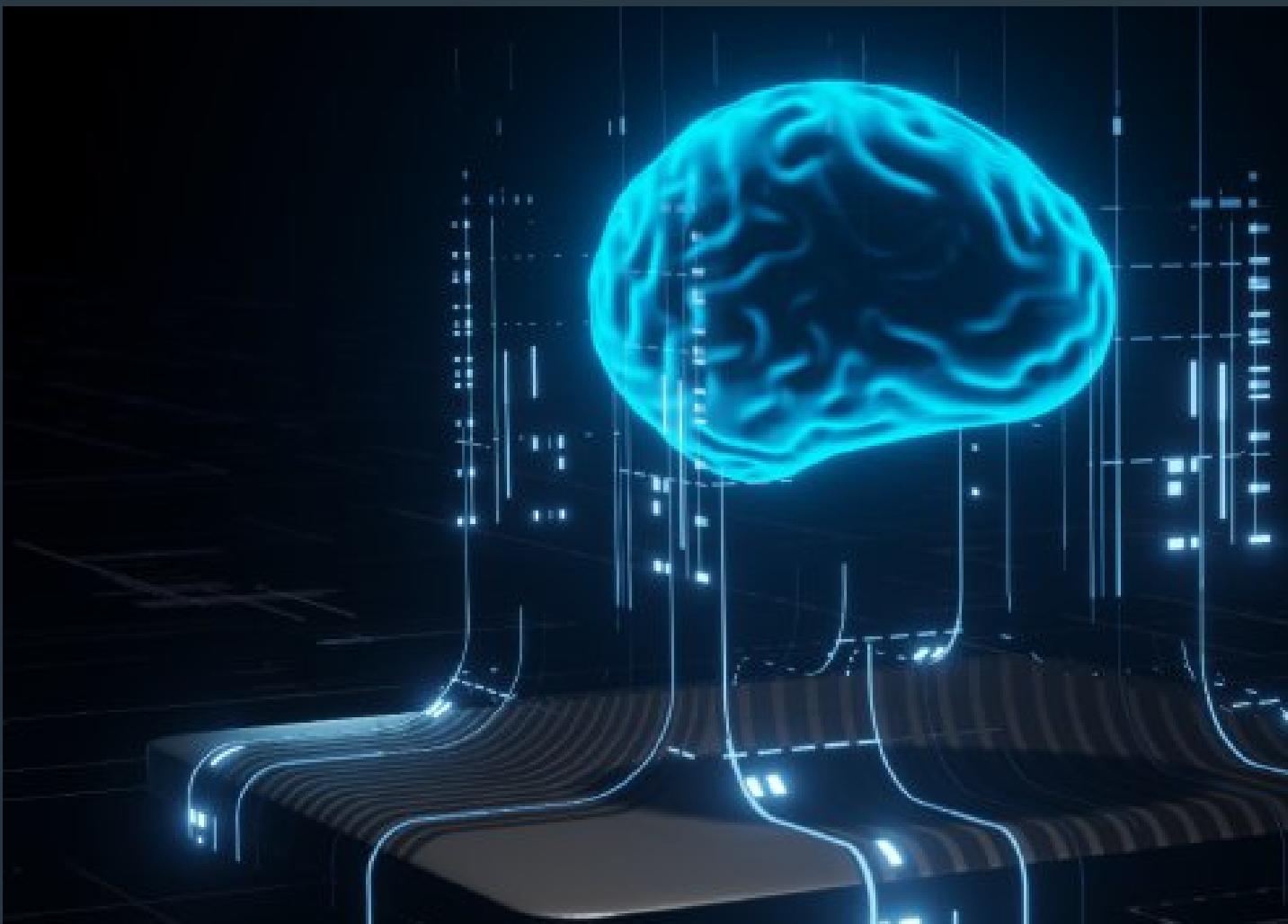MARCH 2022

SEGMENTATION AND ANALYSIS OF ARTIFICIAL INTELLIGENCE JOB MARKET IN INDIA

FeyNN Labs
EXPERIMENT WITH YOUR KNOWLEDGE

Submitted by:

Vaasu Bisht

Nithin Sylesh

Reedham Patel

Nihar Ranjan Sahoo

CONTENT

Abstract

Segmentation is when you isolate potential target audiences to determine which ones will bring the highest return on your marketing efforts. Segmentation is based on specific criteria related to an individual's age, income, subjects of interest and behaviors. When you segment various markets, you learn more about their core values and what will ultimately attract them to your brand.the difference between conventional market segmentation methods and the jobs-based segmentation approach is the primary unit of analysis for grouping customers. For conventional segmentation, the primary unit of analysis is the attributes of customers' themselves. The primary unit of analysis for jobs-based segmentation is a job that customers are trying to get done.

The conventional definition of a market is based on the product and service categories defined by solution providers. Jobs Theory, on the other hand, defines a market as an aggregation of all available solutions, both provider and non-provider, that customers regard as being able to satisfy their needs with respect to getting a job done. Job segmentation gives a company a significant advantage over competitors because they can anticipate the value that customers want—even before customers are aware of certain needs. A company can quickly and efficiently enhance their existing offerings and create new offerings that can satisfy customer needs better than competitive alternatives at the lowest possible cost.

Introduction

The scope of Artificial Intelligence in India is promising. Artificial Intelligence has immense potential to change each sector of the economy for the benefit of society. There is not just one technology under AI, but there are various useful technologies such as self-improving algorithms, machine learning, big data, pattern recognition. Soon, there would hardly be any industry or sector which would be untouched by this powerful tool in India. This is the reason why there has been an increasing demand for Artificial Intelligence online courses in India.

The scope of Machine Learning in India, as well as in other parts of the world, is high in comparison to other career fields when it comes to job opportunities. According to Gartner, there will be 2.3 million jobs in the field of Artificial Intelligence and Machine Learning by 2022. Also, the salary of a Machine Learning Engineer is much higher than the salaries offered to other job profiles.

According to Forbes, the average salary of a Machine Learning Engineer in the United States is US$99,007. In India, it is US$865,257. Let us look at the graph of top job profiles listed by Indeed. There are certain skills that you need to master for becoming a successful Machine Learning Engineer and they are:

● Programming: Programming is one of the important aspects for any Machine Learning enthusiast. For Machine Learning, we generally use R and Python languages. We can learn both. However, the scope of Machine Learning with Python is high.

● Understanding of data structures: The data structure is the core of any software. Thus, it is recommended to have a good grasp of the concepts of data structure.

● Mathematics: We cannot perform computation without mathematics. Therefore, we should have knowledge of applying mathematical concepts into Machine Learning models. These concepts include calculus, linear algebra, statistics, and probability.

● Software engineering: Machine Learning models are built to integrate with the software. Thus, an ML Engineer should have a thorough knowledge of software engineering.

● Data mining and visualization: As we built Machine Learning models on top of various data, it becomes essential to understand the data. For this, a Machine Learning enthusiast must have experience in data visualization and mining.

