

Automated Resume Screeener

By: Aayushi Choudhary, Caleb Tran, Camden Weber, Charlie Hoose,
Justin Gong, Lara Papasian, Taylor Kim



"Resume screening is like looking for a needle in a haystack, except the needle sometimes disguises itself as a piece of hay."

- **Someone famous (totally not chatgpt)**

TABLE OF CONTENTS

01

Problem Statement

What were we trying to achieve?

02

Feature Extraction

What was being analyzed in the resumes?

03

Model Building

How was a subset of resumes selected?

04

Demonstration

Do any of our resumes pass the screener?



Problem Statement

Dataset

Dataset was taken from Kaggle containing 2400+ resumes in string and PDF formats. The data was unlabeled and organized into industries. Industries including accounting, engineering, chef, finance, teacher, and many more

Task

Project was framed as if we were a team of data scientists given these thousands of resumes and were asked to report ~15% of them for the next round.

Objective

Create a metric with specific features of the resumes to create a reasonable and explainable analysis of choosing our subset of resumes

resumes['Resume_str'][0]

" HR ADMINISTRATOR/MARKETING ASSOCIATE\nnHR ADMINISTRATOR Summary Dedicated Customer Service Manager with 15+ years of experience in service Management. Respected builder and leader of customer-focused teams; strives to instill a shared, enthusiastic commitment to customer service. Focused on customer satisfaction Team management Marketing savvy Conflict resolution techniques Training and development Skilled multiions specialist Accomplishments Missouri DOT Supervisor Training Certification Certified by IHG in Customer Loyalty and Marketing by S de General Manager Training Certification Accomplished Trainer for cross server hospitality systems such as Hilton OnQ , Micros Opera PMS eservation System (ORS) , Holiday Completed courses and seminars in customer service, sales strategies, inventory control, loss prevention, safety dership and performance assessment. Experience HR Administrator/Marketing Associate\nnHR Administrator Dec 2013 to Current Co , State Helps to develop policies, directs and coordinates activities such as employment, compensation, labor relations, benefits, training, and prepares employee separation notices and related documentation Keeps records of benefits plans participation such as insurance and pension plan, person s hires, promotions, transfers, performance reviews, and terminations, and employee statistics for government reporting. Advises management in apployee relations issues. Administers benefits programs such as life, health, dental, insurance, pension plans, vacation, sick leave, leave of absence, e. Marketing Associate\nx@o Designed and created marketing collateral for sales meetings, trade shows and company executives. Managed the in-ho consisting of print and media collateral pieces. Assisted in the complete design and launch of the company's website in 2 months. Created an official ock to facilitate interaction with customers. Analyzed rating and programming features of committees to evaluate the effectiveness of marketing c



