

CASE STUDY

Analyzing Job Market Data in Power-BI



INTRODUCTION

In this Power BI case study, we'll explore a real-world job posting dataset to uncover insights for a fictional recruitment company called DataSearch. We'll use Power Query to investigate and clean the data to find out what skills are most in-demand for data scientists, data analysts, and data engineers. We'll then use DAX to build insightful visualizations of our findings. Finally, We'll bring it all together using everything Power BI has to offer to create a business dashboard so that we can answer questions for the Data Search team.

PROJECT OBJECTIVES

Investigate market trends for data science roles

The context An employee recruiting firm Data Search,needs to uncover job trends

The task Discover trends in jobs and skills within the data science industry

The dataset Factitious collection of key details from job postings. The Dataset contains data about 25,114 job postings and 17 columns information about each job posting (Title, Date, Industry, Skills, ...)

Data analysis Pipeline:

Checking Data

Exploring Data

Analyse and Visualise Data

Dashboarding

Communicating Insights

PROCEDURE

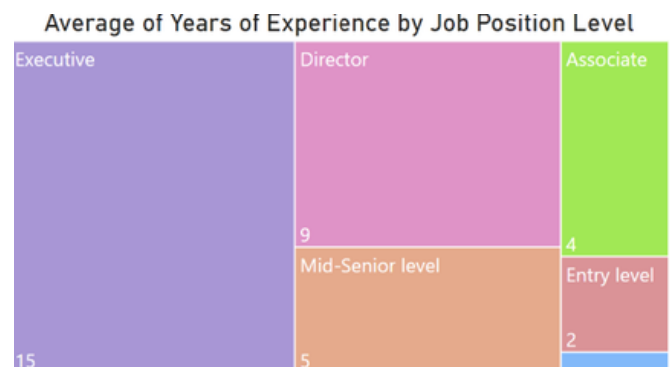
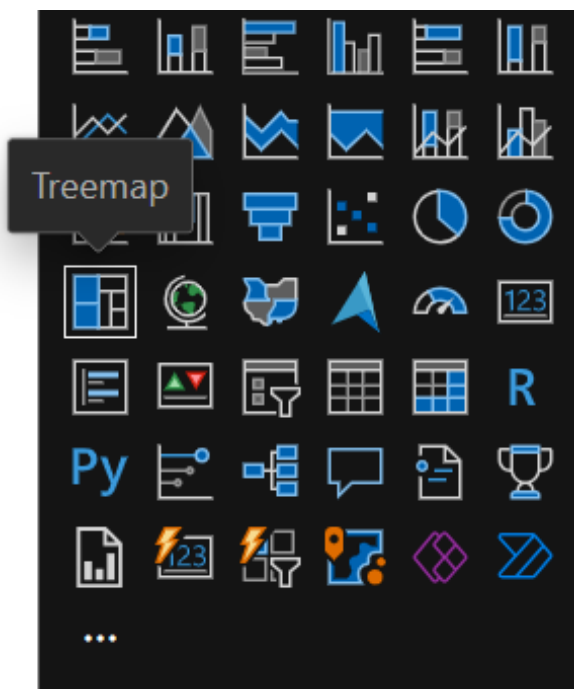
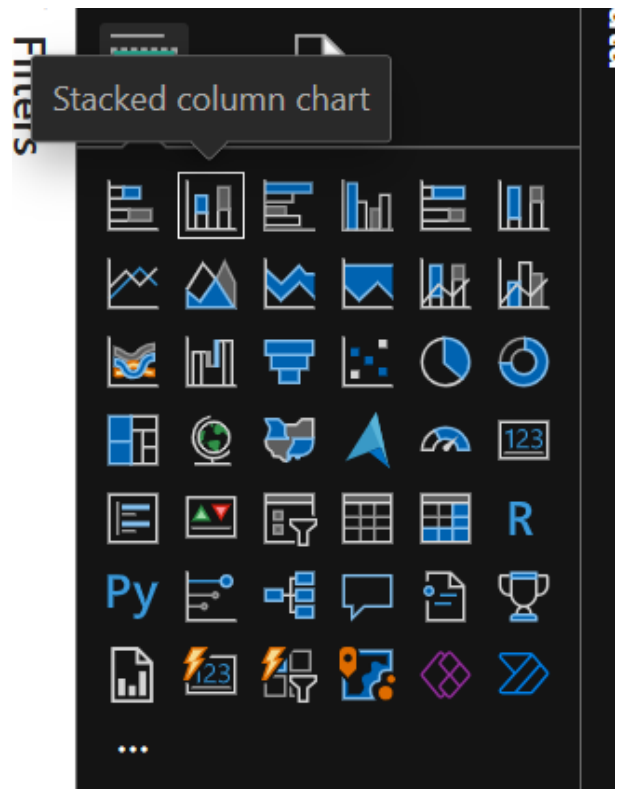
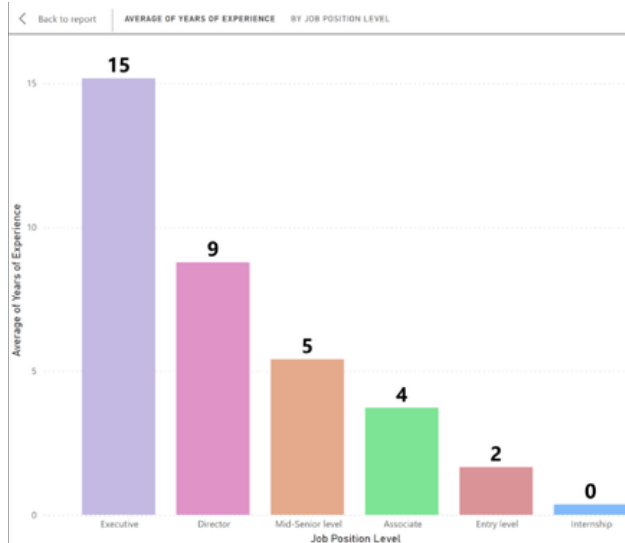
Step 1: Exploratory Analysis

