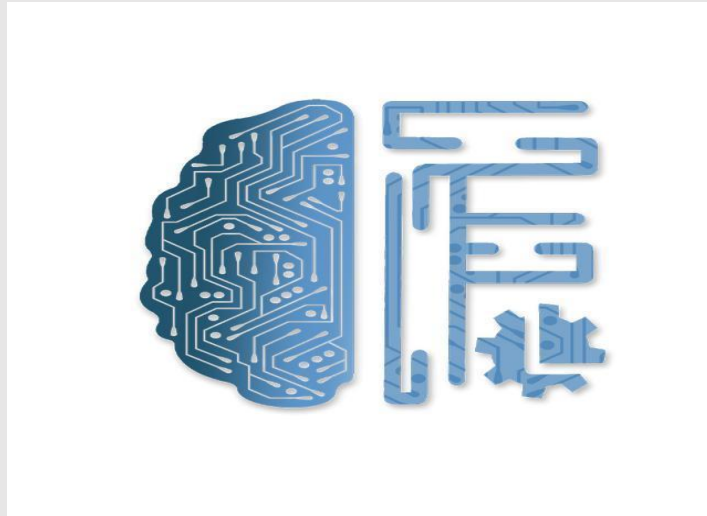


Data Science Job Market Analysis



By-

Team-B

Rahul Rao

Ashish Kumar

Bharti Lendey

Santhanam P

Beerapalli Sai Sanhitha

APPLYING THE FERMI ESTIMATION TECHNIQUE TO JOB ANALYSIS

Job Market

The Job market is the place where the supply and the demand for jobs meet, with the workers or Job providing the services that employers demand. The worker may be anyone who wishes to offer his services for compensation, while the employer may be a single entity or an organization that needs an individual to do a specific job or to complete a task. The worker is then comparable to a seller while the employer is the buyer.

A common factor that connects the two entities is the salary or wage that is agreed to be received by the worker from the employer. In short, this is where workers can find work that suits their skills and qualifications and where both agree on the wages, benefits, and other forms of compensation for the worker.

In the Job market, it is assumed that workers move to where there is a demand for their skills, whether this is in their local region or abroad. Moreover, they are also replaceable, which means that a person who can do the job better can be tapped to take over the other worker's job. Furthermore, salaries are not fixed, meaning they can go up or down, depending on the worker's performance. Wages or compensation is the highest motivating factor in the Job market.

Components of the Job Market

The Job market comprises four components: The Job force population, applicant population, applicant pool, and the individuals selected.

1.Job force population

The Job force population or Job force participation refers to the number of individuals who are available to work in a Job market. It considers all workers who are offering their skills and services for employment regardless of the industry they are in.

2.Applicant population

The second component is the applicant population which refers to the people who are applying for a particular job that suits their expertise and skills. Recruiters look first at the Job market and then look next for individuals who meet the skills and qualifications set for a particular job. For example, the

people looking for IT, graphics design, and similar jobs belong to the same applicant population, which is targeted by recruiters looking for this type of professional.

3.Applicant pool

The third component is the applicant pool, which is the actual number of people who initially signified their interest in applying for a particular job by sending in their resume. It may very well be considered the first part of the selection process where the recruitment department of a specific organization receives applications and screens them to determine who advances to the next round of screening.

4.Individuals selected

The fourth component is the individuals selected, which simply means the individual or individuals who've made it through the screening process and have been hired for the job. Of course, this is judged based on a number of factors, and the person is screened against a carefully determined set of qualifications.

Understanding Job Market Analysis

Job market analysis is an integral part of an organization's recruitment process because it not only helps it find the most qualified workers for the jobs that it offers but also ensures that it provides a competitive compensation package to its workers. This is important in order for an organization to be able to keep its competent workers and, thus, continue its productivity.

Job market analysis involves the following processes:

Identifying the various Job markets for a given type of position. It involves looking at the appropriate Job market based on a specific position.

Checking the market for salaries for a common position. The process involves checking similar positions in the Job market in order to determine if an organization's salary rates are at about the same level.