● Machine Learning algorithms: Along with all these, most importantly, we should have experience in implementing various ML algorithms

Problem Statement:

AI is the ability of a machine to display human-like capabilities such as reasoning, learning, planning, and creativity. AI enables technical systems to perceive their environment, deal with what they perceive, solve problems and act to achieve a specific goal. The computer receives data - already prepared or gathered through its own sensors such as a camera - processes it and responds. Artificial intelligence is being used in hundreds of ways all around us.

It has changed our world and made our lives more convenient and interesting. Some of the many uses of AI you may know include Voice recognition, Self-driving cars, Online shopping, Streaming services, Chatbots, etc. AI is influencing many industries situated across the GLOBE. Even Forbes has predicted in one of its reports that the growth rate of its job roles is more than 70 percent.
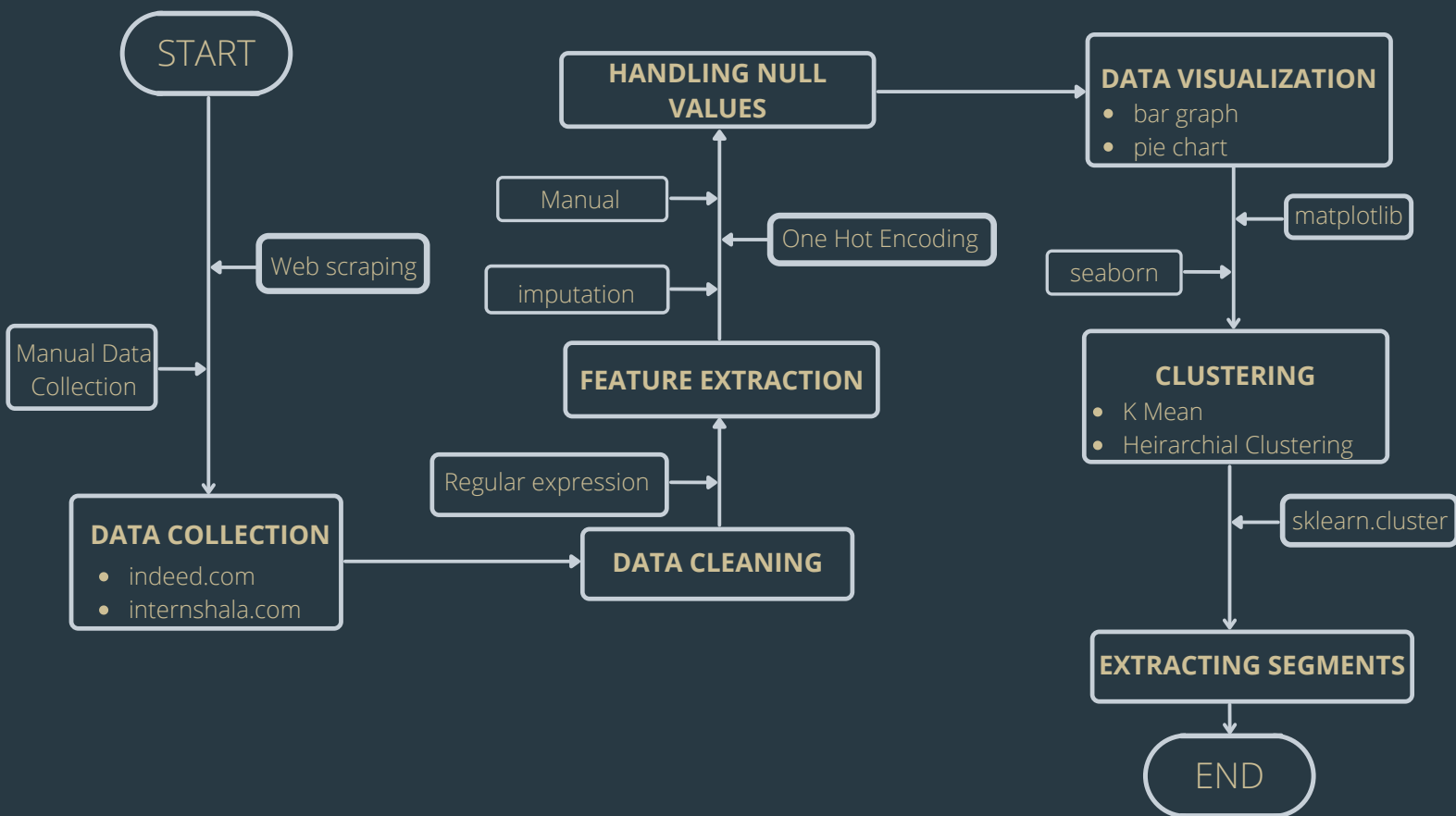
Making a job market segmentation report, which will help aspirants finding fompanies most probable to up come with an AI job an hire an ML Engineer/Data Sciencist etc Applicant in respect to his/her skillset.

=>Data Collection/Scraping based on
1. Geography,
2. Company's field of work,
3. Company size,
4. Upcoming vacancies in respect to company's growth (IPO/Funding etc.)
5. Machine Learning/Data Analysis Skills currently most demanded in the market in respect to:-

  i) Experience required,
  ii) Time required to acquire the skill,
  iii) Vacancies open
 iv) Salary etc. (Interns have the choice to add more segments)

Work Flow:

START → Web scraping

Manual Data Collection →

DATA COLLECTION
- indeed.com
- internshala.com

→ DATA CLEANING

Regular expression → FEATURE EXTRACTION

Manual → HANDLING NULL VALUES

imputation →

One Hot Encoding →

HANDLING NULL VALUES → DATA VISUALIZATION
- bar graph
- pie chart

matplotlib →

seaborn →

DATA VISUALIZATION → CLUSTERING
- K Mean
- Heirarchial Clustering

sklearn.cluster →

CLUSTERING → EXTRACTING SEGMENTS → END

Data Collection:

Data collection is defined as the procedure of collecting, measuring and analyzing accurate insights for research using standard validated techniques. A researcher can evaluate their hypothesis on the basis of collected data. In most cases, data collection is the primary and most important step for research, irrespective of the field of research. The approach of data collection is different for different fields of study, depending on the required information.

