

Data Science Job Salaries

Analyzing Salaries of jobs in the Data Science domain

Motivation Behind Choosing Data Science job salaries data

Data Science is a huge umbrella which contains so many different Jobs. The main aim of this project is to make awareness in the people regarding Data Science Job opportunities and Salaries.

Dataset Details

In this project I have taken the dataset of different jobs in data science and their salaries in the year 2020,2021 and 2022. I have taken this dataset from Kaggle (<https://www.kaggle.com/datasets/greeshmagirish/crime-against-women-20012014-india>). I will analyse this data and try find different aspect of it by using Python libraries Pandas,Matplotlib and Seaborn.

As a first step, let's upload our Jupyter notebook to [Jovian.ml](https://jovian.ml).

```
!pip install jovian --upgrade --quiet
```

```
# Execute this to save new versions of the notebook
import jovian
jovian.commit(project="Data-science-job-salaries1")
```

```
[jovian] Updating notebook "shaikhmisba07/data-science-job-salaries1" on
https://jovian.ai
```

```
[jovian] Committed successfully! https://jovian.ai/shaikhmisba07/data-science-job-salaries1
```

```
'https://jovian.ai/shaikhmisba07/data-science-job-salaries1'
```

Data Preparation and Cleaning

Let's import all the libraries that we are going to use in this project.

```
import pandas as pd
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

Let's download the dataset from Koggle by using opendatasets and then read the file.

```
!pip install jovian opendatasets --upgrade --quiet
```

```
dataset_url = 'https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries'
import opendatasets as od
od.download(dataset_url)
```

Skipping, found downloaded files in "./data-science-job-salaries" (use force=True to force download)

```
ds_salary_df = pd.read_csv('Data Science Job Salaries.csv')
ds_salary_df
```

	Unnamed: 0	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	empl
0	0	2020	MI	FT	Data Scientist	70000	EUR	79833	
1	1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	
2	2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	
3	3	2020	MI	FT	Product Data Analyst	20000	USD	20000	
4	4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	
...	
602	602	2022	SE	FT	Data Engineer	154000	USD	154000	
603	603	2022	SE	FT	Data Engineer	126000	USD	126000	
604	604	2022	SE	FT	Data Analyst	129000	USD	129000	
605	605	2022	SE	FT	Data Analyst	150000	USD	150000	
606	606	2022	MI	FT	AI Scientist	200000	USD	200000	

607 rows × 12 columns

Let us find out the number of rows and column of the particular dataset

```
ds_salary_df.shape
```

(607, 12)

As we can see in the above dataset, there are total 12 columns and 607 rows it means dataset has 607 employee's data. Let's delete the first column which is Unnamed column as there is no way to use this part and add column Salary_in_RS.

```
#Delete first column by using drop() function.
columns = [1,2,3,4,5,6,7,8,9,10]
```

```
ds1_salary_df = ds_salary_df.drop(ds_salary_df.columns[0],axis=1)
ds1_salary_df
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residen
0	2020	MI	FT	Data Scientist	70000	EUR	79833	I
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	,
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	(
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	†
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	I
...	
602	2022	SE	FT	Data Engineer	154000	USD	154000	I
603	2022	SE	FT	Data Engineer	126000	USD	126000	I
604	2022	SE	FT	Data Analyst	129000	USD	129000	I
605	2022	SE	FT	Data Analyst	150000	USD	150000	I
606	2022	MI	FT	AI Scientist	200000	USD	200000	

607 rows × 11 columns

```
#Adding Salary_in_RS column.
ds1_salary_df['Salary_in_RS'] = ds_salary_df.salary_in_usd * 82.54
ds1_salary_df
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residen
0	2020	MI	FT	Data Scientist	70000	EUR	79833	I
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	,
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	(
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	†
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	I
...	
602	2022	SE	FT	Data Engineer	154000	USD	154000	I

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residen
603	2022	SE	FT	Data Engineer	126000	USD	126000	I
604	2022	SE	FT	Data Analyst	129000	USD	129000	I
605	2022	SE	FT	Data Analyst	150000	USD	150000	I
606	2022	MI	FT	AI Scientist	200000	USD	200000	

607 rows × 12 columns

Exploratory Analysis and Visualization

Let's find out the total job titles present in the given data sets

```
len(ds1_salary_df.job_title.unique())
```

50

As we can see there are 50 unique jobs available in Data Science domain,Let's define a helper function that turns a column containing lists of values into a data frame with one column for each possible option.

