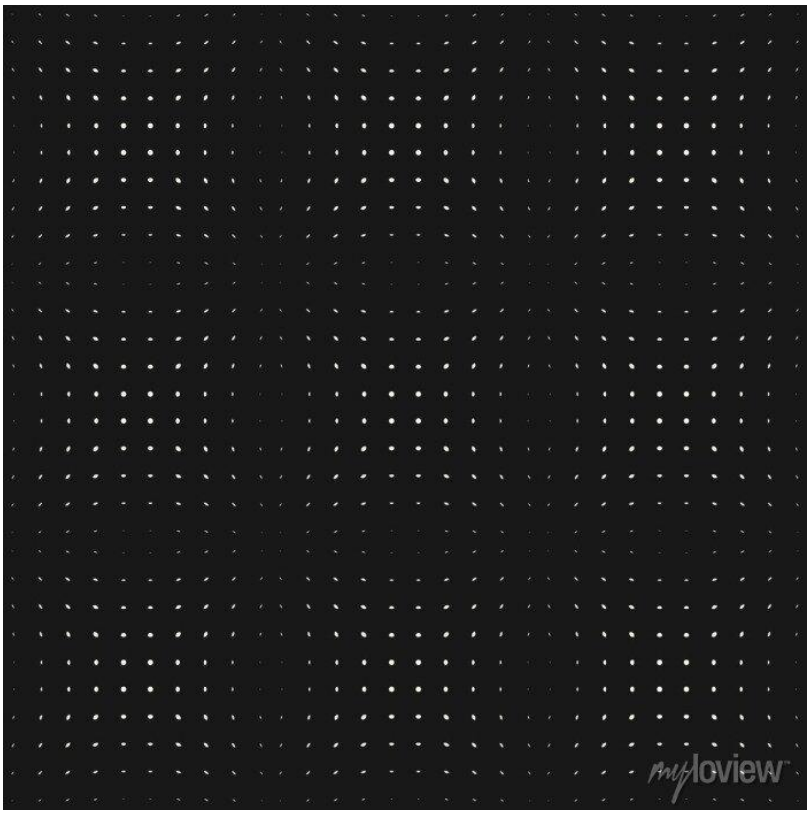
PROJECT - III

FEYNN LABS

III. JOB MARKET



TEAM VISHAKHA

VISHAKHA WARUDKAR
MINTU MEDHI
UJWAL KOTHAPALLY
HARSH LANGADE

PROBLEM STATEMENT

Finding Companies most probable to hire an ML Engineer/Data Analyst Applicant with respect to his/her skillset.

I. INTRODUCTION

The report is the result of the second project given my Feynn Labs based on Market Segmentation. The problem statement above explains the target of this project. The following explains the steps taken to complete the market segmentation based on data acquired from a job posting website called Naukri. This project is an adaptation of the previous study project on market segmentation.

Fermi Estimation

Breaking down the problem using fermi estimation. Let's assume a company has an opening for a Senior Data Scientist job. One of the major attributes this profile requires is the minimum experience of a few years in the assumption that the applicant is exceptionally knowledgeable and can work their way in the industry. Next, the other important factor in determining the probability to hire is the applicant's skillset. Giving weightage to their most honed skill, we will assume that this is Machine Learning/ Data Science. The other skills such as Deep Learning, Data Analysis etc. will be a bonus for this job. In conclusion, using fermi estimation we can assume that a person with Machine Learning as his top skillset and at least 5 years of experience has a high probability of acquiring this job. The problem can be further improved with the addition of the status of the company as well as geographic location correlating the company's work as on-site or remote with that of the applicant's location and their choice of the two.

II. MARKET SEGMENTATION

STEP 1: DECIDING (NOT) TO SEGMENT

The use of machine learning to help a prospective applicant land a job in the market with respect to various factors such as their skills, geographical location, role, etc. With the help of machine learning models it's possible to derive the most compatible roles in companies with the highest probability to hire them.