We'll start this Power BI case study by importing and cleaning the job posting dataset using Power Query. Once we've done this, it's time to do some initial data exploration.

First, I imported and loaded the dataset into Power BI. Then, using Power Query tool, 93% of values for Max and Min pay column were null values. This indicates that more than 90% of companies do not mention Maximum and Minimum of salary in their job postings. After than by sorting the Number of applicant column in a descending order, I understood that the most popular industries were Marketing and Advertisement, Computer Software, and Broadcast Media, respectively.

More Experienced, Higher Level

Next, I want to obtain an idea of the relation between the Years of Experience and the Job Position level (e.g. Entry Level). To do so, I visualised the average Years of Experience and column and the Job Position Level column using a Treemap and a Bar chart.



It is observed that the average experience years required for a Executive level job is around 15 years and you can land an Internship with less than a year experience. At 15.17 years, Executive had the highest Average of Years of Experience and was 4,047.13% higher than Internship, which had the lowest Average of Years of Experience at 0.37 year. Moreover, across all 6 Job Position Levels, Average of Years of Experience ranged from 0.37 to 15.17 years.

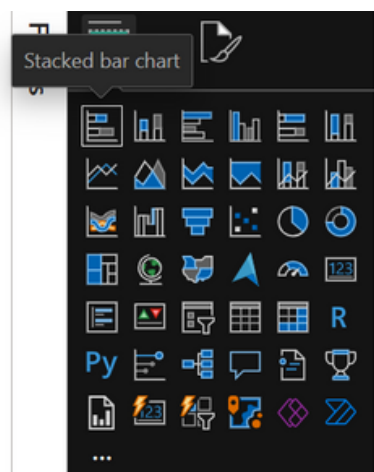
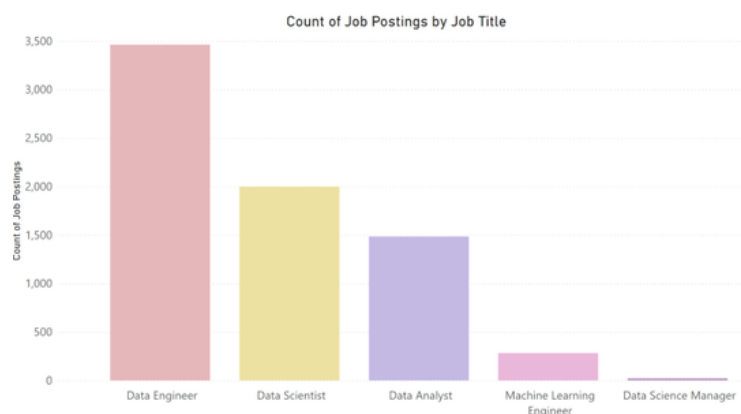
Job Posting Trend Analysis

Next, we are interested in observing the trend of job postings in time. So we visualise the Count of Job Posting IDs over Job Posting Date as time axis.

