Exploring the Demand for Tech Skills

Theodoxea Kwapong

Table of contents

1 INTRODUCTION

1.1 Background

The demand for tech skills in the job market has been steadily increasing, according to the Bureau of Labor Statistics (BLS) and other industry sources. While there are numerous job openings in tech, the necessary skills might not be readily available in the workforce, emphasizing the need for education and training. This highlights that students have to acquire relevant skills to meet the growing demand for tech-related positions.

This project seeks to address this gap by analyzing which tech skills are in high demand, particularly for entry-level positions, and tracking how these skills' relevance evolves over time. By combining data from job postings (via Adzuna API) and Google search trends (via Google Trends API), this report will explore the alignment between job market demand and public interest in these skills.

The complete analysis and code can be found on the project's GitHub repository: https://github.com/TKwapong/project

The insights are valuable for students entering the job market to identify the most relevant skills to focus on and influencing classes they take and certifications they consider ultimately preparing for roles in the technology sector.

2 METHODOLOGY

Two primary data sources were used:

1. Job postings fetched from the Adzuna API
2. Search trends for popular tech skills using Google Trends API through the R package "gtrendsR"

Table 1: Summary of Data Sources Used

Source	Description	Key_Variables	Limitations
Adzuna API	Job postings data for various tech-related roles across the US	Job title, location, salary, description, job level	Limited to jobs from Adzuna; not exhaustive of the job market
Google Trends	Search interest for tech skills over 2018-2023	keywords, interest over time, regions	Limited to a maximum of 5 skills per query

The analysis involved:

- Fetching job postings data for various tech-related roles using Adzuna API and reading queried data into csv for easier retrieval later.
- Cleaning and refining the data to extract insights such as job levels and salaries.
- Performing text analysis on job descriptions to identify key skills and trends.
- Exploring Google Trends data to understand the popularity of tech skills over time.
- Comparison analysis between extracted job posting skills frequency and Google Trends data to determine how well job market demand aligns with public interest.

2.1 Data Sources

2.1.1 Adzuna API

aggregator of job postings data worldwide

The following roles were queried in order to retrieve substantial job postings data from Adzuna API:

- *Data Analyst - Software Engineer - IT Support Specialist - Product Manager - Cybersecurity Analyst - Business Analyst - UX Designer - UI Designer - Data Scientist - Machine Learning - Data Engineer*

To access the Adzuna API, you can sign up here to obtain an **API Key** and **App ID**. The Adzuna API allows a maximum of 50 results per query, and we looped and iterated through 5 pages to gather more comprehensive data across multiple tech roles.

Table 2: Summary of some Key Variables

Variable	Description	Relevance
id	Unique identifier for the job posting	Key for data uniqueness
description	Text description of the job role	Used for text analysis to extract skills
salary_min	Minimum salary for the job posting	Used to calculate average salary
salary_max	Maximum salary for the job posting	Used to calculate average salary
created	Date the job posting was created	Helps filter recent postings
title	Title of the job position	Used to extract job roles
job_level	Categorized level of the job (Entry, Mid, Senior)	Used for analysis by job level

2.1.1.1 Data Management

Since there are quality issues with getting data from the Web, the data was cleaned and structured to facilitate analysis:

- Subsetted job postings to 2023 to get recent data
- Standardized job title and description text and used *Job Description* and *Title* columns to categorize job postings by level: Entry, Mid, and Senior
- Handled missing values by inserting NA for irrelevant columns where applicable. Columns with NA values weren't relevant to project analysis
- Transformed job descriptions column text to lowercase and removed extra spaces in order to prepare text data for accurate tokenization and analysis.

Here's a snapshot of data fetched and cleaned for analysis

```r
# Inspect the data structure
glimpse(job_data_refined)
```

```
Rows: 2,747
Columns: 20
$ salary_min         <dbl> 24879.17, 108049.37, 109494.97, 64252.31, 73924.34~
$ id                 <dbl> 4939409850, 4958151516, 4958151515, 4916955770, 49~
$ `__CLASS__`        <chr> "Adzuna::API::Response::Job", "Adzuna::API::Respon~
$ description        <chr> "ONCOLOGY DATA ANALYST Baltimore, MD SINAI CANCER ~
$ created            <dttm> 2024-11-14 17:40:00, 2024-12-01 12:50:49, 2024-12~
$ salary_max         <dbl> 24879.17, 108049.37, 109494.97, 64252.31, 73924.34~
$ redirect_url       <chr> "https://www.adzuna.com/land/ad/4939409850?se=9psy~
$ title              <chr> "ONCOLOGY DATA ANALYST", "Language Data Analysts",~
$ salary_is_predicted <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ adref              <chr> "eyJhbGciOiJIUzI1NiJ9.eyJpIjoiNDkzOTQwOTg1MCIsInMi~
$ contract_time      <chr> "full_time", "full_time", "full_time", NA, NA, NA,~
$ longitude          <dbl> -76.60944, -77.26776, -77.26776, -73.97850, -118.6~
$ latitude           <dbl> 39.29688, 38.93568, 38.93568, 40.75538, 34.16970, ~
$ query              <chr> "data analyst", "data analyst", "data analyst", "d~
$ category_label     <chr> "Healthcare & Nursing Jobs", "IT Jobs", "IT Jobs",~
$ location_name      <chr> "East Case, Baltimore", "Wolf Trap, Halifax County~
$ company_name       <chr> "LifeBridge Health", "Leidos", "Leidos", "City of ~
$ contract_type      <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
$ job_description    <chr> "oncology data analyst baltimore, md sinai cancer ~
$ job_level          <chr> "Entry Level", "Entry Level", "Entry Level", "Seni~
```

2.1.2 Google Trends

Google Trends data was used to explore the popularity of top skills identified during the text analysis phase. Google search frequency trends data was obtained through the R package "gtrendsR" in order to capture public interest in captured skills, for January 1, 2018 to December 31, 2023.