STEP 2: IDEAL TARGETING SEGMENTS

A company, while hiring, reviews various factors as well as the applicant, while applying for jobs seeks out desirable factors. The most important segments would be the company itself, its geographical location, salary offered and the roles that are available. With respect to the applicant we should consider features such as skill they're proficient in, the amount of experience in the field of application and salary expectation.

STEP 3: COLLECTING DATA

Empirical data forms the basis of data-driven market segmentation.For job market segmentation, we will use either single or multiple segmentation variables each for geographical,demographic and psychographic segmentation.

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

In [2]:
```
data = pd.read_excel("D:\\ML Project\\Feynn_Labs_Project_3\\naukri.xlsx")
data = pd.DataFrame(data)
data.head(5)
```

Out[2]:

	title	title href	subTitle	subTitle href	starRating	reviewsCount	reviewsCount href	elli
0	Global Tax Automation & Operations - Data Scie...	https://www.naukri.com/job-listings-global-tax...	Dell	https://www.naukri.com/dell-jobs-careers-27614	4.1	(2029 Reviews)	https://www.ambitionbox.com/reviews/dell-revie...	3-5
1	Senior Analyst - Data Science	https://www.naukri.com/job-listings-senior-ana...	Tiger Analytics India LLP	https://www.naukri.com/tiger-analytics-india-c...	3.9	(31 Reviews)	https://www.ambitionbox.com/reviews/tiger-anal...	1-(
2	Sr. Analyst - Data Engineering	https://www.naukri.com/job-listings-sr-analyst...	Dell	https://www.naukri.com/dell-jobs-careers-27614	4.1	(2029 Reviews)	https://www.ambitionbox.com/reviews/dell-revie...	3-(
3	Lead - Data Analyst	https://www.naukri.com/job-listings-lead-data-...	CARDINAL HEALTH INTERNATIONAL INDIA PRIVATE LI...	https://www.naukri.com/cardinal-health-interna...	4.3	(14 Reviews)	https://www.ambitionbox.com/reviews/cardinal-h...	
4	Data Analyst	https://www.naukri.com/job-listings-data-analy...	Netomi	https://www.naukri.com/netomi-jobs-careers-545...	3.6	(2 Reviews)	https://www.ambitionbox.com/reviews/netomi-rev...	3-(

5 rows × 21 columns

STEP 4: EXPLORING DATA

The data shown above is collected from web scraping of the Naukri website(Naukri.com). After collecting the data it is to be cleaned so that we can use it in our analysis.
First we renamed the columns as per our convenience…

In [4]:
```
data = pd.DataFrame(data,columns=['title', 'subTitle','ellipsis', 'ellipsis 3', 'fleft', 'fleft 2', 'fleft 3',
       'fleft 4', 'fleft 5', 'fleft 6', 'fleft 7', 'fleft 8'])
data.head(2)
```

Out[4]:

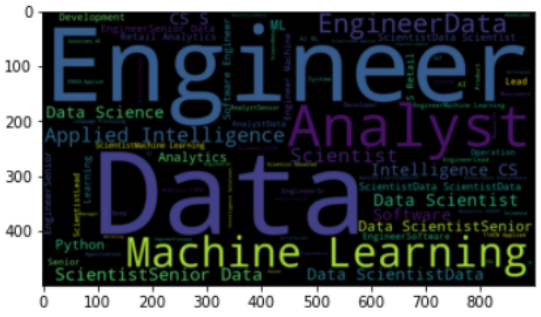
	title	subTitle	ellipsis	ellipsis 3	fleft	fleft 2	fleft 3	fleft 4	fleft 5	fleft 6	fleft 7	fleft 8
0	Global Tax Automation & Operations - Data Scie...	Dell	3-5 Yrs	Bangalore/Bengaluru	Artificial Intelligence	Data Science	Data Analytics	IT Skills	Python	Testing	Machine Learning	RPA
1	Senior Analyst - Data Science	Tiger Analytics India LLP	1-6 Yrs	Chennai	verbal communication	written	python	R	IT Skills	Data Science	Machine Learning	Artificial Intelligence