It is evident that the first quarter of 2020 the number of job positions has been the lowest due to the rise of Covid-19. Between January 2017 and December 2021, Mid-Senior level had the largest increase in Count of Job Postings (174.48%) while Internship had the largest decrease (33.33%). The most recent Count of Job Postings anomaly was in September 2021, when Director had a low of 1. Count of Job Postings for Internship started trending up on February 2021, rising by 300.00% (3) in 4 months.

Filtering for Data Related Job Titles

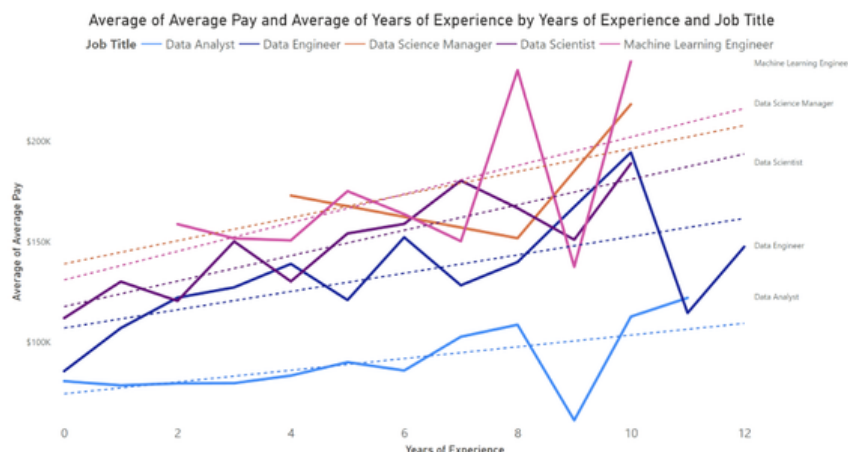
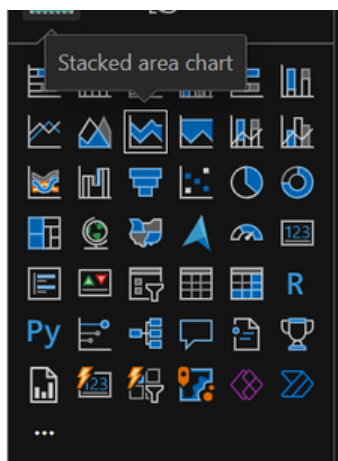
Now it's time to dig deeper into the dataset to gain info about data related jobs: Data Scientist, Data Analyst, Data Engineer, Machine Learning Engineer, and Data Science Manager. We start by analysing the number of job postings for each job title.



At 3,462, Data Engineer had the highest Count of Job Postings and was 14,952.17% higher than Data Science Manager, which had the lowest Count of Job Postings at 23. Data Engineer accounted for 47.74% of Count of Job Postings. Across all 5 Job Titles, Count of Job Postings ranged from 23 to 3,462.

Effect of Years Experience on Salary

In the next step of our EDA, we are interested in identifying influential factors on job market trend. One of the integral parts of a job is the compensation. Here, we analyse the effect of years of experience on each job's offered salary.



It is obvious that the more the years of experience, the higher salary. Also, we observe that the Data Analysts earn the least compared to other data-related jobs. A bit disappointing! However, we must keep in mind that more than 93% of values for Max and Min pay columns (which we used to calculate the Average pay for each job title) were null values, so we cannot draw any major conclusions.

Step 2: Analyzing Market Trends

In our second analysis step, we continue to apply our Power BI skills to extract data insights with tools like DAX and begin to build unique data visualizations.

Key Insights we have Identified So Far:

Job postings are gradually increasing in number.

Data Engineers, Data Scientists, and Analysts are in demand!

Tech industries need data science roles the most.

Salaries are trending upward as expected.

