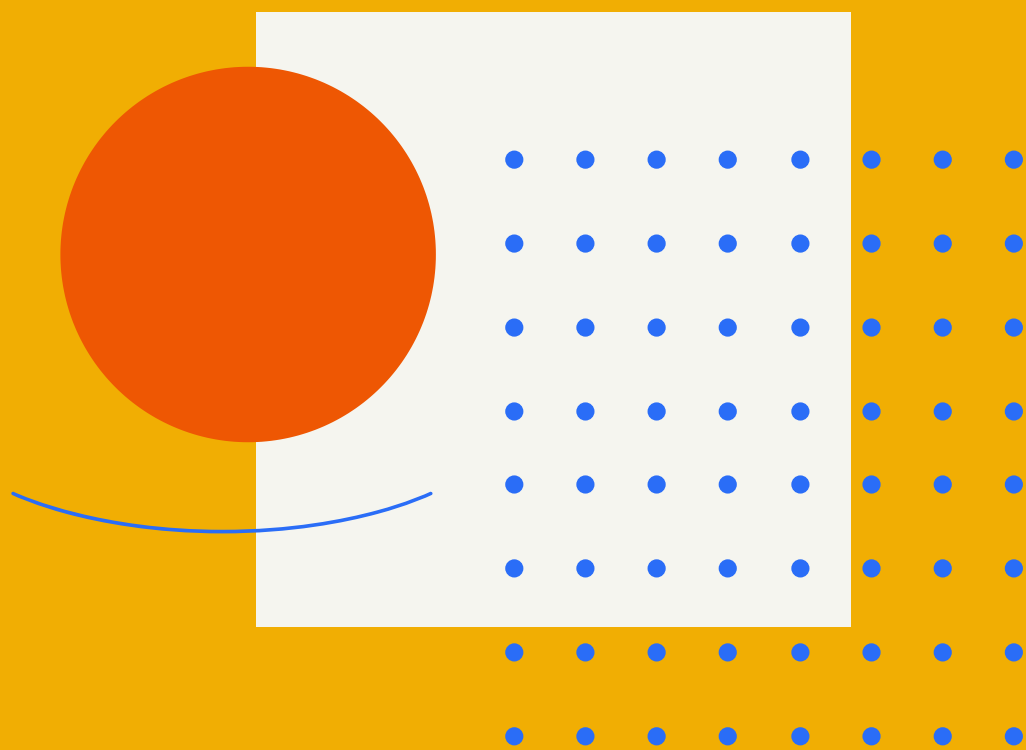




Job & Talent Recommendation Platform

Group X: Annie Ho, Cindy Choi, Constance Wong,
Cooper Li, Sanjay Nair



Meet our Team



Annie Ho



Cindy Choi
xxxxxxxxxxxx



Constance Wong
Constance has
extensive
experience in
Retail sales
analysis.

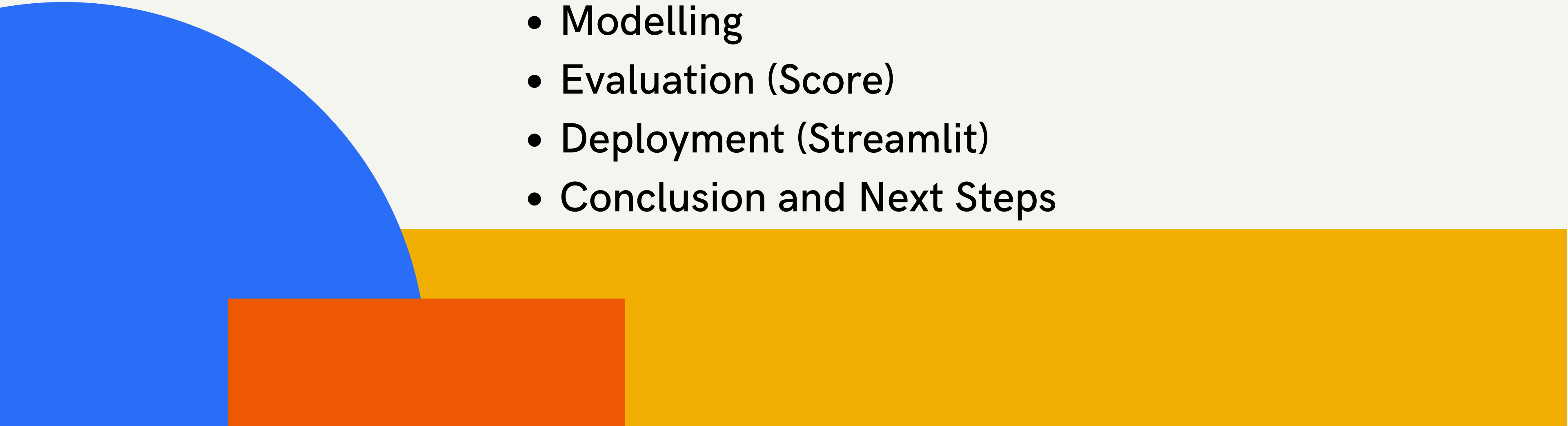


Cooper Li
Former quality
control technician
of semiconductor
and electronic
goods. Currently
seeking
opportunities in
Data Science
industry.



Sanjay Nair
xxxxxxxxxxxx

Table of Contents

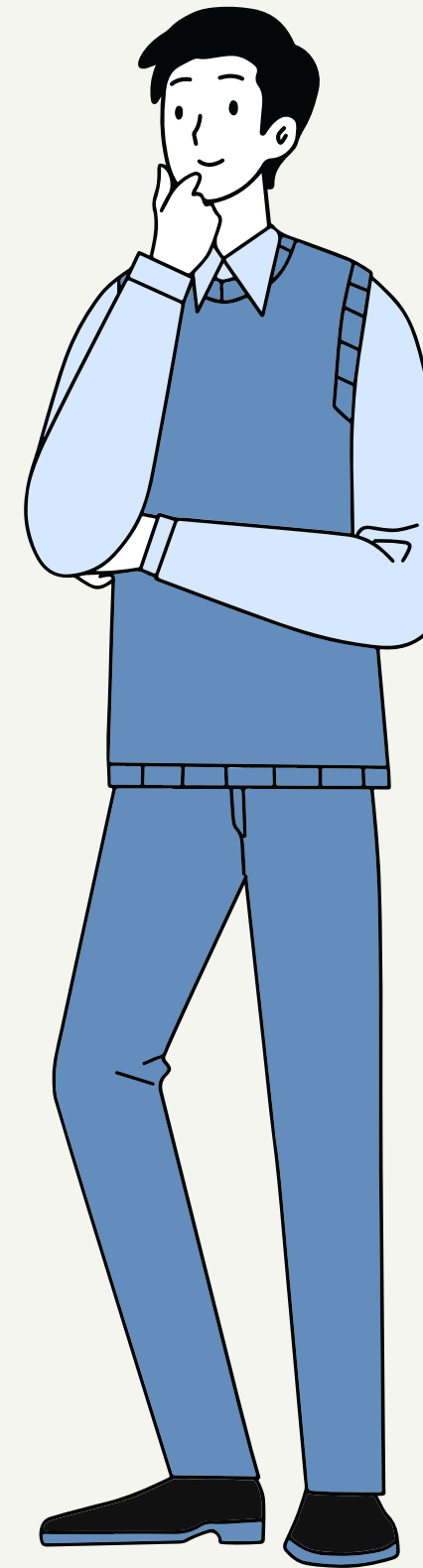
- Project Aim
 - Business Value
 - Milestones
 - Process Flow
 - Data Collection & Preprocessing
 - Modelling
 - Evaluation (Score)
 - Deployment (Streamlit)
 - Conclusion and Next Steps
- 