In [5]:
```
data.rename(columns = {'title':'Work Profile', 'subTitle':'Company Name', 'ellipsis': 'Years of Experience',
                       'ellipsis 3': 'Location', 'fleft' : 'Skill 1', 'fleft 2' : 'Skill 2', 'fleft 3':'Skill 3',
                        'fleft 4': 'Skill 4', 'fleft 5': 'Skill 5', 'fleft 6': 'Skill 6', 'fleft 7':'Skill 7',
                        'fleft 8':'Skill 8'}, inplace = True)
data.head(2)
```

Out[5]:

	Work Profile	Company Name	Years of Experience	Location	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6	Skill 7	Skill 8
0	Global Tax Automation & Operations - Data Scie...	Dell	3-5 Yrs	Bangalore/Bengaluru	Artificial Intelligence	Data Science	Data Analytics	IT Skills	Python	Testing	Machine Learning	RPA
1	Senior Analyst - Data Science	Tiger Analytics India LLP	1-6 Yrs	Chennai	verbal communication	written	python	R	IT Skills	Data Science	Machine Learning	Artificial Intelligence

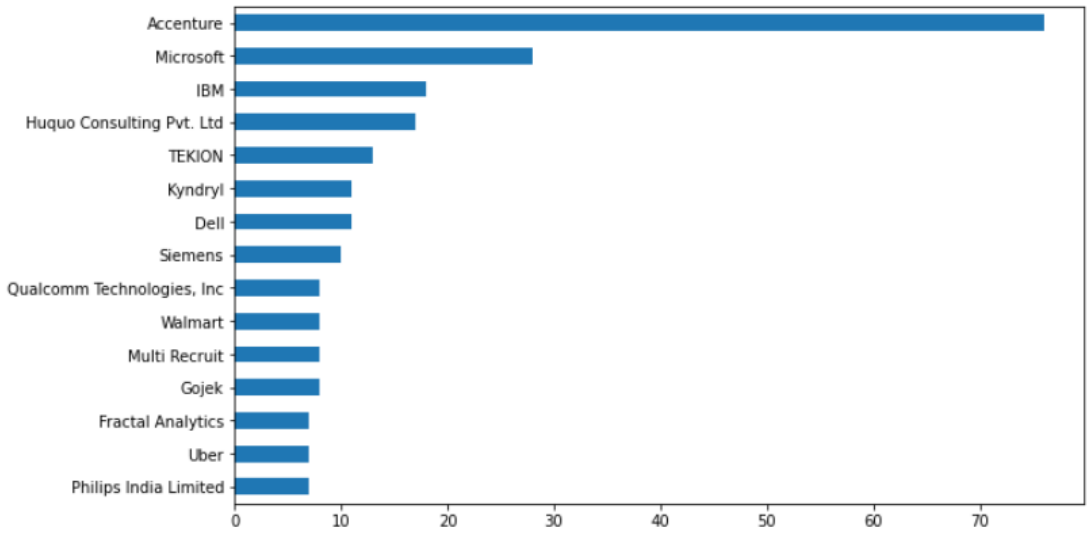
In [6]:
```
from wordcloud import WordCloud
text = "".join(data['Work Profile'].dropna().to_list())

wordcloud = WordCloud(width = 900, height = 500, max_words = 80, mode = 'RGBA').generate(text)
plt.figure()
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis = ("off")
plt.margins(x = 0, y = -0)
plt.show()

