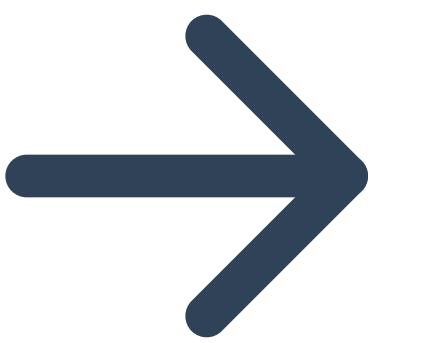


# Map Your Journey

A Simple Guide for International Software Developers on  
Job Opportunities and Living in the U.S



# About

## Business Objective

To analyze the distribution of software engineering job availability, visa sponsorship rates, competitive salaries, and livability factors across different U.S. states. This study aims to provide actionable insights that will assist international students in making strategic decisions regarding their job search and relocation.

We believe that these insights will help both employers to attract top international talent and international students to strategically navigate the U.S. job market.

## Tools Used



## Dataset



### US Software Engineer Jobs

Kaggle | Collected on 22/Jul/2021



### Best States Data

The Motley Fool surveyed | Jan 22, 2024

# Business Questions

## Availability and Salary

How are the software development jobs distributed across states? How about the average annual salary?

## Visa Sponsorship

What is the distribution of visa sponsorship for software engineering roles across different states, and how does it relate to salary?

## Best States for International Software Developers

Which states provide the best combination of sponsored job availability, competitive salaries, and livability for international software developers?

# Prep for Analysis

## Data Profiling

Analyze datasets to identify patterns and inconsistencies.

## ETL Process

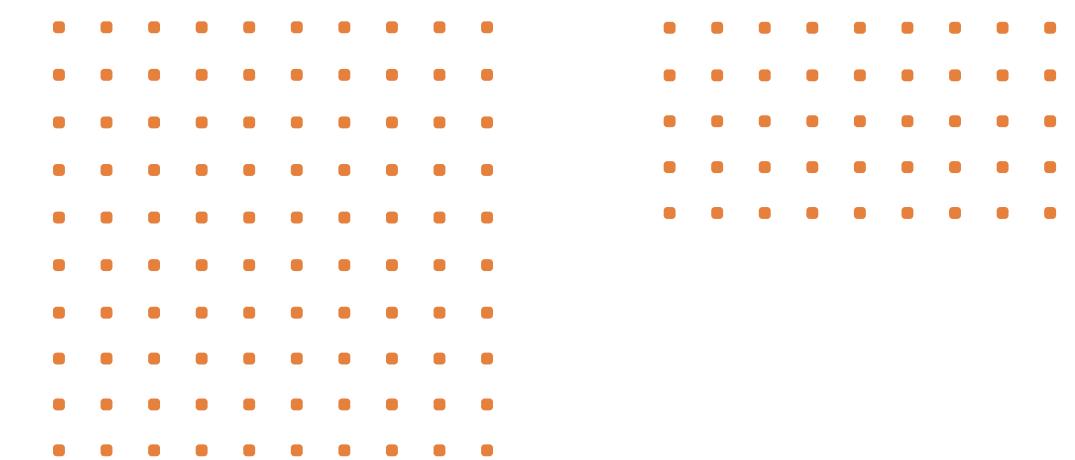
Remove unnecessary columns and standardize salary data across different payment types

