Analysis Of Practo website

Github-
https://github.com/Nakularora05/Python-Project-
/blob/main/pp.ipynb
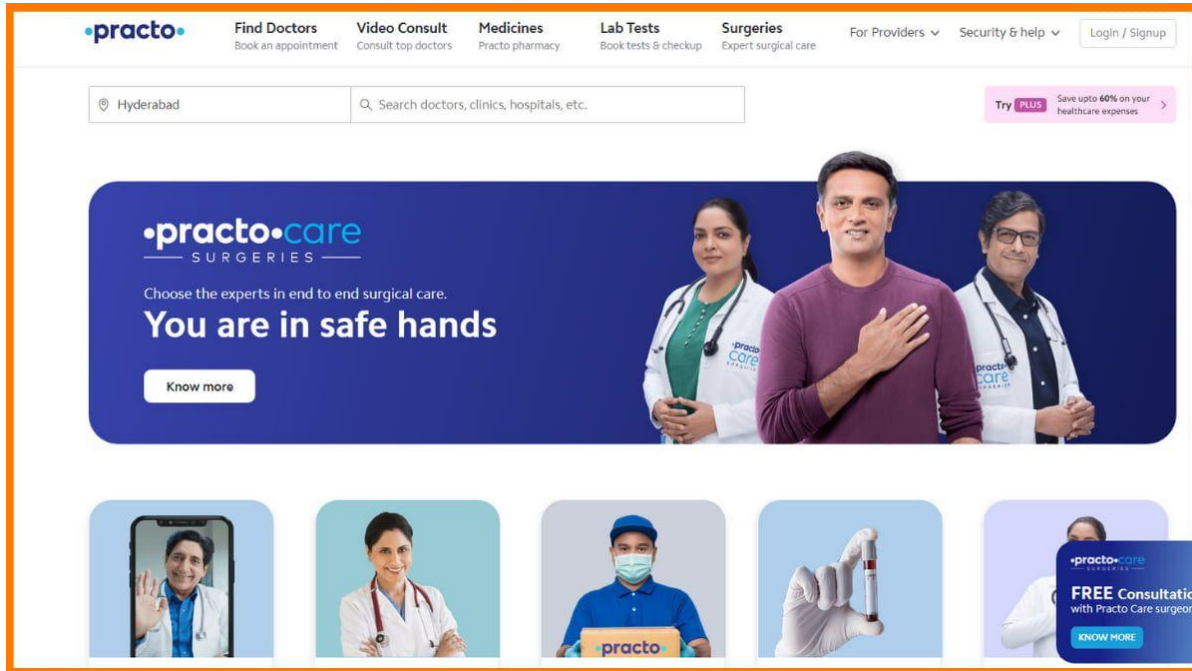
Nakul Arora
045033

INTRODUCTION

Practo is a prominent healthcare technology startup that was founded in 2008 by Shashank ND and Abhinav Lal in Bangalore, India. The company has gained widespread recognition for its innovative platform, which aims to simplify and improve the healthcare experience for both patients and healthcare providers.

Practo offers a range of digital solutions and services that connect patients with doctors, hospitals, diagnostic labs, and pharmacies. Some of the key features and services provided by Practo include: Doctor Discovery ,Telemedicine, Health Records.

Objective:

The objective of web scraping Practo's website is to collect comprehensive and up-to-date data on healthcare professionals in the Delhi region. By extracting information such as doctors' names, specializations, clinic or hospital affiliations, and patient reviews, this web scraping endeavor aims to create a valuable database that can serve various purposes. This data can be used for building directories, assisting patients in finding suitable healthcare providers, conducting market research, or facilitating outreach and partnerships between medical professionals and healthcare-related businesses. The scraped data will provide a valuable resource for improving healthcare accessibility and information dissemination in the Delhi area, ultimately enhancing the healthcare experience for both patients and practitioners.

Web Scraping:

- Web Scraping is a technique to extract a large amount of data from several websites.
- Web Scraping extracts the data from websites in unstructured format.
- It helps to collect these unstructured data and convert it in a structured form.
- Requests library is used for making HTTP requests to a specific URL and returns the response
- Beautiful Soap is used to extract information from the HTML and XML files.

DATA COLLECTION:

Practo -

https://www.practo.com

LIBRARIES USED:

1. Web Scraping Tools
- Python
- Beautiful Soup
- Requests

2. DATA CLEANING AND MANUPULATING
- NumPy
- Pandas
- Regular Expression

3. DATA VISUALIZATION
- Matplotlib
- Seaborn
- Plotly

INTERFACE OF THE WEBSITE ALONG WITH HTML

RAW DATA COLLECTED

```
['Dr. Dhwani Mago', 'Dr. Malvika Sabharwal', 'Dr. Shilpi Juneja', 'Dr. Shivani Sachdev Gour', 'Dr. Parul Katiyar', 'Dr. Keerti
Khetan', 'Dr. Yuvakshi Juneja', 'Dr. Shakuntla Kumar', 'Dr. Shikha Gurnani', 'Dr. Prof. Sadhana Kala']
10
----------------------------------------------------------------------------------------
['Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi']
10
----------------------------------------------------------------------------------------
['Greater Kailash Part 1,', 'Karol Bagh,', 'Greater Kailash Part 1,', 'Greater Kailash Part 1,', 'Defence Colony,', 'Pusa Roa
d,', 'Chirag Delhi,', 'Kingsway Camp,', 'Dwarka,', 'Kalkaji,']
10
----------------------------------------------------------------------------------------
['1000', '1000', '1000', '1000', '1500', '1500', '1000', '800', '1500']
9
----------------------------------------------------------------------------------------
['99% ', '87% ', '100% ', '61% ', '95% ', '91% ', '98% ', '98% ', '98% ', '98% ']
10
----------------------------------------------------------------------------------------
['16', '42', '18', '25', '20', '30', '41', '36', '20', '45']
10
----------------------------------------------------------------------------------------
['Dr. Dheeraj Setia', 'Dr. Geetanjali Marya', 'Dr. Harveen Nakra', 'Dr. Kapil Saroha', 'Dr. Radhika Sahni', 'Dr. Pallav Kinra',
'Dr. Karan Marya', 'Dr. Tanvir Singh', 'Dr. Deepak Nakra', 'Dr. Gunita Singh']
10
----------------------------------------------------------------------------------------
['Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi']
10
----------------------------------------------------------------------------------------
['Panchsheel Park,', 'Greater Kailash Part 2,', 'Punjabi Bagh,', 'Rohini,', 'New Friends Colony,', 'Sheikh Sarai,', 'Greater Ka
ilash Part 2,', 'Chattarpur,', 'Punjabi Bagh,', 'Chattarpur,']
10
----------------------------------------------------------------------------------------
['1000', '800', '500', '500', '500', '800', '500', '500', '500']
9
----------------------------------------------------------------------------------------
['98% ', '98% ', '100% ', '97% ', '99% ', '98% ', '96% ', '96% ', '98% ', '94% ']
10
----------------------------------------------------------------------------------------
['22', '36', '23', '14', '27', '14', '40', '28', '23', '25']
10
----------------------------------------------------------------------------------------
['Dr. Vijay Singhal', 'Dr. Hemant Sharma', 'Dr. Pallav S Kishanpuria', 'Dr. Tanvi Pal', 'Dr. Sirisha Singh', 'Dr. A J Kanwar',
'Dr. Deepti  Shrivastava', 'Dr. Richa Sharma', 'Dr. Divya Aggarwal', 'Dr. Jalpa Jani']
10
----------------------------------------------------------------------------------------
['Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi']
10
----------------------------------------------------------------------------------------
['Punjabi Bagh,', 'Pusa Road,', 'Pusa Road,', 'Pusa Road,', 'Greater Kailash Part 1,', 'Greater Kailash Part 1,', 'Anand Viha
r,', 'Dwarka,', 'Shalimar Bagh,', 'Greater Kailash Part 2,']
10
----------------------------------------------------------------------------------------
['800', '1500', '1500', '1500', '1500', '1500', '800', '1500', '800', '1000']
10
----------------------------------------------------------------------------------------
['93% ', '95% ', '95% ', '97% ', '97% ', '96% ', '95% ', '96% ', '98% ', '96% ']
10
----------------------------------------------------------------------------------------
['18', '48', '19', '19', '29', '52', '28', '26', '21', '21']
10
----------------------------------------------------------------------------------------
```

```
['Dr. Sham Sunder', 'Dr. Rajinder Kumar Singal', 'Dr. Vishal Kaura Aggarwal', 'Dr. Amitabh Khanna', 'Dr. Navneet Kaur', 'Dr. Ha
ri Kumar Yadav', 'Dr. Atul Bhasin', 'Dr. Anuj Gupta', 'Dr. K.S. Walia', 'Dr. (Maj.)Sharad Shrivastava']
10
-------------------------------------------------------------------------------
['Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi']
10
-------------------------------------------------------------------------------
['Safdarjung Enclave,', 'Pusa Road,', 'Pusa Road,', 'Dwarka,', 'Greater Kailash Part 1,', 'Saket,', 'Pusa Road,', 'Munirka,',
'Sarita Vihar,', 'Dwarka Sector 7,']
10
-------------------------------------------------------------------------------
['1500', '1800', '1500', '1200', '500', '800', '1800', '1250', '600', '800']
10
-------------------------------------------------------------------------------
['96% ', '85% ', '100% ', '81% ', '93% ', '100% ', '75% ', '97% ', '90% ', '93% ']
10
-------------------------------------------------------------------------------
['47', '41', '26', '37', '25', '16', '32', '24', '49', '30']
10
-------------------------------------------------------------------------------
['Dr. Neeraj Pasricha', 'Dr. Sneh Khera', 'Dr. Sunil Kumar Dwivedi', 'Dr. C B Singh', 'Dr. Geetika Bansal', 'Dr. Pratima Grove
r', 'Dr. Anika Gupta', 'Dr. Mukta Garg', 'Dr. Gurpreet Juneja', 'Dr. Mridu Avinashi']
10
-------------------------------------------------------------------------------
['Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi', 'Delhi']
10
-------------------------------------------------------------------------------
['Paschim Vihar,', 'Lajpat Nagar 4,', 'Dwarka,', 'Laxmi Nagar,', 'Dwarka Sector 2,', 'Mayur Vihar Ph-I,', 'Rohini Sector 11,',
'Janakpuri,', 'Patel Nagar,', 'Rajouri Garden,']
10
-------------------------------------------------------------------------------
['1000', '330', '400', '500', '700', '500', '500', '500', '500', '700']
10
-------------------------------------------------------------------------------
['93% ', '93% ', '93% ', '90% ', '83% ', '98% ', '100% ', '100% ', '96% ', '100% ']
10
-------------------------------------------------------------------------------
['41', '38', '18', '11', '25', '38', '13', '25', '27', '36']
10
-------------------------------------------------------------------------------
```

