GROUP ASSIGNMENT

ANALYSING THE JOB MARKET FOR BUSINESS ANALYSIS QUALIFICATIONS 2024

GROUP E1

1.Eranga Milan Nakandalage Don – nakey002
2.Hoang Chau Nguyen – nguhy158
3.Martin Ariel Vallazza – valma004
4.Shylyn Garcia – garsy036
5.Ushan Mark Piyasena – piyum001
6.Thi Thu Trang Nguyen – nguty440

ADVANCED BUSINESS INTELLIGENCE AND ANALYTICS

Dr Jan Stanek

10/11/2024

TABLE OF CONTENTS

BUSINESS OBJECTIVES

The objective of this project is to analyse the job market for the position of Business Analyst based on data collected from LinkedIn in 2024. ABC Recruiters need to identify and analyse important trends related to job requirements for this position, including necessary skills, potential employers, and geographical locations of job opportunities.

According to Brainstation (2024), the United States currently lacks between 140,000 and 190,000 people with full data analysis skills. Business Analyst professionals can specialize in many fields such as finance, market analysis, operations research, and cybersecurity. Datacamp (2023) identifies four core tasks of a Business Analyst: solving business problems, optimizing processes, documenting projects, and connecting technical and non-technical teams.

Specifically, the project focuses on the following main objectives:

Identify and analyse patterns related to job requirements for the Business Analyst position, including assessing the most common skills, key employers, and high-demand areas.

Evaluate recruitment needs for Business Analysts across different industries, analysing specific specialties and geographical areas with high demand for this position.

Categorize and assess the importance of necessary technical and non-technical skills, to better understand the actual requirements of employers.

Analyse the relationship between skills and job descriptions, thereby grasping the employers' needs for each specialist area of Business Analysts.

The main challenges that the project faces are:

Processing and analysing large volumes of data from LinkedIn efficiently, while ensuring the accuracy and reliability of analysis results.

Developing methods for classifying and evaluating skills listed in job postings, helping to determine the relevance and priority of each skill.

Combining skill information with job descriptions accurately and effectively, to provide detailed analyses of job requirements for each position.

The project results provide ABC Recruiters with a comprehensive view of market trends and employer requirements for the Business Analyst position. This not only helps improve the company's recruitment strategy but also supports optimizing the candidate evaluation process and developing training programs that align with the rapidly changing labour market needs.

The mind map for the business objective is provided in the appendix.

PROJECT SCOPE

The scope of the project is to identify the demand for skills in business analysis jobs. Through this project, the recruitment company can categorize applicants according to their skills in specific areas.
The data were collected from LinkedIn job postings for three countries. The data source is Kaggle.com, and the job postings were downloaded and uploaded by Asaniczka.
The Kaggle dataset is large and contains a vast number of jobs other than Business Analysis. The project only considers job positions related to Business analysis. Thus, the business analysis jobs are filtered out from the dataset.
The skills can be divided into two primary areas: soft skills and technical skills. Business analysis professionals need to have these two skill categories, and there are several skills under these categories (Datacamp, 2023).

The main soft skills considered are listed below:

- Communication
- Problem-solving
- Critical thinking
- Interpersonal skills

If there are any other significant soft skills in the dataset, they discover through the project.

The technical skills are covered the following major categories:

- Data analysis
- Visualization tools
- Statistical and quantitative skills
- BI tools skills

The data files are spread around the world, and the record count found in the dataset is shown in Table 1.

Country	Number of Records	Percentage
Australia	5,221	3%
Canada	9,741	6%
United Kingdom	8,831	5%
United States	138,251	85%

Table 01: Number of jobs posted by country

When comparing the country's records, the USA holds 85% of the dataset. Other countries account for exceedingly small portions of the dataset, each contributing less than 10%. The project's main objective is to analyse the US job market, and only the US job postings are considered for the analysis.
The project only analyses the job salary, required experience and job type according to the availability in the dataset. At this stage, the salary data is available for 587 records from 5620 total records.

PROJECT SCOPE

Stakeholders:

Business: ABC Recruiters

Course Coordinator

Assumptions:

All business analysis job positions have the term "business analysis" in the job title. The business analysis jobs are filtered from the title using the keywords "business" and "analysis."

Timeline:

Phase 1: Investigate the data files and available resources. Analyse the business requirements to identify the business objective for the recruitment process.

Phase 2: Once the business objectives are understood, this phase focuses on the data source and data quality. The team examines the data source and assesses the accuracy and availability of the data. Further, the project determines the necessary data columns for the analysis variables.

Phase 3: In this phase, the raw data goes through a cleaning process. The raw data contains complexities such as data type mismatches and unnecessary information. The team filters the necessary data and converts them to fit the analytical process.

Phase 4: Preliminary data analysis: The team conducts the first-round analysis of the dataset to identify key characteristics of the business analyst job postings. The project document presents the necessary visualizations and analysis.

Phase 5: Build the model for data analysis. Create relationships between the tables and follow data modelling techniques.

Phase 6: Analyse the data to discover patterns. Detect seasonality and trends if available. Identify any correlations between the variables to build an association rule model for the business analysis job market.

Phase 7: Reflect on the findings through the analysis

DATA EXPLORATION

DATA ACQUIRING:

LinkedIn is a social media platform for professionals. The platform helps professionals showcase their career profiles. Furthermore, some businesses and companies maintain their profiles on LinkedIn. It builds a network among professionals and companies. LinkedIn helps people and companies post available jobs in their companies or for recruitment. A recruitment person can add the job title, job description, and set of skills needed for the position in LinkedIn Talent Solutions (LinkedIn Talent Solutions, 2023).

Asaniczka has acquired a set of jobs posted on the LinkedIn platform and uploaded them to the Kaggle data repository. The entire dataset consists of three CSV files and is available for download from Kaggle.com.

Data source: Asaniczka (2024)

DATA FILES:

- Job_skills.csv

Columns:

job_link – related link for the job posted on LinkedIn.

Job_skills – set of skills needed for the job in comma-separated values.

- Job_summary.csv

Columns:

job_link – related link for the job posted on LinkedIn.

job_summary – job description of the posted job.

- Linkedin_job_posting.csv

Columns:

job_link: related link for the job posted on LinkedIn.

last_processed_time: the last time the job posting was processed.

got_summary: indicates whether the job summary was successfully extracted.

got_ner: indicates Named Entity Recognition was performed on the job posting.

is_being_worked: indicates if the job posting is currently being worked or not.

job_title: title of the job.

company: the company that posted the job.

job_location: location of the job.

first_seen: timestamp indicating when the job was first seen.

search_city: the city used as a search criterion for collecting the job posting.

DATA EXPLORATION

ABOUT THE DATA:

Note: The full data dictionary is in the appendix

The job_link is the key column across all files, used to identify the unique job posted on LinkedIn. The link contains a unique ID number that was separated from the link for data manipulation.

The job_skills column lists all required skills as comma-separated values. This column needs to be transformed to split the skills into different columns and rows. The skills include both soft and technical skills. Therefore, skills need to be categorized during the data processing stage.

We can check the characteristics of the job_skills column.

	Missing Values	Unique Values	Minimum	Maximum
Length of Skill Column	0	162	1	433

Table 02: Statistical Summary of Skill Column

According to the statistics of the column length, some skill names have up to 433 characters. There are no missing values, but the minimum character length is 1. Nevertheless, there is a chance one skill with a character is needed for the job application such as R language. However, both high and low lengths of rows need to be investigated. Therefore, the skill column needs a cleanup process to identify the proper skills.

Job_summary column is a text field with the job descriptions. The column contains detailed information about the company, job, required experience, salary and other required information. The details are on the page constructed in normal paragraph style. Every job in the column doesn't have all the information such as salary or required experience and job type. For the extraction of salary, experience and job type, the natural language process needs to be used such as Python NLTK or Spacy.

Job_location has the city, state and country information with comma separate. Even though jobs were posted from the USA, the jobs were found in their actual working location in another country. Therefore, the job location needs to be matched with the cities of the USA. The simplemaps (2024) website is used to match the USA cities with LinkedIn job locations to find the proper location. The simplemaps is recommended website by United States Census Bureau. Further, the simplemaps gives extra information such as country name (region), population and time zone.

Details of city names	Number of Records
Non-USA jobs	19
Invalid Cities	760
Valid cities	4841

Table 03: Job_location column data validation results

The search_city is the location where the job site. This location has only the city name and no information about the state or region. The search cities are mapped with the simplemaps (2024) data to validate the city names.

Details of city names	Number of Records
Invalid City Names	614

Table 04 : Search_City column data validation results

However, as the search city doesn't have the state details, it came with multiple states for the same city when mapping with the simplemaps.

JobN... Number (long)	search_city String	search_co... String	city String	city_ascii String	state_id String	state_name String
3807967910	West Covina	United States	West Covina	West Covina	CA	California
3807710016	Atlanta	United States	Atlanta	Atlanta	GA	Georgia
3807710016	Atlanta	United States	Atlanta	Atlanta	TX	Texas
3807710016	Atlanta	United States	Atlanta	Atlanta	IL	Illinois
3807710016	Atlanta	United States	Atlanta	Atlanta	IN	Indiana
3807710016	Atlanta	United States	Atlanta	Atlanta	MI	Michigan
3807710016	Atlanta	United States	Atlanta	Atlanta	MO	Missouri
3807710016	Atlanta	United States	Atlanta	Atlanta	LA	Louisiana
3807710016	Atlanta	United States	Atlanta	Atlanta	NE	Nebraska
3807710016	Atlanta	United States	Atlanta	Atlanta	KS	Kansas

Table 05: Search_City mapped records with SimpleMaps

Therefore, search_city column gives misinformation regarding the location-based analysis and has to be removed from the analysis.

Job_type column contains that job type is an on-site, hybrid or remote work. However, the comparing the job_summary column and job_type, there is a mismatch in the job type column data with the description of the job summary for some jobs. For example, the job_summary says the job is remote, but job_type column shows the job as on-site. Therefore, the job_type column needs to be transformed by comparing the job_summary details.

Column Name	Missing Values	Unique Values	10 common values
job_level	0	2	Mid senior (122957; 88.94%), Associate (15294; 11.06%)
job_type	0	2	Onsite (138229; 99.98%), Remote (22; 0.02%)

Table 06: Job_type and Job_level characteristics.

Extracting Data:

Filtering Business Analyst jobs from Dataset

The Business analysis jobs have to be extracted from the LinkedIn dataset.
Using a wildcard search from the job title the related jobs can be filtered to the different datasets.
The wildcard search for the business analysis is (.*Busin.*Analy.*)
The analyse word can be in different forms such as "Analyst, Analyzing, Analysis, Analytical etc… Hence, the "analy' is common for all forms of the word. The Business doesn't have many variation forms but "Busin" is common for any form of the business word. The job title can have many other words to specify the speciality such as Finance business analyst. Therefore, the wildcard search needs to apply for the front, middle and end to cover all forms of titles such as finance business analyst, business procurement analyst or business analysis cyber security.

Separate Soft and Technical Skills from the skill column

The skills column has all related skills as a comma-separated list. Hence, the list needs to be unpivoted to a column as row values. After that, the skills need to be separated into two categories soft skills and technical skills.
Soft skills are mostly commonly known and they can be predefined. Linked Article (Sharma, 2023) showed the highest demand for soft skills such as communication, problem-solving and analytical skills. Hence, soft skills can be found in the known words. The following wild card search flags the skill as a soft skill.

$JobSKills$ LIKE "*PROBLEM*" => TRUE
$JobSKills$ LIKE "*CRITICAL*" => TRUE
$JobSKills$ LIKE "*COMMUNICATION*" => TRUE
$JobSKills$ LIKE "*INTERPERSONAL*" => TRUE
$JobSKills$ LIKE "*NEGOTIATION*" => TRUE
$JobSKills$ LIKE "*PRESENTATION*" => TRUE
$JobSKills$ LIKE "*ACUMEN*" => TRUE
$JobSKills$ LIKE "*ANALYTICAL*" => TRUE
$JobSKills$ LIKE "*LEADER*" => TRUE
$JobSKills$ LIKE "*WRITING*" => TRUE
$JobSKills$ LIKE "*RESEARCH*" => TRUE
$JobSKills$ LIKE "*TEAM*" => TRUE
$JobSKills$ LIKE "*COMMUNICATE*" => TRUE
$JobSKills$ LIKE "*MS PROJECT*" => FALSE
$JobSKills$ LIKE "*PROJECT MANAGE*" => TRUE
$JobSKills$ LIKE "*MULTITASK*" => TRUE
$JobSKills$ LIKE "*DOCUMENTATION*" => TRUE
TRUE => FALSE

The technical skills are very wide and spread in multiple categories. Therefore, it won't be easy to search words only from known words. In this scenario, the look-up table helped to identify the different technical skills for each category.
First of all, the technical skills can be categorized as computer languages, Database, visualisation and analysis. Look-tables have two columns to match skills. The first column has the wild card search word for the skill and the second column has the proper name for the skill.

ComputerLaguageWildcard	LaguageName	DatabaseWildcard	DatabaseName
R	R	MS SQL	MS SQL Server
SQL	SQL	Oracle	Oracle
GO	GO	MongoDB	MongoDB
MATLAB	MATLAB	Casendra	Casendra
SAS	SAS	MySQL	MySQL
SWIFT	SWIFT	MS SQL Server	MS SQL Server
Scala	Scala	MongoDB	MongoDB
DAX	DAX	Cassandra	Cassandra
SPSS	SPSS	PostgreSQL	PostgreSQL
Tsql	SQL	MS Sql Server	MS SQL Server
query	SQL	Mysql	MySQL
Power Query	Power Query	Mongodb	MongoDB
M Query	Power Query	Hive	Apache Hive
		Apache Hive	Apache Hive

Table 07 : Technical skills lookup table to match skills

The wildcard columns are used to search the skills from the dataset and assign the proper technical name under the technical category.

The technical skills that cannot be categorized as language, database, or visualization tools can be investigated by grouping them. Commonly found skills can be searched categories using a wildcard rule-based search as follows..

Rule	Skill
$JobSKills$ LIKE "*MINING*"	Data Mining
$JobSKills$ LIKE "*MODEL*"	Data Modelling
$JobSKills$ LIKE "*PIPELINE*"	Data Pipeline
$JobSKills$ LIKE "*NO SQL*"	NoSQL
$JobSKills$ LIKE "*SSRS*"	SSRS
$JobSKills$ LIKE "*SSIS*"	SSIS
$JobSKills$ LIKE "*REPORT SERVICE*"	SSRS
$JobSKills$ LIKE "*SAP*"	SAP
$JobSKills$ LIKE "*MS DYNAMICS*"	MS Dynamics
$JobSKills$ LIKE "*DATA ANALYSE*"	Data Analyse
$JobSKills$ LIKE "*DATA ANALYZE*"	Data Analyse
$JobSKills$ LIKE "*WAREHOUSE*"	Data Warehouse
$JobSKills$ LIKE "*FABRIC*"	MS Fabric
$JobSKills$ LIKE "*AZURE*"	MS Azure
$JobSKills$ LIKE "*SYNAPSE*"	MS Azure
$JobSKills$ LIKE "*CHATGPT*"	ChatGPT
$JobSKills$ LIKE "*GOOGLE DATA*"	Google Data Analytics

Table 08: Regression expression to match technical skills

After going through the steps, the following outcome comes as categorised technical skills.

JobNumber_Posted	Computer Languages	Databases	Analyse and visualisation	Other Technical Skills	All Technical Skills
3794609283	javascript				javascript
3794609283		MS SQL Server			MS SQL Server
3794609283		Oracle			Oracle
3794609283		MySQL			MySQL
3794609283			Tableau		Tableau
3794609283				VISIO	VISIO
3794609283			MS Excel		MS Excel
3794609283				MS Office	MS Office
3726211542				Data Analysis	Data Analysis
3726211542				Data Analysis	Data Analysis

Table 09: Technical skill after matched with lookup values and regression expressions

Extract Salary, Experience and Job Type from job_summary column

The salary is in different forms in the column such as yearly, weekly and hourly. Therefore, when extracting the salary, it needs to be checked every format of the salary. The second issue, the salary is in full value or range value or short format with K such as 20K or \$10000 – 15000. These all different formats need to be covered in the extraction expression.

The most covered Python expression to extract the salary is the following

Updated regex to capture salary ranges and time frames in a flexible way

salary_regex = 1 to 3 digits with \$ value with optional decimal places: (\\$\d{1,3}(?:,\d{3})

or thousands separator with range indicator "-": (?:\.\d{2})?(?:k|K)?(?:\s*[-to]\s*\\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?)?)

with time period: (?:per\s*(year|week|hour|month|day|hr|w2))

Extracting experience as well is similar to the salary. However, the experience always asks for years. When examining the dataset, there are two types of year formats that can be seen in the records. They are "years" and "yrs". To locate the experience-related section or sentence in the data cell, the expression has to be looked for common words that mean "experience" such as experience, work and professional.

Regular expression to capture years of experience

experience_regex = get years or yrs with digits : (\d+)(?:[-–]\d+)?\+?\s*(?:years?|yrs?)

with the word of experience or similar : (?:experience|work|professional|industry)?|related experience required|directly related experience)?

The job type was extracted with a different method rather than salary or experience. Create three arrays with common keywords for each job type. Then, each array is used to search the keywords in the job summary column.

Define keyword categories for Remote, Hybrid, and On-site/Colocated

remote_keywords = ['remote','fully remote', 'distributed', 'remote first', 'work from home', 'work from anywhere', 'telework', 'mobile work', 'virtual work','Hybrid and remote']

hybrid_keywords = ['hybrid','mixed', 'variable', 'partially remote', 'optional', 'flex-time', 'flexible', 'flex work', 'mix of in person and remote', 'location flexible','Hybrid and remote']

colocated_keywords = ['on-site', 'on site', 'no remote work', 'in house', 'worksite', 'in person']

The result is like the table below:

Experience_years	Salary_value	Salary_time_frame	Job_type_summary
30			Remote
7			Hybrid
3			On-site
			Hybrid
4			Remote
2			Unknown
9			Unknown
5			Remote
5	$90,000	Unknown	Unknown
	$51,900.00 - $74,200.00	year	Remote
5	$90,000 - $110,000	Unknown	Hybrid
8	$76,200.00 - $114,300.00	Unknown	Unknown
8			Hybrid
	$140,000-$160,000	Unknown	Unknown
4	$78,670-$150,330	Unknown	Unknown
8			Unknown
1			Unknown
2			Hybrid
5			Hybrid
4	$126,480.00 - $163,680.00	Unknown	Unknown
5			Unknown
5			Hybrid
100	$120,000	Unknown	Remote

Table 10: Extracted experience, salary, salary type and job type results from the summary column

The Python program would be able to extract the keywords for experience, salary and job type. However, the result has several mixes of values. The experience has 30 and 100 years, there are salaries with a range and salary types are unknown and several unknown job types. Hence, the extracted result needs some manual cleaning and transformation processing to get the workable values.

- If the salary is in a range, it is calculated to get the middle value.

· If the salary is at an hourly rate, it is calculated by assuming normal working hours as 7.5. The equation is hourly salary * 7.5 * 5 * 52

· If the salary is at a weekly rate, it is calculated by multiplying 52

· If the salary is at a monthly rate, it is calculated by multiplying 12

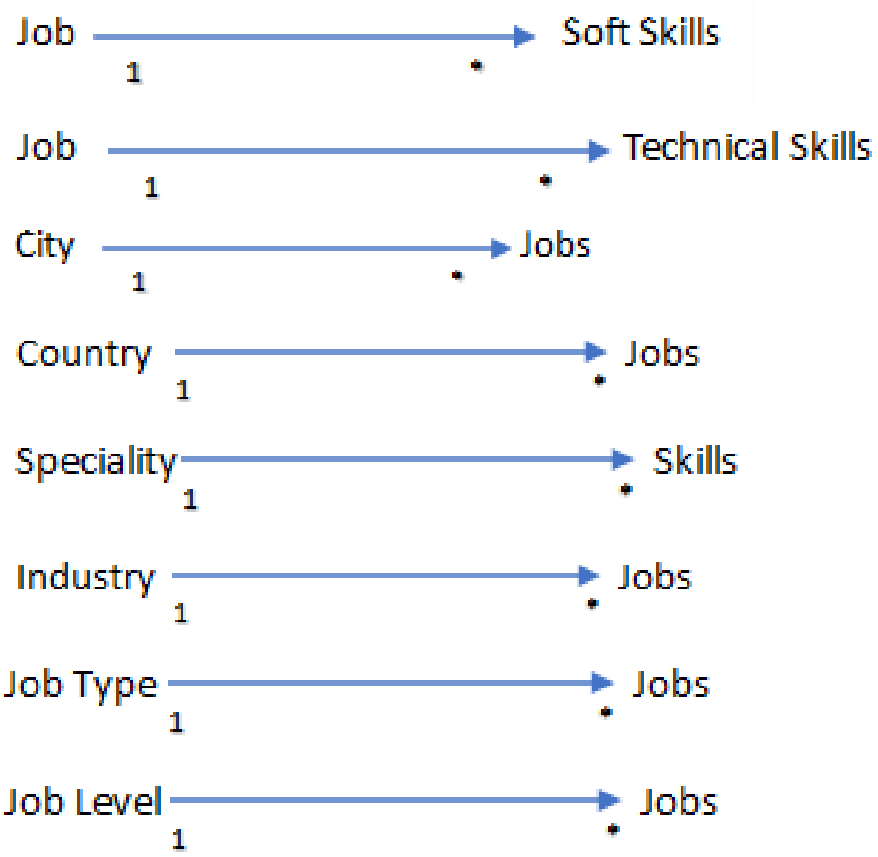
The job speciality was manually defined. After examining the job title, company and summary, each job is allocated to the relevant specialty.

Entities (Dimensions):

The dataset has various entities when analysing Business Analyst jobs. The identified entities are listed below:

- Soft Skills: Personal skills such as communication, problem-solving, and teamwork.
- Technical Skills: Specialized knowledge or expertise for specific tasks like SQL, Power BI, and Python.
- Cities: The job position's related city.
- Countries: The job position's related country.
- Specialty: The job's specific specialty in various industries.
- Professional Services: Includes sectors like Biotechnology Research, Business Consulting, IT Services, and Marketing Services. Key skills include project management, business process modelling, and client relationship management.
- Administrative & Support Services: Includes industries like Facilities Services, Staffing & Recruiting. Business Analysts focus on optimizing workflows with tools like Google Analytics, Power BI, and Python.
- Financial Services: Includes Banking, Insurance, and Investment. Key skills include financial modelling, risk management, and proficiency with tools like SAS, Excel, and SQL.
- Education: Involves higher education institutions where Business Analysts are needed as instructors with strong technical skills such as Power BI, Tableau, and SQL.
- Hospitals & Healthcare: Focuses on enhancing patient care and operational efficiency with healthcare data systems and privacy compliance.
- Job Type: Whether the job is onsite, remote, or hybrid.
- Job Level: The seniority level of the job, such as junior, associate, or senior.
- Experience years: expected experience in years
- Salary Years: Posted salary per year

Relations:



Semantic Data

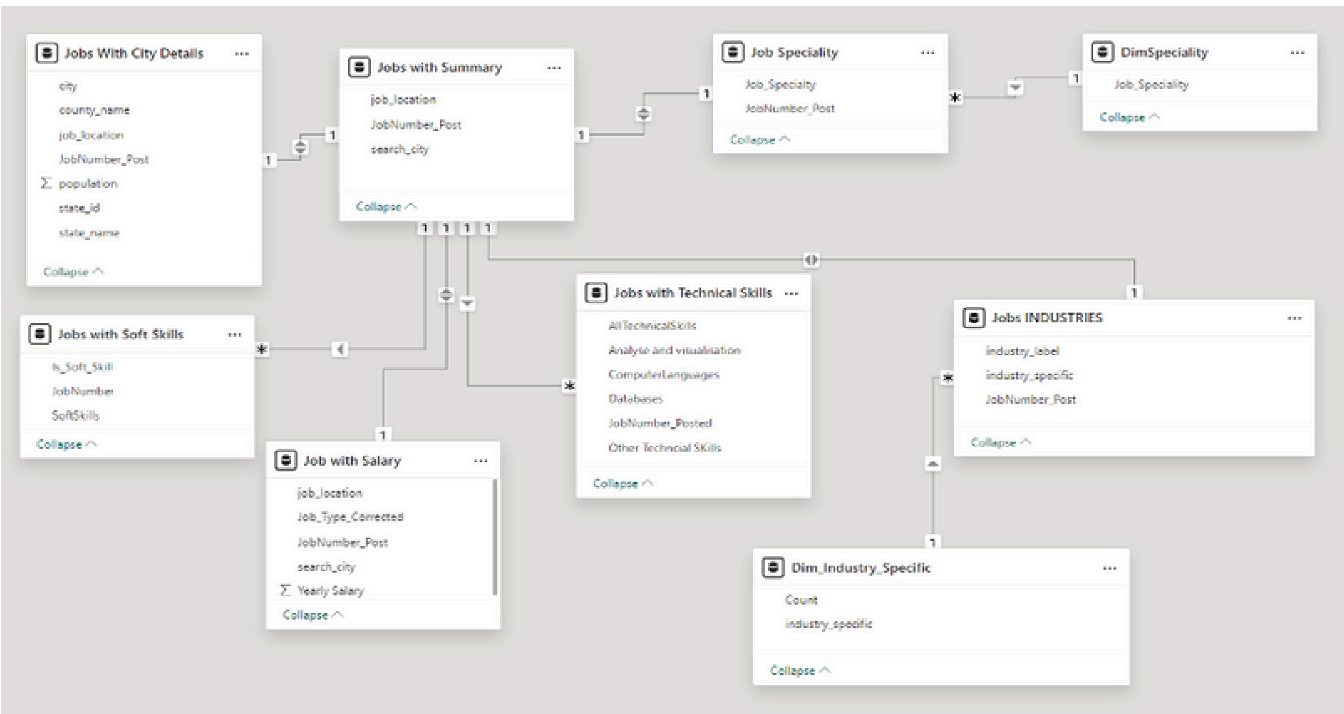


Figure 1: The semantic model of the Business Analysis Skill Analysis

DATA ANALYSIS

Our comprehensive analysis delves into the intricate patterns of Business Analyst (BA) job requirements. We uncover critical insights into job demands, skill sets, employer preferences, and geographic concentrations of BA roles, while also exploring the complex interplay between various job requirements such as technical skills, experience, and locations.

The skills requirements for each business analysis specialty

Business analysis jobs can be categorised into many numbers of speciality according to working environment. Therefore, the skills needed for each specialty may differ from one specialty to another. For example, cyber security analysts need to know about security systems and IT skills, but logistics business analysts should know about ERP systems with logistics knowledge. Therefore, it would be important to find out how the skills are related to each job's specialties.

Figure 2 diagram from IIBA (2022) shows the different specialties that come under the business analysis.

Figure 2: IIBA (2022), Business Analyst Specialty Areas

Examining the LinkedIn dataset for business analysis, several job specialties could be categorized. The job specialties and number of jobs posted under the specialty are the follows.

Specialty	Number of Jobs
Business Analyst	7885
Business Systems Analyst	2015
Senior Business Analyst	1818
Business Process Analyst	1656
Business IT Analyst	1009
Business & Finance Analyst	749
Business Intelligence Analyst	667
Business Operations Analyst	632
Senior Business Systems Analyst	498
Business & Data Analyst	485
Senior Business Process Analyst	436
Senior Business & Finance Analyst	269
Senior Business IT Analyst	216
Junior Business Analyst	198
Healthcare Business Analyst	176
Senior Business Intelligence Analyst	128
Business Support Analyst	98
Business Analyst Professor	96
Business Agile/Scrum Analyst	58
Business Management Analyst	41
Business Application Analyst	31
Senior Business Operations Analyst	24
Senior Business Support Analyst	5
Senior Business Agile/Scrum Analyst	2

Table 11: Job Specialty with number of jobs posted

Skill Requirements: Across industries, the most sought-after skills include data analysis, SQL, Python, communication, and project management. These technical and soft skills are particularly in high demand in finance, IT, and healthcare sectors. Advanced data mining techniques helped us identify these trends and their relative importance in different industries.

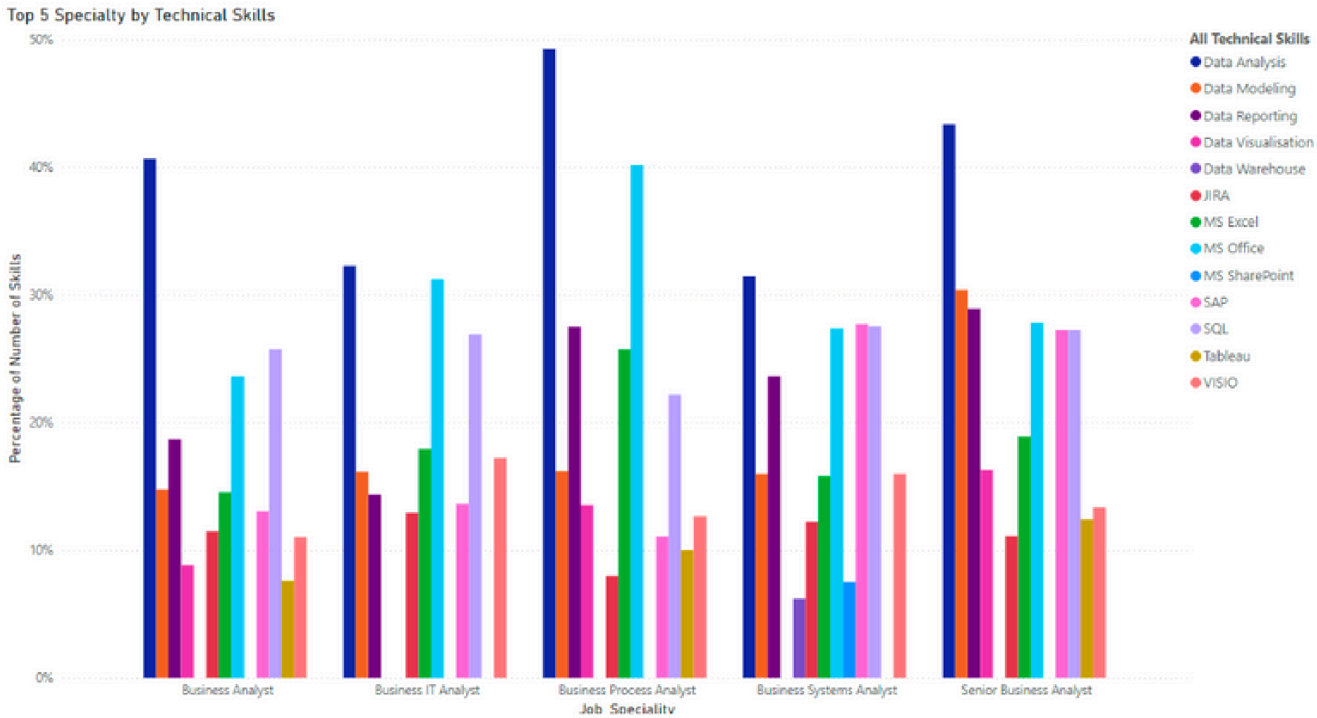


Figure 03 – Number of technical skill requests by job speciality category

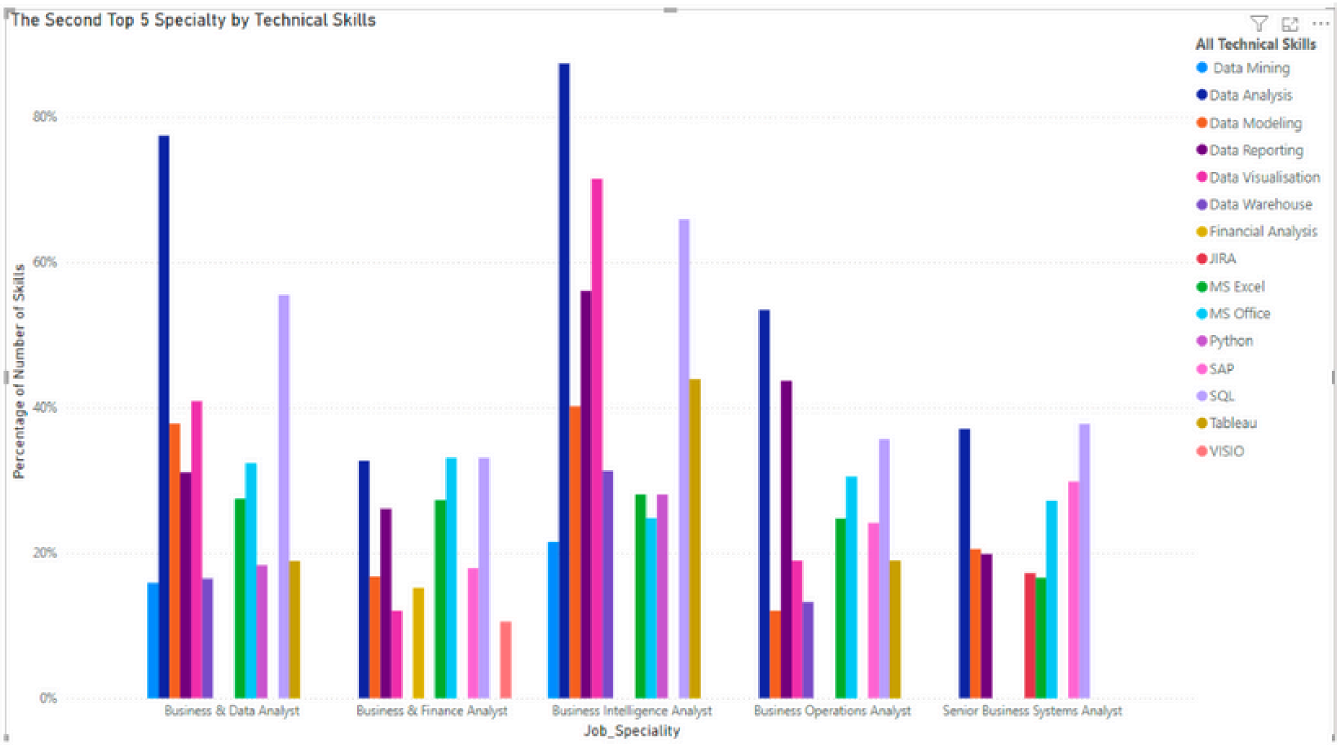


Figure 04: Number of technical skill requests by job speciality category

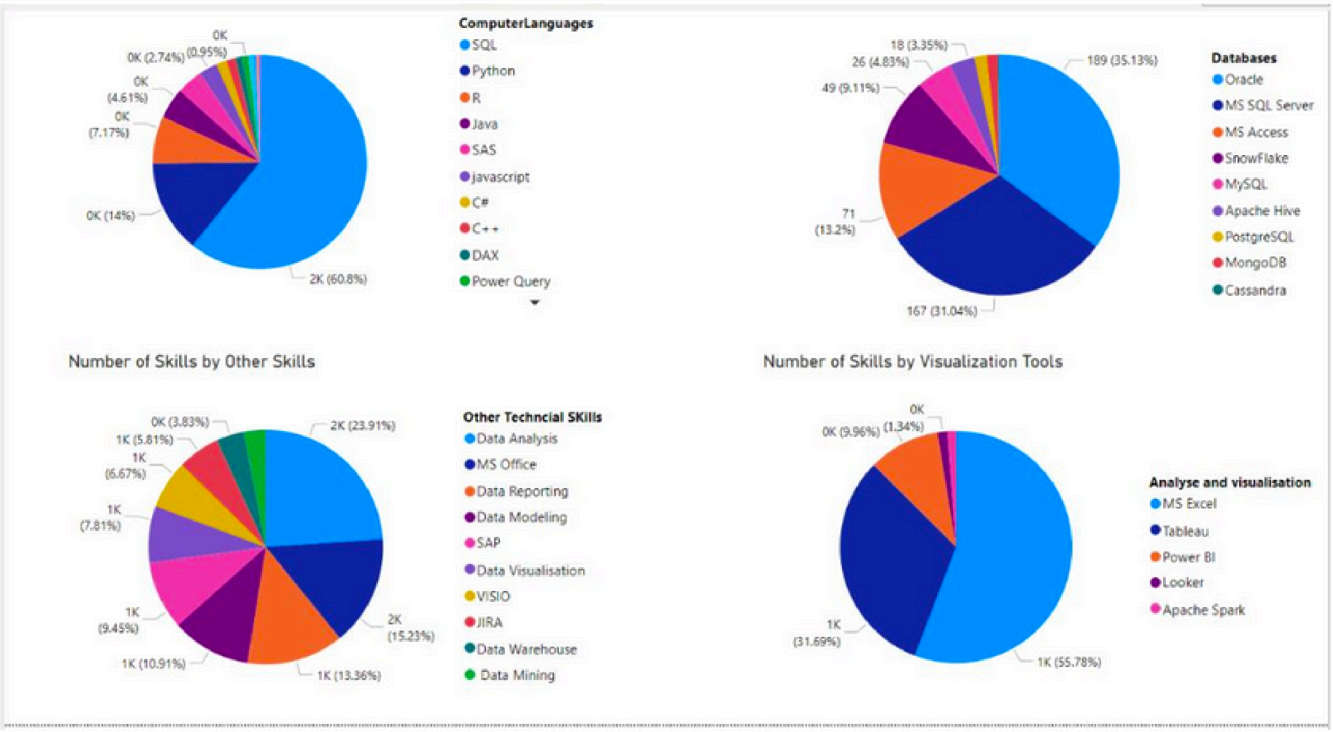


Figure 05 – Demand for technical skills in each speciality category.

