INNOMATICS®
RESEARCH LABS

INNOVATION. AUTOMATION. ANALYTICS

PROJECT ON

EDA on companies in Hyderabad (Ambition box)

about us

- Name : Mohd Ashfaq
- Qualification : B.com computers
- Any work Experience: Fresher

- Data Science models use existing data and can simulate several actions. Thus, companies can devise the path to reap the best business outcomes. Data Science helps organizations identify and refine target audiences by combining existing data with other data points for developing useful insights.

- Name : Shaik Sultan
- Qualification : B.com computers
- Any work Experience: Fresher

- In today's presentation i'll like to explain the Ambition box site which is popular among the job seeker which give the best results for the job seeker

INNOMATICS
RESEARCH LABS

Contents:

- Problem Statements
- Webscraping
- Tools used
- Steps to collect Data
- Raw Data
- Data cleaning steps
- Cleaned Data
- Data Visualization
- Challenges faced
- Conclusion

INNOMATICS
RESEARCH LABS

Problem Statement

- What is Ambition box?
- Companies In Hyderabad
- Types of companies
- Reviews of a company
- Interviews conducted by a company
- Salaries of companies
- Displaying graph with respect to Rating ,Reviews ,Ralaries Companytype , Jobs.

INNOMATICS
RESEARCH LABS

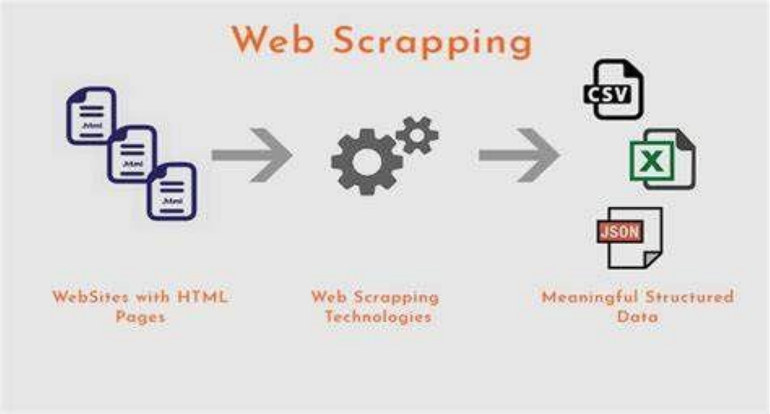
Web scraping

- Web scraping (or data scraping) is a technique used to collect content and data from the internet.
- It is used in a variety of digital businesses that rely on data harvesting.
- The data obtained is mostly unstructured and converted to structured data
- Making an page request to a server
- Extracting and parsing (or breaking down) the website’s code
- Saving the relevant data locally
- Beautiful Soup is a Python library for pulling data out of page and XML files.

Tools used

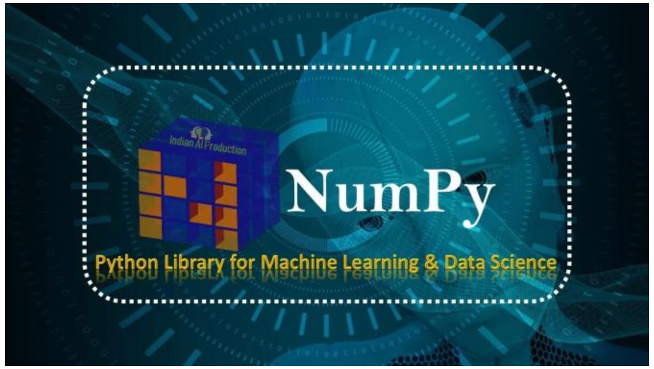
Webscraping Tools(Data Collections)

- Python
- Edge chrome
- BeautifulSoup
- Requests



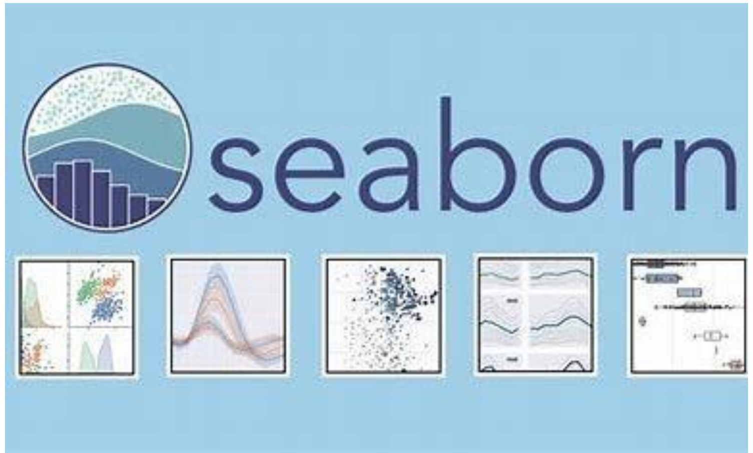
INNOMATICS RESEARCH LABS

- **2. Data cleaning and manupilation**
- Numpy
- Pandas
- Lambda



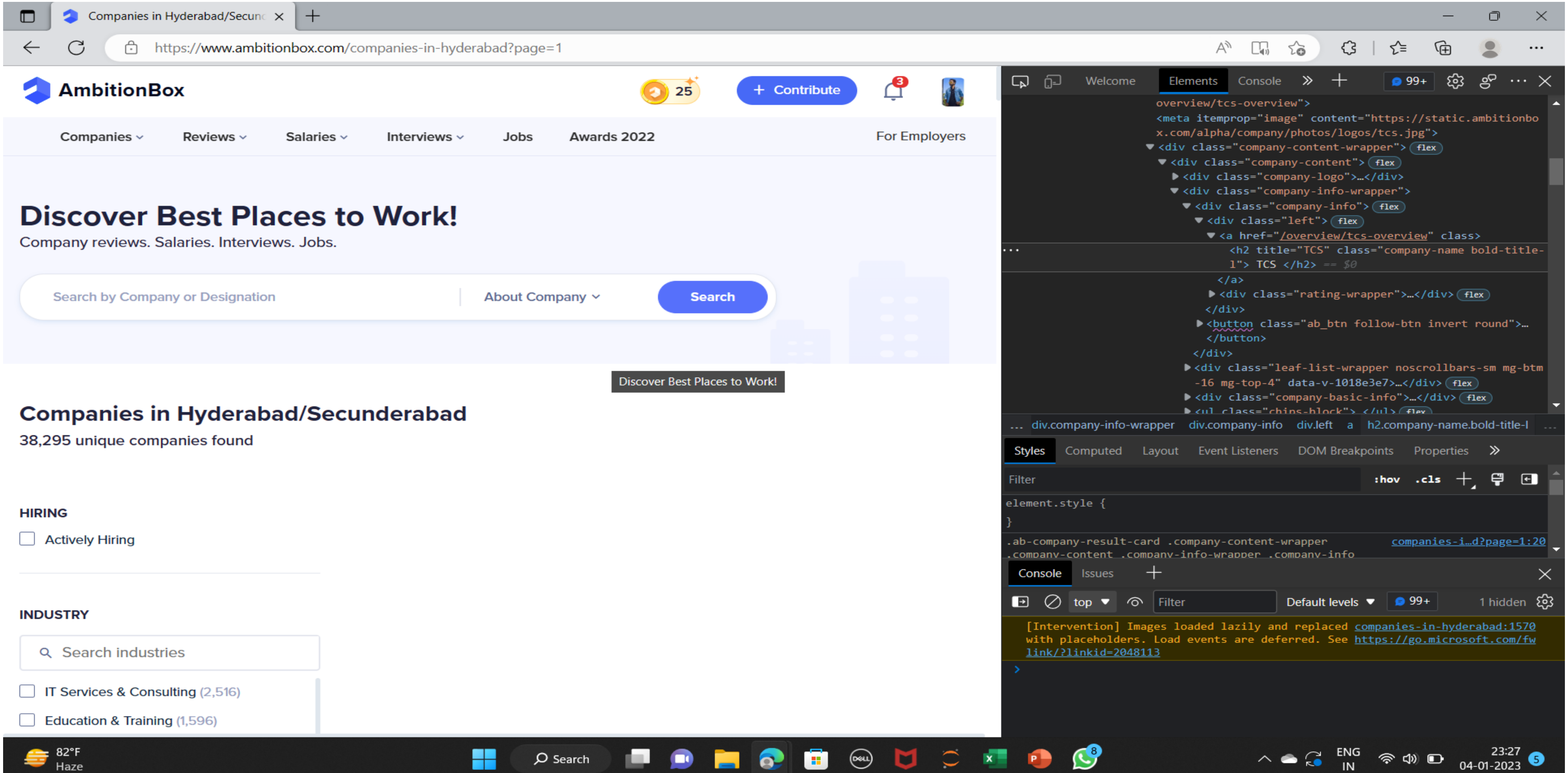
3. Data Visualizations

- Matplotlib
- Seaborn
- Plotly



INNOMATICS
RESEARCH LABS

Ambiton box Website for collecting data



Companies in Hyderabad/Secunc
https://www.ambitionbox.com/companies-in-hyderabad?page=1
AmbitionBox
25
+ Contribute
3
Companies
Reviews
Salaries
Interviews
Jobs
Awards 2022
For Employers

Discover Best Places to Work!