## Data Integration

Integrated multiple datasets using "State" as the joining key

## PowerBI Analysis

Create visualizations including heat maps, donut charts, and bar charts



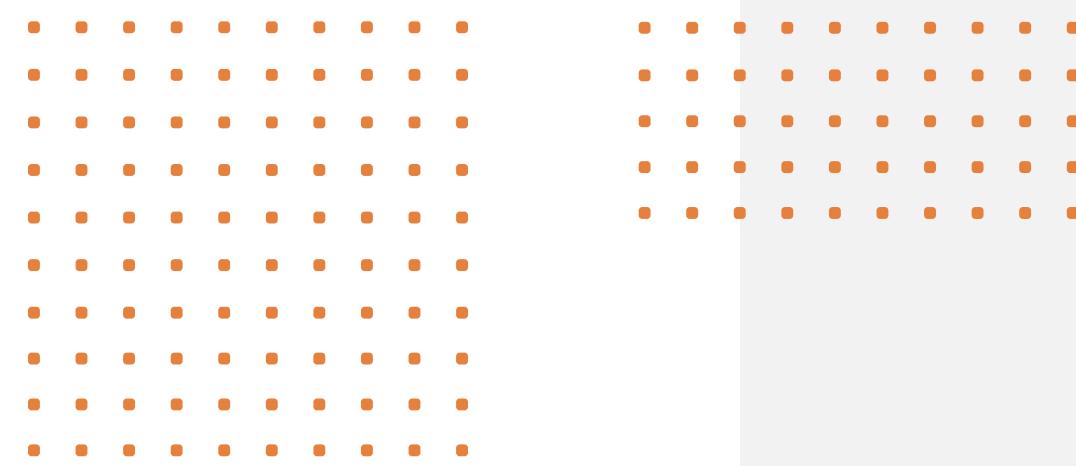
# Data Profiling

## Salary Data

1. Multiple salary formats discovered (annual, hourly, weekly, daily)
2. Salary ranges using hyphenation (e.g., "45000-55000")
3. Some records showing median salaries around 12k
4. Missing salary values

## Location Data

1. Special case with Washington DC data (identified as requiring exclusion)
2. State-level data consistency issues (State is used as a joining key)
3. Location data potentially containing mixed formats of state and city information



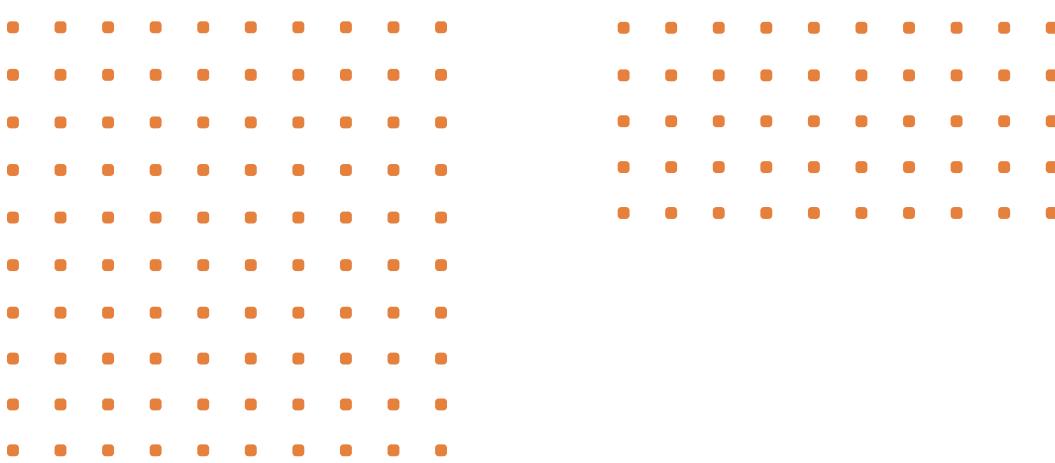
# Data Profiling

## Job Posting

1. Over 1400 records with empty hiring needs
2. Unnecessary or duplicate columns (ie. review\_count, relative\_time, etc.)
3. Multiple types of employment (full-time, part-time, contract, etc.)

## Data Quality

1. Null values across various columns
2. Duplicate columns requiring removal
3. Inconsistent data formats requiring standardization
4. Empty fields in certain categories



# ETL Process

## Extract

1. Removed 20+ unnecessary/duplicate columns including:
  - Review count, relative time, hiring details
  - Job-specific links and IDs
  - Location extras and company details
2. Renamed columns for better clarity
3. Validated and corrected data types

## Transform

1. Standardized salary data across multiple formats:
  - Converted hourly rates ( $\text{hourly} \times 40 \times 52$ )
  - Converted weekly pay ( $\text{weekly} \times 52$ )
  - Converted monthly pay ( $\text{monthly} \times 12$ )
  - Calculated mean values for salary ranges
2. Handled missing values while preserving data integrity
3. Cleaned and standardized text data
4. Split salary ranges into separate columns

## Load

1. Created calculated columns
2. Grouped related data
3. Prepared data structure for visualization
4. Integrated final dataset for PowerBI analysis



# Data Integration

## Integration

1. Used "State" as primary joining key between datasets
2. Excluded Washington DC due to missing retirement statistics
3. Allowed null values to preserve important records

## Data Cleaning

1. Standardized state names across datasets
2. Created consistent state-level aggregations

# Challenges



## Salary Data Standardization

Multiple salary formats and salary ranges requires standardization to ensure consistent comparison

## Missing Values Handling

Allow null values to preserve other information in entries, and replaced null values with mean values

## Data Integration Issues

Ensure consistent state naming and address discrepancies between datasets.