```
def split_multicolumn(col_series):
    result_df = col_series.to_frame()
    options = []
    # Iterate over the column
    for idx, value in col_series[col_series.notnull()].iteritems():
        # Break each value into list of options
        for option in value.split(';'):
            # Add the option as a column to result
            if not option in result_df.columns:
                options.append(option)
                result_df[option] = False
            # Mark the value in the option column as True
            result_df.at[idx, option] = True
    return result_df[options]
```

```
jobs_type_df = split_multicolumn(ds1_salary_df.job_title)
jobs_type_df
```

	Data Scientist	Machine Learning Scientist	Big Data Engineer	Product Data Analyst	Machine Learning Engineer	Data Analyst	Lead Data Scientist	Business Data Analyst	Lead Data Engineer	Lead Data Analyst	...	Finance Data Analyst	F
0	True	False	False	False	False	False	False	False	False	False	...	False	
1	False	True	False	False	False	False	False	False	False	False	...	False	
2	False	False	True	False	False	False	False	False	False	False	...	False	
3	False	False	False	True	False	False	False	False	False	False	...	False	
4	False	False	False	False	True	False	False	False	False	False	...	False	

	Data Scientist	Machine Learning Scientist	Big Data Engineer	Product Data Analyst	Machine Learning Engineer	Data Analyst	Lead Data Scientist	Business Data Analyst	Lead Data Engineer	Lead Data Analyst	...	Finance Data Analyst	F
...
602	False	False	False	False	False	False	False	False	False	False	...	False	
603	False	False	False	False	False	False	False	False	False	False	...	False	
604	False	False	False	False	False	True	False	False	False	False	...	False	
605	False	False	False	False	False	True	False	False	False	False	...	False	
606	False	False	False	False	False	False	False	False	False	False	...	False	

607 rows × 50 columns

There are total 50 different jobs present in the Data Science domain. Let's see which Job title have most opportunities in the year 2020 to 2022.

```
most_job_opportunities =ds1_salary_df.job_title.value_counts(sort=True).head(2)
most_job_opportunities
```

Data Scientist 143
Data Engineer 132
Name: job_title, dtype: int64

Even though there are 50 different jobs present in the field of Data Science,Data Scientist and Data Engineer have most job opportunities.

Now let's analyse Data Scientist and Data Engineer job roles and salaries by using graph.

```
#Data Scientist
Data_Scientist = ds1_salary_df[ds1_salary_df['job_title'].str.contains('Data Scientist')
Data_Scientist
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_resid
0	2020	MI	FT	Data Scientist	70000	EUR	79833	
6	2020	SE	FT	Lead Data Scientist	190000	USD	190000	
7	2020	MI	FT	Data Scientist	11000000	HUF	35735	
10	2020	EN	FT	Data Scientist	45000	EUR	51321	
11	2020	MI	FT	Data Scientist	3000000	INR	40481	
...	
592	2022	SE	FT	Data Scientist	230000	USD	230000	
593	2022	SE	FT	Data Scientist	150000	USD	150000	
596	2022	SE	FT	Data Scientist	210000	USD	210000	

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_resid
598	2022	MI	FT	Data Scientist	160000	USD	160000	
599	2022	MI	FT	Data Scientist	130000	USD	130000	

159 rows × 12 columns

```
#Data Engineer
```

```
Data_Engineer = ds1_salary_df[ds1_salary_df['job_title'].str.contains('Data Engineer')]  
Data_Engineer
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_reside
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	
9	2020	SE	FT	Lead Data Engineer	125000	USD	125000	
16	2020	EN	FT	Data Engineer	4450000	JPY	41689	
17	2020	SE	FT	Big Data Engineer	100000	EUR	114047	
19	2020	MI	FT	Lead Data Engineer	56000	USD	56000	
...	
579	2022	SE	FT	Data Engineer	25000	USD	25000	
582	2022	SE	FT	Data Engineer	220110	USD	220110	
583	2022	SE	FT	Data Engineer	160080	USD	160080	
602	2022	SE	FT	Data Engineer	154000	USD	154000	
603	2022	SE	FT	Data Engineer	126000	USD	126000	

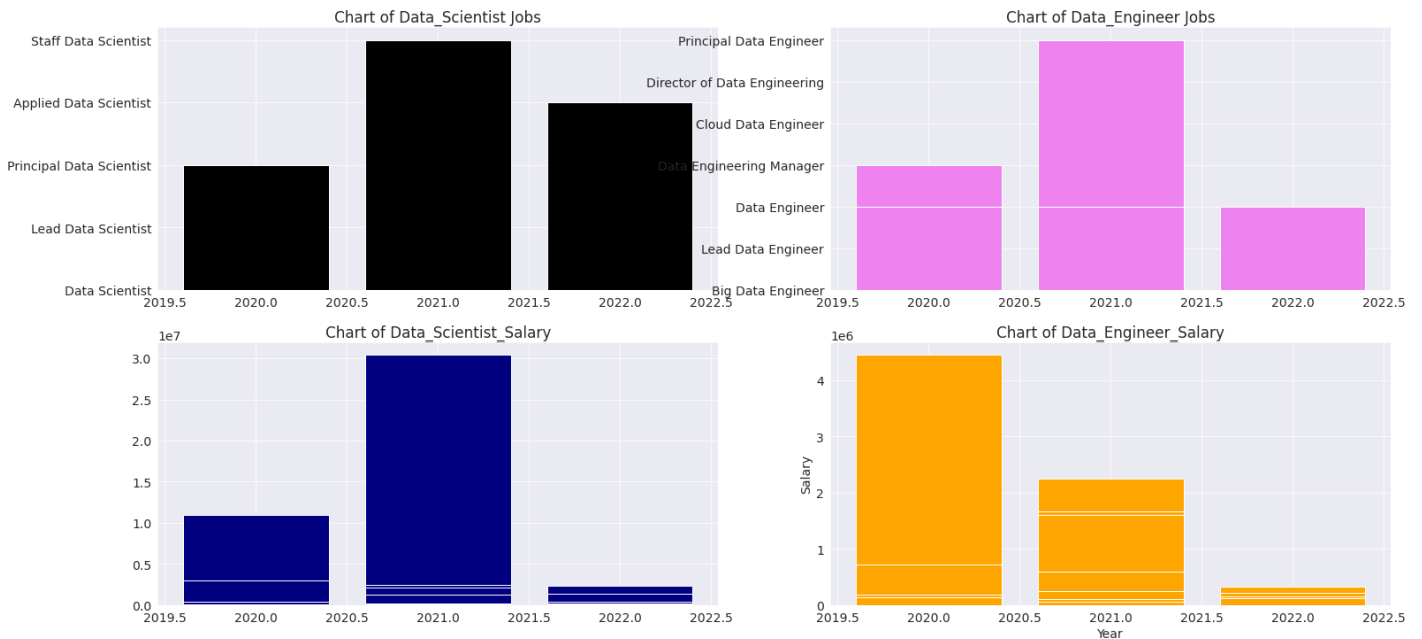
158 rows × 12 columns