Company reviews. Salaries. Interviews. Jobs.

Search by Company or Designation
About Company
Search
Discover Best Places to Work!

Companies in Hyderabad/Secunderabad

38,295 unique companies found

HIRING

□ Actively Hiring

INDUSTRY

□ IT Services & Consulting (2,516)

□ Education & Training (1,596)

Welcome
Elements
Console
99+

```
overview/tcs-overview">
<meta itemprop="image" content="https://static.ambitionbox.com/alpha/company/photos/logos/tcs.jpg">
▼<div class="company-content-wrapper"> flex
  ▼<div class="company-content"> flex
    ▶<div class="company-logo">…</div>
    ▼<div class="company-info-wrapper">
      ▼<div class="company-info"> flex
        ▼<div class="left"> flex
          ▼<a href="/overview/tcs-overview" class>
              <h2 title="TCS" class="company-name bold-title-l"> TCS </h2> == $0
            </a>
          ▶<div class="rating-wrapper">…</div> flex
          </div>
        ▶<button class="ab_btn follow-btn invert round">…
        </button>
        </div>
      ▶<div class="leaf-list-wrapper noscrollbars-sm mg-btm-16 mg-top-4" data-v-1018e3e7>…</div> flex
      ▶<div class="company-basic-info">…</div> flex
      ▶<ul class="chips-block">…</ul> flex
```

… div.company-info-wrapper div.company-info div.left a h2.company-name.bold-title-l …
Styles
Computed
Layout
Event Listeners
DOM Breakpoints
Properties
Filter
:hov
.cls

```
element.style {
}
.ab-company-result-card .company-content-wrapper                companies-i…d?page=1:20
.company-content .company-info-wrapper .company-info
```

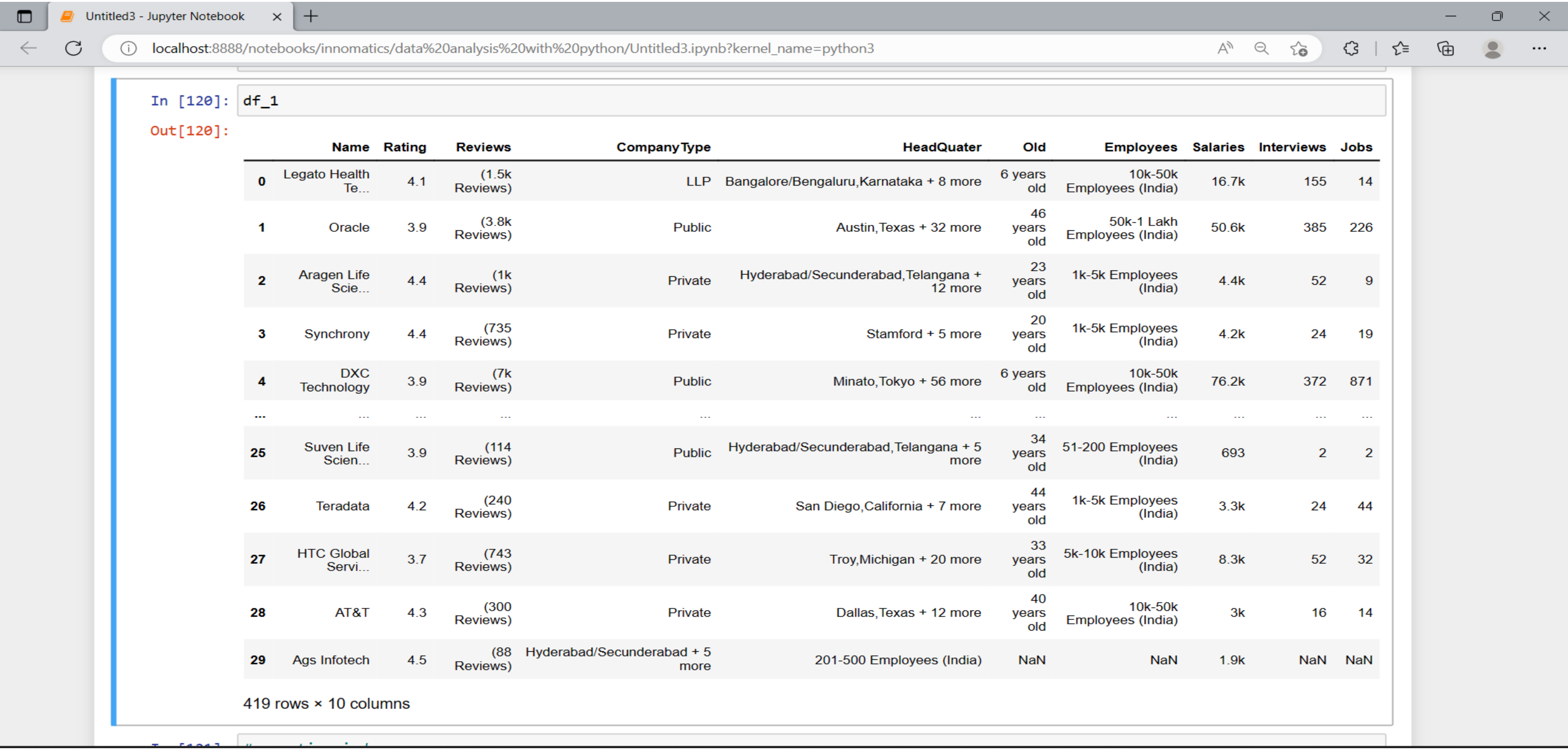
```
[Intervention] Images loaded lazily and replaced    companies-in-hyderabad:1570
with placeholders. Load events are deferred. See https://go.microsoft.com/fwlink/?linkid=2048113
```

Steps to Collect data

- companies in Hyderabad (Ambition box)
- Import Beautiful Soup and all the necessary libraries for scraping the data
- After scraping the raw data we have to check for the column length
- After inserting at a particular value to a column ,we have to make a dictionary and convert it into data frame.
- After converting it into data frame Export into .csv format and read csv file.

INNOMATICS
RESEARCH LABS

Raw Collection from Ambiton box Site


Untitled3 - Jupyter Notebook
localhost:8888/notebooks/innomatics/data%20analysis%20with%20python/Untitled3.ipynb?kernel_name=python3

```
In [120]: df_1
Out[120]:
```