Determining market trends. This step answers questions as to how other organizations are compensating their workers, including their pay practices.

Adjusting salary packages or structure of positions. After checking the salary rates of other organizations and finding out if there is any need for adjustments, the department then makes recommendations for such adjustments and restructuring of positions in the company.

Making consultations with management. This process involves sitting down with management to determine their workforce needs.

Data Science Job Market Analysis

Data Science has become an indispensable part of many businesses and industries. It provides valuable insights into customer behavior that can lead to increased conversions, more detailed market analysis for competitive advantage in pricing strategies or product development, improved operational efficiency, and minimized risk exposure through accurate forecasting models.

The emergence of disruptive technologies like IoT, digital media platforms, smartphones, artificial intelligence, big data analytics, blockchain, and quantum computing has ushered in an era where Data Science will be central to organizational success.

High salary growth in DS and ML

As organizations are turning towards Machine Learning, Big Data, and Artificial Intelligence, the demand for data science roles is seeing a sustained and accelerating upward surge. Since 2012, the Data Science sector has witnessed a massive hike of 650%, far outpacing other sectors.

Therefore, transitioning to data science is a smart move as it fetches far higher comparative returns. For instance, transitioning from a marketing analytics job to a data science job leads to a 37% salary growth on average.

Similarly, the expected salary growth when people switch from digital analytics to data science is 31% on average. For someone transitioning from a data engineering role to a data science role, the salary hike could be as high as 44%.

According to recruitment firm Michael Page's 2021 India Talent Trends report, data science professionals with 3-10 years of experience get annual salaries in the 25-65 lakh range, while those with more experience can command pay packages upwards of 1 crore.

Professionals with over 15 years of experience can get paid up to 1.8 crores. Similarly, the average annual pay hike for data science professionals falls between 20-30% compared with 15-20% for professionals from other backgrounds.

Increasing demand for data science professionals

India Inc is witnessing the rapid digitization of businesses and services, making it the second-largest hub for data science in the world. Analysts predict that the country will have more than 11 million job openings by 2026. In fact, since 2019, hiring in the data science industry has increased by 46%.

Yet, around 93,000 jobs in Data Science were vacant at the end of August 2020 in India. 70% of these vacancies were for positions with less than five years of experience.

Data Source :

Data was scraped from Naukri Job using Selenium web framework and BeautifulSoup package for parsing HTML and XML documents. The data consists of 17 columns and 7967 rows after Data preprocessing.

Out[21]:

	Job_Role	Company	Experience	Location	Key_Skills	Post_History	machine learning	python	data science	analytical
0	data science - machine learning engineer	BCG Attorney Search	5-8 Yrs	Ahmedabad (WFH during Covid)	algorithms,business intelligence,visualization...	5 Days Ago	1	0	0	0
1	faculty for machine learning and big data job ...	MKSSS-AIT, Center for Data Science, Machine Le...	0-3 Yrs	Pune	training,machine learning,big data,sql server ...	7 Days Ago	1	0	0	0
2	computer vision/artificial intelligence/machin...	e-con Systems	0-1 Yrs	Chennai	it skills,python,machine learning,artificial i...	5 Days Ago	1	1	0	0
3	data scientist - image recognition/machine lea...	CarbyneTech India	5-10 Yrs	Hyderabad	image recognition,machine learning,python,hive...	5 Days Ago	1	1	0	0
4	specialist- data analyst- machine learning/spo...	Carrierin Services	3-6 Yrs	Bangalore	hvac,analytical,artificial intelligence,machin...	1 Day Ago	1	1	0	1
...