When comparing technical skills, every specialty has expected data analysis skills as the highest required skill. As a technical skill, SQL has a high demand in most of specialties. Business intelligence and senior business intelligence analysts show a high demand for data visualization and data reporting skills. Further, MS Excel and MS Office are also in demand in posted jobs.

Further considering the technical skills and tools for Business analysis jobs, they can be divided into several categories according to the speciality. Applicants are capable of computer language skills, database skills, visualization tools and analytic concept skills (TechCanvas, 2024). Figure 10 shows how each skill has been demanded in each speciality. SQL is the leading demand for business analyst jobs with 60% portion of the computer language skills. While the SQL demand comes to the Oracle and MS SQL server have high demand on the database speciality. MS Excel is the leader in visualization and analytical tools speciality passing Tableau and Power BI. In other skills, Data analysis came as the highest requested skill for the business analyst. MS Office came to second place however, this needs to be investigated as MS Office is a collection of programs such as MS Excel, and MS PowerPoint. This result is a misleading demand for technical skills.

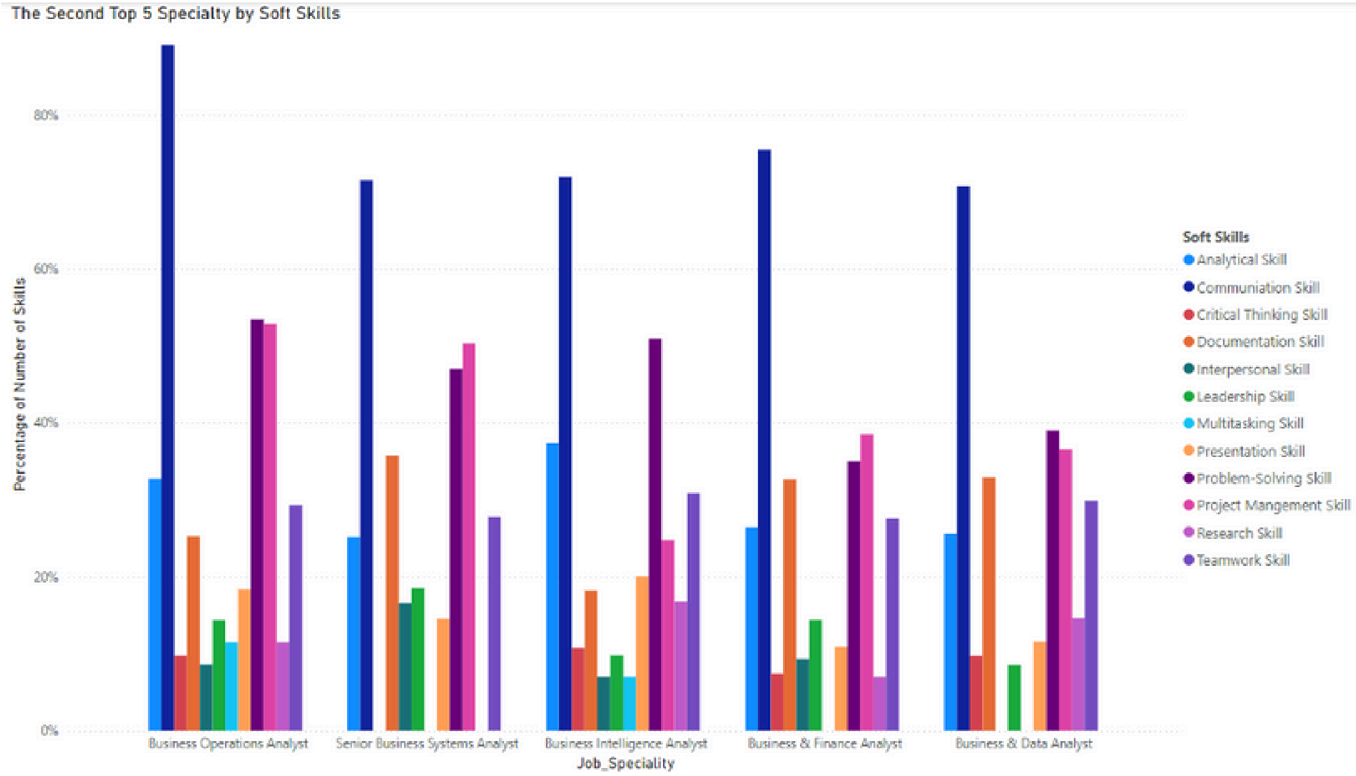


Figure 06: Percentage of each soft skill request by job speciality category

According to Figure 5 and 6, when comparing the top 10 specialties, communication skill is the highest requested skill among every skill. The problem-solving skill shows as the second highest requested skill among the specialties. Documentation, project management and teamwork skills are in third place in different specialties. The senior specialties show a high demand for Leadership skills as well.

Skill requirements for the Speciality

When considering the technical skills and soft skills bar charts in Figures 3,4,5, and 6, there is a significant difference can be seen in the speciality. Therefore, there is a hypothesis that some speciality needs special technical skills or soft skills than others.Hence, the assumption of the null hypothesis is no difference in the mean of the number of skills required for the speciality.

First, check the normality of the number of jobs to decide which test carries the hypothesis test for technical skills and soft skills.

The UNIVARIATE Procedure
Variable: Count_jobs

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.355664	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.356932	Pr > D	<0.0100
Cramer-von Mises	W-Sq	29.62123	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	145.4848	Pr > A-Sq	<0.0050

Table 12: Normality test result of the number of jobs per technical skill in each speciality.

The UNIVARIATE Procedure
Variable: count_jobs

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.381062	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.348226	Pr > D	<0.0100
Cramer-von Mises	W-Sq	11.21321	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	55.98238	Pr > A-Sq	<0.0050

Table 13: Normality test result of the number of jobs per soft skill in each speciality.

According to Tables above, the values are not in normal distribution as all p-values are less than 0.05. Therefore, the assumption of normality has to be rejected. Hence the values don't follow the normality, Kruskal-Wallis test has to be used to check the means differences.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
143.0703	23	<.0001

Table 14: Kruskal-Wallis Test result for technical skills

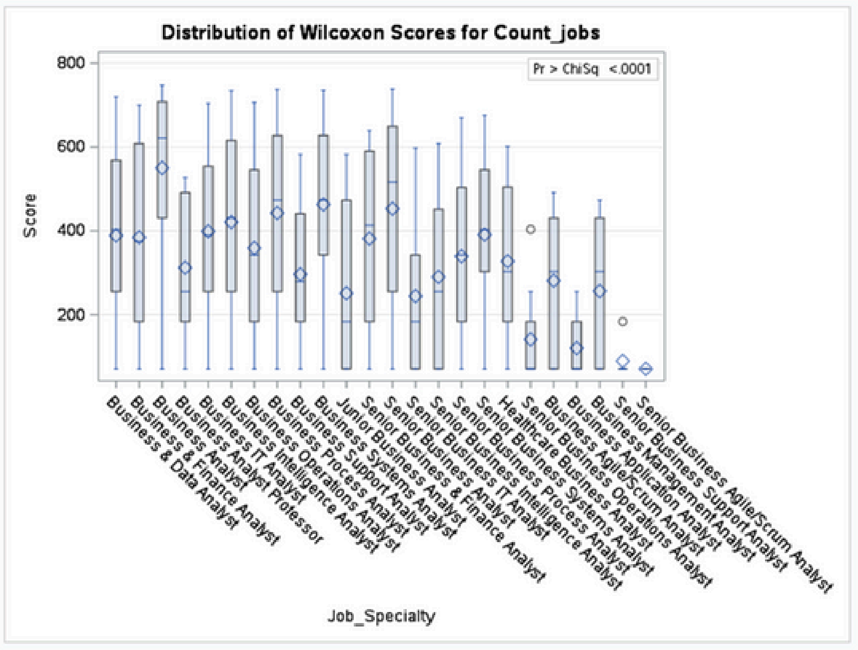


Figure 07: Boxplot of comparison of job specialties Kruskal-Wallis Test result for technical skills

The Kruskal-Wallis test gave chi-square = 143.0703 (data freedom = 23) with a p-value is less than 0.001. Hence, if the p-value is less than 0.05, we can reject the assumption of the null hypothesis of equal means. This means the specialities have different technical skill requirements. Figure 11, the boxplot shows the same results as the boxes are not aligned with the same range.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
175.0076	23	<.0001

Table 15: Kruskal-Wallis Test result for soft skills

Figure 08: Boxplot of comparison of job specialties Kruskal-Wallis Test result for soft skills

According to Table 15 and Figure 08, soft skills also show the same characteristics. The p-value is less than 0.0001 for the Kruskal-Wallis test and the null hypothesis is rejected for the assumption for equal means. The speciality has different means for each soft skill.

Checking the association of skills in each job specialty.

Considering all facts, each job specialty needs to have some special skills in technical and soft skills. Therefore, we can find how the skills are associated with each other for the specialities. LinkedIn job posts have given the skills requirements for each job.

Job Number	Skills Required
3110767987	[Data Visualisation]
3120115540	[Data Analysis]
3148221520	[JIRA, SAP]
3198529598	[Data Analysis, SQL, Python, Tableau, MS Excel, Data Analysis]
3275602758	[SQL, Data Warehouse, Data Analysis]
3282258671	[Data Reporting, Data Modelling, Data Reporting]
3324133236	[SQL, Data Warehouse]
3441455649	[Data Analysis, Data Reporting, Data Visualisation, Tableau, SQL, SQL, DAX, SSRS, SSIS]
3441460823	[SAP]

Table 16: Skills requested for each job

As shown in Table 16, each job has a set of skills from 1 to many. Some skills are listed in the many jobs and some skills are listed in one or two jobs.

The association rules are built according to the transaction of each job.

(Data analysis) -> (SQL)

(SQL) -> (Data Warehouse)

(Data analysis, SQL) -> (Tableau)

To derive the association rules, the minimum support needs to be set up.

The minimum support = * 100

Considering all facts, the minimum support is set to 10% as the items need to be listed 10 times per 100 transactions.

The next fact is the confidence level of the association rule. That tells how confident skills come when selecting another skill. The confidence level is set to 50% as it drops all less confident association rules.

The Knime workflow was used to develop the association rules for job skills.

Note: Please check the appendix for the full workflow diagram

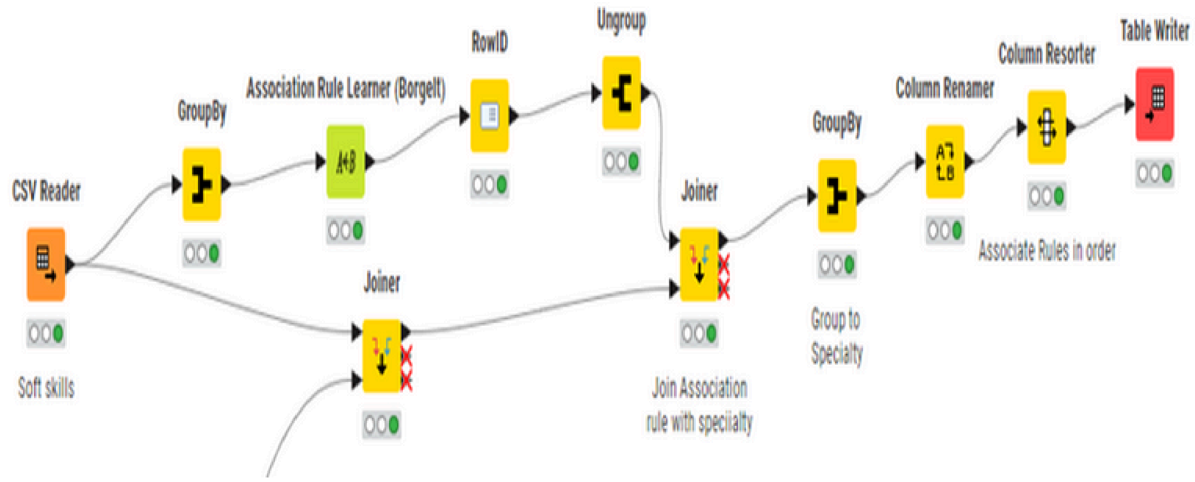


Figure 09: Association rule learner for soft skills

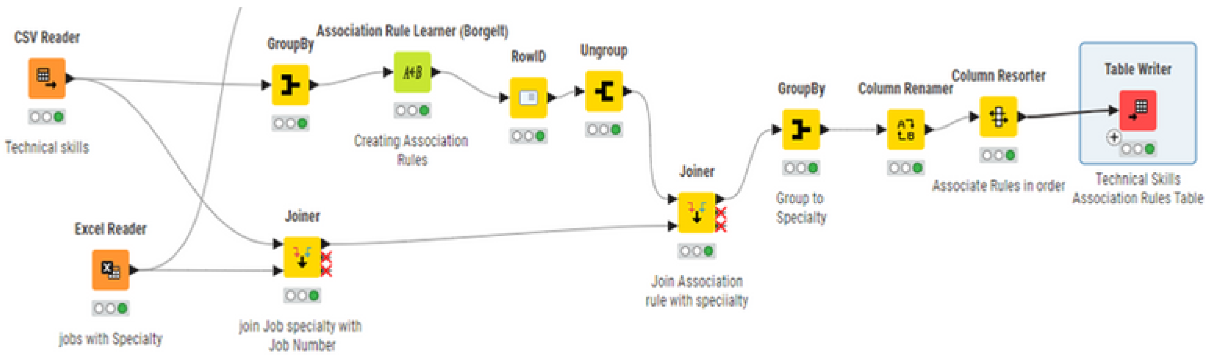


Figure 10: Association rule learner for technical skills

According to Table above, skills' medians are statistically significant. This means some skills have a higher demand than other skills.

With this result, the association rule can be built for the skills. The minimum support level is set for 5.0.

A sample of association rules built for the job skills table is available in the appendix.

Testing Association rules

The association rules can be tested using the Figure 15 Knime workflow.

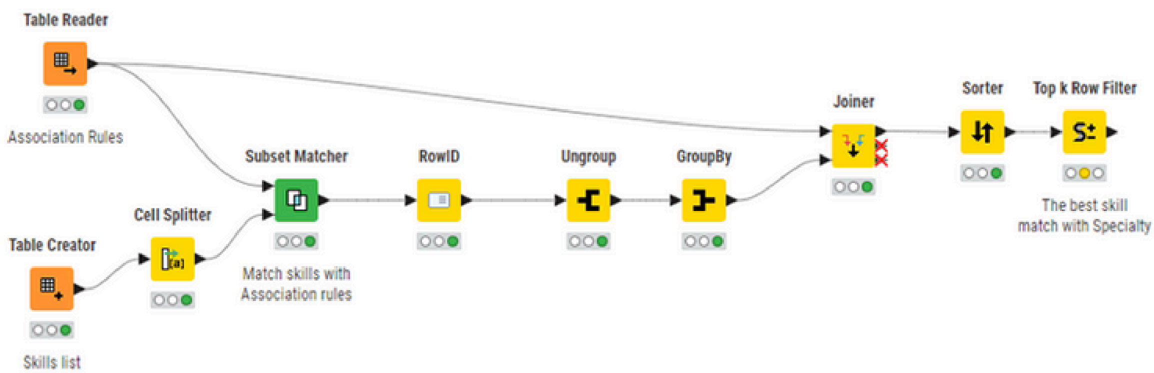


Figure 11: Test Association rule to match the best specialty

The association rule is tested with these skills Python "MS Excel" "Data Analysis" SQL to match the best specialty.

Antecedent Skills	Rule support	Rule Confident	Rule Lift %	Recommended Skills	Recommended Specialty
[SQL, MS Excel]	332	69.2	142.18	Data Analysis	Business & Finance Analyst
[MS Excel, SQL]	332	69.2	142.18	Data Analysis	Business & Data Analyst
[SQL, MS Excel]	332	69.2	142.18	Data Analysis	Business Application Analyst
[Data Analysis]	337	60.7	174.68	SQL	Business Agile/Scrum Analyst
[Data Analysis]	337	60.7	174.68	SQL	Business Application Analyst

Table 17: Association rules match skills and specialties

Geographical Demand

Our analysis of the geographical demand for Business Analysts uses visual renderings to show U.S. cities' demand for soft and technical skills.

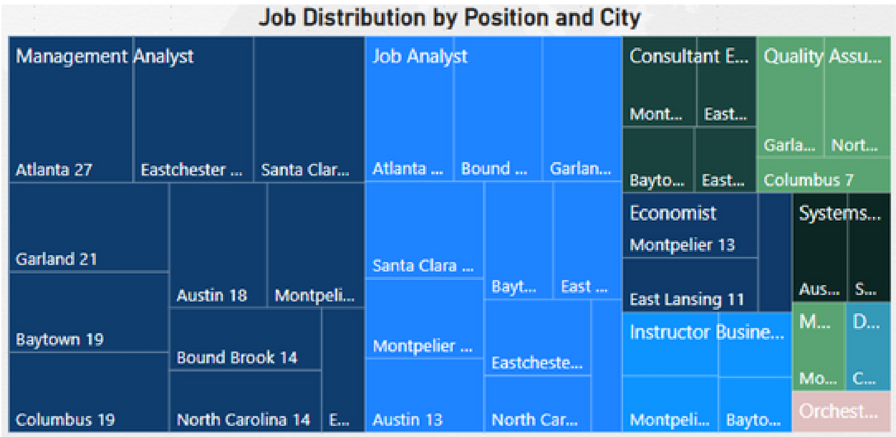


Figure 12 : Job Distribution by Position and City - A city treemap showing the most regularly occurring employment in the disciplines of Management Analyst, Job Analyst, and Consultant Economist.

These insights offer valuable knowledge for job seekers, companies, and educational institutions. Job seekers can focus on top companies and in-demand positions, while companies can adjust their recruitment strategies based on the competition in various cities. Educational institutions can use these insights to align their programs with industry needs.

TOP 10 Softskills for Business Analyst

	Soft Skills / Search Country									
	Analytical Skill	Communiation Skill	Critical Thinking Skill	Interpersonal Skill	Leadership Skill	Presentation Skill	Problem-Solving Skill	Research Skill	Teamwork Skill	Writing Skill
Search City	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States
Atlanta	19	34	4	4	14	6	26	7	17	6
Eastchester	12	27	4	4	3	14	16	6	13	9
Ferguson	12	30	3	3	5	2	21	4	15	8
Garland	22	35	6	8	1	8	19	8	20	11
Montpelier	17	59	4	5	9	12	24	5	33	10
San Diego	17	30	2	4	6	1	17	1	9	5
Santa Clara	11	28	4	1	7	9	12	5	16	4

Figure 13: Top 10 soft skills Highest Demand by Area

Figure 13 presents a highlight table with the distribution of the top 10 soft skills reported by employers when recruiting Business Analysts, based on cities in the United States. The reviewed skills are Analytical Skills, Communication Skills, Critical Thinking Skills, Interpersonal Skills, Leadership Skills, Presentation Skills, Problem-Solving skills, Research & Development skills, Teamwork skills, and Writing Skill. The darkness of colour in each cell represents the demand level for that particular skill in a particular city. For example, Montpelier reportedly requires Communication Skills most of all while Atlanta needs Problem-Solving Skills and Leadership Skills most of all. It is also evident that skills are distributed unevenly across the regions, and this distribution may be informative for soft skill enthusiasts, job seekers, and managers when used to identify trends by region.

TOP10 Technical Skills for Business Analyst

	Search Country / All Technical Skills United States									
Search City	Data Mining	Data Modeling	Data Visualisation	MS Azure	MS Excel	Python	R	SAP	SQL	Tableau
Atlanta	3	4	9		1	3	1	1	10	5
Austin	3	8	8	2	3	12	1	4	18	8
Baytown	2	5	9	2	2	4	2	12	12	5
Beverly	1	13	7		2	2	1		15	8
Columbus	1	10	9	6		3		1	13	3
East Lansing		7	1	13	1				5	
Eastchester	1	9	1	3	3	2		1	8	3
Garland	4	21	6	1	9	2	1	5	20	4
Montpelier		7	5	3	5	1		3	9	1
Santa Clara	4	4	2	1	2	2	1	1	11	5

Figure 14: Top 10 Technical Skills Highest Demand by Area

Here we provide a highlight table indicating the most promising technical skills to acquire Business Analyst jobs in different cities of the USA. The analysed skills include Data Mining, Data Modelling, Data Visualization, MS Azure, MS Excel, Python, R, SAP, SQL and Tableau. Same as in the previous table, the darkness of colour in relation to frequency; the demand value scale is provided under the table. For instance, Garland and Austin have always been most in need of SQL experts while East Lansing needs experts in MS Azure. This graphic is beneficial to help sort out which technical skills are essential in one area over another to help those in their professions and recruiting services understand what specialized skill sets are needed in each region.

Exploring Additional Skills Related to Business Analysts

Our analysis goes beyond traditional Business Analyst skills, employing association rule mining to uncover important skill correlations. For instance, we've found strong associations between data visualization, problem-solving, and SQL in job descriptions. This suggests that roles requiring high-level technical skills also highly value strong analytical thinking and the ability to present findings effectively. We've quantified these associations using metrics such as support, confidence, and lift to provide a robust understanding of skill relationships.

The Salary Effect on Business Analysis Skills

Companies may tend to pay high salaries for the applicant as demand on some special skills. So, in this section, recruitment companies can check how salaries differ for each job type and a number of skills required.

Figure 19 boxplot shows outlier points, which are salaries that exceed the range of the main data set. These salaries are often very high, such as the points near $300,000 and $350,000.

Mean: The average salary value is represented by an "X" in the middle of the box at around $100,000.

Maximum: The highest point in the chart (excluding outliers) represents the maximum salary in the main data set, approximately $175,000.

Minimum: The lowest point in the chart (excluding outliers) indicates the minimum salary, which is around $35,000.

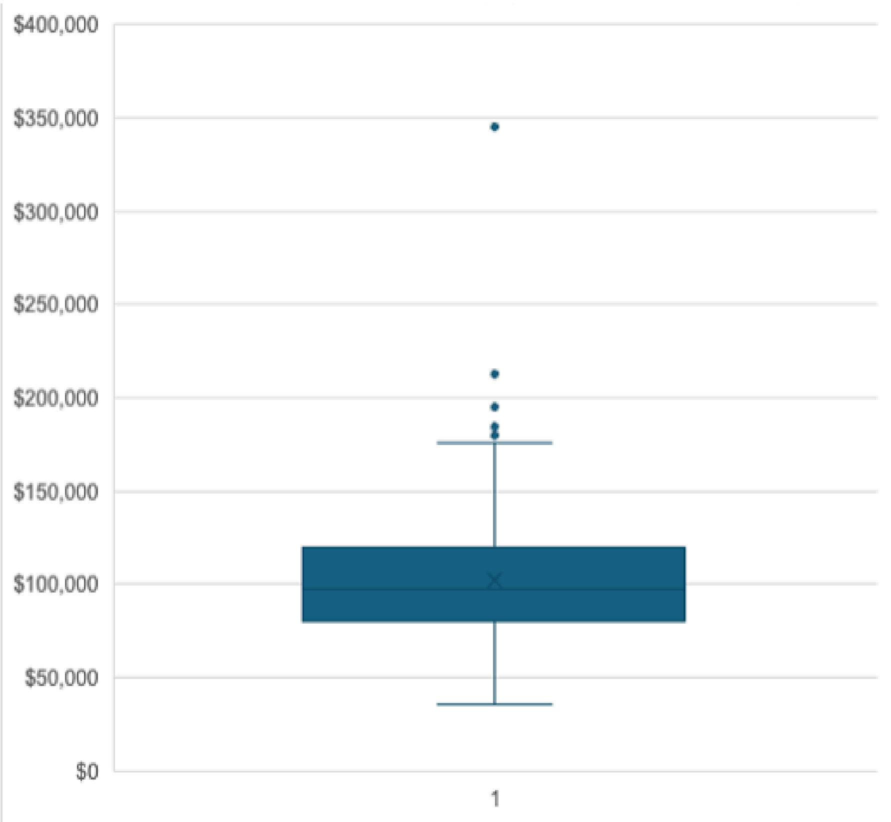


Figure 15: The Boxplot showing annual salary for Business Analyst positions

Based on the data on average salaries by job type, there is a significant difference between Hybrid, Onsite, and Remote work arrangements. Among these, Hybrid jobs have the highest average salary, reaching $102,960.28 per year, followed by Onsite at $102,337.89, and finally Remote at $99,888.52 per year. These figures reflect trends in how businesses value and compensate for different work arrangements and raise the question of whether job flexibility is becoming an important factor in today's labour market.

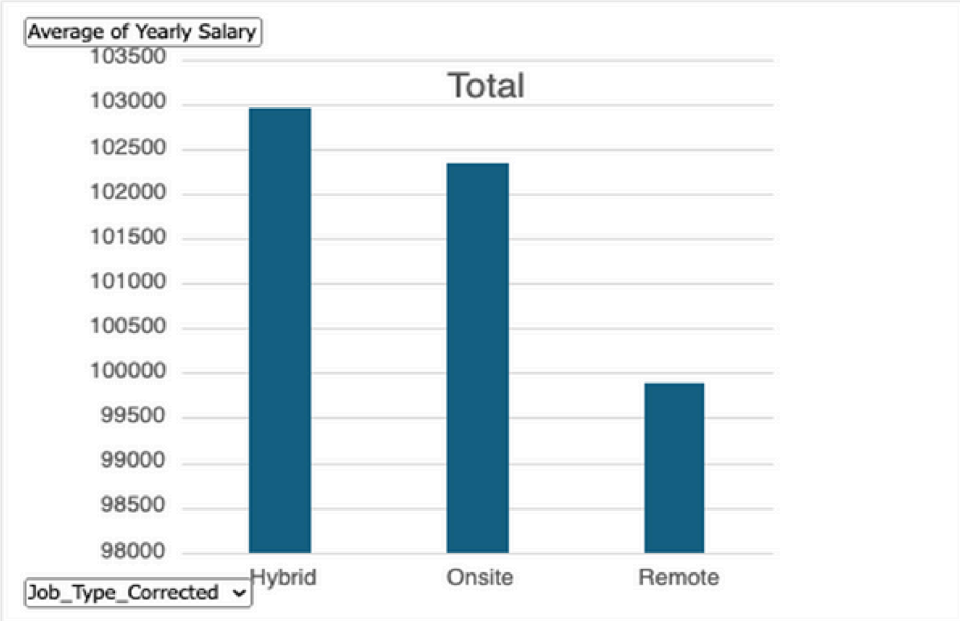


Figure 16: The salary different of each job type

Based on the data on average salaries by job type, there is a significant difference between Hybrid, Onsite, and Remote work arrangements. Among these, Hybrid jobs have the highest average salary, reaching $102,960.28 per year, followed by Onsite at $102,337.89, and finally Remote at $99,888.52 per year. These figures reflect trends in how businesses value and compensate for different work arrangements and raise the question of whether job flexibility is becoming an important factor in today's labour market.

Firstly, the salary preference for Hybrid work arrangements shows a growing trend among businesses in providing flexible working conditions for employees. The combination of office work and remote work not only brings convenience to workers but also creates added value for businesses through optimizing the workspace and increasing productivity. For this reason, Hybrid roles may command higher salaries, as they require the ability to manage work remotely while maintaining face-to-face interactions when necessary. This is a trend that many companies may consider adopting if they want to attract and retain talent in today's competitive work environment.

Secondly, the higher Onsite salary compared to Remote indicates that in-person work is still valued, especially in fields that require the physical presence of employees. These positions often have technical or managerial characteristics, where direct supervision or support still plays an important role in operations. However, the gap between Onsite and Hybrid salaries is not too large, suggesting that companies can be more flexible in offering Hybrid work options without significantly affecting overall salary costs.

Finally, the lower Remote salary may reflect cost savings for businesses when office space and related expenses are not required. However, this could also be a sign of high competition in the remote job market, where the number of workers who can participate in these jobs is very large. With the development of technology, remote work may be a long-term trend, but companies need to find ways to ensure that employees maintain work effectiveness without direct supervision.

Overall, the salary differences between work arrangements reflect changes in the recruitment thinking of businesses, where the element of flexibility is increasingly valued. Companies need to consider adjusting their work policies to ensure productivity, provide the best conditions for employees, and meet the growing need for work-life balance among workers.

Association rules for technical skills to find the best match salary

We have filtered the original data to obtain two quality datasets: Job with Technical Skill and Job Summary Salary. The goal of this step is to build a new dataset that helps illustrate the relationship between the technical skills required for jobs related to data analysis and the corresponding average salary for each skill.

To achieve this goal, the Knime tool was used to perform the process of synthesizing and combining data. Knime allows us to build a visual workflow to integrate the two filtered datasets, thereby creating a new dataset with high analytical value. Through this process, we can identify and evaluate the relationship between skills and salaries, helping to provide valuable insights into the impact of technical skills on income in the field of data analysis. Knime Workflow for Analysis Process The image below illustrates the workflow in Knime used to create the Job Summary Salary data table. This workflow includes steps from importing data from CSV and Excel files, performing grouping, and applying Association Rules to find meaningful relationships. After the data has been processed, the results are recorded in a new data table, ready for deeper analysis.

Knime Workflow:

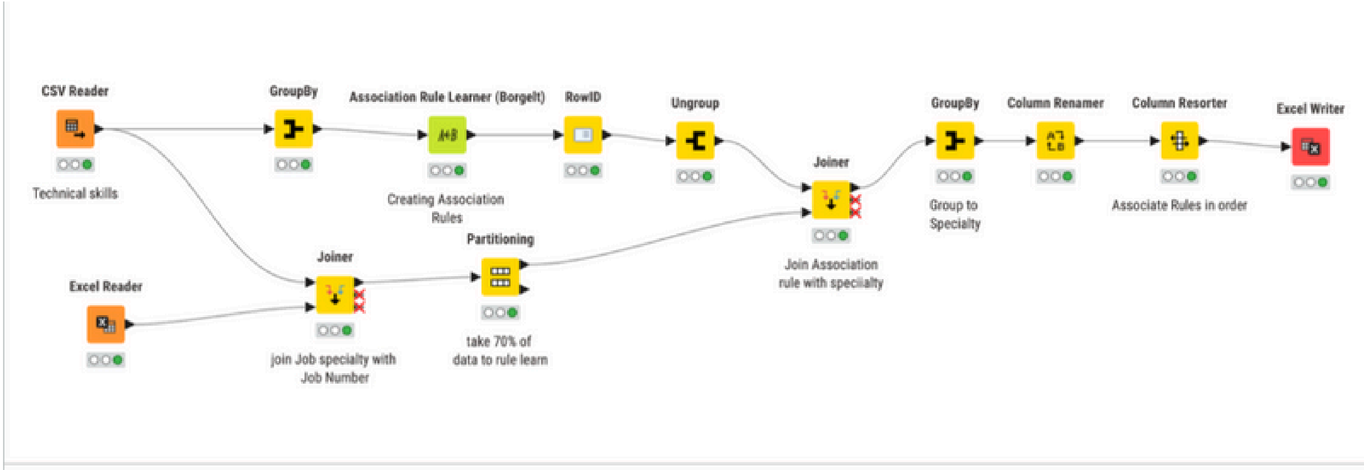


Figure 17: Job Technical Skills and Salary Association Analysis Workflow

Rule Support%	Rule Confident%	Rule Left%	Recommended Technical Skills	Antecedent Technical Skill	Mean Yearly Salary	Recommended Job Type
4.4992	77.3	222.44	SQL	Tableau, Data Visualisation, Data Analysis	111332	Hybrid
4.4992	80	164.45	Data Analysis	Tableau, Data Visualisation, SQL	109465	Onsite
4.4992	80	164.45	Data Analysis	Tableau, Data Visualisation, SQL	108076	Hybrid
4.4992	80	164.45	Data Analysis	Tableau, Data Visualisation, SQL	107011	Remote

Table 18: Sample Technical Skills Associations for Salary Summary

Job Distribution by City and Salary

This section includes American cities with the highest yearly earnings based on pay scales and job distribution. Information on employment distribution and pay can be exposed to stakeholders' location-based salary concentration trends and highly compensated work opportunities.

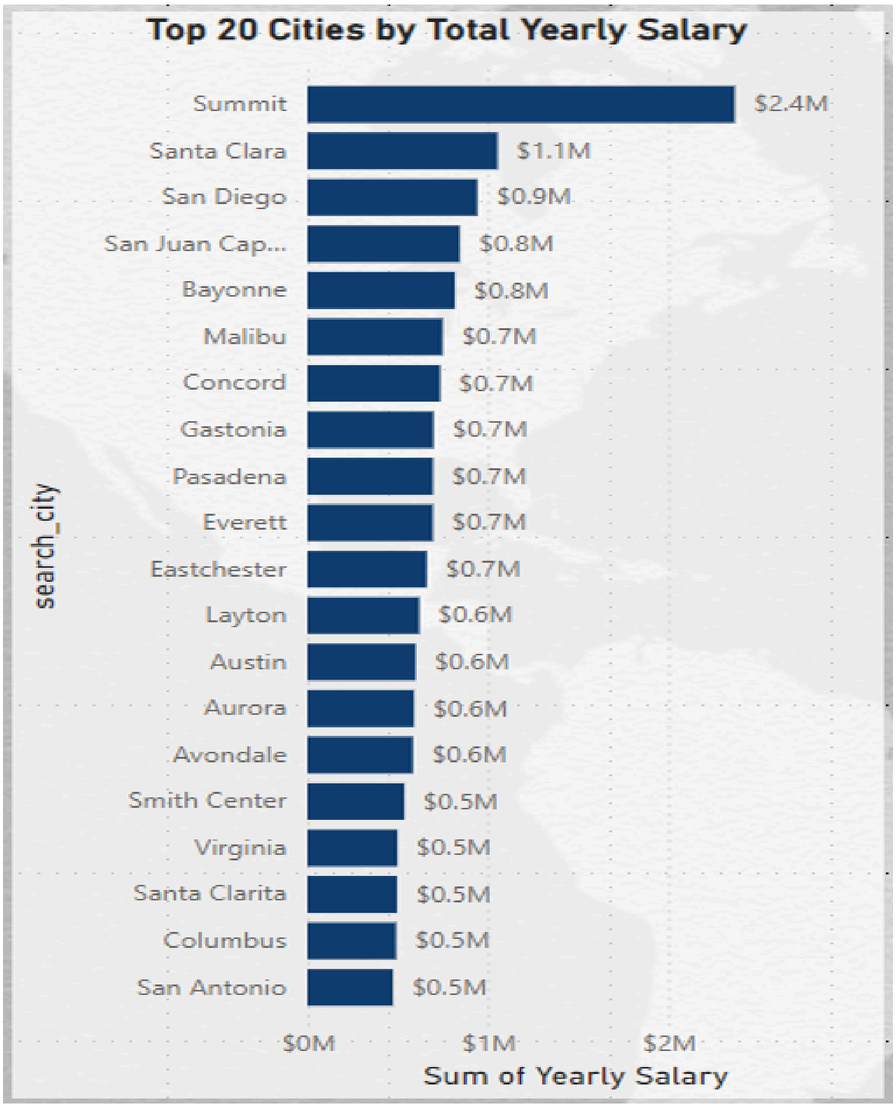


Figure 18: Top 20 Cities by Total Yearly Salary

Description: The bar chart "Top 20 Cities by Total Yearly Salary" ranks the cities with highest total pay. Summit pays \$2.37 million; Santa Clara \$1.05 million; San Diego \$0.94 million. These figures suggest that businesses seeking local talent and job seekers choose places with more highly paid employment.

Insights: Reflecting a spectrum of high-paying job areas, technology centres and less-travelled sites, the top cities by total annual income are Santa Clara and San Diego reflect high-income California; conversely, Concord and Gastonia offer competitive pay in particular industries.

The numbers show Santa Clara, Summit, and Everett pay high overall annual incomes across levels. This can be the outcome of some areas' very profitable sectors or main industries. Santa Clara lives in Silicon Valley, where demand for IT experts is high and pay rises following this. Paying top dollar for outstanding individuals, aerospace and financial services could affect Everett and Summit. Targeting these cities could offer profitable employment, even if businesses may have to pay competitively to retain qualified staff.

General Insights

A market study revealed for business analysts' significant cities and strengths. Montpelier and Columbus are looking for business analysts. This emphasizes the significance of having a complete set since soft and technical skills are required.