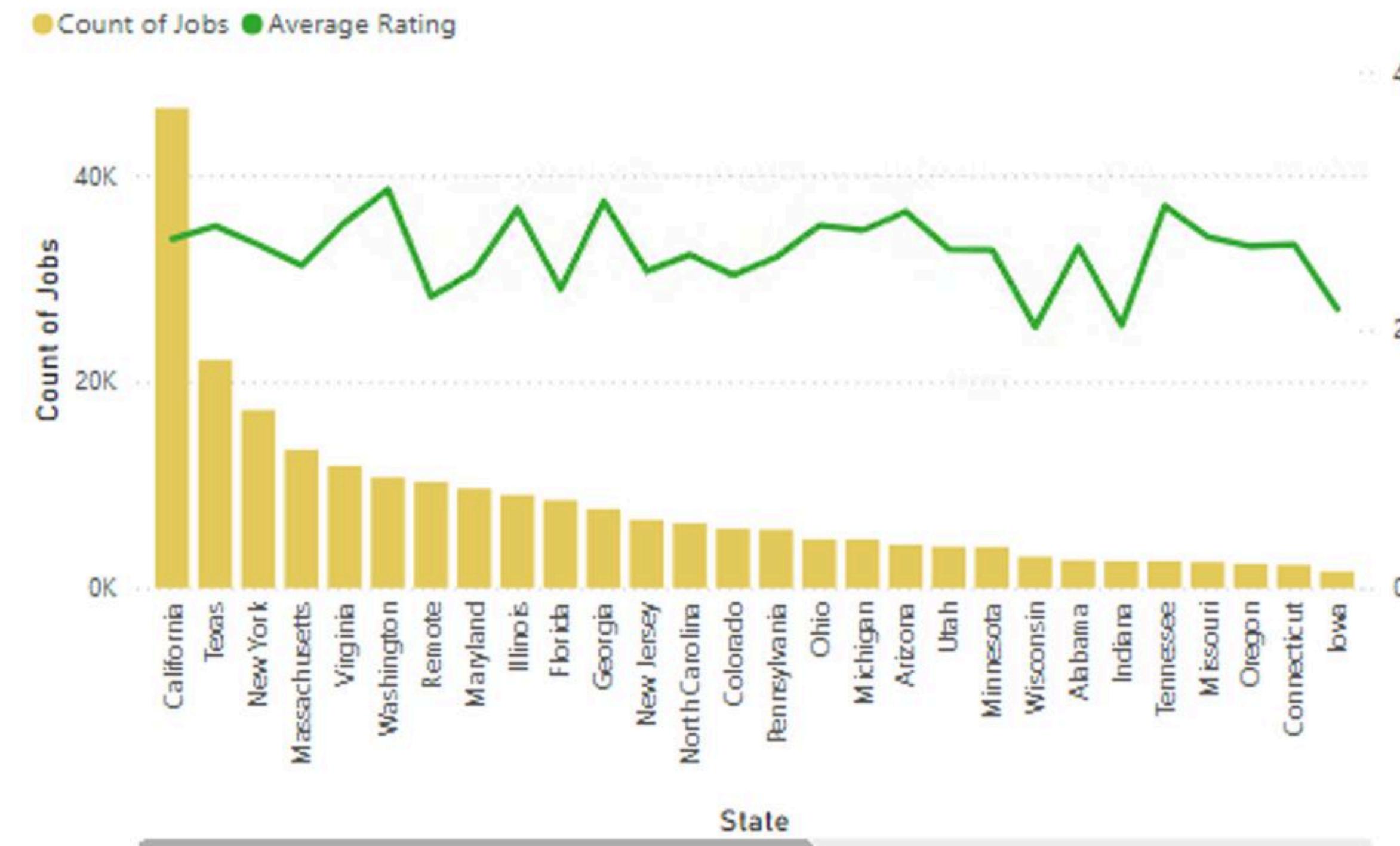
## Data Type Inconsistency

Multiple columns contained mixed data types, requiring careful parsing and standardization