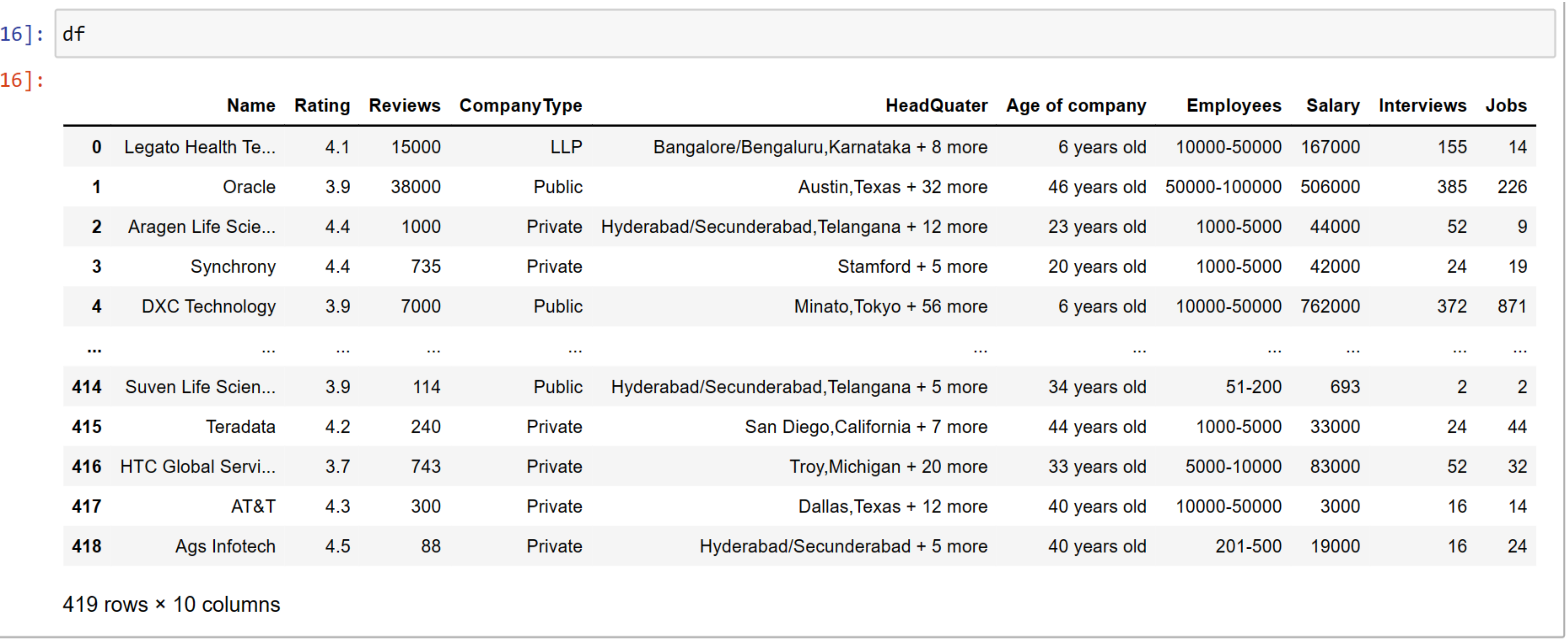
	Name	Rating	Reviews	CompanyType	HeadQuater	Old	Employees	Salaries	Interviews	Jobs
0	Legato Health Te...	4.1	(1.5k Reviews)	LLP	Bangalore/Bengaluru,Karnataka + 8 more	6 years old	10k-50k Employees (India)	16.7k	155	14
1	Oracle	3.9	(3.8k Reviews)	Public	Austin,Texas + 32 more	46 years old	50k-1 Lakh Employees (India)	50.6k	385	226
2	Aragen Life Scie...	4.4	(1k Reviews)	Private	Hyderabad/Secunderabad,Telangana + 12 more	23 years old	1k-5k Employees (India)	4.4k	52	9
3	Synchrony	4.4	(735 Reviews)	Private	Stamford + 5 more	20 years old	1k-5k Employees (India)	4.2k	24	19
4	DXC Technology	3.9	(7k Reviews)	Public	Minato,Tokyo + 56 more	6 years old	10k-50k Employees (India)	76.2k	372	871
...	...	...	...	...	...	...	...	...	...	...
25	Suven Life Scien...	3.9	(114 Reviews)	Public	Hyderabad/Secunderabad,Telangana + 5 more	34 years old	51-200 Employees (India)	693	2	2
26	Teradata	4.2	(240 Reviews)	Private	San Diego,California + 7 more	44 years old	1k-5k Employees (India)	3.3k	24	44
27	HTC Global Servi...	3.7	(743 Reviews)	Private	Troy,Michigan + 20 more	33 years old	5k-10k Employees (India)	8.3k	52	32
28	AT&T	4.3	(300 Reviews)	Private	Dallas,Texas + 12 more	40 years old	10k-50k Employees (India)	3k	16	14
29	Ags Infotech	4.5	(88 Reviews)	Hyderabad/Secunderabad + 5 more	201-500 Employees (India)	NaN	NaN	1.9k	NaN	NaN

419 rows × 10 columns

Data Cleaning Steps

- Check for Duplicates
- Drop the duplicate column and unnecessary columns.
- Removing irrelevant data
- Fix structural errors
- Deal with missing Data
- Filter out data
- Validate our data
- Creating dictionary and saving it into csv format and
- again Reading it for Data visualization

Cleaned data



16]:

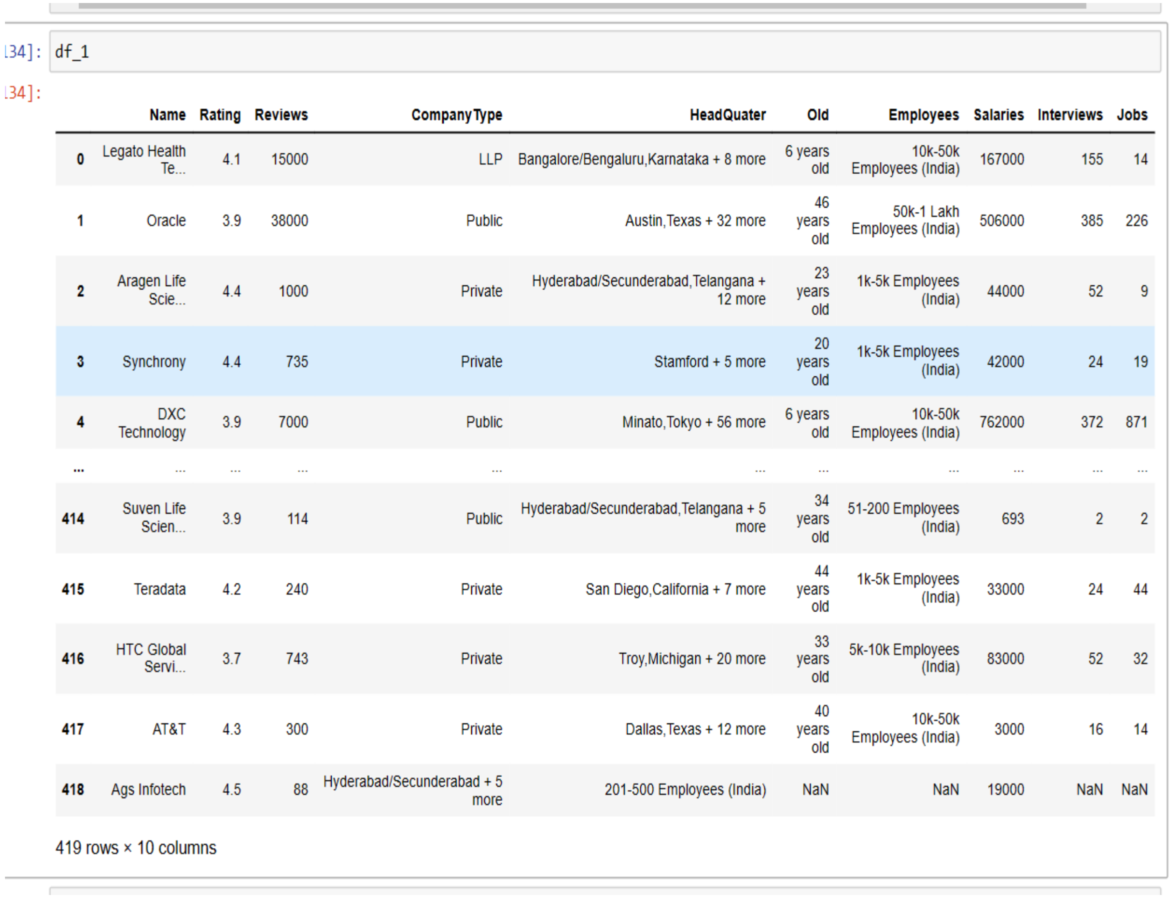
```
df
```

16]:

	Name	Rating	Reviews	CompanyType	HeadQuater	Age of company	Employees	Salary	Interviews	Jobs
0	Legato Health Te...	4.1	15000	LLP	Bangalore/Bengaluru,Karnataka + 8 more	6 years old	10000-50000	167000	155	14
1	Oracle	3.9	38000	Public	Austin,Texas + 32 more	46 years old	50000-100000	506000	385	226
2	Aragen Life Scie...	4.4	1000	Private	Hyderabad/Secunderabad,Telangana + 12 more	23 years old	1000-5000	44000	52	9
3	Synchrony	4.4	735	Private	Stamford + 5 more	20 years old	1000-5000	42000	24	19
4	DXC Technology	3.9	7000	Public	Minato,Tokyo + 56 more	6 years old	10000-50000	762000	372	871
...	...	...	...	...	...	...	...	...	...	...
414	Suven Life Scien...	3.9	114	Public	Hyderabad/Secunderabad,Telangana + 5 more	34 years old	51-200	693	2	2
415	Teradata	4.2	240	Private	San Diego,California + 7 more	44 years old	1000-5000	33000	24	44
416	HTC Global Servi...	3.7	743	Private	Troy,Michigan + 20 more	33 years old	5000-10000	83000	52	32
417	AT&T	4.3	300	Private	Dallas,Texas + 12 more	40 years old	10000-50000	3000	16	14
418	Ags Infotech	4.5	88	Private	Hyderabad/Secunderabad + 5 more	40 years old	201-500	19000	16	24