```
fig, axes = plt.subplots(2,2,figsize=(25,12))  
axes[0,0].set_title("Chart of Data_Scientist Jobs")  
axes[0,0].bar(Data_Scientist.work_year, Data_Scientist.job_title, color = 'black');  
plt.xlabel('Year')  
plt.ylabel('Domain')  
  
axes[0,1].set_title("Chart of Data_Engineer Jobs")  
axes[0,1].bar(Data_Engineer.work_year,Data_Engineer.job_title,color='violet');  
plt.xlabel('Year')  
plt.ylabel('Domain')  
  
axes[1,0].set_title("Chart of Data_Scientist_Salary")  
axes[1,0].bar(Data_Scientist.work_year, Data_Scientist.salary, color = 'navy');
```

```
plt.xlabel('Year')
plt.ylabel('Salary')

axes[1,1].set_title("Chart of Data_Engineer_Salary")
axes[1,1].bar(Data_Engineer.work_year,Data_Engineer.salary,color='orange');
plt.xlabel('Year')
plt.ylabel('Salary')
```

Text(0, 0.5, 'Salary')



From above graphs,we can conclude below things.

- 1.For Data_Scientist,most of the jobs are available in 2021 followed by 2022 and 2020. However highest salaries offered in the year 2021 followed by 2020 and 2022.
- 2.For Data_Engineer,most of the jobs are available in 2021 followed by 2020 and 2022. However highest salaries offered in the year 2020 followed by 2021 and 2022.