CLEANED DATA

Dr. Divya Aggarwal 800	98%	Dermatologist		21	Delhi	Shalimar Bagh,	SKINOS
Dr. Gaurav Bhardwaj Clinic	Dermatologist 1500	100%	11	Delhi	Pusa Road,	Twacha Skin
Dr. Arihant Surana peciality Hospital	Dermatologist 1000	96%	23	Delhi	Defence Colony,	Max Super S
Dr. Rajinder Kumar Singal er Speciality Hospital	General Physician 1800	85%	41	Delhi	Pusa Road,	BLK-Max Sup
Dr. Vishal Kaura Aggarwal 1500	100%	General Physician		26	Delhi	Pusa Road,	+ 1 more
Dr. Amitabh Khanna er Speciality Hospital	General Physician 1200	81%	37	Delhi	Dwarka,	BLK-Max Sup
Dr. Navneet Kaur eciality Center	General Physician 500	93%	25	Delhi	Greater Kailash Part 1,	Diabetes Sp
Dr. Hari Kumar Yadav 800	100%	General Physician		16	Delhi	Saket,	+ 1 more

DATA CLEANING:

- Check for Duplicates
- Drop the duplicate column and unnecessary columns.
- Checking and removing special characters.
- Identifying and imputing missing values.
- Data Type conversion.