419 rows × 10 columns

Raw Collection from Ambiton box Site

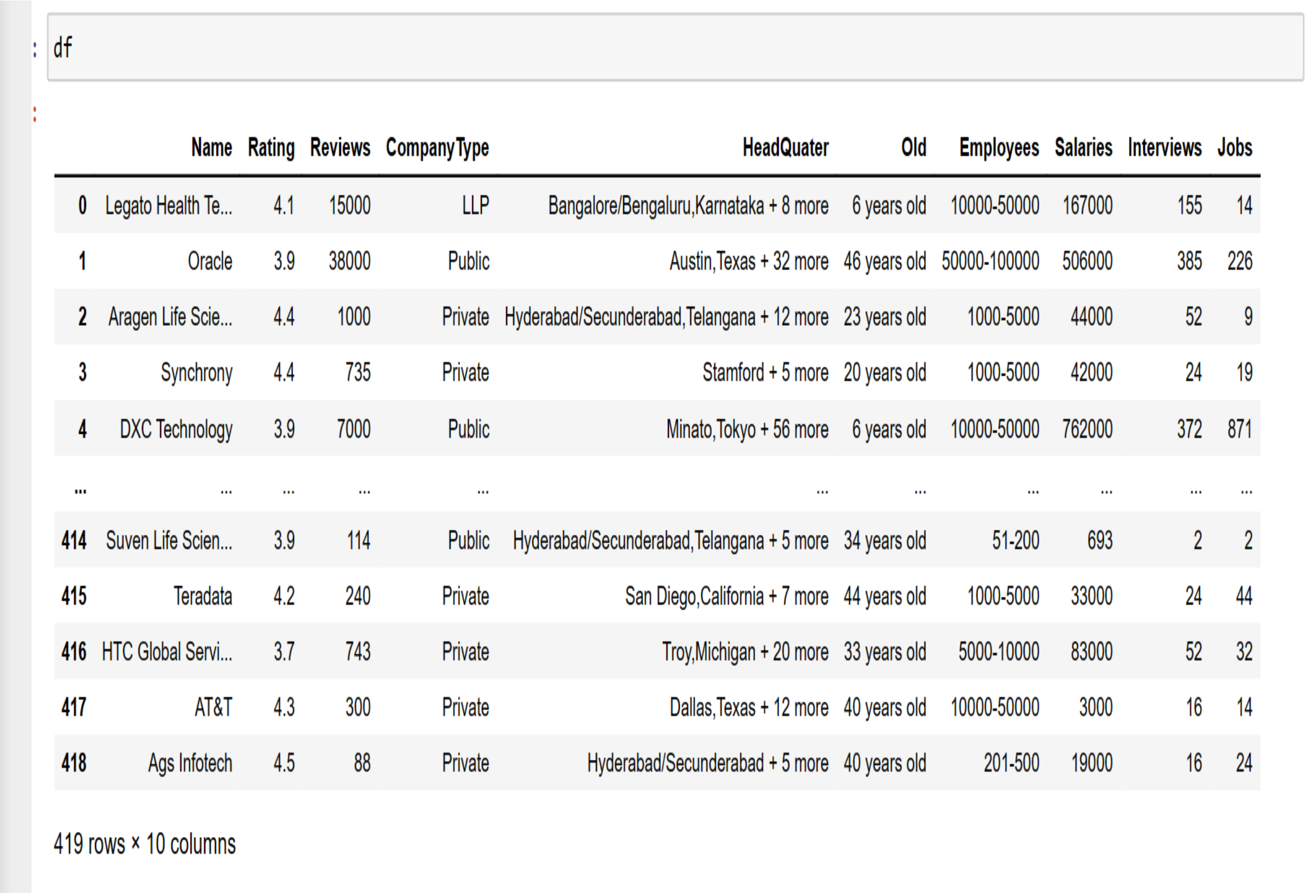


```
134]: df_1
134]:
```

	Name	Rating	Reviews	CompanyType	HeadQuater	Old	Employees	Salaries	Interviews	Jobs
0	Legato Health Te...	4.1	15000	LLP	Bangalore/Bengaluru,Karnataka + 8 more	6 years old	10k-50k Employees (India)	167000	155	14
1	Oracle	3.9	38000	Public	Austin,Texas + 32 more	46 years old	50k-1 Lakh Employees (India)	506000	385	226
2	Aragen Life Scie...	4.4	1000	Private	Hyderabad/Secunderabad,Telangana + 12 more	23 years old	1k-5k Employees (India)	44000	52	9
3	Synchrony	4.4	735	Private	Stamford + 5 more	20 years old	1k-5k Employees (India)	42000	24	19
4	DXC Technology	3.9	7000	Public	Minato,Tokyo + 56 more	6 years old	10k-50k Employees (India)	762000	372	871
...	...	...	...	...	...	...	...	...	...	...
414	Suven Life Scien...	3.9	114	Public	Hyderabad/Secunderabad,Telangana + 5 more	34 years old	51-200 Employees (India)	693	2	2
415	Teradata	4.2	240	Private	San Diego,California + 7 more	44 years old	1k-5k Employees (India)	33000	24	44
416	HTC Global Servi...	3.7	743	Private	Troy,Michigan + 20 more	33 years old	5k-10k Employees (India)	83000	52	32
417	AT&T	4.3	300	Private	Dallas,Texas + 12 more	40 years old	10k-50k Employees (India)	3000	16	14
418	Ags Infotech	4.5	88	Hyderabad/Secunderabad + 5 more	201-500 Employees (India)	NaN	NaN	19000	NaN	NaN

419 rows × 10 columns

Cleaned data



```
: df
```

	Name	Rating	Reviews	CompanyType	HeadQuater	Old	Employees	Salaries	Interviews	Jobs
0	Legato Health Te...	4.1	15000	LLP	Bangalore/Bengaluru,Karnataka + 8 more	6 years old	10000-50000	167000	155	14
1	Oracle	3.9	38000	Public	Austin,Texas + 32 more	46 years old	50000-100000	506000	385	226
2	Aragen Life Scie...	4.4	1000	Private	Hyderabad/Secunderabad,Telangana + 12 more	23 years old	1000-5000	44000	52	9
3	Synchrony	4.4	735	Private	Stamford + 5 more	20 years old	1000-5000	42000	24	19
4	DXC Technology	3.9	7000	Public	Minato,Tokyo + 56 more	6 years old	10000-50000	762000	372	871
...	...	...	...	...	...	...	...	...	...	...
414	Suven Life Scien...	3.9	114	Public	Hyderabad/Secunderabad,Telangana + 5 more	34 years old	51-200	693	2	2
415	Teradata	4.2	240	Private	San Diego,California + 7 more	44 years old	1000-5000	33000	24	44
416	HTC Global Servi...	3.7	743	Private	Troy,Michigan + 20 more	33 years old	5000-10000	83000	52	32
417	AT&T	4.3	300	Private	Dallas,Texas + 12 more	40 years old	10000-50000	3000	16	14
418	Ags Infotech	4.5	88	Private	Hyderabad/Secunderabad + 5 more	40 years old	201-500	19000	16	24

419 rows × 10 columns

Data Visualization

- **Univariate Analysis** :
- Univariate analysis explores each variable in a data set, separately.
- 1) Categorical
- 2) Numerical
- **Bivariate Analysis :**
- Bivariate analysis is a kind of statistical analysis in which two variables are observed against each other.
- **Categorical & Categorical**
- Cross tab, Count plot, Stacked/Group Bar chart.
- **Categorical & Numerical**
- Bar chart, Group By, Pivot.
- **Numerical & Numerical**
- Scatter plot, Heat map etc