Needed skills and city demand are listed below:

City	Top Skills Required	Demand Level
Montpelier	Communication, Data Modeling, SQL	High
Columbus	Analytical Skills, Problem-Solving	High
Austin	Teamwork, Data Visualization	Moderate
Santa Clara	Leadership, Technical Writing	Moderate
Atlanta	Presentation Skills, Power BI	Moderate

Table 19: Needed skills and city demand

Key Insights from the Data:

1. City-Specific Demand: The great demand for Montpelier and Columbus business analysts suggests possible expansion in the industry. Often disregarded in favor of bigger cities, these little towns offer decent jobs in less saturated markets.
2. Skill Balance: Businesses seek soft as well as technological skills. Top skills were always communication and data modelling, which emphasizes the great value of technical expertise and precise presentation of difficult material.
3. Interpersonal and Technical Competencies: Business analysts have to be adaptive in SQL, Power BI, leadership, and problem-solving. Their capacity to balance numerous roles qualifies them as important in many different fields.

Data show Montpelier has a lot of jobs for business analysts. While some say New York and San Francisco control this field, other cities also generate employment. The apparently broader labor market than expected allows professionals to grow outside of IT hubs.

Key Findings on City Prominence:

City	Expected Prominence	Actual Demand	Key Skills in Demand
Montpelier	Low	High	Communication, Analytical Skills
Columbus	Moderate	High	Problem-Solving, Data Modeling
New York	High	Moderate	Business Analytics, Technical Writing
San Francisco	High	Moderate	Data Visualization, Leadership
Santa Clara	Moderate	High	Technical Skills, Teamwork

Table 20: Key Findings on City Prominence

Analysis and Implications:

1. Unexpected Prominence of Smaller Cities: Here surprise came from Montpelier, a city not known for business analysts. This suggests that smaller towns could have competitive job markets and growing companies with value for corporate analytics.
2. Competitive Advantage in Smaller Markets: Perhaps more than smaller cities, big cities could compete. Job seekers could so look at occupations in fields with less competition but high demand.
3. Strategic Considerations for Job Seekers: Job seekers should go outside traditional sites given this dispersion. Smaller cities could provide better chances for professional growth away from the saturation of large metropolis.

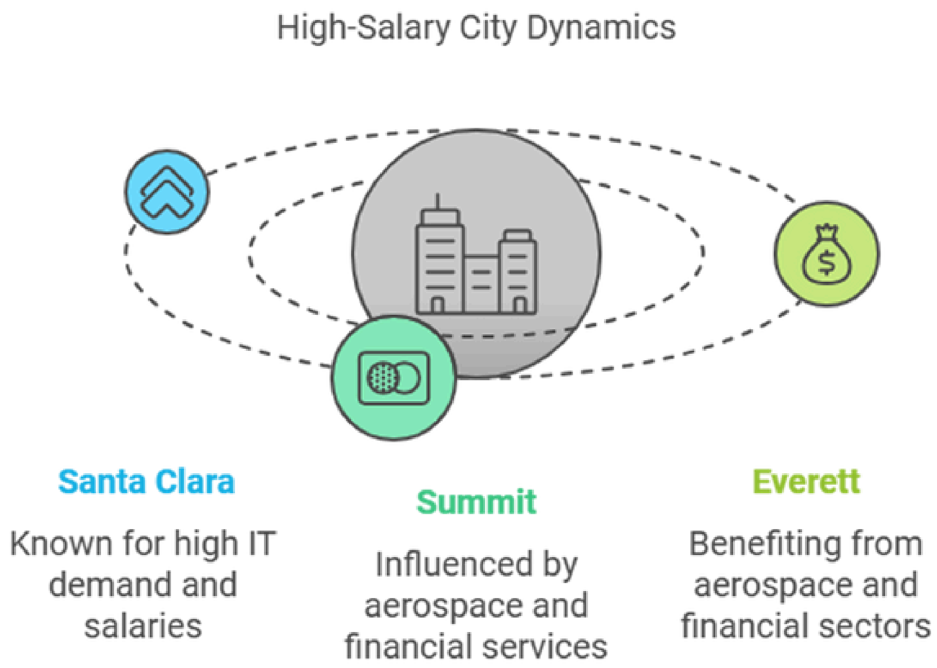


Figure 19: High-Salary City Dynamics Overview

Regional Skill Demands

The demand for regional abilities is somewhat different. Seattle and New York, centers for banking and IT, call for tech knowledge. Businesses in these specialized cities call upon engineers, data analysts, and software programmers. Businesses that depend on soft abilities frequently prioritize interpersonal communication, customer service, and healthcare among other things. Job seekers could find it helpful to match their qualifications to the leading industries in these areas, and companies could use these particular talents to boost hiring.

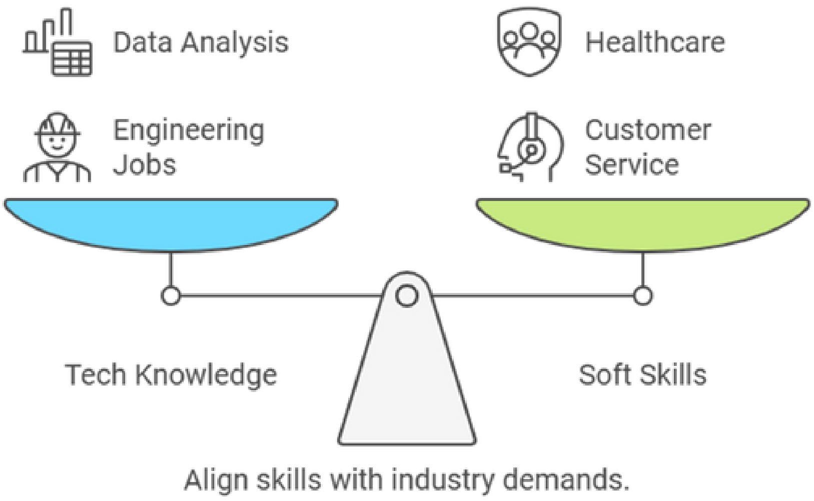


Figure 20: Regional Skill Demand Balance: Tech vs. Soft Skills

Reflection for Job Location

This study of employment distribution, salary trends, and qualifying requirements can help job searchers, businesses, and governments all over many American cities.

For Job Seekers

Target Cities with High Demand for Specific Skills: Targeting skill-demanding cities will assist job seekers. The "Technical Skills Job Distribution by City" graphic displays techies either working remotely in tech hotspots or cities with more job openings.

Leverage Qualifications for Higher Salaries: Candidates for master's and PhD degrees should search for employment in areas where their particular work would be adequately compensated. Based on qualifying tables, North Carolina (PhDs) and Boulder (Master's) pay more but lure highly qualified people.

Figure 21: Job Search Strategy for High-Demand Skills and Advanced Qualifications

Industry-Specific Requirements

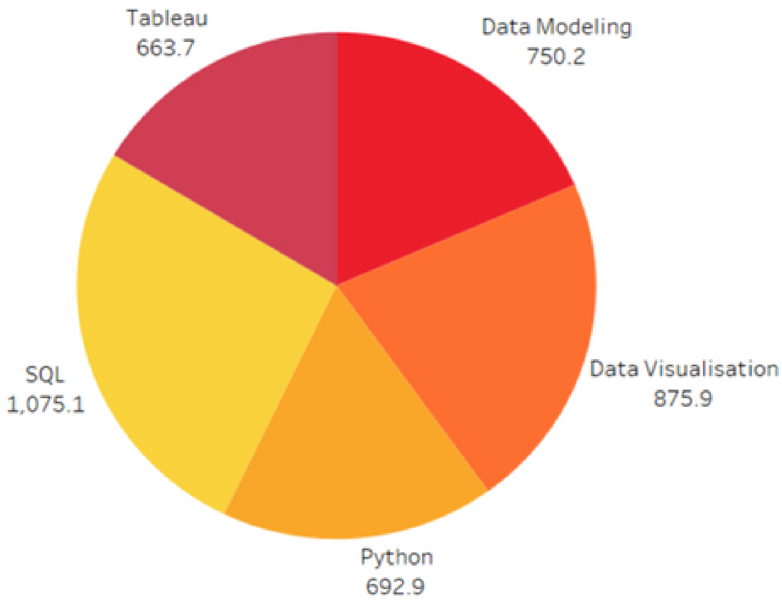


Figure 22: Top 5 Technical Skills by Industries

The most in-demand technical skills across industries are SQL, Data Visualization, Data Modeling, Python, and Tableau, reflecting the widespread need for data analysis, interpretation, and visualization capabilities in Business Analyst roles.

Industry-Label Technical Skills Analysis

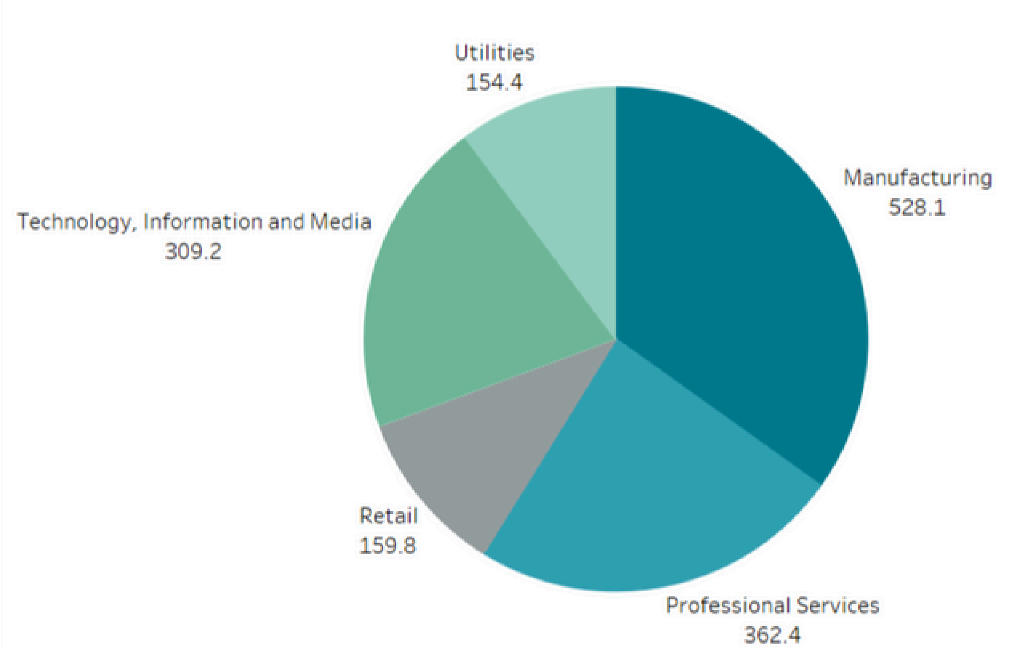


Figure 23: Top 5 Label Industries Requiring Technical Skills

The top technical skills are consistently in demand across several industries, including Utilities, Retail, Manufacturing, Professional Services, and Technology, Information, and Media. Each of these fields relies heavily on technical skills to enhance productivity, streamline operations, and remain competitive.

Industry-Specific Technical Skills Analysis

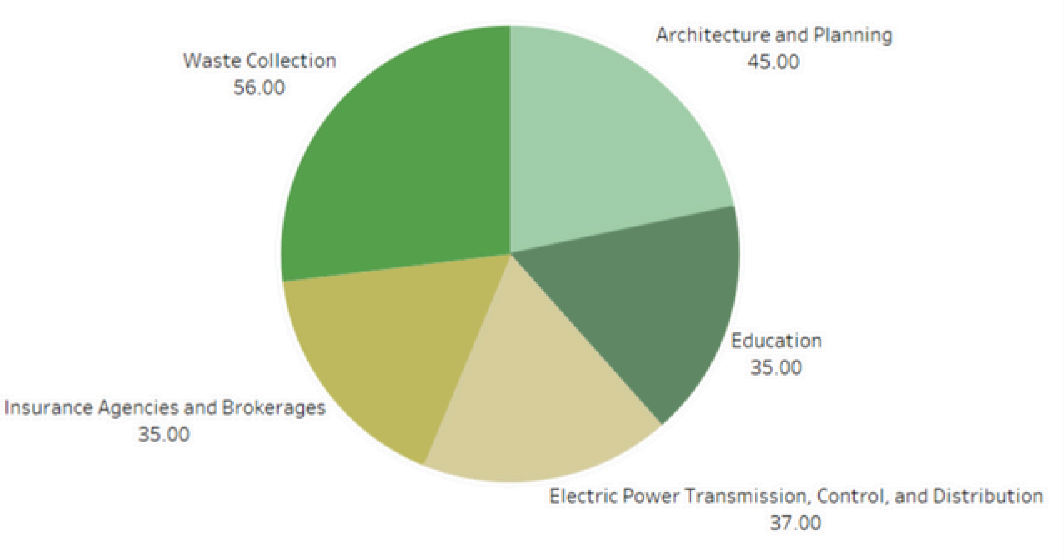


Figure 24: Top 5 Specific Industries Requiring Technical Skills

The top five specific industries with high demand for technical skills include Architecture and Planning, Education, Electric Power Transmission, Control, and Distribution, Insurance Agencies and Brokerages, and Waste Collection. Architecture and Planning rely on data modeling and visualization for design and project planning, while Education utilizes data visualization and SQL to analyze student performance and improve outcomes. The Electric Power industry emphasizes data modeling and SQL for managing complex networks efficiently. Insurance Agencies require SQL and data visualization for effective risk assessment and client insights. Waste Collection has a strong need for data analysis skills to enhance operational efficiency and comply with regulatory standards.

Soft Skills Analysis

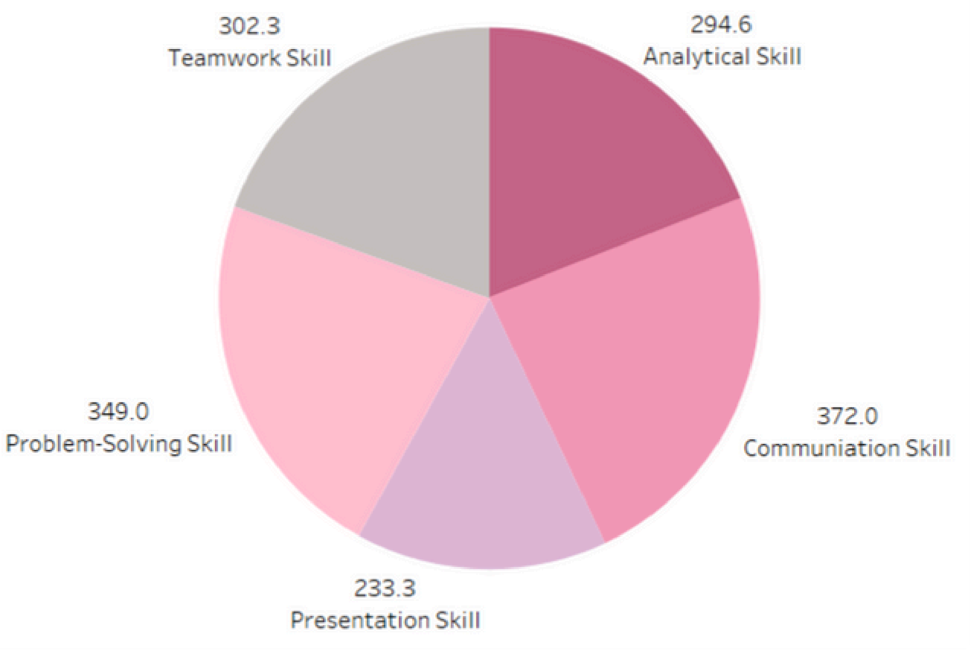


Figure 25: Top 5 Soft Skills Required

The most listed soft skills across industries for Business Analyst roles include Communication, Problem-Solving, Teamwork, Analytical Skills, and Presentation Skills. These skills are essential for effective collaboration, critical thinking, and the ability to convey insights clearly, all of which are vital in a data-driven business environment.

Industry-Label Soft Skills Analysis

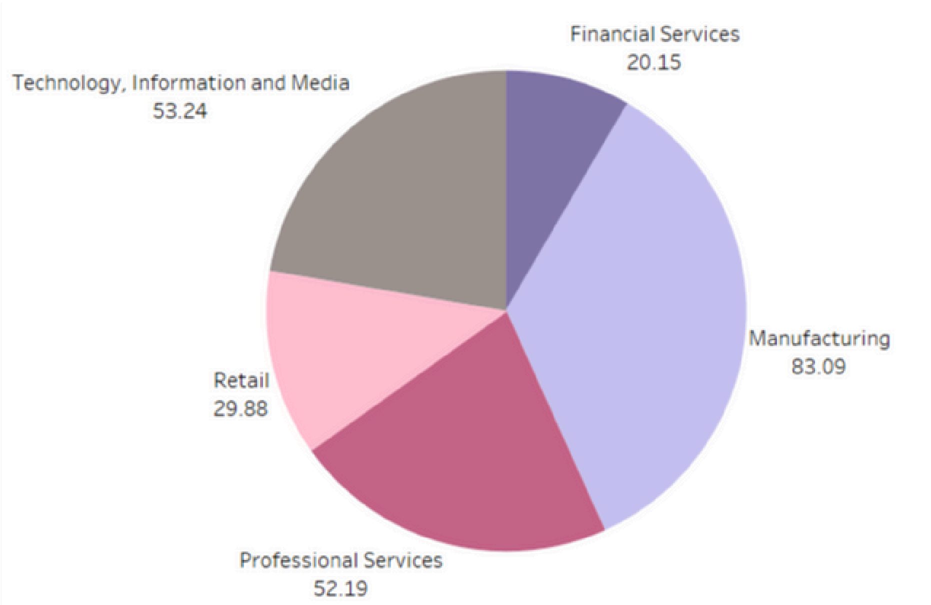


Figure 26: Top 5 Label Industries Requiring Soft Skills

Manufacturing has the highest demand for soft skills, essential for managing operations. Professional Services and Technology, Information, and Media also require strong soft skills for client interactions and teamwork. Retail and Financial Services show a lower emphasis on soft skills, suggesting a greater focus on technical skills in these fields. Each of these industries relies strongly on soft skills to improve collaboration, enhance problem-solving, and maintain a competitive edge.

Industry-Specific Soft Skills Analysis

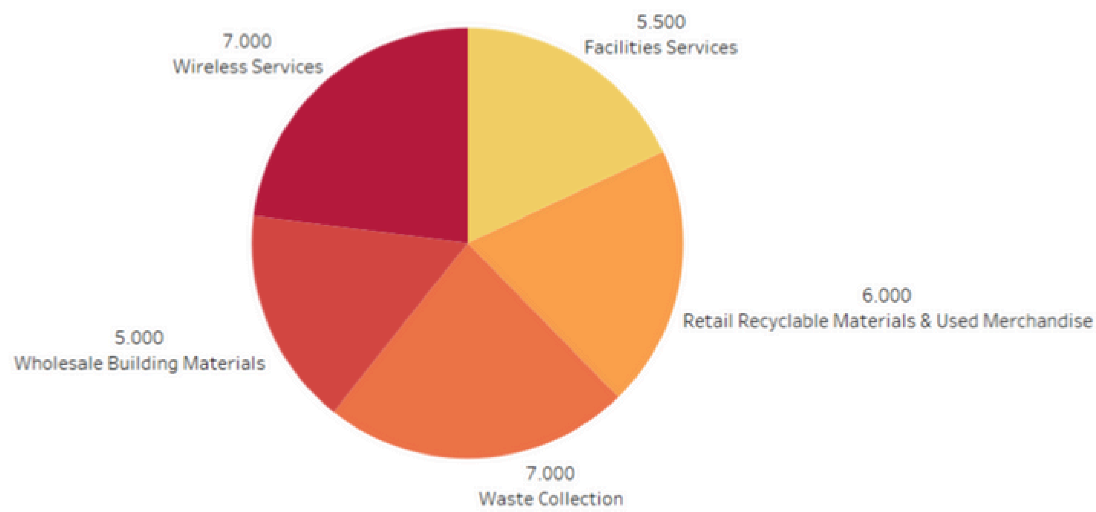


Figure 27: Top 5 Specific Industries Requiring Soft Skills

Wireless Services and Waste Collection industries show a high need for communication and teamwork, essential for managing customer interactions and coordinating field operations. Retail Recyclable Materials & Used Merchandise, along with Facilities Services, emphasize problem-solving and organizational skills, supporting the efficient management of inventory and compliance with environmental regulations. Wholesale Building Materials also highlight a demand for problem-solving, given the need to manage logistical challenges and customer relationships effectively.

Key Insights
Core Technical Skills: SQL and Data Visualization are universally required, serving as foundational skills across nearly all industries.
Non-Traditional Data Industries: Industries like Waste Collection and Architecture increasingly value data-driven decision-making, highlighting an expanding role for analytics.
Soft Skills Emphasis: Communication, Problem-Solving, and Analytical Skills are universally critical, underscoring the importance of effective collaboration and adaptability in Business Analysis.

The following table, derived from our comprehensive data analysis, compares the required skills and tools across different industries:

Industry	Key Skill Requirements	Important Tools	Project Management Methodologies
Finance	Understanding of financial principles, financial modeling, risk analysis	SAP, Oracle Financials, Excel	Traditional, Waterfall
Information Technology (IT)	Understanding of software development processes, system integration, programming skills	Jira, Confluence, Git	Agile, Scrum
Healthcare	Understanding of medical terminology, privacy laws, electronic health record (EHR) management	EHR, HL7 technology, FHIR, Tableau	Six Sigma, Lean
Retail	Understanding of consumer behavior, supply chain management, inventory optimization, CRM management	CRM systems, POS systems, Google Analytics, Shopify, Tableau	Agile, Lean
Manufacturing	Understanding of process optimization, supply chain management, quality control, lean manufacturing	ERP systems (SAP, Oracle), IoT platforms, MES (Manufacturing Execution Systems), Tableau	Six Sigma, Lean, Agile

Table 21: Key Skills by Industry for Business Analysts

By rigorously examining these association rules and applying advanced statistical techniques, we've gained deep insights into how specific skills, experiences, and qualifications align with particular job specialties and locations. This data-driven knowledge empowers both job seekers and employers to optimize their search and recruitment strategies effectively.

Our comprehensive analysis of Business Analyst jobs, leveraging advanced data mining and machine learning techniques, reveals significant variations in skill requirements across industries and locations. Through a sophisticated combination of skills gap analysis, industry-specific demand modelling, and geo-trend exploration, we've identified key patterns and associations that illuminate how employers value different skill sets.

This analysis underscores the critical need for modern technical skills such as data analysis, SQL, and AI, coupled with strong communication abilities, to remain competitive in the rapidly evolving job market. Our findings provide a robust foundation for strategic decision-making in career development, hiring practices, and educational curriculum design within the Business Analysis field.

The hypothesis for the industry-specific analysis posits that Business Analyst roles exhibit substantial variations across different sectors. This proposition underscores the critical importance of identifying and understanding the unique skill sets and domain knowledge required in each industry. Our investigation aims to delve into the distinct requirements for Business Analysts in key sectors such as Finance, Information Technology (IT), Healthcare, Retail, and Manufacturing. This hypothesis implies that success in Business Analysis is contingent upon acquiring industry-specific expertise and mastering tools tailored to each sector's unique challenges and operational environments. By exploring these sector-specific demands, we aim to provide valuable insights for both aspiring Business Analysts and employers seeking to optimize their recruitment strategies.

Skills Gap Analysis

Through comparative analysis of job requirements and candidate profiles, we've identified a significant gap in technical skills, particularly in data analysis and visualization tools. This gap is especially pronounced in sectors like finance and IT. Our predictive models suggest that addressing this gap is crucial for meeting current and future job market demands.

As per Figure 7 the entire market for BA with PhD level is comprised of 6 available roles. Of which two are required for roles in the Education industry as Professors.

PhD Level Qualifications

Job Specialty	Search City	PhD	PhD, Master
Business & Finance Analyst	Austin		■
Business Analyst Professor	Aurora	■	
	Boulder	■	
Business Process Analyst	North Carolina	■	
Healthcare Business Anal..	Levittown		■
Senior Business Analyst	Tarrytown		■

Figure 28: PhD Level Qualifications

Salary and geographical analysis

The salary and geographical analysis indicate that the top 10 cities in terms of available roles are as per Figure 29.

Job Location	
Austin, TX	8
Bellevue, WA	8
Charlotte, NC	14
Denver, CO	9
Irvine, CA	9
Los Angeles, CA	8
Manhattan, NY	9
New York, NY	30
San Diego, CA	14
Seattle, WA	18

Figure 29: Top 10 cities in terms of available roles

New York is clearly the market with the most roles available, however from a population perspective Denver is the city with more roles per capita as per Figure 9 and Los Angeles is the most competitive market with 1 role every 479,364 habitants.

City	Roles Available	Population	Roles per capita
Austin	8	1027845	128,481
Bellevue	8	152347	19,043
Charlotte	14	910204	65,015
Denver	9	75665	8,407
Irvine	9	314621	34,958
Los Angeles	8	3834915	479,364
Manhattan	9	1645867	182,874
New York	30	8335897	277,863
San Diego	14	1428042	102,003
Seattle	18	771888	42,883

Figure 30: Comparison of Job Roles Per Capita in Major U.S. Cities

Salary ranges

Figure 10 shows that the range within the top 10 locations is between 95,000 thousand and 110,000 thousand a year. Which is slightly above the lowest end of the range for the Other cities however, the top end of the range is slightly below. With Other cities being closest to 120,000 thousand.

This paper examines salary trends for Business Analyst roles across various U.S. locations and industries, drawing on LinkedIn job postings and data from Kaggle. The findings provide several insights into the Business Analyst job market:

Location-Based Salary Trends:

- Regional Differences: Salary levels vary widely by location, with states like California and New York offering average salaries above $120,000. These higher figures may be due to increased demand or higher living costs in these areas.
- Demand-Driven Wages: States with a high volume of job listings, such as California and Texas, report elevated average salaries, indicating a relationship between demand and compensation. However, some areas show an oversupply of jobs with lower pay, suggesting market saturation may drive wages down in certain regions.

Figure 31: Comparison of Salary Ranges for Top 10 Locations and Other Cities

Reflection

The analysis of the job skills shows a very high association with Data analysis and SQL for the technical skills. It is necessary to know SQL, Data Analysis and Data Visualization to work in the Business Analysis field for any speciality. LinkedIn article (Shubham A. 2023) says SQL needs to extract and manipulate large datasets to understand large datasets and analyse them. Further, SQL always helps to aggregate and summarize data to easily understand patterns and trends. Another point is sometimes data needs to be merged and joined from different tables or sources. SQL is the best way to get joined products from different sources.

Data visualization is another main skill for business analysts. Without knowing the data visualisation, it is not easy to communicate information with the clients. The visualization helps to identify and isolate the factors related to the client. Further, customers can see the trends or patterns in their data to identify where went wrong or success (Hashemi-Pour, Brush, Burns 2024).

As data analysis and data visualization show high demand and association in skills, MS Excel and Tableau are also in slightly high demand in technical skills. In the Medium article, Sapra (2023) express that it is vital to know MS Excel and Tableau applications to data clean, transform, analyse and visualize.

The Knime association rules learner does not give a good result in finding highly associated skills for each speciality. The initial analysis shows there is a significant difference in means of skills in speciality. Therefore, this analysis can further drill down to more categories to find missing connections. It would need some decomposition to databases, languages and visualization tools for the technical skills.

The job market is evolving rapidly, with clear trends towards big data, AI, and flexible work arrangements. Businesses must adapt by offering competitive salaries, investing in talent with the right technical skills, and providing growth opportunities. Education, experience, and geographic factors all play a crucial role in shaping compensation and job opportunities, and companies that understand and act on these trends are better positioned to succeed in the competitive landscape.

Further, This analysis outlines industry-specific skill needs for Business Analyst roles, showing both common and unique demands across sectors. Aligning training and hiring with these insights can help meet rising expectations for data literacy and collaboration, ensuring Business Analysts are prepared to succeed across industries.

DATA QUALITY ASSESSMENT

1. Completeness

Completeness refers to the extent to which all required data is present in the dataset. Our analysis revealed that approximately 4% of data is missing from the Job_summary and Job_skills columns. This missing data could potentially impact our ability to accurately identify and analyze required skills for Business Analyst positions. Missing records are spread evenly, affecting about 4% of the dataset.

- Impact: Missing data can distort insights, particularly if the missing data is concentrated in critical roles or regions.

- Solution: We employed multiple imputation techniques using information from other variables to estimate missing values, helping maintain data integrity while minimizing bias in our analysis.

2. Accuracy

Accuracy measures how closely the data represents real-world constructs. Our findings include:

- Source: The data was collected from LinkedIn, a reputable platform for professional networking and job postings, lending credibility to our dataset.

- Challenges: We identified potential errors due to user-input inconsistencies, such as inconsistent city/state abbreviations (e.g., "New York NY" vs. "NYC NY") and typos in job titles.

Job_location has some city names that didn't match with the simplemap data. The job_location data has some location for small town and different countries.

Description	Number of Records
Non-USA jobs	19
Invalid Cities	760
Valid cities	4841

valid data = 4841

Accuracy score = (4841/ 5621) * 100 = 86%

Impact: Minimum

- Impact: Such errors could affect regional demand analysis or skill classification, leading to misleading insights.

- Solution: We implemented a series of data validation rules and cross-referencing techniques, including spot checks of random job postings, to flag and correct inaccuracies.

3. Consistency

Consistency ensures that data remains uniform across the dataset. Our analysis shows that:

- The data columns maintain a consistent format, with no significant formatting errors detected.
- Certain fields, such as job skills and locations, were harmonized to ensure consistency in naming conventions, improving accuracy in region-specific analyses.
- Impact: Inconsistent data (e.g., job levels such as "Senior" vs. "Mid-Senior") can lead to poor segmentation and difficulty in forecasting job demand by level or type.
- Solution: The job skills can be searched and replaced with the proper name using lookup tables. The lookup table has different forms of skill names to search and match proper names for the skill. For the location, use the simplemaps website to download the proper USA city name and state details. The simplemaps data is matched the LinkedIn dataset and get the standard state name and city names.

4. Validity

The search_city column has the city name without the state where the city belongs. When the city match with the simplemap data, the search_city gives multiple states for the same city. Therefore, the search_city data doesn't have validity for the data analysis.

- Impact: High
- Solution: Remove the seach_city column from the dataset.

5. Uniqueness

Uniqueness refers to the absence of duplicate records within the dataset. Our analysis identified only 4 duplicate records, accounting for less than 1% of the dataset, indicating a high level of uniqueness.

- Impact: Duplicate records can artificially inflate the demand for certain job roles or locations, distorting the final analysis.
- Solution: We applied a deduplication process using exact and fuzzy matching techniques, ensuring that each job posting was only represented once in the final dataset.

6. Timeliness

Timeliness measures how current and relevant the data is. We ensured that:

- All job postings were current and relevant at the time of analysis, with the oldest timestamp being from January 2024.
- A timestamp field was added to each record, allowing us to track the data's freshness and perform time-based trend analyses.

- Impact: Outdated data could skew insights and provide an inaccurate reflection of the current job market.
- Solution: By ensuring that all records are up-to-date, we can confidently analyse recent trends and patterns in the job market. Data Analysis Plan of Industry-Specific Needs in Business Analysts

REFERENCES

Asaniczka. (2024). 1.3M LinkedIn jobs and skills 2024 [Data set]. Kaggle. https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Brainstation. (2024). The U.S. data analysis skills gap. https://www.brainstation.io

Bouloudnine, S. (2024, June 1). Complete LinkedIn industry list 2024. lobstr.io. Retrieved November 3, 2024, from https://www.lobstr.io/blog/linkedin-industry-list

Clark, S., & Murphy, D. (2019). Interoperability challenges in healthcare: The role of Business Analysts. Journal of Health Informatics, 24(2), 88-102.

Datacamp. (2023). Business Analyst core tasks and responsibilities. https://www.datacamp.com

Dice. (2024). Business Analyst job posting. LinkedIn. https://www.linkedin.com/jobs/view/business-analyst-at-dice-3805943874

Excel Macros for Dummies (3rd ed.). (2022). John Wiley & Sons.

Hashemi-Pour, C., Brush, K., Burns, E., (2024). https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization#:~:text=Why%20is%20data%20visualization%20important,to%20be%20understandable%20by%20anyone.

Harris, M. (2020). The future of business analysis in the age of AI. Tech Trends Journal, 15(3), 102-120.

Institute of Business Analysis (IIBA). (2021). The in-demand skills for Business Analysts. https://www.iiba.org

Jones, L., & Patel, R. (2020). The evolving role of Business Analysts in healthcare. Healthcare Management Review, 39(3), 147-159.

LinkedIn Talent Solutions. (2023). How LinkedIn Talent Solutions works. https://business.linkedin.com/talent-solutions

Microsoft Excel Power Pivot & Power Query for Dummies (2nd ed.). (2022). John Wiley & Sons.

Miller, D., & Zhang, Y. (2020). The role of Business Analysts in optimizing manufacturing processes. Manufacturing Technology Review, 15(4), 122-137.

Sapra, B. (2023, May 1). https://medium.com/analysts-corner/do-you-need-to-excel-in-data-analytics-before-moving-into-a-ba-role-981b64902991

Smith, A. (2022). E-commerce and the evolving role of Business Analysts in retail. Journal of Digital Retail, 18(3), 45-59.

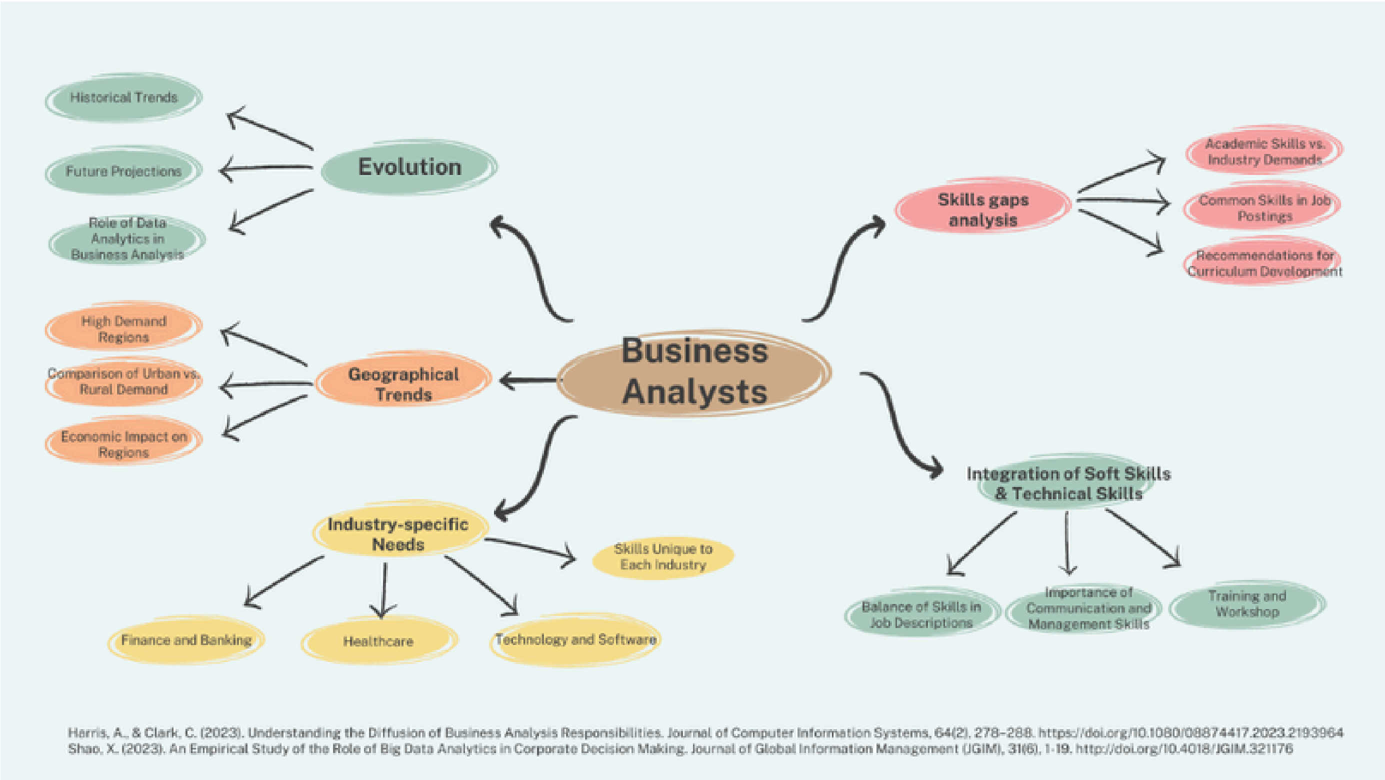
Smith, J. (2020). The role of Business Analysts in financial modeling. Journal of Financial Analysis, 45(2), 123-136.

Shubham, A. (2023, March 01). https://www.linkedin.com/pulse/sql-queries-know-business-analyst-shubham-a/

Williams, S., & Zhang, H. (2021). Bridging the gap between business and technology: The role of IT Business Analysts. Tech Insights Journal, 34(1), 54-68.