```



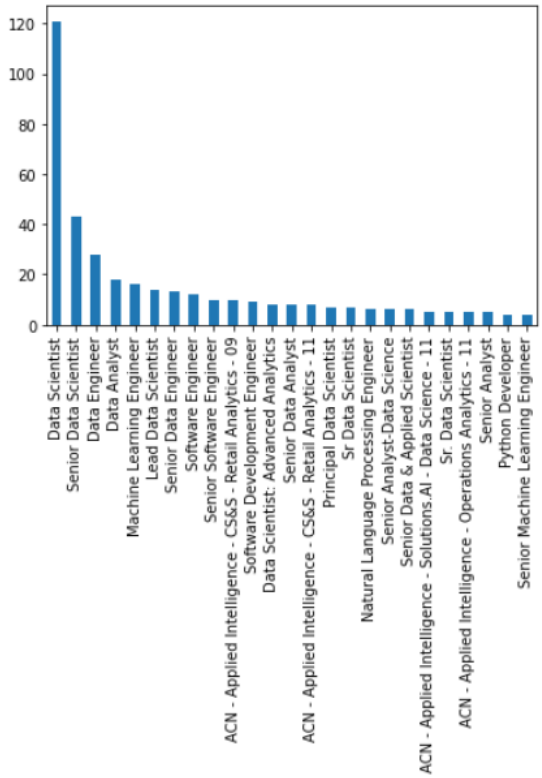
Word Cloud is a data visualization technique used to represent text data in which the size of each word indicates its frequency or importance. Important text data points can be highlighted using cloud words.

Here the word cloud shows the Work Profile, Large size of text will show its importance and frequency.

```
In [7]: 1 data['Company_Name'].value_counts().head(15).sort_values().plot.barh(figsize = (10,6));
```




```
In [5]: 1 data['Work Profile'].value_counts().head(25).plot.bar();
```



..and then we have extracted top 10 most frequent companies in terms of job posting

Then we split location and experience to keep only one value and further preprocessing was done to extract the valuable data.

```
In [220]: df_location.head()
```

Out[220]:

Name	Location	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6	Skill 7	Skill 8	Min Experience	Work_F
Dell	Bangalore	Artificial Intelligence	Data Science	Data Analytics	IT Skills	Python	Testing	Machine Learning	RPA	3	data_a
Dell	Bangalore	Azure	Data Engineering	ECS	IT Skills	Data Science	Machine Learning	Cloud	MS SQL	3	data_a
IBM	Bangalore	IT Skills	Java	Python	Software Development	Data Science	Machine Learning	Cloud	Computer science	1	data_a
KION	Bangalore	IT Skills	Python	Testing	Data Science	Project Management	Tableau	Power BI	Business Analyst	2	data_a
KION	Bangalore	Product Development	R	Business Solutions	Algorithms	Computer Science	Big Data	Data Analyst	Machine Learning	2	data_a

The following is the final data frame used for the analysis.

```
In [224]: df1.head()
```

Out[224]:

	Company_Name	Skill 1	Min Experience	Work_Profile
0	Dell	Artificial Intelligence	3	data_analyst
2	Dell	Azure	3	data_analyst
10	IBM	IT Skills	1	data_analyst
18	TEKION	IT Skills	2	data_analyst
28	TEKION	Product Development	2	data_analyst

```
In [225]: df1.to_csv("df1")
```

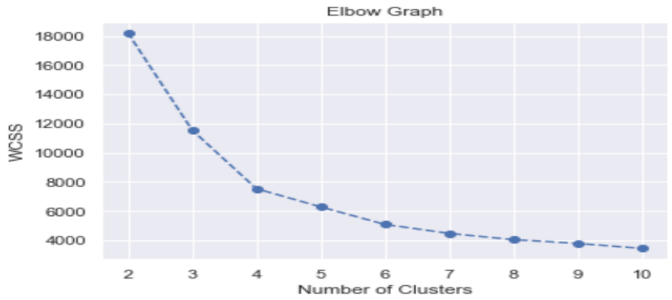
STEP 5: EXTRACTING SEGMENTS

Many segmentation methods can be used to extract job market segments using clustering algorithms, where cluster signifies the segments. While deciding the optimum algorithm for extracting the segments we need to consider Data Set characteristics like number of companies, number of segmentation variables, scale of segmentation variables etc and Segmentation characteristics like similarities in the segment and differences between the segments, number and size of segments etc.