Scenario 1

Peter is an job-seeker currently looking for a job

However, he is struggling with the following:

- 1) Poor job market
- 2) Low confidence
- 3) Finding a job that matches his skills



Scenario 2

Chloe is a job recruiter hired to help companies find the right candidate

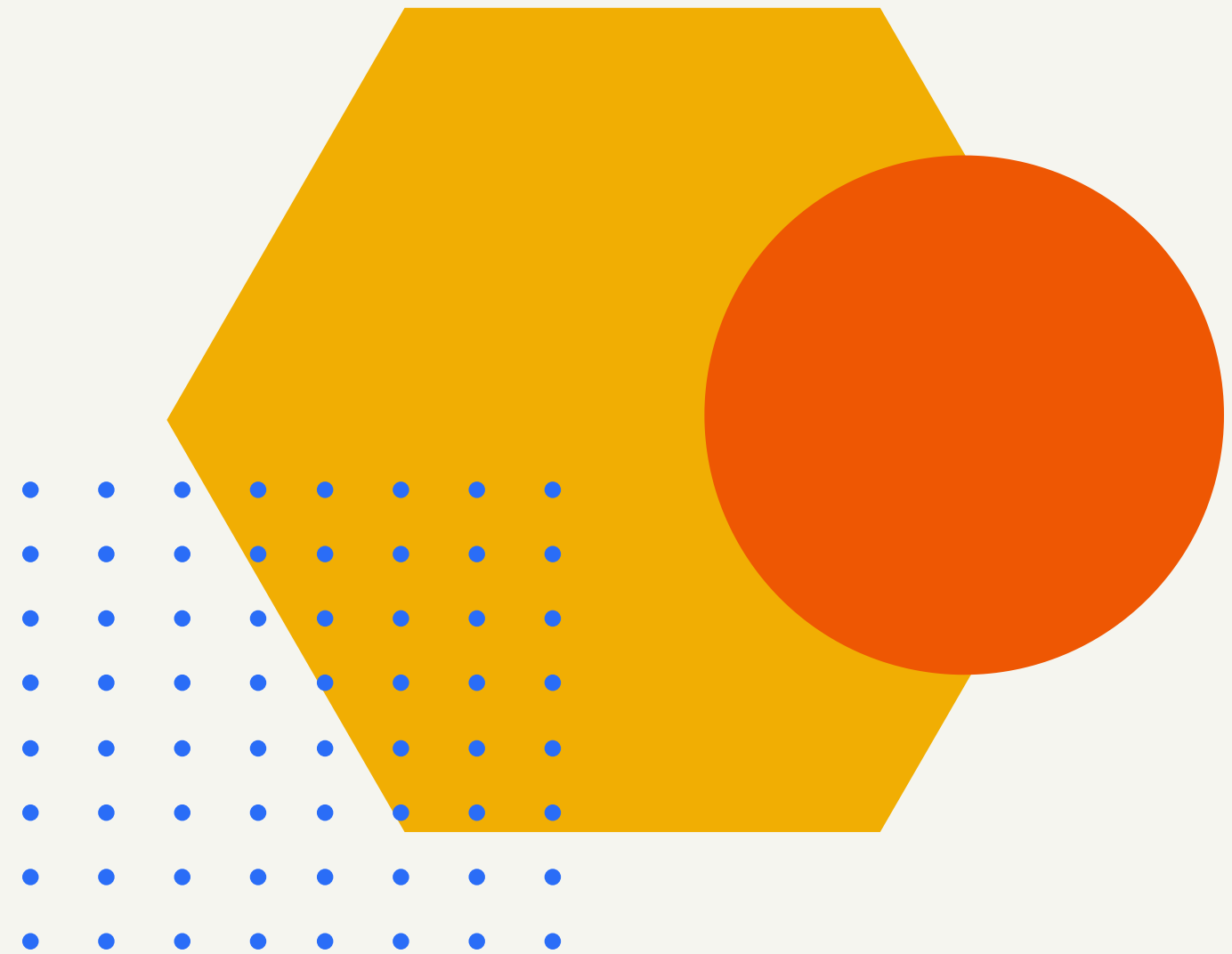
However, she wishes struggling with the following:

- 1) Finding the time to read through all the CVs
- 2) Matching the right candidate to the right position