02

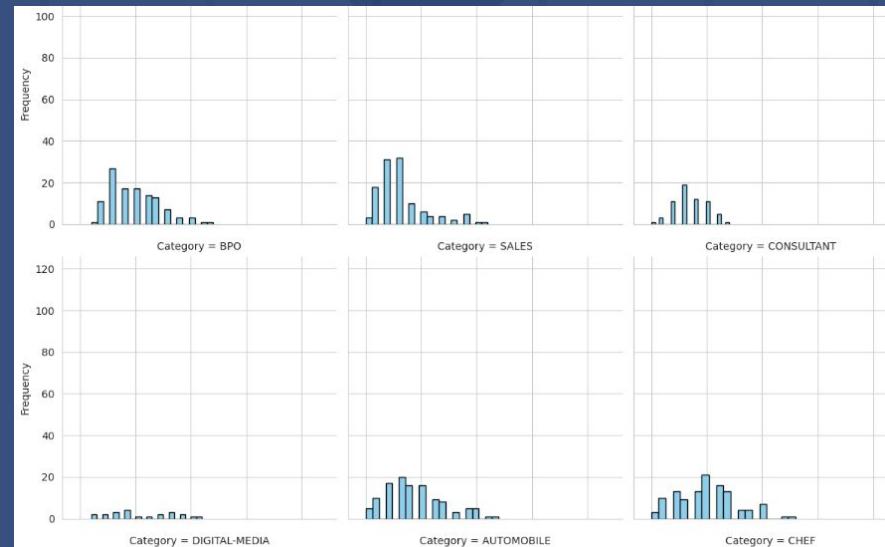
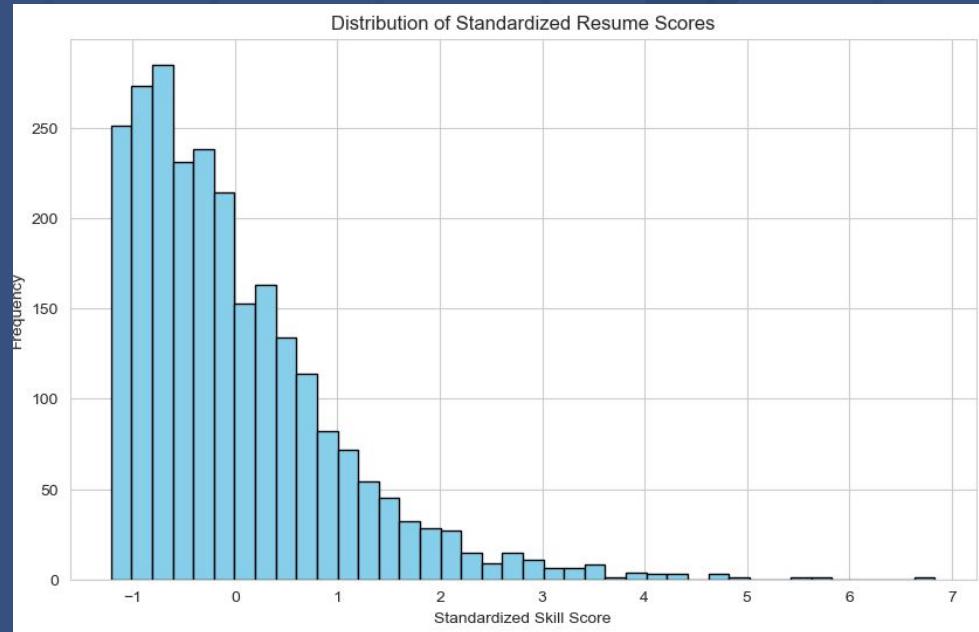
FEATURE EXTRACTION

Skills Feature

- **Goal:** Score each individual resume based on the skills it listed



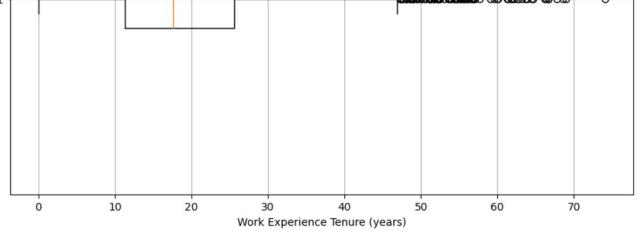
Distribution of Standardized Resume Scores



'FINANCE': [

"Financial analysis", "Accounting", "Financial reporting", "Budgeting", "Forecasting", "Risk management", "Investment analysis", "Portfolio management", "Financial planning", "Taxation", "Auditing", "Internal controls", "Compliance", "Regulatory reporting", "Corporate finance", "Mergers and acquisitions", "Capital markets", "Derivatives", "Equities", "Fixed income", "Financial modeling", "Valuation", "Business analysis", "Cash flow management", "Cost accounting", "Managerial accounting", "Financial systems", "ERP systems", "SAP", "Oracle", "Hyperion", "QuickBooks", "Xero", "Financial software", "Data analysis", "Quantitative analysis", "Statistical analysis", "Financial projections", "Strategic