APPENDICES

MIND MAP



DATA DICTIONARY

Job_link
Definition: The URL link of the data downloaded from LinkedIn. The link contains the title and unique job number.
Format: String
Comment: This column is the primary key to all three data files.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

JobNumber_Posted
Definition:The Unique number from the job_link extracted as a separate value.
Format: String
Comment:Data must be unique and derived from the job_link.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_title
Definition:The title of the job listed in the LinkedIn.
Format: String
Comment:The job_title is used to filter business analyse jobs from the title.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_skills
Definition: The list of skills related to the job_title.
Format: String
Comment:Skills are listed as comma-separate
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Is_Soft_Skill
Definition:The skill defined as a soft skill or not.
Format: Boolean
Comment:The skills have been identified as soft skills with predefined values such as communication, problem-solving and analytical skills. This column is a flagged to separate soft skills from other skills
Code: 1 – True (Soft skill)
0- – False (non-soft skill)
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

SoftSkills
Definition:The soft skills category.
Format: String
Comment:The soft skills are derived from the Job_skills column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

ComputerLanguges
Definition:The technical skills identified as a computer language such as SQL, Python, R. Non-computer language skills showed as null
Format: String
Comment:The skills are derived from Job_skills column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Databases
Definition:The technical skills identified as databases such as Oracle, MS SQL server and SnowFlake.
Format: String
Comment:The skills are derived from Job_skills column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

AnalyseAndVisualTools
Definition:The technical skills identified as AnalyseAndVisualTool such as Tableau, Power BI or MS Excel.
Format: String
Comment:The skills are derived from Job_skills column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

OtherTechnicalSkills
Definition:Technical skills cannot be categorised as languages, databases or visualisation tools.
Format: String
Comment:The skills are derived from Job_skills column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

AllTechnicalSkills
Definition:languages, databases, visualisation tools and OtherTechnicalSkills are combined.
Format: String
Comment: The skills are derived from ComputerLanguges, Databases, AnalyseAndVisualTools, and OtherTechnicalSkills columns after joining.

Job_Speciality
Definition:The posted job relevant speciality such as finance, cyber security or system analysis.
Format: String
Comment: Manually categorised each job according to the job title and search area.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024
Job_summary
Definition:The summary of the job listed in the LinkedIn.
Format: String
Comment:The experience has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Experience_years
Definition:How many years of experience need for the job.
Format: Number
Comment:The experience has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

YearlySalary
Definition:The yearly salary for the listed job.
Format: currency
Comment:The salary has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type
Definition:Type of the job such as remote or on-site
Format: String
Comment:The job type as remote, one-site or hybrid
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type_summary
Definition:Type of the job such as remote or on-site
Format: String
Comment:The job type has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type_corrected
Definition:corrected job type to use for the analyse
Format: String
Comment:comparing the job_type_summary and job_type column derived the correct job type.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_location
Format: String
Definition: The location of the job, including city and country.
Example: “New York, USA”, “Melbourne, Australia”
Source: Asaniczka (2024).

Search_city
Format: String
Definition: The city where the job listing was found or specified during the search query.
Example: “San Francisco”
Source: Asaniczka (2024).

Search_country
Format: String
Definition: The country where the job listing was found or specified during the search query.
Example: “USA”, “Canada”
Source: Asaniczka (2024).

The Association Rules for Job skills and Specialty

	Rule Support	Rule Confident%	Recommended Technical Skills	Recommended Job Speciality
Data Analysis, Python	231	88.5	SQL	Business & Finance Analyst
Data Analysis	231	88.5	SQL	Business Agile/Scrum Analyst
Data Analysis, Python	231	88.5	SQL	Business Analyst
Python, Data Analysis	231	88.5	SQL	Business Analyst Professor
Data Analysis	231	88.5	SQL	Business Application Analyst
Python, Data Analysis	231	88.5	SQL	Business IT Analyst
Data Analysis, Python	231	88.5	SQL	Business Intelligence Analyst
Data Analysis	231	88.5	SQL	Business Management Analyst
Data Analysis, Python	231	88.5	SQL	Business Operations Analyst
Python, Data Analysis	231	88.5	SQL	Business Process Analyst
Data Analysis, Python	231	88.5	SQL	Business Support Analyst
Data Analysis, Python	231	88.5	SQL	Business Systems Analyst
Python, Data Analysis	231	88.5	SQL	Senior Business Operations Analyst
Data Analysis, Python	231	88.5	SQL	Senior Business Process Analyst
Data Analysis, Python	231	88.5	SQL	Senior Business Systems Analyst
Python	328	85.9	SQL	Business & Data Analyst
Python	328	85.9	SQL	Business & Finance Analyst

Python	328	85.9	SQL	Business Analyst
Python	328	85.9	SQL	Business Analyst Professor
Python	328	85.9	SQL	Business IT Analyst
Python	328	85.9	SQL	Business Intelligence Analyst
Python	328	85.9	SQL	Business Operations Analyst
Python	328	85.9	SQL	Business Process Analyst
Tableau, MS Excel	148	55.8	Data Visualisation	Senior Business Intelligence Analyst
MS Excel	148	55.8	Data Visualisation	Senior Business Operations Analyst
MS Excel, Tableau	148	55.8	Data Visualisation	Senior Business Process Analyst
MS Excel, Tableau	148	55.8	Data Visualisation	Senior Business Systems Analyst
Data Analysis, Python	144	55.2	Tableau	Business & Data Analyst
Data Analysis, Python	144	55.2	Tableau	Business & Finance Analyst
Data Analysis	144	55.2	Tableau	Business Agile/Scrum Analyst
Data Analysis, Python	144	55.2	Tableau	Business Analyst
Python, Data Analysis	144	55.2	Tableau	Business Analyst Professor
Data Analysis	144	55.2	Tableau	Business Application Analyst
Python, Data Analysis	144	55.2	Tableau	Business IT Analyst
Data Analysis, Python	144	55.2	Tableau	Business Intelligence Analyst
Data Analysis	144	55.2	Tableau	Business Management Analyst
Data Analysis, Python	144	55.2	Tableau	Business Operations Analyst

Table: 22 Association rules table sample

APPENDICES

INDIVIDUAL CONTRIBUTIONS

Eranga (nakey002): Eranga worked on the data transformation process. The downloaded data from Kaggle was uploaded to the Knime workflow and manipulated to obtain clean, structured data. Separate CSV outputs were created for technical skills, soft skills, geographical job lists, and industrial data. Eranga downloaded SimpleMap data to quality-check job posts' locations and created a Knime workflow to join SimpleMap location data with LinkedIn posts. Furthermore, he created a Power BI dashboard from the transformed data to generate visualizations for the group. Eranga checked data quality and analyzed the data structure for team use. In the report, Eranga contributed to writing project objectives, scope, data understanding, data dictionary, and preliminary data analysis. Eranga's main focus was on business analyst skills with a job specialty. SAS code was created to check the normality of data and statistical significance of job specialties.

Ushan (piyum001): In the Geographical Trends section, I presented a high degree of analysis for the trends that govern USA-based data on job distribution, salary trends, skill demands, and regional differences across major American cities. Key findings showed unusual growth in cities like Montpelier and Columbus, which are emerging as new hubs against traditionally known cities like New York and San Francisco. It showed a balanced requirement between soft skills, which were communication and problem-solving, and technical skills comprising SQL and data modeling. Cities like Santa Clara and Seattle required more technically oriented people, whereas for Atlanta, the requirement was more biased toward soft skills. The salary analysis has shown highly paid locations, such as Santa Clara and Summit, which have been driven by industries such as technology and finance. This, in turn, has indicated strategic locations for both job seekers and employers. This will be done via visualizations: choropleth maps for hotspots of jobs, bar and treemap charts for illustrating demands for skills, box plots for showing variance in the count of jobs, and Sankey diagrams for flow across job levels, skills, and locations. The holistic approach gave quite clear insights into the regional trends, thus providing some data-driven recommendations to job seekers, employers, and educational institutions on how to align their strategies to match the emerging patterns in the business analytics job market.

Chau (nguhy158): In this assignment, I contributed to writing on several important areas as follows. I initiated and coordinated the mind map work and structured the major elements of the project to create a framework that we could work from. I also supported the team in constructing the data dictionary and checking the quality of data that we included in our dataset. Furthermore, as a subject matter expert, I focused on the skill demand by geo-location and was able to capture market-specific trends of soft and technical skills for business analyst professionals. Salary analysis was also included; I considered how salary distribution changes with location and industry, providing an understanding of the spatial patterns of pay and the particularisms of the field. Altogether, these contributions supported the comprehensive foundation for our assignments that correlated with the objectives of the project findings.

Shy (Garsy036): I contributed to this report by manually inputting and classifying job listings into broader industry labels and specific industries, ensuring accurate categorization for analysis. Additionally, I classified different types of Business Analysts based on job descriptions, providing a detailed breakdown of role-specific insights. These efforts, combined with the analysis of technical and soft skill requirements, allowed the group to identify shared and unique demands across varying industries.

Martin (Valma004): I contributed to the Skills Gap Analysis, which involved examining the balance between soft and technical skill requirements across various roles. Additionally, I analyzed formal academic qualifications in comparison to industry-specific certifications, identifying gaps between what employers seek and the qualifications candidates typically hold. This work helps provide a clearer understanding of the alignment between job market demands and available talent.

Trang (Nguty440): I actively participated in group activities and excellently completed assigned tasks. Specifically, I was responsible for manually calculating the annual salary for each job position based on the description. I also contributed to researching and writing the "Industry-Specific Analysis" section with my teammate Shy. Furthermore, I conducted the analysis for "The Salary Effect on Business Analysis Skills" section in the group project. Additionally, I created the "Job Technical Skills and Salary Association" data table using Knime Workflow. Notably, I took on a key role in finalizing the end product. My work included optimizing the presentation of tables and images while applying advanced presentation techniques to enhance visual quality and readability. Through continuous efforts and dedication, I made significant contributions to creating a high-quality final product that fully demonstrates the capabilities and professionalism of the entire group.

APPENDICES

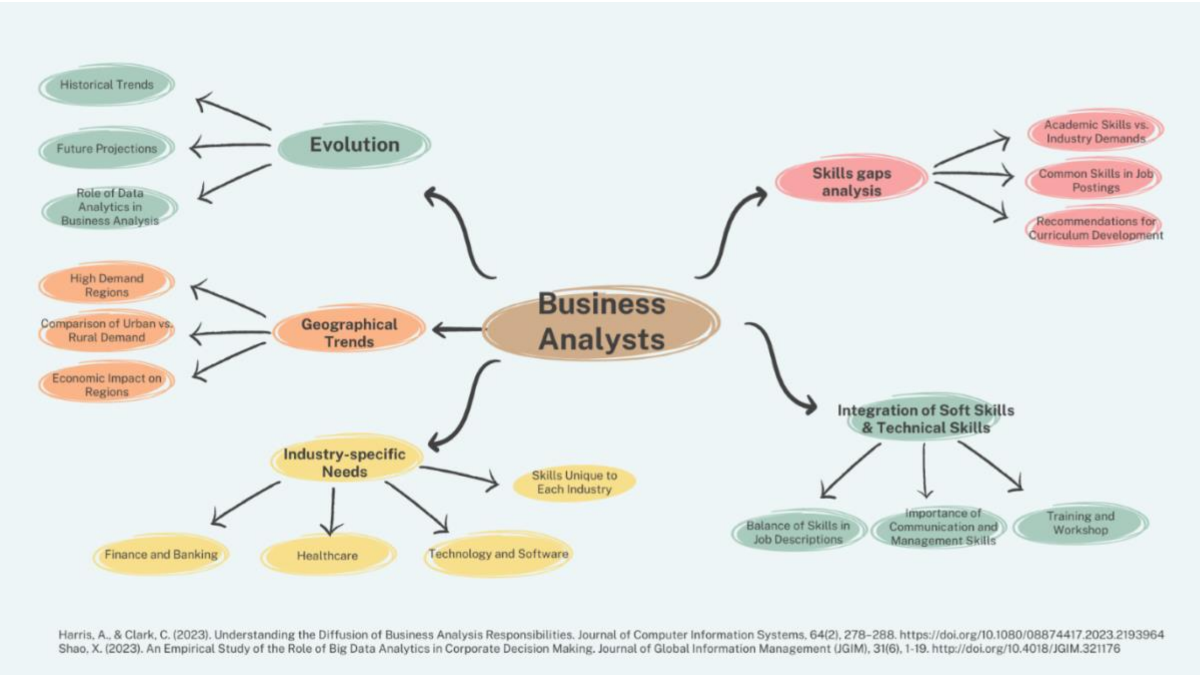
LIST OF FIGURE

LIST OF TABLE

Individual Part

Hoang Chau Nguyen – Nguhy158

The main focus of this project is to compare and explore the current salary offer in Business Analyst Jobs within different business sectors and in different regions in USA. LinkedIn job postings are used as the source of the data with textual and numerical features involving job title, industry, salary, and locations. Due to the fact that provided dataset was sets of text values which makes possible only few analyses, the new additional dataset considering Business Analyst's salaries divided into ranges was included in the case and added to the analysis as well.



The purpose of this research will therefore be to see how much Business Analysts are paid according to the industry and the location, to get a clear picture of trends and flows in the job market.

The new dataset which we grabbed from Kaggle includes more formalized information about Business Analyst positions, such as estimated wages, job titles, company names, positions' locations, industries, and company rating. Here are the key attributes:

Job Title: In other words, it can bear the name of the job position as follows: Business Analyst, Senior Business Analyst, and so on.

Salary Estimate: Pay scale (in form of a range such as $50,000 – $100,000).

Company Name: The name of the company that is offering the job.

Location: The city and the state in which the job is located.

Industry: Industry of the offered job, for instance Information technology, finance, or health services.

Data Dictionary

Column Name	Description
Job Title	Title of the business analyst job (e.g., Senior Business Analyst).
Salary Estimate	Salary range for the job, in formats like "$50K-$100K".
Company Name	Name of the company offering the job.
Location	The geographical location of the job.
Industry	Industry sector of the job posting.
Min Salary	The minimum salary extracted from the Salary Estimate column.
Max Salary	The maximum salary extracted from the Salary Estimate column.
Average Salary	The computed average salary between Min Salary and Max Salary.

Data Cleaning

Extract Salary Information:

Select salary estimate and derive the minimum and maximum values of the salaries.

Some of the figures contain symbols such as $ and K, they should be converted to numeric format by stripping off the symbols and multiplying the figure by 1000.

Example:

Salary Estimate: $50K-$100K

Min Salary: 50,000

Max Salary: 100,000

Average Salary: (50,000 + 100,000) / 2 = 75,000

Remove Unnecessary Columns:

Exclude columns that may not affect salary such as 'Job Description'.

Handle Missing Data:

Scan for any N/A's or NULLs with respect to salary, location and/or industry and deal with them either by deletion or by putting in estimates.

Standardize Text:

Pay particular attention to job titles and locations as well as the company names; do not have the same job name in two different locations, such as, "NYC" and "New York").

Data Quality Assessment

Completeness: Checks for missing values in fields like salaries or location, ensure key fields are completed

Accuracy: Salary data are validated and converted to numeric values

Consistency: Job title, job location and industry are correctly spelt and also match across the data set.

Uniqueness: Same entries for a job are cleared to avoid mislead the study.

Timeliness: The data was collected to match current jobs.

Data Modeling

We remain more concerned with how to model the data in a way that will help to analyze the salary based on location and industry. The quantitative attributes which must be used in drawing analysis include Min-Salary, Max-Salary, and Average-Salary.

New Columns:

Min Salary: From Salary Estimate extract the minimum amount that you will accept in the company.

Max Salary: Extract the maximum salary.

Average Salary: Compute the average salary.

Categorization:

By Location: The defined should then filter their results based on any geographic area (e.g., New York, California) so as to identify geographical trends.

By Industry: There are Corporate IT, Corporate Finance, and Corporate Healthcare industries etc; sort business analyst salaries according to the industry in order to compare which industries pay Business Analysts more or less.

Analysis

Skills demand analysis:

Soft skills by Location:

TOP 10 Softskills for Business Analyst

	Soft Skills / Search Country									
	Analytical Skill	Communiatio n Skill	Critical Thinking Skill	Interpersonal Skill	Leadership Skill	Presentation Skill	Problem-Solving Skill	Research Skill	Teamwork Skill	Writing Skill
Search City	United States	United States	United States	United States	United States	United States	United States	United States	United States	United States
Atlanta	19	34	4	4	14	6	26	7	17	6
Eastchester	12	27	4	4	3	14	16	6	13	9
Ferguson	12	30	3	3	5	2	21	4	15	8
Garland	22	35	6	8	1	8	19	8	20	11
Montpelier	17	59	4	5	9	12	24	5	33	10
San Diego	17	30	2	4	6	1	17	1	9	5
Santa Clara	11	28	4	1	7	9	12	5	16	4

Figure 15 – Top 10 soft skill Highest Demand by Area

The following figure presents a highlight table with the distribution of the top 10 soft skills reported by employers when recruiting Business Analysts, based on cities in the United States. The reviewed skills are Analytical Skill, Communication Skill, Critical Thinking Skill, Interpersonal Skill, Leadership Skill, Presentation Skill, Problem-Solving Skill, Research & Development Skill, Teamwork Skill, and Writing Skill. The darkness of color in each cell represents the demand level for that particular skill in a particular city. For example, Montpelier reportedly requires Communication Skills most of all while Atlanta needs Problem-Solving Skills and Leadership Skills most of all. It is also evident that skills are distributed unevenly across the regions, and this distribution may be informative for soft skill enthusiasts, job seekers, and managers when used to identify trends by region.

Technical Skills by Location

TOP10 Technical Skills for Business Analyst

	Search Country / All Technical Skills United States									
Search City	Data Mining	Data Modeling	Data Visualisation	MS Azure	MS Excel	Python	R	SAP	SQL	Tableau
Atlanta	3	4	9		1	3	1	1	10	5
Austin	3	8	8	2	3	12	1	4	18	8
Baytown	2	5	9	2	2	4	2	12	12	5
Beverly	1	13	7		2	2	1		15	8
Columbus	1	10	9	6		3		1	13	3
East Lansing		7	1	13	1				5	
Eastchester	1	9	1	3	3	2		1	8	3
Garland	4	21	6	1	9	2	1	5	20	4
Montpelier		7	5	3	5	1		3	9	1
Santa Clara	4	4	2	1	2	2	1	1	11	5

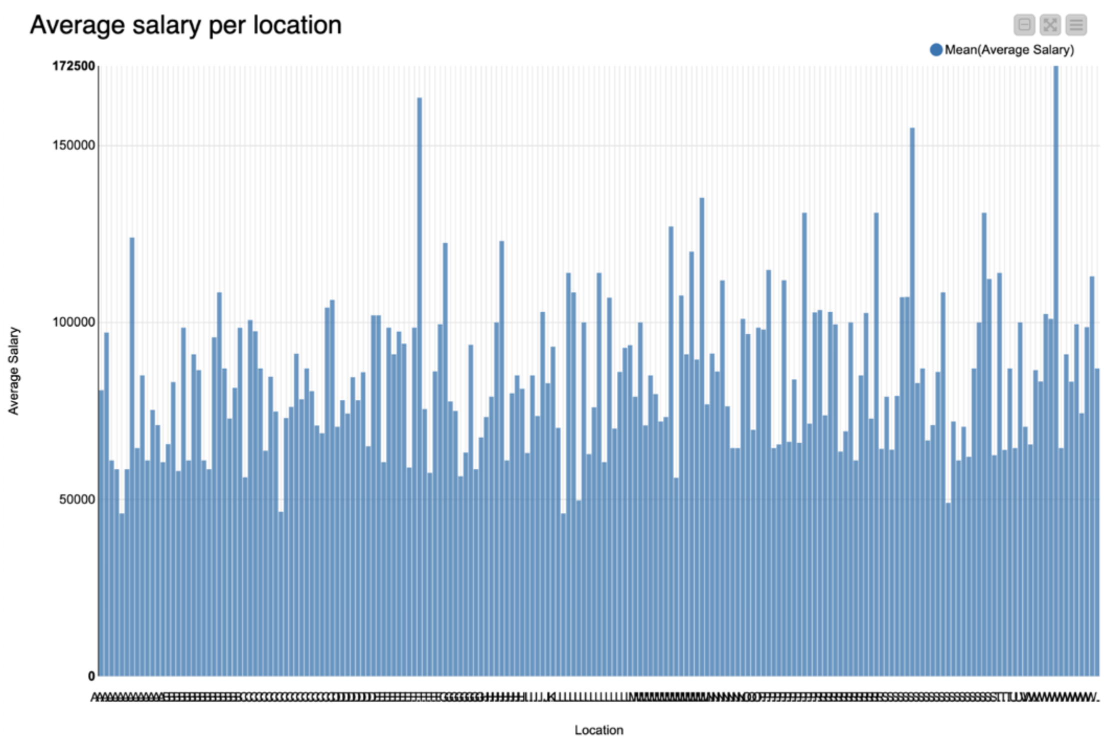
Figure 16: Top 10 Technical Skills Highest Demand by Area

Here we provide a highlight table indicating the most promising technical skills to acquire Business Analyst jobs in different cities of the USA. The analysed skills include Data Mining, Data Modelling, Data Visualization, MS Azure, MS Excel, Python, R, SAP, SQL and Tableau. Same as in the previous table, the darkness of colour in relation to frequency; the demand value scale is provided under the table. For instance, Garland and Austin have always been most in need of SQL experts while East Lansing needs experts in MS Azure. This graphic will be beneficial to help sort out which technical skills are essential in one area over another to help those in their professions and recruiting services understand what specialized skill sets are needed in each region.

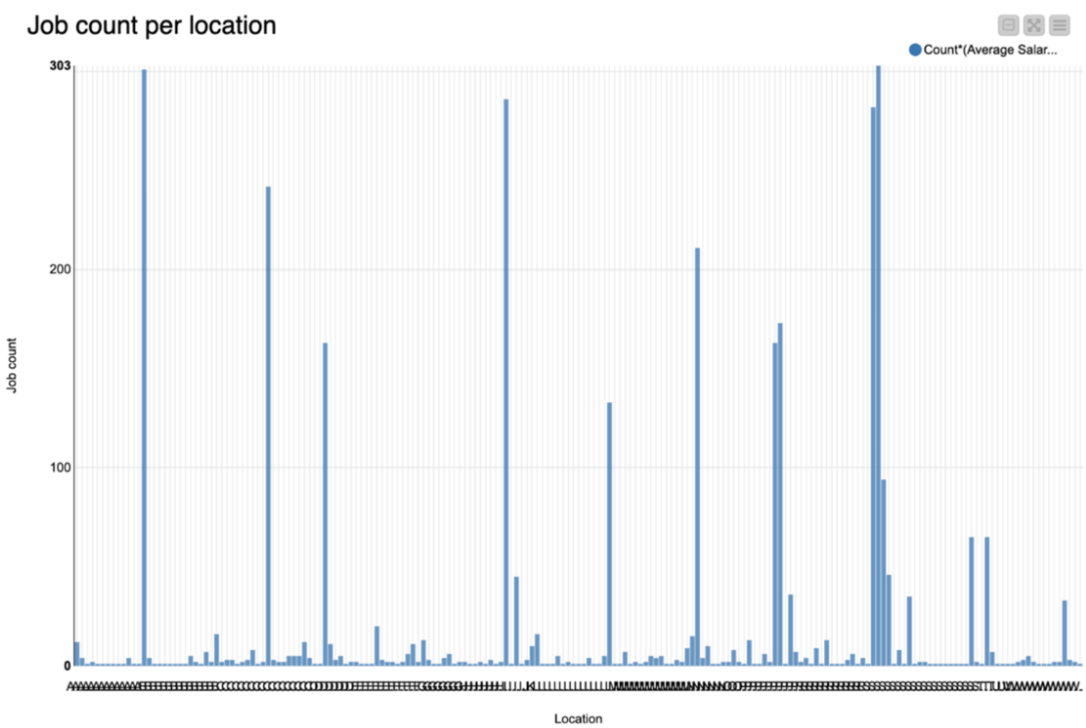
Salary analysis:

Salary by Location:

We analyze how salaries for Business Analyst roles vary by city or state. This helps us understand which geographic areas are offering the highest pay for these roles.



From the chart, it is evident that the average salary differs greatly depending on the location. Some regions have average salaries which are much higher (e.g., California, Pennsylvania), and this may mean high demand, stiff competition in a particular area or regions with a high cost of living.

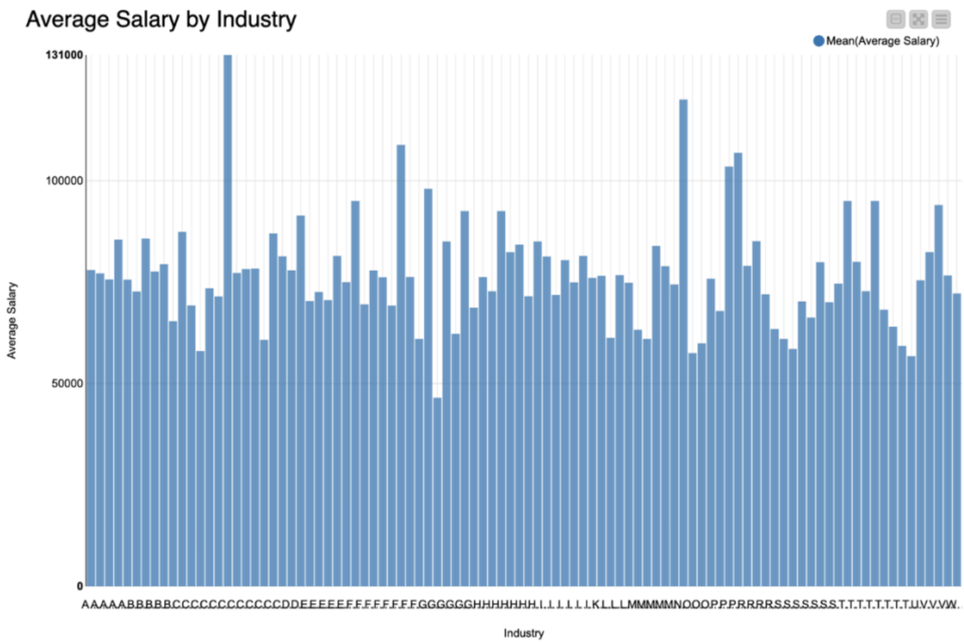


The job count chart shows where most Business Analyst jobs are available. Certain cities or states seem to have far more job opportunities such as California and Texas.

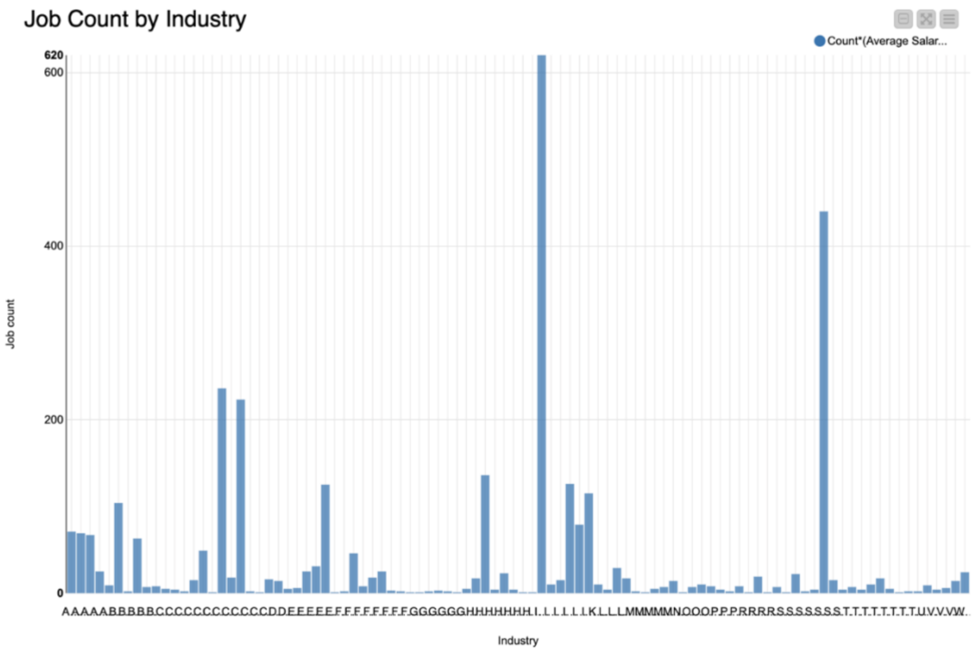
States with high job counts but lower average salaries might indicate an oversupply of jobs, while states with both high salaries and job counts could indicate a candidate's market.

Salary by Industry:

Next, we explore how salaries for Business Analysts vary across different industries (e.g., Finance, Healthcare, IT). This can help determine which industries are the most lucrative for Business Analyst positions.



Here we can see that industry in commercial equipment & repairs, news outlet, and food & beverage are the top industries with the highest salary for Business Analyst positions.

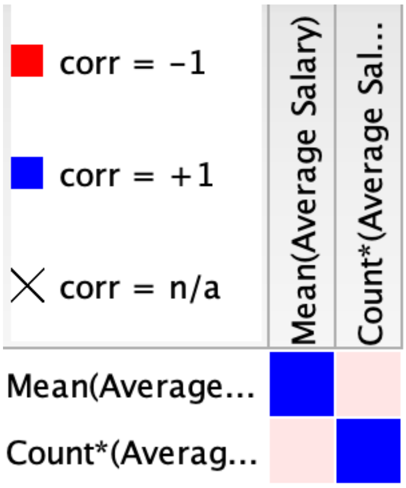


Industries have the highest number of job postings, indicating high demand for Business Analysts which are IT Services, Outsourcing and Computer Hardware and Software.

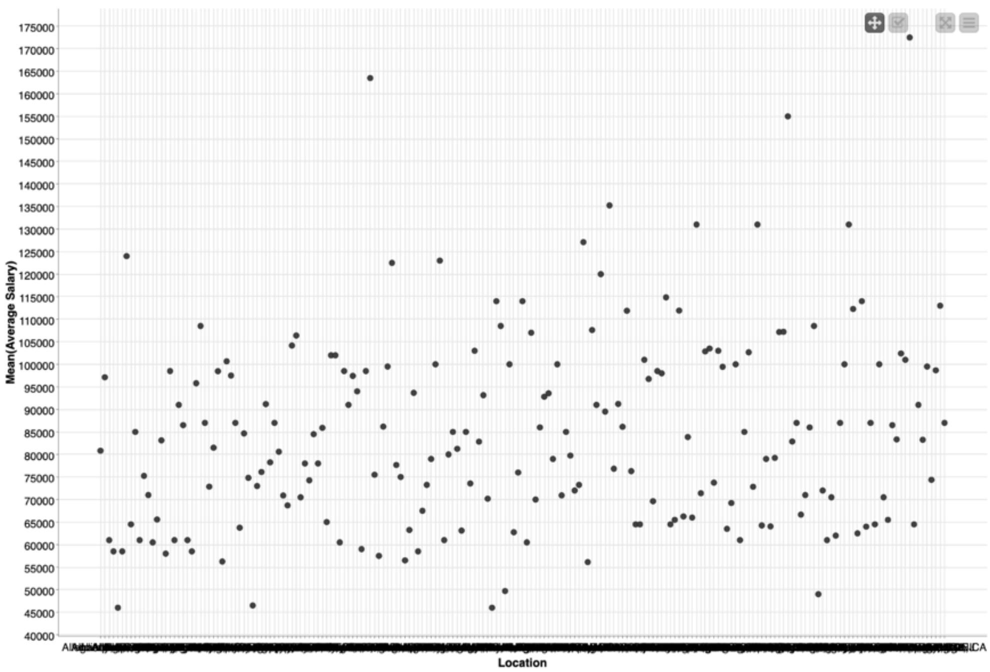
Some industries may have high demand but offer lower salaries, while others might have fewer job postings but offer more competitive salaries.

Correlation Analysis:

Correlation by Location



The blue squares indicate a positive correlation between job count and average salary. This suggests that locations with higher job counts also tend to offer higher salaries for Business Analysts. A strong positive correlation implies that there's a high demand for Business Analysts in these areas, which is likely driving up the salaries. There are no red squares, meaning there is no strong negative correlation between job count and salary. This indicates that in the locations analyzed, an increase in job count does not result in a significant decrease in salary.



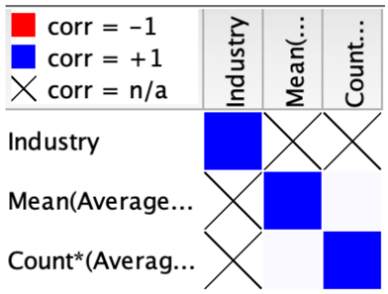
Scatter Plot (for Location vs Salary)

The scatter plot shows that average salaries for Business Analysts are distributed within a broad range (from around 40,000 to over 170,000).

While some locations offer relatively high salaries (above 120,000), most locations seem to cluster between 50,000 and 100,000.

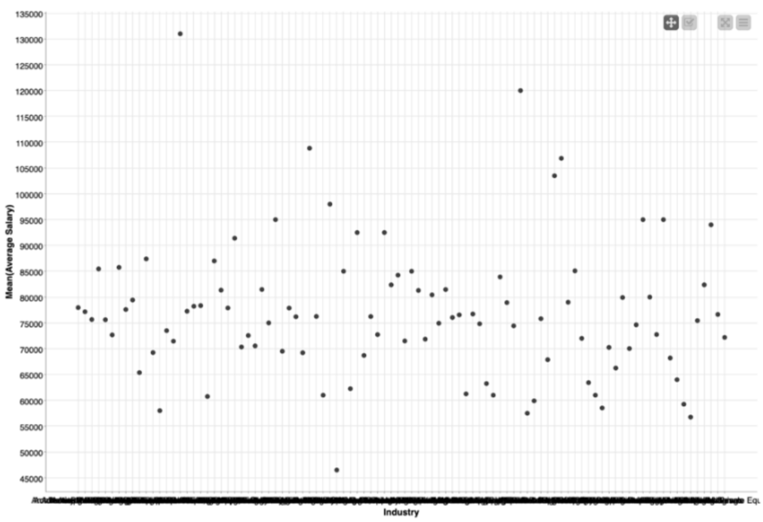
A few locations are offering significantly higher salaries, as indicated by the points above 150,000. These might represent cities with high living costs (e.g., New York, San Francisco) or industries requiring highly specialized Business Analysts. Based on this scatter plot, there doesn't appear to be a clear upward or downward trend when moving from one location to another. The salaries are dispersed across a variety of locations, showing that no single region drastically dominates in terms of salary.

Correlation by Industry:



Similar to the location-based analysis, we see a positive correlation between job count and average salary for industries.

This means industries with more job postings also tend to offer higher salaries for Business Analysts. For example, industries like finance or IT (which tend to have higher demand) might be driving up salaries due to the competitive market for skilled analysts. Again, there are no strong negative correlations, suggesting that an increase in the number of job postings does not decrease the average salary in any industry.



Scatter Plot (for Industry vs Salary)

Most industries appear to have average salaries ranging between 60,000 and 90,000. There are a few industries offering salaries above 100,000, but they are more spread out.

There are some industries offering higher salaries (around 120,000 and 130,000), but these appear to be outliers rather than a general trend. These might be highly specialized industries, or industries in high-demand fields like technology or finance.

Based on this plot, there doesn't seem to be a strong, clear trend showing that certain industries consistently offer higher or lower salaries. The salary variation appears to be relatively even across different industries.

The industries with fewer job postings and higher salaries may be those requiring more specialized knowledge or skills. On the other hand, industries with a larger number of job postings seem to offer more moderate salaries.

Conclusion

This paper provides an extended discussion of the salary dynamics in various locations and industries related to Business Analyst positions in the United States based on the job postings on LinkedIn provided with the data found on Kaggle. The research reveals several key insights about the Business Analyst job market:

Salary by Location:

Regional salary variations: The data show that salaries vary significantly depending on the location of the organization. The highest general average pay comes from states like California and New York where the general average pay is above 120000 dollars, which might be due to high demand or high costs of living.

Demand-Driven Salaries: States with high density job listings such as; California and Texas also presented higher average wages implying that the demand for a certain type of job is directly proportional to how much the employers are willing to pay for the job. However, some of the places feature high numbers of jobs with slashed salaries, suggesting that there is excess demand for jobs in these areas.

Salary by Industry:

Top-Paying Industries: Industry professionals from Commercial Equipment & Repairs, News Outlets, and Food & Beverage pay Business Analysts some of the highest average salaries. These must probably require some form of expertise in certain fields related to specific industries.

Demand Across Industries: Job advertisement for Business Analysts is high in IT and Outsourcing industries mainly due to the advancement of these industries and technological-based services.

Salary Variability: Some industries have average salaries ranging between $60,000 and $90,000; however, some industries, there are extreme cases, whose salaries are above $120,000. These cases are particularly possible to be unique positions in such areas of specialization as finance or technology.

Correlation Analysis:

Positive Correlation Between Job Count and Salary: The results of both the location based and the industry based studies reveal that the ‘number of jobs posted in the field is positively related to salary’. Certain areas that post larger quantities of jobs tend to pay better and this can be attributed to the fact that there is stiff competition for human resource in those sectors.