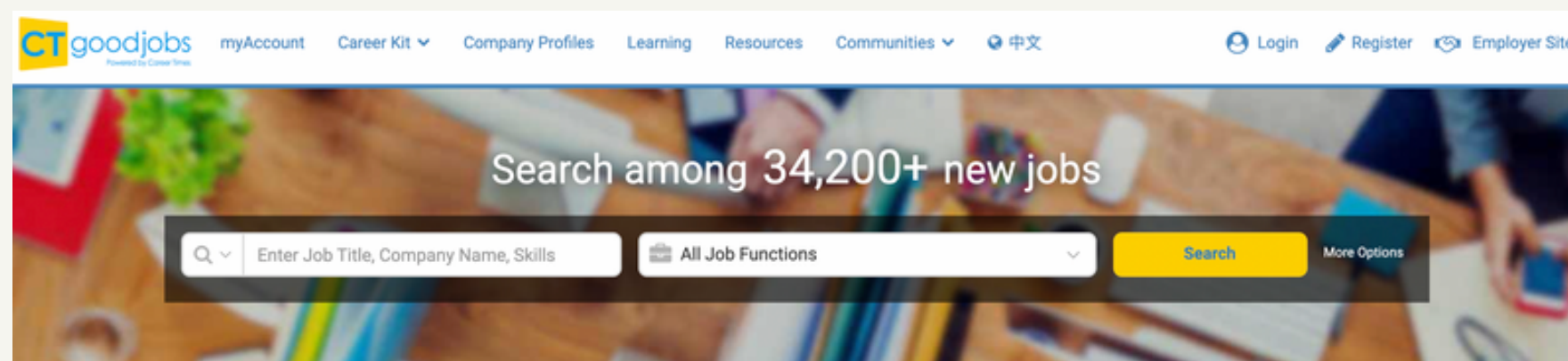
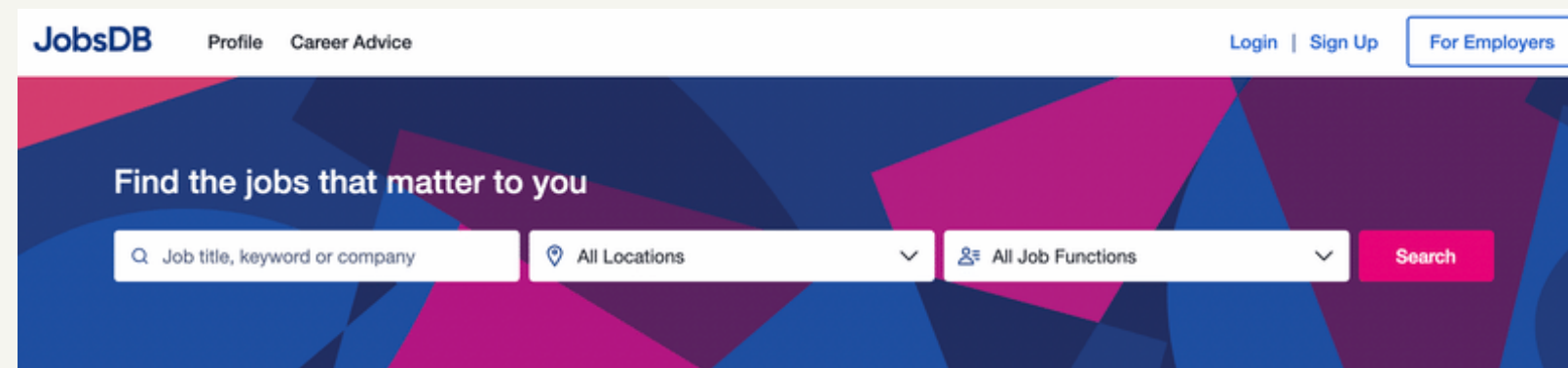
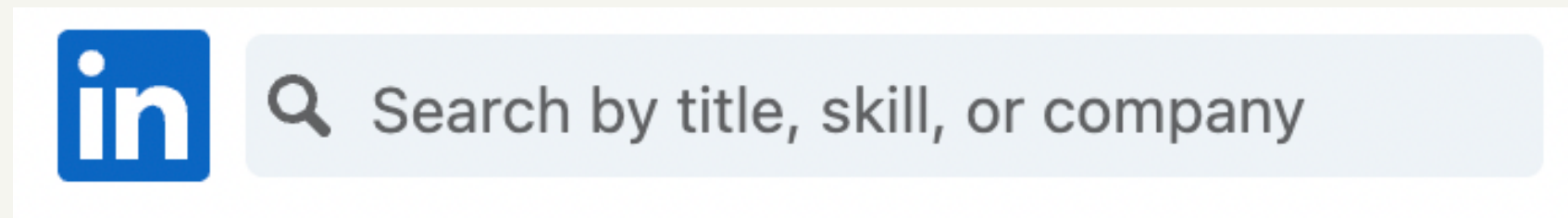
Project Aim

- Assist **applicants** to search for potential jobs
- Assist **recruiters** to find talents



Business Value

Jobmigo brings the job world to you

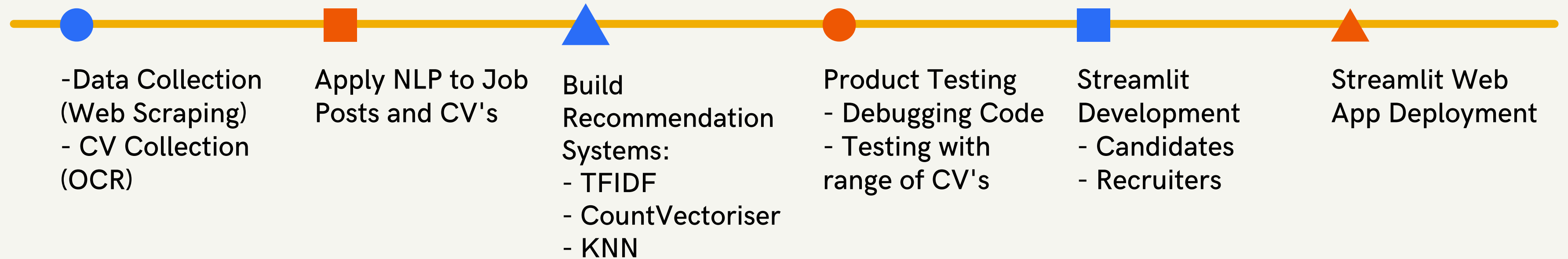


- **Simplicity - Drag and Drop**
- **Time is money**
- **Tailor-made**



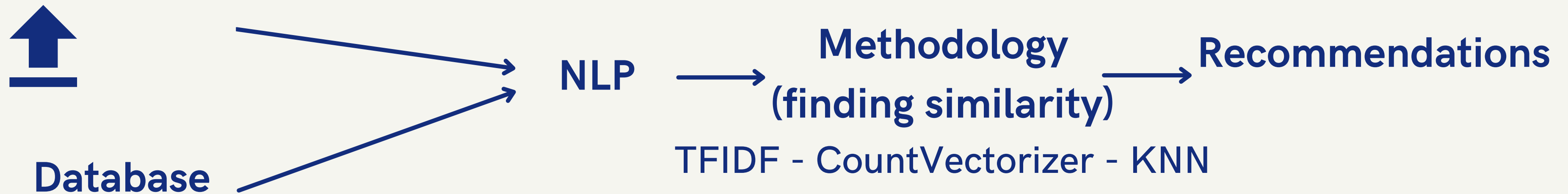
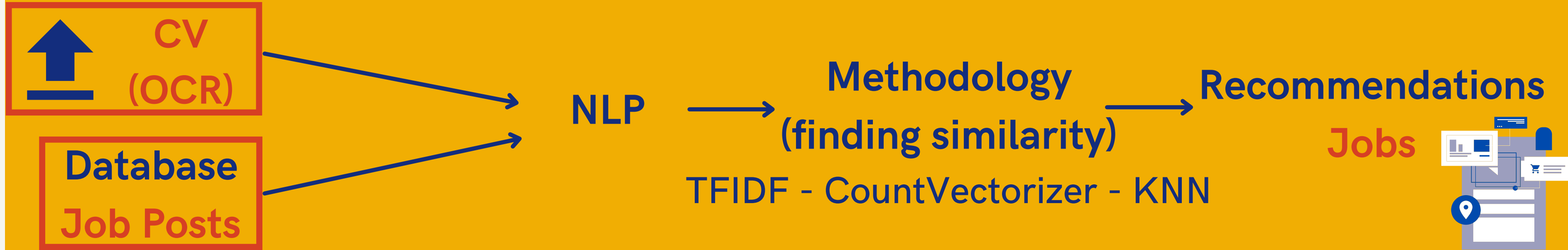
Current Job Search by Title or keyword

Milestones:



Process Details

Applicants View



Data Overview

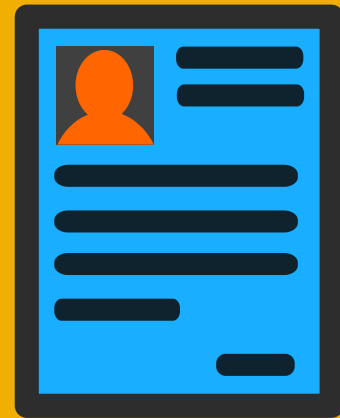
CV

VS

Job Posts

Applicants View

Any CV



Recruiters View

Source: Friends & Google
CV: 54



Applicants View

Source: JobsDB
Method: Web Scraping
Jobs: 20,137
Job Functions: 5



Recruiters View

Any Job Posts



Data Collection - Webscraping

The screenshot shows the JobsDB website interface. At the top, there's a navigation bar with 'JobsDB', 'Profile', and 'Career Advice'. Below this is a search bar with the placeholder 'Job title, keyword or company', a location filter set to 'All Locations', and a button for 'Information'. There are also filters for 'Salary', 'Job type', 'Date posted', and a 'Sort By Relevance' dropdown. The main content area displays a list of jobs, with the first one being 'Assistant IT Manager' at 'Wang On Group' in 'Kowloon Bay', posted '2h ago'. A detailed view of this job is shown on the right, including the company logo, job title, company name, location, and posting time. An orange box highlights the job details, and an orange arrow points from this box to a list of data fields on the right.