In our second step, we are going to analyse our data to provide insights on the following:

Correlation between skills and job title

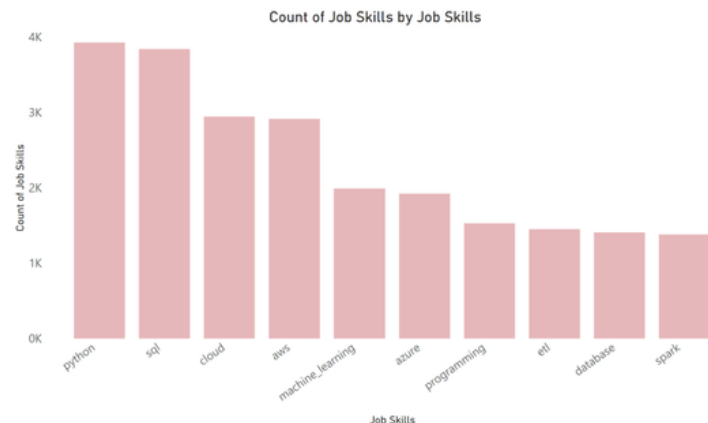
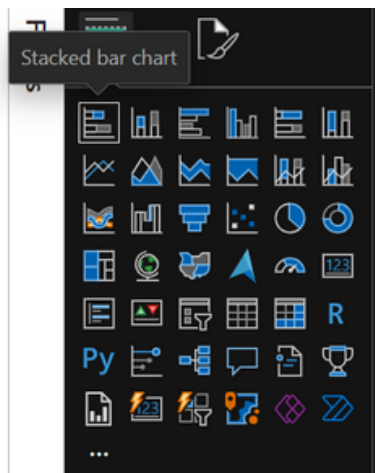
Top industries and companies

Potential recommendations

Skill analysis

Using only the two columns Job Posting ID and Job Skills, we create a second Table called Job Skills. In the new table we remove the extra characters in the column Job Skills such as brackets [], quotes ", and extra spaces. I have also used the Trim and Clean features to remove the extra spaces in each cell. Then, we split this column into rows to separate the list of skills for each job on a separate row.

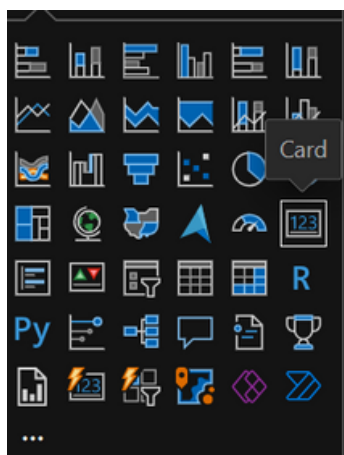
Next, I visualised the count of Job Skills to gain an initial idea of the most required skills.



We observe that the count of Job Skills was highest for python at 3,931, followed by sql and cloud. Python accounted for 7.44% of Count of Job Skills. Across all 165 Job Skills, Count of Job Skills ranged from 1 to 3,931.

Skill Likelihood in Job Posting

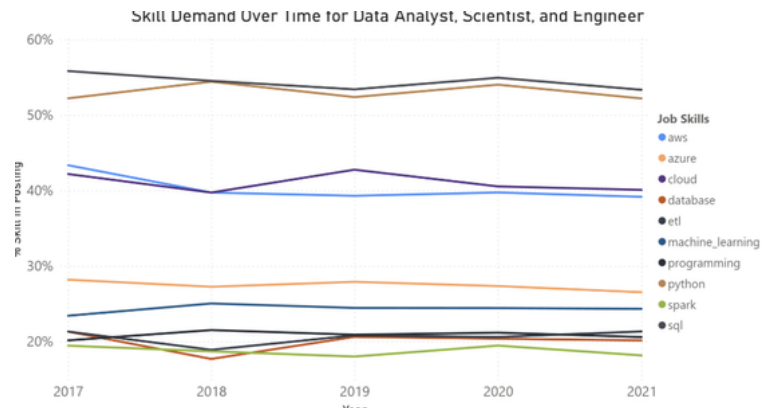
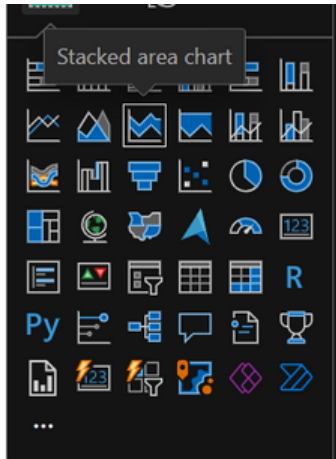
Here, I used DAX to identify the probability of a certain skill occurrence in a job posting. To do so, we use the basic probability principles. The probability of a certain skill existing in a job ad, can be calculated at the count of all the job postings that contain this skill over the total number of all job postings. Therefore, I created three measures called Skill Count, Posting Count, and % of Skill in Posting that calculated the required chance. Then I created a Matrix to visualise the result. I also added a slicer, so that it would be easier to filter for a specific skill.