Average Experience	Salary	Frequency
6	1250000	21
2	625000	158
0	625000	160
8	1250000	333
4	1250000	158
...
6	1250000	158
6	1250000	333
4	1250000	140
4	1250000	279
4	1250000	279

Data Preprocessing:

3- columns were preprocessed since it had many unique values .The columns which were preprocessed were Average Experience, Salary, Company Name, Location. Company Name had approx. 5,000 unique values which was converted from string to numeric as Frequency column so companies which occurred frequently had more weightage .

```
In [69]: data_frequency_map=data['Company'].value_counts().to_dict()

In [70]: data_frequency_map
{'SAP India Pvt.Ltd': 164,
 'Cunesoft India Private Limited (a Phleglobal Company)': 162,
 'Infosys': 161,
 'e-con Systems': 160,
 'AIM recruits': 159,
 'Carrierin Services': 158,
 'JioSaavn': 158,
 'MKSSS-AIT, Center for Data Science, Machine Learning and AI': 158,
 'Zycus Infotech': 157,
 'Quinnox Consultancy Services Limited': 157,
 'eClerx': 140,
 'Dell': 43,
 'Diverse Lynx': 34,
 'CHANGE LEADERS CONSULTING': 33,
 'Career Maker': 33,
 'First Employer': 28,
 'Intel': 27,
 'Qualcomm Technologies, Inc': 26,
 'Walmart': 25,
 '...': 25}

In [71]: #Now we replace company labels in dataset
data['Frequency']=data['Company'].map(data_frequency_map)
```

Since salary and Experience have High correlation when it comes to IT Jobs. Salary column values were created using Industry standard salaries based on Experiences.

```
In [64]: def sal(text):
        if text==0 or text==1 or text==2 or text==3:
            return 625000
        elif text==4 or text==5 or text==6 or text==7 or text==8:
            return 1250000
        elif text==9 or text==10 or text==11 or text==12 or text==13 or text==14 or text==15:
            return 2200000
        elif text==16 or text==17 or text==18 or text==19 or text==20:
            return 2900000
        elif text==21 or text==22 or text==23 or text==24 or text==25:
            return 4000000

In [65]: data['Salary']=data['Average Experience'].apply(sal)
```

```

#Updating Location
NS_df['Location'] = NS_df['Location'].replace('Bangalore/Bengaluru', 'Bangalore', regex=True)
NS_df['Location'] = NS_df['Location'].replace('Delhi / NCR(WFH during Covid)', 'Delhi (WFH during Covid)')
NS_df['Location'] = NS_df['Location'].replace('Ahmedabad(Bodakdev)(WFH during Covid)', 'Ahmedabad (WFH during Covid)')
NS_df['Location'] = NS_df['Location'].replace('Hyderabad/Secunderabad', 'Hyderabad', regex=True)
NS_df['Location'] = NS_df['Location'].replace('Gurgaon/Gurugram', 'Gurgaon', regex=True)
NS_df['Location'] = NS_df['Location'].replace('Delhi / NCR ', 'Delhi')

#Removing unwanted characters in Skills

```

Top Skills were selected to reduce the Unique Values :

```

[2]: NS_df = pd.read_csv('Naukri_FF.csv')

[3]: #Checking dataset
NS_df = pd.read_csv('Naukri_FF.csv')

# Selecting Specific role
NS_df['Job_Role'] = NS_df['Job_Role'].str.lower()
J_R = ['data analyst', 'data scientist', 'machine learning', 'data science', 'deep learning', 'data engineer']
NS_df['Imp'] = NS_df.apply(lambda n: 1 if any(word in n.Job_Role for word in J_R) else 0, axis=1)
NS_df = NS_df[NS_df['Imp']==1]
NS_df = NS_df.drop(['Imp'], axis=1)
NS_df = NS_df.drop(['Salary'], axis=1)
#Dropping empty Skills row
NS_df = NS_df[NS_df['key_skills']!='[]']

```

Segment Extraction :

The Machine learning Techniques used were Principle Component Analysis(PCA) to reduce the Dimensionality curse .17 columns were reduced to 2 Principle components.