To begin with, collecting data we should signify that empirical data forms the basis of both commonsense and data-driven job segmentation as it provides a clear base to work on. This data enables us to create job segments as well as give a grand picture and explanation of these segments. Although data is a valuable asset for every organization, it does not serve any purpose until analyzed or processed to get the desired results
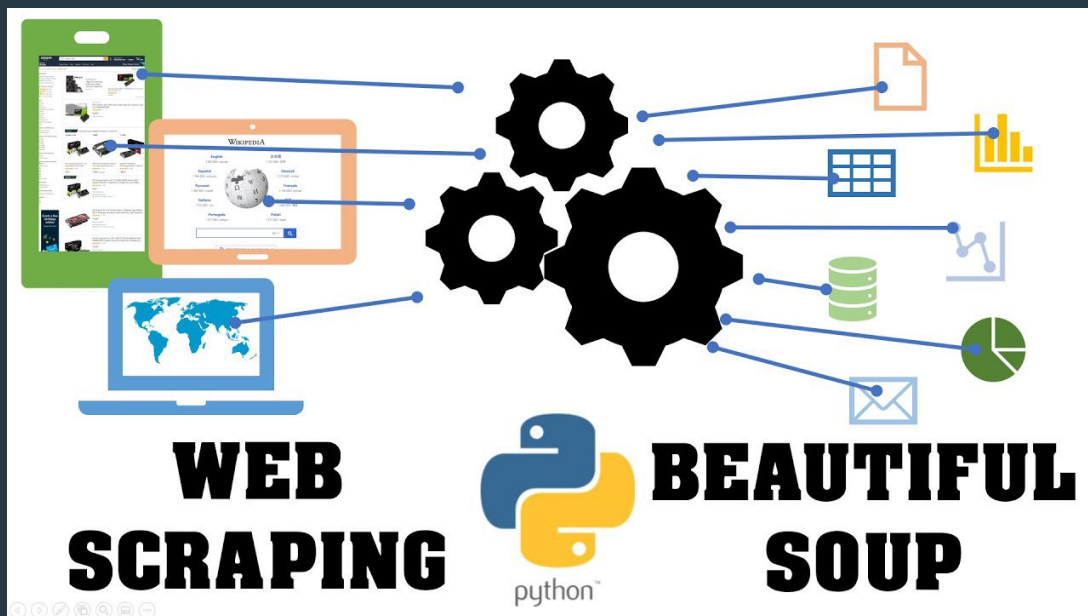
This report is focused on analyzing Artificial Intelligence job in Indian market and draw valuable insights, the data is extracted from mainly two leading employment websites for the year 2022 :
1)Indeed.com
2)Intershala.com

We used webscraping techniques like Beautiful Soup and Auto Scraperto extract features like Company name, skills requirements, location, Job Title, Description, etc from the above mentioned sites. We also did a bit of manual data collection and added many features in our dataset.

RAW DATASET:-
https://github.com/vaasu2002/Feynn-Labs/blob/main/Job%20Segmentation/DATASET/Indeed.csv

https://github.com/vaasu2002/Feynn-Labs/blob/main/Job%20Segmentation/DATASET/internshalla_final.csv

FINAL DATASETS:-
https://github.com/vaasu2002/Feynn-Labs/blob/main/Job%20Segmentation/Final_data.csv

Data Pre-processing

DESCRIPTION OF WORK

PHASE 1
Data Cleaning

PHASE 2
Feature Extraction

PHASE 3
Handling Null Values

When it comes to creating a ML model, data preprocessing is the first step marking the initilization of the process. We didn't use any dataset available on internet but webscrapped the dataset from two job protals. The problem with such real life dataset is that it is incomplete, inconsistent and often lacks some attributes which we then tried to add manually. Preprocessing helped us to clean the the format, organize the raw data there-bu making it ready-to-go for segmentation task.

DATA CLEANING -

We used numpy and pandas library to perform data cleaning. We removed useless features like data posted etc from the dataset and removed many inconsistency in feature columns like Job Role and Location. In Location feature Some jobs requirements had multiple vacancies in multiple cities so we duplicated those ,thus increasing the data, data for better segmentation. We segricated city and state for better segmentation.

```python
pattern_data_science = ['Data Scientist','Data Science']

for x in pattern_data_science:
  for i in range(len(l)):
    m = re.findall(x, l[i])
    if(len(m)!=0):
      l[i] = pattern_data_science[0]
```

A regular expression is a pattern that the regular expression engine attempts to match in the input text. A pattern consists of one or more character literals, operators, or constructs. In the Job Title feature, we used Regular Expression for the key job roles. This was done to add consistency. Other features present in the job title were not ignored but were added in new columns. Eg - Data Scientist (AWS specialization) job title was changed to Data Scientist and we added a new column that said AWS needed.

FEATURE EXTRACTION:-

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. We extracted many new features like Cloud Requirement, Linux OS Requirement, R or Python Requirements, and SQL database requirements. It yields better results than applying machine learning directly to the raw data.

One hot encoding is a process of converting categorical data variables so they can be provided to machine learning algorithms to improve segmentation. One hot encoding is a crucial part of feature engineering (extraction) for machine learning. One hot encoding is useful for data that has no relationship to each other. Machine learning algorithms treat the order of numbers as an attribute of significance. It provides more nuanced predictions than single labels. While this is helpful for some ordinal situations, some input data does not have any ranking for category values, and this can lead to issues with predictions and poor performance. That's when one hot encoding saves the day. One hot encoding makes our training data more useful and expressive, and it can be rescaled easily. By using numeric values, we more easily determine a probability for our values.

Missing values occur because of various factors like missing completely at random, missing at random or missing not at random. All these may result from system malfunction during data collection or human error during data pre-processing. Nevertheless, it is important to deal with missing values before analysing data since ignoring or omitting missing values may result in biased or misinformed analysis or wrong segmentations.

One way to handle this problem is to get rid of the observations that have missing data. A better strategy would be to impute the missing values. In other words, we need to infer those missing values from the existing part of the data. There are three main types of missing data:

- Missing completely at random (MCAR)
- Not missing at random (NMAR)
- Missing at random (MAR)

Imputation Using (Mode) Values:
This works by finding the Mode (as we are dealing with categorical data only in our segmentation task) of the non-missing values in a column and then replacing the missing values within each column separately and independently from the others. It can only be used with numeric data.

Imputation Using KNN:
The k nearest neighbours is an algorithm that uses 'feature similarity' to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. This can be very useful in making predictions about the missing values by finding the k's closest neighbours to the observation with missing data and then imputing them based on the non-missing values in the neighbourhood.