Job Title	Data Analyst			Data Engineer			Data Science Manager		
	Skills Count	Posting Count	% Skill in Posting	Skills Count	Posting Count	% Skill in Posting	Skills Count	Posting Count	% Skill in Posting
chainer	1485			3462					
dart	1485			3462					
delphi	1485			3462					
jse/jee	1485			3462					
lua	1485			3462					
nuix	1485			3462					
objective_c	1485			3462					
pascal	1485			3462					
shogun	1485			3462					
vb.net	1485			3462					
agile	1	1485	0.07%	1	3462	0.03%			
agile/scrum	1	1485	0.07%	1	3462	0.03%			
ai	1	1485	0.07%	1	3462	0.03%	1		
ai/ml	1	1485	0.07%	1	3462	0.03%	1		
airflow	1	1485	0.07%	1	3462	0.03%			
ajax	1	1485	0.07%		3462				
alteryx	1	1485	0.07%	1	3462	0.03%			
anarchy	1	1485	0.07%	1	3462	0.03%			
Total	146	1485	9.83%	153	3462	4.42%	42		

Trends in skills over time

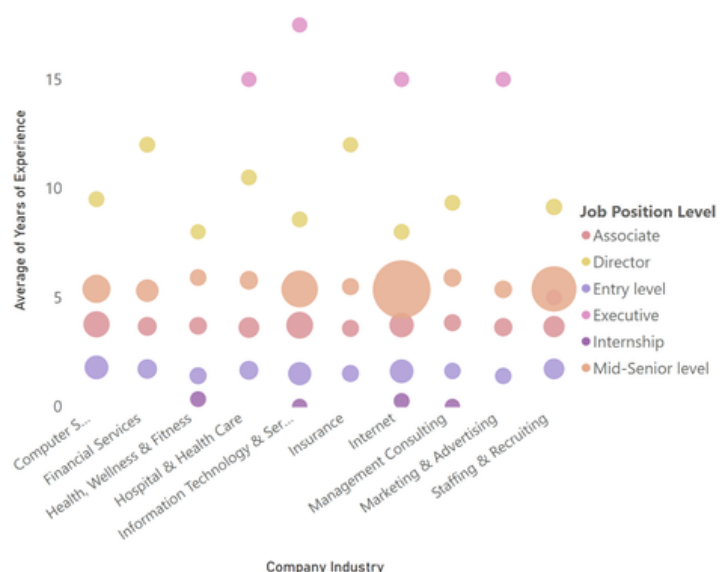
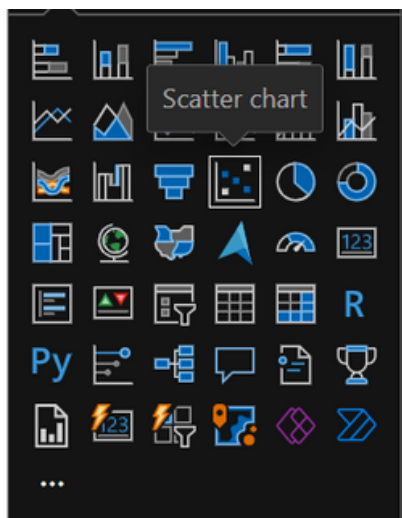
Next, let's analyse the trend of skills in job postings over time. This can give us a good idea of what skills are more in demand in certain time periods. We want a time-series graph, so we use a Line Chart to visualise the % Skill in Postin over time by using Posting Date and use Job Titles from our Job Skills table as a legend. The result is the following graph:



We can observe that although the metrics have fluctuations on a short-term basis, the top 10 skills maintain at consisted deman over the past few years. Moreover, across Job Skills, spark had the most interesting recent trend and started trending down on 2017, falling by 6.61% (1.29%) in 4 years.