3 ANALYSIS

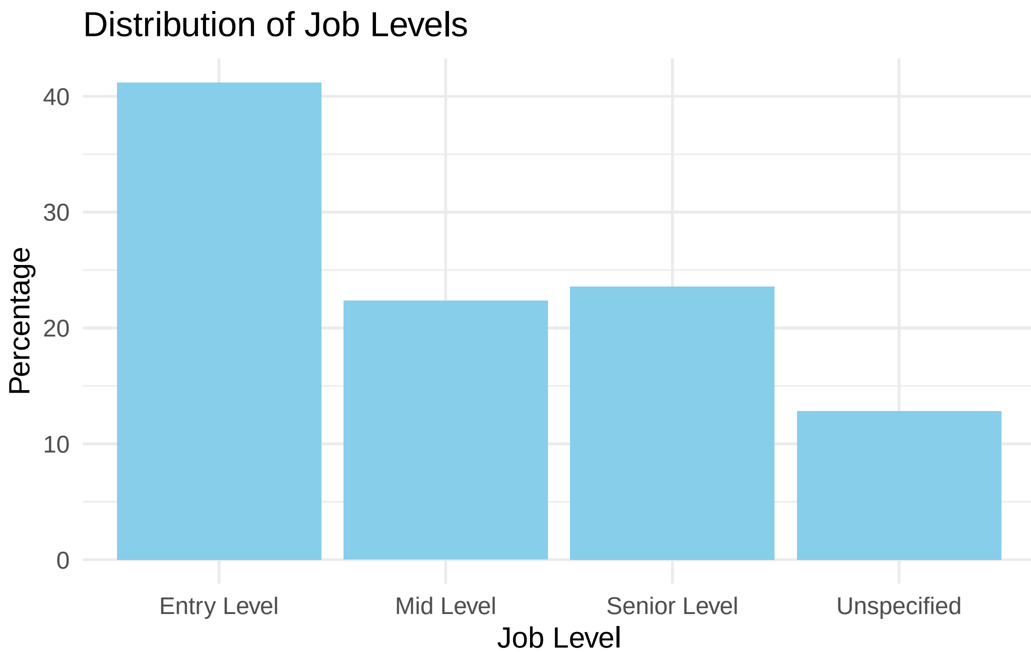
3.1 Data Analysis: Insights from Job Postings

Performed initial data analysis to visualize and understand the extracted API data better.

3.1.1 Distribution of Job Levels

Job Level Distribution: Analyzed the proportion of job postings at each level to better understand distribution of data obtained.

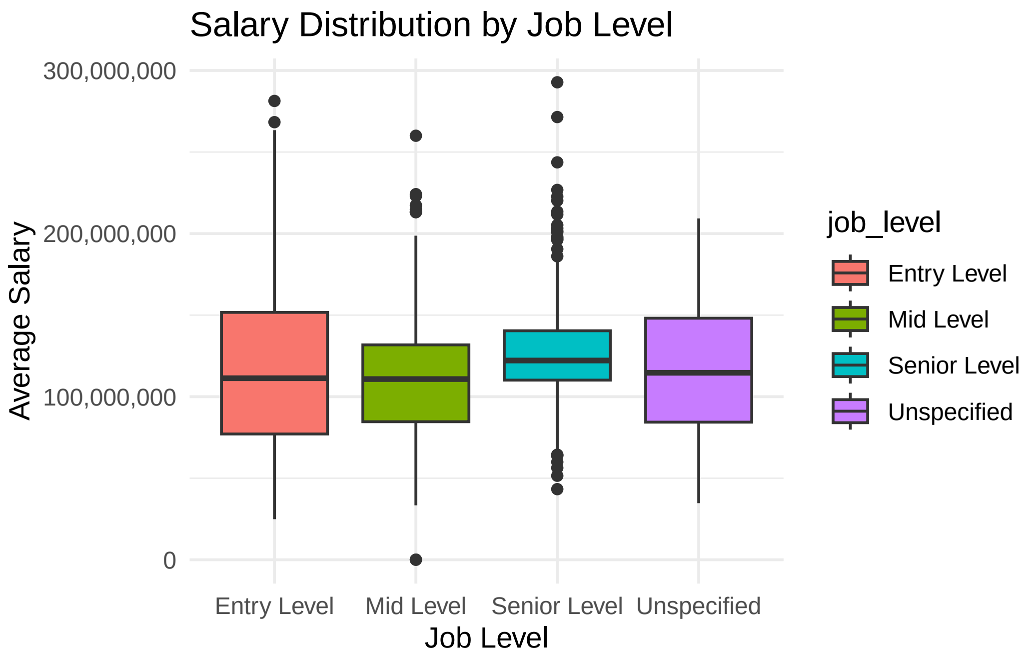
```
# A tibble: 4 x 3
  job_level         n percentage
  <chr>         <int>      <dbl>
1 Entry Level    1132       41.2
2 Mid Level       614       22.4
3 Senior Level    648       23.6
4 Unspecified     353       12.9
```



Great to cumulatively have over 60% of job postings data being entry & mid level due to primary target audience being students.