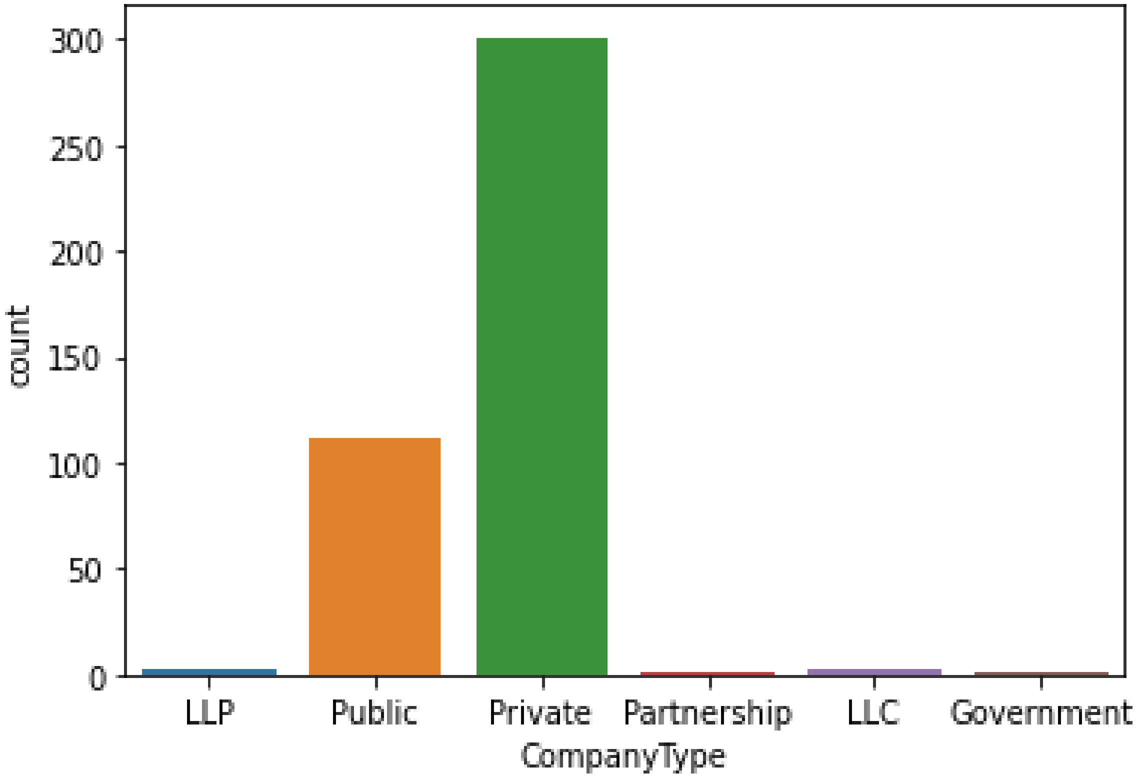
INNOMATICS
RESEARCH LABS

Multi-variate Analysis :

Multi-variate analysis is a kind of statistical analysis in which more than two variables are observed against each other.

1. Two Categorical and One Numerical
2. Two Numerical and One Categorical
3. Three or more Numerical

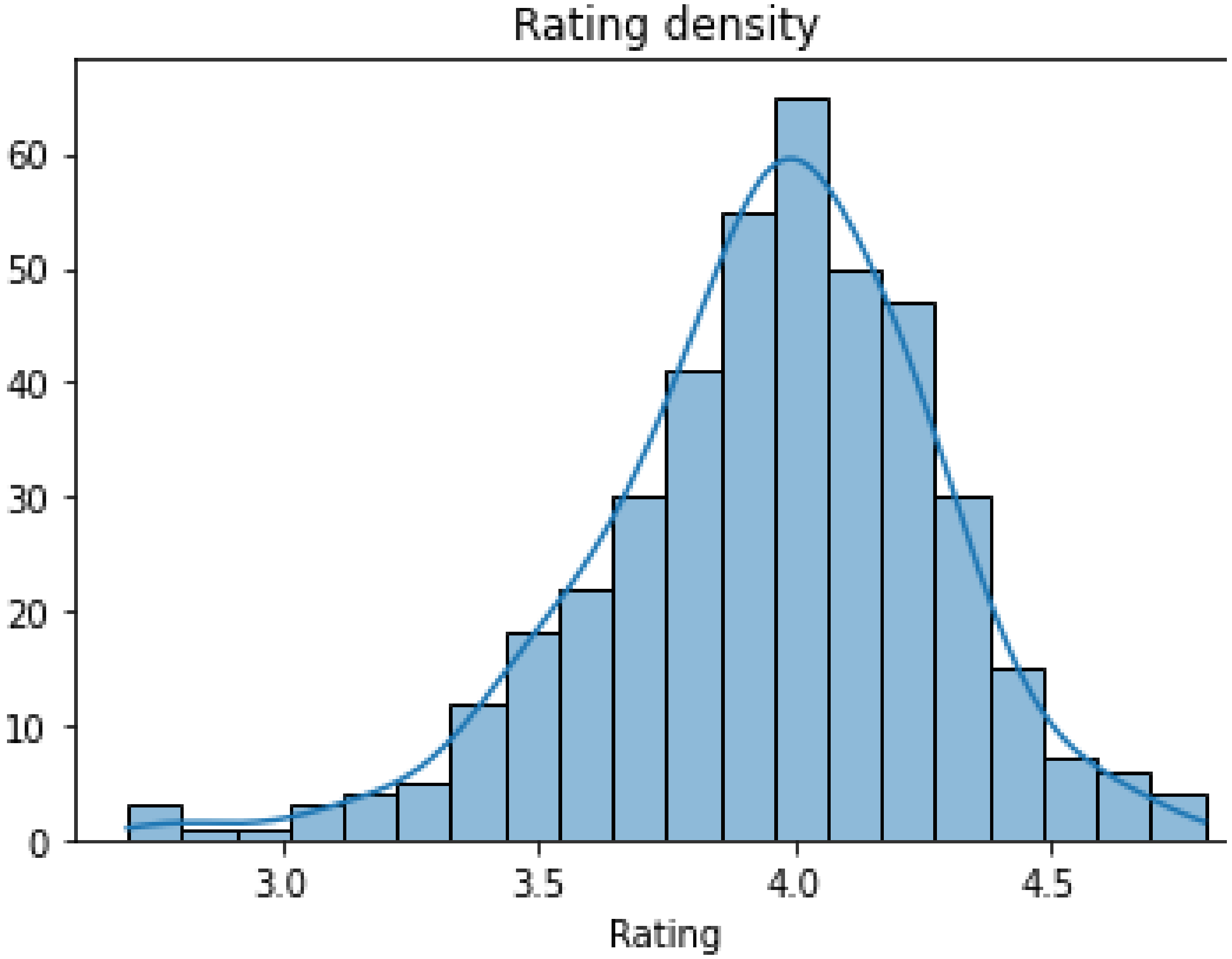
Types of companies category Distribution(Univariate-cat)



- From the graph we can say that we have maximum number of companies in private company in our data

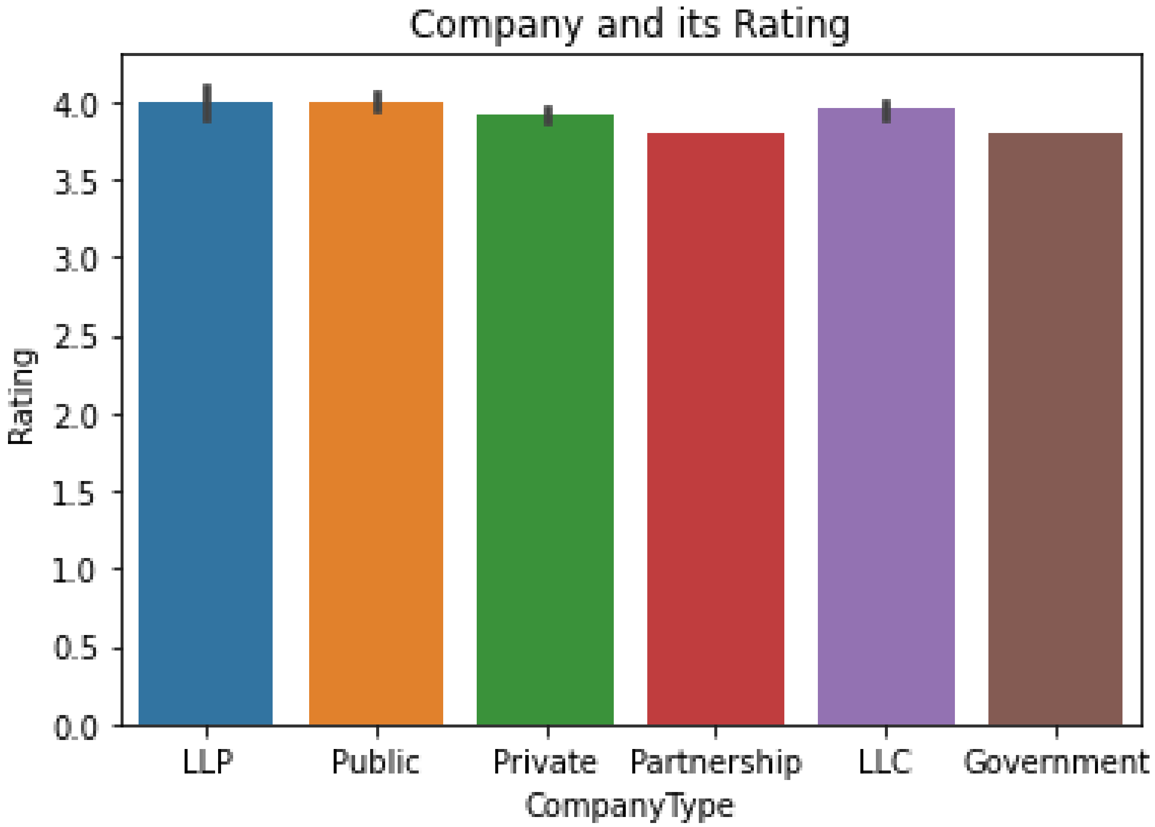
Density of rating(univariate-num)