Deep dive into key job descriptions

Next, we are interested in exploring the companies and industries that are looking for data-related roles. To do this, I visualised the Average years of experience and posting count by job position level and company industry in a scatter plot, along with a table containing posting count, job title and company name. This way, I can find out which industry and company is mostly looking for which job position level and with how many years of experience.



RESULT AND OBSERVATION

It is evident that there is a large demand for Mid-senior scientists with an average of 5.37 years of experience in the internet industry.

Posting Count	Job Title	Company Name
994	Data Engineer	Toptal
673	Data Scientist	Toptal
183	Data Engineer	Jobot
120	Data Engineer	Recruiting from Scratch
97	Data Scientist	Jobot
89	Data Engineer	Dice
50	Data Analyst	Perficient
43	Data Analyst	Kraken Digital Asset Exchange
43	Data Engineer	Blackbaud
40	Data Scientist	Harnham
37	Data Analyst	Dice
37	Data Analyst	Jobot
36	Machine Learning Engineer	Jobot
35	Data Engineer	Aha!
32	Data Engineer	84.51°
32	Data Scientist	Mozilla
31	Data Analyst	New Relic, Inc.
30	Data Engineer	Degreed
29	Data Analyst	WW
28	Data Analyst	NoviMarket
7252		

Using the table, we observe that the company Toptal is hiring a large number of Data Engineers (994) and Data Scientist (673). Staffing & Recruiting had the highest Average of Years of Experience (4.96) and Internet had the highest Posting Count (5052).

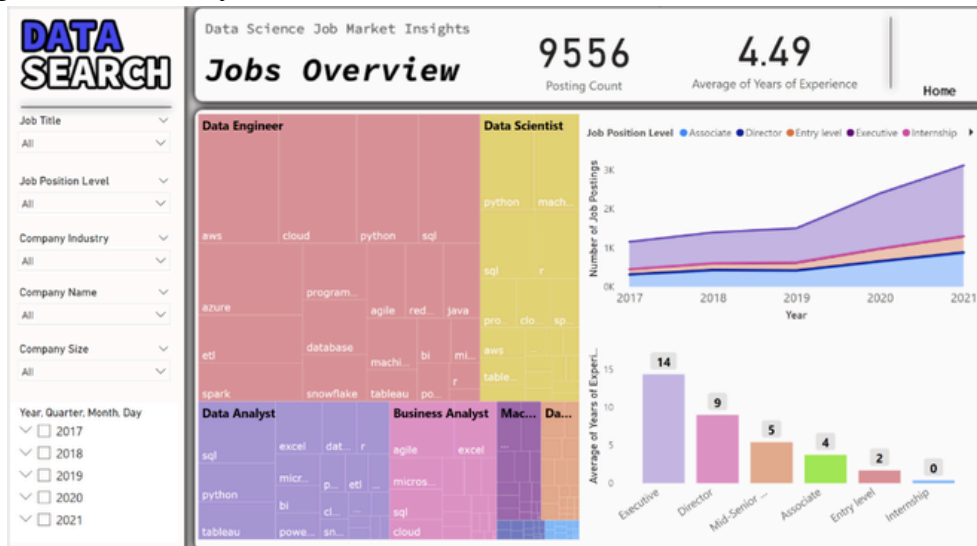
Step 3: Building Dashboard

Now it's time to hand over the final deliverable. We'll combine all of the job listing insights we've discovered into an interactive Power BI dashboard for the DataSearch team.

Begin dashboard templates

We decide to summarise our findings in three pages Jobs, Skills, Company, introduced in the Home page. We used template pictures for the background to make our final dashboard more organised.