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover, PCA is an unsupervised statistical technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

```

[8]: pca = PCA(n_components = 2)
X_train = pca.fit_transform(scaleddata)

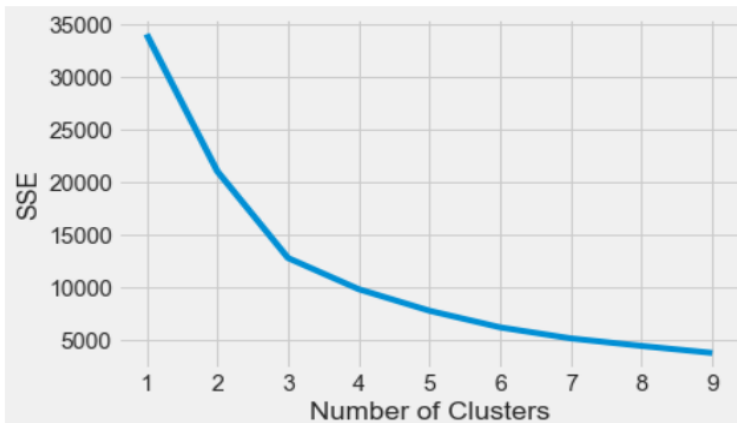
```

After PCA,K-means Clustering Machine learning technique was used and 3 segment's were obtained

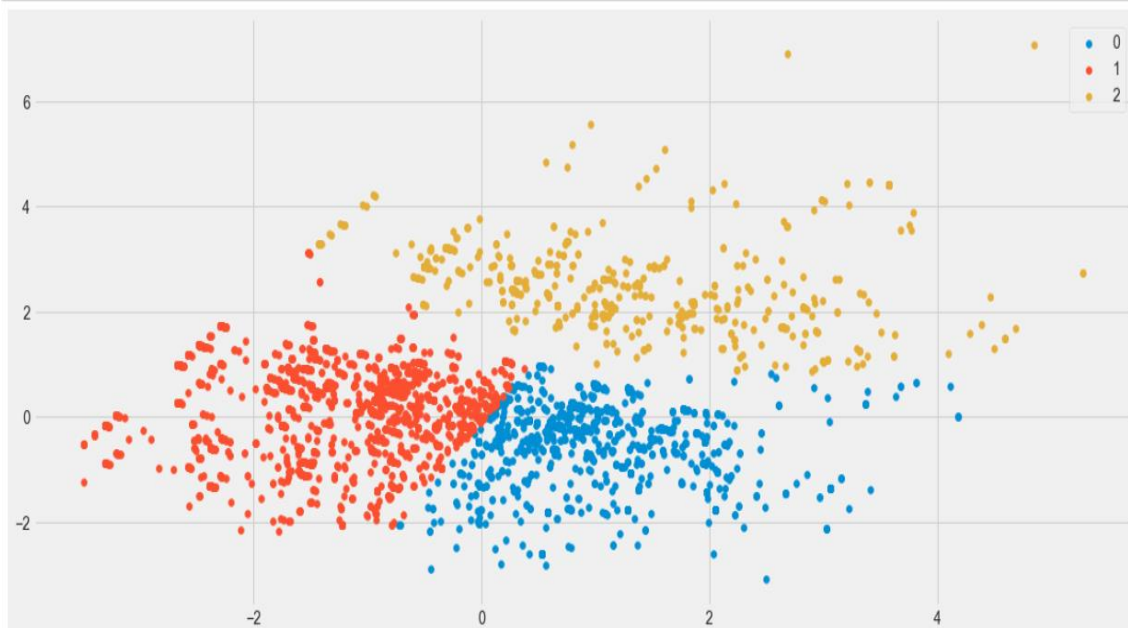
K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. In this topic, we will learn what is K-means clustering algorithm, how the algorithm works, along with the Python implementation of k-means clustering.

```
In [79]: k_range = range(1,10)
sse = []
for k in k_range:
    km = KMeans(n_clusters = k)
    km.fit(X_train)
    sse.append(km.inertia_)
```