Work Experience Feature



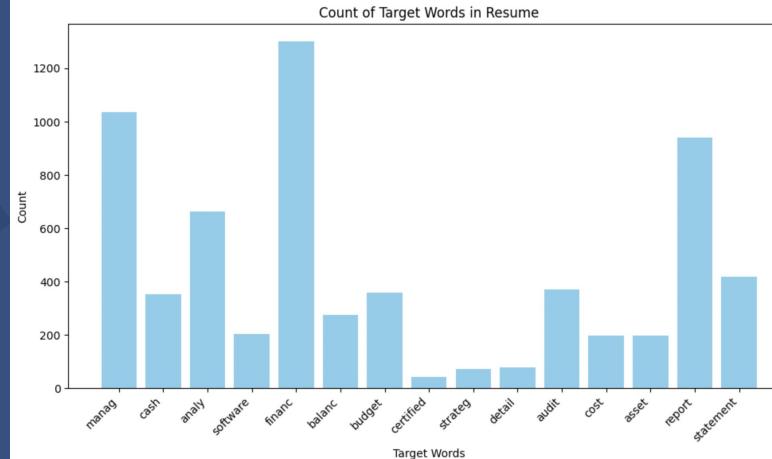
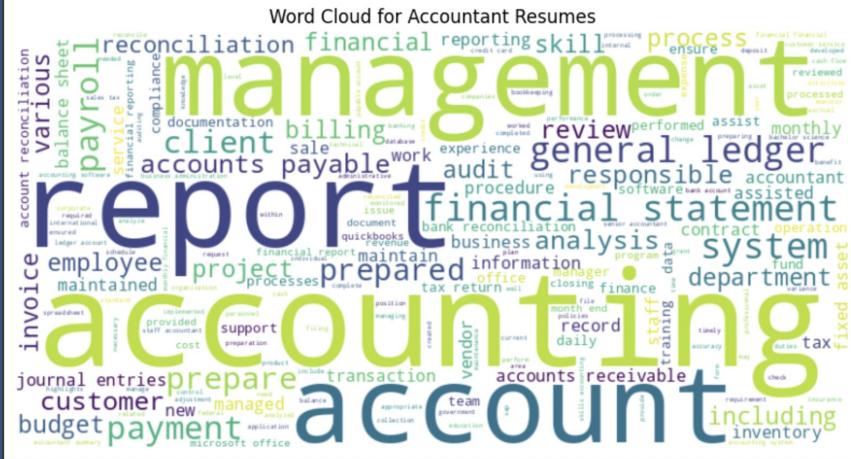
- Goal: find the total amount of years of work experience on someone's resume



count	2484.000000
mean	19.710959
std	11.983834
min	0.000000
25%	11.342466
50%	17.591781
75%	25.600000
max	74.054795

- Approach: scrape the resume string for dates of work tenure (e.g., "June 2020 to February 2023") per job, find the difference in days between the two, divide by 365, then do so for every job on the resume.

Industry Keywords

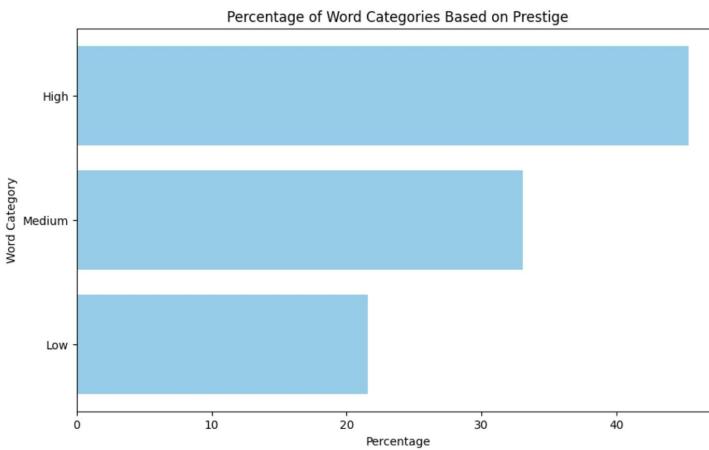


```
target_words = ['manag', 'cash', 'analy', 'software', 'financ', 'balanc', 'budget',
               'certified', 'strateg', 'detail', 'audit',
               'cost', 'asset', 'report', 'statement']
```