JobsDB Profile Career Advice

Job title, keyword or company All Locations Information

Salary Job type Date posted Sort By Relevance

1-30 of 8,057 jobs

Assistant IT Manager
Wang On Group
Kowloon Bay
2h ago

Finance and IT Manager
Abdoolally Ebrahim Services Limited
Kwun Tong
3h ago

Assistant IT Manager
Wang On Group
Kowloon Bay
Posted 3 hours ago

Job Highlights

1	JobID
2	title
3	company
4	location
5	salary
6	post date
7	job highlights
8	job description
9	company overview
10	career level
11	qualification
12	experience
13	job type
14	industry
15	benefit
16	webpage
17	sal_low
18	sal_high
19	All

Data Processing

Salary

HK\$11,000 - HK\$25,000 /month —————> 11000 , 25000

Phone number & Email

**Stephen
Greet**
Senior Data Scientist

Brooklyn, NY 11222

(123) 456-7890

stephen@beamjobs.com

LinkedIn: [/in/stephen-greet](#)

Blog: [2020 Data Science Resume Guide](#)

Experience

EdTech Company

January 2018 - Present, New York NY

Senior Data Scientist

Worked closely with the product team to build a production recommendation engine in Python that improved the average length on page for users and resulted in \$325k in incremental annual revenue

