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. ##Checkout: https://www.practo.com/kakinada/doctors 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') 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").ge exp.append(doc details[j].find("div", class ="uv2-spacer--xs-top").get text() Location.append(doc details[j].find("div", class = "u-bold u-d-inlineblock u-ve fees.append(doc details[j].find("span", attrs={'data-qa-id': 'consultation fee hosp.append(doc details[j].find("span", attrs={'data-qa-id': 'doctor clinic ne 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) 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 Experience Hospital General Physician Dr. B. Manohar Prasad 400 23 Ramaraopet **Apollo Hospitals** Dr. S Veerabhadra Ophthalmologist 150 Ramaraopet Apollo Hospitals 58 Raju Dr. P.P.Chatterjee Orthopedist 2 20 300 Ramaraopet **Apollo Hospitals** Srinivasa Multi 3 Dr. D.V.S.Kiran Raju 300 Dentist 11 Temple Street Speciality Dental Hospital Dr. M.V. Kiran Kumar Neurosurgeon 300 Apollo Hospitals 4 Ramaraopet 5 Dr. G.V.Rao General Surgeon 300 14 Ramaraopet Apollo Hospitals Dr. G.S.R. Murthy **Apollo Hospitals** 6 Radiologist 20 300 Ramaraopet 7 Dr. D. Sudheer kumar Cardiologist 17 500 Ramaraopet Apollo Hospitals Dr. K. K. Kiran Kumar 8 Cardiologist 500 10 Ramaraopet **Apollo Hospitals** Reddy Dr. T.Prem Kumar 500 Cardiologist 6 Ramaraopet Apollo Hospitals 10 300 Kakinada Bazar Dr. A. Neerja Homoeopath 14 Positive Homeopathy Dr. R. D. Naik Cardiothoracic Surgeon 22 300 Ramaraopet Apollo Hospitals Dr. P.Ravinder Nephrologist 300 Apollo Hospitals 19 Ramaraopet Dr. A.Ravi Kumar 7 300 Gastroenterologist Ramaraopet **Apollo Hospitals** Pulmonologist Apollo Hospitals Dr. P. Raju 300 Ramaraopet Dr. N. V. Srikanth 15 Ophthalmologist 12 150 Ramaraopet Apollo Hospitals 16 Dr. Harshavardhan Urologist 6 300 Ramaraopet Sai Sudha Hospital Dhanwin Child Neuro 17 300 **APSP Camp** Dr. Manojna Battina Pediatrician Sp Dhanwin Child Neuro Dr. Siva Rama Gandhi Neurologist 14 300 **APSP Camp** Care Dr. Yasaswi Curewell 19 Homoeopath 3 300 Atchutapuratriam Mamidisetti Homoeopathy Dr. Amrutha Rao Sudha Urology and 20 400 Gynecologist/Obstetrician 9 Ramaraopet Thota **Andrology Hospital** Dr. K.V.Siva Rama 21 300 Homoeopath 6 Suryaraopeta Dr.Care Homeopathy Krishna Dr. Vijaykumar Naranji 400 General Surgeon 13 Temple Street Signia Precision Clinic Dr. AM Reddy 999 23 Homoeopath 24 Suryaraopeta Dr.Care Homeopathy Dr. T Kiran Kumar 300 Kakinada Bazar Positive Homeopathy Homoeopath 27 ##using requests module and Python's webscraping tool beautifulsoup we have gathered ##loaded it into a pandas dataframe. Hereafter we perform data analysis of this data **Exploratory Data Analysis** df.dtypes object Names Dept object Experience object Fees object Location object object Hospital dtype: object 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 string Names string Dept object Experience Fees object Location string Hospital string dtype: object df['Fees']=df['Fees'].astype(str).astype(int) df.dtypes Names string Dept string Experience object Fees int32 Location string Hospital string dtype: object df Location **Names** Hospital Dept Experience Fees General Physician 0 Dr. B. Manohar Prasad 23 400 **Apollo Hospitals** Ramaraopet Dr. S Veerabhadra Ophthalmologist 58 150 **Apollo Hospitals** Ramaraopet Raju 2 Dr. P.P.Chatterjee Orthopedist 20 300 **Apollo Hospitals** Ramaraopet Srinivasa Multi Speciality Dental 3 Dr. D.V.S.Kiran Raju Dentist 300 11 Temple Street Hospital 4 Dr. M.V. Kiran Kumar