- From the plot we can see that the density of rating in different company of ratio is between 3.5 to 4.5



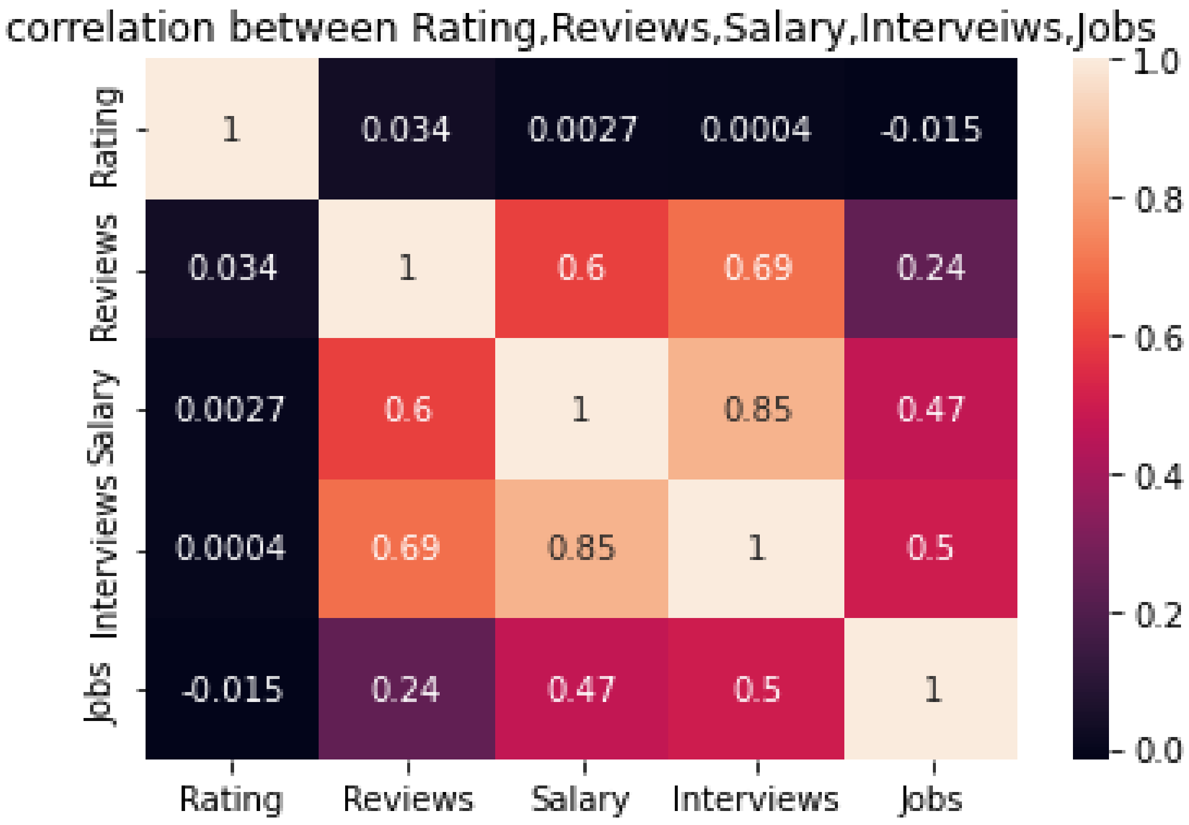
INNOMATICS
RESEARCH LABS

Rating and company type (bivariate- cat and Num)



- From graph we can analyse that almost of the companies have higest nearby equally rating

Correlation (Multivariate)

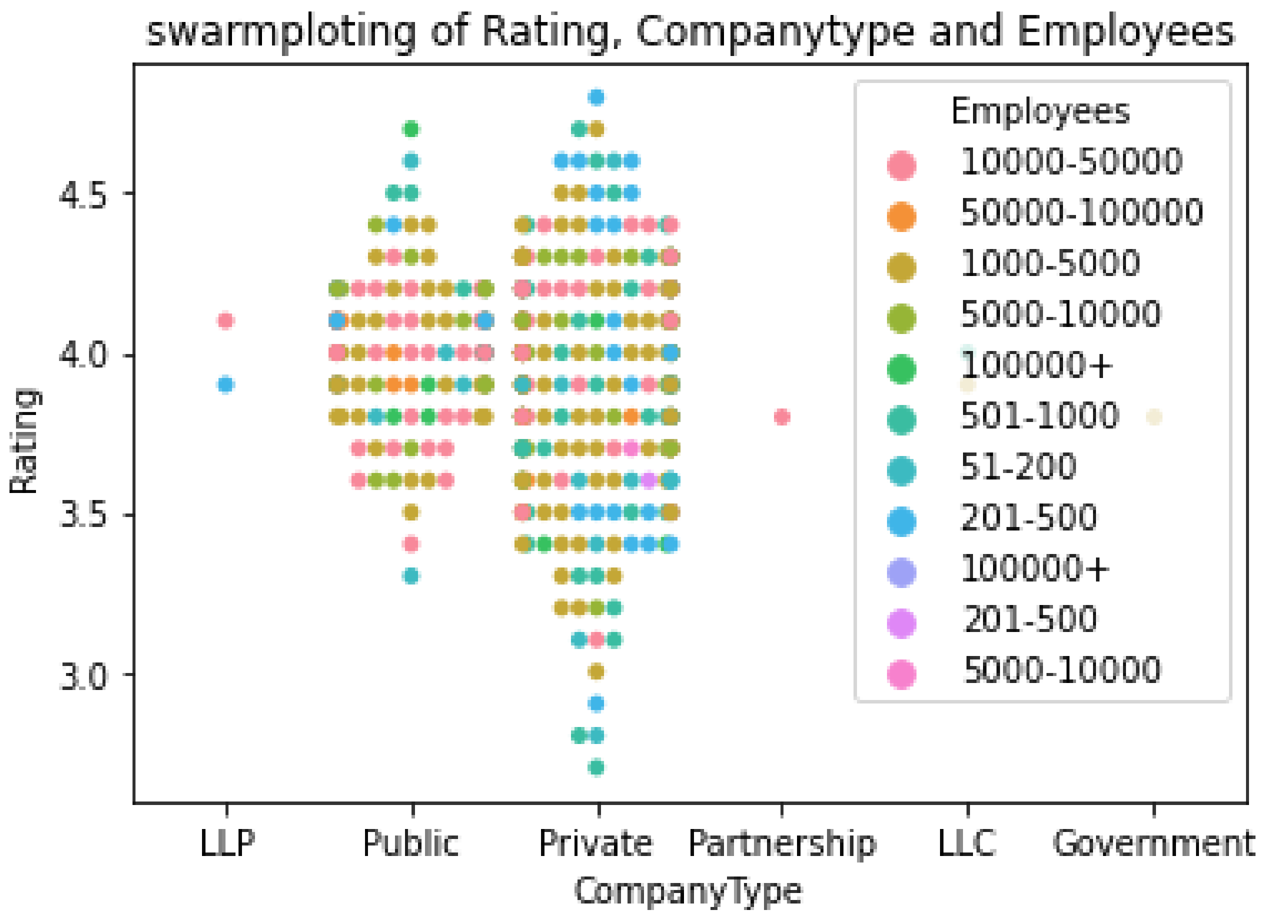


	Rating	Reviews	Salary	Interviews	Jobs
Rating	1	0.034	0.0027	0.0004	-0.015
Reviews	0.034	1	0.6	0.69	0.24
Salary	0.0027	0.6	1	0.85	0.47
Interviews	0.0004	0.69	0.85	1	0.5
Jobs	-0.015	0.24	0.47	0.5	1