Summary statistics of numeric columns

```
summary_stats = combined_df.describe()
print(summary_stats)
```

```
               Doctor_name  Specialist Experience City_name        Area  \
count                   59          59         59        59          59
unique                  49           5         27         1          29
top     Dr. Pratima Grover  Homoeopath         38     Delhi  Pusa Road,
freq                     2          20          6        59           6

                               Hospital_name Consultation_fee Recommendation
count                                     59               59             59
unique                                    24               14             16
top     BLK-Max Super Speciality Hospital                 1000            93%
freq                                      12               14             13
```
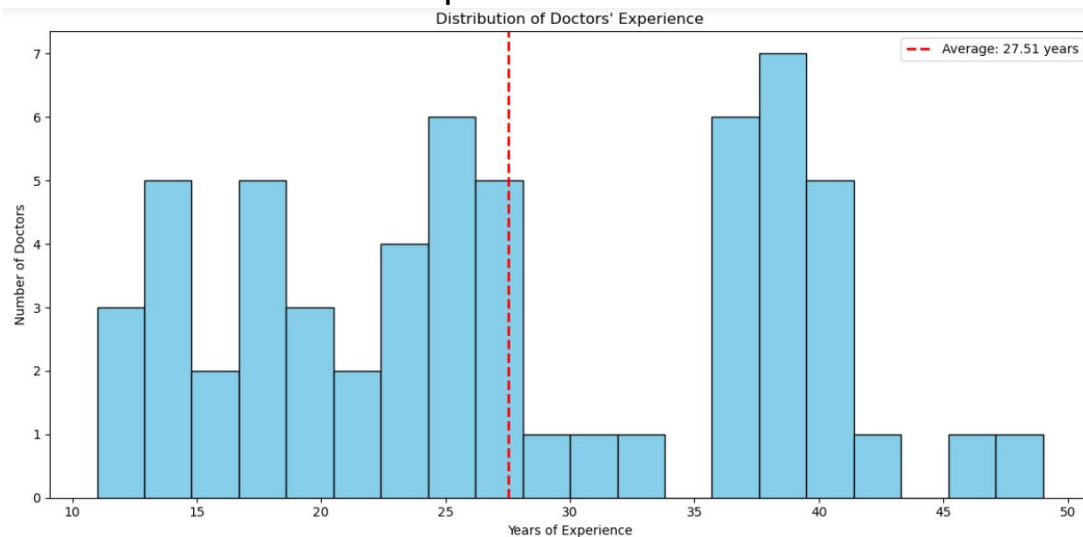
```
3]: # Summary Statistics
    summary_stats = combined_df.describe()
    print("\nSummary Statistics:")
    print(summary_stats)
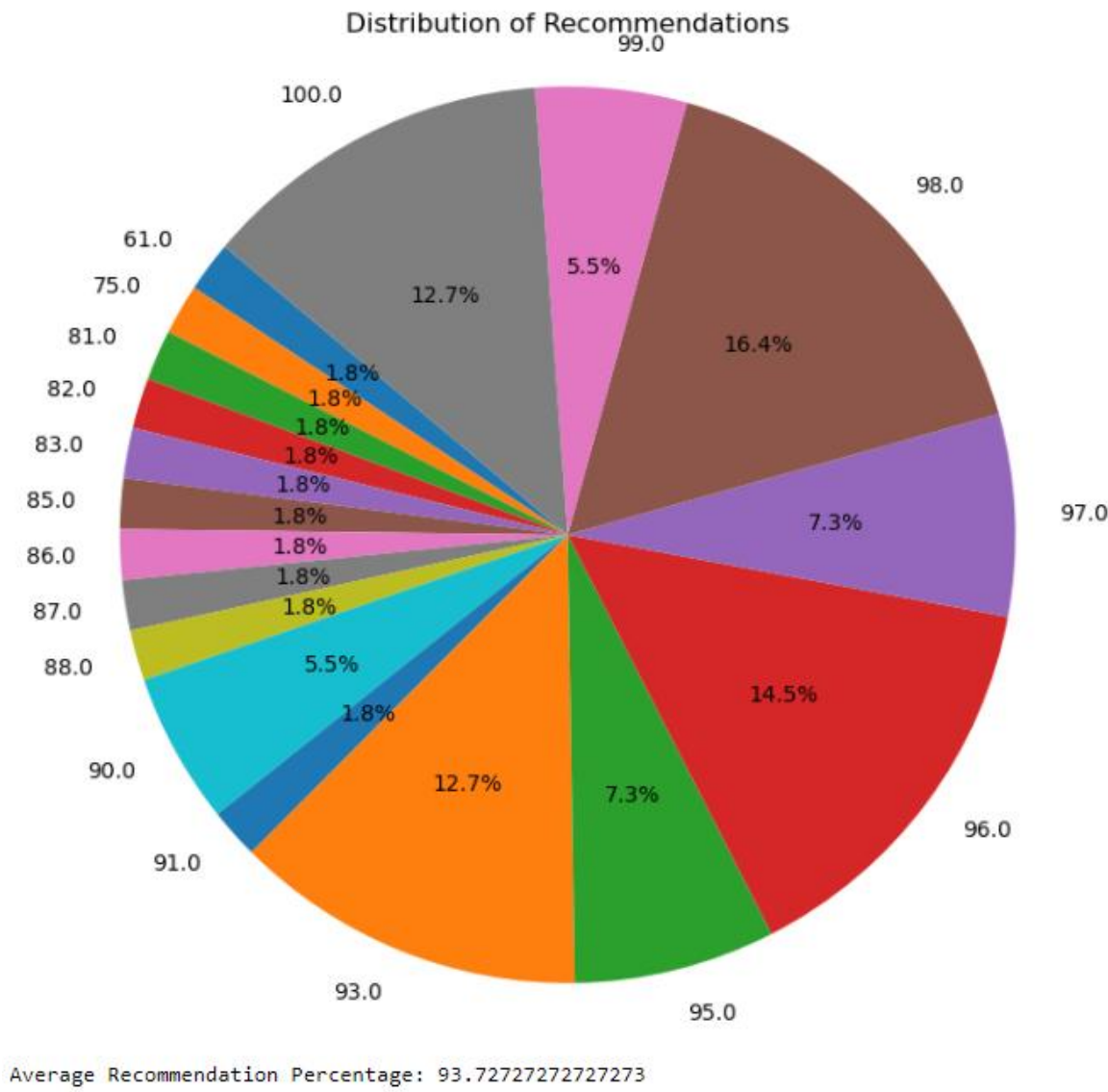```