3.1.2 Salary Analysis

Salary Distribution: Compared salaries across Entry, Mid, and Senior levels to know what current figures are based on retrieved data.




```
# A tibble: 4 x 3
  job_level     avg_min_salary avg_max_salary
  <chr>         <chr>          <chr>
1 Entry Level   114467.46      115118.67
2 Mid Level     109001.69      110235.77
3 Senior Level  125176.81      125689.78
4 Unspecified   117119.06      117811.44
```

3.1.2.1 Text Analysis for Skills

In this section, we perform various text analysis techniques to extract key insights from the job descriptions, including the identification of key skills, skill counts by job level, bigram and trigram analysis.

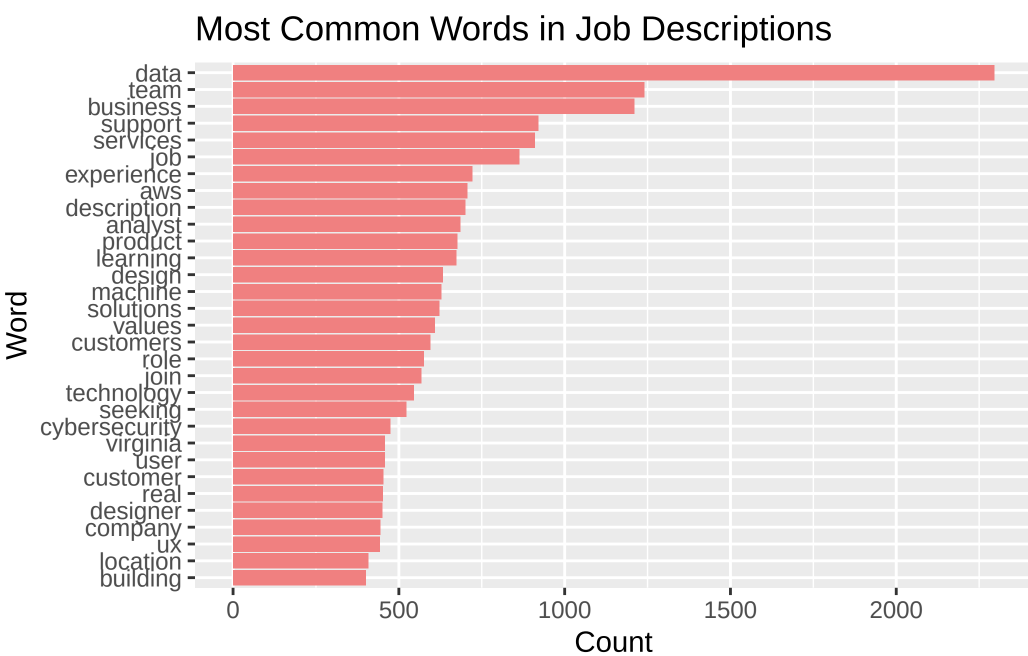
3.1.2.1.1 Process

- Tokenized job descriptions into individual words
- Removed stop words
- Further filtered for technical skills using a predefined and dynamically generated list.
- Performed bigrams and trigram analysis to further identify the top skills and phrases.

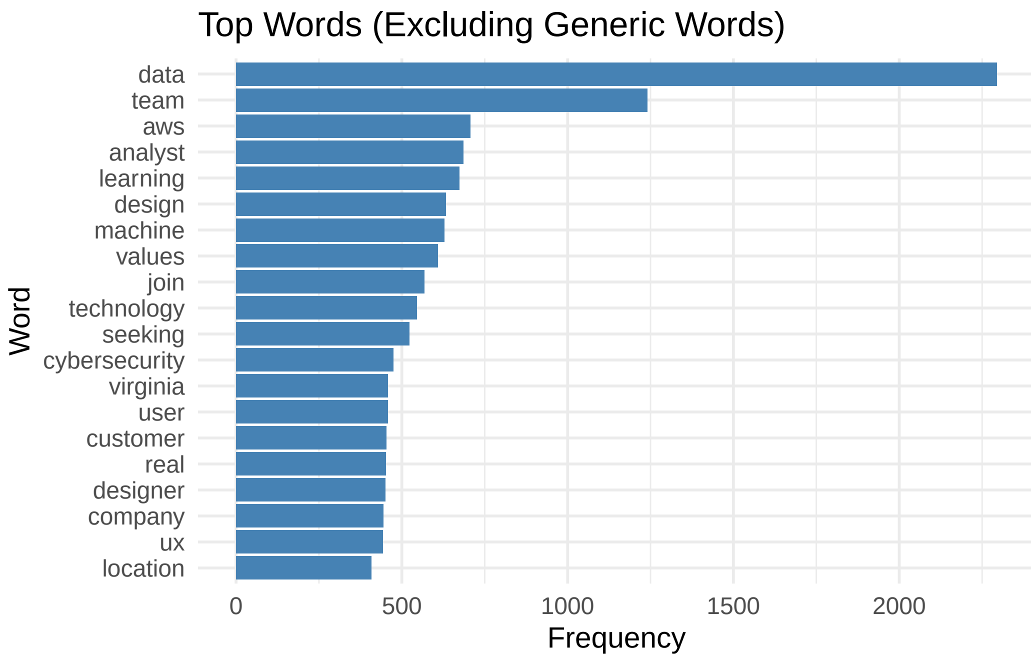
3.1.2.1.2 Text Analysis

```
[1] "Top 10 words"
```

```
# A tibble: 10 x 2
   word            n
   <chr>       <int>
 1 data         2295
 2 team         1240
 3 business     1210
 4 support       920
 5 services      910
 6 job           863
 7 experience    721
 8 aws           706
 9 description   700
10 analyst       685
```