KNN with 30 neighbors
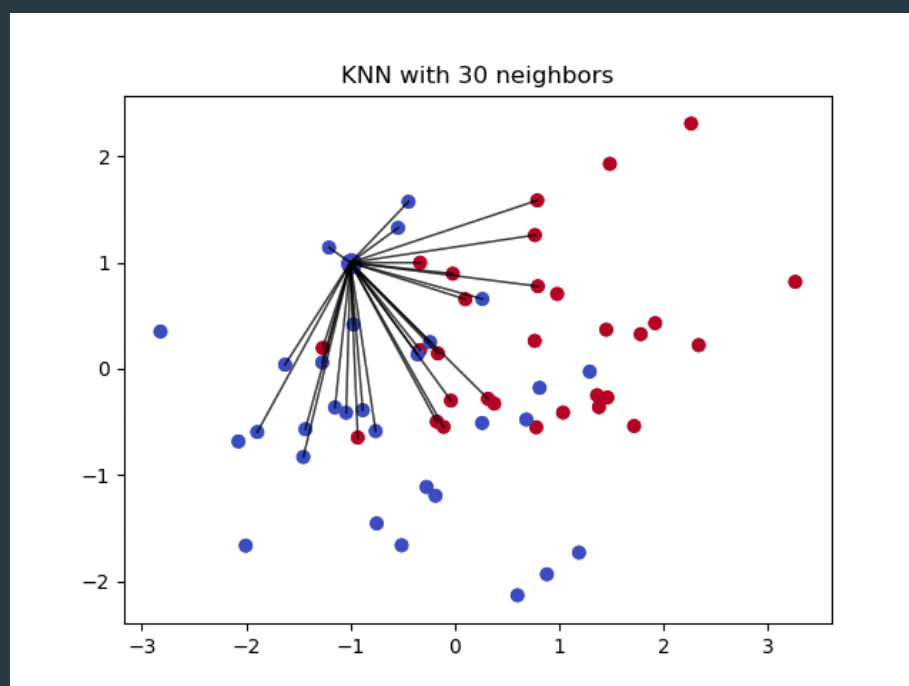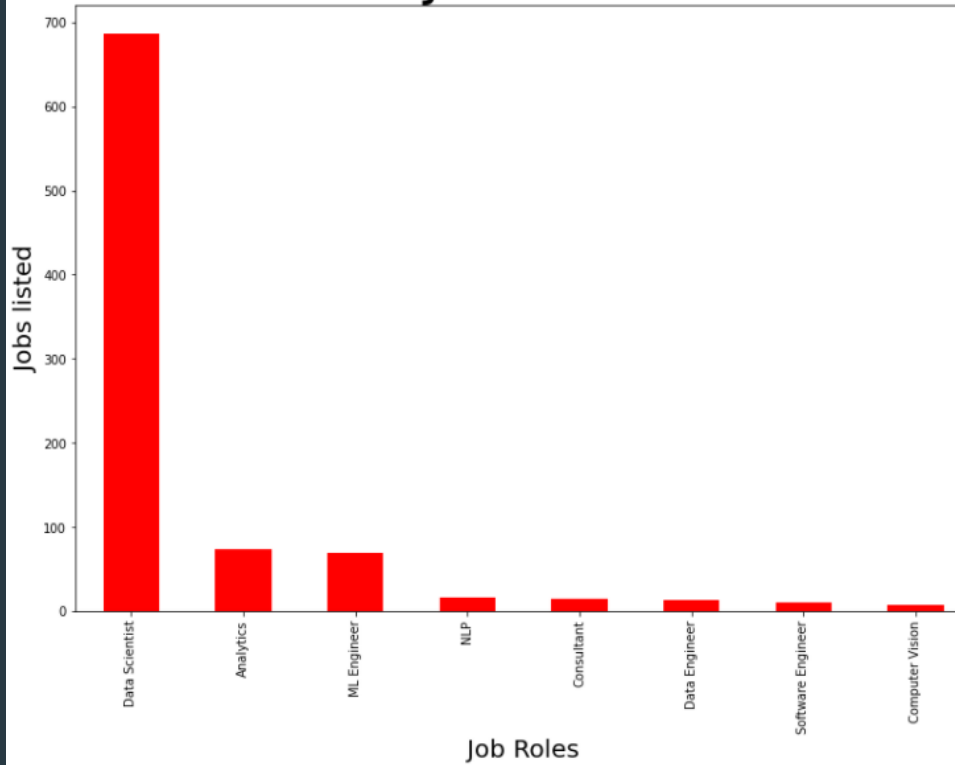
Analysis and Visualization
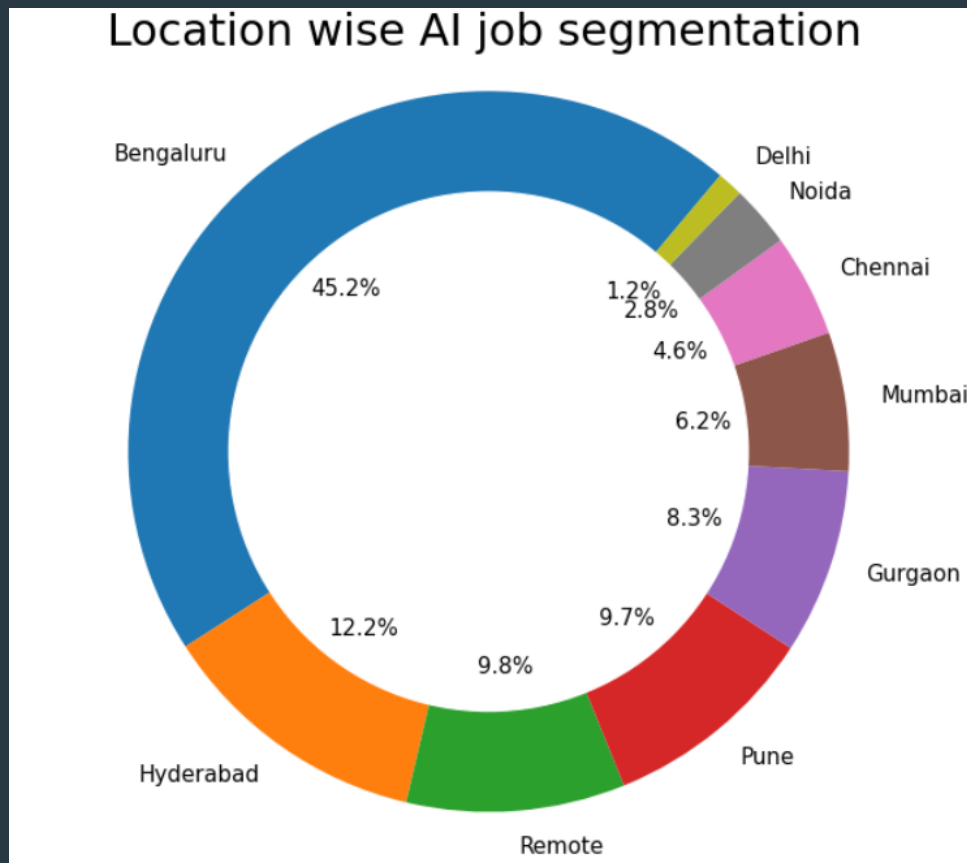
There are thousands of ML jobs available currently but not every place provides equal opportunity. Currently Bangalore (also known as Asia's silicon valley) seems to be at a boom providing a significant 45.2% of the jobs available in the market. Hyderabad, Pune and Delhi NCR (most notably Gurgaon) are the next major contributors in the market holding share of 12.2%, 9.7% and 8.3% respectively. The covid situations are getting better but as work from home culture and flexibility in the work has gained acceptance, Remote jobs accounts for 9.8% of share in the industry. Other places like Mumbai and Chennai are also contributing to the market.

```
df.Location.apply(pd.Series).stack().str.strip().value_counts()[:9].plot.pie(figsize=(12,10),startangle=50,autopct='%1.1f%%',fontsize=15)
plt.title("Location wise AI job segmentation",fontsize=30)

centre_circle = plt.Circle((0,0),0.72,color='white',fc='white',linewidth=1.25)
fig = plt.gcf()
fig.gca().add_artist(centre_circle)
plt.axis('equal')
plt.show()
```

AI job Roles

Jobs listed vs Job Roles

Data Scientist, Analytics, ML Engineer, NLP, Consultant, Data Engineer, Software Engineer, Computer Vision