```
Summary Statistics:
        Experience Consultation_fee
count   59.000000   59.000000
mean   27.508475   934.067797
std   10.284526   408.188254
min   11.000000   330.000000
25%   18.500000   550.000000
50%   26.000000   1000.000000
75%   37.500000   1150.000000
max   49.000000   2000.000000
```
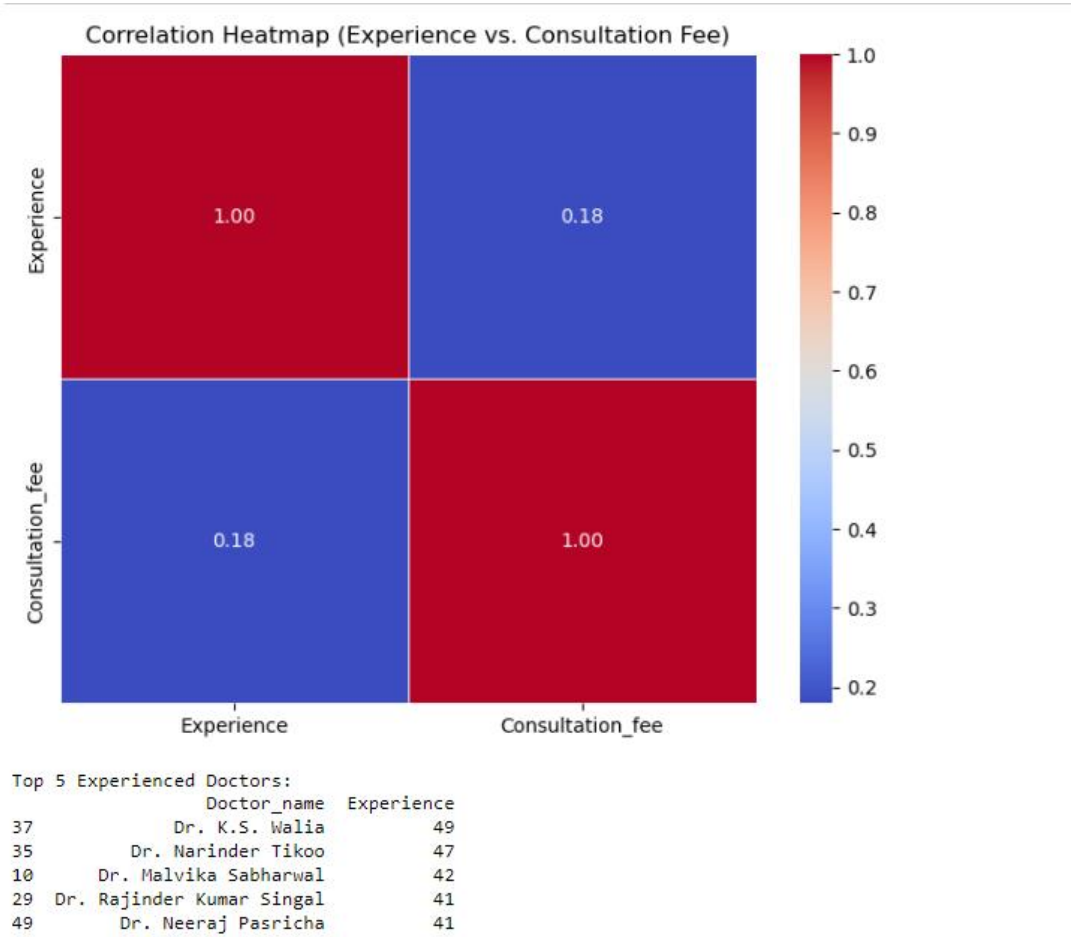
# Distribution of doctors experience