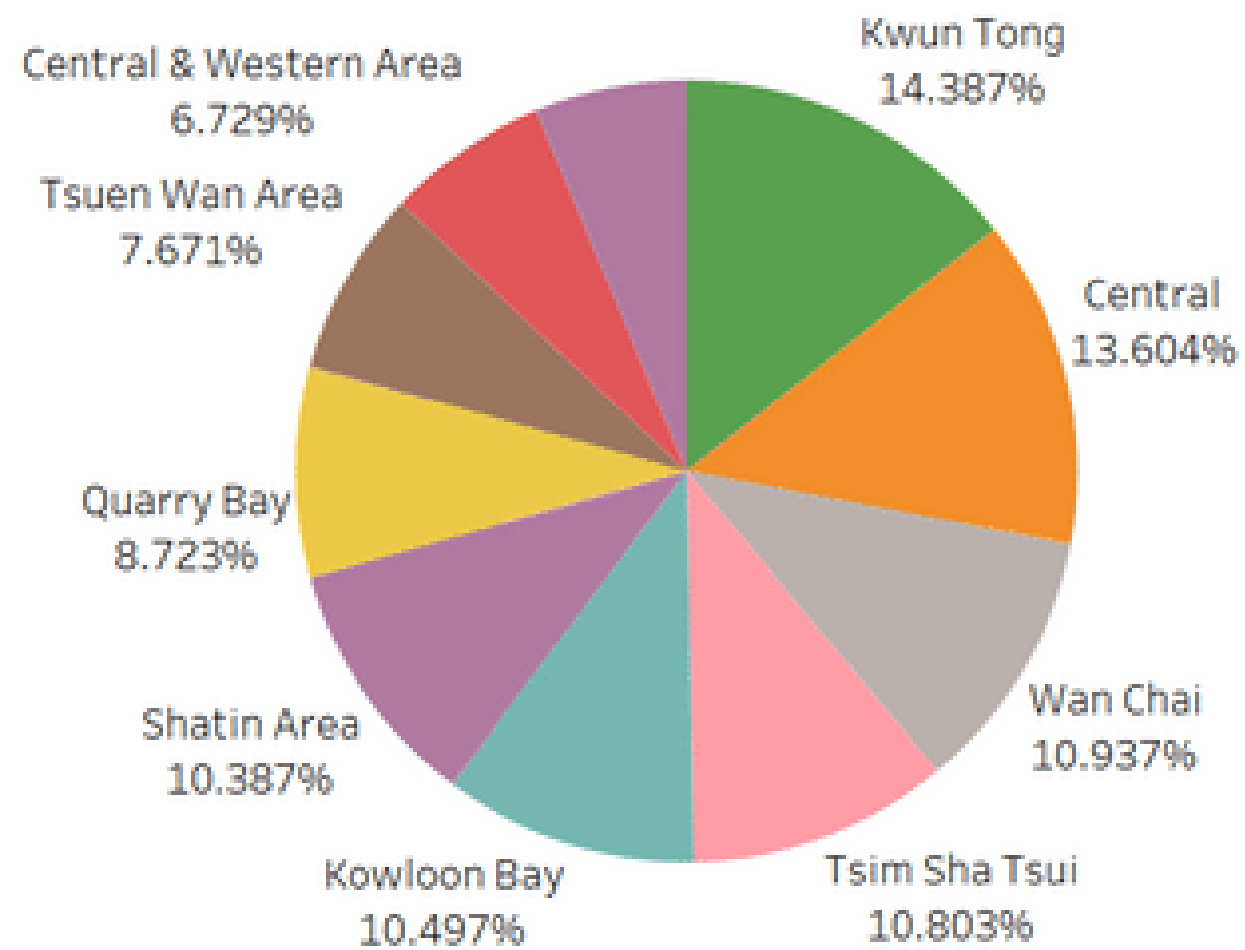
extract

(123) 456-7890

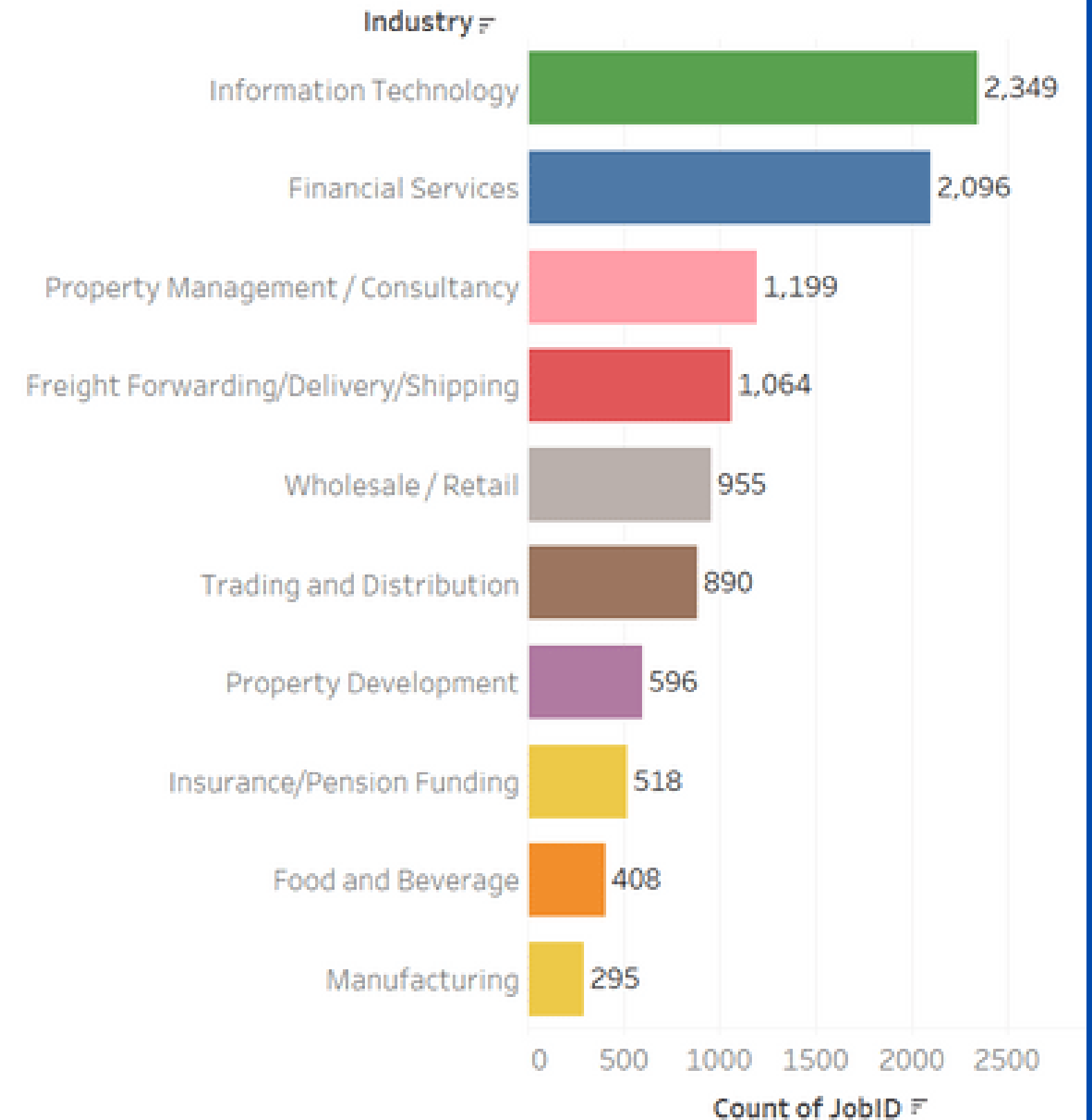
stephen@beamjobs.com

Exploratory Data Analysis (EDA)

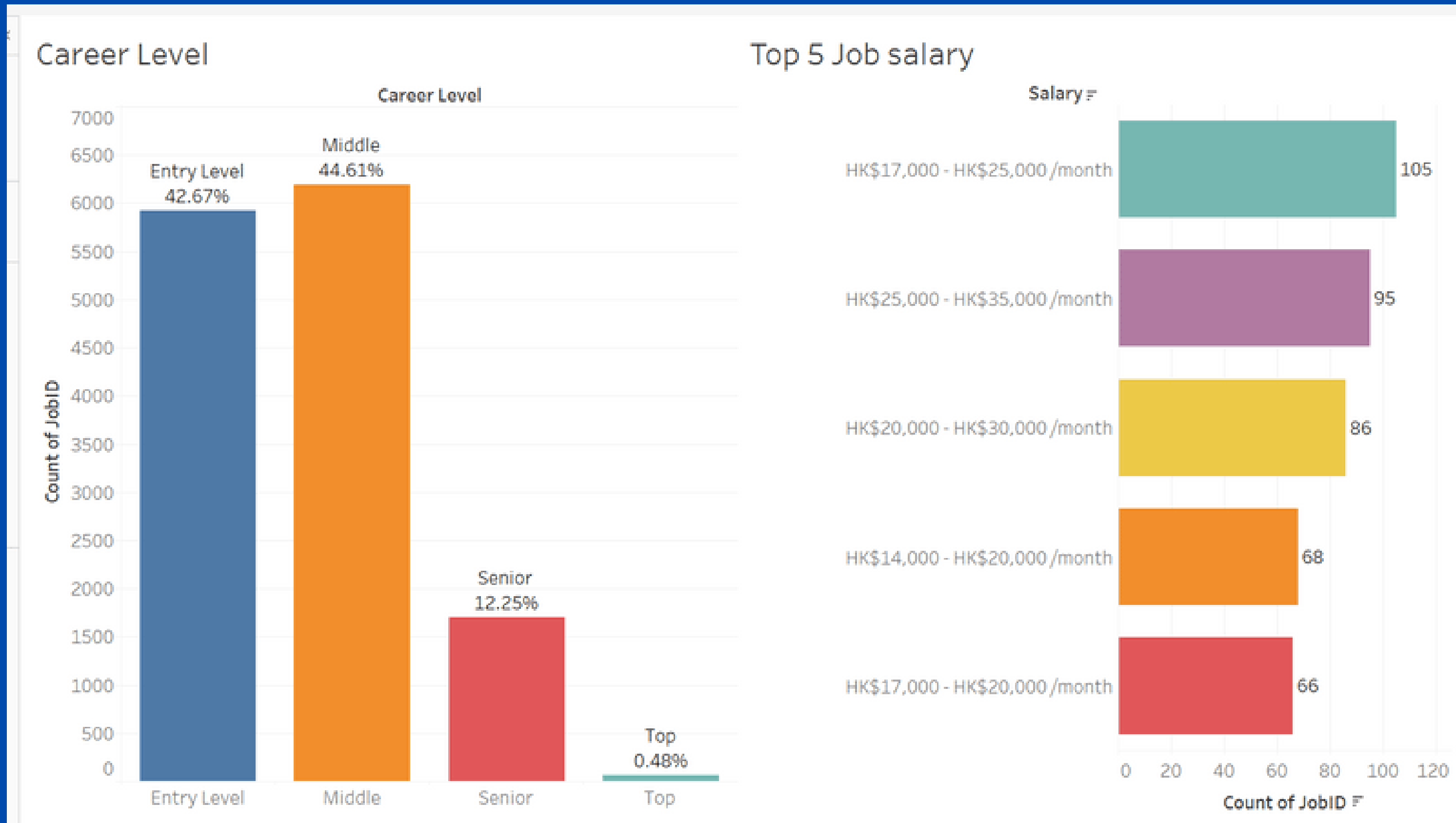
Top 10 Job locations in HK



Top 10 industries



Exploratory Data Analysis (EDA)



OCR

Optical Character Recognition

What is OCR?