ID	Resume_str	Resume_html	Category	Weighted_Points
1808	15906625 ACCOUNTANT Professional Summary...	<div class="fontsize fontface vmargins hmargin...>	ACCOUNTANT	18
1809	30361788 ACCOUNTANT Summary Innovat...	<div class="fontsize fontface vmargins hmargin...>	ACCOUNTANT	15
1810	21031285 ACCOUNTANT Highlights ...	<div class="fontsize fontface vmargins hmargin...>	ACCOUNTANT	17
1811	25749150 ACCOUNTANT Professional Sum...	<div class="fontsize fontface vmargins hmargin...>	ACCOUNTANT	19
1812	12338274 ACCOUNTANT Summary To pursue...	<div class="fontsize fontface vmargins hmargin...>	ACCOUNTANT	18

Work Position Feature

- Goal: Extract the most recent work position
 - Approach: Parse through the HTML and find all sections relating to jobs and extract the job title
 - Application: Rank the prestige of the job titles and use it for our model!



Word Cloud of Most Common 50 Words

A word cloud visualization where the size of each word represents its frequency or importance. The words are color-coded by category. The most prominent words are 'business', 'director', 'consultant', 'manager', 'sales', and 'specialist'. Other visible words include 'accountant', 'construction', 'supervisor', 'marketing', 'public', 'engineering', 'information', 'analyst', 'officer', 'intern', 'operations', 'coordinator', 'job', 'sr', 'found', 'relations', 'hr', 'technology', 'customer', 'service', 'associate', 'assistant', 'representative', 'title', 'lead', 'designer', 'instructor', 'digital', 'engineer', 'not', 'technician', 'advocate', 'banking', 'executive', and 'teacher'.

Examples of job prestige:

- High: Manager, Lead
 - Medium: Engineer, Accountant
 - Low: Intern

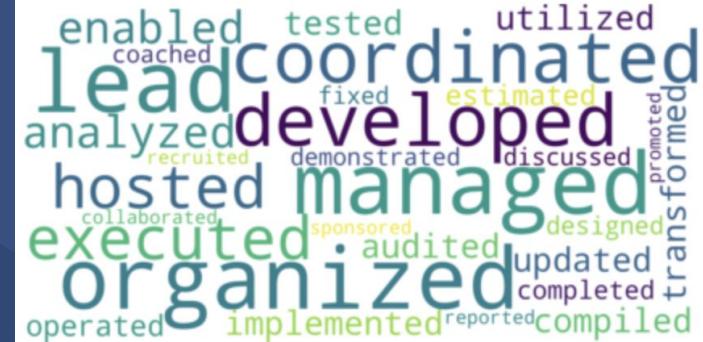
Sentiment Analysis

- Goal: Determine whether the resume sections display the desired tones of *professionalism* and *confidence*.
- Approach: Perform NLP using 'textblob' to count the occurrences of professionalism and confidence keywords in each resume.
- Application: Use results to measure strength of candidate's writing quality.

```
print(experience_metrics_df.head(11))
```