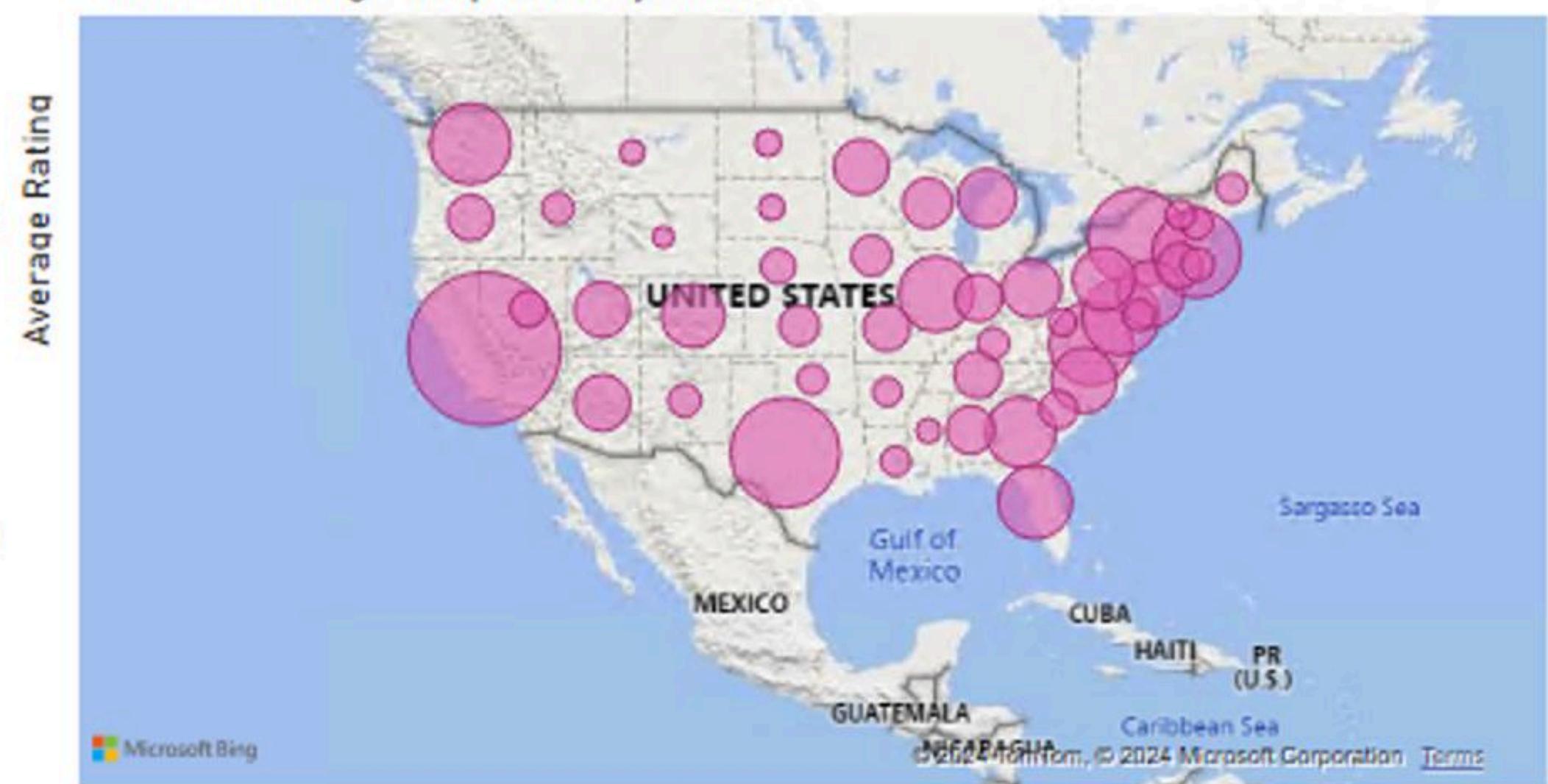
# PowerBI Analysis

## - Availability

Job