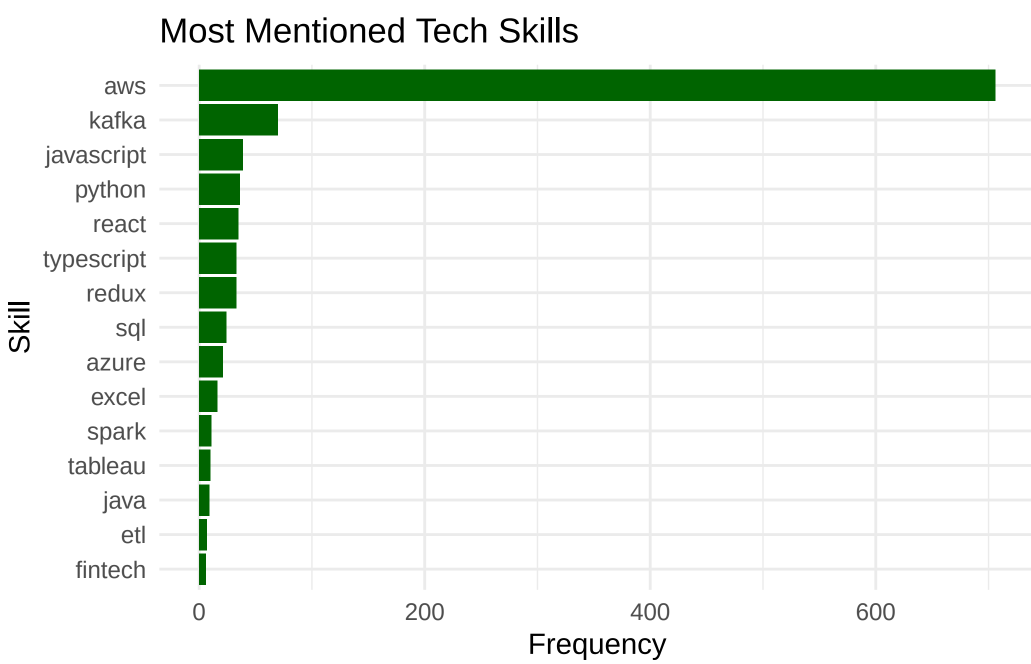


Proceed to further analyse text to sift out generic words like job, description, support, experience, product, position, business, services, customers, role etc.



3.1.2.1.3 Filtering for Technical Skills

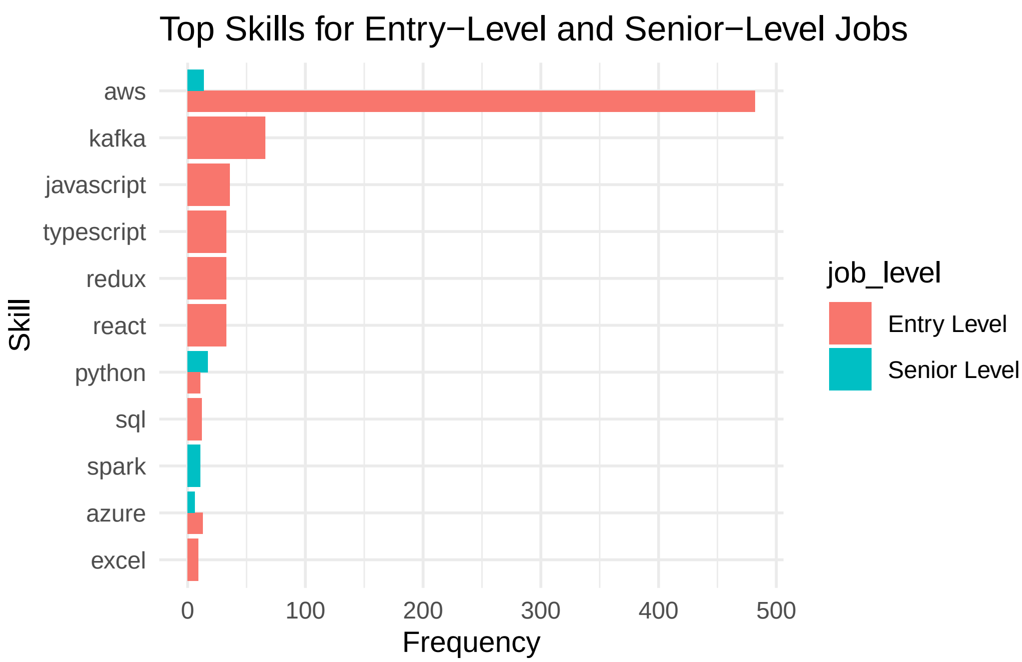
Focused on extracting and counting the frequencies of tech skills in job descriptions by filtering with a dictionary of tech skills.



```
# A tibble: 15 x 2
   word           n
   <chr>      <int>
 1 aws          706
 2 kafka         70
 3 javascript    39
 4 python        36
 5 react         35
 6 redux         33
 7 typescript    33
 8 sql           24
 9 azure         21
10 excel         16
11 spark         11
12 tableau       10
13 java           9
14 etl            7
15 fintech        6
```

3.1.2.1.4 Analyzing Skills by Job Level

Here, we examine the tech skills by different job levels, focusing on entry-level roles.