No Negative Correlation: Thus, there is no solid evidence of a negative relationship at the level of the whole economy, that is, an increase in job offers does not result in a sharp decrease in pay.

Key Findings from Scatter Plots:

Salary Distribution: The scatter plots imply that most of the Business Analyst positions salaries are within $50,000-$ $100,000. It pointed out that figures of more than $120,000 are more likely to occur in specialized sectors or places where costs are high.

Even Distribution: There are no clear trends in one certain area or industry majoring in offering the highest or the lowest salaries where the data points are more or less equally spread geographically and by industries.

In conclusion, this research can contribute insights regarding determining Business Analyst remuneration within the USA. It is even noted that high paid areas are the places that require these positions, although other factors can influence the salary, such as the degree of specialization, for instance, or cost of living for example. Such dynamic analysis should be useful to the job seekers and employers who expect to set their salary benchmarks and hiring standards in the Business Analyst job market accordingly.

Shylyn Garcia – Garsy036

CRAAP Test:

A Longitudinal Analysis of Job Skills for Entry-Level Data Analysts

Link: https://tinyurl.com/TianxiDong

Currency: 2020

Relevance: Is relevant to understanding the evolution of job skills required for data analysts and business intelligence analysts.

Authority: Journal of Information Systems Education

Accuracy: Used text mining approach and longitudinal data analysis, which are valid research methods.

Purpose: To inform educational institutions and instructors about evolving industry demands to adjust accordingly.

Investigation of Essential Skills for Data Analysts: An Analysis Based on LinkedIn

Link: https://tinyurl.com/TaowenLeETAL

Currency: 2023

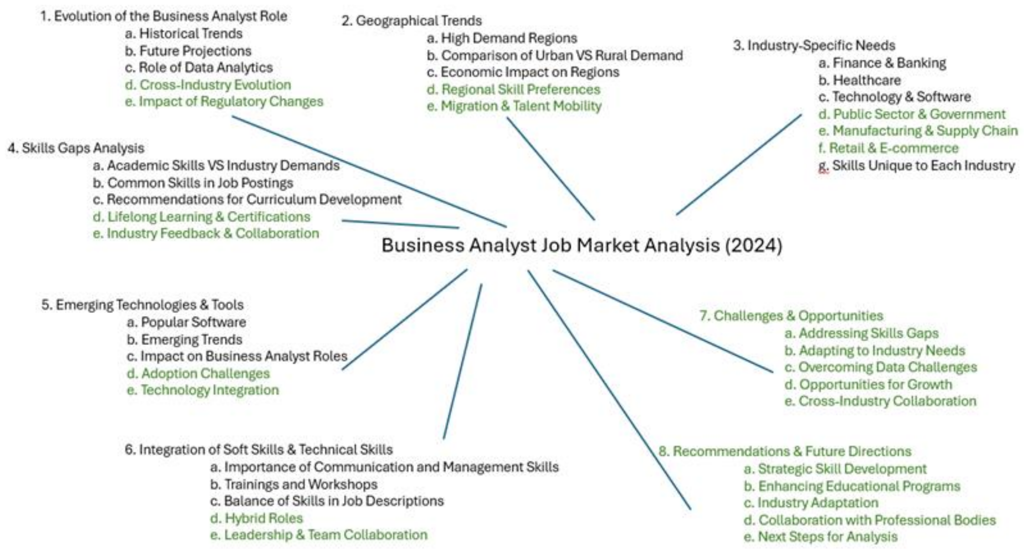
Relevance: Is relevant to understanding the skills gap between academic studies and industry expectations for data analysts.

Authority: Journal of Global Information Management (JGIM)

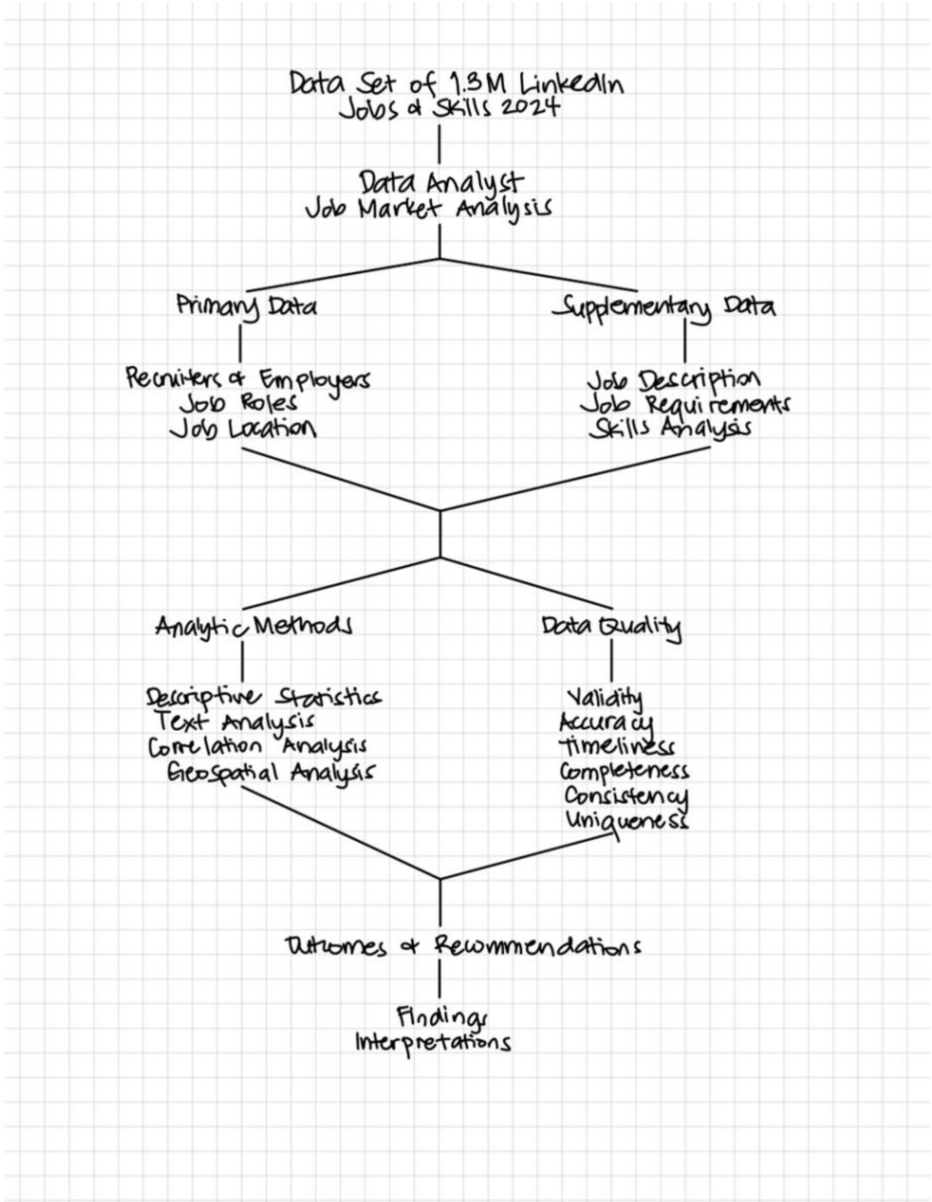
Accuracy: Used recognized data analysis methods like Distribution Analysis and Cross-Tabulation.

Purpose: Appears to be informative and highlights the gap between education and actual industry needs for Data Analysts

Mind Map: Based on Hoang Chau Nguyen's



Mind Map:



References/Bibliography:

Kumar, A. & Kumar, T. V. (2021). View Materialization Over Big Data. International Journal of Data Analytics (IJDA), 2(1), 61-85. http://doi.org/10.4018/IJDA.2021010103 García, S.,

Ramírez-Gallego, S., Luengo, J. et al. (2016). Big data preprocessing: methods and prospects. Big Data Anal 1, 9. https://doi.org/10.1186/s41044-016-0014-0

Data Dictionary

job_number
Format: number
Definition: refers to a unique identifier for the job listing, extracted from the job link.
Example: “3800443444”, “3728878892”, “3802336668”

Sources: Asaniczka, (2024).

job_title
Format: string
Definition: refers to the title of the job listing indicating the role.
Example: “Business Analyst”, “Systems Analyst”, “Senior Analyst”
Sources: Asaniczka, (2024).

company
Format: string
Definition: refers to the name of the company or entity offering the job.
Example: “Amazon”, “BioSpace”, “Cloudflare”
Sources: Asaniczka, (2024).

job_location
Format: string
Definition: refers to the location where the job is based, typically a city, state, or country.
Example: “Omaha, NE”, “Chicaho, IL”, “Melbourne, VIC”
Sources: Asaniczka, (2024).

job_level
Format: string
Definition: refers to the required level of experience for the job.
Example: “Associate”, “Mid-Senior”, “Internship”
Sources: Asaniczka, (2024).

job_type
Format: boolean
Definition: refers to the type of employment offered.
Example: “On-site” or “Off-site”
Sources: Asaniczka, (2024).

job_link
Format: string
Definition: refers to the URL to the original job listing on LinkedIn.
Example: “https://www.linkedin.com/jobs/view/business-analyst-at-dice-3805943874”
Sources: Asaniczka, (2024).

job_summary
Format: string
Definition: refers to a brief overview or summary of the job listing.
Example: Can be lengthy text; includes job responsibilities, skills & qualifications, company overview, etc.
Sources: Asaniczka, (2024).

job_skills

Format: string
Definition: refers to the skills needed for the job, such as data analysis, data management, consulting, etc.
Example: "SQL", "Excel", "Data Analysis"
Sources: Asaniczka, (2024).

first_seen
Format: date (YYYY-MM-DD)
Definition: refers to timestamps or dates that indicate when a particular job listing was first encountered.
Sources: Asaniczka, (2024).

last_processed_time
Format: date (YYYY-MM-DD HH:mm:ss.SSSSSSZZ)
Definition: refers to timestamps or dates that indicate when a particular job listing was last processed.
Sources: Asaniczka, (2024).

search_city
Format: string
Definition: refers to the city or county where the job listing was found, or the country specified during the search query.
Example: "California", "Baytown", "Pompano Beach"
Sources: Asaniczka, (2024).

search_country
Format: string
Definition: refers to the country where the job listing was found, or the country specified during the search query.
Example: "United States", "Canada", "Australia"
Sources: Asaniczka, (2024).

search_position
Format: string
Definition: refers to the specific job title or role that was targeted during the search query.
Example: "Job Analyst", "Senior Analyst", "Intelligence Analyst"
Sources: Asaniczka, (2024).

industry_id
Format: number
Definition: refers to the unique identifier assigned to each industry
Example: "1737", "42", "96"
Sources: lobstr.io, (2024).

industry_hierarchy
Format: string

Definition: refers to the structured representation of the relationship between industries, often indicating sub-industries within a broader industry
Example: "Technology, Information and Media > Media and Telecommunications > Broadcast Media Production and Distribution"
Sources: lobstr.io, (2024).

industry_label
Format: string
Definition: refers to the main category of an industry, representing a broader sector.
Example: "Professional Services", "Hospitals and Health Care", "Construction"
Sources: lobstr.io, (2024).

industry_specific
Format: string
Definition: refers to the specific industry that falls under the broader industry_label
Example: "Insurance", "Retail", "Chemical Manufacturing"
Sources: lobstr.io, (2024).

Preliminary Data Analysis

Objective:

The primary objective of this project is to analyze and provide an overview of the current job market for Business Analyst roles based on 1.3MLinkedIn Job & Skills listings. The analysis will focus on the demand for business analysts across various industries, the key skills required, geographical distribution, job levels, and emerging trends in 2024. The insights gathered will help identify opportunities, skills gaps, and industry-specific needs for business analysts.
Data Set: 1.3M LinkedIn Jobs & Skills 2024 (LINK)
Key Columns Analyzed:
job_number, job_title, company, job_location, job_link, job_summary, and job_skills
Key Findings:

- **Most Common Job Titles:** Business Analyst and Business Systems Analyst
  - **Insight:** The data shows a high demand for both general Business Analysts and specialized roles like Data Analysts. Mid-Senior positions are also frequently advertised, indicating opportunities for more experienced professionals.
- **Location Distribution:** United States, Australia, Canada
  - **Insight:** The United States dominate the job listings.
- **In-Demand Skills:** Excel, SQL, Data Analysis, PowerBI, etc.
  - **Insight:** Technical skills, especially in Excel, SQL, Data Analysis, PowerBi and others, are in high demand. Communication skills are also important, showing the need for Business Analysts to clearly share data insights.
- **Trends:** There has been a steady increase in job postings within the month of January 2024.
  - **Insight:** The increase in postings suggests growing demand for Business Analysts.

Descriptive Analysis:

- **Job Distribution:** Analyze the distribution of business analyst job listings across different industries, locations, and job levels.
- **Skills Analysis:** Identify the most required skills for business analyst positions, including both technical and soft skills.
- **Geographical Trends:** Map the geographical distribution of business analyst jobs, highlighting regions with the highest demand.
- **Industry-Specific Needs:** Determine which industries require data analysts in sectors such as Finance & Banking, Healthcare, Technology & Software, etc.

Report:

Manually input and categorize Business Analyst listings into specific industries.

Manually input and categorize different types of Business Analyst (eg. IT Business Analyst, Business Process Analyst, Security Analyst, etc.)

Data Analysis

Summary:

Business analyst roles continue to be in high demand as organizations increasingly rely on data to guide their decisions. This analysis explores 1.3 million LinkedIn job listings from 2024, looking at general trends in business analyst postings, such as industry distribution, required skills, geographical trends, and key qualifications.

Introduction:

Business analysts bridge the gap between IT and business by analyzing processes, identifying needs, and providing data-driven recommendations. As data becomes more important in decision-making, demand for business analysts is rising across industries, although requirements differ by sector.

This report analyzes industry-specific needs for business analysts based on 2024 LinkedIn job listings. It covers:

- Data Overview: A general summary of business analyst job postings across industries.
- Skills Analysis: Common and industry-specific skills required.
- Qualifications: Educational and experience requirements.
- Soft and Technical Skills: How industries balance the need for both soft and technical skills in business analysts.

Data Overview:

The dataset used for this analysis contains job postings from LinkedIn in 2024, with the following columns:

job_number
job_title
company
job_level

job_type
job_link
job_summary
job_skills
job_hiarchy
job_label
job_specific

Note: The dataset did not include an industry column, so an additional dataset was combined to align companies with their respective industries.

Skills Analysis:

Professional Services

Key Industry-Specific Skills:

Consulting Expertise: An ability to translate client needs into actionable business solutions.

Project Management: Experience with project management tools.

Client Relationship Management: Expertise in managing stakeholder expectations and delivering projects on time.

Qualifications:

Degrees or certifications in Business Administration, Finance, or any related field.

Administrative & Support Services

Key Industry-Specific Skills:

Process Automation: Familiarity with administrative and automation tools.

Analytics: Proficiency in analytics tools.

Workflow Optimization: An ability to analyze administrative processes and recommend improvements.

Qualifications:

Degrees or relevant certifications in Business Administration, Human Resources, or Management Information Systems.

Financial Services

Key Industry-Specific Skills:

Financial Modeling: Expertise in tools like Excel, SAS, and SQL for building financial models.

Risk Management: Proficiency in identifying and managing financial risks, along with knowledge of regulatory standards.

Regulatory Compliance: Strong understanding of laws and regulations affecting the financial sector.

Qualifications:

Degrees or certifications in Finance, Economics, or Business.

Education

Key Industry-Specific Skills:

Educational Technology: Familiarity with student information systems.

Data-Driven Decision Making: Experience with data analysis tools to inform decisions regarding student performance and institutional improvement.

Process Engineering: Ability to optimize administrative and academic workflows.

Qualifications:

Degrees or certifications in Education Administration, Information Systems, or Business.

Hospitals & Healthcare

Key Industry-Specific Skills:

Healthcare Data Management: Expertise in managing healthcare data, including compliance with regulations.

Process Improvement: Knowledge in healthcare operations.

Electronic Health Records (EHR): Experience with EHR systems.

Qualifications:

Degrees or certifications in Healthcare Administration, Public Health, or Information Systems.

Integration of Soft & Technical Skills:

Industries need business analysts to have both technical skills and soft skills like teamwork, communication, and leadership. They must work across teams and manage different stakeholders.

Conclusion:

Industry-specific needs for business analysts differ, with each sector requiring unique skills and qualifications. As industries evolve, business analysts will play an even more important role, especially in fast-changing fields like Professional Services, Administrative & Support Services, and Hospitals & Healthcare.

Actual Part of Report:

5.2 Industry-Specific Needs

Business analysts are in high demand across various industries, each with defined criteria for skill sets and qualifications. The main aspect and function of a business analyst is to bring together information technology and business. Over time, industry-specific demands such as system management, education, certification, and the like, have greatly influenced the skill set, tools, and requirements necessary to do the job thus, business analysts must align their competencies accordingly.

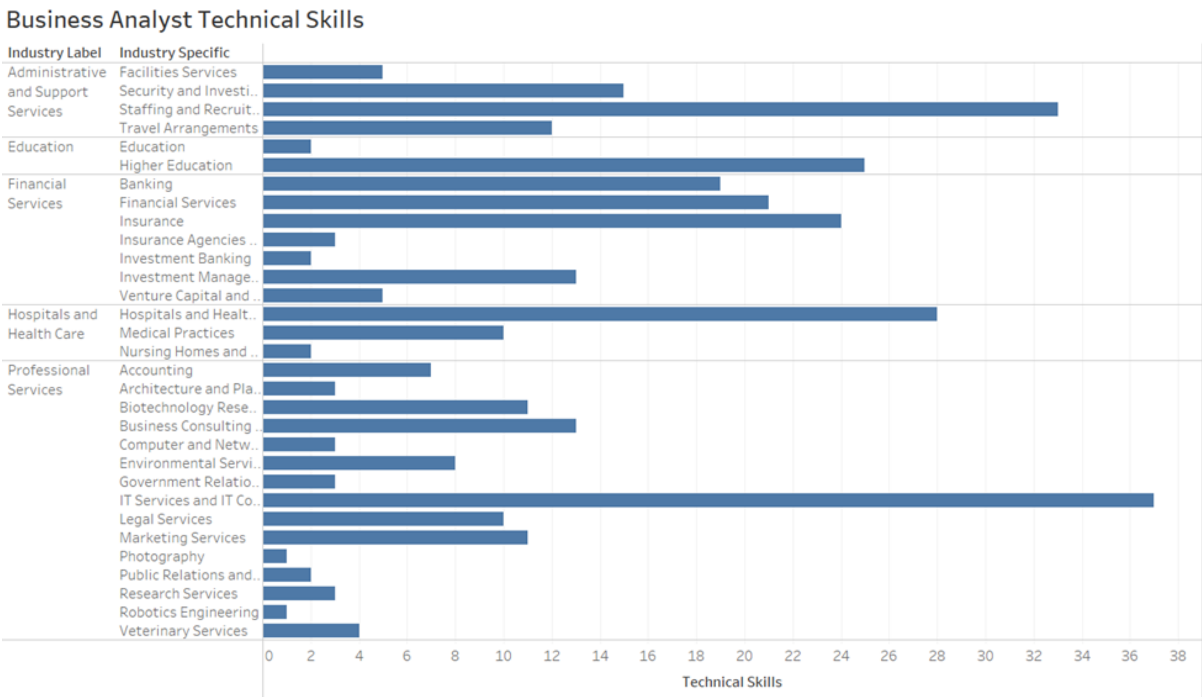


Figure – Business Analyst Technical Skills Needed per Industry

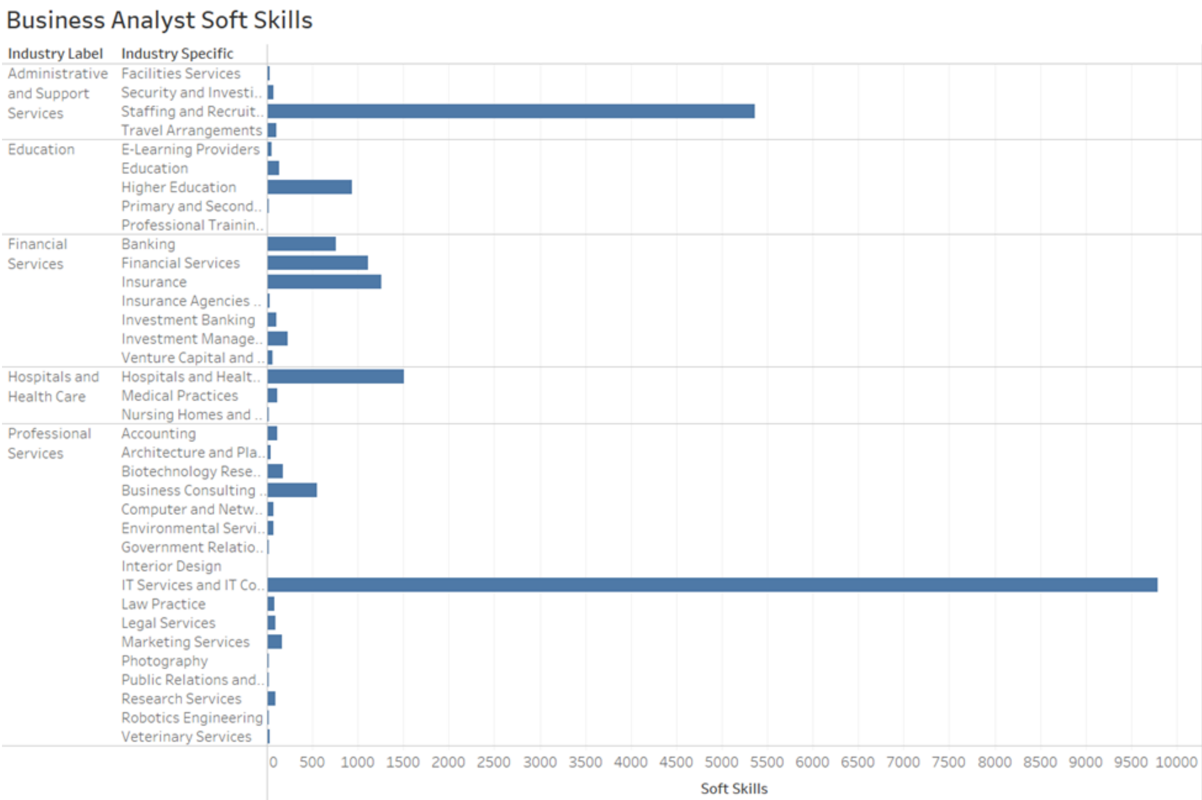


Figure – Business Analyst Soft Skills Needed per Industry

5.2.1 Data Overview

Data analyzed were Business Analyst roles based on 1.3M LinkedIn Job & Skills listings. Additional industry-alignment datasets were also used to map companies to their respective industries. Specific industries that were considered are Professional Services, Administrative & Support Services, Financial Services, Education, and Hospitals & Healthcare, as they recorded the highest volume of recruitment and job listings for business analysts.

5.2.2 Industry Overview

5.2.2.1 Professional Services

In the Professional Services industry, which includes sectors like Biotechnology Research, Business Consulting & Services, Government Relations Services, IT Services & IT Consulting, Marketing Services, and Research Services, business analysts play an important role in facing clients and customers where they are tasked with identifying and solving various business problems. Their responsibility is to assess client needs, translate these into actionable strategies, and deliver solutions that improve efficiency, productivity, or profitability. Business analysts in these sectors must possess strong consulting skills and the ability to manage complex relationships, which often requires balancing the interests of various parties.

A typical demand in this sector is for business analysts to have a deep understanding of project management methodologies, as they are frequently required to manage large-scale, multi-phase projects. They are also expected to be proficient in business process modeling and systems integration. Another critical skill is client relationship management, where analysts must navigate complex client needs and deliver actionable solutions and insights. This requires excellent communication skills to ensure that both technical and non-technical clients and stakeholders understand the proposed solutions.

5.2.2.2 Administrative & Support Services

In Administrative & Support Services, which includes industries such as Facilities Services, Staffing & Recruiting, and Travel Arrangements, business analysts focus on improving the internal operations of organizations. Their role is to streamline workflows, introduce new technologies, and enhance the efficiency of administrative processes. Analysts in this sector are often tasked with optimizing business functions, improving reporting systems, and reducing bottlenecks in operational workflows.

A much-needed demand in this industry is expertise in tools such as PowerBI, Google Analytics, or Python. Business analysts work on workflow optimization, identifying inefficiencies in administrative processes and proposing solutions that reduce operational costs and time. They must also be skilled in analytics, which involves using data to assess performance, retention, and strategies.

5.2.2.3 Financial Services

As per the Financial Services industry, which includes Banking, Financial Services, Insurance, Insurance Agencies and Brokerages, Investment Banking, Investment Management, and Venture Capital & Private Equity Principals, business analysts help organizations manage and navigate regulatory challenges and financial risks effectively. In this sector, business analysts must be well-versed in financial modeling, regulatory compliance, and risk management. Their tasks often involve analyzing financial data to create forecasts, improve profitability, and ensure compliance with financial regulations and laws.

In addition to technical expertise in data analysis and financial reporting, business analysts in this sector are expected to have a deep knowledge and understanding of regulatory standards and risk management frameworks, as they are tasked with identifying areas where their organization may be exposed to financial risks and designing systems or processes to eliminate or mitigate those risks. Familiarity with tools like SAS, Excel, and SQL is essential for conducting complex financial analyses.

5.2.2.4 Education

In the Education sector, which consists of General Education and Higher Education, there has been an increasing demand for professors and teachers specializing in business analytics. Educational institutions are looking for experts who can teach students how to analyze data, make data-driven decisions, and implement analytical tools in various business contexts. As more universities and colleges expand their programs in data science and analytics, business analysts with teaching expertise are needed to prepare the next generation of professionals in the field.

Job listings for business analytics professors often require not only a strong academic background in business analytics, data science, or any related field, but also practical and technical experience with tools such as Power BI, Tableau, SQL, and Python. These educators are expected to design courses that integrate theoretical knowledge with real-world applications, ensuring that students are well-prepared for careers in industries that rely on data and fact-based decision making.

5.2.2.5 Hospitals & Healthcare

The Hospitals & Healthcare sector, which include Medical Practices and Nursing Homes & Residential Care Facilities, business analysts are integral to improving patient care and operational efficiency. They play a role in managing the flow of data, ensuring that systems comply with laws and regulations, and improving clinical and operational processes through data analysis.

A common demand in this industry is proficiency with healthcare data management systems, which are used to maintain and analyze patient records.

Business analysts must also be skilled in process improvement methodologies, which help to reduce inefficiencies in healthcare operations, such as patient wait times or resource allocation. And lastly, analysts in healthcare must ensure that hospitals and clinics comply with data privacy regulations while optimizing the delivery of healthcare services.

5.2.3 Skills Analysis

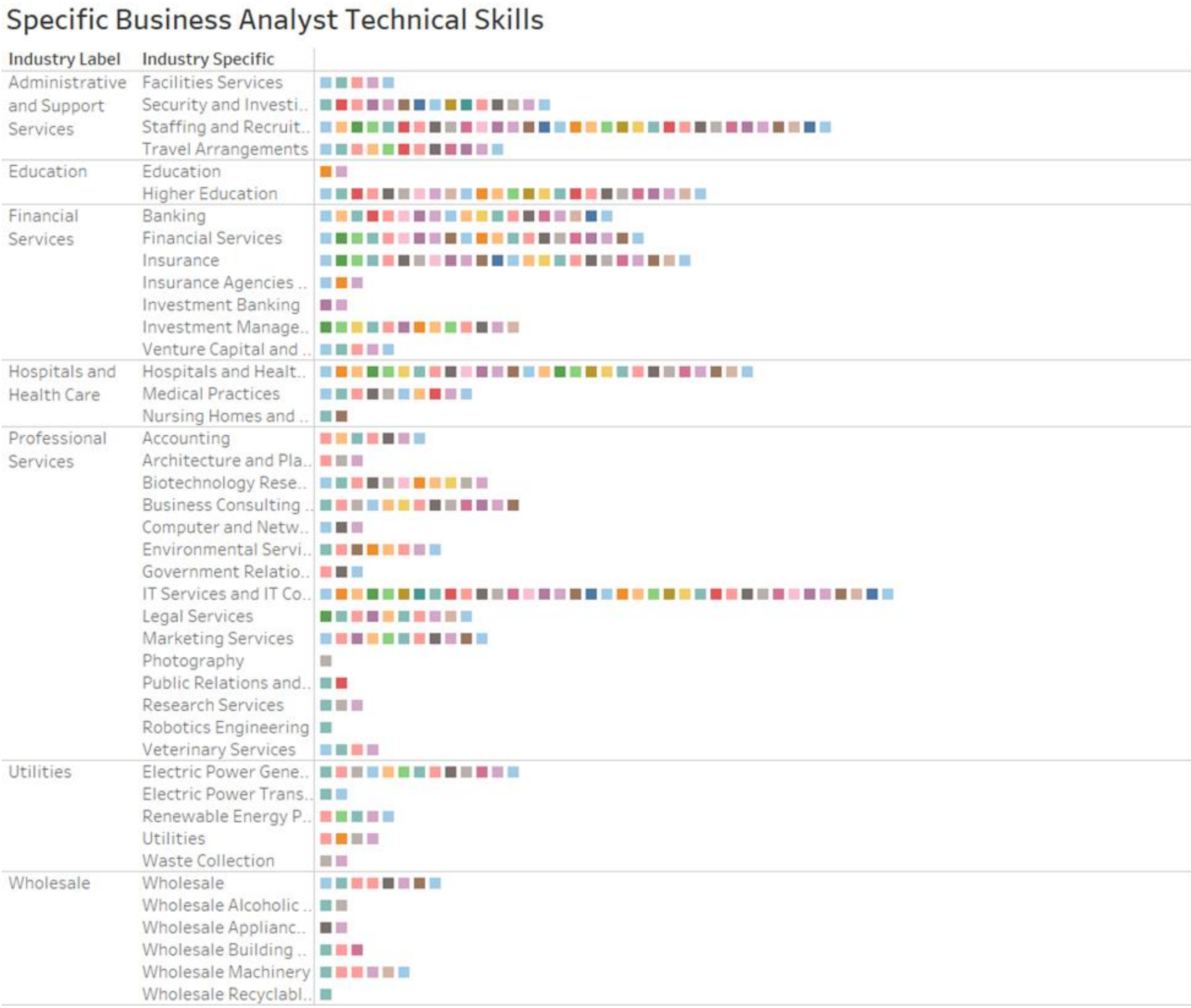


Figure – Specific Technical Skills Needed as a Business Analyst

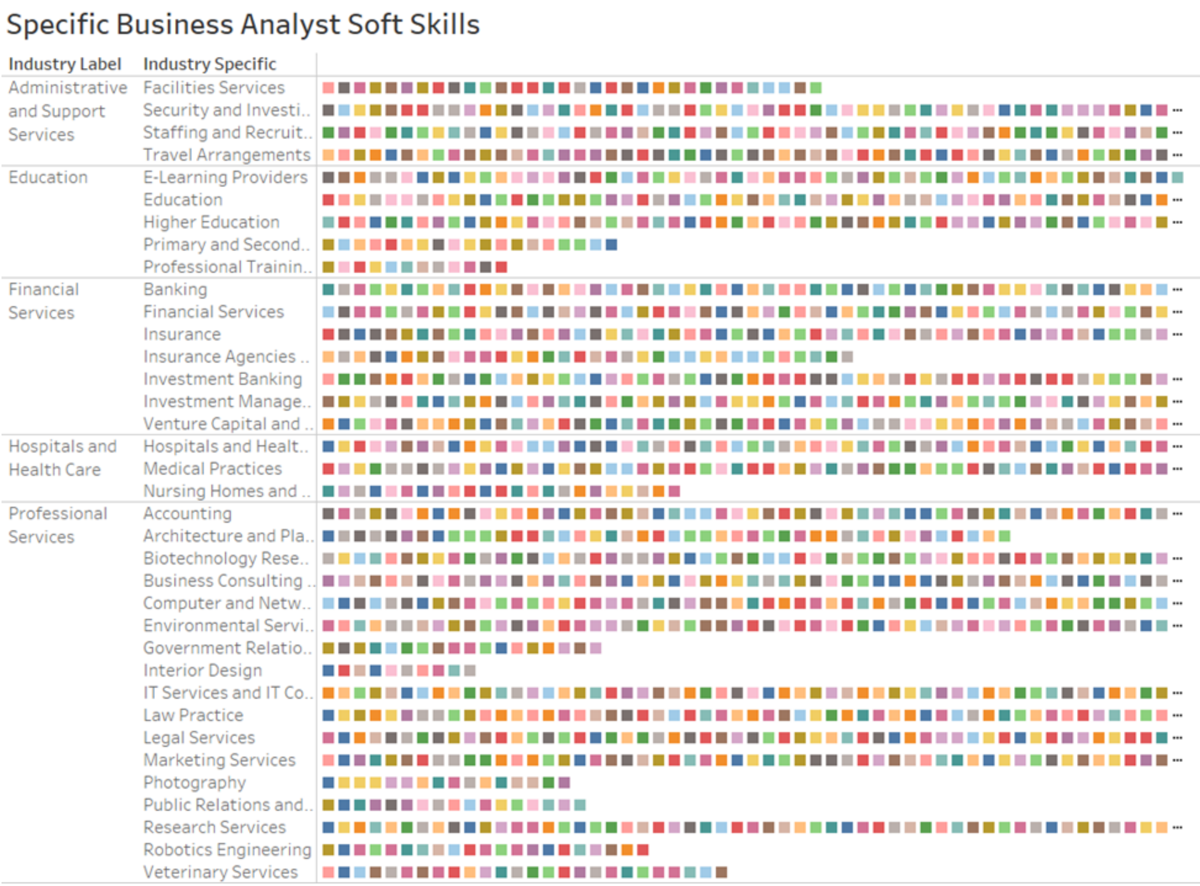


Figure – Specific Soft Skills Needed as a Business Analyst

5.2.3.1 General Skills

There are several universally required skills for business analysts across industries. First, data analysis is critical, as business analysts must be able to analyze large datasets and deliver actionable insights. While strong communication skills, both written and verbal, are essential for conveying findings and recommendations clearly. Problem-solving is another key skill, as business analysts need to apply analytical thinking to overcome business challenges. And finally, project management experience, particularly with methodologies like Agile or Scrum, is valuable for overseeing project timelines and ensuring successful outcomes.

5.2.3.2 Industry-Specific Skills

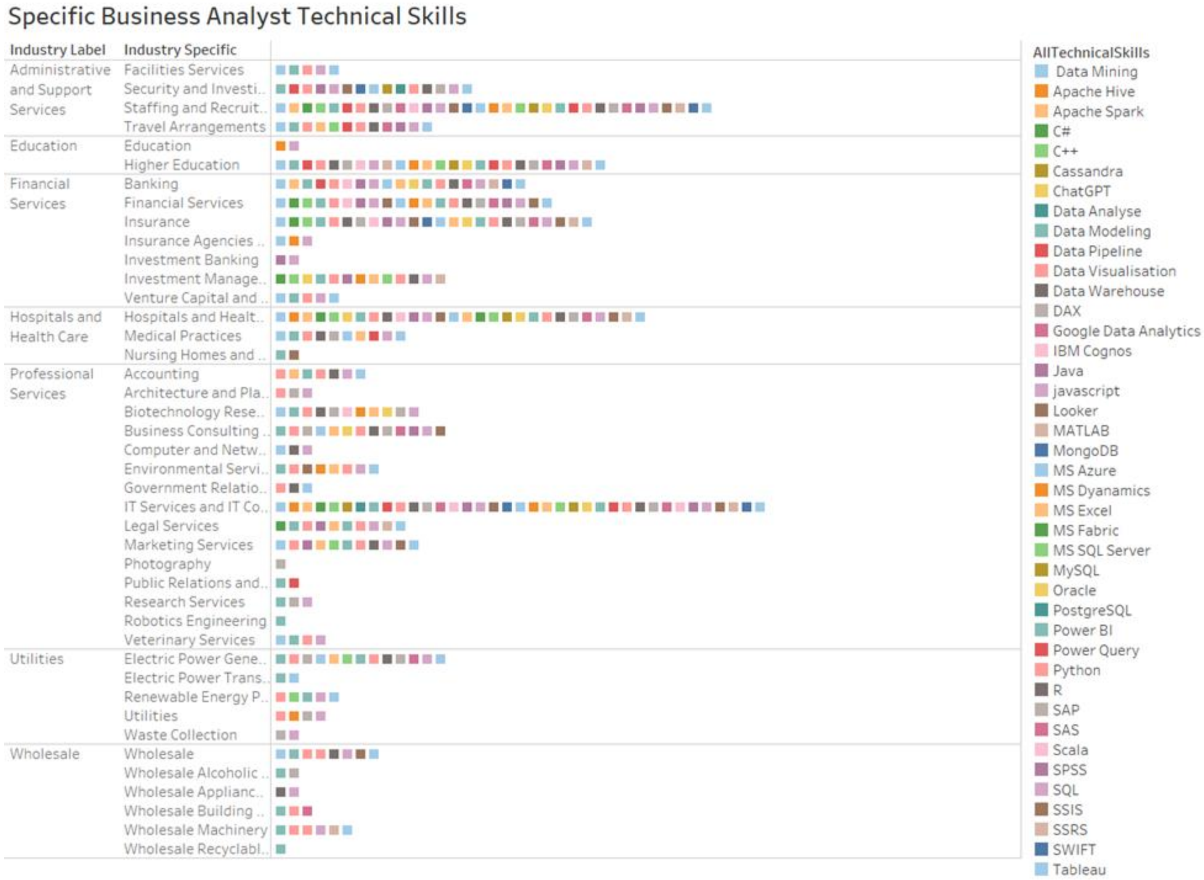
Different industries demand specialized skills for business analysts.

