

In [3]:

```
1 ! pip install pandoc
```

Requirement already satisfied: pandoc in c:\users\manpr\anaconda3\lib\site-packages (2.3)

Requirement already satisfied: plumbum in c:\users\manpr\anaconda3\lib\site-packages (from pandoc) (1.8.2)

Requirement already satisfied: ply in c:\users\manpr\anaconda3\lib\site-packages (from pandoc) (3.11)

Requirement already satisfied: pywin32 in c:\users\manpr\anaconda3\lib\site-packages (from plumbum->pandoc) (302)

In [4]:

```
1 import docx2txt
```

## Data Analyst Job

In [5]:

```
1 job_description_DA=docx2txt.process('Data Analyst job description.docx')
```

In [3]:

```
1 print (job_description_DA)
```

Data Analyst job description

Data Analyst responsibilities include:

Interpreting data, analyzing results using statistical techniques

Developing and implementing data analyses, data collection systems and other strategies that optimize statistical efficiency and quality

Acquiring data from primary or secondary data sources

## Data Scientist Job

In [4]:

```
1 job_description_DS=docx2txt.process('DATA SCIENTISTS JOB DESCRIPTION.docx')
```

In [5]:

```
1 print (job_description_DS)
```

## DATA SCIENTISTS JOB DESCRIPTION

### Data Scientist Roles and Responsibilities:

Data scientist roles and responsibilities include:

Data mining or extracting usable data from valuable data sources

Using machine learning tools to select features, create and optimize classifiers

Carrying out preprocessing of structured and unstructured data

Enhancing data collection procedures to include all relevant information for developing analytic systems

Processing, cleansing, and validating the integrity of data to be used for analysis

Analyzing large amounts of information to find patterns and solutions

Developing prediction systems and machine learning algorithms

Presenting results in a clear manner

Propose solutions and strategies to tackle business challenges

Collaborate with Business and IT teams

### Data Scientist Skills

You need to master the skills required for data scientist jobs in various industries and organizations if you want to pursue a data scientist career. Let's look at the must-have data scientist qualifications. Key skills needed to become a data scientist:

**Programming Skills** - knowledge of statistical programming languages like R, Python, and database query languages like SQL, Hive, Pig is desirable. Familiarity with Scala, Java, or C++ is an added advantage.

**Statistics** - Good applied statistical skills, including knowledge of statistical tests, distributions, regression, maximum likelihood estimators, etc. Proficiency in statistics is essential for data-driven companies.

**Machine Learning** - good knowledge of machine learning methods like k-Nearest Neighbors, Naive Bayes, SVM, Decision Forests.

**Strong Math Skills (Multivariable Calculus and Linear Algebra)** - understanding the fundamentals of Multivariable Calculus and Linear Algebra is important as they form the basis of a lot of predictive performance or algorithm optimization techniques.

**Data Wrangling** - proficiency in handling imperfections in data is an important aspect of a data scientist job description.

**Experience with Data Visualization Tools** like matplotlib, ggplot, d3.js., Tableau that help to visually encode data

**Excellent Communication Skills** - it is incredibly important to describe findings to a technical and non-technical audience.

Strong Software Engineering Background

Hands-on experience with data science tools

Problem-solving aptitude

Analytical mind and great business sense

Degree in Computer Science, Engineering or relevant field is preferred

Proven Experience as Data Analyst or Data Scientist

## Data Engineer Job

In [6]:

```
1 job_description_DE=docx2txt.process('JOB DESCRIPTION FOR DATA ENGINEER.docx')
```

In [7]:

```
1 print(job_description_DE)
```

## JOB DESCRIPTION FOR DATA ENGINEER

### Job brief

We are looking for an experienced data engineer to join our team. You will use various methods to transform raw data into useful data systems. For example, you'll create algorithms and conduct statistical analysis. Overall, you'll strive for efficiency by aligning data systems with business goals.

To succeed in this data engineering position, you should have strong analytical skills and the ability to combine data from different sources. Data engineer skills also include familiarity with several programming languages and knowledge of learning machine methods.

If you are detail-oriented, with excellent organizational skills and experience in this field, we'd like to hear from you.

### Responsibilities

Analyze and organize raw data

Build data systems and pipelines

Evaluate business needs and objectives

Interpret trends and patterns

Conduct complex data analysis and report on results

Prepare data for prescriptive and predictive modeling

Build algorithms and prototypes

Combine raw information from different sources

Explore ways to enhance data quality and reliability

Identify opportunities for data acquisition

Develop analytical tools and programs

Collaborate with data scientists and architects on several projects

Requirements and skills

Previous experience as a data engineer or in a similar role

Technical expertise with data models, data mining, and segmentation techniques

Knowledge of programming languages (e.g. Java and Python)

Hands-on experience with SQL database design

Great numerical and analytical skills

Degree in Computer Science, IT, or similar field; a Master's is a plus

Data engineering certification (e.g IBM Certified Data Engineer) is a plus

## Input ur resume

In [9]:

```
1 Resume1=docx2txt.process(input())
```

data-analyst resume 1.docx

In [10]:

```
1 content = [job_description_DA, Resume1]
2
3
4 from sklearn.feature_extraction.text import CountVectorizer
5 cv = CountVectorizer()
6 count_matrix = cv.fit_transform(content)
7
8
9 from sklearn.metrics.pairwise import cosine_similarity
10 mat = cosine_similarity(count_matrix)
11 print(mat)
```

```
[[1.          0.65982032]
 [0.65982032 1.          ]]
```

In [11]:

```
1 print('Resume Matches by ' + str(mat[1][0]*100) + '%' + ' for data analyst position')
```

Resume Matches by 65.98203223065642% for data analyst position:

In [12]:

```
1 content = [job_description_DS, Resume1]
2
3
4 from sklearn.feature_extraction.text import CountVectorizer
5 cv = CountVectorizer()
6 count_matrix = cv.fit_transform(content)
7
8
9 from sklearn.metrics.pairwise import cosine_similarity
10 mat = cosine_similarity(count_matrix)
11 print(mat)
```

```
[[1.          0.63170411]
 [0.63170411 1.          ]]
```



In [13]:

```
1 print('Resume Matches by ' + str(mat[1][0]*100) + '%' + ' for data science position')
```

Resume Matches by 63.17041106461077% for data science position:

In [14]:

```
1 content = [job_description_DE, Resume1]
2
3
4 from sklearn.feature_extraction.text import CountVectorizer
5 cv = CountVectorizer()
6 count_matrix = cv.fit_transform(content)
7
8
9 from sklearn.metrics.pairwise import cosine_similarity
10 mat = cosine_similarity(count_matrix)
11 print(mat)
```

```
[[1.          0.64566453]
 [0.64566453 1.          ]]
```

In [15]:

```
1 print('Resume Matches by ' + str(mat[1][0]*100) + '%' + ' for data engineer position')
```

Resume Matches by 64.56645299824838% for data engineer position:

# THANK YOU

In [ ]:

```
1
```