A mechanical conversion of images of typed or handwritten word into machine-encoded text



SCANNED FILE



OCR SOFTWARE



EDITABLE FILE


OCR

Optical Character Recognition

CV pdf version




text version - string





Diana Dawa


Data Scientist


Data Scientist with 4+ years of broad-based experience in building data-intensive applications, overcoming complex architectural, and scalability issues in diverse industries. Proficient in predictive modeling, data processing, and data mining algorithms, as well as scripting languages, including Python and Java. Capable of creating, developing, testing, and deploying highly adaptive diverse services to translate business and functional qualifications into substantial deliverables.


 diana@novoresume.com

 044 412 2020

 Palo Alto, CA

 diana-dawa.com

 linkedin.com/in/diana.dawa

 diana.dawa

WORK EXPERIENCE

Data Scientist

Future Energy Ltd.

06/2018 - Present

Palo Alto, CA

- Develop action plans to mitigate risks in decision making while increasing profitability by leveraging data science.
- Drive the interaction and partnership between the managers to ensure active cooperation in identifying as well as defining analytical needs, and generating the pull-through of insights with the business.
- Build predictive models using various machine learning tools to predict the possibility of equipment failure.
- Develop algorithms using Natural Language Processing and Deep Learning models for predictive maintenance.
- Design algorithms to track and detect anomalies in multiple sensors data for the Energy Industry.
- Demonstrate knowledge and execution of application programming interface development and test automation.

Data Analyst

THETA Financial Group

01/2016 - 05/2018

San Francisco, CA

- Utilized analytical and technical expertise to provide insights and proposals to support business improvements.
- Evaluated analytical model findings in the Global Monitoring Report, the company's flagship product.
- Conducted business analysis to understand business needs and requirements to translate into conceptual designs.
- Actively engaged in the quantitative analysis of sophisticated modeling to address business issues.

Business Analyst / Statistician

Maxicare Healthcare

02/2013 - 12/2015

- Conducted business process analysis and identified critical issues, gaps, and needs for an established process center.
- Developed Key Performance Indicators (KPI) and presented it to the management and led to the execution plan.
- Analyzed and produced KPI reports allowing to monitor field service engineer and customer care center closely.
- Led training sessions on the software developed and presented it to the management for approval of deployment.

GENERAL SKILLS

Data Visualization

Machine Learning

Deep Learning

Pattern Recognition

Database Structures & Algorithms

Statistical Analysis

Data Preparation

Quality Management

Agile Methodologies

TECHNICAL SKILLS

Operating System:

Windows, MacOS, Linux

Database/Server:

My SQL, Postgres, SQL Server

Programming Language:

Python, scikit-learn, Python, OpenCV, D3.js, H2O.ai, Spark, Hadoop, R Programming, Django, Angular.js, HTML, SQL, JavaScript, PHP

Other Software/Tools:

Tableau, Deep Learning, Machine Learning, IP Cameras, AWS Services, Microsoft Azure

CERTIFICATES

Certification for Applied Data Analytics (2019)

Cloudera Data Science Essentials Certificate (2018)

Essentials of High Performance and Parallel Statistical Computing with R (2018)


EDUCATION


Master of Science in Computer Science and Informatics


San Francisco University


2014 - 2016


INTERESTS

 Video Games

 Music

 Renewable Energy

 Artificial Intelligence



Summary
Motivated and determined individual from a Data Science and Machine Learning background. My logical and analytical skills, as well as recognizing the importance of teamwork for opportunities where I can deploy data analytic skills.

Technical Skills
● Proficient: Python, Pandas, Keras-Tensorflow, Machine Learning, Deep Learning, Tableau
● Familiar with: SQL

Xccelerate Full-time Data Science and Machine Learning Immersive Bootcamp

Graduate from Xccelerate's Full-time Data Science Bootcamp, developing various data both individually and in a team. Some of the Projects:

House Ownership Scheme Market Analysis (Web-scraping, Pandas)
● Deployed web-scraping technique on various websites and obtained valuable
● Deployed Pandas visualization tools to analyze housing estate information.
● Identified significant insights in the public housing estate market, i.e. locations

Flight Ticket Price Prediction (Machine Learning, Pandas)
● Performed techniques such as Pandas and Regular Expression on a large dataset phase.
● Trained and deployed various regression models to conduct price prediction.
● Models could predict prices with more than 75%.

Facial Detection and Recognition System (Deep Learning)
● Deployed facial detection technique and captured faces from real-life picture
● Developed and optimized model with convolution neural network (CNN) for image prediction
● Model achieved more than 85% on gender prediction and 83% on emotion

Work Experience
● Laboratory Technician at Smith and Associates Far East Ltd and testing
Conduct quality control on electronic components via visual inspection and
Detect counterfeit electronic components and report relevant discrepancies

● Technician at Crestron Asia Limited products 11/ 2015
Perform testing and repairing of company's products

Education
● Bachelor of Engineering in Chemical Engineering, University of Birmingham
● Brockenhurst College A-Levels: Mathematics, Chemistry, Physics, Chinese, Mandarin

Language: Native in Cantonese, fluent in English and familiar with Mandarin

NLP

Natural Language Processing

Graduate science/ both from machine learning individually and Xccelerate's in projects a team. Full-time Data Science Bootcamp, developing various data Projects:

House Ownership Scheme Market Analysis (Web-scraping, Pandas) and obtained valuable dataset for further analysis. ● Deployed web-scraping technique on various websites ● Deployed Pandas visualization tools to analyze housing estate information. ● Identified significant insights in the public housing estate market, i.e. locations and pricing.

Flight Ticket Price Prediction (Machine Learning, Pandas) on a large datas et during data preprocessing phase. ● Performed techniques such as Pandas and Regular Expression ● Trained and deployed various regression models to conduct price prediction. ● Models could predict prices with more than 75%.

Facial Detection and Recognition System (Deep Learning) and captured faces from real-life pictu res ● Developed and optimized model with convolution neural network (CNN) for image predi ction. ● Model achieved more than 85% on gender prediction and 83% on emoti on prediction.

Work Experience at Smith and Associates Far East Ltd 07/ 2017 - 06/ 2020 Conduct quality control on electronic components via visual inspection and testing Detect counterfeit electronic components and report relevant discrepancies



```
'github',  
'linkedin',  
'summary',  
'motivated',  
'determined',  
'individual',  
'data',  
'science',  
'machine',  
'learning',  
'bootcamp',  
'experienced',  
'python',  
'data',  
'analysis',  
'machine',
```

4. Bigram Collection Finder:

- Get keywords from the list of tokenised words

```
(( 'data', 'science'), 4)  
(( 'deep', 'learning'), 2)  
(( 'electronic', 'component'), 2)  
(( 'facial', 'detection'), 2)  
(( 'full-time', 'data'), 2)  
(( 'housing', 'estate'), 2)  
(( 'machine', 'learning'), 6)  
(( 'prediction', 'model'), 2)  
(( 'price', 'prediction'), 2)  
(( 'science', 'machine'), 2)
```

```
['data science',  
 'deep learning',  
 'electronic component',  
 'facial detection',  
 'full-time data',  
 'housing estate',  
 'machine learning',  
 'prediction model',  
 'price prediction',  
 'science machine']
```