In Professional Services, client management, project management, and consulting methodologies are important. For Administrative & Support Services, process automation, HR analytics, and workflow improvement are key areas of focus. While in Financial Services, business analysts must understand regulatory compliance, financial modeling, and risk management. In the Education sector, education, certifications, skills in educational technology, data-driven decision-making, and academic optimization are essential. And in Hospitals & Healthcare, knowledge of Electronic Health Record (EHR) systems, healthcare data management, and process improvement is important for success.

5.2.4 Qualifications

While a bachelor's degree is the minimum requirement for business analysts, many industry-specific roles require additional certifications to meet the specialized demands of the field. These certifications help professionals develop targeted skills and expertise that are crucial for success in particular sectors.

Appendix:



Legend for Figure-- Specific Technical Skills Needed as a Business Analyst

Legend for Figure-- Specific Soft Skills Needed as a Business Analyst

FOR ASSIGNMENT 02:

Data Analysis Plan of Industry-Specific Needs in Business Analysts

Introduction

Business analyst roles have become pivotal across industries as organizations increasingly turn to data-driven decision-making. Our analysis, which focuses on 1.3 million LinkedIn job listings from 2024, explores trends, skill requirements, and industry-specific needs in business analyst postings.

Background and Motivation

The demand for business analysts is growing as organizations rely more on data to guide decisions. Analyzing 2024 LinkedIn job listings reveals industry-specific needs and required skills for business analysts. This insight helps businesses refine hiring strategies and individuals focus on the right skills for career growth.

Material for Analysis

Data Description

Report is based on a dataset of 1.3 million LinkedIn job postings for Business Analyst positions from 2024. Each job posting includes fields such as: job_number, job_title, company, job_level, job_type, job_link, job_summary, job_skills, job_hiarchy, job_label, and job_specific.

The analysis focuses on extracting data related to technical and soft skills requirements across various industries. For clarification, "Specific Industry" falls under the broader classification of "Industry Label." Each industry is categorized under a broad Industry Label (e.g., Manufacturing, Financial Services) and further classified into Specific Industries (e.g., Waste Collection, Retail, Education).

Data Quality

Completeness: The dataset had a small portion of missing or inconsistent data in fields such as job_skills, job_label, job_specific. These inconsistencies were cleaned up by employing interpolation techniques where appropriate and filtering out incomplete rows where necessary.
Accuracy: The data was sourced from LinkedIn job postings, ensuring relevance and currency for 2024.
Consistency: Industry labels and specific industries are systematically categorized, facilitating reliable analysis of industry-specific needs.

Analytic Approach

Data Processing

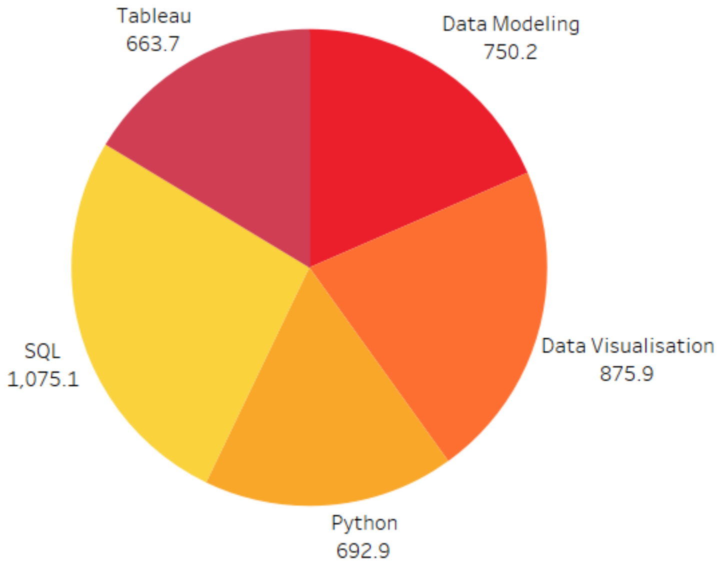
Our analysis plan involved splitting the dataset into soft and technical skills. We used rule-based classification to categorize the job postings into relevant industries and assigned weights to each skill depending on frequency and industry demands. The key tools used were KNIME, Python and Tableau for data processing and visualization and Excel for normalization and comparison of skill frequencies.

Visualization

A series of graphs and charts were created, showing the demand for specific technical and soft skills within industry labels and specific industries.

Findings

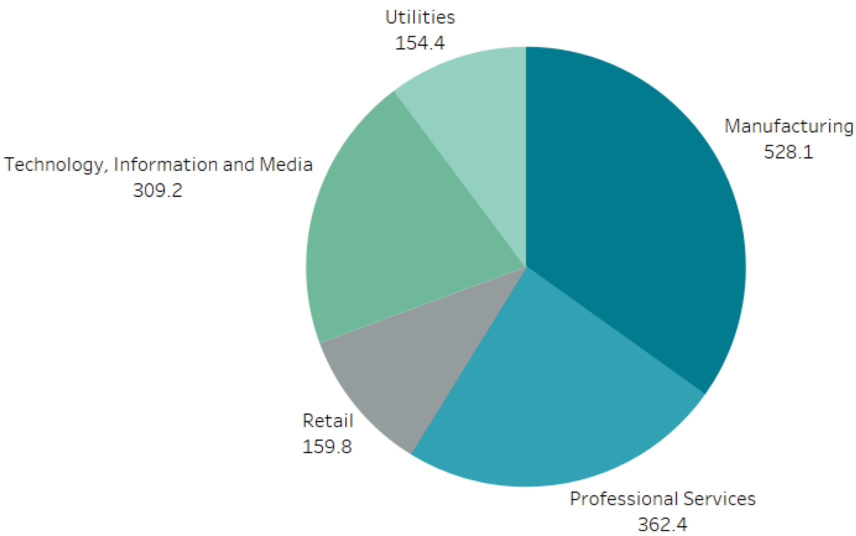
Technical Skills Analysis



Graph 01: Top 5 Technical Skills Required

The most in-demand technical skills across industries are SQL, Data Visualization, Data Modeling, Python, and Tableau, reflecting the widespread need for data analysis, interpretation, and visualization capabilities in Business Analyst roles.

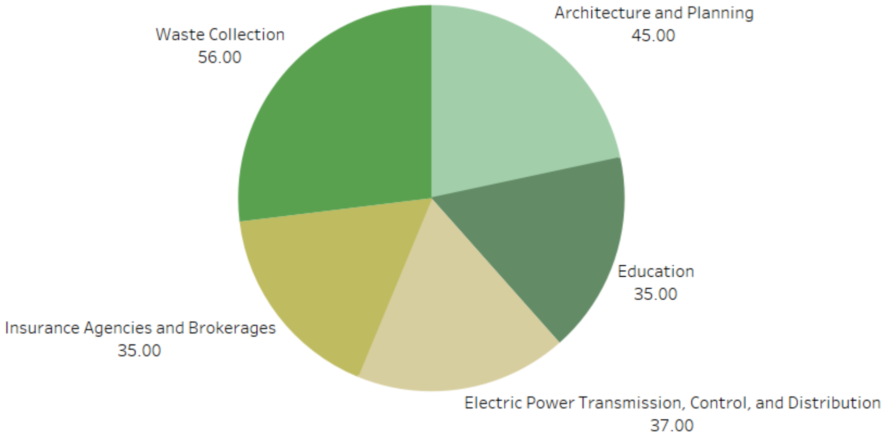
Industry-Label Technical Skills Analysis



Graph 02: *Top 5 Label Industries Requiring Technical Skills*

The top technical skills are consistently in demand across several industries, including Utilities, Retail, Manufacturing, Professional Services, and Technology, Information, and Media. Each of these fields relies heavily on technical skills to enhance productivity, streamline operations, and remain competitive.

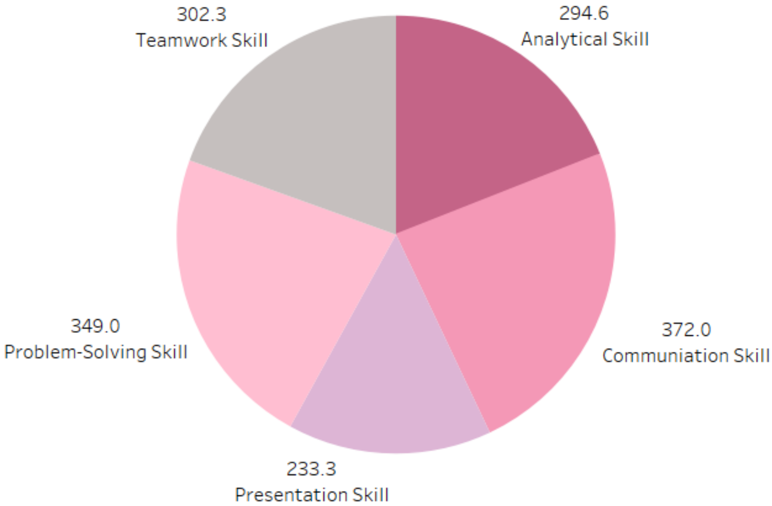
Industry-Specific Technical Skills Analysis



Graph 03: Top 5 Specific Industries Requiring Technical Skills

The top five specific industries with high demand for technical skills include Architecture and Planning, Education, Electric Power Transmission, Control, and Distribution, Insurance Agencies and Brokerages, and Waste Collection. Architecture and Planning rely on data modeling and visualization for design and project planning, while Education utilizes data visualization and SQL to analyze student performance and improve outcomes. The Electric Power industry emphasizes data modeling and SQL for managing complex networks efficiently. Insurance Agencies require SQL and data visualization for effective risk assessment and client insights. Waste Collection has a strong need for data analysis skills to enhance operational efficiency and comply with regulatory standards.

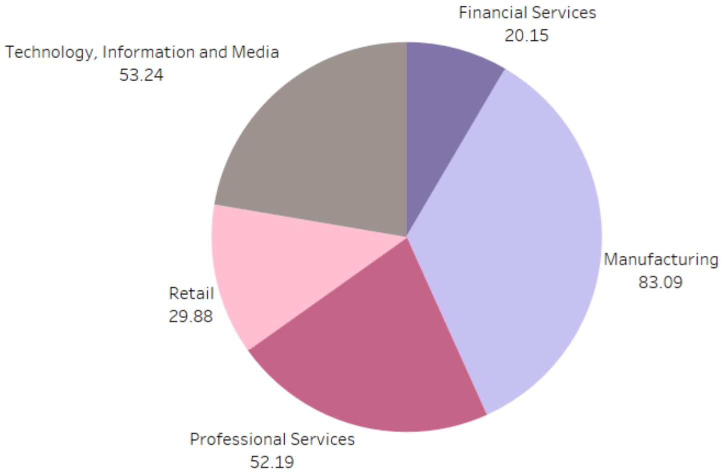
Soft Skills Analysis



Graph 04: Top 5 Soft Skills Required

The most listed soft skills across industries for Business Analyst roles include Communication, Problem-Solving, Teamwork, Analytical Skills, and Presentation Skills. These skills are essential for effective collaboration, critical thinking, and the ability to convey insights clearly, all of which are vital in a data-driven business environment.

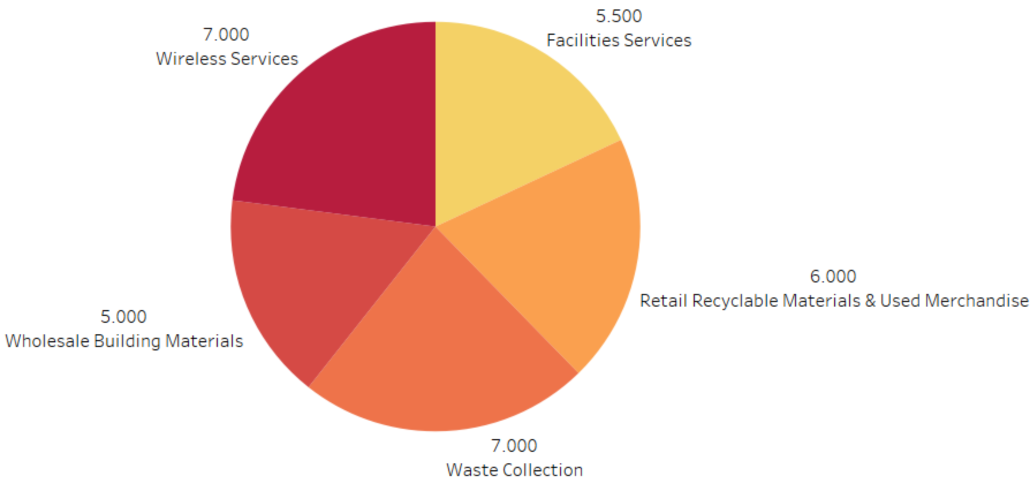
Industry-Label Soft Skills Analysis



Graph 05: Top 5 Label Industries Requiring Soft Skills

Manufacturing has the highest demand for soft skills, essential for managing operations. Professional Services and Technology, Information, and Media also require strong soft skills for client interactions and teamwork. Retail and Financial Services show a lower emphasis on soft skills, suggesting a greater focus on technical skills in these fields. Each of these industries relies strongly on soft skills to improve collaboration, enhance problem-solving, and maintain a competitive edge.

Industry-Specific Soft Skills Analysis



Graph 06: Top 5 Specific Industries Requiring Soft Skills

Wireless Services and Waste Collection industries show a high need for communication and teamwork, essential for managing customer interactions and coordinating field operations. Retail Recyclable Materials & Used Merchandise, along with Facilities Services, emphasize problem-solving and organizational skills, supporting the efficient management of inventory and compliance with environmental regulations. Wholesale

Building Materials also highlights a demand for problem-solving, given the need to manage logistical challenges and customer relationships effectively.

Key Insights

Core Technical Skills: SQL and Data Visualization are universally required, serving as foundational skills across nearly all industries.

Non-Traditional Data Industries: Industries like Waste Collection and Architecture increasingly value data-driven decision-making, highlighting an expanding role for analytics.

Soft Skills Emphasis: Communication, Problem-Solving, and Analytical Skills are universally critical, underscoring the importance of effective collaboration and adaptability in Business Analysis.

Interpretation

The findings suggest that while there is a core set of technical and soft skills valued across industries, certain industries have unique skill requirements based on their operational focus and strategic priorities. The demand for skills like SQL and Python indicates a broad need for data literacy, while other skills in areas like communication or teamwork reflect tailored needs. There are also the rising data needs in non-traditional sectors which indicate a trend towards data-driven decision-making across all fields.

Recommendations

The analysis shows the growing need for a blend of soft and technical skills across industries. Cross-training programs that teach both data analysis and project management can help business analysts work better across teams. Workshops on new data tools are also important for industries moving toward data-focused operations. Partnering with universities to develop courses and internships can ensure a steady flow of talent with the right mix of technical and soft skills. This shift highlights the value of employees who can adapt and collaborate effectively in a data-driven environment.

Conclusion

This analysis outlines industry-specific skill needs for Business Analyst roles, showing both common and unique demands across sectors. Aligning training and hiring with these insights can help meet rising expectations for data literacy and collaboration, ensuring Business Analysts are prepared to succeed across industries.

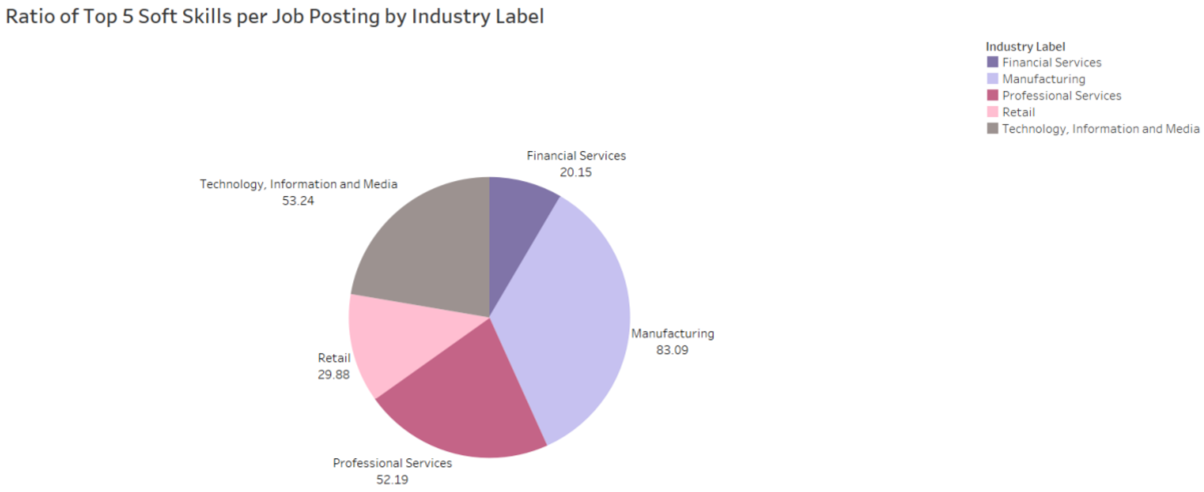
References

Bouloudnine, S. (2024, June 1). *Complete LinkedIn industry list 2024*. lobstr.io. Retrieved November 3, 2024, from https://www.lobstr.io/blog/linkedin-industry-list

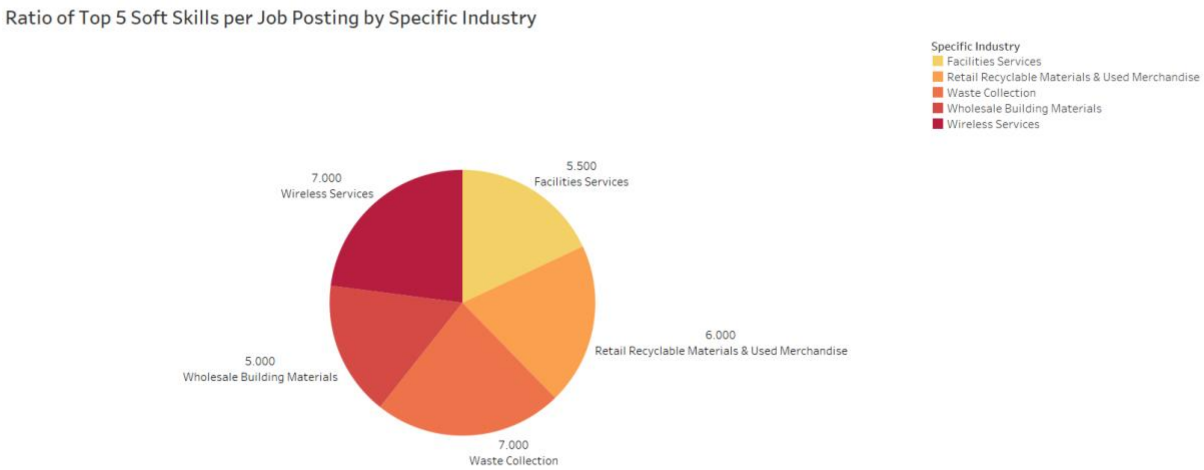
Excel Macros for Dummies (3rd ed.). (2022). John Wiley & Sons.

Microsoft Excel Power Pivot & Power Query for Dummies (2nd ed.). (2022). John Wiley & Sons.

Appendices

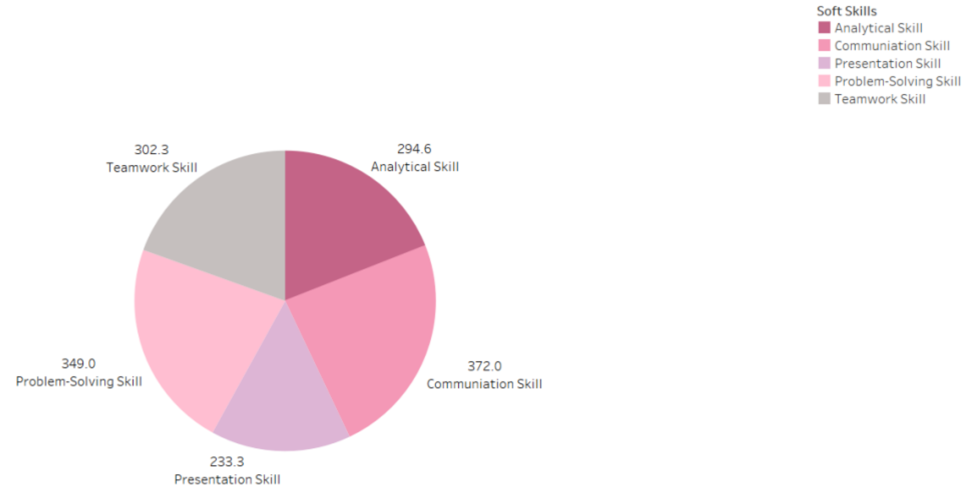


Pie Graph 01: Ratio of Top 5 Soft Skills per Job Posting by Industry Label



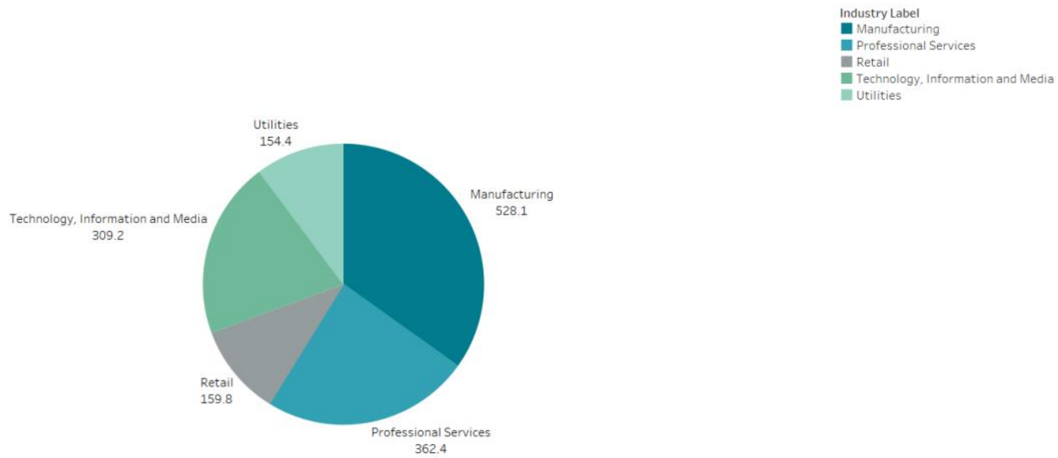
Pie Graph 02: Ratio of Top 5 Soft Skills per Job Posting by Specific Industry

Top 5 Soft Skills Required

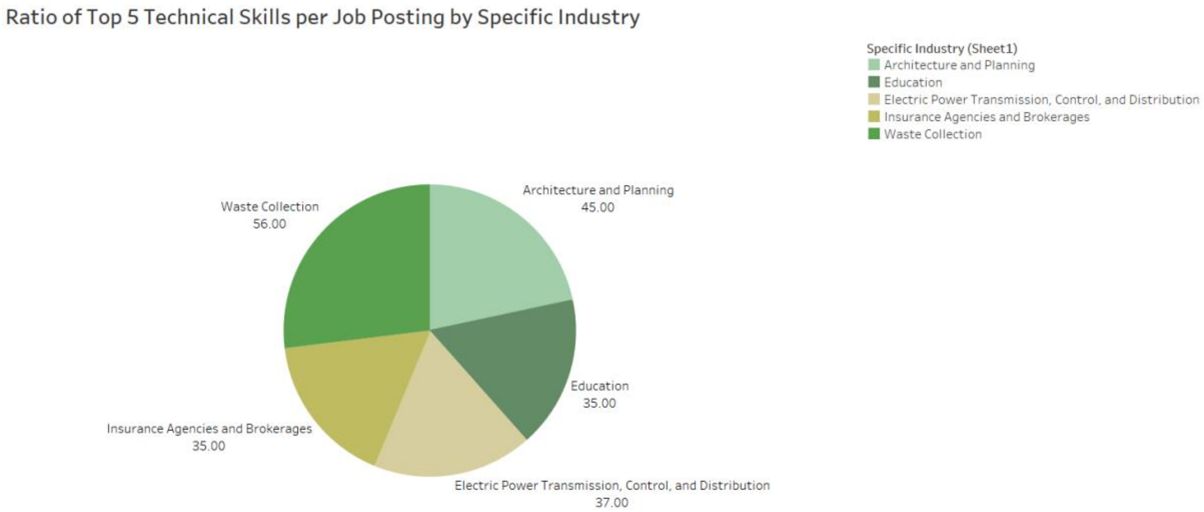


Pie Graph 03: Top 5 Soft Skills Required

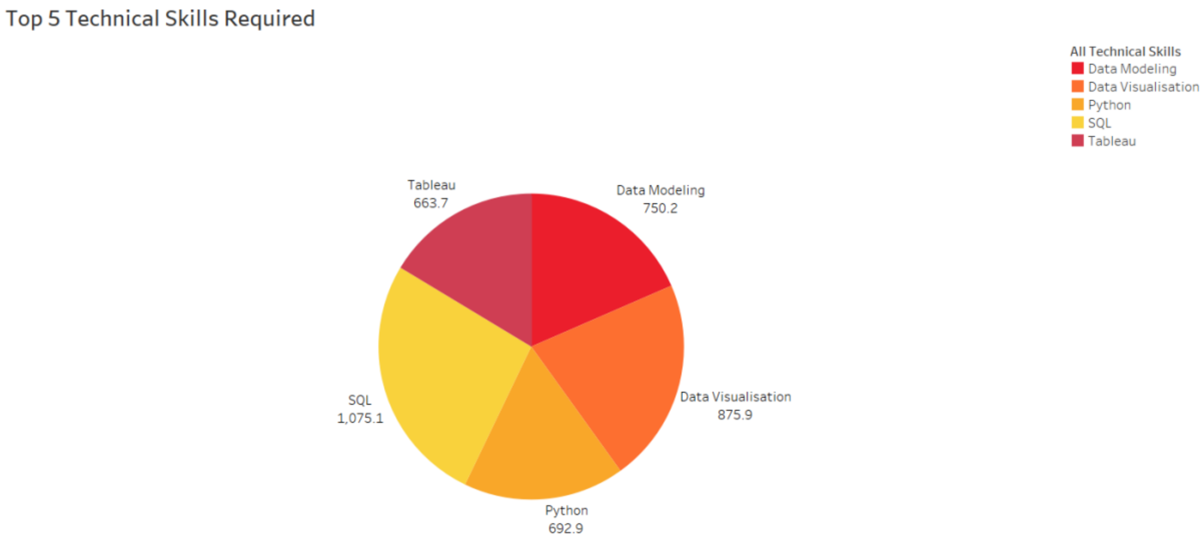
Ratio of Top 5 Technical Skills per Job Posting by Industry Label



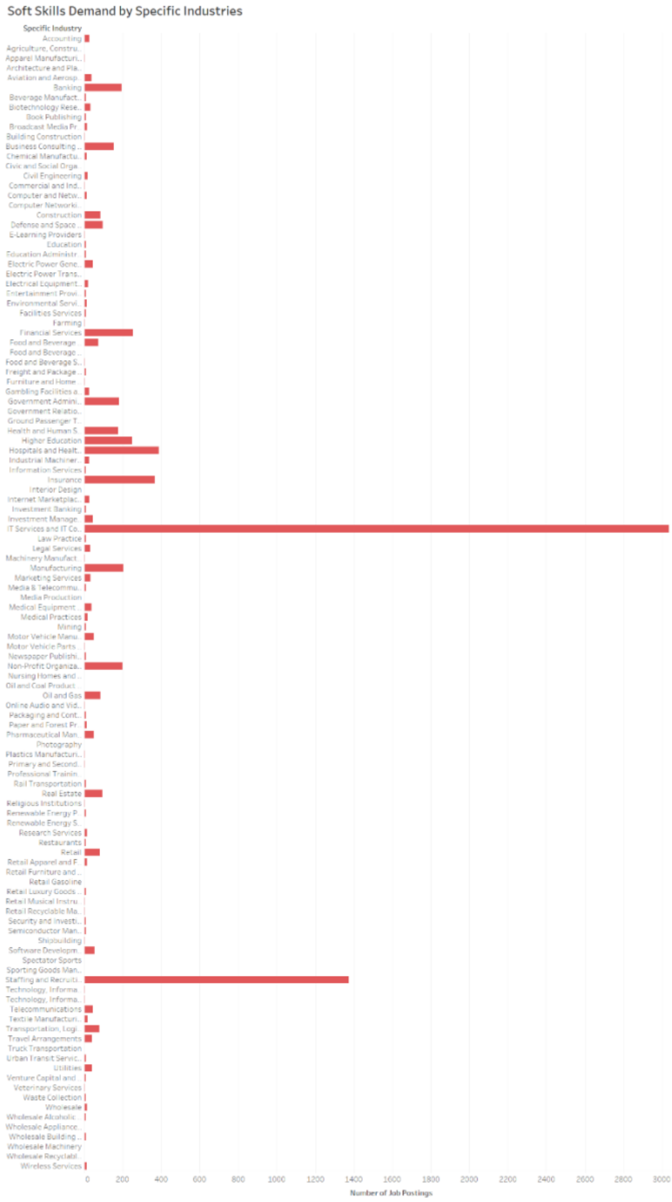
Pie Graph 04: Ratio of Top 5 Technical Skills per Job Posting by Industry Label



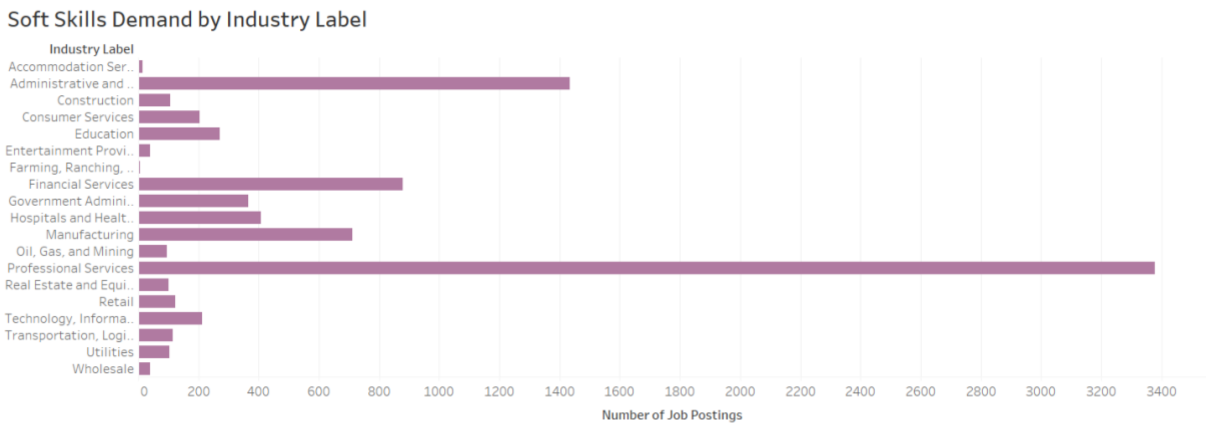
Pie Graph 05: Ratio of Top 5 Technical Skills per Job Posting by Specific Industry



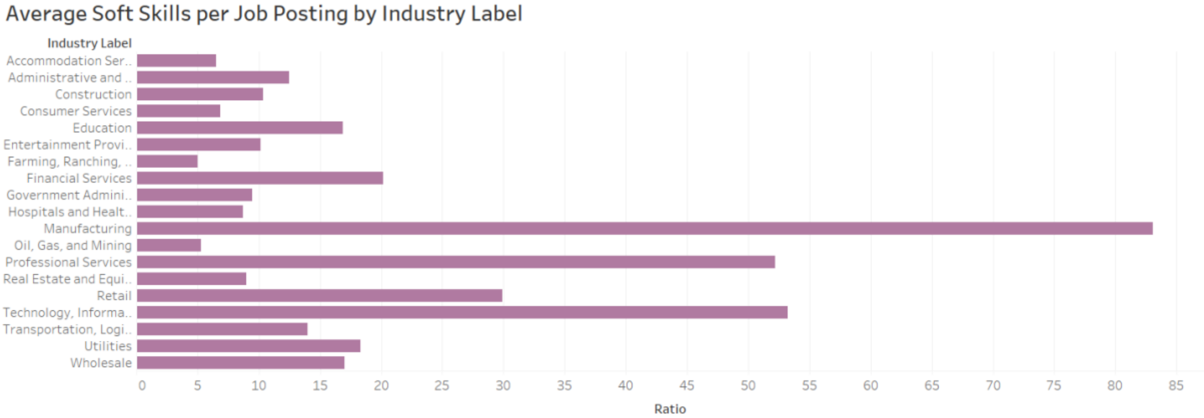
Pie Graph 06: Top 5 Technical Skills Required



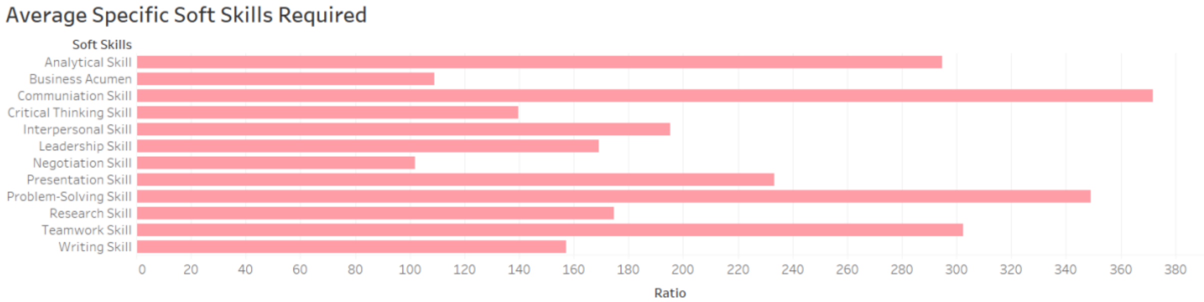
Bar Graph 01: Soft Skills Demand by Specific Industries



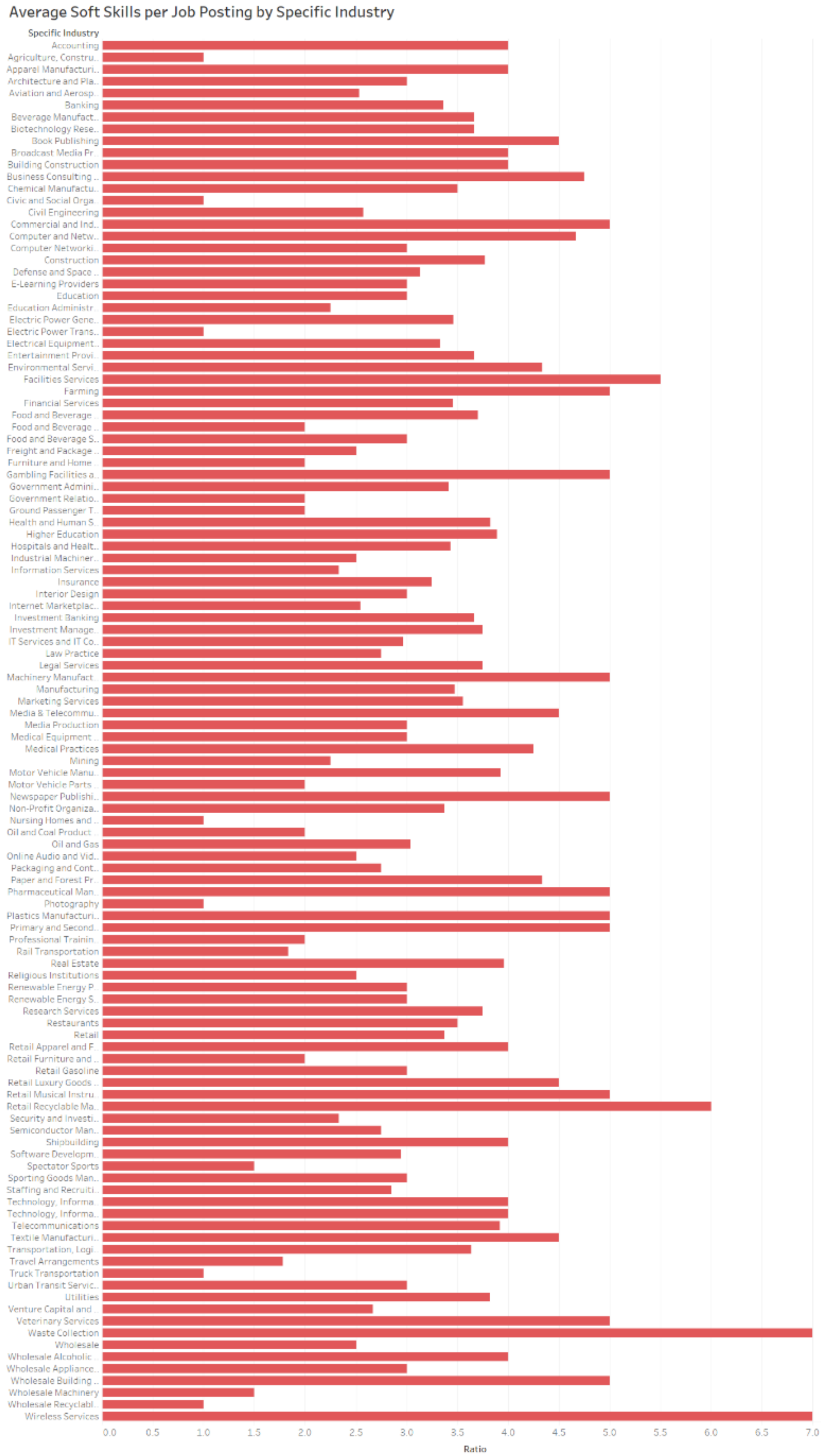
Bar Graph 02: Soft Skills Demand by Industry Label



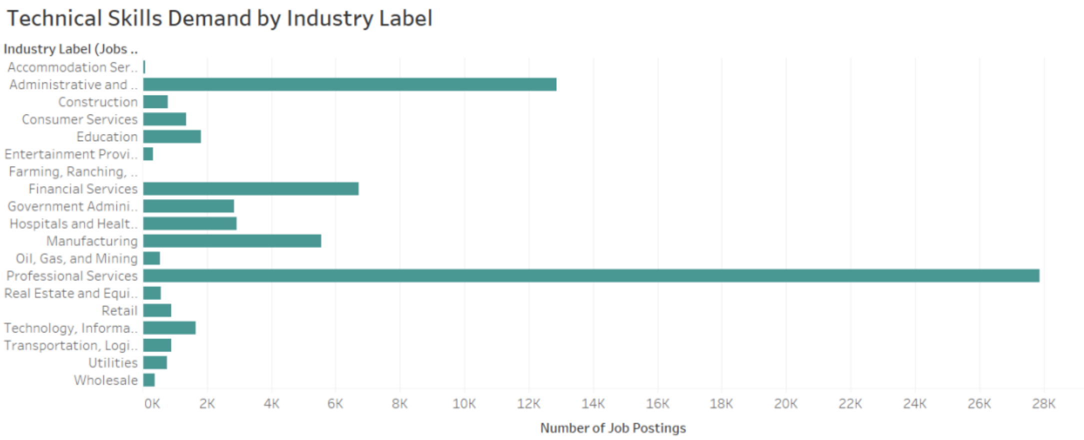
Bar Graph 03: Average Soft Skills per Job Posting by Industry Label



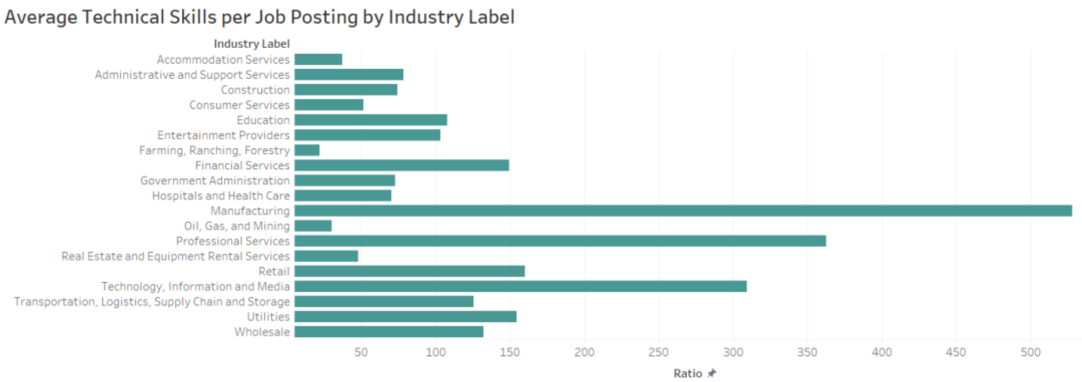
Bar Graph 04: Average Specific Soft Skills Required



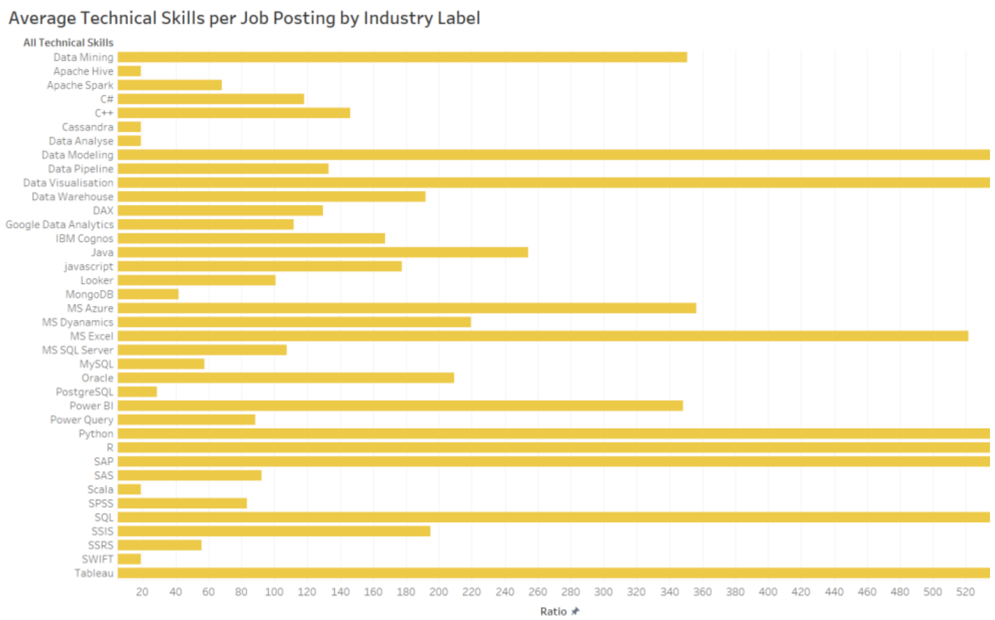
Bar Graph 05: Average Soft Skills per Job Posting by Specific Industry



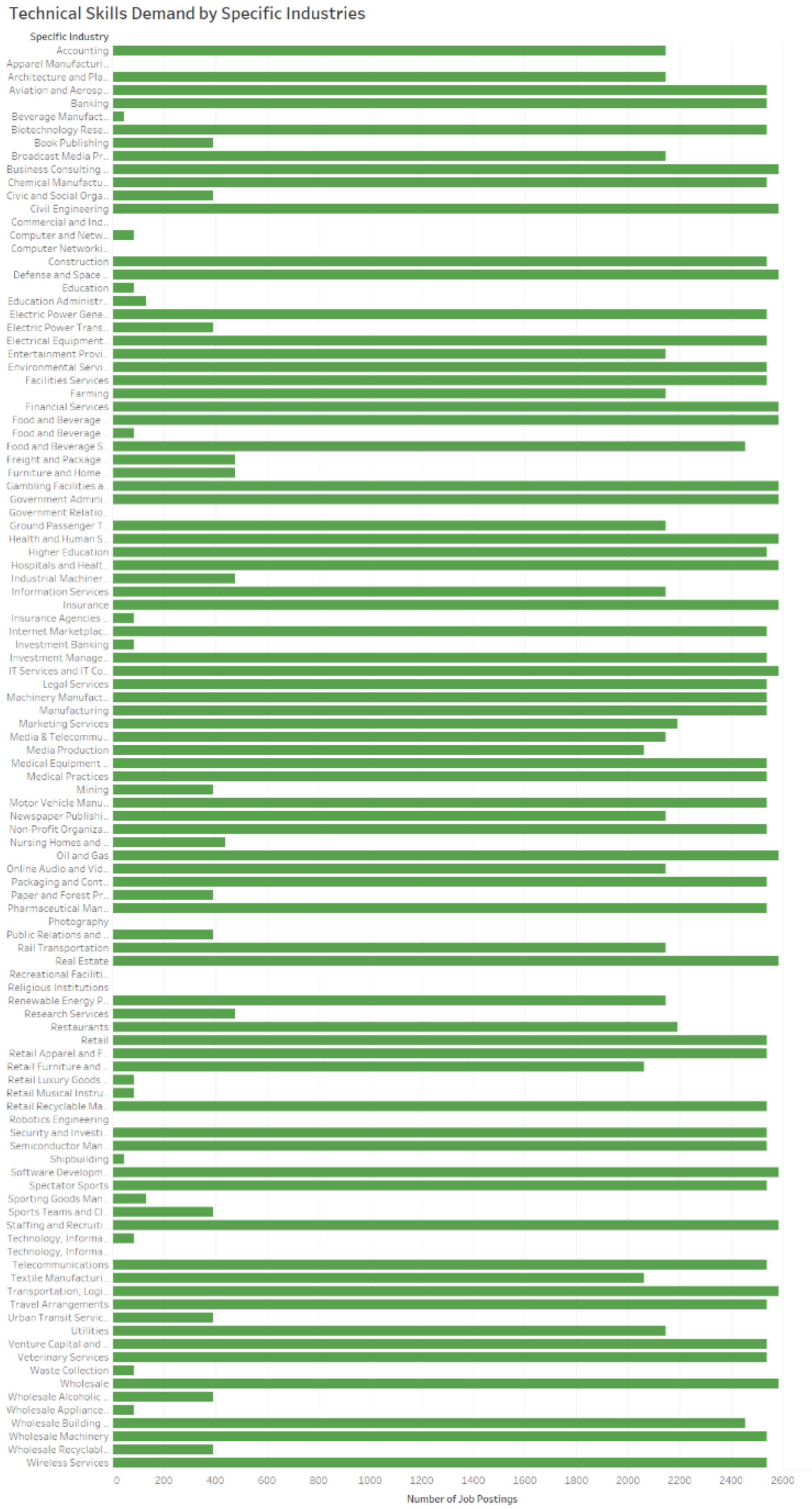
Bar Graph 06: Technical Skills Demand by Industry Label



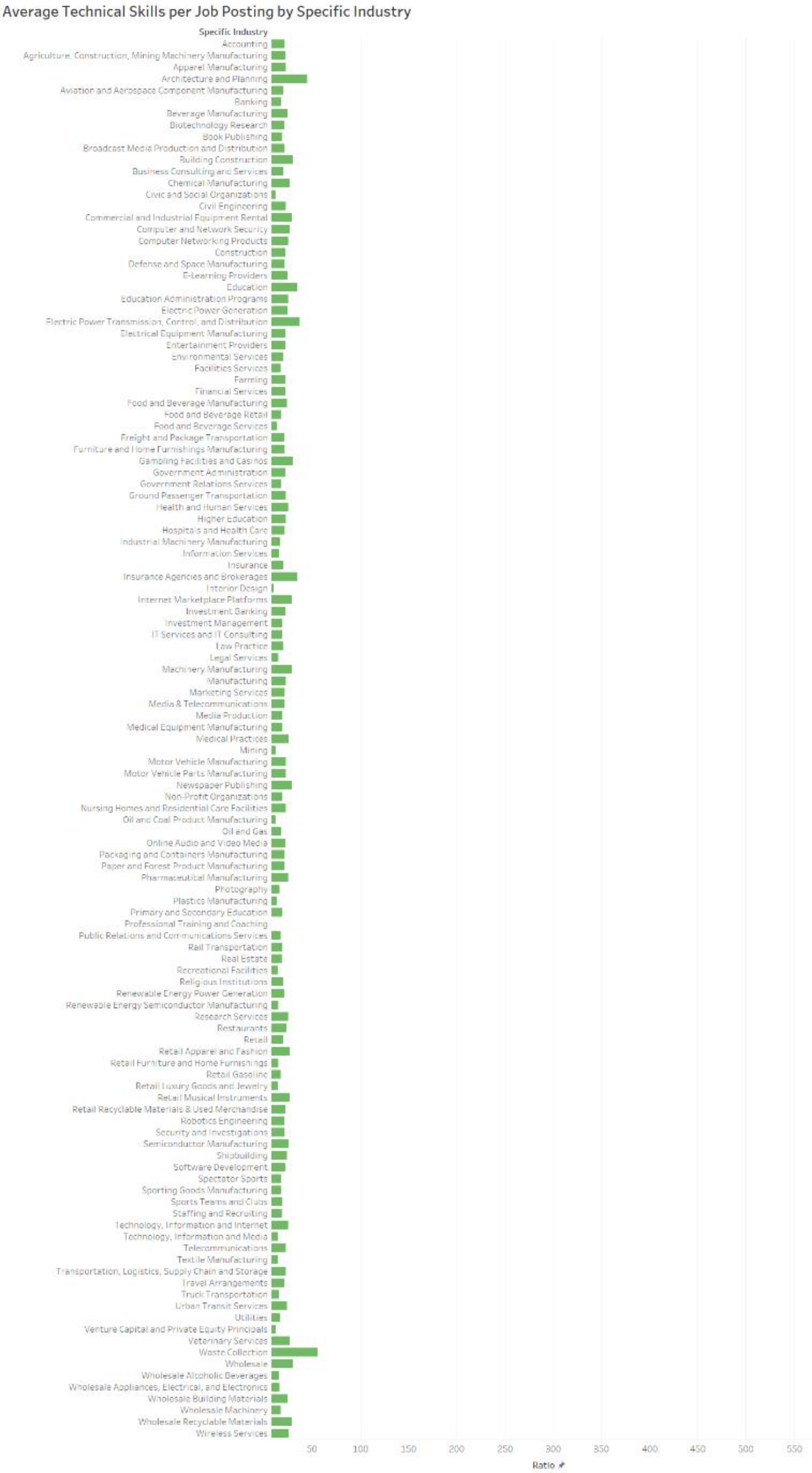
Bar Graph 07: Average Technical Skills per Job Posting by Industry Label



Bar Graph 08: Average Specific Technical Skills Required



Bar Graph 09: Technical Skills Demand by Specific Industries



Bar Graph 10: Average Technical Skills per Job Posting by Specific Industry

The goal of this report is to analyze key trends and patterns in the job market, focusing on technical skills, job types, salaries, and qualifications. Based on the data provided and the analysis of various charts, the report presents factual insights into the current labor market. This content will focus on the results of the analysis and their interpretations, providing a thorough examination of the factors affecting the job market, including salary trends, required skills, and geographical and educational influences.

1. Data scope

The project focuses on identifying skill requirements in the field of business analysis. The results will assist recruitment companies in categorizing candidates according to the necessary skills in each specific area.

Data is collected from job postings on LinkedIn across three countries, with the source provided by Kaggle.com uploaded by Asaniczka. Although the initial data includes various types of jobs, the project only focuses on positions related to business analysis. Irrelevant postings will be removed.

Two main skill groups considered are soft skills and technical skills. Important soft skills include communication, problem-solving, critical thinking, and interpersonal skills. Technical skills include data analysis, use of visualization tools, statistical skills, and BI tools.

The data is predominantly distributed in the US, accounting for about 85% of the entire dataset. The remainder comes from countries such as Australia, Canada, and the United Kingdom, each accounting for less than 10%. Therefore, the project focuses on analyzing the job market in the US, including factors such as salary levels, experience requirements, and job types.

Assumption: The LinkedIn data sample is random and representative of the main characteristics of the population. Business analyst positions are filtered through the keywords "business" and "analysis" in the job title.

Project phases:

1. Survey data and identify business requirements.
2. Evaluate data quality and identify necessary columns for analysis.
3. Clean data to fit the analysis process.
4. Conduct preliminary analysis to detect key characteristics of the job market.
5. Build data analysis models and connect data.
6. Analyze trends and find relationships between variables.

2. Data Quality Evaluation

Dataset	Analysis	Key Findings
Job_skills	Missing Values	- job_link: Complete
		- job_skills: 429 missing

	Duplicates	None
	Data Types	Object (strings)
	Cleaning	Saved as cleaned_data_file.csv
Job_summary	Missing Values	None
	Duplicates	None
	Data Types	Object (strings)
	Cleaning	Saved as cleaned_job_summary.csv
Linkedin_job_postings	Missing Values	- company: 3
		- job_location: 5
		- Others: Complete
	Duplicates	None
	Data Types	Object (non-numeric)
	Cleaning	Saved as cleaned_linkedin_job_postings.csv
Conclusion	Data Cleaning Outcome	Data cleaned and saved. All datasets exhibit high quality, uniqueness, and completeness, ready for further analysis.

The Data Quality assessment was conducted to validate the readiness and reliability of three key datasets—Job_Skill_20, Job_Summary, and LinkedIn_Job_Postings—prior to detailed analysis. The goal of this quality check is to confirm that each dataset is dependable and properly structured to support robust analysis.

Completeness and Management of Missing Values:

- In Job_Skill_20, 429 entries in the job_skills column are missing, indicating incomplete skill information for some jobs. This gap may influence skill-specific analyses, yet these records retain value for job-related insights and were kept for comprehensive coverage.

- Both Job_Summary and LinkedIn_Job_Postings demonstrate near-complete data integrity. LinkedIn_Job_Postings contains minimal missing entries, with company missing in 3 instances and job_location in 5 instances. These minor gaps were deemed non-disruptive to the dataset's overall utility.

Uniqueness and Duplication Check:

- A duplicate check across all three datasets confirmed that each job entry is unique, eliminating risks of inflated or biased findings due to repeated entries. This validation supports accurate job frequency distributions and guarantees unique representation of each job role.

Data Type Consistency:

- Data type uniformity was verified, with all columns identified as object type (string data) across each dataset. While numeric columns were absent, standardizing data types is critical for seamless handling of text data. This consistency is particularly beneficial for text-based analytical techniques, ensuring all categorical and descriptive data is appropriately formatted.

Data Cleaning Process:

- Each dataset was processed to remove inconsistencies and standardize formatting, and was then saved as a new, cleaned file. This cleaning ensures the datasets are primed for analysis, focusing on textual attributes without introducing artifacts from initial data irregularities.

Conclusion: The Data Quality review confirms that the Job_Skill_20, Job_Summary, and LinkedIn_Job_Postings datasets meet necessary standards for completeness, uniqueness, and consistency. This foundational quality check ensures that all three datasets are analytically sound and ready for the subsequent stages of data processing and model development.

Note: Code for data quality assessment is included in the appendices.

3. Data Exploration

During the Exploratory Data Analysis process, we conducted an in-depth analysis of three datasets: *Job_Skill_20*, *Job_Summary*, and *Linkedin_Job_Postings*. The objective of this analysis was to understand the structure, quality, and characteristics of each dataset, thereby identifying missing values, duplicate rows, and the distribution of key attributes. Through the process of examining data columns and reviewing common values, we were able to assess data integrity, identify important information patterns, and guide subsequent detailed analysis steps—especially for the text attributes that these tables contain.

Dataset	Analysis Content	Detailed Information
Job_Skill	Missing Values	- job_link: No missing values.- job_skills: 429 missing values, meaning some jobs lack skill data, which could impact skill-based analyses.
	Duplicate Rows	No duplicate rows. This indicates that the data is unique for each job entry.
	Descriptive Statistics	- job_link: 259,276 unique values, no duplicates. - job_skills: 258,847 values, with 258,155 unique values. Some skills are commonly required across various jobs,

		while most jobs have specific skill requirements.
	Common Value Distribution	Common skill clusters in job_skills, e.g., “Front Counter, DriveThru, Outside Order Taker...” appear frequently, often related to industries like hospitality, retail, and customer service.
	Additional Information	There are no numeric columns in the dataset, so histograms cannot be generated for numerical data. Analysis will mainly focus on textual columns.
Job_Summary	Missing Values	- job_link and job_summary: No missing values. All jobs have full URLs and job descriptions.
	Duplicate Rows	No duplicate rows, ensuring each job is unique.
	Descriptive Statistics	- job_link: 259,466 unique values. - job_summary: 259,466 values, with 215,045 unique values. Some job descriptions are repeated across multiple roles, especially for large companies. For instance, “Dollar General Corporation has been delivering...” appears 921 times.
	Additional Information	There are no numeric columns, so histograms cannot be generated. The data is primarily textual, meaning analysis will focus on text attributes.
Linkedin_Job_Postings	Missing Values	- company: 3 missing values.- job_location: 5 missing values.- Other columns have no missing values, ensuring

		most job information is complete.
	Duplicate Rows	No duplicate rows, confirming each job is unique.
	Descriptive Statistics	- job_link: 269,690 unique values.- last_processed_time: 144,585 unique processing times, with the most common time "2024-01-19 09:45:09.215838+00" appearing 125,099 times. - got_summary and got_ner: Two values, "t" and "f" (processed or not processed). - job_title: 152,302 unique values, with the most common job being "LEAD SALES ASSOCIATE-FT." - company: 42,101 companies, with "Health eCareers" being the most frequent. - job_location: 17,419 locations, with "New York, NY" as the most common. - first_seen: 6 unique dates, with "2024-01-14" as the most common. - search_city and search_country: Most common values are "Baytown" and "United States." - job_level and job_type: Most common values are "Mid senior" and "Onsite."
	Additional Information	There are no numeric columns, so histograms cannot be generated.

Overall, the data exploration process from the three tables *Job_Skill_20*, *Job_Summary*, and *Linkedin_Job_Postings* has provided a comprehensive view of the data characteristics and quality. The identification of missing values and common phrases has helped clarify important aspects of the labor market and required skills. Although there were no numerical columns for distribution analysis, the rich text attributes of each table present opportunities for deeper exploration of common skill patterns, job positions, and job descriptions. The information gathered from this stage will serve as a foundation for in-depth analysis steps and help optimize data-driven decisions in the context of current recruitment and skill requirements.

Note: Code for data exploration is included in the appendices.

4. Data Analysis

Building a Dataset from Technical Skills and Salary Levels

We have filtered the original data to obtain two quality datasets: **Job with Technical Skill**and **Job Summary Salary**. The goal of this step is to build a new dataset that helps illustrate the relationship between the technical skills required for jobs related to **data analysis** and the corresponding average salary for each skill.
To achieve this goal, the **Knime** tool was used to perform the process of synthesizing and combining data. Knime allows us to build a visual workflow to integrate the two filtered datasets, thereby creating a new dataset with high analytical value. Through this process, we can identify and evaluate the relationship between skills and salaries, helping to provide valuable insights into the impact of technical skills on income in the field of data analysis. **Knime Workflow for Analysis Process** The image below illustrates the workflow in **Knime** used to create the **Job Summary Salary** data table. This workflow includes steps from importing data from CSV and Excel files, performing grouping, and applying Association Rules to find meaningful relationships. After the data has been processed, the results are recorded into a new data table, ready for deeper analysis.
Knime Workflow:

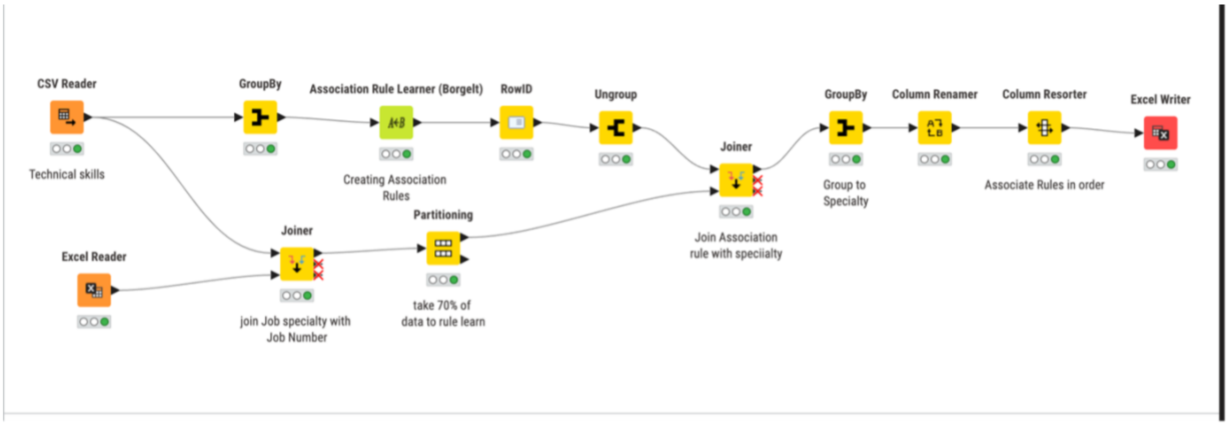


Figure: **Job Technical Skills and Salary Association Analysis Workflow**
Table: **Job Technical Skills Salary Summary**

RuleID	RuleSupport	RelativeRuleSuppo	RuleConfident%	Mean(A	Mean(R	RuleLift	RuleLeft	Mean(A	Mean(Relative	Recomende	AntecedentTechnicalSkill	MeanYearlySalary
Row39	133	2.8769	92.4	144	3.11	2.657	265.7	1607	34.761	SQL	Python, Tableau, Data Analysis	102977.4912
Row11	128	2.7688	91.4	140	3.03	2.6302	263.02	1607	34.761	SQL	R, Python, Data Analysis	100764.4679
Row45	136	2.9418	90.7	150	3.24	2.6083	260.83	1607	34.761	SQL	Python, Data Visualisation, Data Analysis	103838.1915
Row25	124	2.6822	84.4	147	3.18	1.734	173.4	2249	48.648	Data Analysis	Data Mining, Data Visualisation, SQL	109434.8805
Row60	124	2.6822	83.8	148	3.2	1.7222	172.22	2249	48.648	Data Analysis	Tableau, Data Visualisation, MS Excel	106508.4913
Row112	140	3.0283	83.3	168	3.63	1.713	171.3	2249	48.648	Data Analysis	Data Visualisation, Data Reporting, SQL	103827.7752
Row87	127	2.7471	82.5	154	3.33	1.6952	169.52	2249	48.648	Data Analysis	Tableau, Data Reporting, SQL	103121.0066
Row83	163	3.5258	81.1	201	4.35	1.667	166.7	2249	48.648	Data Analysis	Tableau, MS Excel, SQL	104437.4215
Row107	140	3.0283	80.9	173	3.74	1.6635	166.35	2249	48.648	Data Analysis	Data Visualisation, MS Excel, SQL	105179.8829
Row56	119	2.5741	80.4	148	3.2	2.3131	231.31	1607	34.761	SQL	Tableau, Data Visualisation, MS Excel	106508.4913
Row27	124	2.6822	80	155	3.35	5.0456	504.56	733	15.856	Data Visualisati	Data Mining, SQL, Data Analysis	103800.2771
Row72	208	4.4992	80	260	5.62	1.6445	164.45	2249	48.648	Data Analysis	Tableau, Data Visualisation, SQL	108895.4958
Row10	128	2.7688	78.5	163	3.53	1.6142	161.42	2249	48.648	Data Analysis	R, Python, SQL	104888.8014
Row88	127	2.7471	77.9	163	3.53	2.2414	224.14	1607	34.761	SQL	Tableau, Data Reporting, Data Analysis	100740.7012
Row26	124	2.6822	77.5	160	3.46	2.2295	222.95	1607	34.761	SQL	Data Mining, Data Visualisation, Data Analysis	105340.327
Row73	208	4.4992	77.3	269	5.82	2.2244	222.44	1607	34.761	SQL	Tableau, Data Visualisation, Data Analysis	105214.6419
Row124	138	2.9851	77.1	179	3.87	1.5847	158.47	2249	48.648	Data Analysis	MS Excel, Data Reporting, SQL	100701.7848
Row84	163	3.5258	76.9	212	4.59	2.2119	221.19	1607	34.761	SQL	Tableau, MS Excel, Data Analysis	101570.617
Row44	136	2.9418	76.8	177	3.83	1.5794	157.94	2249	48.648	Data Analysis	Python, Data Visualisation, SQL	107529.4293
Row93	116	2.5092	75.8	153	3.31	1.5585	155.85	2249	48.648	Data Analysis	Data Modeling, Data Visualisation, SQL	110737.4954
Row38	133	2.8769	74.7	178	3.85	1.5359	153.59	2249	48.648	Data Analysis	Python, Tableau, SQL	106770.05
Row0	116	2.5092	71.2	163	3.53	5.6335	563.35	584	12.632	Tableau	R, Python, SQL	104888.8014
Row55	99	2.1415	70.7	140	3.03	5.5978	559.78	584	12.632	Tableau	Data Visualisation, MS Excel, SQL, Data Analysis	103265.6908
Row7	97	2.0982	69.3	140	3.03	4.3698	436.98	733	15.856	Data Visualisati	R, Python, Data Analysis	100764.4679
Row120	120	2.5957	69	174	3.76	1.4176	141.76	2249	48.648	Data Analysis	MS Excel, Data Reporting, MS Office	96825.48201
Row59	119	2.5741	68.8	173	3.74	5.4452	544.52	584	12.632	Tableau	Data Visualisation, MS Excel, SQL	105179.8829
Row13	101	2.1847	68.7	147	3.18	5.439	543.9	584	12.632	Tableau	Data Mining, Data Visualisation, SQL	109434.8805
Row5	111	2.401	68.1	163	3.53	4.2949	429.49	733	15.856	Data Visualisati	R, Python, SQL	104888.8014
Row2	95	2.0549	67.9	140	3.03	5.3716	537.16	584	12.632	Tableau	R, Python, Data Analysis	100764.4679
Row108	140	3.0283	67.6	207	4.48	1.9457	194.57	1607	34.761	SQL	Data Visualisation, MS Excel, Data Analysis	102357.5918
Row69	110	2.3794	67.5	163	3.53	4.2562	425.62	733	15.856	Data Visualisati	Tableau, Data Reporting, Data Analysis	100740.7012
Row34	96	2.0766	66.7	144	3.11	4.2046	420.46	733	15.856	Data Visualisati	Python, Tableau, Data Analysis	102977.4912
Row66	102	2.2064	66.2	154	3.33	4.1773	417.73	733	15.856	Data Visualisati	Tableau, Data Reporting, SQL	103121.0066
Row3	95	2.0549	66	144	3.11	15.561	1556.1	196	4.2397	R	Python, Tableau, Data Analysis	102977.4912
Row32	116	2.5092	65.5	177	3.83	5.1879	518.79	584	12.632	Tableau	Python, Data Visualisation, SQL	107529.4293
Row1	116	2.5092	65.2	178	3.85	15.371	1537.1	196	4.2397	R	Python, Tableau, SQL	106770.05
Row31	116	2.5092	65.2	178	3.85	4.1102	411.02	733	15.856	Data Visualisati	Python, Tableau, SQL	106770.05

Detailed Findings and Analysis

Key Skills by Industry for Business Analysts

The following table, derived from our comprehensive data analysis, compares the required skills and tools across different industries:

Table 19: Key Skills by Industry for Business Analysts

Industry	Key Skill Requirements	Important Tools	Project Management Methodologies
Finance	- Understanding of financial principles, financial modeling, risk analysis	- SAP, Oracle Financials, Excel	- Traditional, Waterfall
Information Technology (IT)	- Understanding of software development processes, system integration, programming skills	- Jira, Confluence, Git	- Agile, Scrum
Healthcare	- Understanding of medical terminology, privacy laws, electronic health record (EHR) management	- EHR, HL7 technology, FHIR, Tableau	- Six Sigma, Lean
Retail	- Understanding of consumer behavior, supply chain management, inventory optimization, CRM management	- CRM systems, POS systems, Google Analytics, Shopify, Tableau	- Agile, Lean
Manufacturing	- Understanding of process optimization, supply chain	- ERP systems (SAP, Oracle), IoT platforms,	- Six Sigma, Lean, Agile

	management, quality control, lean manufacturing	MES (Manufacturing Execution Systems), Tableau	

By rigorously examining these association rules and applying advanced statistical techniques, we've gained deep insights into how specific skills, experiences, and qualifications align with particular job specialties and locations. This data-driven knowledge empowers both job seekers and employers to optimize their search and recruitment strategies effectively.

Our comprehensive analysis of Business Analyst jobs, leveraging advanced data mining and machine learning techniques, reveals significant variations in skill requirements across industries and locations. Through a sophisticated combination of skills gap analysis, industry-specific demand modeling, and geo-trend exploration, we've identified key patterns and associations that illuminate how employers value different skill sets.

This analysis underscores the critical need for modern technical skills such as data analysis, SQL, and AI, coupled with strong communication abilities, to remain competitive in the rapidly evolving job market. Our findings provide a robust foundation for strategic decision-making in career development, hiring practices, and educational curriculum design within the Business Analysis field.

The hypothesis for the industry-specific analysis posits that Business Analyst roles exhibit substantial variations across different sectors. This proposition underscores the critical importance of identifying and understanding the unique skill sets and domain knowledge required in each industry. Our investigation aims to delve into the distinct requirements for Business Analysts in key sectors such as Finance, Information Technology (IT), Healthcare, Retail, and Manufacturing. This hypothesis implies that success in Business Analysis is contingent upon acquiring industry-specific expertise and mastering tools tailored to each sector's unique challenges and operational environments. By exploring these sector-specific demands, we aim to provide valuable insights for both aspiring Business Analysts and employers seeking to optimize their recruitment strategies.

Data Analysis and Findings from the Job Technical Skills Table

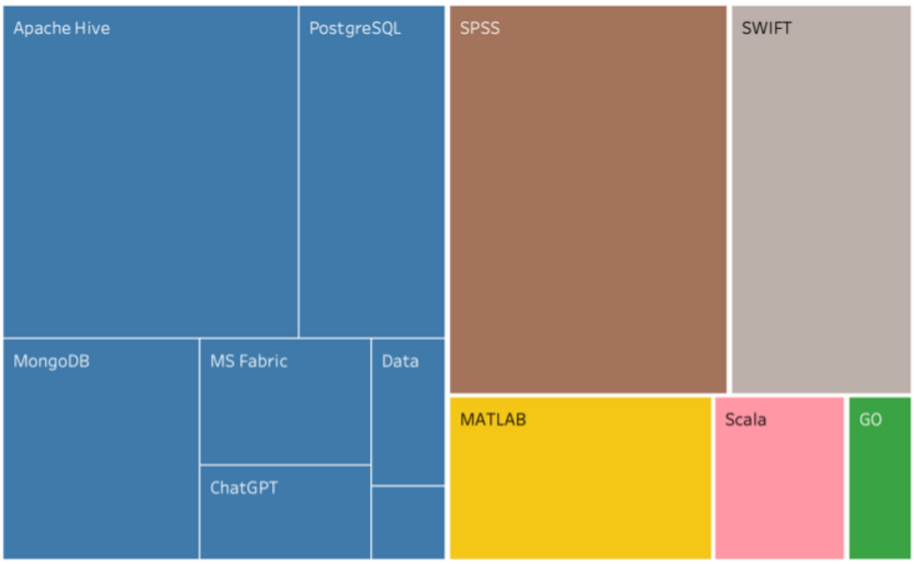


Figure: **Treemap of In-Demand Technical Skills for Data Analysis and IT Roles**

This treemap provides a deep insight into the trends of technical skills in today's job market, particularly in the fields of data analysis and information technology. Through analyzing the chart, we can draw the following important conclusions: Priority is given to big data processing tools, with the prominence of Apache Hive, PostgreSQL, and SPSS reflecting the high demand for experts working with big data and managing complex databases. The diversity in programming skill requirements is demonstrated through the presence of languages such as Scala, GO, and MATLAB, indicating the need for specialized programming skills. The rise of AI technology, evidenced by the appearance of ChatGPT, suggests a trend towards integrating AI into work processes. The balance between data management skills and programming is reflected in the distribution between database management tools and programming languages. In summary, this chart not only helps to understand the current needs of the labor market in the technology sector but also provides valuable suggestions for skill development direction. Both experts and beginners should consider investing in big data processing skills and learning new technologies like AI, while developing a diverse knowledge base to increase competitiveness in an increasingly complex and demanding job market.