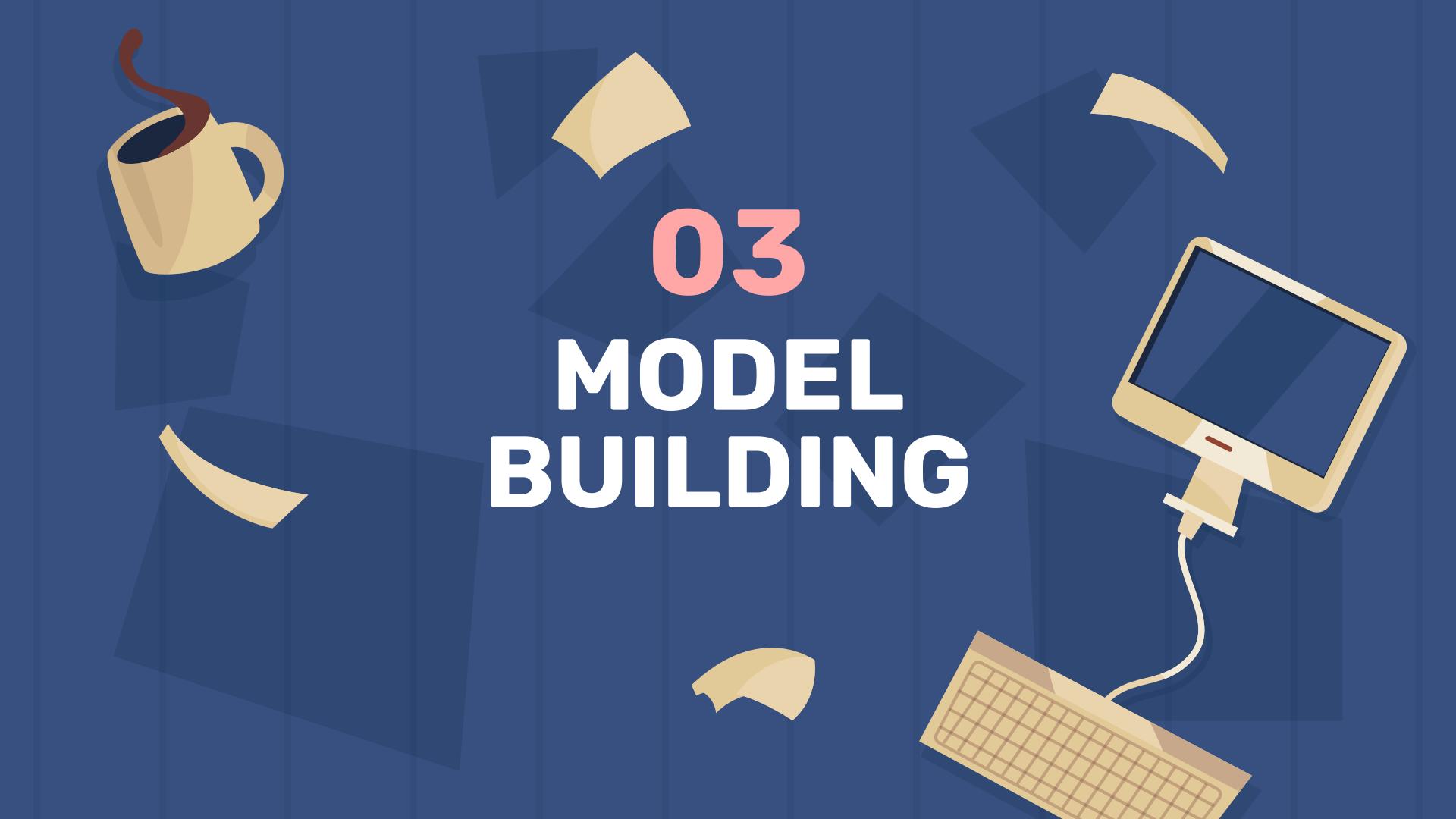
	Resume_ID	Professionalism_Count	Confidence_Count
0	16852973	6	4
1	22323967	23	17
2	33176873	11	11
3	27018550	3	2
4	17812897	4	5
5	11592605	11	12
6	25824789	6	5
7	15375009	31	12
8	11847784	16	3
9	32896934	2	2
10	29149998	14	1