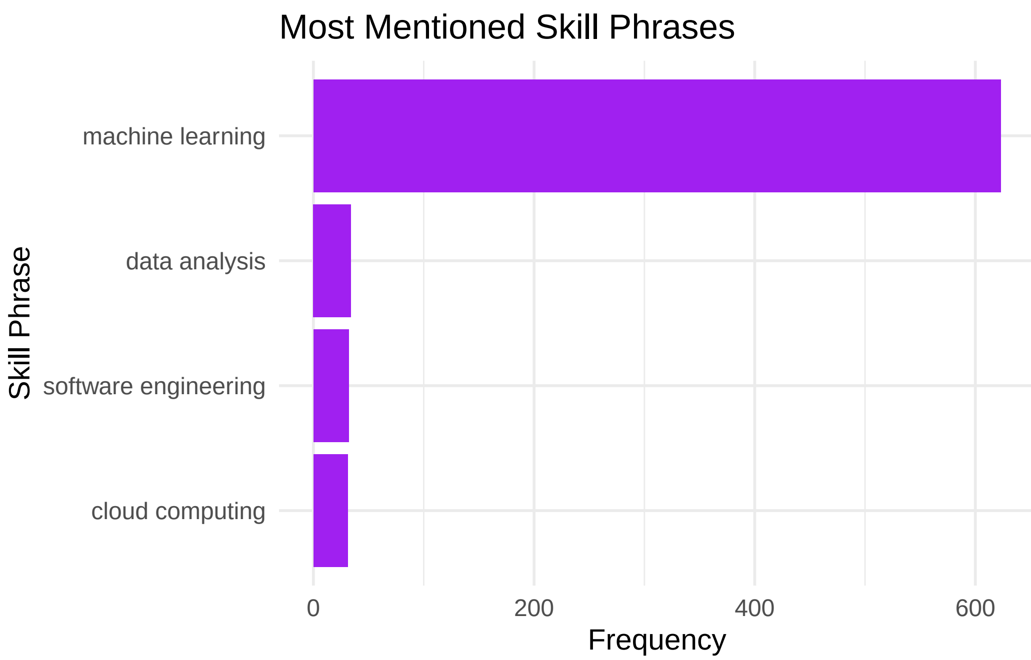
3.1.2.1.5 Bigram and Trigram Analysis (Skill Phrases)

In addition to single words, we will now analyze pairs (bigrams) and triplets (trigrams) of words to capture more phrases like "machine learning" et al that are often used in job descriptions. This provides a richer understanding of the skills demanded in the job market so not to miss relevant skills.

```
# A tibble: 20 x 2
   bigram                n
   <chr>             <int>
 1 we are              810
 2 machine learning    623
 3 our values          587
 4 in the              535
 5 you will            510
 6 do you              508
 7 is a                491
 8 looking for         454
 9 of the              449
10 of our              433
11 will be             401
12 job description     386
13 part of             383
14 and the             378
15 as a                374
16 responsible for     366
17 for a               322
18 united states       319
19 to join             308
20 in a                304
```

```
# A tibble: 20 x 2
   trigram                      n
   <chr>                    <int>
 1 states of america          267
 2 united states of           267
 3 us intelligence community   256
 4 are looking for            249
 5 we are looking             245
 6 we are seeking             233
 7 to join our                232
 8 be part of                 223
 9 part of a                  201
```

```
10 looking for an              194
11 at capital one              183
12 be responsible for          182
13 you will be                 178
14 a fast paced                164
15 looking for a               164
16 amazon web services         163
17 web services aws            163
18 it support specialist       162
19 you'll be part              162
20 in a fast                   159
```




```
# A tibble: 4 x 2
  bigram                   n
  <chr>                <int>
1 machine learning       623
2 data analysis           34
3 software engineering    32
4 cloud computing         31
```

3.1.2.1.6 Outcome

This analysis helped identify which tech skills are most frequently mentioned in job descriptions and how they correlate with different job levels. The use of bigrams and trigrams further helped understand the combination of skill phrases that employers are looking for.

- The top skills extracted include AWS, Kafka, Javascript, and Typescript with they all being apparent for entry level roles.
- Top skill phrases include "machine learning," "data analysis," and "software engineering."
- Going through the trigrams list *"amazon web services"* made the most sense but already accounted for with AWS in earlier process so won't be repeated

These findings were used to conduct the **Google Trends** popularity demand search.

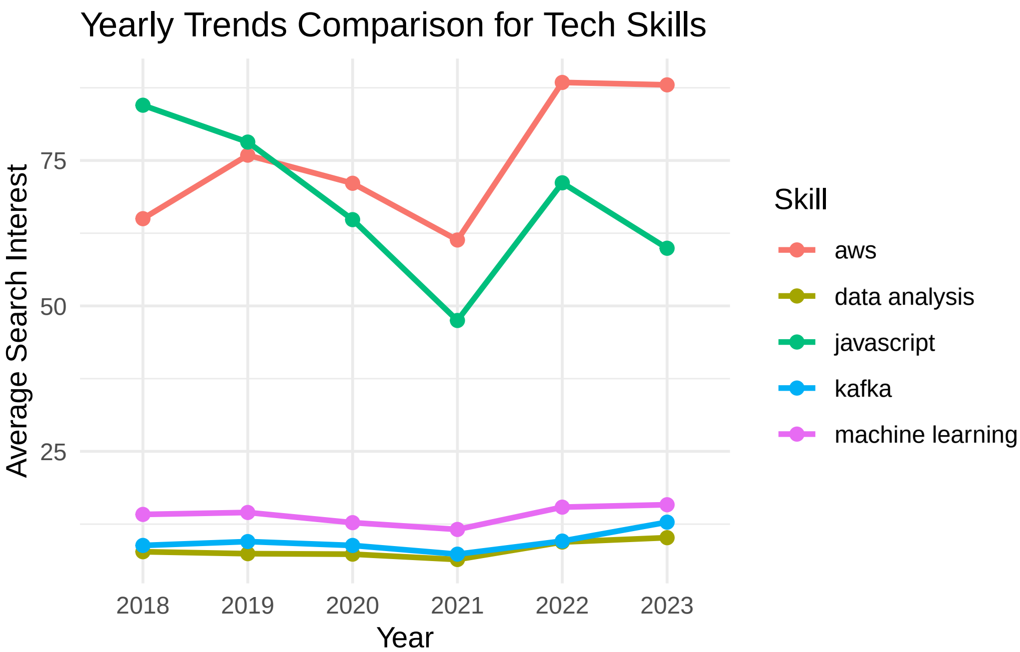
3.1.3 Trend Analysis: Insights from Google Trends

Based on the text analysis, we explored the popularity of the top skills extracted from job descriptions using GoogleTrends query between 2018 - 2023.