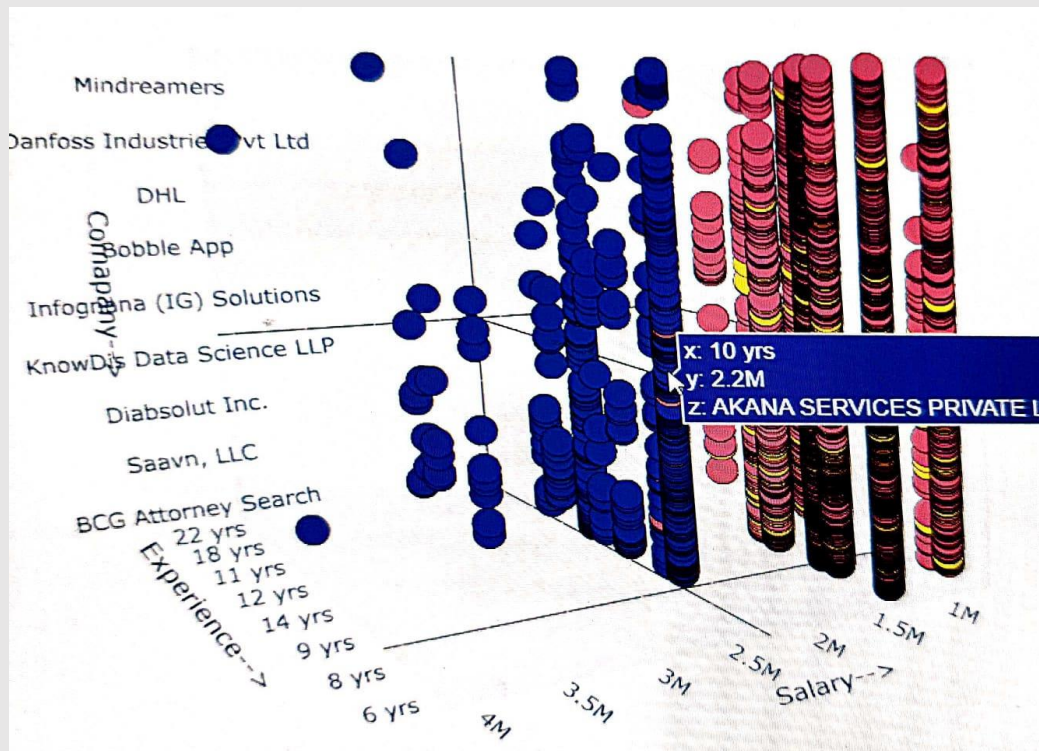
```
In [80]: plt.style.use("fivethirtyeight")
plt.plot(range(1, 10), sse)
plt.xticks(range(1, 10))
plt.xlabel("Number of Clusters")
plt.ylabel("SSE")
plt.show()
```



```
In [83]: u_labels = np.unique(label)
plt.figure(figsize=(18,8))
for i in u_labels:
    plt.scatter(X_train[label == i , 0] ,X_train[label == i , 1] , label = i)
plt.legend()
plt.show()
```