```
Average Experience: 27.51 years
Minimum Experience: 11 years
Maximum Experience: 49 years
```

Distribution of Recommendations

Distribution of Recommendations

Average Recommendation Percentage: 93.72727272727273

Correlation Heatmap (Experience vs Consultation Fee)

Correlation Heatmap (Experience vs. Consultation Fee)
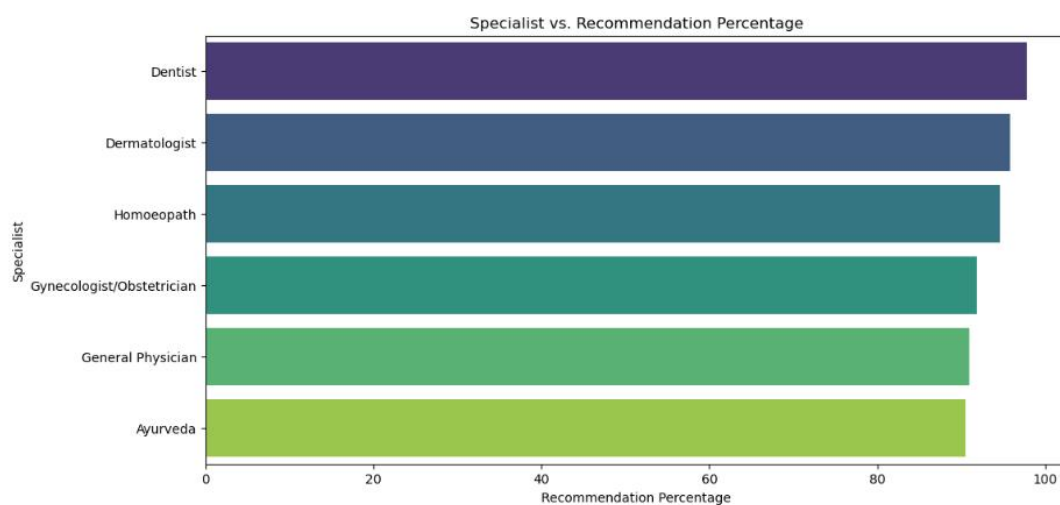
```
Top 5 Experienced Doctors:
              Doctor_name  Experience
37          Dr. K.S. Walia          49
35        Dr. Narinder Tikoo        47
10      Dr. Malvika Sabharwal       42
29  Dr. Rajinder Kumar Singal       41
49        Dr. Neeraj Pasricha       41
```

Top 5 Doctors with the Highest Years of Experience:

	Doctor_name	Experience
37	Dr. K.S. Walia	49
35	Dr. Narinder Tikoo	47
10	Dr. Malvika Sabharwal	42
16	Dr. Yuvakshi Juneja	41
29	Dr. Rajinder Kumar Singal	41

Top 5 Doctors with the Highest Consultation Fees:

	Doctor_name	Consultation_fee
17	Dr. Shakuntla Kumar	2000
29	Dr. Rajinder Kumar Singal	1800
34	Dr. Atul Bhasin	1800
15	Dr. Madhu Ahuja	1500
20	Dr. Pallav S Kishanpuria	1500

Specialist vs Recommendation Percentage
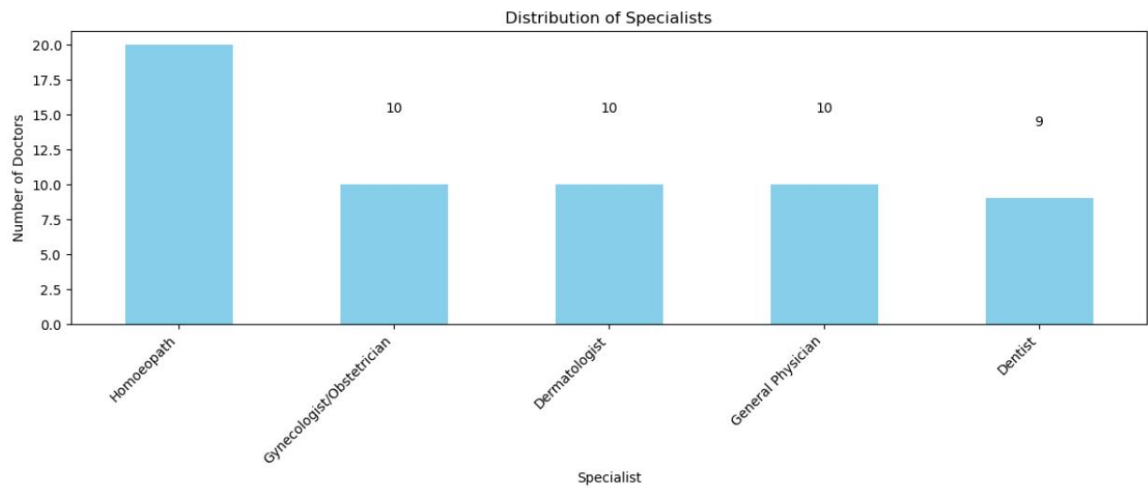
Difference in Recommendations Based on Specialist:

	Specialist	Recommendation
1	Dentist	97.777778
2	Dermatologist	95.800000
5	Homoeopath	94.600000
4	Gynecologist	91.888889
3	General Physician	91.000000
0	Ayurveda	90.571429

Managerial Implications and insights that can be derived from this data:
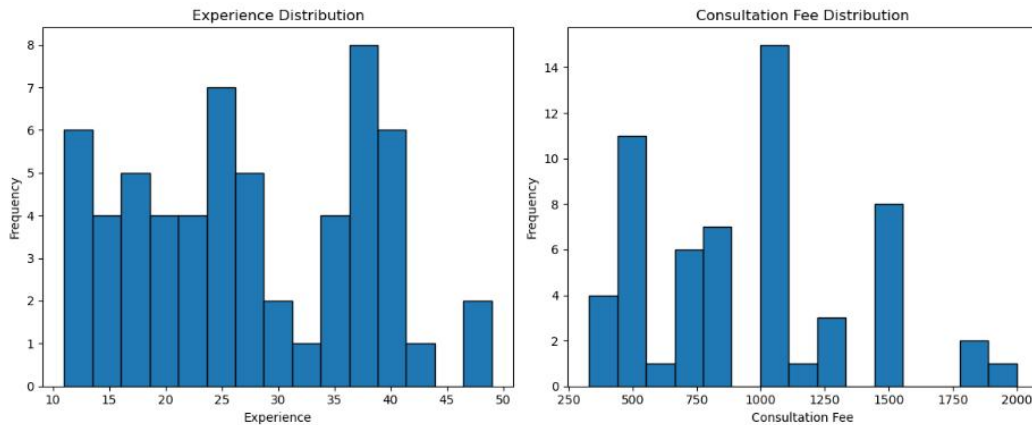
Specialty Analysis:

Healthcare organizations can use this data to understand the distribution of doctors across different specialties. For example, it's clear that there are dentists, gynecologists/obstetricians, dermatologists, general physicians, and homeopaths in the dataset.
Hospitals can use this information to assess the availability of specialists in different areas

Distribution of Specialists

Experience vs. Recommendation:

Healthcare institutions can analyze the relationship between a doctor's years of experience and their patient recommendations. This can help in identifying whether more experienced doctors tend to receive higher recommendations. Hospitals might consider promoting doctors with high recommendations as part of their quality assurance efforts.