Count and Rating by State



Count of Hiring Companies by State



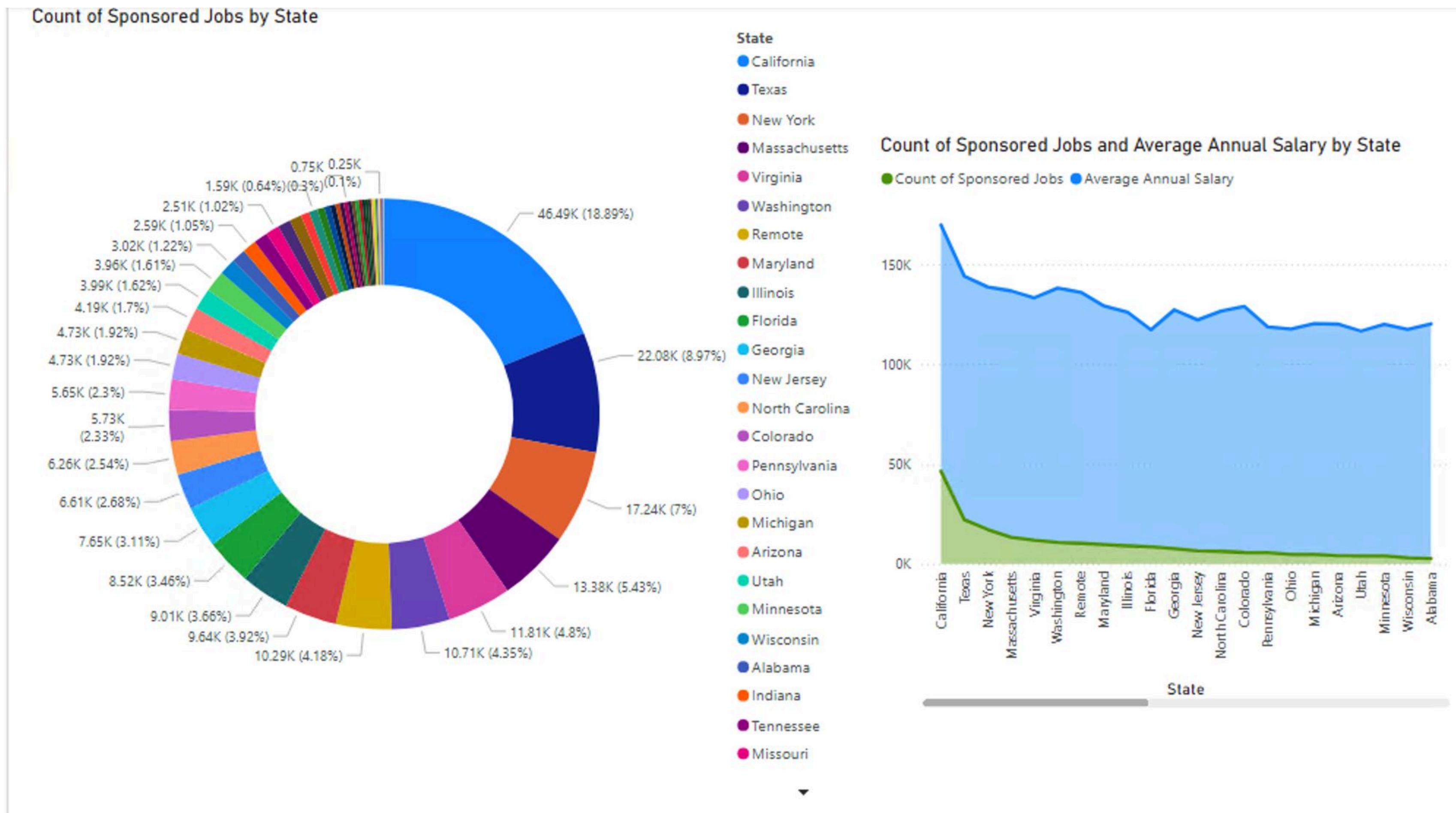