Neurosurgeon 17 300 Apollo Hospitals Ramaraopet 5 Dr. G.V.Rao General Surgeon 14 300 Ramaraopet **Apollo Hospitals** 6 Dr. G.S.R. Murthy Radiologist 20 300 Apollo Hospitals Ramaraopet Dr. D. Sudheer kumar 500 Ramaraopet Cardiologist 17 **Apollo Hospitals** Dr. K. K. Kiran Kumar 8 10 500 Cardiologist Ramaraopet **Apollo Hospitals** Reddy 9 6 500 Dr. T.Prem Kumar Cardiologist Ramaraopet Apollo Hospitals 10 Dr. A. Neerja Homoeopath 14 300 Kakinada Bazar Positive Homeopathy Dr. R. D. Naik Cardiothoracic Surgeon 300 11 22 Ramaraopet **Apollo Hospitals** Apollo Hospitals 12 Dr. P.Ravinder 19 300 Nephrologist Ramaraopet 13 Dr. A.Ravi Kumar Gastroenterologist 7 300 **Apollo Hospitals** Ramaraopet 14 Dr. P. Raju Pulmonologist 17 300 Apollo Hospitals Ramaraopet Dr. N. V. Srikanth Ophthalmologist 15 12 150 Ramaraopet Apollo Hospitals 16 Dr. Harshavardhan 6 300 Sai Sudha Hospital Urologist Ramaraopet **Dhanwin Child Neuro** 17 Dr. Manojna Battina Pediatrician Sp 300 **APSP Camp Dhanwin Child Neuro** 18 Dr. Siva Rama Gandhi 14 300 **APSP Camp** Neurologist Care Dr. Yasaswi Curewell 19 Homoeopath 3 300 Atchutapuratriam Mamidisetti Homoeopathy Sudha Urology and Dr. Amrutha Rao Gynecologist/Obstetrician 20 9 400 Ramaraopet Andrology Hospital Thota Dr. K.V.Siva Rama 21 6 300 Homoeopath Suryaraopeta Dr.Care Homeopathy Krishna Dr. Vijaykumar Naranji 22 13 400 Signia Precision Clinic General Surgeon Temple Street 23 Dr. AM Reddy Homoeopath 24 999 Suryaraopeta Dr.Care Homeopathy 24 Dr. T Kiran Kumar Homoeopath 27 300 Kakinada Bazar Positive Homeopathy df.loc[12]['Experience'] 1191 df.iat[17, df.columns.get loc('Experience')] = 11 df [14]: Out[14]: **Names Dept** Location Hospital Experience Fees 0 Dr. B. Manohar Prasad General Physician 400 23 Ramaraopet Apollo Hospitals Dr. S Veerabhadra 1 58 150 Ophthalmologist Ramaraopet **Apollo Hospitals** Raju 2 Dr. P.P.Chatterjee Orthopedist 20 300 Ramaraopet **Apollo Hospitals** Srinivasa Multi 300 3 Dr. D.V.S.Kiran Raju Dentist 11 Temple Street Speciality Dental Hospital 4 Dr. M.V. Kiran Kumar Neurosurgeon 300 17 Ramaraopet Apollo Hospitals 5 Dr. G.V.Rao General Surgeon 300 Ramaraopet **Apollo Hospitals** 14 20 6 Dr. G.S.R. Murthy Apollo Hospitals Radiologist 300 Ramaraopet Dr. D. Sudheer kumar 7 Cardiologist 500 **Apollo Hospitals** 17 Ramaraopet Dr. K. K. Kiran Kumar 8 500 Cardiologist 10 Ramaraopet Apollo Hospitals Reddy 500 9 6 Dr. T.Prem Kumar Cardiologist Ramaraopet **Apollo Hospitals** 300 10 Homoeopath 14 Positive Homeopathy Dr. A. Neerja Kakinada Bazar Dr. R. D. Naik 11 Cardiothoracic Surgeon 300 Ramaraopet **Apollo Hospitals** 22 Dr. P.Ravinder 12 300 Nephrologist 19 Ramaraopet Apollo Hospitals Dr. A.Ravi Kumar 300 **Apollo Hospitals** Gastroenterologist 7 Ramaraopet 13 14 Dr. P. Raju 300 **Apollo Hospitals** Pulmonologist 17 Ramaraopet Dr. N. V. Srikanth **Apollo Hospitals** 15 Ophthalmologist 150 Ramaraopet 12 300 Ramaraopet 16 Dr. Harshavardhan Urologist 6 Sai Sudha Hospital Dhanwin Child Neuro **APSP Camp** 300 17 Dr. Manojna Battina Pediatrician 11 Dhanwin Child Neuro 18 Dr. Siva Rama Gandhi Neurologist 14 300 **APSP Camp** Care Dr. Yasaswi Curewell 19 Homoeopath 3 300 Atchutapuratriam Mamidisetti Homoeopathy Dr. Amrutha Rao Sudha Urology and 20 9 400 Gynecologist/Obstetrician Ramaraopet Thota **Andrology Hospital** Dr. K.V.Siva Rama 21 300 Dr.Care Homeopathy 6 Homoeopath Suryaraopeta Krishna Dr. Vijaykumar Naranji Signia Precision Clinic 22 General Surgeon 13 400 Temple Street Suryaraopeta Dr.Care Homeopathy 23 Dr. AM Reddy Homoeopath 999 24 Dr. T