- From the plot we can conclude that it has positive relationship and negative relationship
- And we can say that as reviews increases interviews also increases

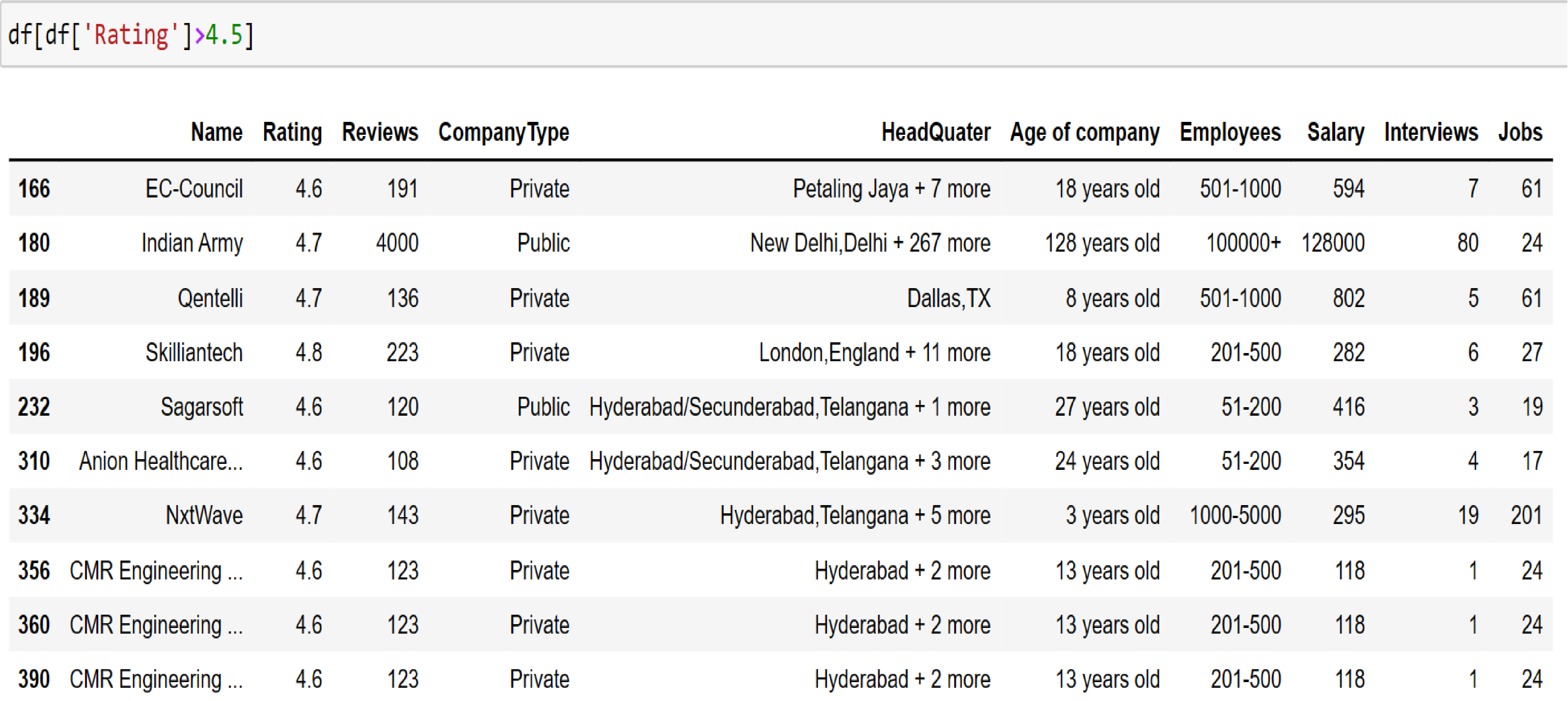
INNOMATICS
RESEARCH LABS

Max number of company type with salaries and and rating (Multivariate-cat vs num vs object)



- From the graph we can conclude that most of the rating lie between 3.5 to 4.5 and private companies more rating compare to other and the no of employees in 201-500 group

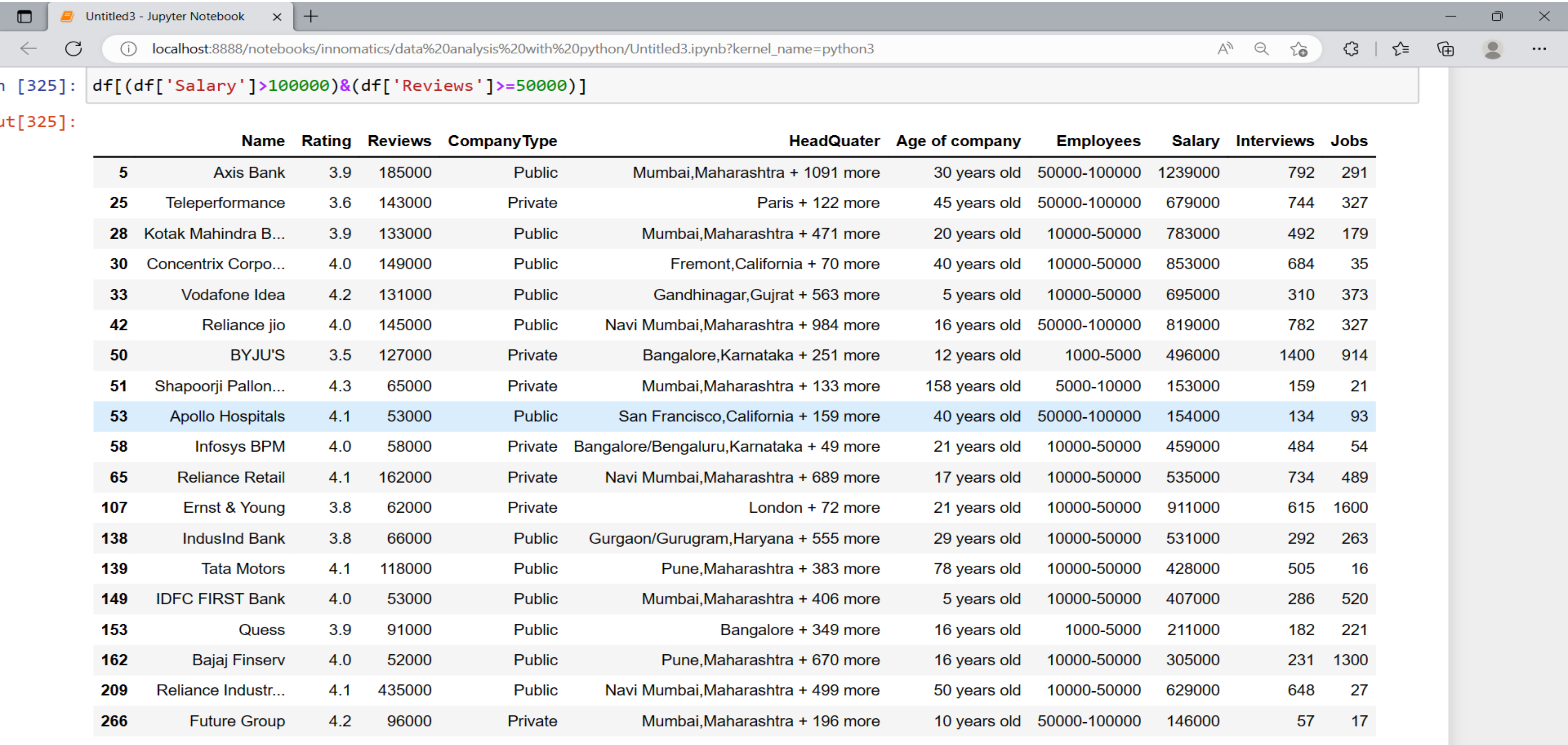
Scenario 1
suppose a person wants know a company with more than 4.5 rating



```
df[df['Rating']>4.5]
```