Professionalism Word Cloud



Confidence Word Cloud

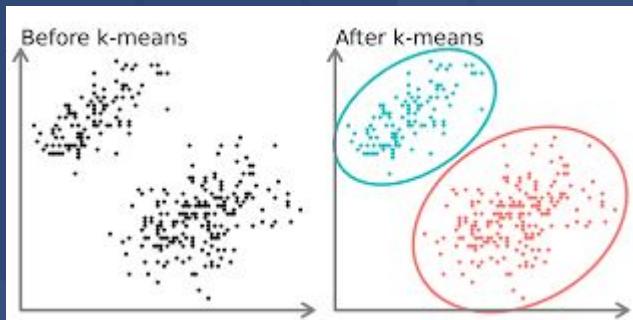


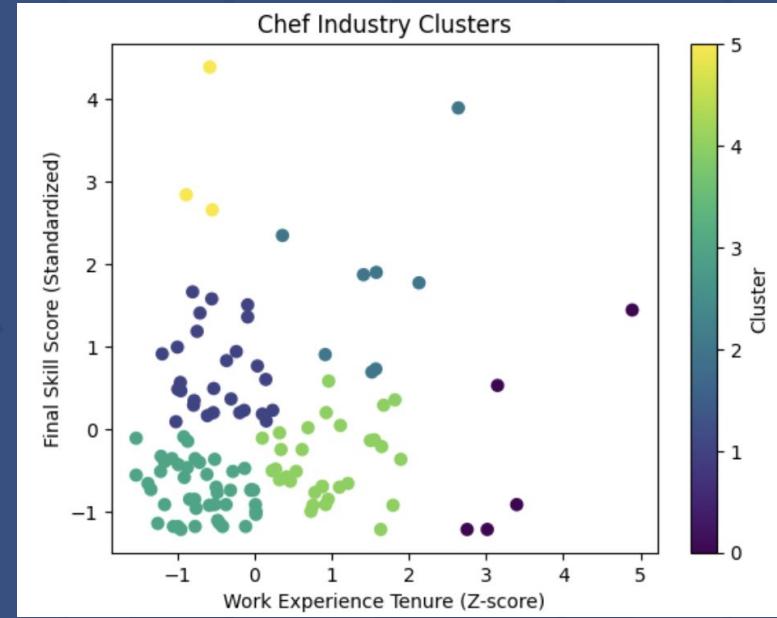
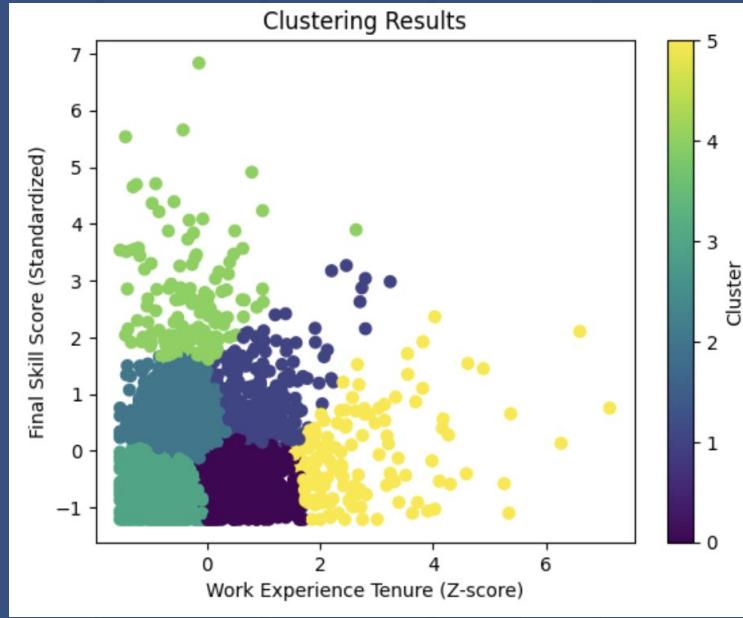


