



# WHICH DATA ROLE

Kara Bethany Liu

# MOTIVATION

- JOB-SEEKERS:  
IDENTIFY ROLES  
THEY ARE MOST  
SUITED FOR
- EMPLOYERS:  
ROUGH  
IDENTIFICATION  
FOR THE RIGHT FIT
- I WANT YOU TO  
HIRE ME





data scientist resumes  
Sort by: relevance - date

**Last Updated**  
within last 6 months

**Job Titles**  
 Data Scientist 1300+  
 Data Scientist Intern 100+  
 Senior Data Scientist 100+  
 Data Analyst 60+  
 Lead Data Scientist 30+  
 Research Assistant 10+

**Company**  
 Metis 10+  
 IBM 10+  
 Booz Allen Hamilton 10+  
 THE NIELSEN 10+  
 Deutsche Bank 8  
 Community Health Systems 7  
 Microsoft 7  
 Santander Bank 7  
 ConnectOne Bank 7  
 Inspira Health Network 6  
 Dialog Direct 5  
 WSFS Bank 5  
 Comcast 5  
 JP Morgan Chase 5

**Years of Work Experience**  
 Less than 1 year 100+  
 1-2 years 300+  
 3-5 years 1000+

**what**  
Job title, skills, or company  
  
**where**  
City, state, or zip

Limit search to:  Job Title  Skills  Companies  Field of Study  
4,487 resumes



#### Search for your next candidate

Ready to contact a candidate? Purchase a [subscription](#) today!  
Create an Indeed account or login to get started.

#### Data Scientist - Recently Updated

Data Scientist - Utah State University

Data Scientist - CAPGEMINI

Utah State University, Master of Science

Updated: Sep 5

#### Data Scientist - Austin, TX - Recently Updated

Data Scientist - Cerebri AI

Data Scientist - Smarter Sorting

University of Manchester, Ph.D.

Updated: Sep 4

#### Data Scientist - Macomb, MI - Recently Updated

Data Scientist - Utility Company (confidential)

Data Scientist - TATA CONSULTANCY SERVICES (TCS)

Data Scientist - Boeing (TCS Contract)

OAKLAND UNIVERSITY, Bachelor of Arts

Updated: Sep 13

#### Data scientist - Singapore - Recently Updated

Data scientist - National health group singapore

Data Scientist

Nanjing university

Updated: Aug 31

#### Data Scientist - Silver Spring, MD - Recently Updated

Data Scientist - Jacobs Engineering Inc

D. S. Institute, LLC

# DATA

- ~85,000 RESUMES

- DATA ANALYSTS | DATA SCIENTISTS | DATA ENGINEERS

- Labeled Dataset

- FuzzyWuzzy

# EXAMPLE RESUME

## (My Resume)

### KARA BETHANY LIU

DATA SCIENTIST | ANALYST

✉ karabethanyliu@gmail.com ☎ +1-415-670-0729  
📍 San Francisco, CA  
in kara-bethany-liu-20840895

#### EDUCATION

National University of Singapore  
Economics 2015

CFA Level 1  
Completed CFA Level 1 Exam 2016

#### SKILLS

**PROGRAMMING:** Python, SQL, R, MongoDB, AWS, Tableau, Flask  
**PACKAGES:** Pandas, SKLearn, Matplotlib, Numpy, BeautifulSoup, Selenium, Plotly, Spacy  
**MODELING:** Regression / Classification, Unsupervised Learning / Clustering, Feature Engineering, NLP, Data Wrangling  
**COMMUNICATION:** Negotiations, Report Writing  
**LANGUAGES:** English (Native), French (Professional), German (Basic), Mandarin (Basic)

#### SUMMARY

Data scientist | analyst with experience negotiating Free Trade Agreements for Singapore, the financial services sector (tax and venture cap) and background in quantitative economics. My data projects include the use of statistics, hypothesis testing, and machine learning in order to analyze and predict human behavior.

#### EXPERIENCE

**JOBSCAN.CO**  
Data Science Intern

Nov./2018 to Current

Customizing a Named Entity Recognizer (NER) for job advertisements by training a convolutional neural network model with Spacy.

**METIS**  
Data Science Fellow

San Francisco  
July/2018

Developed several data projects as part of a selective 12-week data science immersive used to hone, expand, and contextualize skills in statistical modeling and machine learning.

- Job Matching: Deciding by Data - Built a model to determine what kind of data-related role (analytics, science, engineering) a resumé is best matched with. Cleaned data from Indeed.com and stored onto MongoDB. Utilised various NLP tools such as FuzzyWuzzy and W2V. Ran supervised categorical algorithms on AWS. Created a Flask WebApp on Heroku.
- Chatting with Bots - Analyzed and recommended suitable findings from messages sent to various chatbots deployed over Facebook by Prime.ai. Merging and cleaning Messages, Followers and Pages json files stored on AWS. Unsupervised learning (K-Means) and NLP. TSNE plots to visualize clusters.
- Crime in SF - Analyzed neighborhoods in SF that would require targeted police attention. AWS and PostgreSQL to manage data. Supervised categorical algorithms. Tableau and Flask web app visualizations.
- Predicting Hotel Prices on Booking.com - Scrapped data off booking.com using BeautifulSoup and Selenium. Python/Pandas for feature engineering and transformation. Seaborn to visualize data.