	Name	Rating	Reviews	CompanyType	HeadQuater	Age of company	Employees	Salary	Interviews	Jobs
166	EC-Council	4.6	191	Private	Petaling Jaya + 7 more	18 years old	501-1000	594	7	61
180	Indian Army	4.7	4000	Public	New Delhi,Delhi + 267 more	128 years old	100000+	128000	80	24
189	Qentelli	4.7	136	Private	Dallas,TX	8 years old	501-1000	802	5	61
196	Skilliantech	4.8	223	Private	London,England + 11 more	18 years old	201-500	282	6	27
232	Sagarsoft	4.6	120	Public	Hyderabad/Secunderabad,Telangana + 1 more	27 years old	51-200	416	3	19
310	Anion Healthcare...	4.6	108	Private	Hyderabad/Secunderabad,Telangana + 3 more	24 years old	51-200	354	4	17
334	NxtWave	4.7	143	Private	Hyderabad,Telangana + 5 more	3 years old	1000-5000	295	19	201
356	CMR Engineering ...	4.6	123	Private	Hyderabad + 2 more	13 years old	201-500	118	1	24
360	CMR Engineering ...	4.6	123	Private	Hyderabad + 2 more	13 years old	201-500	118	1	24
390	CMR Engineering ...	4.6	123	Private	Hyderabad + 2 more	13 years old	201-500	118	1	24

Scenario 2
if a person wants a company with >100000 salaries and reviews with >=50000


Untitled3 - Jupyter Notebook
localhost:8888/notebooks/innomatics/data%20analysis%20with%20python/Untitled3.ipynb?kernel_name=python3

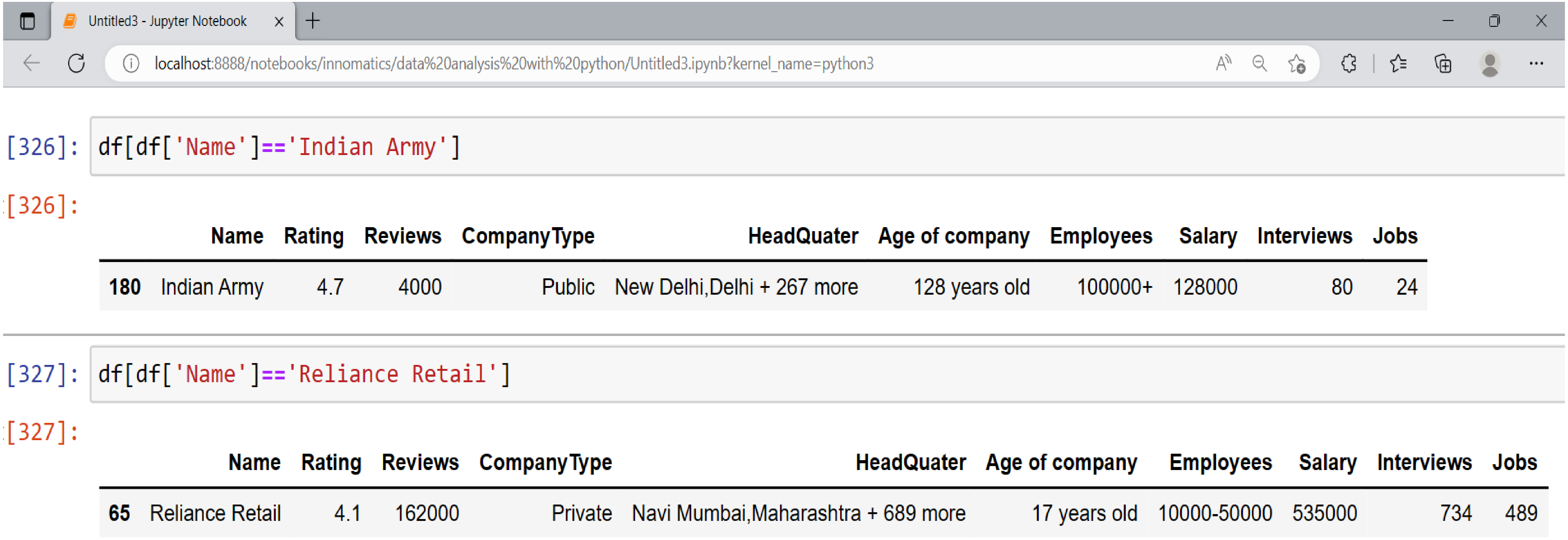
```
n [325]: df[(df['Salary']>100000)&(df['Reviews']>=50000)]
```

ut[325]:

	Name	Rating	Reviews	CompanyType	HeadQuater	Age of company	Employees	Salary	Interviews	Jobs
5	Axis Bank	3.9	185000	Public	Mumbai,Maharashtra + 1091 more	30 years old	50000-100000	1239000	792	291
25	Teleperformance	3.6	143000	Private	Paris + 122 more	45 years old	50000-100000	679000	744	327
28	Kotak Mahindra B...	3.9	133000	Public	Mumbai,Maharashtra + 471 more	20 years old	10000-50000	783000	492	179
30	Concentrix Corpo...	4.0	149000	Public	Fremont,California + 70 more	40 years old	10000-50000	853000	684	35
33	Vodafone Idea	4.2	131000	Public	Gandhinagar,Gujrat + 563 more	5 years old	10000-50000	695000	310	373
42	Reliance jio	4.0	145000	Public	Navi Mumbai,Maharashtra + 984 more	16 years old	50000-100000	819000	782	327
50	BYJU'S	3.5	127000	Private	Bangalore,Karnataka + 251 more	12 years old	1000-5000	496000	1400	914
51	Shapoorji Pallon...	4.3	65000	Private	Mumbai,Maharashtra + 133 more	158 years old	5000-10000	153000	159	21
53	Apollo Hospitals	4.1	53000	Public	San Francisco,California + 159 more	40 years old	50000-100000	154000	134	93
58	Infosys BPM	4.0	58000	Private	Bangalore/Bengaluru,Karnataka + 49 more	21 years old	10000-50000	459000	484	54
65	Reliance Retail	4.1	162000	Private	Navi Mumbai,Maharashtra + 689 more	17 years old	10000-50000	535000	734	489
107	Ernst & Young	3.8	62000	Private	London + 72 more	21 years old	10000-50000	911000	615	1600
138	IndusInd Bank	3.8	66000	Public	Gurgaon/Gurugram,Haryana + 555 more	29 years old	10000-50000	531000	292	263
139	Tata Motors	4.1	118000	Public	Pune,Maharashtra + 383 more	78 years old	10000-50000	428000	505	16
149	IDFC FIRST Bank	4.0	53000	Public	Mumbai,Maharashtra + 406 more	5 years old	10000-50000	407000	286	520
153	Quess	3.9	91000	Public	Bangalore + 349 more	16 years old	1000-5000	211000	182	221
162	Bajaj Finserv	4.0	52000	Public	Pune,Maharashtra + 670 more	16 years old	10000-50000	305000	231	1300
209	Reliance Industr...	4.1	435000	Public	Navi Mumbai,Maharashtra + 499 more	50 years old	10000-50000	629000	648	27
266	Future Group	4.2	96000	Private	Mumbai,Maharashtra + 196 more	10 years old	50000-100000	146000	57	17

Scenario 3

suppose a person wants to compare two companies between these two he wants to know which one has rating and other info


Untitled3 - Jupyter Notebook
localhost:8888/notebooks/innomatics/data%20analysis%20with%20python/Untitled3.ipynb?kernel_name=python3

```
[326]: df[df['Name']=='Indian Army']
```

[326]:

	Name	Rating	Reviews	CompanyType	HeadQuater	Age of company	Employees	Salary	Interviews	Jobs
180	Indian Army	4.7	4000	Public	New Delhi,Delhi + 267 more	128 years old	100000+	128000	80	24

```
[327]: df[df['Name']=='Reliance Retail']
```

[327]:

	Name	Rating	Reviews	CompanyType	HeadQuater	Age of company	Employees	Salary	Interviews	Jobs
65	Reliance Retail	4.1	162000	Private	Navi Mumbai,Maharashtra + 689 more	17 years old	10000-50000	535000	734	489

INNOMATICS
RESEARCH LABS

Conlusion :

- By studing the data we can know about various companies info based on rating, salaries , reviews , company type, interviews jobs and company’s age

INNOMATICS
RESEARCH LABS

INNOMATICS
RESEARCH LABS