```
Experience Summary:
count    59.000000
mean     27.508475
std      10.284526
min      11.000000
25%      18.500000
50%      26.000000
75%      37.500000
max      49.000000
Name: Experience, dtype: float64

Consultation Fee Summary:
count      59.000000
mean      934.067797
std       408.188254
min       330.000000
25%       550.000000
50%      1000.000000
75%      1150.000000
max      2000.000000
Name: Consultation_fee, dtype: float64
```
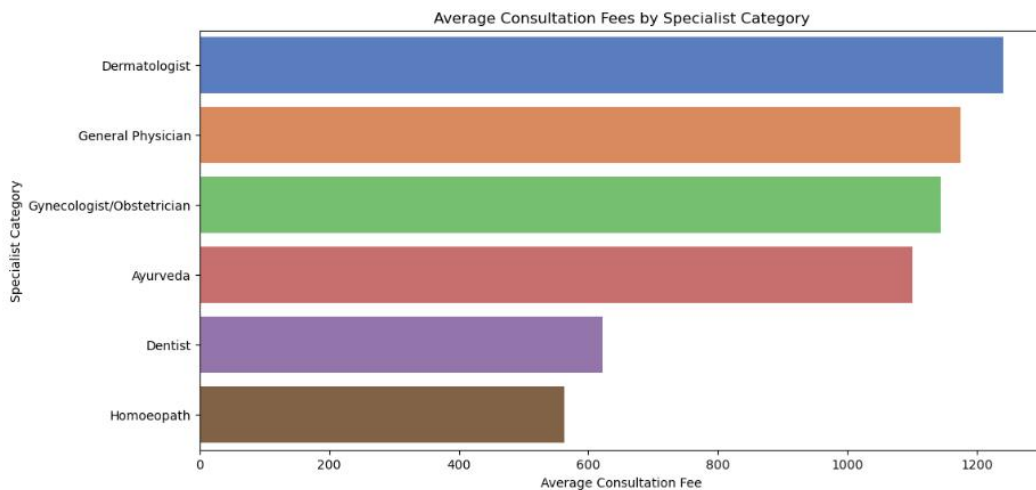
Consultation Fee Analysis:

Analyzing consultation fees across different specialties and areas can help healthcare organizations set competitive pricing for their services.
Hospitals can also identify whether there is a correlation between consultation fees and the percentage of patient recommendations.

Average Consultation Fees by Specialist Category

Location-Based Insights:

Healthcare facilities can analyze the distribution of doctors in different areas and cities to optimize the allocation of medical staff.
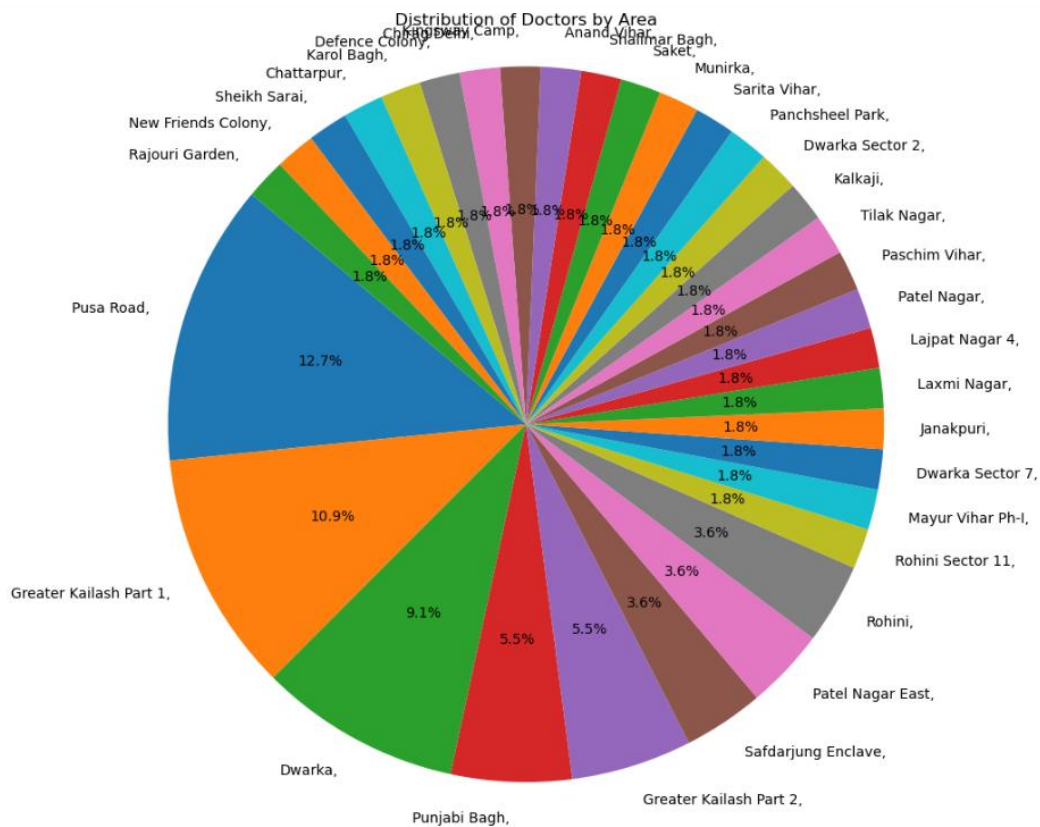It can help patients find doctors in their preferred locations.
Hospital Selection:

Patients can use this data to compare hospitals based on consultation fees and doctor recommendations. For instance, some patients may prefer hospitals with highly recommended doctors.
Hospitals can evaluate their competitiveness by comparing their consultation fees and doctor recommendations with other institutions.
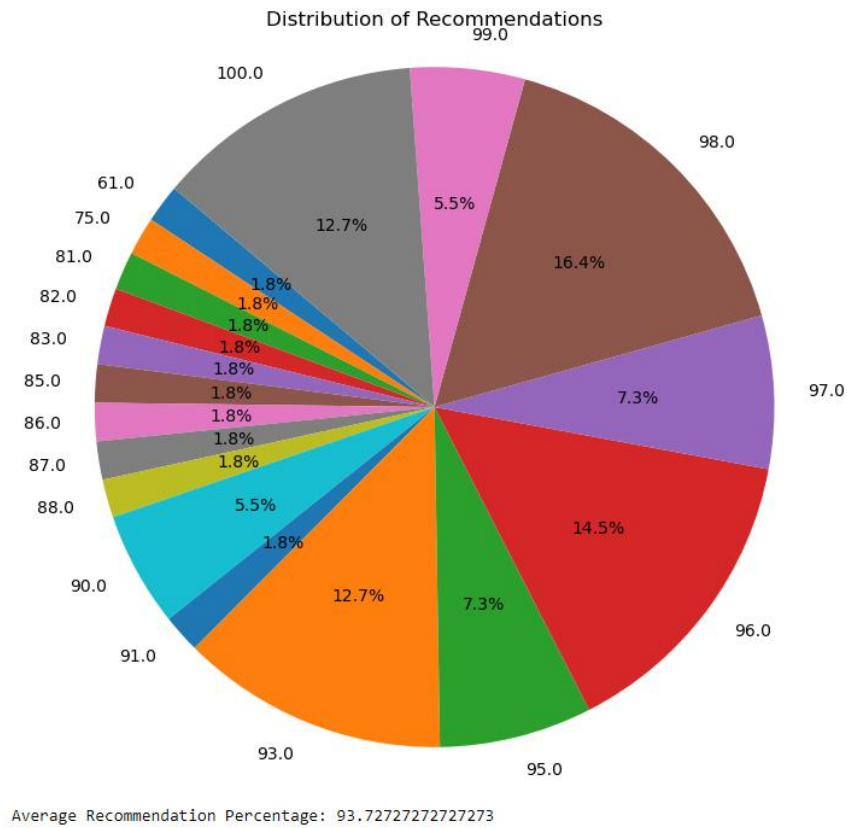
Hospital-wise Consultation Fees:

	Hospital_name	Consultation_fee
14	Dr. Sudha's Ayurveda Kendra	1500.000000
9	Dr Anjana Kalia's Diet clinic	1500.000000
1	Aashlok Hospital	1500.000000
22	SCI International Hospital	1375.000000
2	Apollo Spectra Hospital	1250.000000
15	Hakim Hari Kishan Lal Dawakhana	1100.000000