3.1.4 Process

- Queried Google Trends using the top 5 skills identified from the text analysis.
- Visualized trends in search interest over time and regional variations.

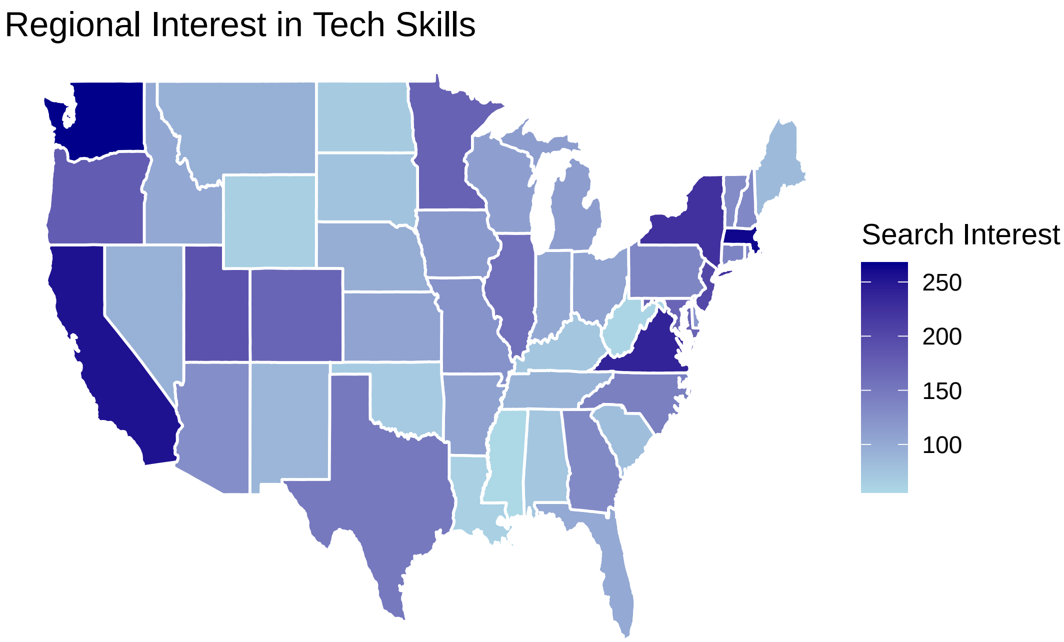
3.1.5 Yearly Trends: Search interest over time.

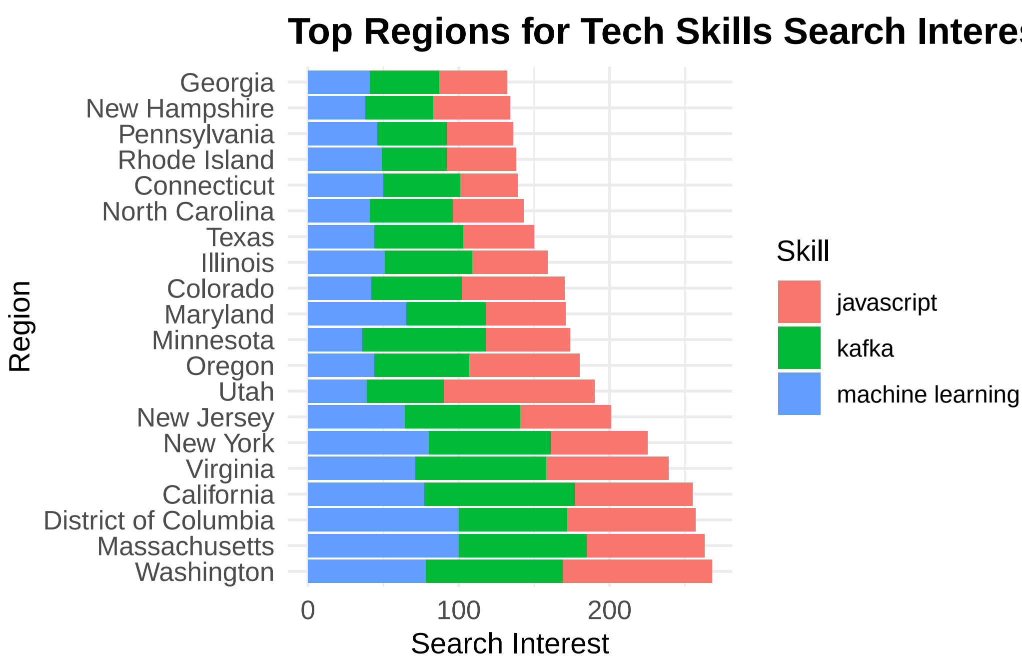


3.1.5.1 Growth Rate between Start of 2018 and End of 2023

```
# A tibble: 5 x 2
  keyword          growth
  <chr>             <dbl>
1 aws                35.4
2 data analysis      31.2
3 javascript        -29.1
4 kafka              45.3
5 machine learning   11.8
```