In order to decide the optimum number of clusters we used the “elbow” method. As per elbow curve, the optimum number of clusters are 4.

Hence for our analysis the optimum number of segments is 4.

```
In [230]: import seaborn as sns
          sns.set()
          plt.plot(range(2, 11), WCSS,'o--')
          plt.title("Elbow Graph")
          plt.xlabel("Number of Clusters")
          plt.ylabel("WCSS")
          plt.show()
```



STEP 6: PROFILING SEGMENTS

Profiling is all about differentiating segments based on variables so that it can be observed how the groups are segmented based on segregating criteria. The goal of the profiling is to differentiate between groups with proper distinction so that each group has some unique characteristics and the cluster has an ideal solution.

For instance, what is the average min-experience of each segment, and then on comparing the segments, can understand one group has a high experience, another group has low experience. This way can see that the groups are getting differentiated on the basis of experience or experience is segregating the segments.

For profiling of the segments we have counted the nos of each cluster in the dataset.

```
In [149]: data_groupded=df1.groupby(by=["Clusters"]).mean().reset_index()
          data_groupded.rename(columns={"Clusters":"Applicant_Groups"},inplace=True)
          data_groupded
```

Out[149]:

	Applicant_Groups	Min Experience
0	0	3.038462
1	1	3.585714
2	2	2.333333
3	3	3.576271

```
In [150]: profilling=pd.merge(data_groupded,CustomerMagnitue_df, on="Applicant_Groups")
          profilling.round()
```

Out[150]:

	Applicant_Groups	Min Experience	Applicant Group Magnitude
0	0	3.0	26
1	1	4.0	70
2	2	2.0	45
3	3	4.0	59

Then we grouped the other segments as per the clusters (Applicant_groups) to find them in relation to other segments like “Skill 1” and “Min Experience” for each cluster.

As final step of profiling we merged the Applicant_magnitude data frame with data grouped dataframe with common key as Applicant_Groups (clusters)

```
In [143]: df1["Clusters"]=df1_cluster.labels_
          df1.head()
```

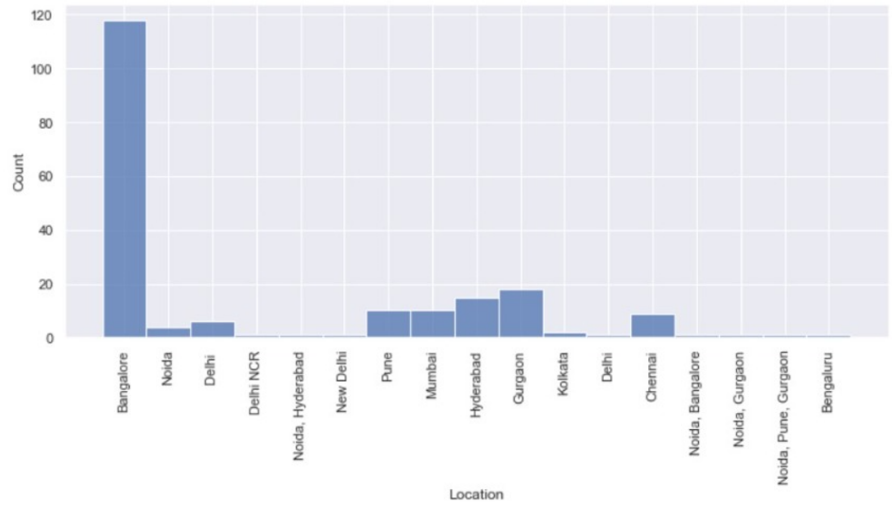
Out[143]:

	Company_Name	Skill 1	Min Experience	Work_Profile	Clusters
0	Dell	Artificial Intelligence	3	data_analyst	3
1	Dell	Azure	3	data_analyst	3
2	IBM	IT Skills	1	data_analyst	1
3	TEKION	IT Skills	2	data_analyst	1
4	TEKION	Product Development	2	data_analyst	2