03 MODEL BUILDING

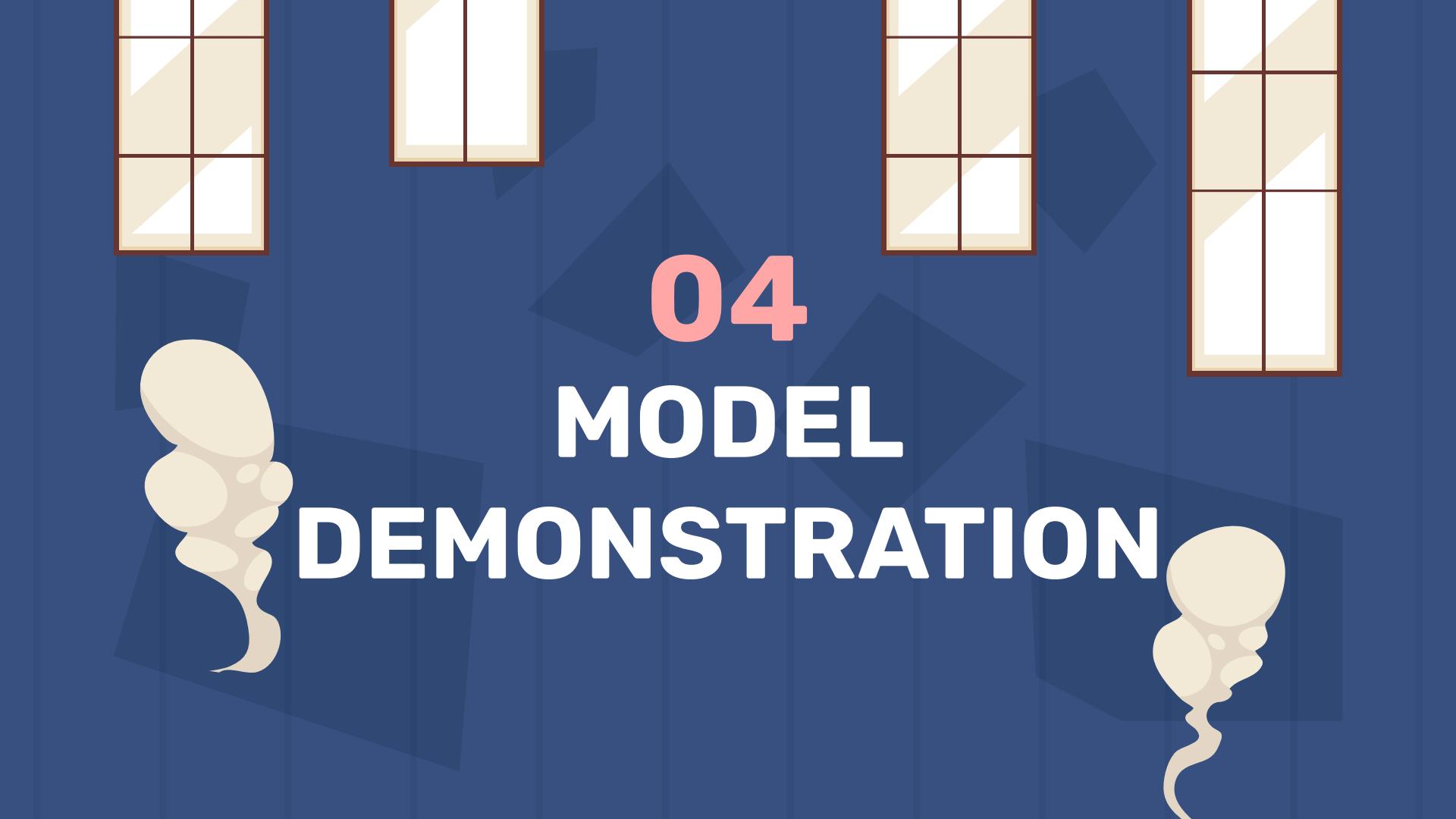
K-Means Clustering Modeling

- Unsupervised learning model (Unlabeled data)
- Sorts data points into clusters based on distance metrics
- Clusters: 6
 - Chosen to obtain ~15% acceptance
- Selected industries: Engineering, chef, accountants





	work_experience_tenure_zscore	final_skill_score_standardized	in_top_university	Professionalism_Count	Confidence_Count	recent_job_level_numeric	Weighted_Points
cluster							
1	0.097384	0.375005	0.238095	7.571429	5.666667	1.714286	24.285714
2	0.056838	0.710161	0.233333	3.233333	1.133333	1.766667	28.200000
3	-0.083444	1.391650	0.322581	9.516129	5.354839	2.064516	31.387097
4	-0.334723	0.178106	0.157895	2.789474	1.473684	1.315789	17.315789
5	0.102925	0.558985	0.153846	17.000000	7.230769	2.230769	30.461538
6	0.795810	2.261046	0.500000	17.750000	22.750000	2.250000	30.500000



04 MODEL DEMONSTRATION



Method 1:

- Insert resume into training data and retrain clusters
- Reselect cluster of choice