```
Salary_count = ds1_salary_df.groupby('job_title')[['salary']].count()
Salary_count
```

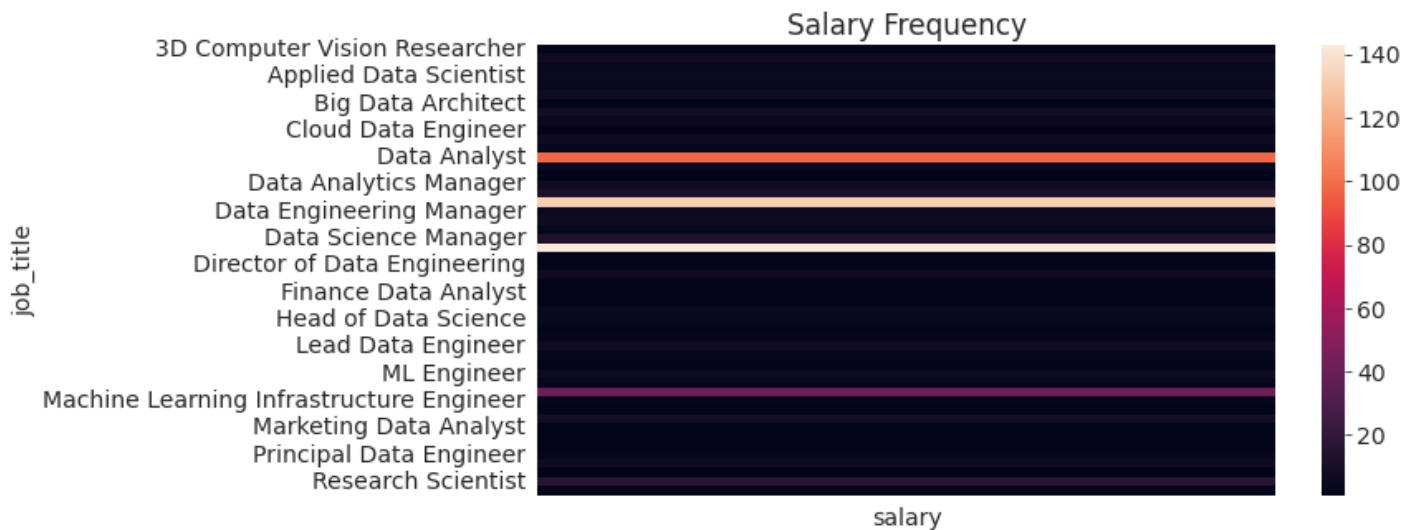
	salary
job_title	
3D Computer Vision Researcher	1
AI Scientist	7
Analytics Engineer	4
Applied Data Scientist	5
Applied Machine Learning Scientist	4
BI Data Analyst	6
Big Data Architect	1
Big Data Engineer	8
Business Data Analyst	5
Cloud Data Engineer	2

	salary
job_title	
Computer Vision Engineer	6
Computer Vision Software Engineer	3
Data Analyst	97
Data Analytics Engineer	4
Data Analytics Lead	1
Data Analytics Manager	7
Data Architect	11
Data Engineer	132
Data Engineering Manager	5
Data Science Consultant	7
Data Science Engineer	3
Data Science Manager	12
Data Scientist	143
Data Specialist	1
Director of Data Engineering	2
Director of Data Science	7
ETL Developer	2
Finance Data Analyst	1
Financial Data Analyst	2
Head of Data	5
Head of Data Science	4
Head of Machine Learning	1
Lead Data Analyst	3
Lead Data Engineer	6
Lead Data Scientist	3
Lead Machine Learning Engineer	1
ML Engineer	6
Machine Learning Developer	3
Machine Learning Engineer	41
Machine Learning Infrastructure Engineer	3
Machine Learning Manager	1
Machine Learning Scientist	8
Marketing Data Analyst	1
NLP Engineer	1
Principal Data Analyst	2
Principal Data Engineer	3
Principal Data Scientist	7
Product Data Analyst	2

salary	
job_title	
Research Scientist	16
Staff Data Scientist	1

Let us use seaborn to take help of heatmap to represent the frequency of salaries offered by each job in Data Science Domain.

```
plt.figure(figsize=(10,5))
plt.title("Salary Frequency")
sns.heatmap(Salary_count);
```



We can conclude from above heatmap Data scientist has maximum salary and maximum job opportunities in the year 2020 to 2022.

Asking and Answering Questions

Q1.Which jobs have most opportunities and which jobs have least opportunities in Data Science domain?

```
jobs_type_df
```

	Data Scientist	Machine Learning Scientist	Big Data Engineer	Product Data Analyst	Machine Learning Engineer	Data Analyst	Lead Data Scientist	Business Data Analyst	Lead Data Engineer	Lead Data Analyst	...	Finance Data Analyst	F
0	True	False	False	False	False	False	False	False	False	False	...	False	
1	False	True	False	False	False	False	False	False	False	False	...	False	
2	False	False	True	False	False	False	False	False	False	False	...	False	
3	False	False	False	True	False	False	False	False	False	False	...	False	
4	False	False	False	False	True	False	False	False	False	False	...	False	
...	
602	False	False	False	False	False	False	False	False	False	False	...	False	
603	False	False	False	False	False	False	False	False	False	False	...	False	

	Data Scientist	Machine Learning Scientist	Big Data Engineer	Product Data Analyst	Machine Learning Engineer	Data Analyst	Lead Data Scientist	Business Data Analyst	Lead Data Engineer	Lead Data Analyst	...	Finance Data Analyst	F
604	False	False	False	False	False	True	False	False	False	False	...	False	
605	False	False	False	False	False	True	False	False	False	False	...	False	
606	False	False	False	False	False	False	False	False	False	False	...	False	

607 rows × 50 columns

Let's find out the job opportunities percentage for each role.