# PowerBI Analysis

## -Salary



# PowerBI Analysis

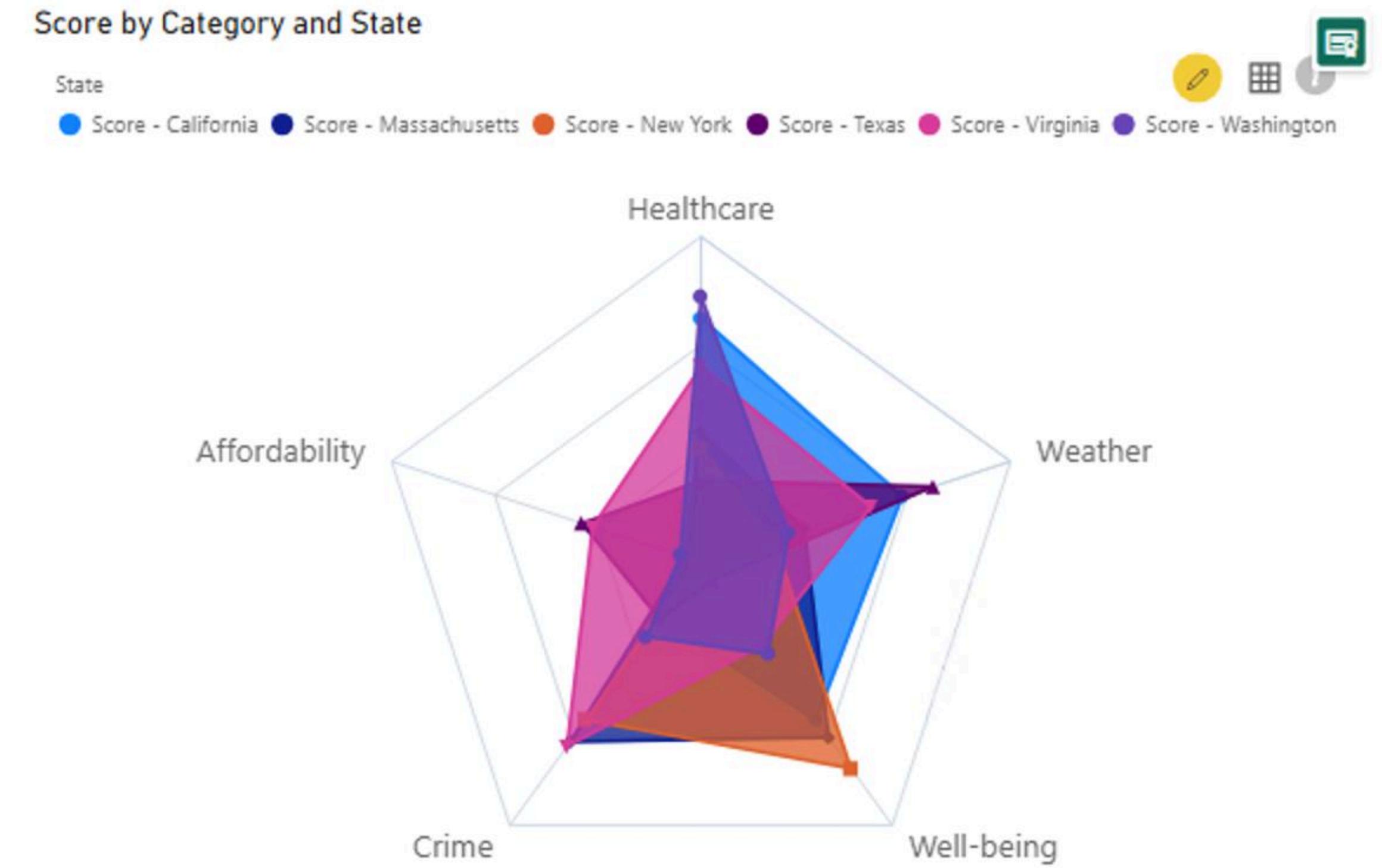
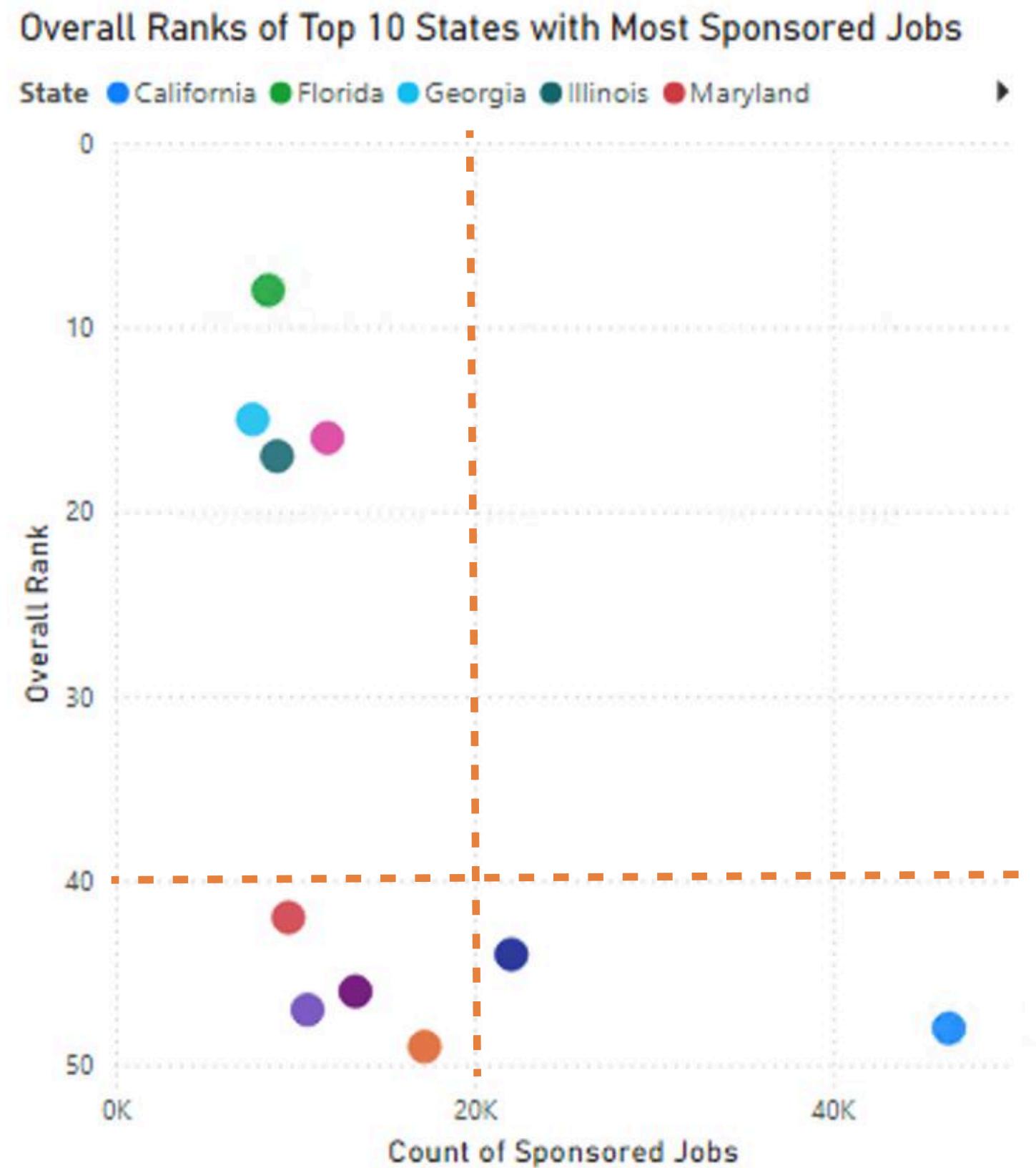
-Sponsorship