Companies with most job offer

- GO-JEK — 18.0%
- Airtel India — 6.2%
- Novartis — 6.2%
- Standard Chartered — 7.8%
- ADCI - Karnataka — 7.8%
- Sutherland — 8.6%
- JPMorgan Chase Bank, N.A. — 9.4%
- Microsoft — 10.2%
- Accenture — 10.9%
- Deloitte — 14.8%
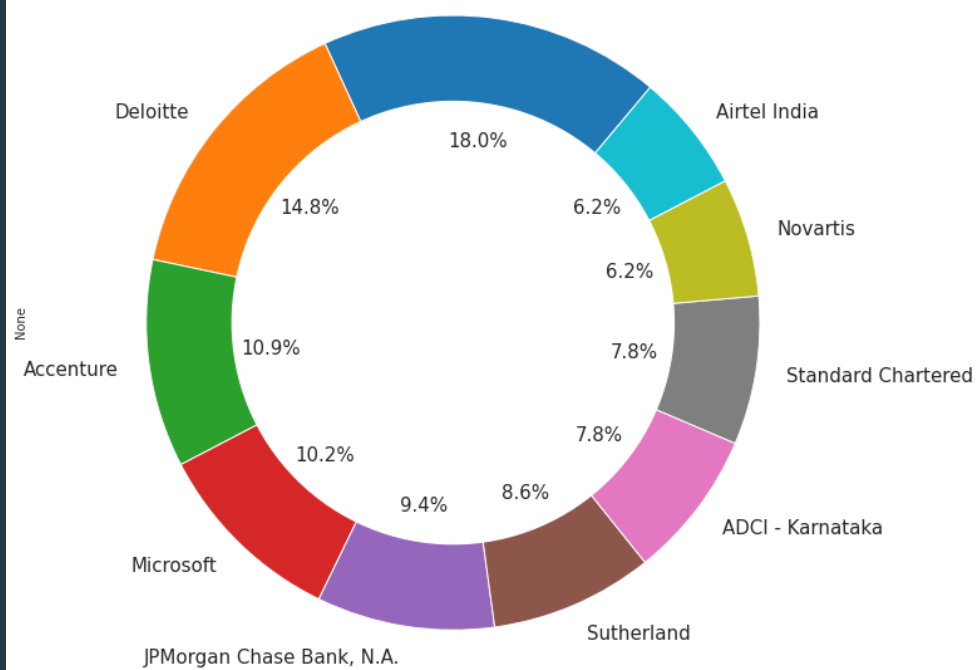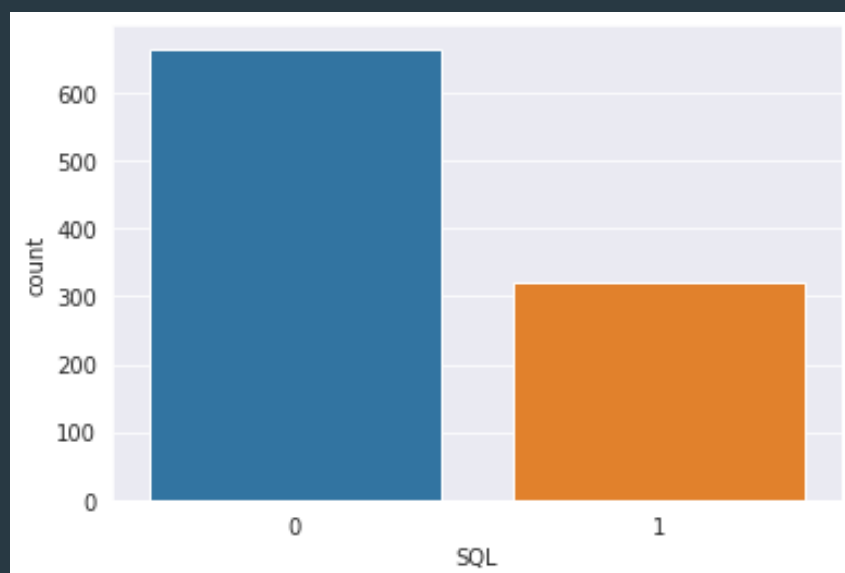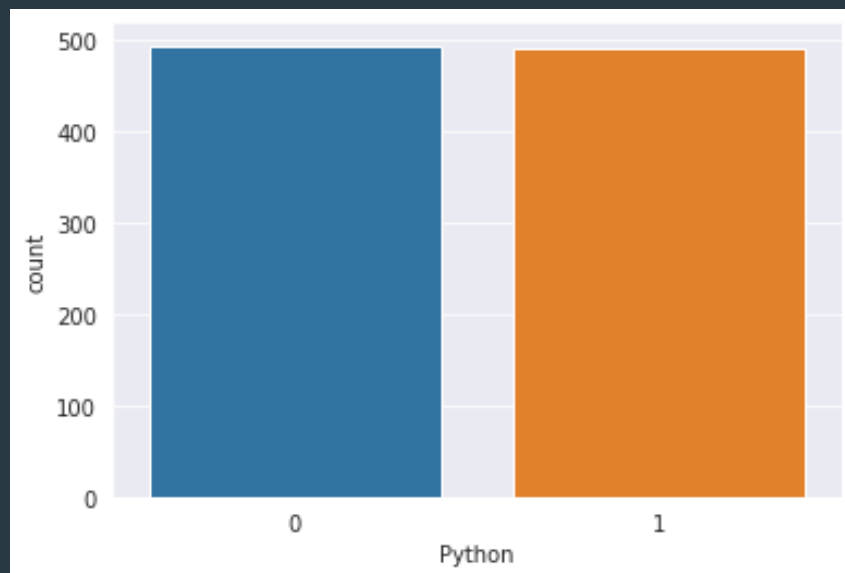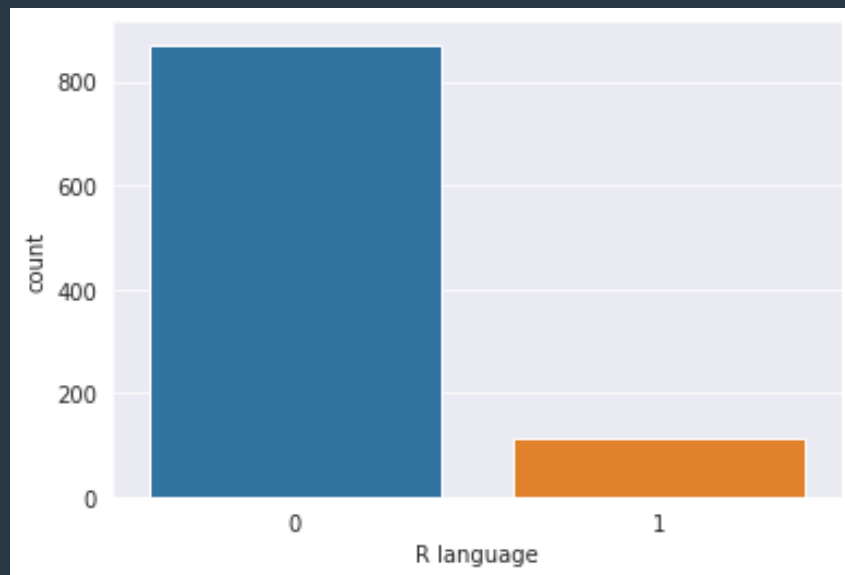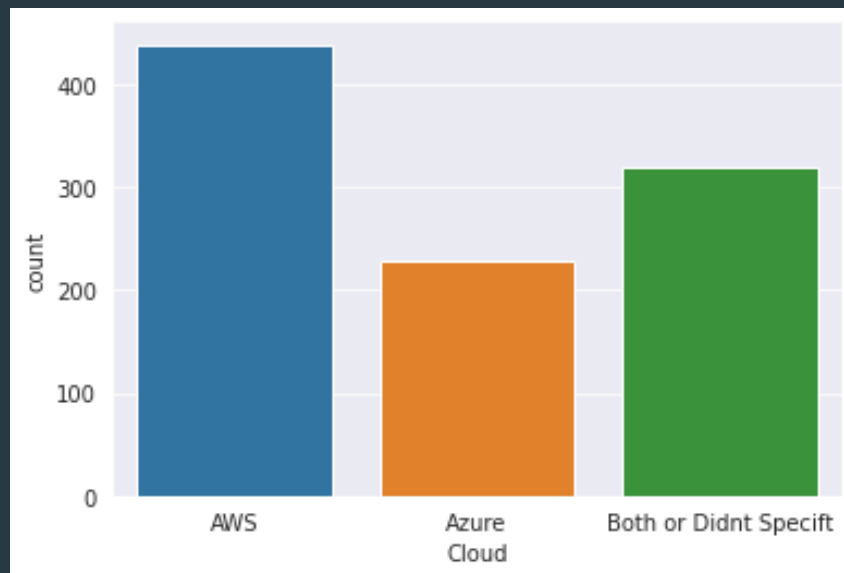
From the analysis, we concluded that most recruiters want proficiency in the Python programming language. Also having experience in handling cloud services like AWS and Azurealso is really helpful. AWS and Azure have been relentless and taken the top honors for a while now. Thus having AWS Certification or Azure Certification which hold immense value in the current Cloud market from a career perspective. Now we did a little analysis between AWS and Azure. Both Azure and AWS Pricing models offer pay as you go structure. AWS charges on hourly basis whereas Azure charges on per minute basis. When it comes to short-term subscription plans, Azure gives you a lot more flexibility. If we are looking for durability AWS has Amazon RDS whereas Azure has Azure SQL Server Database. Amazon RDS supports different database engines like MariaDB, Amazon Aurora, MySQL, Microsoft SQL, PostgreSQL, and Oracle whereas when it comes to Azure, SQL Server Database is based on SQL as the name suggests. Both AWS and Azure provide long-running and reliable storage services. AWS has services like AWS S3, EBS, and Glacier whereas Azure Storage Services have Blob Storage, Disk Storage, and Standard Archive. With Azure, it uses temporary storage and page blobs for VM volume. Azure has a Block Storage option as a counterpart to S3 in AWS. In addition, Azure also provides two types in their storage, cold and hot storage.