```
Percentage_df = jobs_type_df.mean().sort_values(ascending = False)*100
Percentage_df
```

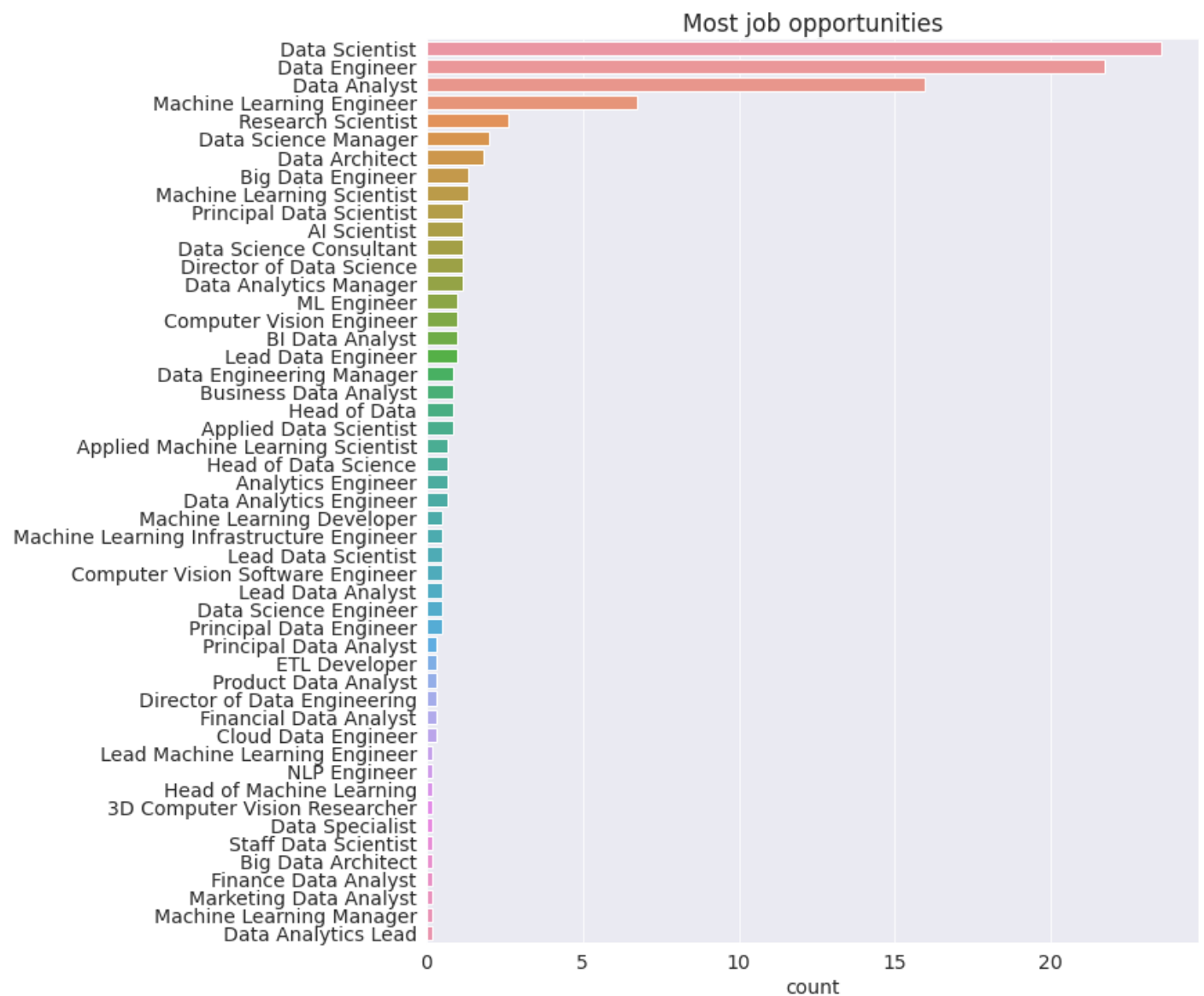
Data Scientist	23.558484
Data Engineer	21.746293
Data Analyst	15.980231
Machine Learning Engineer	6.754530
Research Scientist	2.635914
Data Science Manager	1.976936
Data Architect	1.812191
Big Data Engineer	1.317957
Machine Learning Scientist	1.317957
Principal Data Scientist	1.153213
AI Scientist	1.153213
Data Science Consultant	1.153213
Director of Data Science	1.153213
Data Analytics Manager	1.153213
ML Engineer	0.988468
Computer Vision Engineer	0.988468
BI Data Analyst	0.988468
Lead Data Engineer	0.988468
Data Engineering Manager	0.823723
Business Data Analyst	0.823723
Head of Data	0.823723
Applied Data Scientist	0.823723
Applied Machine Learning Scientist	0.658979
Head of Data Science	0.658979
Analytics Engineer	0.658979
Data Analytics Engineer	0.658979
Machine Learning Developer	0.494234
Machine Learning Infrastructure Engineer	0.494234
Lead Data Scientist	0.494234
Computer Vision Software Engineer	0.494234
Lead Data Analyst	0.494234
Data Science Engineer	0.494234
Principal Data Engineer	0.494234
Principal Data Analyst	0.329489
ETL Developer	0.329489

Product Data Analyst	0.329489
Director of Data Engineering	0.329489
Financial Data Analyst	0.329489
Cloud Data Engineer	0.329489
Lead Machine Learning Engineer	0.164745
NLP Engineer	0.164745
Head of Machine Learning	0.164745
3D Computer Vision Researcher	0.164745
Data Specialist	0.164745
Staff Data Scientist	0.164745
Big Data Architect	0.164745
Finance Data Analyst	0.164745
Marketing Data Analyst	0.164745
Machine Learning Manager	0.164745
Data Analytics Lead	0.164745

dtype: float64

Let's visualize barplot for the same.