Designing a dashboard layout



Exploring top skills for jobs

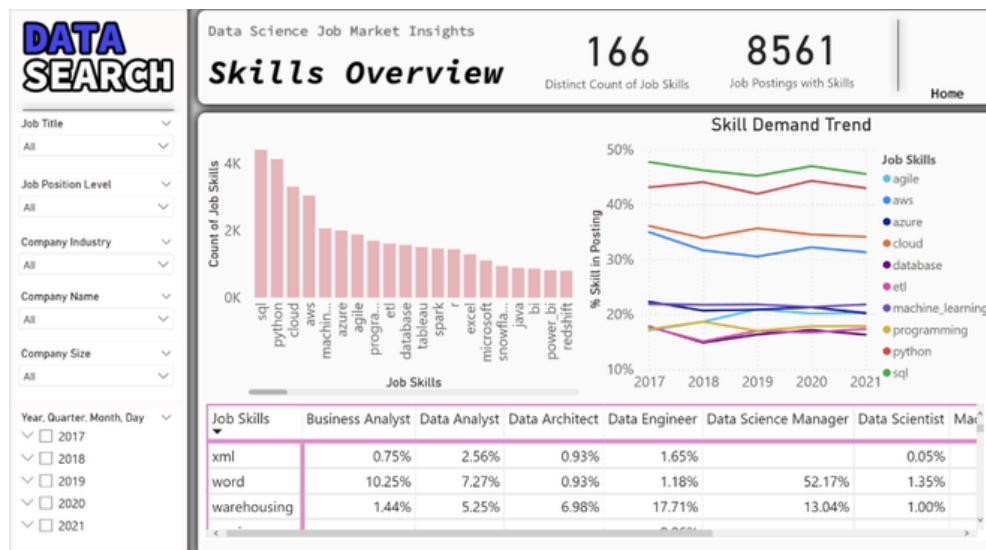
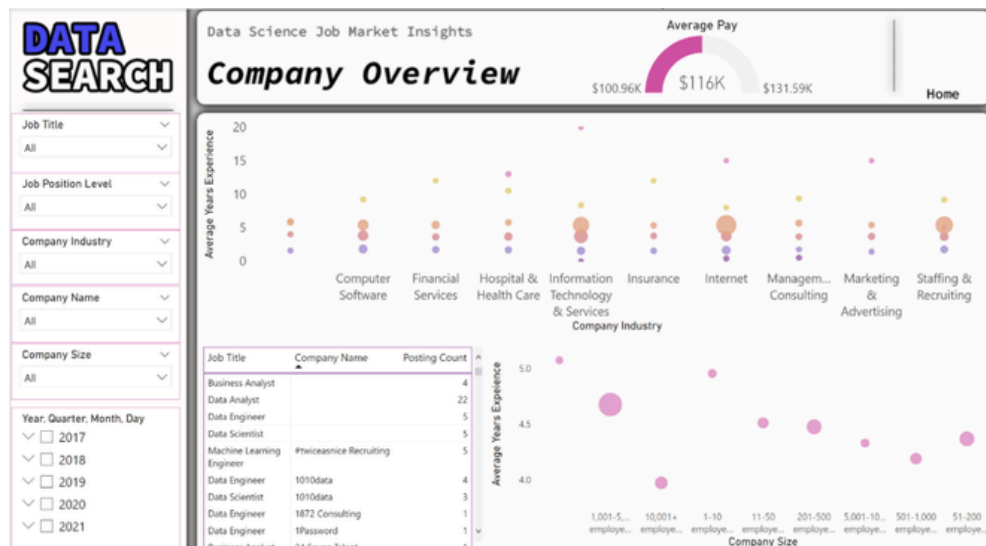