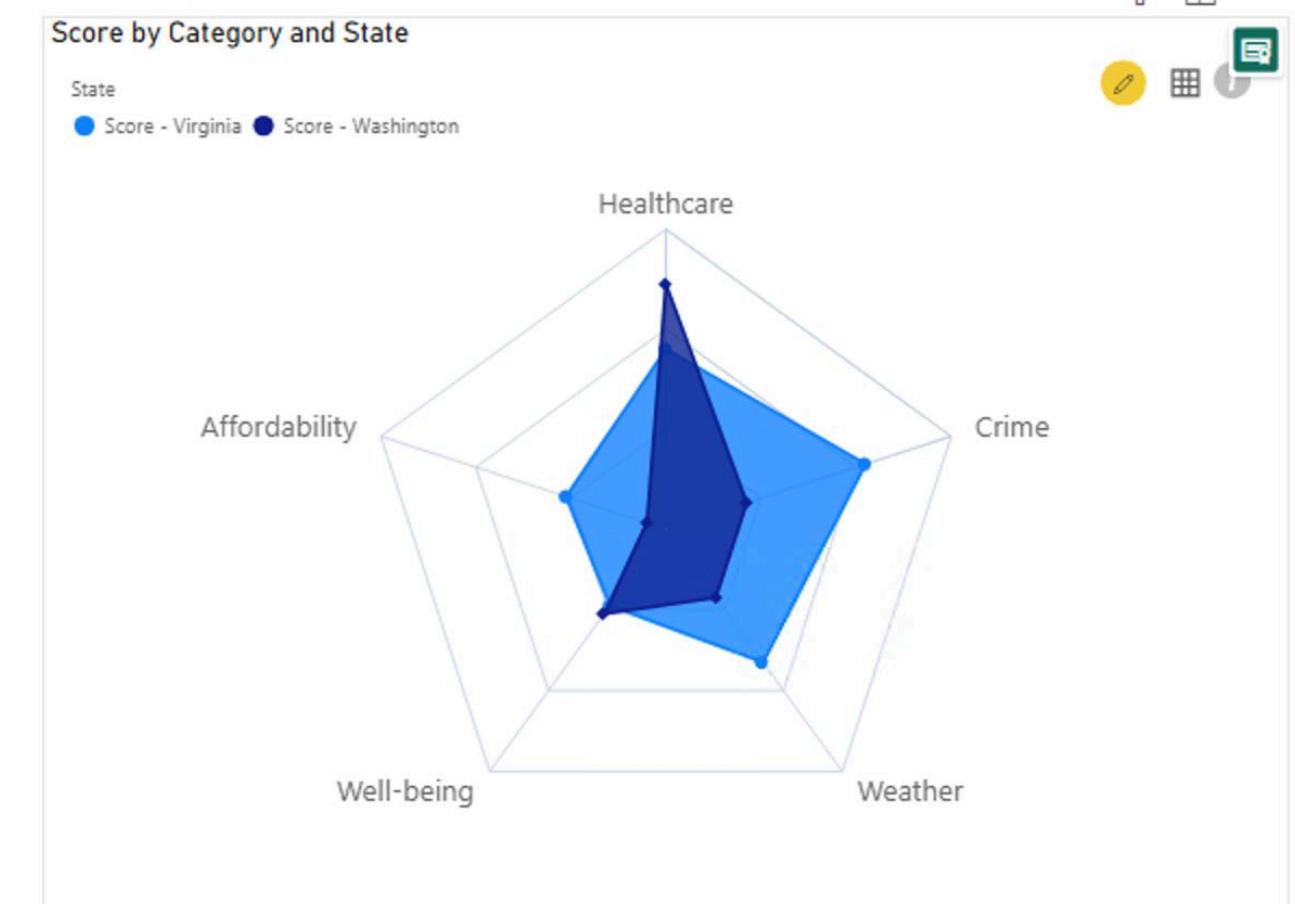
# PowerBI Analysis



## -Top states

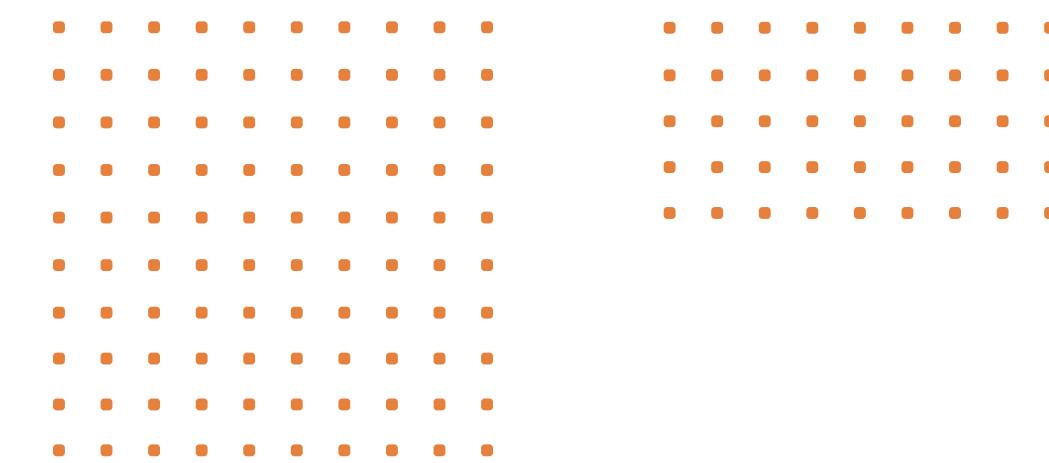


# Top 6 States





# Future Scope



01

Predictive analysis – Azure ML

02

Additional data integration

# Thank you!