SQL and other database management skills also come handy and are preferred by recruiters .
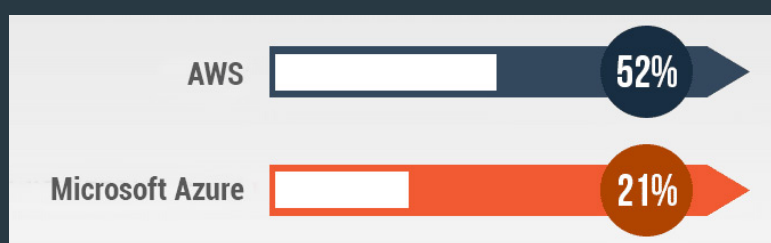
Segment Extraction

Clustering is an unsupervised machine learning technique that divides the population or data points into several groups or clusters such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups.

Distance-based method

In order to find groups of similar consumers, one needs a notion of similarity or dissimilarity, mathematically speaking: a distance measure.
A distance measure has to comply with a few criteria. One criterion is symmetry, that is:

$$d(x, y) = d(y, x)$$

A second criterion is that the distance of a vector to itself and only to itself is 0:

$$d(x, y) = 0 \Leftrightarrow x = y.$$

In addition, most distance measures fulfill the so-called triangle inequality:

$$d(x, z) \leqslant d(x, y) + d(y, z)$$

The triangle inequality says that if one goes from x to z with an intermediate stop in y, the combined distance is at least as long as going from x to z directly. The most common distance measure used in market segmentation analysis is Euclidean distance.

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

Manhattan or absolute distance:

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$

Improved k-Means

Instead of randomly initializing starting points i.e. - segment representative, we would initialize with starting points that are evenly spread over whole data and they will better represent the entire data otherwise if points are consumers being located very close to each other, the solution will be stuck with the problem of local-optimum. Good representatives are close to their segment members; the total distance of all segment members to their representatives is small.

```python
from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=5, random_state=42,algorithm='elkan',init='k-means++').fit(X)

df1['Clusters'] = kmeans.labels_

df2 = pd.read_csv('Final_data.csv')

df2 = df2[['Company Name','Location.1','JOB','Python','SQL','R language','Linux','Cloud','0','Revenue','No of employee']]
df2['Tablue'] = df2['0']
df2.drop('0',axis = 1,inplace = True)

df2['Clusters'] = kmeans.labels_

df2.head()
```

	Company Name	Location.1	JOB	Python	SQL	R language	Linux	Cloud	Revenue	No of employee	Tablue	Clusters
0	Bosch Group	Bengaluru	Data Scientist	0	0	0	0	AWS	77.721	31000	0	2
1	Standard Chartered	Chennai	Data Scientist	1	0	0	1	AWS	14.701	85000	0	4
2	Kaplan	Bengaluru	Data Scientist	1	0	0	0	AWS	1.3	12000	0	2
3	Corteva Agriscience	Hyderabad	Data Scientist	1	0	0	1	AWS	14.271	21000	0	2
4	UnitedHealth Group	Bengaluru	Data Scientist	1	0	0	0	AWS	285.1	300000	0	3

```python
Cluster1 = df2[df2['Clusters'] == 0]
Cluster2 = df2[df2['Clusters'] == 1]
Cluster3 = df2[df2['Clusters'] == 2]
Cluster4 = df2[df2['Clusters'] == 3]
Cluster5 = df2[df2['Clusters'] == 4]
```

Profiling and describing potential segments

Using k-means unsupervised machine learning algorithm we made 8 clusters from our dataset. The fifth cluster is the biggest.
Cluster -
https://github.com/vaasu2002/Feynn-Labs/blob/main/Job%20Segmentation/Cluster%20Tables.pdf

CLUSTER 1


```
] Cluster1
```

	Company Name	Location.1	JOB	Python	SQL	R language	Linux	Cloud	Revenue	No of employee	Tablue	Clusters
29	Unilever	Bengaluru	Data Scientist	1	0	0	0	AWS	52.444	149000	0	0
35	Oracle	India	Data Scientist	0	0	0	0	AWS	40.47	132000	0	0
38	Unilever	Bengaluru	Data Scientist	1	0	0	0	AWS	52.444	149000	0	0
53	Intel	Bengaluru	Software Engineer	0	0	0	0	AWS	79.02	121100	0	0
77	Apple	Bengaluru	ML Engineer	0	1	1	1	Azure	365.82	1,54,000	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...
919	Unilever	Bengaluru	Data Scientist	1	1	0	0	AWS	52.444	149000	0	0
942	Oracle	Bengaluru	Data Scientist	1	0	0	0	AWS	40.47	132000	0	0
968	ZF	Hyderabad	ML Engineer	1	0	0	0	Azure	41.45	147797	0	0
972	GE Corporate	Bengaluru	ML Engineer	1	0	0	0	Azure	74.196	168000	0	0
976	GE Corporate	Bengaluru	ML Engineer	0	0	0	1	Azure	74.196	168000	0	0