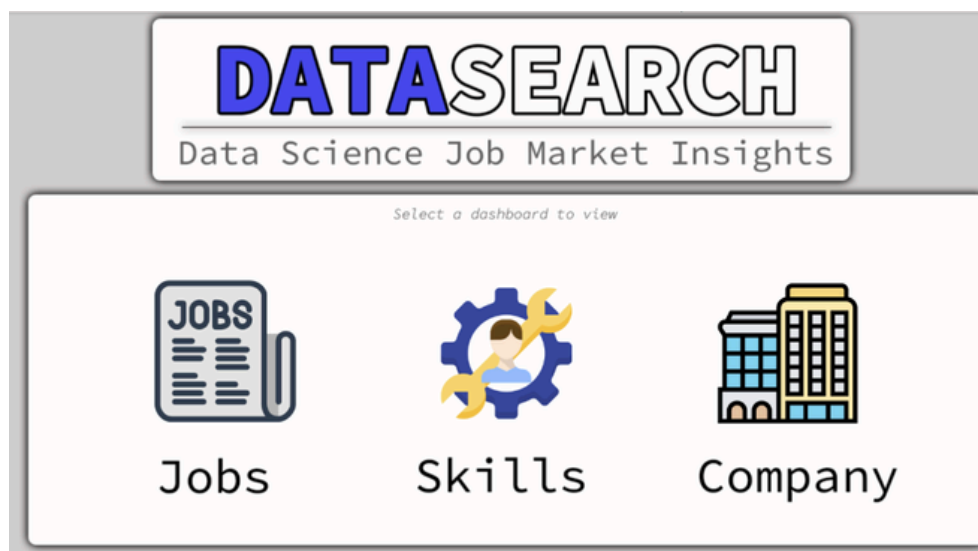
fig.12

Salary gauge for companies

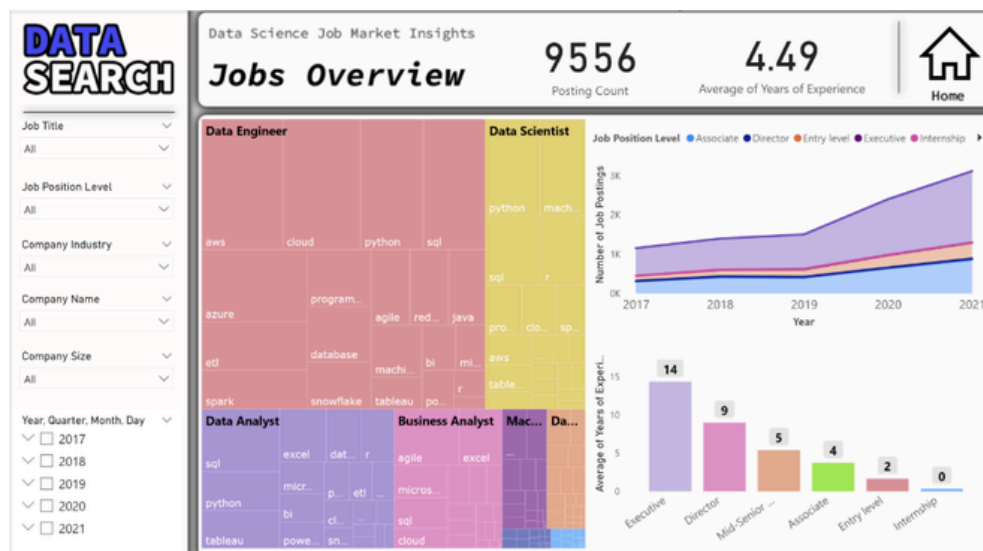


Adding interactivity with bookmarks

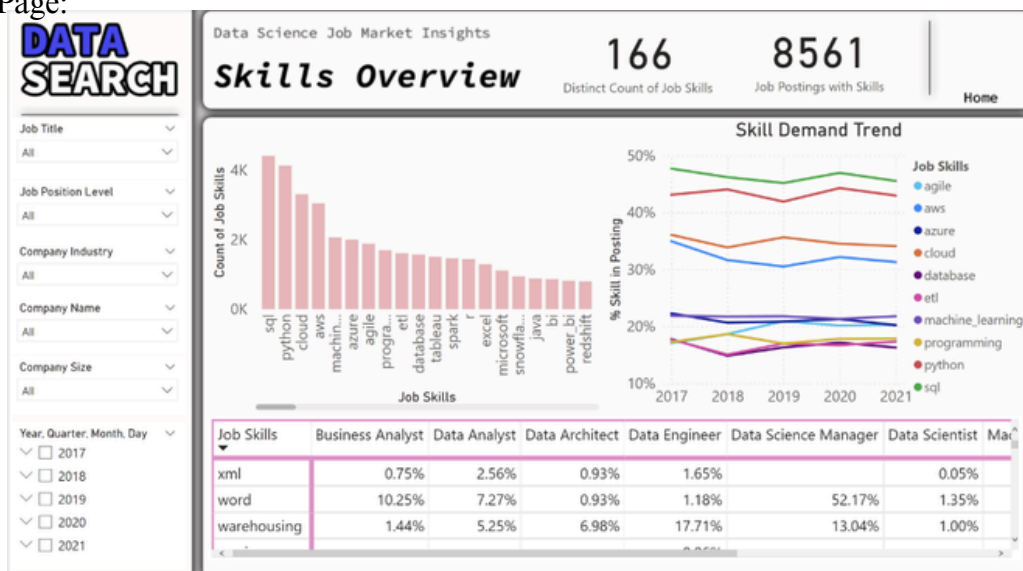
With the help of icons and bookmarking them, I was able to make my dashboard even more interactive. This is the final result where the icons are clickable and make working with the dashboard a lot easier for the end-user. This is the final result: Home Page:



Jobs Page:



Skills Page:



Company Page:

