Career Recommendation Engine

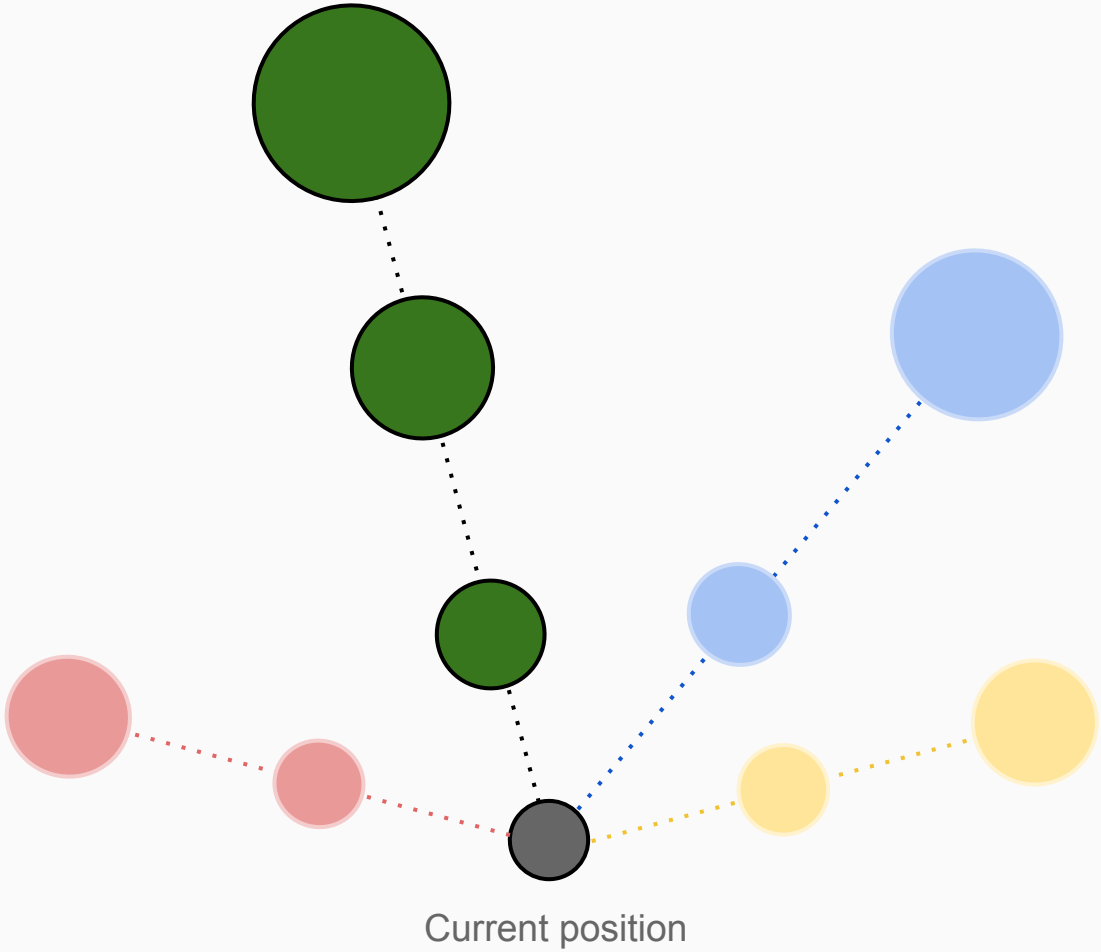
Data Science Immersive
Seattle - 01

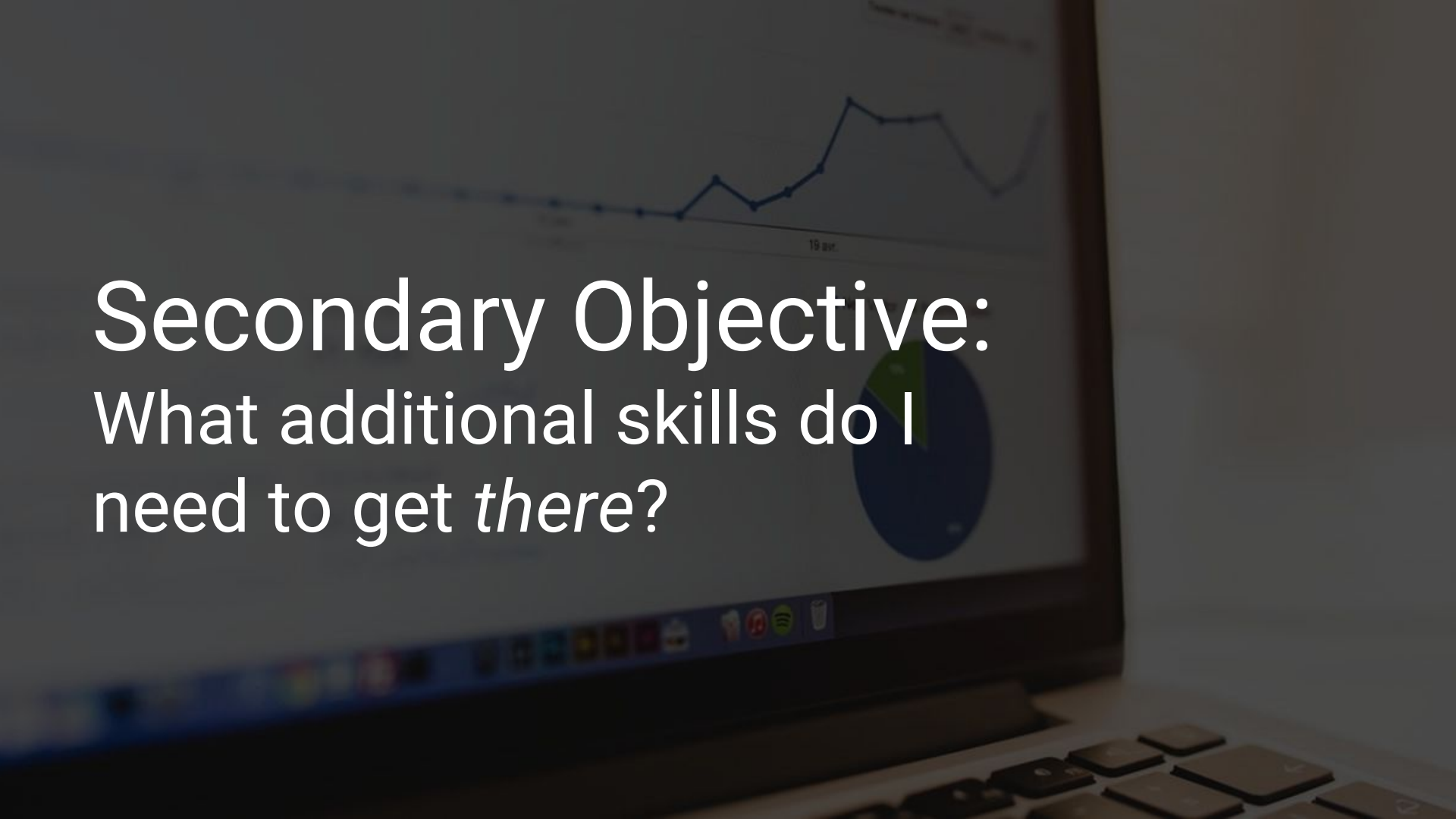
Primary Objective:

What is the next step in my career?

Primary Objective:

What career can I move into with my current skills?



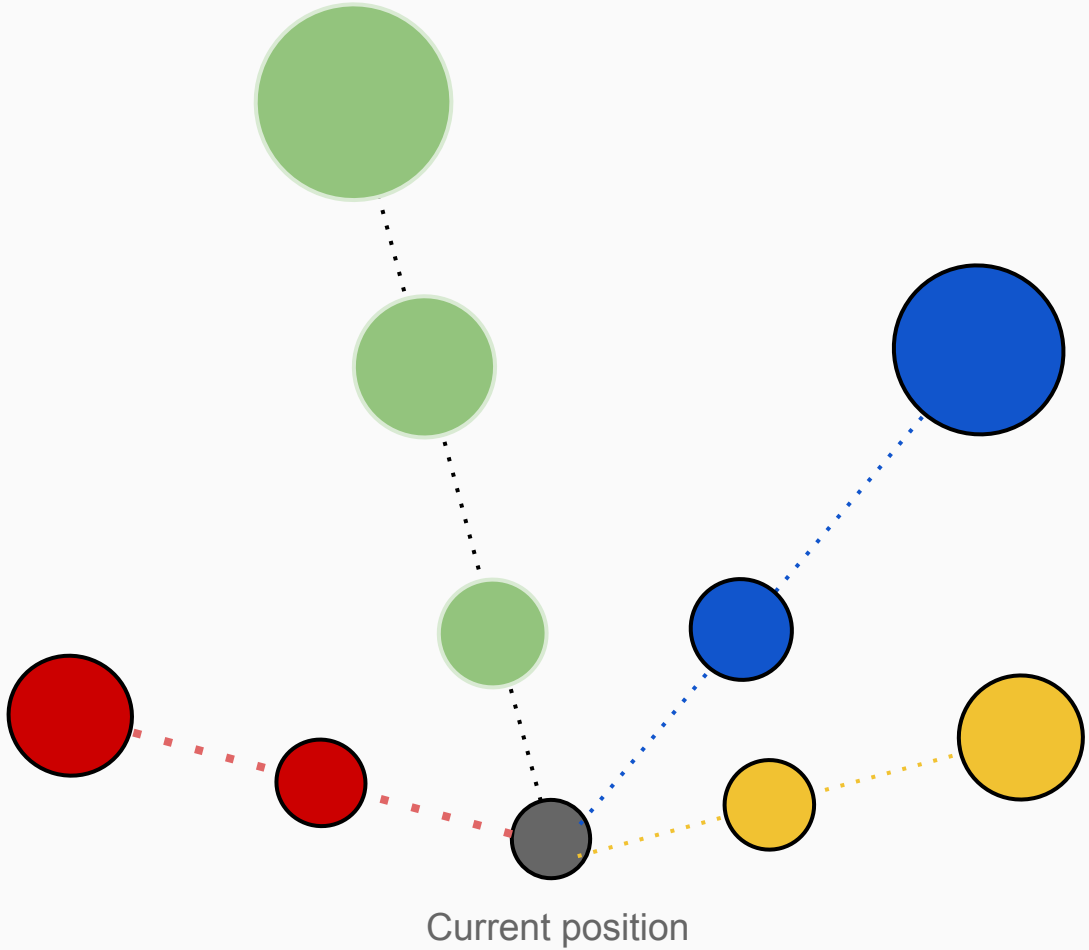


Secondary Objective:

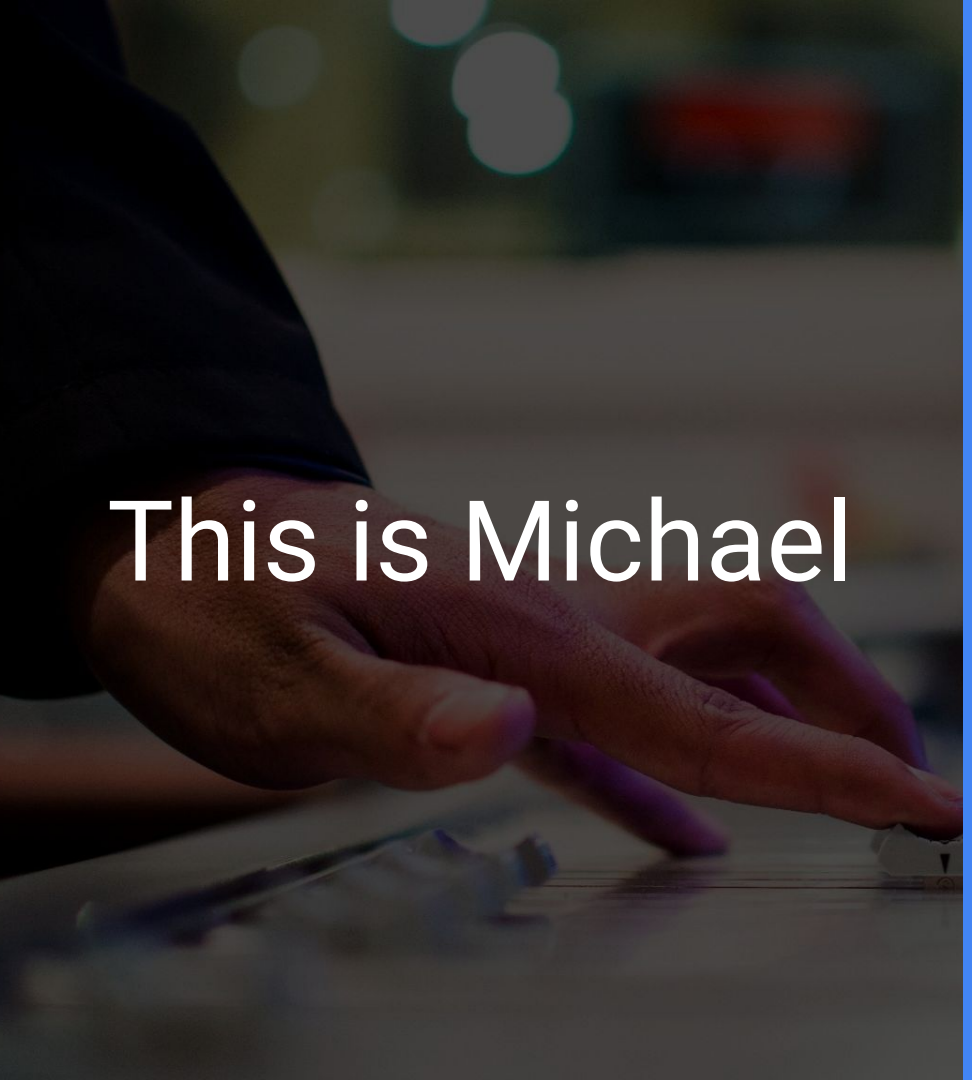
What additional skills do I need to get *there*?

Secondary Objective:

What new career can I move into, and what skills do I need to get there?



Product Demo



This is Michael

Previous positions:

- Founded small company in college
- Inventory analyst
- Brand specialist

Skills list:

- Marketing
- Business
- SQL/Python
- Data Analysis
- Negotiations

Recommendations for Michael: Within his field

Title	Score
Senior Product Marketing Manager	0.322
Assistant Professor	0.227
Product Management Consultant	0.216
Business Data Analyst	0.215
Senior Marketing Research Analyst	0.196

Business Data Analyst Skills	Rank
Platforms	1
Network	2
IRI	3
Analyzing	4
Software	5

Recommendations for Michael: Outside of His Skill Set

Title	Score
Assistant Director of HR	0.013
National Manager - Talent Acquisition	0.011
Research Assistant Intern	0.006
Assistant Human Resources Manager	-0.007
Recruiter Consultant	-0.007

Talent Acquisition Skills	Rank
Recruitment	1
Hiring	2
Swift	3
SOP	4
Outsourcing	5

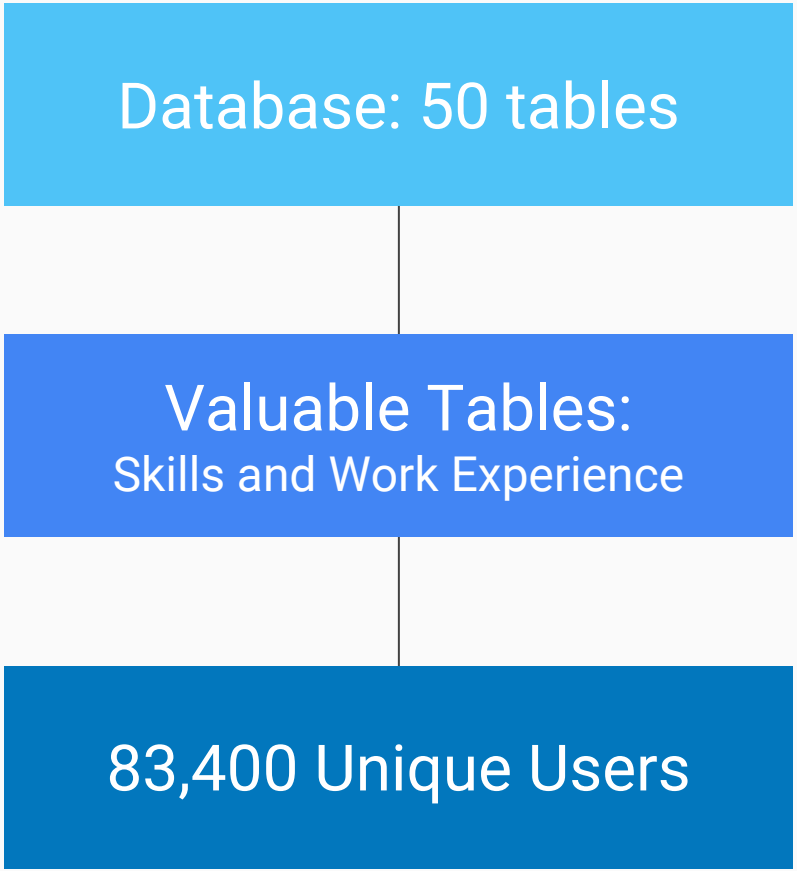
Exploratory Data
Analysis

Tables of Interest

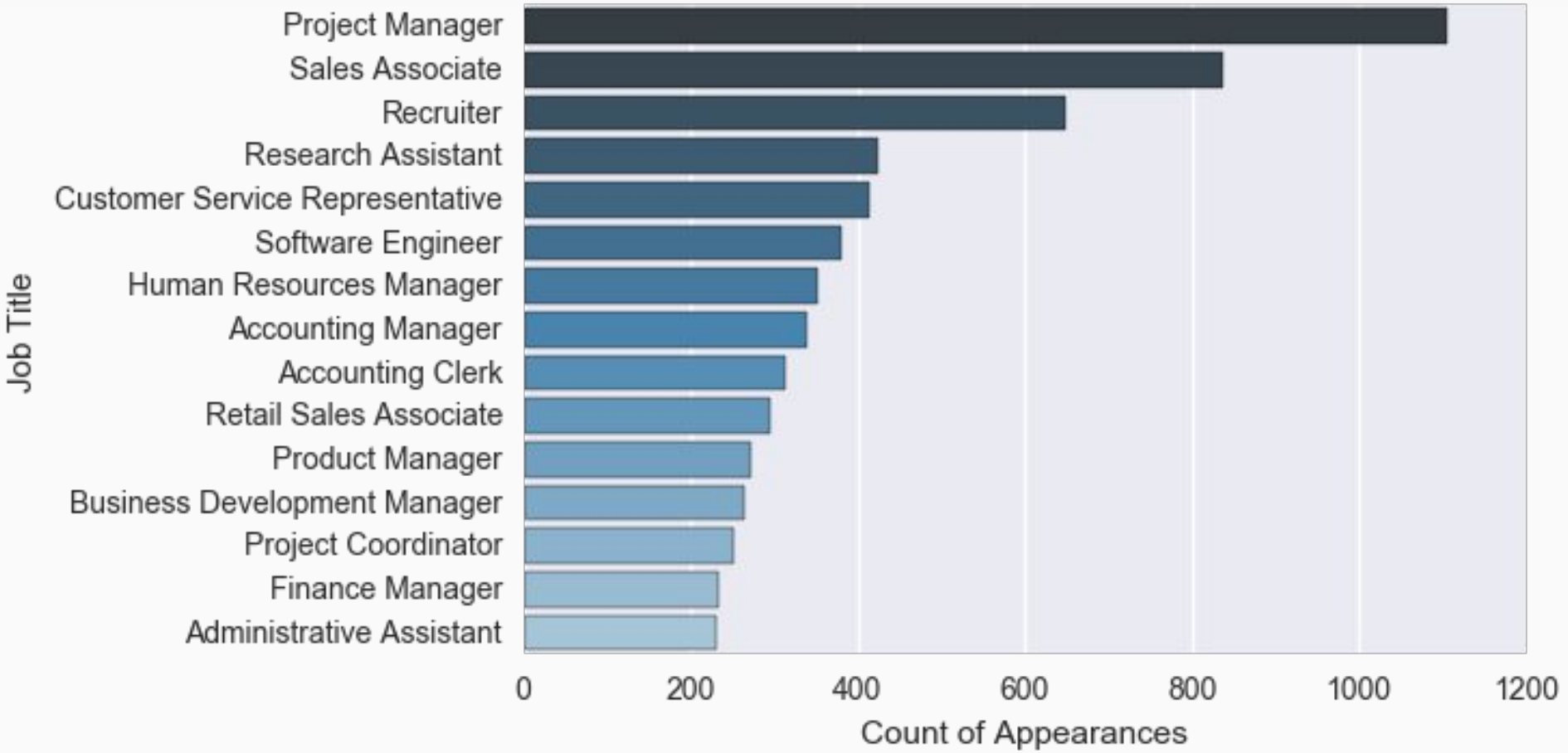
Database: 50 tables

Table Name	Number of Rows
work_experience	430,000
skills	140,000
resumes	260,000
education	140,000
certifications	22,000
awards	9,000
jobs	30

Applicable Data



Most Common Job Titles



Skills Word Cloud

Challenges

Ambiguous User Generated Data

Ambiguous and subjective resume content . For example:

- Account Manager vs Account Management
- Computer Scientist vs Software Engineer
- Buyer vs Vendor Manager
- Microsoft Word vs Word vs MS Word

Horizontal and Vertical ‘Landing Zone’

Horizontal: Similar fields or industries

- Accounting to Human Resources
- Mechanical Engineer to Software Engineer

Vertical: Experience level

- Junior to Senior
- Supervisor to Manager

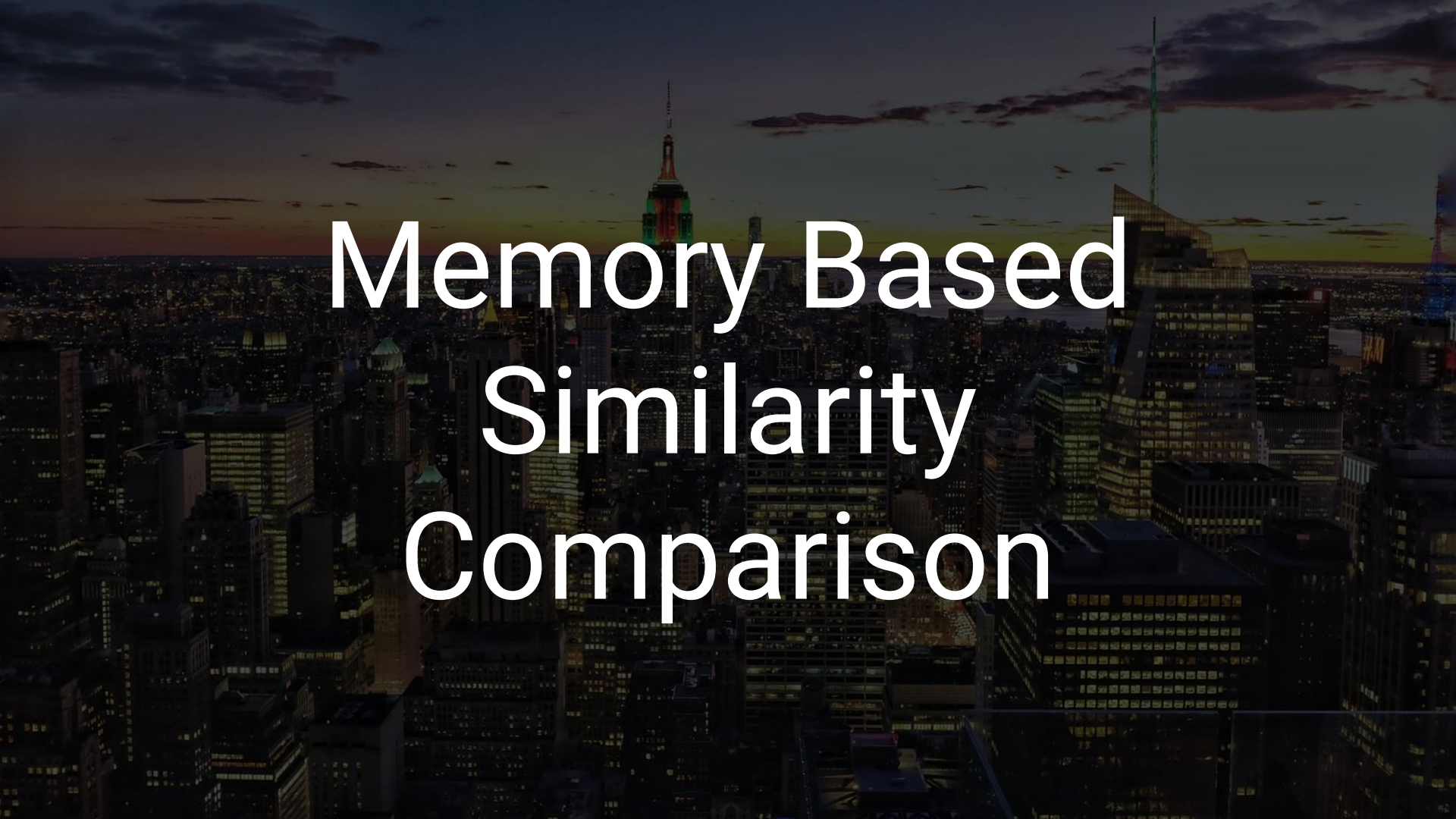
Targeting the Sweet Spot

We need to keep suggestions relevant , but not *too* relevant.

A computer engineer doesn’t want recommendations for software development positions. A more appropriate recommendation would be a technical product management position.

Two Solutions

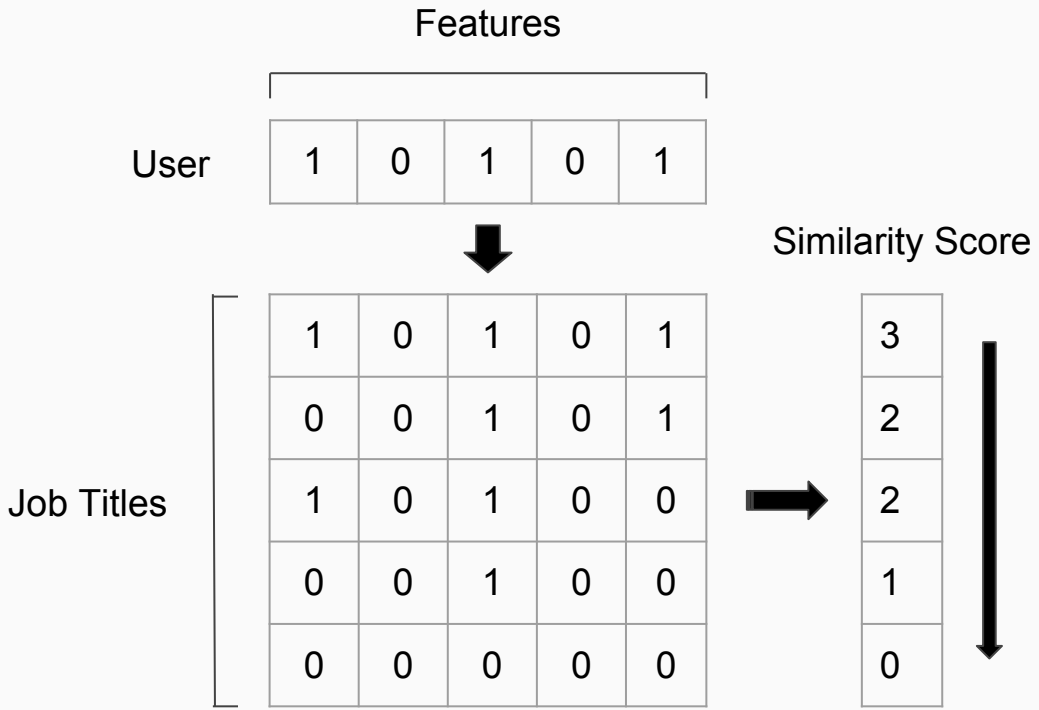
We Built Both


Memory Based
Similarity
Comparison

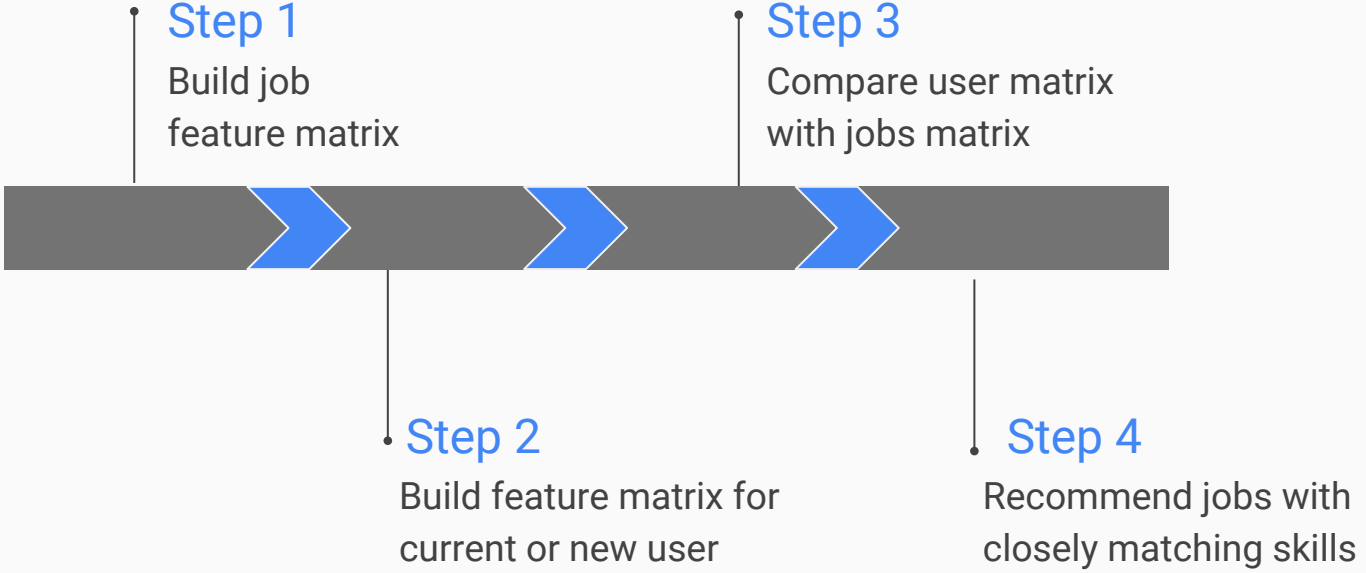
Similarity Comparison

- Content-based filtering recommendation engine
- Cosine similarity between user and jobs
- Tf-idf representation - vector space matrix
- Easy to implement and scale
- Simple but powerful

Cosine Similarity



Similarity Comparison - Implementation



Similarity Comparison - Data

- Extract work experience, skills and job titles from SQL
- Group skills by user
- Group user's work experiences and keep latest job title
- Create user table - Merge work experience and skills per user
- Create job table - Select most frequently occurring job titles with their description and skills

`User Table`

	exp_skills
resume_id	
623	• Reinvigorated revenue in South Florida and t...
624	Maximizing engagement and consumer loyalty by ...
625	Responsibilities • Maintain, extract, and prep...
626	• Identify and initiate contact with advertisi...
627	• Account analysis and preparation. • Customer...

`Jobs Table`

	exp_skills
title	
account coordinator	Coordinated activities in the New Business dep...
account executive	National Advertising sales of all iHeartMedia ...
account executive/recruiter	Built relationships with new business prospect...
account manager	Duties • Managing finances (A/R, A/P, Gen...
account representative	Telephone salesAdvertising sales of print yell...

Similarity Comparison - Vector Matrix

Job feature matrix

	abandoned	abbott	abiding	abilities	ability	able	abnormal	abreast	abroad	absence
title										
account coordinator	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000
account executive	0.0	0.0	0.0	0.012549	0.030606	0.026981	0.0	0.0	0.0	0.006602
account executive/recruiter	0.0	0.0	0.0	0.000000	0.024254	0.000000	0.0	0.0	0.0	0.000000
account manager	0.0	0.0	0.0	0.000000	0.033123	0.004326	0.0	0.0	0.0	0.006351
account representative	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.0	0.0	0.0	0.000000

User feature Vector

	abandoned	abbott	abiding	abilities	ability	able	abnormal	abreast	abroad	absence	...
resume_id											
625	0.0	0.0	0.0	0.0	0.024467	0.0	0.0	0.0	0.0	0.0	...

Similarity Comparison - Examples and Limitation

```
recommend_user(626)
```

```
Current Job Title :
    Digital Advertising Sales Executive

Top 10 Recommended Jobs :
    sales representative
    account executive
    business development manager
    marketing manager
    business development representative
    marketing consultant
    advertising sales representative
    senior account manager
    marketing coordinator
    account manager
```

```
recommend_newuser(resume)
```

```
Top 10 Recommended Jobs :
    product manager
    project manager
    program manager
    business analyst
    engineering manager
    software engineer
    senior business analyst
    software developer
    senior product manager
    senior project manager
```

Limitations

- Job recommendations that closely match the user's skill set
- Not able to suggest lateral career changes

Hybrid
Recommendation Engine

Hybrid Recommendation

Step 1
Build job feature matrix

Step 2
Create clusters on job titles

Step 3
Train model for cluster assignment of new user

Step 4
Create new user feature matrix

Step 5
Predict cluster for new user

Step 6
Suggest job titles and needed skills from each cluster to use

Hybrid Recommendation

Step 1 - Building job title feature matrix

- Select job titles that occurred at least three times
- Cleaned skills and descriptions to remove ‘filler’ words
- Job skills and descriptions vectorized
- Resulting matrix became the job title feature matrix

	personnel administration	labor	employment laws	staff recruitment	training	organizational management	employee relations	conflict resolution	regulatory compliance	powerpoint	...
0	0.0	0.091733	0.0	0.0	0.303941	0.0	0.0	0.0	0.0	0.065525	...
1	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.000000	...
2	0.0	0.000000	0.0	0.0	0.067620	0.0	0.0	0.0	0.0	0.000000	...

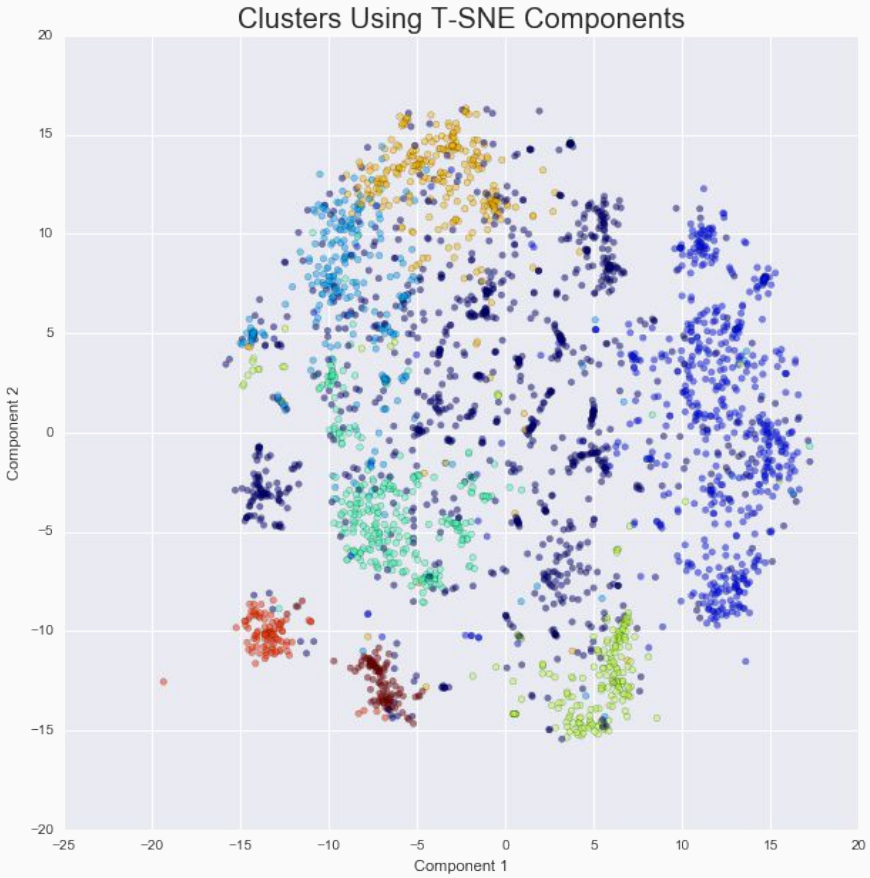
Hybrid Recommendation

Step 2 - Cluster job titles

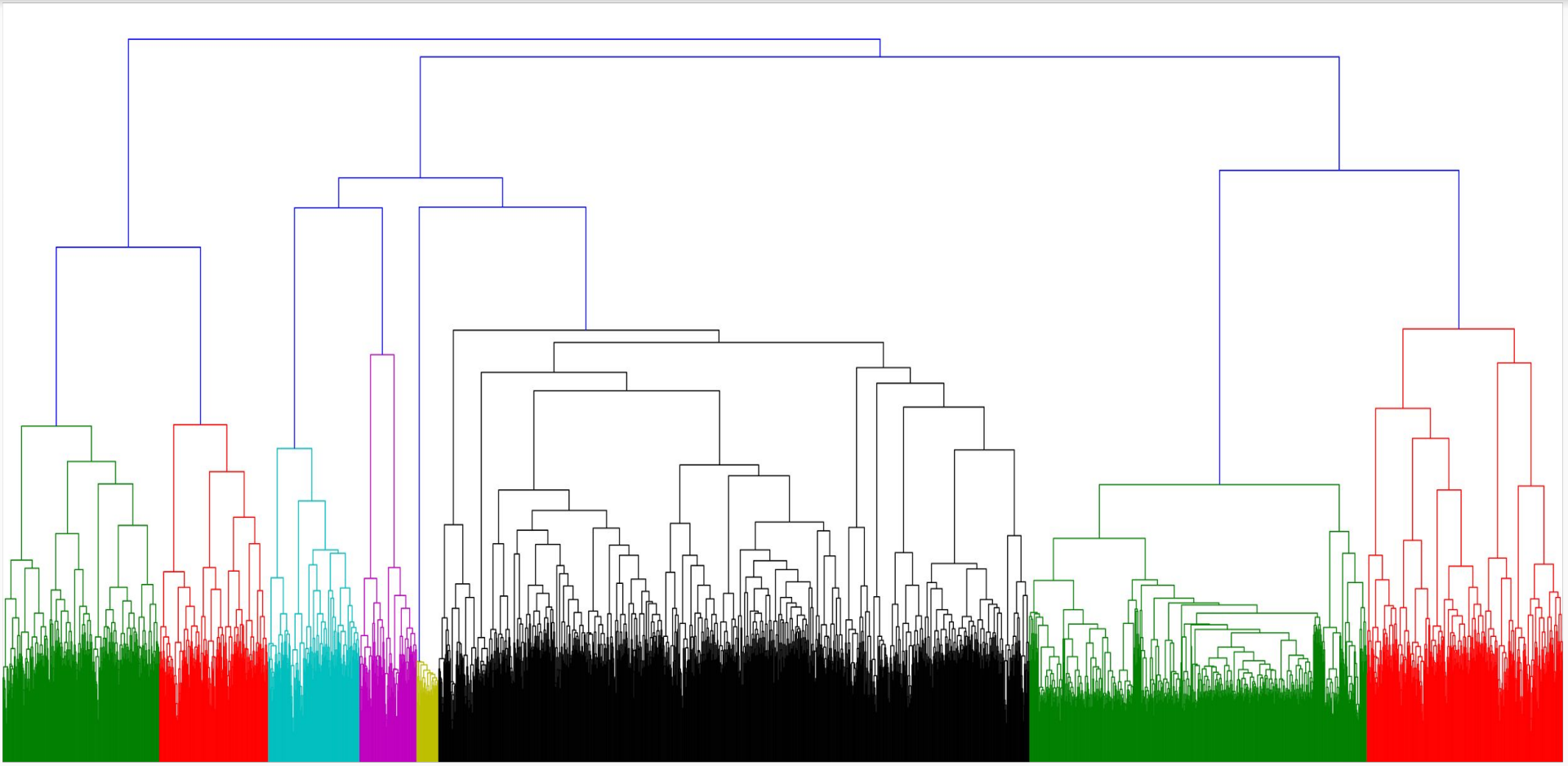
- Principal component analysis (PCA) performed on feature matrix
  - This reduces amount of features and allows the clustering to run more efficiently
- Resulting matrix from PCA fed into agglomerative clustering algorithm
- Clusters are groups of jobs most related by skills
  - Intracluster recommendations = vertical career movement (usually)
  - Intercluster recommendations = horizontal career movement (again, usually)

Job Title Clustering Scatterplot

- Cluster 0 - Business, health, academia, labor
- Cluster 1 - Software, data analysis, tech
- Cluster 2 - Management, administration
- Cluster 3 - Marketing and arts
- Cluster 4 - Finance
- Cluster 5 - PR, brand representative, service
- Cluster 6 - Recruiting
- Cluster 7 - Human Resources



Job Title Clustering Dendrogram



Hybrid Recommendation

Step 3 - Train logistic regression model

- A multinomial logistic regression model was trained
  - Features - job title matrix
  - Target - clusters
- Mean accuracy of model was 0.865

Hybrid Recommendation

Step 4 - Create new user feature matrix

- User’s resume text imported, vectorized, and decomposed using PCA
  - Same methods used as with job title matrix for consistency

Hybrid Recommendation

Step 5 - Assign cluster to new user

- Run user’s feature vector through logistic regression model to obtain cluster number prediction

Hybrid Recommendation

Step 6 - Create function to recommend job titles and skills

- Very similar to simpler model at this point
  - Uses cosine similarity
- Displays top suggestions for each cluster
  - This provides user with a diverse set of job titles
- Also displays the most common skills for each job title that the user is lacking

Results

- Due to subjectivity of the results, defining an accuracy metric for a recommendation system can be difficult.
- Initial observations indicate reasonable results

Future Improvements

- LightFM
- Deep Learning
- Advanced NLP to standardize user input
- More features and feature engineering
- Incorporate career path
- Leverage outside data sources

Client Recommendations

- Standardized input format for user data
- Improve jobs dataset features by leveraging external sources, ie: LinkedIn
- Invest heavily in feature engineering
- Incorporate distributed systems and big data architecture at scale

The team

Data Scientists with diverse backgrounds.

Corey Young

Data scientist with background in microbiology

Sadashiva Shetty

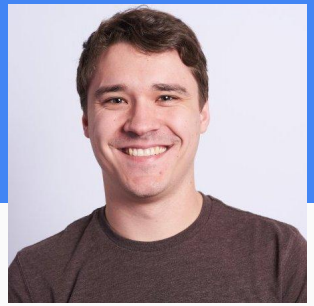
Digs deep into every problem

Matt Russell

Global traveller and blockchain enthusiast

Daniel Naylor

WSU history grad turned coder



Jesse Royalty

Outdoors enthusiast and college grad with a passion for data science