Interactive 3-dimensional Visualization



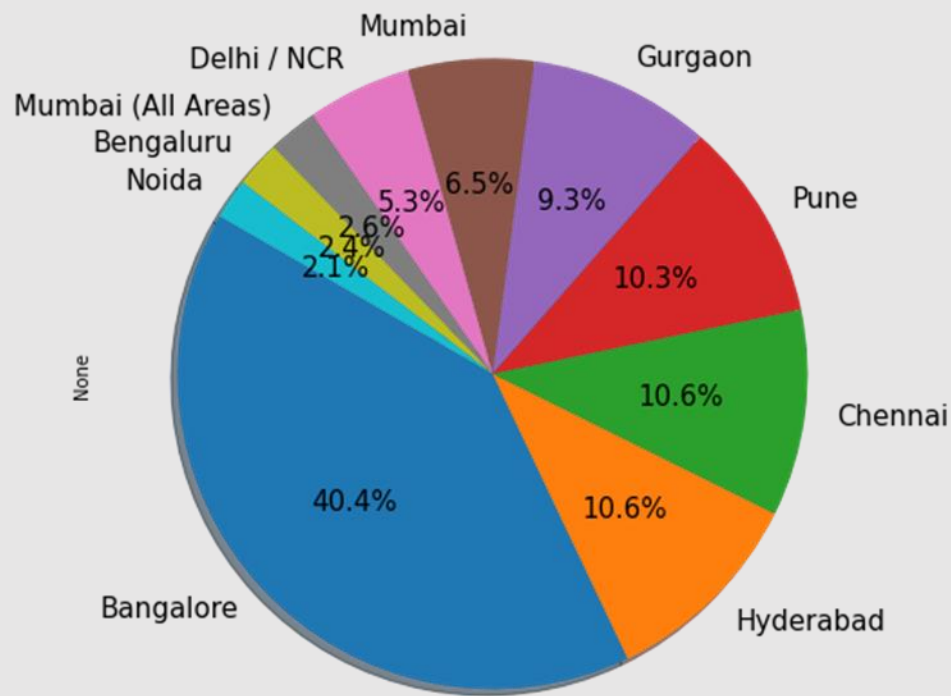
Market Profile

Definitions of a market segment profile are often quite similar, as shown by the following two definitions:

- A [market segment] profile might may information about lifestyle patterns, attitudes towards product attributes as well as brands, product-use habits, geographic locations, and demographic characteristics.
- A market segment profile explains the similarities among potential customers within a segment and also explains the differences among people as well as organizations in different segments.

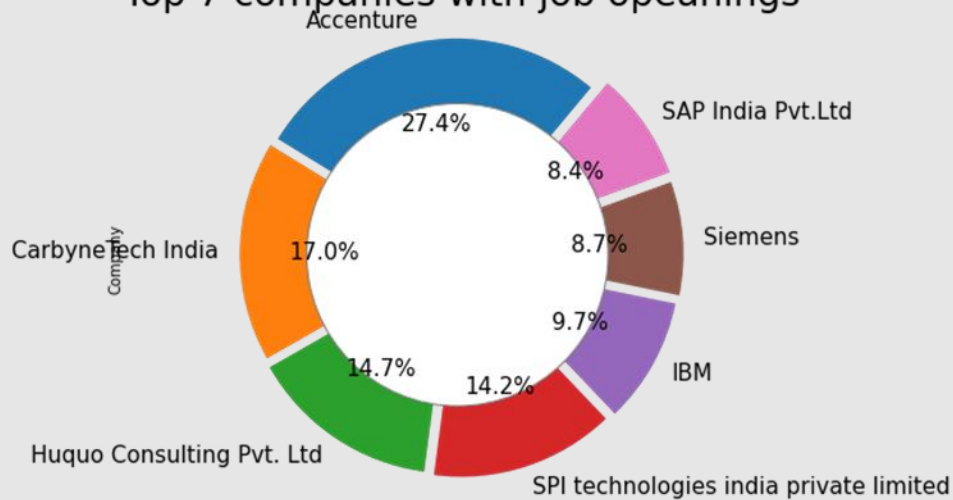
In our project:

Geographic distribution,

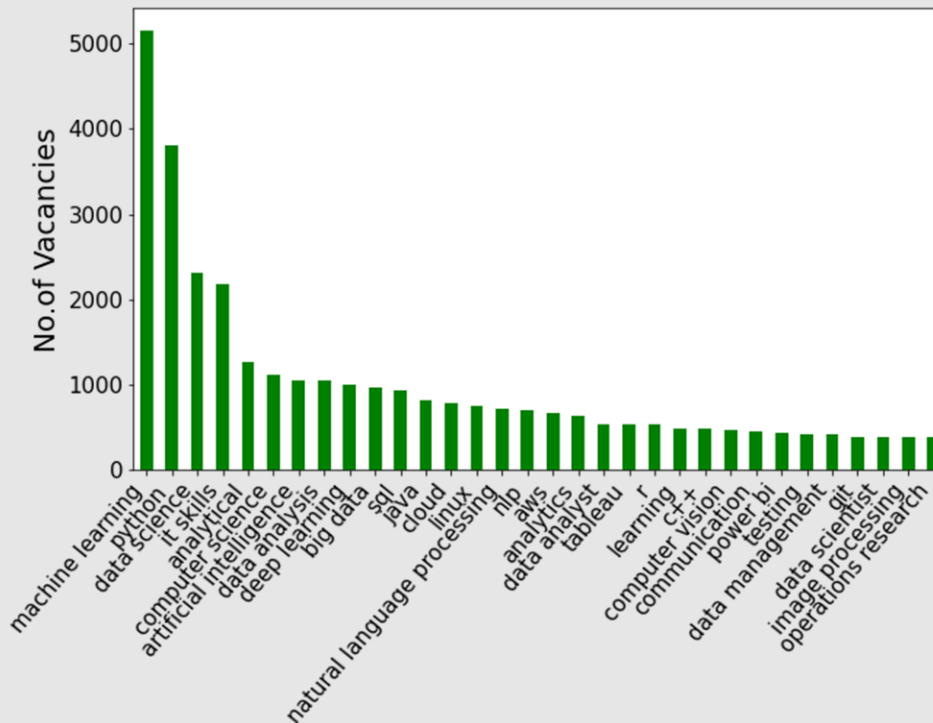
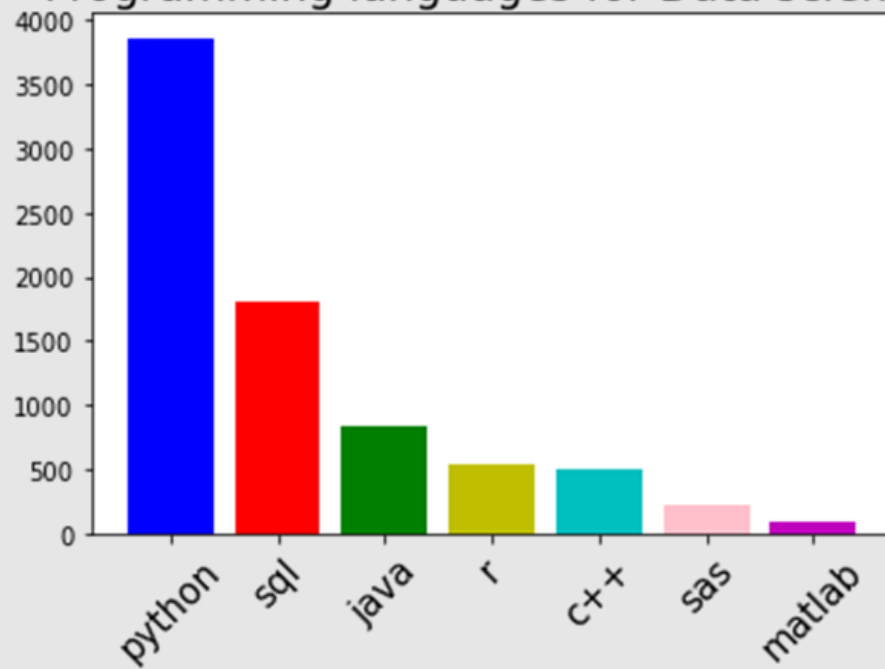