```
plt.figure(figsize=(10,12))
sns.barplot(x = Percentage_df,y=Percentage_df.index)
plt.title("Most job opportunities")
plt.xlabel('count');
```



Data scientist,Data engineer and Data analyst have most job opportunities and Data Analytics Lead,Machine Learning Manager and Marketing Data Analyst have least job opportunities in the year 2020 to 2022.

Q2.How many jobs are there for entry level employees? what are the top 5 salaries have been given to them?

```
EN = ds1_salary_df.groupby('experience_level')[['job_title']].count()
EN
```

job_title	
experience_level	
EN	88
EX	26
MI	213
SE	280

There are total 88 jobs available for entry level employees in the year 2020 to 2022.

Now let's make the dataframe for entry level employee

```
EN_df = ds1_salary_df.loc[ds1_salary_df['experience_level'] == 'EN']
EN_df
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_resi
5	2020	EN	FT	Data Analyst	72000	USD	72000	
10	2020	EN	FT	Data Scientist	45000	EUR	51321	
12	2020	EN	FT	Data Scientist	35000	EUR	39916	
16	2020	EN	FT	Data Engineer	4450000	JPY	41689	
18	2020	EN	FT	Data Science Consultant	423000	INR	5707	
...
512	2022	EN	FT	Data Engineer	65000	USD	65000	
514	2022	EN	FT	Data Analytics Engineer	20000	USD	20000	
521	2022	EN	FT	Computer Vision Engineer	10000	USD	10000	
600	2022	EN	FT	Data Analyst	67000	USD	67000	
601	2022	EN	FT	Data Analyst	52000	USD	52000	

88 rows × 12 columns

Let's find out the top five salaries for entry level employees.

```
EN_Salaries = EN_df.sort_values('salary', ascending = False).head()
EN_Salaries
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_reside
16	2020	EN	FT	Data Engineer	4450000	JPY	41689	
109	2021	EN	FT	Data Engineer	2250000	INR	30428	
94	2021	EN	FT	Data Scientist	2200000	INR	29751	
253	2021	EN	FT	Data Scientist	2100000	INR	28399	
239	2021	EN	FT	Data Engineer	1600000	INR	21637	

The highest salary for EN employee is RS. 3441010 as Data Engineer in small scale company, which is surprisingly quite good amount.

Now let's plot graph for the same.

```
plt.title('Salaries for entry level employees')
plt.plot(EN_Salaries.salary, 'bo')
```

[<matplotlib.lines.Line2D at 0x7fdb400f95b0>]



Q3. How many companies provided remote working during these years of pandemic?

```
remote_working_companies = ds1_salary_df.remote_ratio.value_counts(ascending = True)
remote_working_companies
```

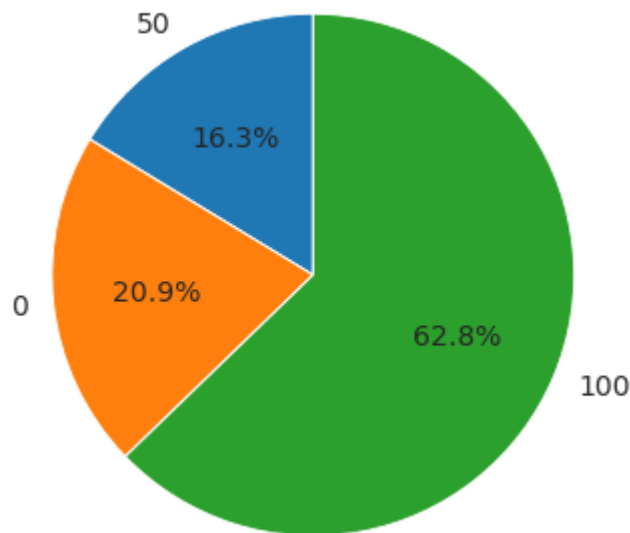
```
50      99
0      127
100     381
```

Name: remote_ratio, dtype: int64

It is good to know that most of the companies had offered remote [working.To](#) understand this better let's find out percentage of remote working in pie chart.