Kiran Kumar 27 300 24 Homoeopath Kakinada Bazar Positive Homeopathy df.dtypes Names string Dept string Experience object int32 Fees Location string Hospital string dtype: object In [16]: df['Experience'] = df['Experience'].astype(str).astype(int) df.dtypes Out[17]: Names string string Dept Experience int32 int32 Fees Location string Hospital string dtype: object **Questions Asked:** 1. Number of doctors listed on Practo 2. Department having highest number of Doctors 3. Hospital max listed on Practo 4. Max, Min and Avg fee by department 5.Practo listings by location 6. Number of Doctors by Departmnent ##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 ##Q2. df['Dept'].value counts() Out[21]: Homoeopath Cardiologist 3 Ophthalmologist General Surgeon Neurosurgeon Dentist Gynecologist/Obstetrician Nephrologist Pediatrician Radiologist Pulmonologist Cardiothoracic Surgeon Neurologist Gastroenterologist Urologist General Physician Orthopedist Name: Dept, dtype: Int64 In []: | ##Interestingly Homeopathy doctors are dominant in Practo listings Kakinada ##Q3. df['Hospital'].value_counts() Out[22]: Apollo Hospitals 14 Positive Homeopathy 2 Dhanwin Child Neuro Care Dr.Care Homeopathy Sai Sudha Hospital Signia Precision Clinic Sudha Urology and Andrology Hospital Srinivasa Multi Speciality Dental Hospital 1 Curewell Homoeopathy Name: Hospital, dtype: Int64 In [23]: ##Interesting note: Apollo hospitals is max listed on Practo website while others are In [41]: ##Q5. Locs=df['Location'].value counts() Out[41]: Ramaraopet 16 Temple Street Suryaraopeta Kakinada Bazar APSP Camp Atchutapuratriam 1 Name: Location, dtype: Int64 ##Ramaraopeta has highest number of Practo listings, 16. In [42]: ##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 Doctor=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 Doctor) Department having max fees on Practo Kakinada is 23 Name: Dept, dtype: string Doctor having max fees on Practo Kakinada is 23 Dr. AM Reddy Name: Names, dtype: string ##Max fees is collected by a Homeopathy Doctor in Kakinada pertaining to Practo Kakin 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_Doctor=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_Doctor) Department having min fees on Practo Kakinada is 1 Ophthalmologist Ophthalmologist Name: Dept, dtype: string Doctor having min fees on Practo Kakinada is 1 Dr. S Veerabhadra Raju Dr. N. V. Srikanth Name: Names, dtype: string In [63]: ##Min fees 150 is collected by an Opthamologist 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() Number of doctors by Department count General Surgeon Radiologist Gastroenterologist Cardiothoracic Surgeor Dept ##Plotting doctors by Experience In [67]: 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() Number of Doctors by Experience 3.0 2.5 2.0 1.0 0.5 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() Hospital breakup 12 10 count Srinivasa Multi Speciality Dental Hospital Dhanwin Child Neuro Care Sudha Urology and Andrology Hospital Signia Precision Clinic Sai Sudha Hospital Conclusion 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() Location breakup 16 14 12 10 count **Emple Street** APSP Camp Suryaraopeta Kakinada Bazar Atchutapuratriam Location ##There are only 25 doctors listed on Practo Kakinada, we can clearly say that this we ##Homeopathy Docotrs are (five of them) are dominant on Practo when compared to other ##Max fees 999 is collected by a Homeopathy Doctor ##Apollo hospitals is max listed on Practo with 14 listings ##Ramarao Peta location had max listings on Practo