Top companies with most data science job openings

Top 7 companies with job opeanings

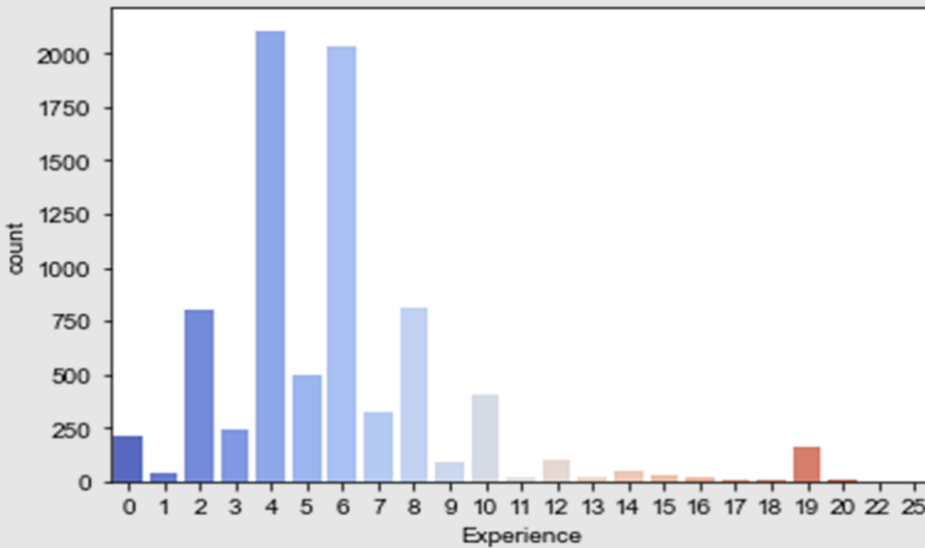


Programming languages for Data science



Top Skills for Data science

Year wise experience distribution

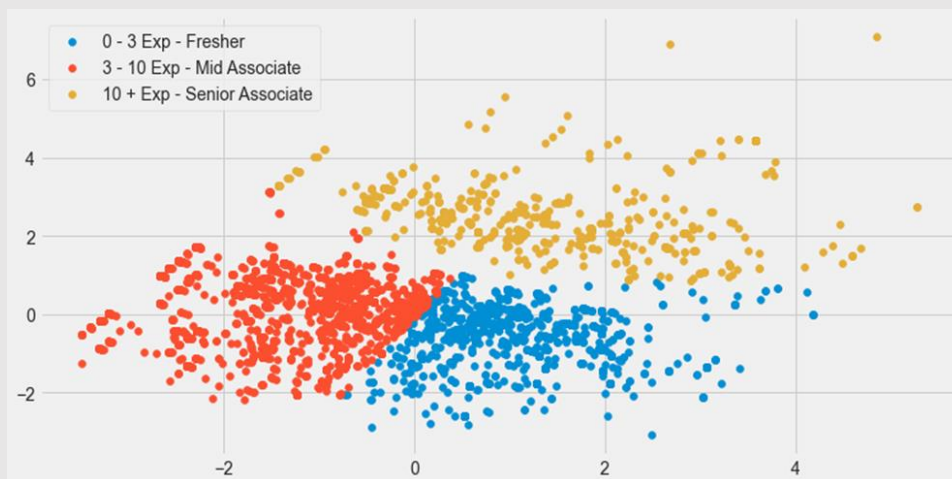


Segmentation

market segmentation is the practice of dividing your target market into approachable groups. Market segmentation creates subsets of a market based on demographics, needs, priorities, common interests, and other psychographic or behavioral criteria used to better understand the target audience.

When it comes to the project, we use PCA for reducing the dimensions of the training data. Then use K means clustering algorithm to cluster the data into 3.

- 0-3 Exp – Fresher
- 3-10 Exp – Mid Associate
- 10+ Exp – Senior Associate



7. Customizing the market mix:

The data science job market has been analyzed and clustered into groups based on experience, salary, key skills and company. There are a total of 3 clusters -

Cluster Blue : The job positions falling into this cluster are suitable for recent graduates and freshers, who are at the beginning stages of their career with 0-3 years of experience.

Cluster Red : The job positions available in this cluster are recommended for those with mid experience level i.e., 3-10 years of experience. These jobs can help them grow and accelerate their career to new heights. Majority of the job positions are from medium-large sized companies.

Cluster Yellow : The potential targets for job openings in this cluster are senior employees with work experience of 10 years or more.

Marketing mix is a combination of tools often known as 4 P's of marketing

The goal is come up with an ideal marketing strategy that takes all these 4 P's into consideration.

For the data science job market, the product is job openings in the field. Price is the salary offered for that job position by the company. Place is the location of job and promotion typically includes the strategies to target and reach out to the potential customers based on the other P's.

Following the segmentation-targeting-positioning (STP) approach we can understand in detail the various job clusters formed by our model. Based on that

we can follow the 4P's model of marketing mix to identify the potential targets for each cluster and achieve the customization of the data science job market .



Github Link-<https://github.com/RahulRao333/Data-Science-Job-Market-Analysis->