71 rows × 12 columns

	Company Name	Location	Job Title	Skills
CLUSTER- 1	Bosch Group of pvt. Ltd. Corteva Agriscience Ltd. Orange Business Services Expleo Technology pvt. Ltd. Smith's Detection Ltd. DBS bank Master card Qualfon pvt. Ltd. Sutherland group enterprises Visa Group HDFC Life Bank Philips India Ltd. Uber Technology Inc. Signify Innovations India Ltd. Ford Global Business Services Lam Research Boston Consulting Group The Brose Group Maruti Suzuki India Ltd Amazon Dev Centre India - Hyd HGS India Limited S&P Global pvt. Ltd. Assurant India Clariant Chemicals India Ltd.	Bengaluru Hyderabad Pune Chennai Mumbai Gurgaon	Data Scientist Analytics Associate Data Scientist ML engineer S/w Engineer Sr Data Scientist Manager	Python Machine Learning SQl R language Linux AWS, Azure

CLUSTER 2

CLUSTER- 2	Micron Technology Syneos Health Clinical Bayer AG PayPal India Ltd. Goldman Sachs India Merkle Inc. Vodafone Idea pvt. Ltd Yara International Hella India Lighting Ltd. VMware Inc. Unisys Corporation TARGET Corporation General Mills Eli Lilly Company Thomson Reuters Carlson Wagonlit Travel	Airoli Maharashtra Ahmedabad Tamil Nadu New Delhi Telangana	Data Engineer ML Scientist Jr. Data Scientist Big data Engineer Product Management AI Engineer Data Science Team Lead	Python Deep Learning SQl R language Linux AWS, Azure

CLUSTER 3

	Company Name	Location	Job Title	Skills
CLUSTER- 3	JPMorgan Chase Bank, N.A. IBM Infosys Limited AT&T The Walt Disney Company (Corporate) HSBC KPMG	Bengaluru Mumbai Pune Ahmedabad Kochi Kolkata Hyderabad	Data Scientist Analytics ML and Python Developer ML Engineer AI Engineer Technical Architect Data Science Consultant	Python Machine Learning Deep Learning AWS and Azure

CLUSTER 4

	Company Name	Location	Job Title	Skills
CLUSTER- 4	Standard Chartered pvt. Ltd Maersk Pvt. Ltd. Kyndryl Company Honeywell Intel Pvt. Ltd. SAP s/w company Shell India Vodafone India Ltd. thyssenkrupp India Pvt. Ltd. GSK India The Michelin Group Medtronic Pvt. Ltd. Novartis Ltd. Couche-Tard ABB Group HDFC Bank India Emerson Electric Genpact India Pvt. Ltd. Caterpillar Inc. EATON Corporation	Chennai Pune Hyderabad Bengaluru Gurgaon Telangana Nanakramguda Mumbai Noida Hadapsar	Data Scientist Analytics s/w Engineer CV Engineer ML Engineer ADV. AI Engineer Manager predictive Modelling Data Science Consultant Data Engineer	Python Machine Learning Deep Learning Linux SQL AWS and Azure

CLUSTER 6

|	Company Name	Location	Job Title	Skills
CLUSTER-6	HP Pvt. Ltd. Ecolab Inc. Hewlett Packard Enterprise Jacobs Eng. Group Cummins Inc. CGI Inc. BNY Mellon Corporation NatWest Group AmEx GE Aviation ExxonMobil Corporation Morgan Stanley Pvt. Ltd. Alstom Company Netradyne Tech. India Ltd. Atlassian Corporation IHS Markit Ltd. Klevu Oy India Pvt. Ltd. Airtel India Ltd. Blackbuck Sol. Pvt. Ltd. IQVIA Inc. Pfizer Inc. Fidelity Investments Ltd.	Bengaluru Navi Mumbai Pune Delhi Gurgaon Mumbai Remote in Gurgaon Remote in Bengaluru India Bellandur Sholinganallur Chennai Hyderabad	Data Scientist Analytics s/w Engineer ML Engineer AI Scientist HR Business Analyst Data Engineer Sr. Statistical Programmer Data Integration Specialist	Python Machine Learning Deep Learning AWS and Azure

CLUSTER 5

	Company Name	Location	Job Title	Skills
CLUSTER- 5	Kaplan Inc. 3SC Solutions Jumio Corporation Microland Tech. Company Onward Group Indium Software Narayana Health citius tech AIBridge ML Pvt. Ltd. Confluent India Pvt. Ltd. Global Industrial Solutions Dunnhumby India Coinbase Global Inc. Umlaut AG Company Fractal Analytics Tredence Inc. Casepoint Pvt. Ltd. Cliksource Corporation CityLink Technology Solutions Sofar Ocean Tech. Walmart Global Tech India Oportun Inc Crowe LLP Cygnet Infotech Vistex Asia-Pacific Private Ltd. Beinex Pvt. Ltd. Kwalee Pvt. Ltd. MABZONE IT SOLUTIONS Paktolus Solution LLC The Luxury Closet LLC RedSeer Management Consult ing Canum Infotech Cyient Info Tech. Agoda Company Open Insights Solutions LLC Tekion Corporation mPokket s/w Company GO-JEK India Autodesk Inc. Ignitarium Technology Solutio ns McAfee Enterprise Surya Systems Ayla Networks Qualys Tech. QuestAlliance Corporation Exapas Consultant ArrayGen Technologies ADCI - Karnataka GD Research Center Spiceworks Tech. Texas Instruments Myntra.com Collins Aerospace Aptus Data LAbs WHITE FORCE OUTSOURCING SERVICES PRIVATE LIMITED Mobiquity Consultancy Siemens Technology and Servi ces Private Limited Impiger Technologies AISPL - Karnataka Schneider Electric Ola Cabs Great Learning Wells Fargo Acuity Knowledge Partners Dun & Bradstreet HERE Technologies Coursera State Street Harman International NEC Opportunities Institutional Shareholder Servi ces Expedia Group Lingaro Group GE Digital Paytm Bank LeadSquared Pvt. Ltd, Inspirion Digital Solutions GE Healthcare RELX Group Elpis IT Solutions Fundnel Pte Ltd	Bengaluru India Jaipur Chennai Remote Remote in Goa Hyderabad Temporarily Remote Mumbai Gurgaon Pune Surat San Remote in India Noida Kochi Mohali Mysore Delhi Coimbatore Remote in Gurgaon Karnataka Indore Remote in Hyderabad Remote in Bengaluru Kanayannur Navi Mumbai Vadodara Chandigarh Agra Kolkata Thiruvananthapuram Bhubaneshwar West Bengal Chittoor New Delhi Punjab Ahmedabad Kandivali Kanpur Remote in Chennai Remote in Mumbai Powai Remote in Noida Remote in Rajastan Jamshedpur Tamil Nadu Kanayannur Navi Mumbai Vadodara Chandigarh Agra Kolkata Thiruvananthapuram Bhubaneshwar West Bengal Chittoor New Delhi Punjab Ahmedabad	Data Scientist Sr. Deaccession Scientist Sr. Data and Applied Scientist Data Analyst s/w Engineer Data Engineer CV Engineer ML Engineer Applied Scientist Data Analyst AI Engineer DL Engineer AI/ML Developer Data Science Consultant Applied Scientist Data Controller Security Analyst Sr. AI Analyst Technical Engineer AI/ML data lead Big data Analyst	Python R Language Linux Computer vision Machine Learnin Deep Learning SQL NLP AWS and Azure