**MINISTRY OF TRADE AND INDUSTRY (SINGAPORE)**  
Assistant Director

Singapore  
Jan./2017 to June/2018

- Negotiated market access and rules of origin chapters in Free Trade Agreements (FTAs) for Singapore with countries such as Sri Lanka, New Zealand, Russia and Mexico.
- Analyzed and determined Singapore's industry and company interests to strategically front them in FTAs by consulting with companies from various industries and agencies and analyzing tariff data and the origin of Singapore goods.
- Negotiated the Fisheries Subsidies Agreement at the WTO, which aims to encourage sustainable fishing practices by monitoring and disciplining them.
- Handled other trade in goods issues like the implementation of the Trade Facilitation Agreement and recourse for violations in the WTO or any trade agreements.

**KPMG | QUANTERA GLOBAL**  
Transfer Pricing Intern | Consultant

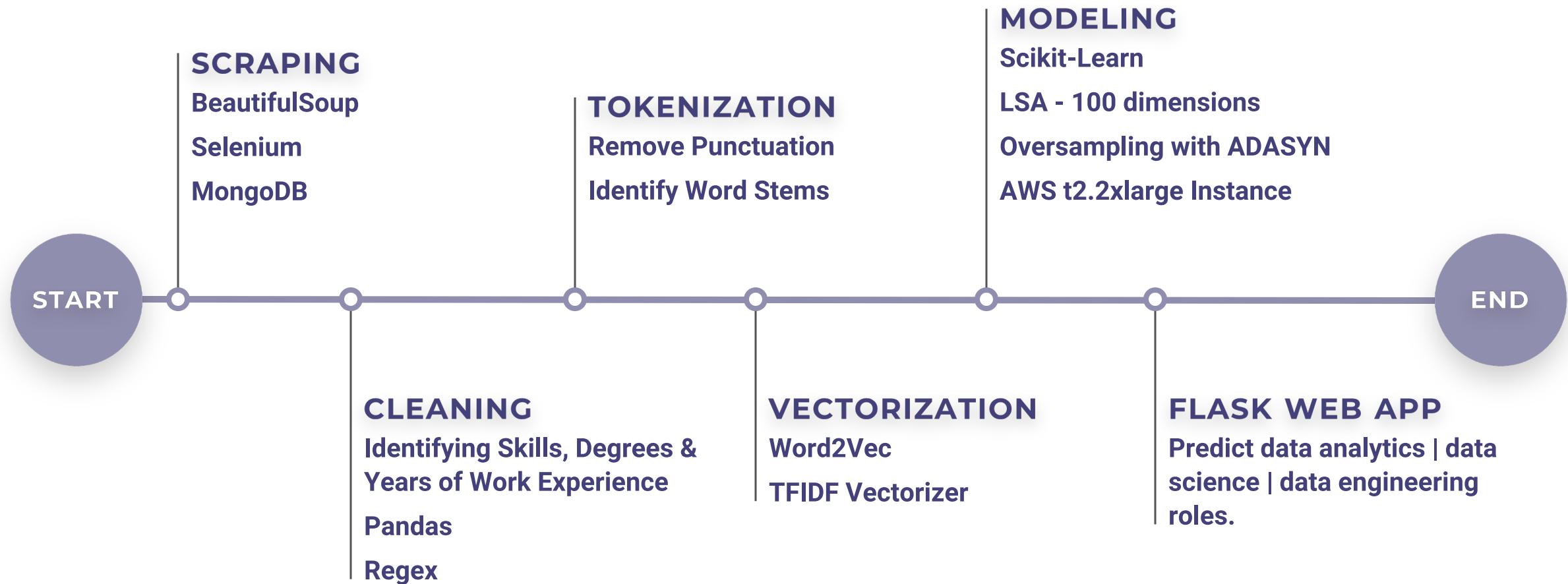
Singapore  
May/2015 to July/2015, Apr./2016 to Nov./2016

Advised MNCs on suitable profit levels for taxation purposes after conducting financial analyses using indicators best suited for the industries they operate in.

**IMPACT INVESTMENT EXCHANGE**  
Corporate Finance Apprentice

Singapore  
Aug./2015 to Dec./2015

# CLEANING AND PROCESSING



# BEST MODEL

Support Vector Classifier - TFIDF

- **RESULTS**

- 0.95 Accuracy score on Train Set
- 0.83 Accuracy score on Test Set
- Overfitting - need more data.
- 0.33 Dummy Score

- **OTHER MODELS**

- Gradient Boosting Classifier
- Log Regression Classifier
- With W2V



# INTRODUCING THE “WHICH DATA ROLE?” APP

<https://whichdatarole.herokuapp.com/>

Video

# THANK YOU FOR YOUR TIME!



## KARA BETHANY LIU

- [karabethanyliu@gmail.com](mailto:karabethanyliu@gmail.com)
- [/in/kara-bethany-liu-20840895/](https://www.linkedin.com/in/kara-bethany-liu-20840895/)