```
plt.figure(figsize = (12,6))
plt.title('Remote working ratio')
plt.pie(remote_working_companies, labels = remote_working_companies.index, autopct='%1.1f
```

Remote working ratio



From above Pie chart we can conclude, 62.8% companies provide complete remote working, 16.3% companies provide partial remote working and 20.9% companies does not provide remote working.

Q4.How many employees are living in their own country? and how many employees are migrated to another country for jobs?

```
Emp_living_inOwn_Country = ds1_salary_df.loc[ds1_salary_df['employee_residence'] == ds1_salary_df['employee_residence']]
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residen
0	2020	MI	FT	Data Scientist	70000	EUR	79833	I
1	2020	SE	FT	Machine Learning Scientist	260000	USD	260000	,
2	2020	SE	FT	Big Data Engineer	85000	GBP	109024	(
3	2020	MI	FT	Product Data Analyst	20000	USD	20000	+
4	2020	SE	FT	Machine Learning Engineer	150000	USD	150000	I
...
601	2022	EN	FT	Data Analyst	52000	USD	52000	(
602	2022	SE	FT	Data Engineer	154000	USD	154000	I
603	2022	SE	FT	Data Engineer	126000	USD	126000	I
604	2022	SE	FT	Data Analyst	129000	USD	129000	I

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residen
605	2022	SE	FT	Data Analyst	150000	USD	150000	I

556 rows × 12 columns

```
Emp_living_inOwn_Country.shape
```

(556, 12)

From 667 employees, 556 employees living in their own country. Let's find out the employees which have migrated to another country for good job opportunities.

```
Emp_Notliving_inOwn_Country = ds1_salary_df.loc[ds1_salary_df['employee_residence'] !=  
Emp_Notliving_inOwn_Country
```

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_re
17	2020	SE	FT	Big Data Engineer	100000	EUR	114047	
19	2020	MI	FT	Lead Data Engineer	56000	USD	56000	
32	2020	SE	FT	Data Scientist	60000	EUR	68428	
40	2020	MI	FT	Data Scientist	45760	USD	45760	
53	2020	EN	FT	Data Engineer	48000	EUR	54742	
54	2020	SE	FL	Computer Vision Engineer	60000	USD	60000	
61	2020	MI	FT	Data Engineer	130800	USD	130800	
70	2020	MI	FT	Data Scientist	55000	EUR	62726	
73	2021	EX	FT	BI Data Analyst	150000	USD	150000	
82	2021	MI	FT	Applied Data Scientist	68000	CAD	54238	
84	2021	EX	FT	Director of Data Science	130000	EUR	153667	
89	2021	SE	FT	Data Analyst	80000	USD	80000	
90	2021	SE	FT	Marketing Data Analyst	75000	EUR	88654	
96	2021	EN	PT	AI Scientist	12000	USD	12000	
99	2021	MI	FT	Computer Vision Software Engineer	81000	EUR	95746	

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_re
102	2021	MI	FT	BI Data Analyst	11000000	HUF	36259	
113	2021	EN	PT	AI Scientist	12000	USD	12000	
120	2021	MI	FT	Big Data Engineer	60000	USD	60000	
132	2021	MI	FT	Applied Machine Learning Scientist	38400	USD	38400	
149	2021	SE	FT	Cloud Data Engineer	160000	USD	160000	
156	2021	MI	FT	Data Scientist	160000	SGD	119059	
179	2021	MI	FT	Data Scientist	420000	INR	5679	
182	2021	MI	FT	Data Engineer	22000	EUR	26005	
186	2021	SE	FT	Data Analytics Engineer	50000	USD	50000	
187	2021	EX	FT	Data Science Consultant	59000	EUR	69741	
188	2021	SE	FT	Data Engineer	65000	EUR	76833	
190	2021	SE	FT	Data Science Manager	152000	USD	152000	
198	2021	SE	FT	Data Science Manager	4000000	INR	54094	
200	2021	MI	FT	Data Scientist	52000	EUR	61467	
203	2021	SE	FT	Research Scientist	50000	USD	50000	
208	2021	MI	FL	Data Engineer	20000	USD	20000	
212	2021	MI	FT	Data Engineer	48000	GBP	66022	
213	2021	EN	FT	Big Data Engineer	435000	INR	5882	
224	2021	SE	FT	Machine Learning Scientist	225000	USD	225000	
244	2021	EN	FT	AI Scientist	1335000	INR	18053	
264	2021	MI	FT	Data Scientist	21600	EUR	25532	
265	2021	SE	FT	Lead Data Engineer	160000	USD	160000	
273	2021	EN	FT	Machine Learning Engineer	85000	USD	85000	