CLUSTER 7

	Company Name	Location	Job Title	Skills
CLUSTER-7	UnitedHealth Group Deloitte Pvt. Ltd. EY India Siemens India Tesco PLC PepsiCo Inc.	Bengaluru Gurgaon Pune Noida Hyderabad Kolkata Thiruvananthapuram	Data Scientist Analytics ML Engineer Data Engineer Data Architect	Python Machine Learning NLP Deep Learning AWS and Azure

CLUSTER 8

	Company Name	Location	Job Title	Skills
CLUSTER-8	Unilever PLC Oracle Corporation Apple Inc. Mercedes-Benz Group AG Microsoft Corporation Dell Technologies NTT DATA ZF Group Accenture PLC. Google Inc. GE Renewable Energy Aditya Birla Group GE Corporate	Chennai Pune Hyderabad Bengaluru Gurgaon Telangana Nanakramguda Mumbai Noida Hadapsar	Data Scientist Sr. Data and Applied Scientist Data Analyst s/w Engineer CV Engineer ML Engineer Data Science Consultant Security Analyst Sr. AI Analyst Technical Engineer AI/ML data lead	Python Machine Learning Deep Learning SQL AWS and Azure
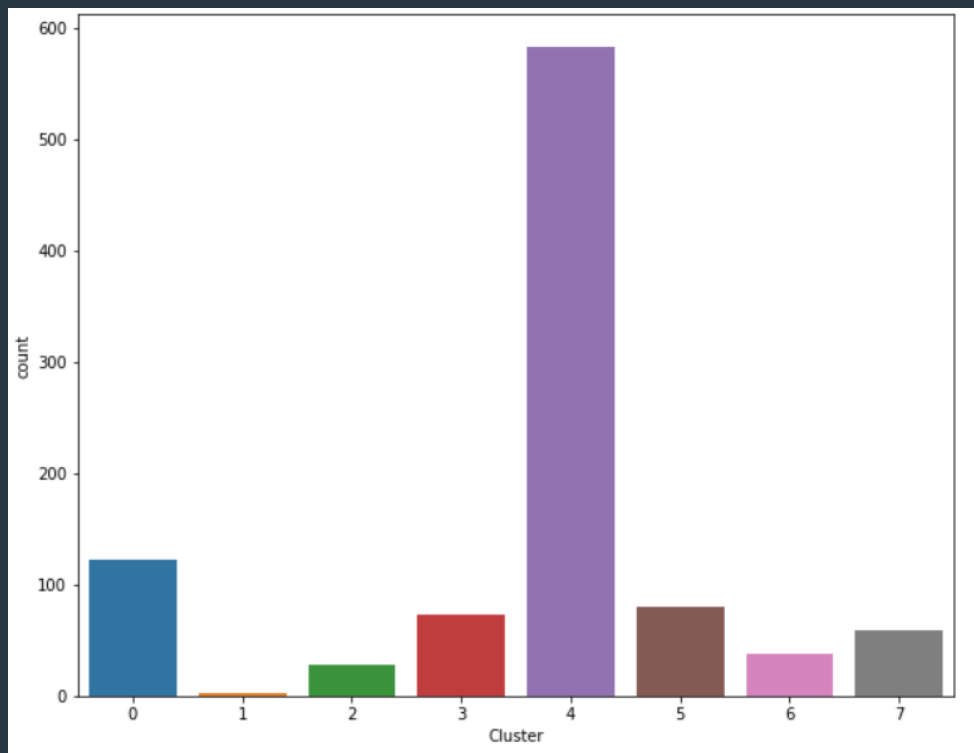
```
## Herarical Clustering
from scipy.cluster.hierarchy import linkage, dendrogram
merg = linkage(X.iloc[:,0:-1],method="ward")
plt.figure(figsize=(10,8))
dendrogram(merg,leaf_rotation = 40)
plt.xlabel("data points")
plt.ylabel("euclidean distance")
plt.show()
```
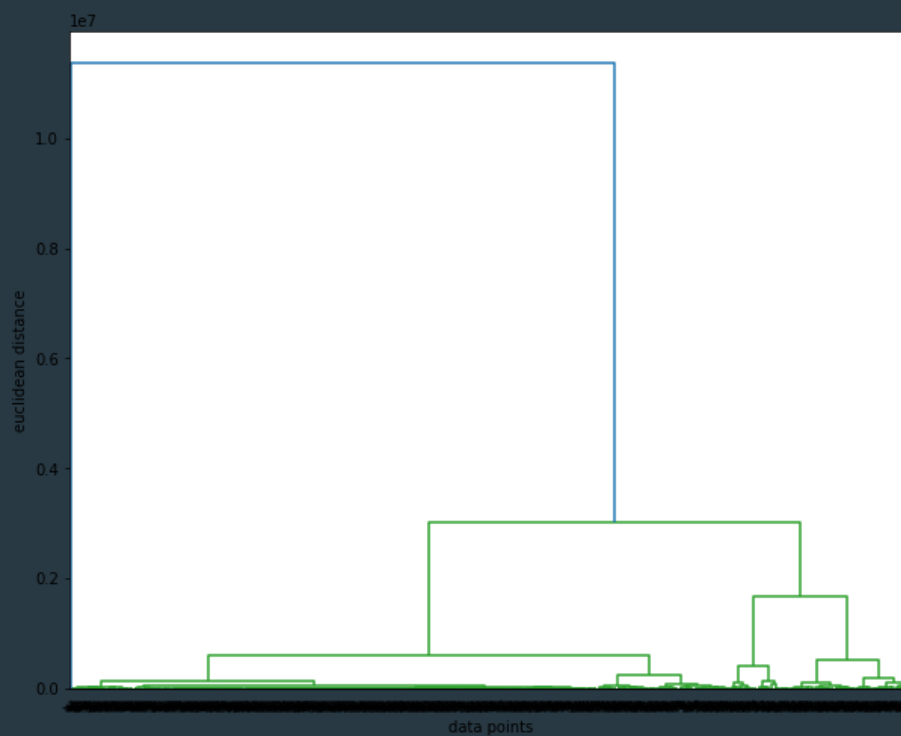
Link to github:-

WEB SCRAPING-

https://github.com/vaasu2002/Feynn-Labs/tree/main/Job%20Segmentation/WebScraping

DATASET-

https://github.com/vaasu2002/Feynn-Labs/tree/main/Job%20Segmentation/DATASET

SEGMENTATION-

https://github.com/vaasu2002/Feynn-Labs/blob/main/Job%20Segmentation/Segmentation%20.ipynb

https://github.com/vaasu2002/Feynn-Labs/tree/main/Job%20Segmentation