1. Word tokenization

- Tokenize words into a list of strings

2. Stopwords:

- Remove words like ('i', 'me', 'my', 'myself', 'we')

3. Lemmatization:

- Return the words back to the original form

Recommendation Systems

1

**Count
Vectorizer**

**Cosine
Similarity**

2

TF-IDF

**Cosine
Similarity**

3

TF-IDF

KNN

CountVectorizer + Cosine Similarity

	1000	be	for	great	greatest	is	it	lasagna	life	love	loved	of	the	thing	times	to
sentence 1	0	0	0	0	1	1	0	0	1	1	0	1	1	1	0	0
sentence 2	0	1	0	2	0	1	1	0	0	1	1	0	0	0	0	1
sentence 3	0	0	0	0	1	1	0	0	0	1	0	0	1	1	0	0
sentence 4	1	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0

CountVectorizer:

- It converts a set of strings into a sparse matrix
- One-hot encode the text document

TF-IDF + Cosine Similarity

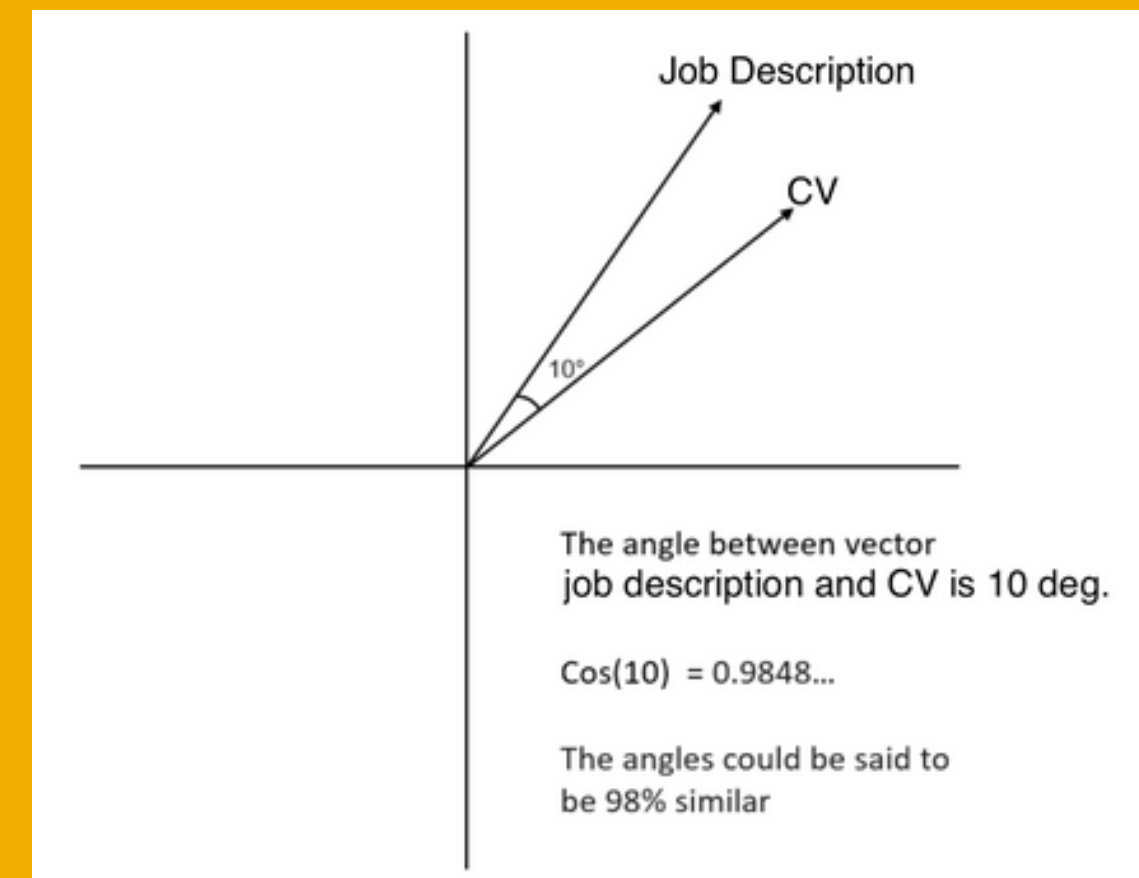
Term Frequency - Inverse Document Frequency (TF-IDF):

- Term Frequency, TF: The number of times a term occurs in a document
- Inverse Document Frequency, IDF: It measures how important a term is

sentence 1	earth is the third planet from the sun				
sentence 2	Jupiter is the largest planet				
Word	TF (Sentence 1)	TF (Sentence 2)	IDF	TF*IDF (sentence 1)	TF*IDF (Sentence 2)
earth	1/8	0	$\log(2/1)=0$	0.0375	0
is	1/8	1/5	$\log(2/2)=0$	0	0
the	2/8	1/5	$\log(2/2)=0$	0	0
third	1/8	0	$\log(2/1)=0.3$	0.0375	0
planet	1/8	1/5	$\log(2/2)=0$	0	0
from	0	0	$\log(2/1)=0.3$	0	0
sun	1/8	0	$\log(2/1)=0.3$	0.0375	0
largest	0	1/5	$\log(2/1)=0.3$	0	0.06
Jupiter	0	1/5	$\log(2/1)=0.3$	0	0.06

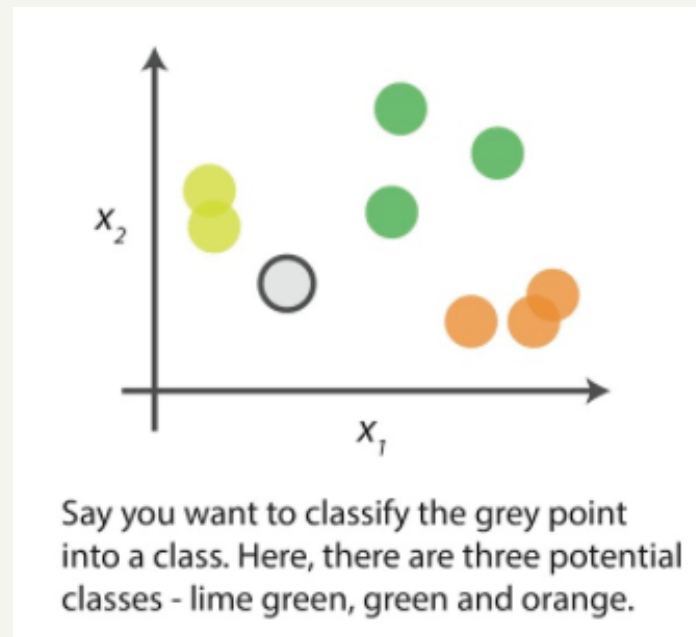
Cosine Similarity:

- It is a measurement that quantifies the similarity between two or more vectors (i.e. vectorised words)
- Vectorised word documents
 - CV
 - Job description