STEP 7: DESCRIBING SEGMENTS

Now, we describe various segments formed. The descriptor variable used here is ‘Location’. Histogram is plotted to analyze company locations distribution for each segment.

```
plt.figure(figsize=(12,5))
sns.histplot(df_location['Location'])
degrees = 90
plt.xticks(rotation=degrees);
```



The Company Location distribution shows that most of the companies are situated in ‘Bangalore’. Other than Bangalore, ‘Pune’,’Mumbai’,’Hyderabad’,’Gurgaon’ and ‘Chennai’ have a significant number of companies.

```
clust_1 = pd.DataFrame()
clust_2 = pd.DataFrame()
clust_3 = pd.DataFrame()
clust_4 = pd.DataFrame()
```

```
df1.head()
```

	Company_Name	Skill 1	Min Experience	Work_Profile	Clusters
0	Dell	Artificial Intelligence	3	data_analyst	3
1	Dell	Azure	3	data_analyst	3
2	IBM	IT Skills	1	data_analyst	1
3	TEKION	IT Skills	2	data_analyst	1
4	TEKION	Product Development	2	data_analyst	2

```
df_location=pd.read_csv("df_location")
```

```
df_location.head()
```

	Unnamed: 0	Company_Name	Location	Skill 1	Skill 2	Skill 3	Skill 4	Skill 5	Skill 6	Skill 7	Skill 8	Min Experience	Work_Profile
0	0	Dell	Bangalore	Artificial Intelligence	Data Science	Data Analytics	IT Skills	Python	Testing	Machine Learning	RPA	3	data_analyst
1	2	Dell	Bangalore	Azure	Data Engineering	ECS	IT Skills	Data Science	Machine Learning	Cloud	MS SQL	3	data_analyst
2	10	IBM	Bangalore	IT Skills	Java	Python	Software Development	Data Science	Machine Learning	Cloud	Computer science	1	data_analyst
3	18	TEKION	Bangalore	IT Skills	Python	Testing	Data Science	Project Management	Tableau	Power BI	Business Analyst	2	data_analyst
4	28	TEKION	Bangalore	Product Development	R	Business Solutions	Algorithms	Computer Science	Big Data	Data Analyst	Machine Learning	2	data_analyst

```
df1["Location"]=df_location["Location"]
```

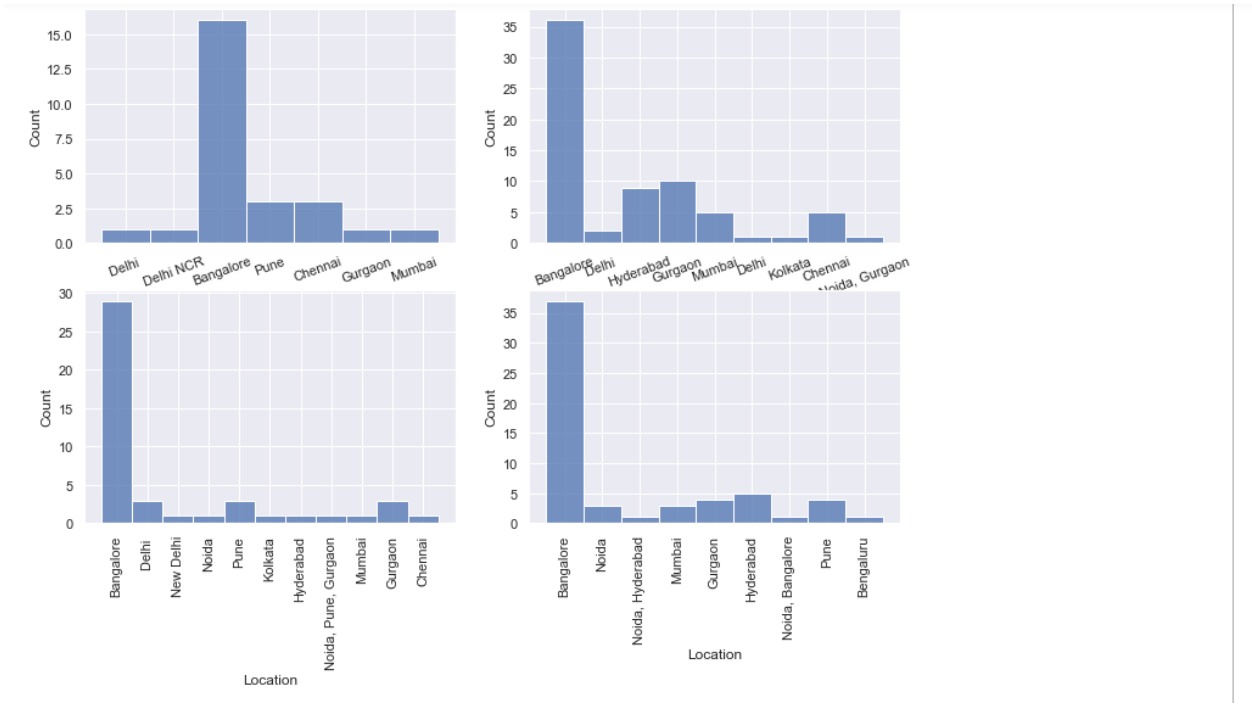
```
df1.head()
```

```
clust_1 = df1[df1['Clusters'] == 0]
clust_2 = df1[df1['Clusters'] == 1]
clust_3 = df1[df1['Clusters'] == 2]
clust_4 = df1[df1['Clusters'] == 3]
```

```
clust_1.head()
```