3.1.5.2 Regional Interest: Geographic distribution of search interest.

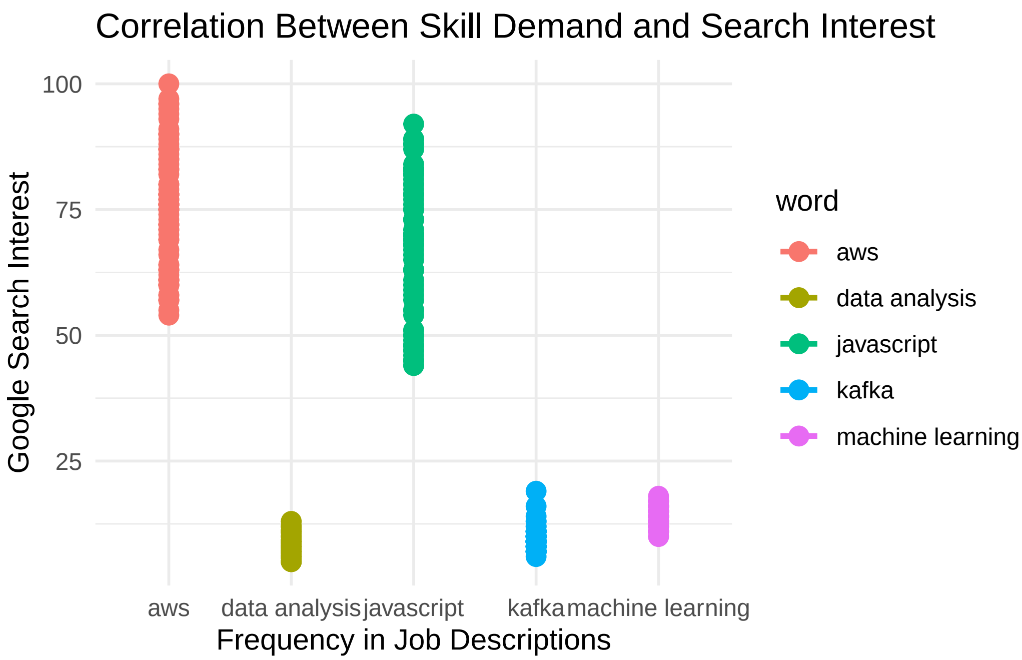




This suggests that students interested in developing competitive tech skills may benefit from focusing their job search and skill development efforts in these regional markets, where the demand appears to be strongest. Conversely, areas with lower relative demand and search interest may indicate opportunities for educational institutions to better align their curriculum and training programs to meet the evolving needs of local tech employers.

3.1.6 Comparison Analysis Between Job Market Demand(Adzuna API) and Search Interest(GTrends)

This analysis provides valuable insights into the alignment between employer demand for certain technical skills and public interest/awareness of those skills.



The correlation analysis revealed disparities between job market demand and public interest in technical skills. This could be as a result of GTrends API containing other search interests that may not necessarily be the job market demand.

Due to varying sample sizes and scope of Adzuna and GTrends data, to ensure fair comparisons across skills, we normalized the frequency and search hits using a 0-100 scale and computed z-scores. This approach preserves relative differences while providing a standardized metric for analysis. Normalized metrics allow us to classify skills into categories such as 'High Demand' or 'Moderate Interest,' providing a nuanced understanding of the interplay between market needs and public awareness.

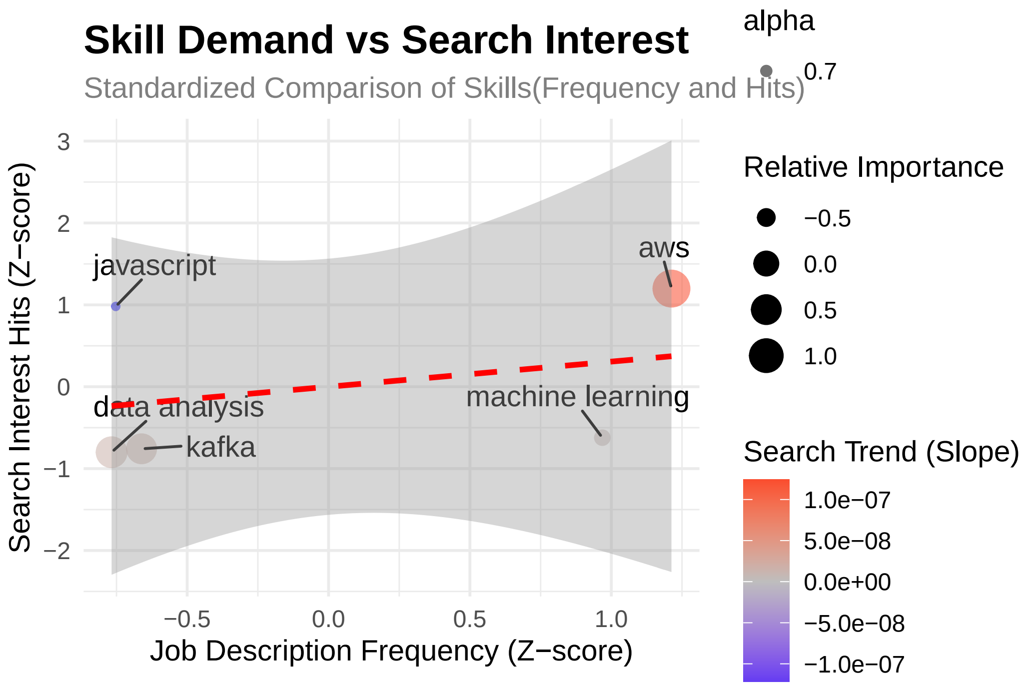


Table 3: Skill Comparative Analysis: Demand and Interest Insights

Skill	Job_Description_Freq	Avg_Search_Hits	Rel_Importance	Demand_Level	Search_Level
aws	706	74.958333	1.4535033	High Demand	High Interest
kafka	70	9.486111	0.5011235	Low Demand	Low Interest
javascript	39	67.680556	-0.7382529	Low Demand	Moderate Interes
machine learning	623	14.041667	-0.6020200	Moderate Demand	Low Interest
data analysis	34	8.083333	0.6136791	Low Demand	Low Interest

```
[1] "Correlation between Job Description Frequency and Search Hits: 0.307"
```