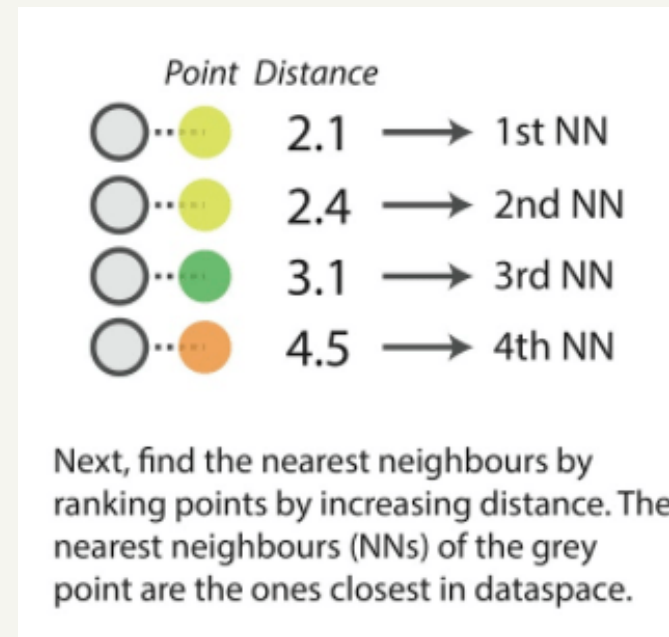


TF-IDF + KNN

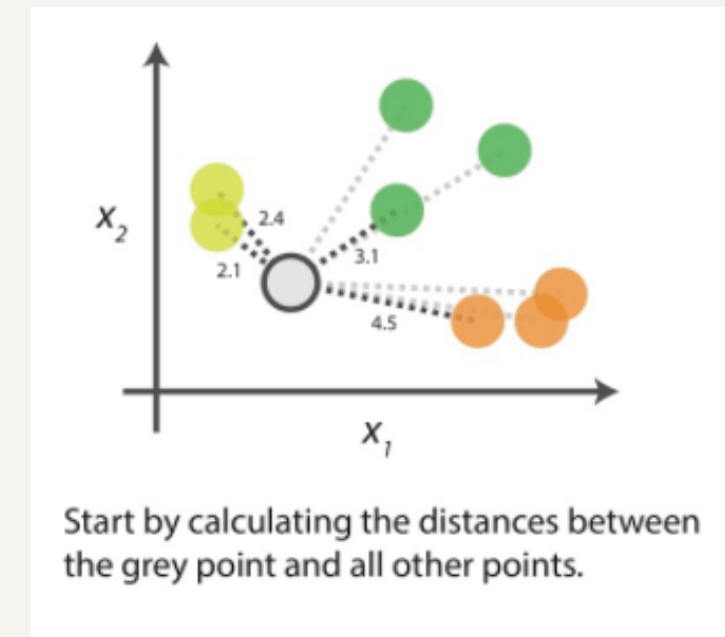
1. Look at the data



2. Find the neighbours



3. Calculate distances




KNN

- Supervised machine learning model with applying TF-IDF
- Calculating the distance between input text and training text
- Highest recommended CV/JD will be the one with the shortest distance to the input text

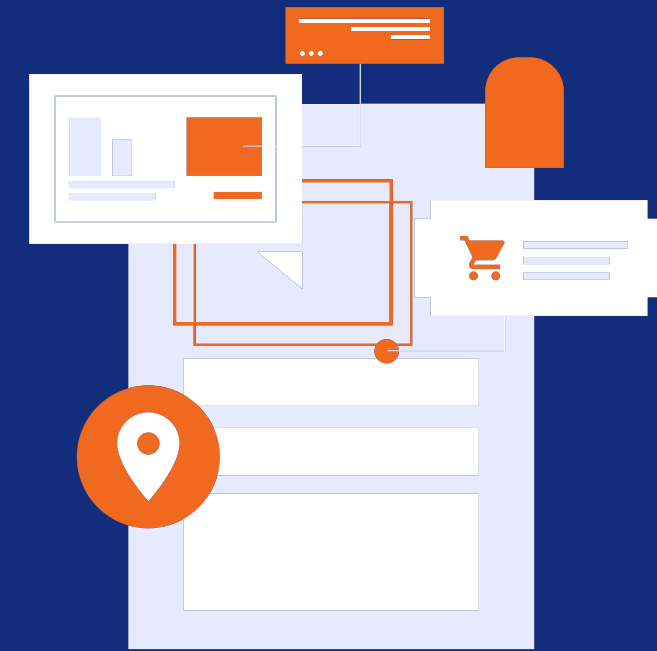
Results:

Combined 3 methods Job Recommendation list



	JobID	title	KNN	TF-IDF	CV	Final
0	641	Senior Machine Learning Engineer/Expert (Plann...	0.333333	0.333333	0.189947	0.856614
3	7357	Associate Data Scientist/ Data Scientist - Top...	0.27341	0.275167	0.278731	0.827308
7	7259	Data Insights Analyst (Big Data- Machine Learn...	0.249845	0.252081	0.325352	0.827278
2	3942	Software Engineer Trainee (AI / GIS)	0.275054	0.276773	0.248643	0.80047
10	857	Associate Data Scientist	0.214888	0.217617	0.310602	0.743106
8	7312	Artificial Intelligence (AI) R&D Engineer	0.242279	0.244644	0.238304	0.725226
9	5717	Data Science Manager - Machine Learning/AI 2...	0.239048	0.241464	0.212282	0.692794
4	12794	Vice President, Big Data Analytics & Machine L...	0.270547	0.272368	0.142904	0.685819
12	3247	Machine Learning Engineer/ Data Scientist	0.19946	0.202323	0.259108	0.66089
14	7582	Data Scientist Team Lead	0.178896	0.181858	0.299642	0.660395

Streamlit Demo





Job Recommendation

Upload your CV



Drag and drop file here

Limit 200MB per file • PDF

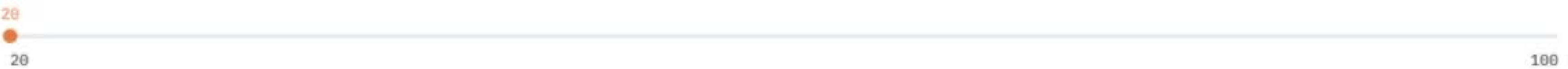
Browse files

Career level

Entry Level x Middle x Senior x Top x

Not Specified x

Number of Job Recommendations:



Conclusions

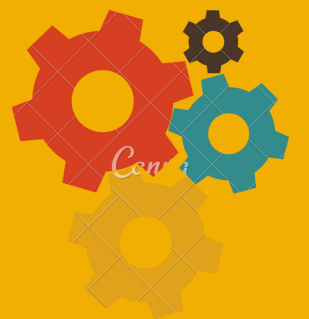


Recruiters Benefits - less bias & less time for screening CV

Challenge: deploy Tableau in python

Applicants Benefits -
Opportunities
Faster
Tailor-made features

Simplify the seeking process - Drag and Drop



Next steps



Further develop recommendation system to return accurate relevant results



Engage with job posting companies to get access to updated database / API



Invite applicants to join the CV pool



Build a chatbot in the streamlit to make our web app more interactive



More features on the dashboard



Thank you!

Q & A