Justin's Resume

JUSTIN GONG	
Links	Address
linkedin.com/in/justingong03 gitlab.com/justingong03	609-651-9901 10 Meadowlark Dr, Plainsboro, NJ 08536
EDUCATION	
University of California, Los Angeles	Los Angeles, 2021 - 2025
Mathematics and Statistics (B.S.), Minor in Data Science Engineering	
GPA: 3.9/4.0, Cumulative: Jan 2023	
Relevant Courses:	
Python & Technologies for Data Science, Database Management System, Data Structures & Algorithms (Computer Science),	
Introduction to Machine Learning, Statistical Theory, Mathematical Foundations in Data Science & ML, Statistical Models & Data Mining	
EXPERIENCE	
AT&T - Data Engineer Intern (Technology Development Program)	June 2023 - Sept. 2023
Ingested IBM HR records using PostgreSQL and Flask, allowing users to retrieve socially connected field-based reports	
Researched and implemented data structures including GIN, GIST, and R-tree indices to maximize real-time search and computation speed for large-scale datasets, resulting in a 50% reduction in search time for specific queries.	
Experimented with proof of concept integration of NLP models, which expanded tool-based search results by 20%.	
Developed automated sentiment classification and priority application by fine-tuning transformer models from Huggingface, achieving a 47% increase in ticket classification accuracy and a 10% reduction using our item-poker patch detection.	
Data Science Unites - Machine Learning Projects Lead	January 2024 - present
Monitored team of 8 to leverage convolutional neural networks with TensorFlow and PyTorch to architect facial recognition models for various applications, such as attendance tracking and access control systems.	
Collaborated with cross-functional teams to implement and validate machine learning models across multiple domains.	
Crafted multi-platform deliverables to present to 50+ individuals including Sr. Stakeholders, Tableau dashboard, and PowerPoint.	
DataRobot at UCLA - Vice President	March 2023 - present
Managed an organization of over 200 members, leading 10+ concurrent internal projects to achieve club-wide goals.	
Spearheaded the creation of a comprehensive executive project framework to organize internal club-wide tasks.	
OKTV - Volunteer Data Consultant	Sept. 2022 - March 2023
Utilized Python to clean and normalize data pipeline of user app data, improving the integration into modeling process.	
Developed ML-based collaborative filtering recommendation system correlated with an ~15% increase in user app retention.	
PROJECTS	
NBA Player Analysis Predictor, 2022-2023 Season	July - August, 2022
Developed end-to-end machine learning pipeline in Python, employing web scraping, data cleaning, exploratory data analysis, feature engineering, data preprocessing, model building (feature selection, engineering, etc.), hyperparameter tuning, and deployment analysis.	
Corporated multiple regression algorithms (SVR, Ridge, LASSO, Elastic Net, and more), achieving 75% prediction accuracy.	
Created a user-friendly web application using Streamlit, integrating a heatmap visualization for player performance and more.	
Music Recommendation System: KNN-Based Playlist Enhancer	June - July 2023
Designed a recommendation system for music tracks and AI建模 on 50+ million data points using Python and Java.	
Developed sophisticated model using K-Means, Apriori algorithm and item-based matrix factorization, achieving 75% user success rate.	
MISCELLANEOUS	
Tools: Python (Sqlite, Tensorflow, PyTorch), SQL, Tableau, C - Java, JavaScript, Microsoft Office Tools (Excel, PowerPoint), Git, Flask, Shiny.	
Skills: Advanced data cleaning, analysis, and visualization; database management; query optimization; and analysis; machine learning, deep learning, and generative AI; A/B Testing and business analytics; object-oriented programming and basic web development.	
Honors: DataVis National Finalist (American Statistical Association - 2023), Dean's Honors List (UCLA - 2023)	
Hobbies/Interests: Basketball, traveling, journaling, music, weightlifting, hiking, spikeball, hiking	

Method 2:

- Utilize previously trained clusters
- Predict cluster of new resume and reject/accept

Justin's Feature Scores

justin_resume	
work_experience_tenure_zscore	-1.07692
final_skill_score_standardized	1.59541
in_top_university	1.00000
Professionalism_Count	13.00000
Confidence_Count	3.00000
recent_job_level_numeric	1.00000
Weighted_Points	25.95792
dtype:	float64

Justin's Cluster

3

THANK YOU!!!