3	BLK-Max Super Speciality Hospital	1100.000000
28	+ 1 more	1086.666667
29	+ 2 more	1000.000000
26	The Dental Roots	1000.000000
24	Skin Hair & You	1000.000000
21	Pasricha Homeo Care Clinics	1000.000000
17	Jeewan Mala Hospital	1000.000000
30	+ 3 more	1000.000000
13	Dr. Sahni's Dental Clinic	800.000000
7	Dr AJ Kanwar Skin Clinic	800.000000
27	Zenith Skin Hair & Dental Centre	800.000000
6	Diabetes Speciality Center	800.000000
0	ART Fertility Clinics	800.000000
12	Dr Singh Clinic	700.000000
8	Dr Anika's Homoeo Health Care	700.000000
19	Marya's Dental and Implant Centre	650.000000
20	Max Super Speciality Hospital - Saket East Wing	600.000000
10	Dr Geetika's Homeopathic Clinic	500.000000
18	Kapil Dental Clinic and Implant Center	500.000000
25	Southpoint Dentistry	500.000000
16	Holistic Homoeopathic Care	500.000000
5	Dentem the Dental & Orthodontic Clinic	500.000000
4	Dental World	500.000000
23	SD Homoeopathy Multispecality Clinic	400.000000
11	Dr Khera's Homoeopathic Clinic	330.000000

Distribution of Doctors by Area

Recommendation Analysis:

Healthcare organizations can track the average recommendation percentage for different specialists. This information can be used for quality assessments and improving patient satisfaction.

It's important to note that doctors with high recommendation percentages may attract more patients.

Distribution of Recommendations

Average Recommendation Percentage: 93.72727272727273

Hospital Reputation:

Hospitals can use the percentage of patient recommendations as a measure of their overall reputation and quality of service. Higher percentages may attract more patients.

Marketing and Promotion:

Hospitals can use the data to identify high-performing doctors for marketing and promotional activities to attract more patients to their facilities.