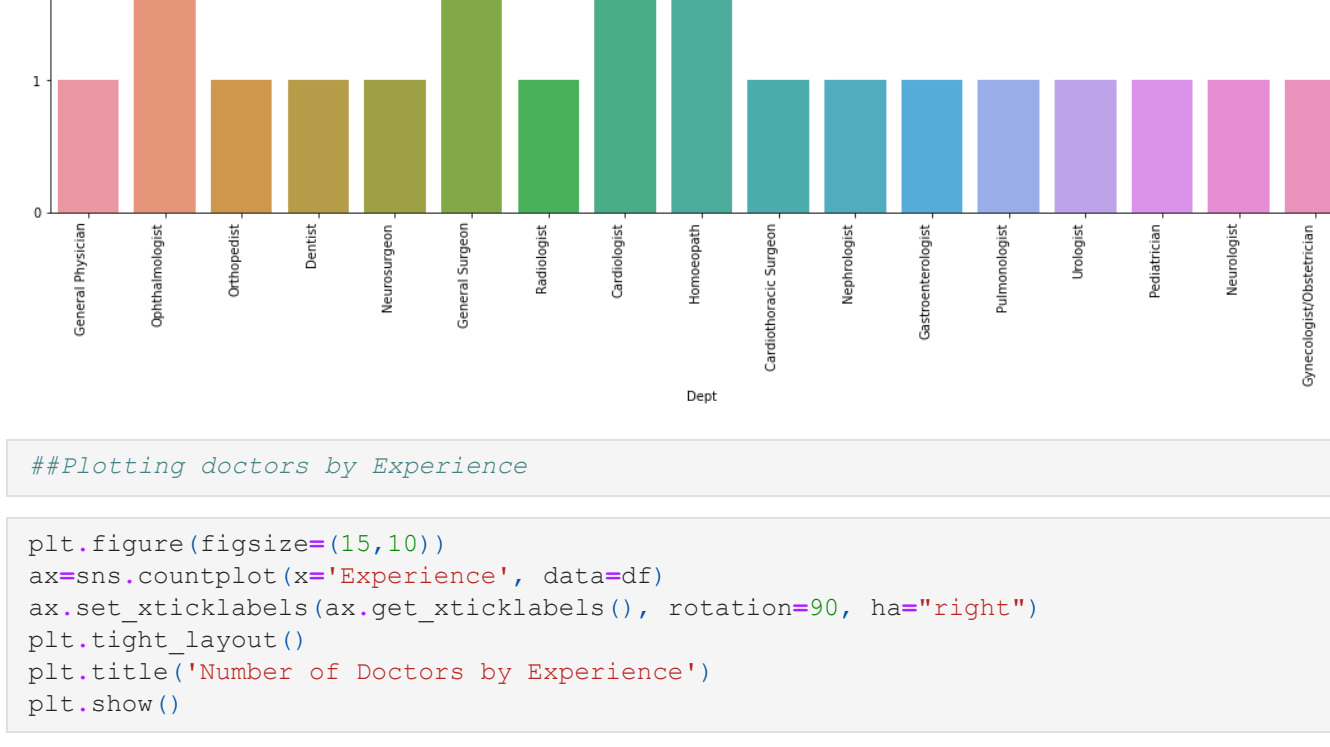


```
In [67]: ##Plotting doctors by Experience
```

```
In [68]: plt.figure(figsize=(15,10))
ax=sns.countplot(x='Experience', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.tight_layout()
plt.title('Number of Doctors by Experience')
plt.show()
```



```
In [69]: plt.figure(figsize=(15,10))
ax=sns.countplot(x='Hospital', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.tight_layout()
plt.title('Hospital breakup')
plt.show()
```



Conclusion

```
In [70]: plt.figure(figsize=(15,10))
ax=sns.countplot(x='Location', data=df)
ax.set_xticklabels(ax.get_xticklabels(), rotation=90, ha="right")
plt.tight_layout()
plt.title('Location breakup')
plt.show()
```



```
In [ ]: ##There are only 25 doctors listed on Practo Kakinada, we can clearly say that this w
##Homeoeopathy Doctors are (five of them) are dominant on Practo when compared to other
##Max fees 999 is collected by a Homeoeopathy Doctor
##Apollo hospitals is max listed on Practo with 14 listings
##Ramaraopeta location had max listings on Practo
```