This scatter plot compares the standardized job description frequency and search interest for top skills. Each point represents a skill, with its position reflecting demand (x-axis) and interest (y-axis). Point size indicates relative importance, while color shows the search trend (red for increasing, blue for decreasing). Skills in the top-right quadrant, such as AWS, exhibit high demand and high interest, whereas Machine Learning highlights a disconnect with high demand but moderate search interest. JavaScript's decline in search interest is noteworthy despite its utility.

3.1.6.0.1 Outcome

The correlation coefficient between job description frequency and search hits suggests a moderate positive relationship between the two metrics, indicating that skills with higher demand tend to also have higher search interest.

From the table:

- "aws" emerges as a dominant skill, with high demand, high search interest, and significant growth (32.8%) over the past five years. This indicates it is a highly relevant and sought-after skill. Industries such as **finance**, **e-commerce**, and **tech startups** increasingly rely on AWS for their cloud infrastructure.
- "kafka" and "data analysis" have relatively lower demand and interest, but are seeing positive growth trends, suggesting they may be emerging as areas of increasing importance.
- "javascript" is an interesting case, with moderate search interest but declining growth, indicating decreasing relevance as compared to other skills. This is likely due to be the rise of newer frameworks like React and Vue.js, which build upon JavaScript but provide more specialized functionality for web development.
- "machine learning" has high demand but moderate search interest and growth, hinting at a potential disconnect between job market needs and public/professional interest.

4 CONCLUSION

4.1 Findings

1. *Tech Skills in Demand*: AWS, data analysis, javascript, kafka and machine learning dominate based on job descriptions.
2. *Trends Over Time*: Kakfka and AWS show significant growth in search interest according to GTrends API.
3. *Regional Variations*: GTrends API shows that certain tech hubs, such as the District of Columbia, California, Massachusetts, New York, Virginia and Washington, exhibit higher levels of search interest for skills.
4. *Salary Insights*: Senior roles command significantly higher salaries, with noticeable gaps across job levels.

The findings help inform educational program development and career guidance decisions for primarily students to ensure alignment between market demands and skill development efforts.

- *Students aiming for high-demand roles like AWS Engineer should focus on certifications, as the job postings indicate strong demand despite lower search interest. Relevant certifications, such as AWS Certified Solutions Architect and AWS Certified Developer, are highly valued by employers, further driving demand for professionals with AWS expertise.*
- *Skills like "machine learning" and "data analysis" show strong search interest, suggesting areas for professional development and visibility*
- *Also, in understanding these geographic variations in tech skill popularity, students can make more informed decisions about where to target their career aspirations, while educators can identify regional gaps and tailor their program offerings accordingly.*

4.2 Limitations

While the findings from this analysis provide valuable insights, certain limitations should be noted:

- **Sample Scope**: The job description data is based on a sample of job postings retrieved from the Adzuna API. The dataset is not fully representative of the entire job market and may be biased towards specific sectors, regions, or job platforms. Additionally, the Adzuna API limits the number of hits per day (250), which may restrict the completeness of the data. Some skills may be underrepresented due to the limited query terms, and the dataset may not cover all industries or regions.
- **Skill Extraction**: The identification of skills in job postings relied on pre-defined keywords and phrases. While this approach is useful, it may overlook emerging or less commonly mentioned skills that do not fit within the chosen keyword set. Additionally, the keyword-based text analysis may miss nuances in how skills are described across different job postings.
- **Google Trends Data**: The Google Trends data reflects public search behavior, but it may not directly correlate with industry demand or skill acquisition. Search volume for keywords can be influenced by various factors unrelated to actual hiring needs. For instance, individuals may search for skills due to curiosity or academic interest without necessarily pursuing a career in that field. Furthermore, people may search for tools or technologies not directly related to job openings, which can skew the data.
- **Normalization Assumptions:** The normalization process used to compare different skills assumes a linear relationship between job description frequency and search interest. This simplification may not fully capture the complexity of how these two metrics interact in the real world. Z-scores standardize but may obscure small variations in data.
- **Data Availability**: Certain sites like LinkedIn, Glassdoor, and Google Jobs block scraping, limiting the potential for gathering data from these comprehensive job boards.

4.3 Future Work

To address these limitations, and improve the comprehensiveness of future analyses future research could:

1. Expand the list of roles and skills in API queries.
2. Expand the scope of job posting data by integrating multiple APIs or datasets from different job boards, ensuring a more comprehensive view of the job market trends and skill demands.
3. Implement more sophisticated natural language processing (NLP) techniques, such as topic modeling or deeper clustering, could be used to capture a broader and more diverse range of skills mentioned in job descriptions. This would improve the richness of skill extraction.
4. Future research could extend this analysis to a global scale, comparing skill demand in different countries or regions. While this study focuses on the U.S. market, other countries (e.g., India, China, or European markets) may have varying skill demands, providing a broader understanding of global job market trends