	Company_Name	Skill 1	Min Experience	Work_Profile	Clusters	Location
6	Huquo Consulting Pvt. Ltd	data analysis	3	machine_learning	0	Delhi
7	Huquo Consulting Pvt. Ltd	data analysis	3	machine_learning	0	Delhi NCR
20	Microsoft	Spark	5	data_science	0	Bangalore
32	Microsoft	deep learning	5	data_science	0	Bangalore
37	Gojek	Text mining	4	data_analyst	0	Bangalore

```
plt.figure(figsize=(12,12))

#plt.tight_layout(pad=6.0)
plt.subplot(3,2,1)
sns.histplot(clust_1['Location'])
degrees = 20
plt.xticks(rotation=degrees)
plt.subplot(3,2,2)
sns.histplot(clust_2['Location'])
degrees = 20
plt.xticks(rotation=degrees)
plt.subplot(3,2,3)
sns.histplot(clust_3['Location'])
degrees = 90
plt.xticks(rotation=degrees)
plt.subplot(3,2,4)
sns.histplot(clust_4['Location'])

degrees = 90
plt.xticks(rotation=degrees);
```



From these histograms, company locations distribution for each segment can be visualized. It is observed that the 1st segment has most of its companies situated in Bangalore, Pune and Chennai.In the 2nd segment most of the companies are situated in Bangalore,Gurgaon,Hyderabad,Mumbai and Chennai. In the 3rd segment most of the companies are situated in Bangalore,Pune,Gurgaon and Delhi. The 4th segment has most of its companies situated in Bangalore,Hyderabad,Mumbai,Gurgaon,Noida and Pune.

STEP 8: SELECTING (THE) TARGET SEGMENT(S)

After profiling and describing the segments, now we select the target segment.

We choose the target segment considering a job applicant/candidate as our potential stakeholder.

```
df2 = pd.read_csv("Segments_Evaluation_Data.csv")
```

```
df2
```

	Customer_Groups	Average_Minimum_Experience	Rating(Value)_Min_Experience	Segment_Attractiveness	Percentage_Correspondance_of_available_work_profile
0	0	3.04	8	8	
1	1	3.58	7	7	
2	2	2.33	3	3	
3	3	3.57	7	7	