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_re
281	2021	EN	FT	Research Scientist	100000	USD	100000	
396	2022	MI	FT	Machine Learning Engineer	80000	EUR	87932	
417	2022	SE	FT	Data Science Engineer	60000	USD	60000	
488	2022	MI	FL	Data Scientist	100000	USD	100000	
489	2022	EN	CT	Applied Machine Learning Scientist	29000	EUR	31875	
490	2022	SE	FT	Head of Data	200000	USD	200000	
494	2022	SE	FT	Data Scientist	100000	USD	100000	
496	2022	EN	FT	Data Engineer	52800	EUR	58035	
502	2022	EN	FT	Data Scientist	40000	USD	40000	
506	2022	MI	FT	Applied Machine Learning Scientist	75000	USD	75000	
515	2022	MI	FT	Data Scientist	48000	USD	48000	
521	2022	EN	FT	Computer Vision Engineer	10000	USD	10000	
606	2022	MI	FT	AI Scientist	200000	USD	200000	

```
Emp_Notliving_inOwn_Country.shape
```

```
(51, 12)
```

From 667 employees only 51 employees have been migrated to another countries for jobs. Most of the employees are living in their own country.

Q5.What are the counts for each employment type? and what are the salaries given to lowest count employment type?

```
ET = ds1_salary_df.employment_type.value_counts()
ET
```

From 607 employees,588 employees are full time employees, 10 employees are part time employees,5 employees are contract employees and only 4 employees are freelancer. Now let's visualize this in bar plot.

```
plt.figure(figsize=(10,5))
sns.barplot(x = ET,y=ET.index)
```

```
plt.title("Employment Type")
plt.xlabel('count');
```

As we can see most of the employees are full time and least of the employees working as a freelancer. Let's find out what are salaries provided for freelancer in Data Science domain.

First, let's make data frame for freelancer employment.

```
FL_df = ds1_salary_df.loc[ds1_salary_df['employment_type'] == 'FL']
FL_df
```

Now let's sort the salaries to see clear picture.

```
FL_Salaries = FL_df.sort_values('salary', ascending = False)
FL_Salaries
```

Even though there are very few folks work as Freelancer but the highest salary offered is RS.8254000, which is quite good. There is no entry level employee working as freelancer in Data Science domain. If you want to work at your own space and time then first you should take some experience as full time employee then you can earn handsome amount as Freelancer.

Now let's visualize salaries of freelancer on basis of their job titles.

```
plt.figure(figsize=(15,5))
plt.title("Salaries and Job title for Freelancer Employment")
plt.bar(FL_Salaries.job_title, FL_Salaries.Salary_in_RS);
plt.xlabel('Job title')
plt.ylabel('Salaries');
```

Highest salary for Freelancer in Data Science domain is Data Scientist followed by Computer Vision Engineer, Data Engineer and Machine learning Scientist.

Q6. Which country has highest and lowest count for working in Data Science Domain?

```
Highest_CC = ds1_salary_df.employee_residence.value_counts().head()
Highest_CC
```

```
Lowest_CC = ds1_salary_df.employee_residence.value_counts().tail()
Lowest_CC
```

Let's plot Pie charts for both Highest country count and lowest country count.

```
plt.figure(figsize = (12,6))
plt.title('Employee Residence Country')
plt.pie(Highest_CC, labels = Highest_CC.index, autopct='%1.1f%%', startangle = 60);
```

72.2% people who work in Data Science domain are from US followed by UK and India.

```
plt.figure(figsize = (12,6))
plt.title('Employee Residence Country')
plt.pie(Lowest_CC, labels = Lowest_CC.index, autopct='%1.1f%%', startangle = 60);
```

BG,CH,NZ,PH and HR have least people(20%) working in DataScience Domain.

Q7. What is the maximum salary(In Rupees) offered in data science domain in the year 2020 to 2022?

```
ds1_salary_df.Salary_in_RS.max()
```

Quite handsome highest salary have been offered in Data [Science.So](#) one should consider data Science as a career path.

Inferences and Conclusion

The main aim of the project was to analyse the different jobs available in the year 2020 to 2022.

We also did a deep analysis through charts and by raising important questions. Let us go through some of the important analysis, we have done through this project -

1. There are total 50 different jobs available in Data Science domain in the span of 2020-2022.
2. Out of this 50 different jobs, maximum opportunities available for Data Scientist,Data Engineer and Data Analyst.
3. Most of the companies had provided remote working in these pandemic.
4. Maximum Salary offered in Data Science Domain is RS.49524000.
5. Senior level employee have maximum jobs available followed by Mid level and Entry level.
6. Medium scale companies offering more jobs in Data Science domain.
7. Most jobs available in USA and least jobs available in China.

References

Reference -

- 1) All of my doubts has been clared at <https://stackoverflow.com>
- 2)For knowing more about pandas and it's functions in deatil - <https://pandas.pydata.org>
- 3) For more ideas on Matplotlib and it's library- <https://matplotlib.org>
- 4)Also for many coding related doubts, refferd to - <https://www.w3schools.com>