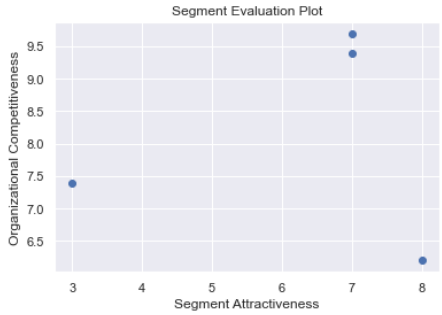
The segment attractiveness criterion used here is ‘Min Experience’ and the criterion used for evaluating Organizational Competitiveness is ‘Work Profile’. As both,segment attractiveness and Organizational Competitiveness, have only one criterion, these criteria get a weightage of 100%.For segment attractiveness we compare the average minimum experience for each segment with average minimum experience calculated considering the entire dataset. While considering organizational competitiveness we take only 5 four work profiles - machine_learning,data_analyst,data_science,applied_intellegence and analytics. We consider percentage correspondence of the work profiles distribution for a segment with respect to work profiles distribution for the entire dataset.

The segments which have their average minimum experience required closer to the average of the entire dataset are given higher rating for segment attractiveness and segments which have higher percentage correspondence of work profiles when compared with that of the entire dataset are given a higher rating for organizational competitiveness.

Finally, using this data we choose that segment which has higher segment attractiveness and higher organizational competitiveness.

```
x1 = np.asarray(df2['Segment_Attractiveness']).astype(np.float32)
y1 = np.asarray(df2['Organizational_Competitiveness']).astype(np.float32)
plt.scatter(x1,y1)
plt.xlabel('Segment Attractiveness')
plt.ylabel('Organizational Competitiveness')
plt.title('Segment Evaluation Plot')
```

```
Text(0.5, 1.0, 'Segment Evaluation Plot')
```



As observed from the above scatter plot, our target segment is the 4th segment i.e the one with segment attractiveness score of 7 and organizational competitiveness score of 9.7. But, the 2nd segment i.e. the one with segment attractiveness score of 7 and organizational competitiveness score of 9.4 is close enough as well.

STEP 9: CUSTOMIZING THE MARKET MIX

➔ PLACE:
Place in the job marketing mix is the location where your job or service is actually located. Develop the habit of reviewing and reflecting on the exact location where the job meets the skills. Sometimes changes in place can lead to a rapid change in work life balance.

➔ PRODUCT:
Here skills are considered as ‘product’. This segment has a considerable market for freshers. The most wanted skills are having knowledge about programming languages, Machine Learning, AI, Big Data Tools, BI Tools, Cloud Platforms. Hence the applicants could concentrate on strengthening their knowledge in the above mentioned domains.

➔ PROMOTION:
In this case, Promotion is nothing but the Years of Experience. More years of experience is responsible to Salary package. It may open a number of jobs with excellent skills.

➔ PRICE:
Salary is considered as the Price here. With the number of years of experience and excellent skills, Price/Salary will get decided accordingly. In this era of the world, Skills are more important than anything else to get a job.

STEP 10: EVALUATION & MONITORING

The Objective of Evaluating the effectiveness of the Job Market Segmentation Strategy is to determine whether to develop a customized marketing mix for one or more segments that yields the expected benefits for the Organization.

Evaluation of Job Market For ML Engineer and Data Analyst.

We have to consider three factors while Evaluating different JobMarket Segments.

- **Segment Size & Growth**

  Size of the job market segment must be right for that market segment to be large and it must include characteristics like skills, years of experience, etc of the applicant. The size of the job market should be consistent with the area of interest of the applicant and work profile of the organization..

- **Segment Structural Attractiveness**

  Every candidate needs to evaluate the requirements of the job market. Job market can be more attractive or less attractive. If the Job market is large it will be considered as more attractive when not much larger it will be less attractive.

- **Company Objective & Resources**

  After completion of job market analysis and all strategic and tactical marketing activity have been initiated, the success of job market strategy should be evaluated and the job market should be carefully monitored on a continuous basis.

III. REPLICATION OF CASE STUDY IN PYTHON

https://github.com/vishakhawarudkar96/Feynn_Labs_Final_Project.git