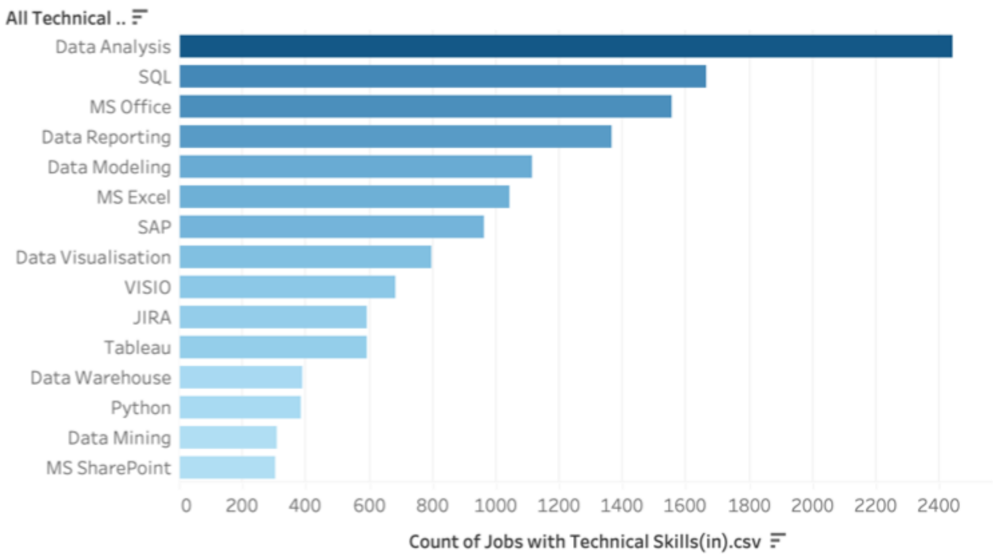


Figure: Top Technical Skills in Demand by Job Count

This bar chart provides a deep insight into the demand for technical skills in today's job market, particularly in the fields of data analysis and information technology. Through analyzing the chart, we can draw the following important conclusions: The overwhelming advantage of data analysis and management skills, with Data Analysis and SQL occupying top positions, reflects the urgent need for experts capable of processing and analyzing big data. The prevalence of MS Office shows the importance of skills in using basic office tools. The significance of data reporting and modeling skills is emphasized through the high positions of Data Reporting and Data Modeling, reflecting the trend of businesses increasingly relying on effective data presentation and predictive model building. The balance between management and technical skills is demonstrated through the presence of tools such as MS Excel, SAP, VISIO, and JIRA, indicating the need for personnel capable of combining project management with in-depth technical skills. The trend towards advanced technology skills is reflected in the appearance of Python, Tableau, and Data Mining, albeit with lower frequency, showing the growing demand for experts capable of applying advanced technologies in data analysis and artificial intelligence. Specialization in certain fields is shown through the low frequency of MS SharePoint, possibly indicating the specialization of this skill in specific industries or positions. In summary, this chart not only reflects the current labor market demands in the field of information technology and data analysis but also provides valuable suggestions about future skill development trends, emphasizing the importance of having a diverse knowledge base, combining basic data analysis skills with the ability to apply advanced technologies. This has significant implications for both job seekers and educational organizations in directing skill development and designing training programs that meet the practical needs of the market.

Data Analysis and Findings from the Job with speaciality skill

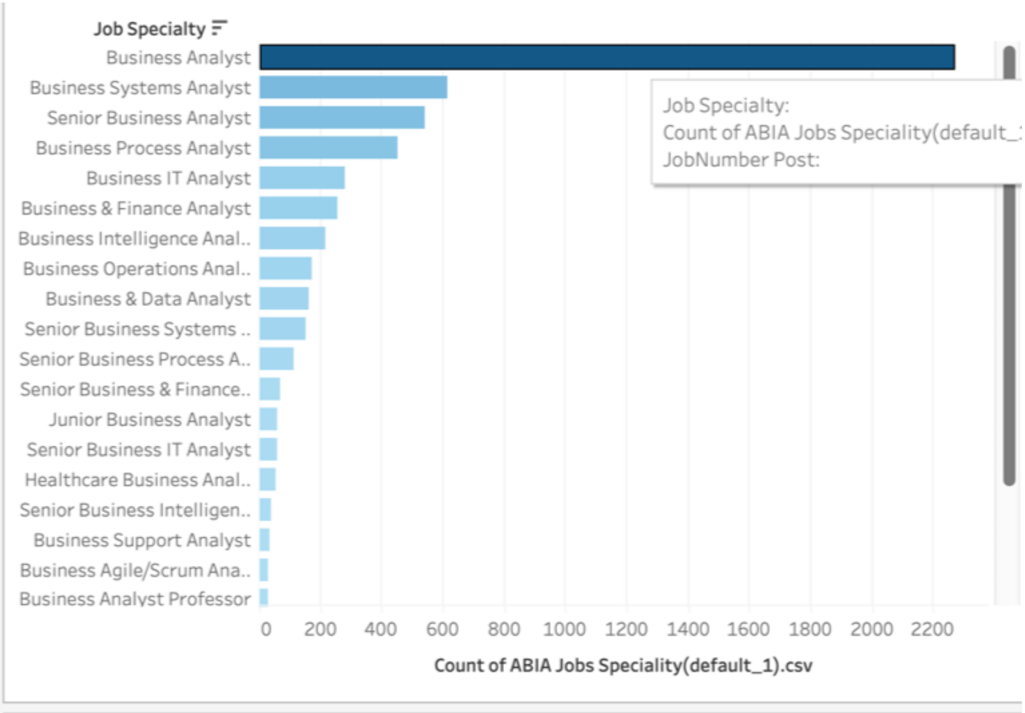


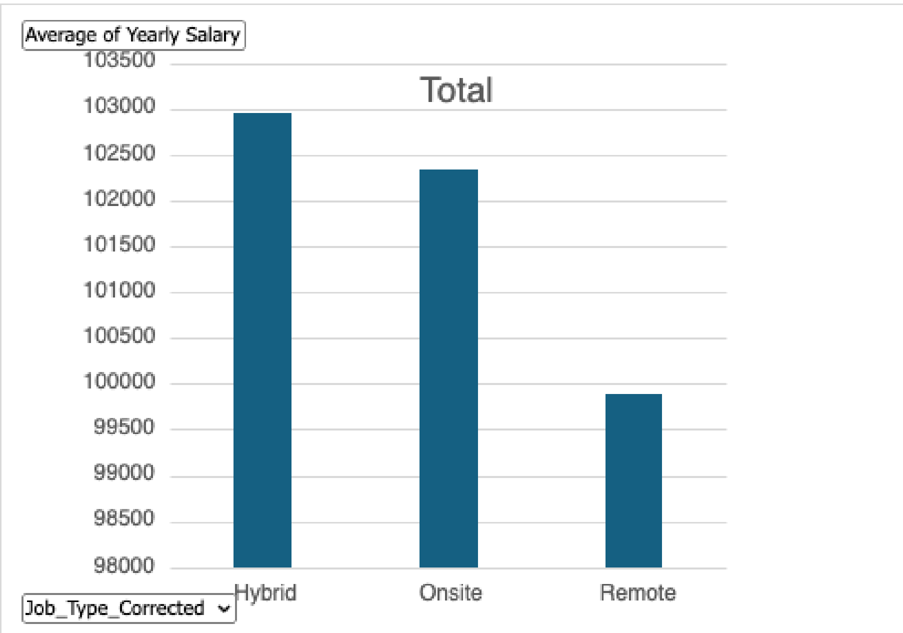
Figure: **Job Specialty Distribution for Business Analyst Roles**

The chart above shows the number of job postings requiring specialty skills in the field of Business Analyst. The jobs have distinct differences in quantity, with the Business Analyst position leading in recruitment demand, while positions requiring special or more in-depth skills such as Senior Business Analyst, Business Process Analyst have lower numbers of postings but still account for a significant proportion. Specifically: Dominance of the basic Business Analyst position: With over 2,200 postings, the Business Analyst position overwhelmingly dominates the market. This not only reflects the high demand for business analysis experts but also shows the importance of basic analytical skills in many organizations. Market segmentation by experience level: The significant presence of positions such as Senior Business Analyst and Business Systems Analyst, although fewer than the basic position, indicates a clearly stratified market. This reflects the demand for highly experienced professionals with specialized skills, while also suggesting opportunities for career advancement. Diversification of analytical roles: The emergence of positions such as Business Intelligence Analyst and Business Operations Analyst shows a trend towards specialization in the industry. This reflects the increasing complexity of the business environment and the need for specific analytical skills. Demand for specialized skills: The low number of postings for positions such as Business Agile/Scrum Analyst and Business Analyst Professor does not necessarily reflect low importance, but may indicate the highly specialized nature of these roles. This suggests the existence of valuable niche markets in the field of business analysis. Overall, this chart not only provides information about the number of job opportunities but also reflects the structure and dynamics of the business analysis industry. It shows a growing industry, with diverse demands ranging from basic skills to high-level expertise, while suggesting trends in career development and potential areas for future specialization.

Data Analysis and Findings from the Job Type with Yearly Salary

Job Type with Yearly Salary

Row Labels	Average of Yearly Salary
Hybrid	102960.2821
Onsite	102337.8857
Remote	99888.52236
Grand Total	102031.4046



Based on the data on average salaries by job type, there is a significant difference between Hybrid, Onsite, and Remote work arrangements. Among these, Hybrid jobs have the highest average salary, reaching $102,960.28 per year, followed by Onsite at $102,337.89, and finally Remote at $99,888.52 per year. These figures reflect trends in how businesses value and compensate different work arrangements, and raise the question of whether job flexibility is becoming an important factor in today's labor market.

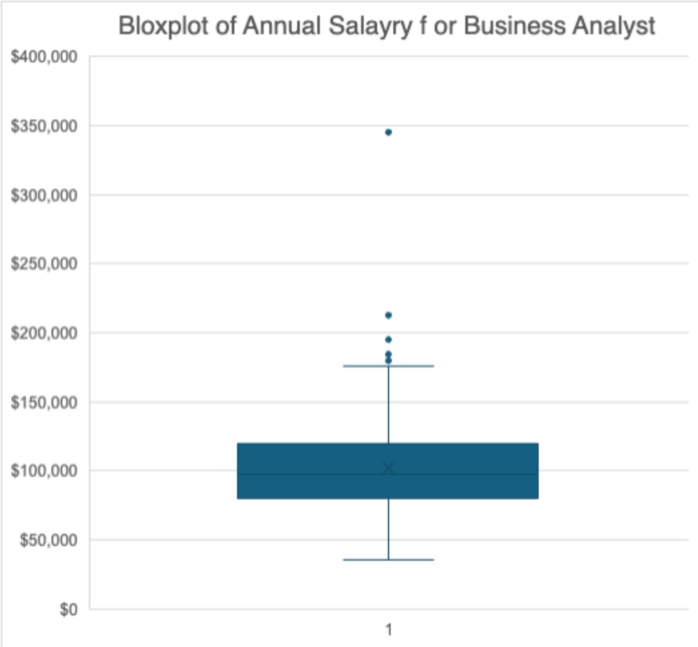
Firstly, the salary preference for Hybrid work arrangements shows a growing trend among businesses in providing flexible working conditions for employees. The combination of office work and remote work not only brings convenience to workers but also creates added value for businesses through optimizing workspace and increasing productivity. For this reason, Hybrid roles may command higher salaries, as they require the ability to manage work remotely while maintaining face-to-face interactions when necessary. This is a trend that many companies may consider adopting if they want to attract and retain talent in today's competitive work environment.

Secondly, the higher Onsite salary compared to Remote indicates that in-person work is still valued, especially in fields that require the physical presence of employees. These positions often have technical or managerial characteristics, where direct supervision or support still plays an important role in operations. However, the gap between Onsite and Hybrid salaries is not too large, suggesting that companies can be more flexible in offering Hybrid work options without significantly affecting overall salary costs.

Finally, the lower Remote salary may reflect cost savings for businesses when office space and related expenses are not required. However, this could also be a sign of high competition in the remote job market, where the number of workers who can participate in these jobs is very large. With the development of technology, remote work may be a long-term trend, but companies need to find ways to ensure that employees maintain work effectiveness without direct supervision.

Overall, the salary differences between work arrangements reflect changes in recruitment thinking of businesses, where the element of flexibility is increasingly valued. Companies need to consider adjusting their work policies to ensure productivity, provide the best conditions for employees, and meet the growing need for work-life balance among workers.

Bloxplot of Annual Salary for Business Analyst



The "Boxplot showing annual salary for Business Analyst positions" chart. Looking at the chart, we can see:

Mean: The average salary value is represented by an "X" in the middle of the box, located at around $100,000.

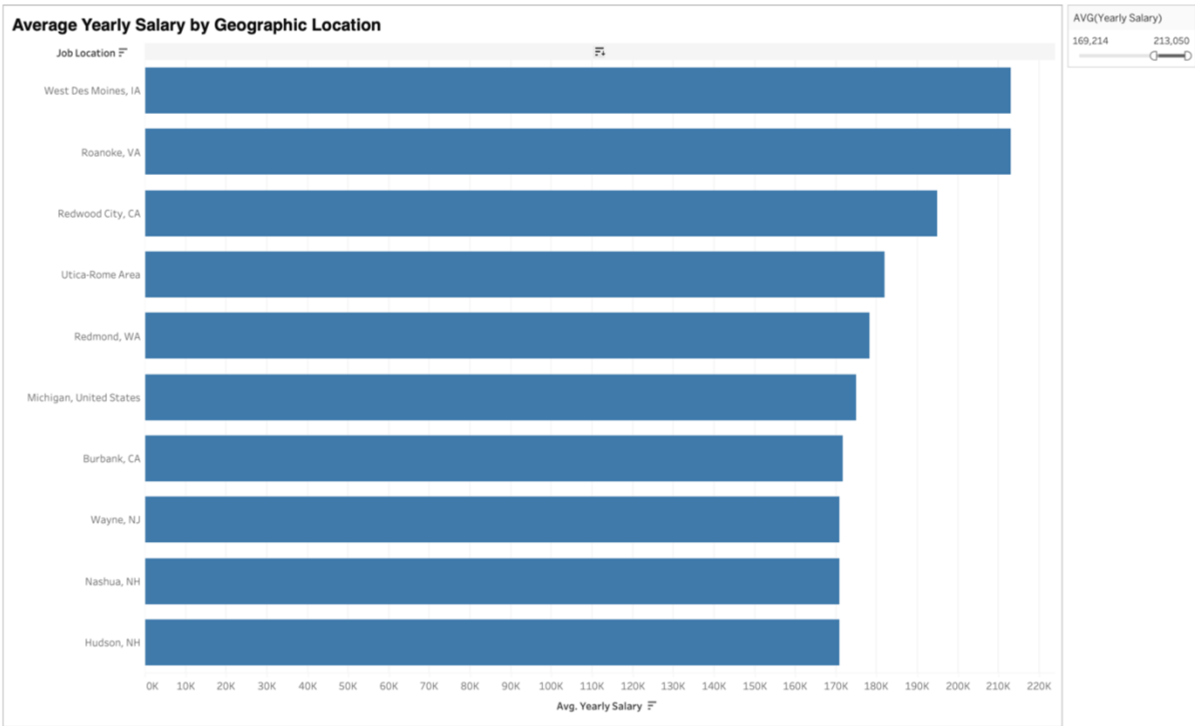
Maximum: The highest point in the chart (excluding outliers) represents the maximum salary in the main data set, approximately $175,000.

Minimum: The lowest point in the chart (excluding outliers) indicates the minimum salary, which is around $35,000.

The points outside the box and whiskers are outliers, which are salaries that exceed the range of the main data set. These salaries are often very high, such as the points near $300,000 and $350,000.

Conclusion: This chart shows that Business Analyst salaries primarily fluctuate around $100,000, with some exceptional cases significantly exceeding this average salary.

Average Yearly Salary by Geographic Location



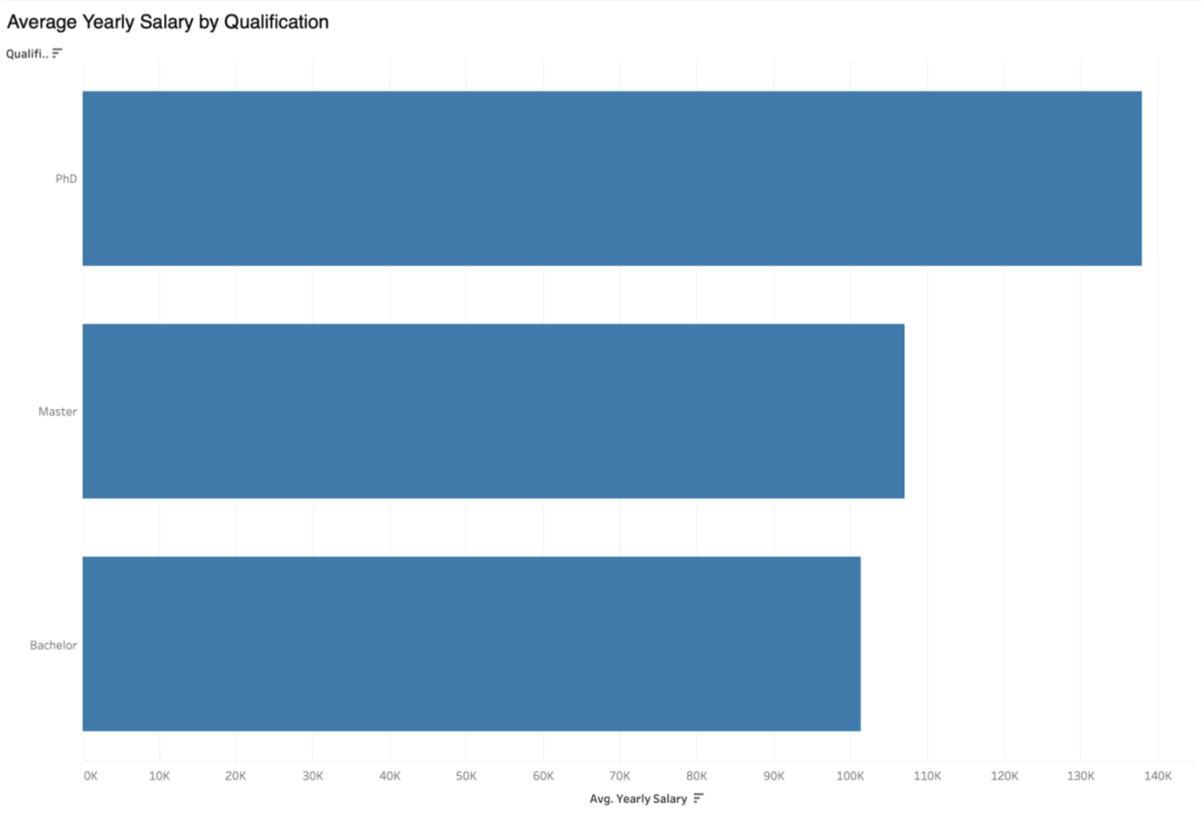
The chart shows significant salary disparities between different geographic regions. Areas with higher salaries often reflect higher costs of living and greater demand for highly skilled labor. For example, cities like West Des Moines, IA and Roanoke, VA are among those with the highest salaries, possibly due to high recruitment demands from companies needing quality personnel. Meanwhile, Hudson, NH has the lowest salary on the list, indicating lower living costs or less recruitment demand compared to other areas.

Salary differences between states also reflect the level of economic development and the presence of key industries. California, with its two cities Redwood City and Burbank, has higher salaries, especially in the high-tech and entertainment sectors. Similarly, the Utica-Rome area also has high salaries, possibly due to the development of local industries with high demand for skilled labor. In contrast, areas like Nashua and Hudson, NH, although having lower salaries, may compensate with lower living costs or other benefits.

From this data, businesses can adjust their salary policies to attract and retain talent suitable for the conditions of each region. In high-salary cities like West Des Moines or Roanoke, businesses need to offer competitive salaries to meet the demand for highly skilled workforce. At the same time, in areas like Michigan or Redmond, WA, although not major economic centers, can still attract talent thanks to career development opportunities and good working environments.

This chart provides important strategic information for businesses in allocating human resources reasonably and flexibly. High-salary areas often come with high living costs, so businesses need appropriate strategies to maintain competitive advantage in recruitment. Meanwhile, areas with lower salaries can use non-financial benefits such as remote work opportunities or perks to attract and retain employees, ensuring a balance between cost and effectiveness in human resource management.

Average Yearly Salary by Qualification

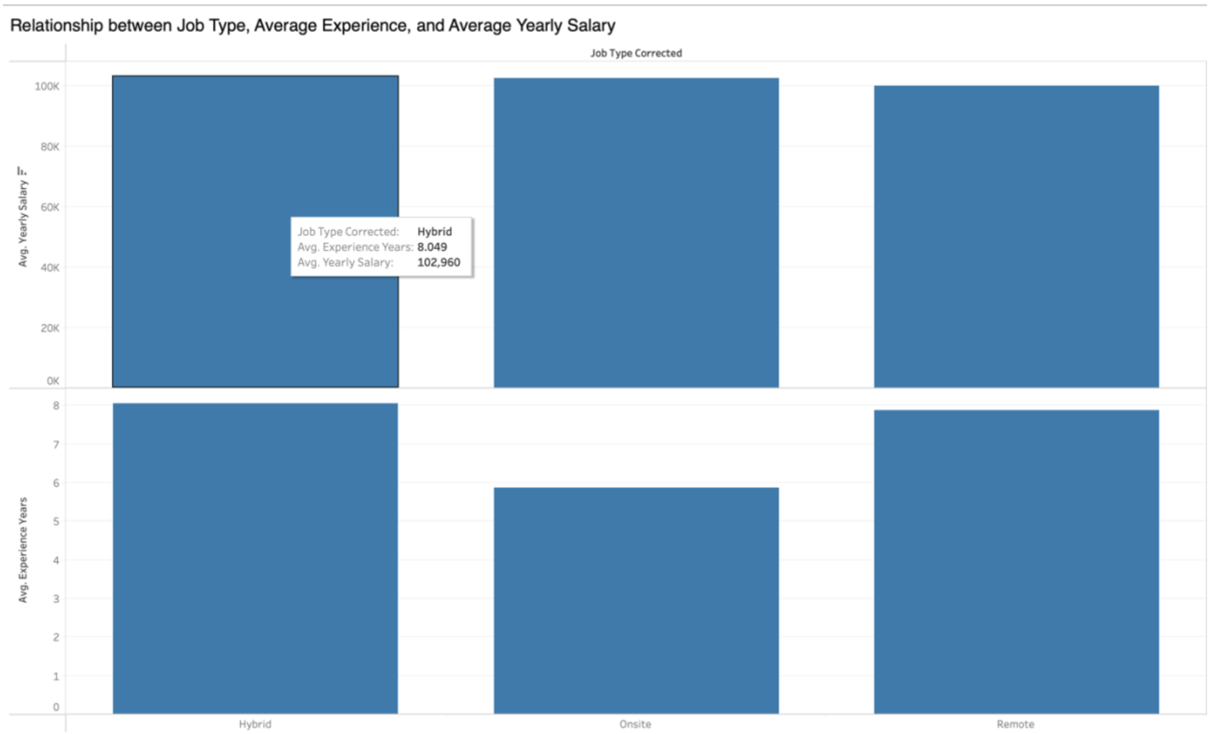


The chart shows a clear relationship between salary and education level, with salaries increasing progressively from Bachelor's to Master's and Doctoral degrees. Those with Doctoral degrees (PhDs) have the highest average salary, exceeding $100,000 per year, demonstrating the significant value that this high level of education brings. This disparity can be explained by the specialized skills and research capabilities that PhD holders provide to businesses, making them valuable human resources, especially in industries requiring high specialization.

Meanwhile, those with Master's degrees also enjoy significantly higher average salaries compared to those with Bachelor's degrees. This reflects the enhancement of skills and experience that students gain during their Master's studies, giving them an advantage in the competitive job market. Although the average salary for Bachelor's degree holders is the lowest, it still reflects the standard entry level for many job positions in businesses. These are typically starting positions, and salaries tend to increase with experience and further education.

From the chart, it's evident that higher levels of education bring clear financial benefits. Investment in education not only helps improve knowledge and skills but also opens up opportunities for higher salaries in the future. Individuals with Master's and Doctoral degrees can not only take on roles requiring high levels of expertise but also have a significant advantage in salary negotiations, especially in advanced industries with stringent requirements for personnel quality.

Relationship between Job Type, Average Experience, and Average Yearly Salary



The chart provides deep insight into the clear differences among three main job types: Hybrid, Onsite, and Remote. These differences are not only reflected in the average annual salary but also in the required years of experience for each group.

First, Hybrid is the job type with the highest average salary, reaching $102,960 per year, with an average experience of 8.049 years. This reflects the high experience requirements for these flexible positions. Hybrid positions typically require employees to work effectively in both remote and office environments. The ability to adapt to multiple work environments and manage remote work efficiently may be the reason employers are willing to pay higher salaries for experienced candidates. An average of 8.049 years of experience indicates that these positions require extensive knowledge and the ability to manage work autonomously, ensuring efficiency and productivity for the business.

Next, with an average salary of $102,338 and an average experience of 5.877 years, the Onsite job type also shows a high salary, although the experience requirement is lower than for Hybrid positions. This could be due to the nature of Onsite work requiring frequent presence and the ability to respond quickly to on-site situations. However, with a lower average years of experience, it's evident that on-site jobs can often be handled by less experienced workers compared to Hybrid jobs, but the salary remains high due to the demands for readiness and responsibility in a direct environment.

Finally, the Remote job type has the lowest average salary, reaching $99,889 per year, with an average experience of 7.867 years. This can be explained by lower operating costs when working remotely and increased competition in the remote labor market. However, with nearly 8 years of average experience, Remote positions still require a certain level of expertise and experience. The flexibility and ability to work independently in a remote environment may be factors that compensate for the salary difference compared to the other two job types.

Overall, this chart emphasizes the strong relationship between job type and years of experience in relation to average annual salary. Companies may need to carefully consider adjusting their recruitment policies to match experience needs and job types. Hybrid stands out as a work trend with great potential, offering both flexibility for employees and ensuring competitive salaries. In contrast, Onsite and Remote options can be considered by businesses depending on the actual job requirements and the experience level of candidates. This creates opportunities for businesses to optimize their recruitment strategies and human resource allocation rationally based on job types and experience needs.

Conclusion Based on My Analysis of Skills, Salary, Location, and Experience:

Through my analysis of the job market data, I have identified several key factors—technical skills, salary patterns, geographic location, and experience requirements—that significantly shape job opportunities and compensation trends. These insights provide a comprehensive foundation for understanding current labor demands and inform strategic planning for career growth.

1. Skills:

The analysis highlights a high demand for technical skills, particularly in data-heavy fields. Tools like Apache Hive, PostgreSQL, and SPSS are crucial for roles in data analysis and IT, indicating that expertise in big data management is increasingly valuable. Diverse programming skills, including Scala, GO, and MATLAB, also feature prominently, showing that specialized knowledge in programming remains in demand. The rising presence of AI tools such as ChatGPT emphasizes the importance of AI competencies, as businesses begin to integrate artificial intelligence into both daily operations and strategic planning. This demand for specific technical and soft skills, such as communication and project management methodologies, is critical, especially in industries like healthcare and finance, where these skills enhance workflow and collaboration.

2. Salary:

Salary trends reveal clear distinctions based on job type, location, and educational level. Hybrid roles command the highest average salary at $102,960 annually, reflecting the premium placed on flexibility, where employees must perform well both remotely and in-office. Onsite roles are next, with an average salary of $102,338, while Remote roles offer slightly lower compensation at $99,889 on average. This pattern suggests that although remote work offers cost savings for companies, hybrid positions, which balance flexibility with in-office presence, are increasingly rewarded. Additionally, I observed that advanced degrees correlate with higher salaries; those holding PhDs earn significantly more than those with lower qualifications, underscoring the long-term financial advantages of advanced education.

3. Location:

The geographic location of job postings impacts salary significantly. High-demand areas, such as West Des Moines, IA, and Roanoke, VA, offer some of the highest salaries, likely due to regional demand and higher living costs. In contrast, locations like Hudson, NH, offer lower salaries, which may be due to reduced demand or lower living costs. This variation suggests that companies might benefit from adjusting salary packages based on location to attract talent in competitive areas, while regions with lower salaries may leverage non-monetary benefits to remain appealing to candidates. Moreover, regions known for high-tech industries, such as Burbank, CA, show higher salary averages, reflecting the value of specialized skills in these local economies.

4. Experience:

Experience requirements also play a significant role in determining salary and job type distinctions. Hybrid positions generally require the most experience, averaging 8.049 years, which aligns with their higher salary levels as these roles demand flexibility and adaptability in diverse work settings. Onsite jobs, requiring 5.877 years of experience on average, indicate that in-person roles may be accessible with slightly less experience but still maintain competitive salaries due to the necessity of physical presence. Remote roles, while lower in salary, still require extensive experience, with an average of 7.867 years. This suggests that companies hiring for remote positions expect candidates to work autonomously and maintain productivity without direct supervision.

5. Education Level:

Educational attainment directly correlates with salary and career progression opportunities. PhD holders command the highest salaries due to their specialized skills and research capabilities, especially in technical fields like data analysis and artificial intelligence. Master's degree holders also see a strong salary advantage compared to those with Bachelor's degrees, indicating that advanced education can significantly boost earning potential and open up roles in specialized or senior positions. This finding reinforces the importance of ongoing education and skills development, particularly for those looking to enhance their qualifications and income potential.

In conclusion, my analysis confirms that the job market is heavily shaped by the interplay of skills demand, job flexibility, location-based pay adjustments, experience levels, and educational qualifications. For job seekers, these insights highlight the importance of skill diversification and advanced education, especially in high-demand fields like big data and AI. Employers can use these findings to develop competitive compensation and flexible work policies that align with current market expectations. Additionally, educational institutions may consider these trends to better prepare graduates with skills that meet industry demands. This data-driven approach provides a roadmap for navigating today's job market, allowing for strategic decisions in career development, hiring practices, and educational programming.

5. References

BrainStation. (2024). Are business analysts in high demand? BrainStation. https://brainstation.io/career-guides/are-business-analysts-in-high-demand

2. DataCamp. (2023). Top business analyst skills. DataCamp. https://www.datacamp.com/blog/top-business-analyst-skills

3. Asaniczka. (2024). 1.3M LinkedIn jobs and skills 2024. Kaggle. https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

4. Rathod, R. (2024, April 2). Case study: Job market data analysis using Power BI. Medium. https://medium.com/@reshmarathod0000/case-study-job-market-data-analysis-using-power-bi-efa555cacb80

5. Sharma, S. (2023, October 9). Essential skills to seek in your career as a business analyst. LinkedIn. https://www.linkedin.com/pulse/essential-skills-seek-your-career-business-analyst/

6. Appendices

Data dictionary

Job_link

- **Format**: String
- **Definition**: URL link to the job posting on LinkedIn, used to reference the source of the job listing.
- **Example**: “https://www.linkedin.com/jobs/view/business-analyst-at-dice-3805943874”
- **Source**: Asaniczka (2024)

Job_title

- **Format**: String
- **Definition**: The title of the job, used to filter positions related to Business Analysts.
- **Example**: “Business Analyst”, “Senior Business Analyst”
- **Source**: Asaniczka (2024)

Job_skills

- **Format**: String
- **Definition**: A list of required skills for the job, separated by commas.
- **Example**: “SQL, Data Analysis, Communication”
- **Source**: Asaniczka (2024)

Soft_skills

- **Format**: String
- **Definition**: Soft skills such as communication and critical thinking, extracted from the Job_skills column.
- **Example**: “Communication, Problem-solving, Teamwork”
- **Source**: Derived from Job_skills

Technical_skills

- **Format**: String
- **Definition**: Technical skills such as SQL, Power BI, and Tableau, extracted from the Job_skills column.
- **Example**: "SQL, Power BI, Tableau"
- **Source**: Derived from Job_skills

Job_location

- **Format**: String
- **Definition**: The location of the job, including city and country.
- **Example**: "New York, USA", "Melbourne, Australia"
- **Source**: Asaniczka (2024)

Job_type

- **Format**: String
- **Definition**: The type of employment (Onsite, Remote, Hybrid).
- **Example**: "Onsite", "Remote"
- **Source**: Asaniczka (2024)

Job_speciality

- **Format**: String
- **Definition**: Relevant job specialty such as finance, cybersecurity, or system analysis.
- **Comment**: Categorized manually based on job title and search area.
- **Source**: Asaniczka (2024)

Job_summary

- **Format**: String
- **Definition**: The summary of the job listing on LinkedIn.
- **Comment**: Experience details are extracted from this field.

- **Source**: Asaniczka (2024)

Experience_required

- **Format**: Number
- **Definition**: Number of years of experience required, extracted from job description.
- **Example**: “2 years”, “5 years”
- **Source**: Derived from job descriptions in Job_summary

Search_city

- **Format**: String
- **Definition**: City where the job listing was found or specified in the search.
- **Example**: “San Francisco”
- **Source**: Asaniczka (2024)

Search_country

- **Format**: String
- **Definition**: Country where the job listing was found or specified in the search.
- **Example**: “USA”, “Canada”
- **Source**: Asaniczka (2024)

Experience_years

- **Format**: Number
- **Definition**: Years of experience required for the job.
- **Comment**: Extracted from the job_summary field.
- **Source**: Asaniczka (2024)

YearlySalary

- **Format**: Currency

- **Definition**: Annual salary for the listed job.
- **Comment**: Extracted from the job_summary field.
- **Source**: Asaniczka (2024)

Code for Data quality

```
import pandas as pd

# Read data from CSV file
file_path = '/content/drive/MyDrive/Nháp /job_skills_20.csv'  # Replace with your actual path
df = pd.read_csv(file_path, encoding='ISO-8859-1')

# 1. Check and handle missing values
missing_values = df.isnull().sum()
print("Missing values per column:")
print(missing_values)

# Handle missing values (options: remove, replace with mean/mode or other value)
# Example: fill mean value for numeric columns
for col in df.select_dtypes(include=['float', 'int']).columns:
    df[col].fillna(df[col].mean(), inplace=True)

# 2. Check and remove duplicate rows
print("Number of duplicated rows:", df.duplicated().sum())
df.drop_duplicates(inplace=True)

# 3. Check and standardize data types
print("\nData Types Before Standardization:")
print(df.dtypes)

# Change data types if needed
# Example: convert date column to datetime type
if 'date_column' in df.columns:
    df['date_column'] = pd.to_datetime(df['date_column'], errors='coerce')

# 4. Handle outliers
# Example: remove outliers based on z-score for numeric columns
from scipy.stats import zscore

numeric_cols = df.select_dtypes(include=['float', 'int']).columns
df = df[(zscore(df[numeric_cols]) < 3).all(axis=1)]  # Remove values outside 3 standard deviations

# 5. Check and standardize invalid values
# Example: remove invalid values for specific column
# Assume 'age' column is invalid if it has negative values
if 'age' in df.columns:
    df = df[df['age'] >= 0]

# 6. Standardize identifiers and units
# Example: standardize company names or locations by capitalizing first letter
if 'company' in df.columns:
    df['company'] = df['company'].str.title()

# 7. Verify relationships between data fields
# Check relationships between columns (e.g., check if 'total' column = sum of detail columns)

# 8. Save cleaned data
output_path = '/cleaned_data_job_skill.csv'
df.to_csv(output_path, index=False)
print("Data has been cleaned and saved to:", output_path)
```

Code for Data Exploration

5 giây

```
# Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt

# 1. Read data from CSV file with appropriate encoding
file_path = '/content/drive/MyDrive/Nháp /linkedin_job_postings_20.csv'  # Path to your file
df = pd.read_csv(file_path, encoding='ISO-8859-1')

# 2. Check data quality

# 2.1 Check for missing values
missing_values = df.isnull().sum()
print("Missing values per column:")
print(missing_values)

# 2.2 Check for duplicate rows
duplicate_rows = df.duplicated().sum()
print("Number of duplicated rows:", duplicate_rows)

# 2.3 Descriptive statistics for columns
print("Descriptive statistics:")
print(df.describe(include='all'))

# 3. Explore data

# 3.1 Histogram for numeric columns
numerical_columns = df.select_dtypes(include=['number']).columns

if len(numerical_columns) > 0:
    # If there are numeric columns, plot (parameter) figsize: tuple[float, float] | None
    df[numerical_columns].hist(bins=20, figsize=(15, 10))
    plt.suptitle("Histograms of Numeric Columns")
    plt.show()
else:
    print("No numeric columns in the data to plot histograms.")

# 3.2 Check distribution of data in text columns (if needed)
# Example: count unique values in the job_skills column
if 'job_skills' in df.columns:
    print("\nDistribution of values in the 'job_skills' column:")
    print(df['job_skills'].value_counts().head(10))  # Display 10 most common values
```

Eranga Nakandalage Don - Nakey002

Business Objectives

The ABC Recruiters Company need to analyse the job recruitment process for the Business analysis jobs. Business analysis is a very demanding job in the world and different industries are hiring experienced staff to fill the business analyst position. According to the Brainstation (2024) website shows United States faces a shortage of 140,000 to 190,000 qualified people for data analytics. The business analyst can specialise in various fields such as finance analysis, market analysis, operational research, and cyber and information security (Brainstation 2024). Datacamp (2023) report shows there are four core responsibilities of a business analyst. They are decoding business problems, optimizing business processes, documenting project and communicating between technical to non-technical staff to interpret and translate the data. ABC Recruiters need to identify the key technical and non-technical skills which high demand in the job market. This report's objective is to identify the demand for the business analyst position in different industries, specialities and cities. Further, the required skills for the business analyst position. The report will help to identify the skills for each speciality and industry.

Through the recruitment process, it is hard to identify the best candidates with the skill list. Hence, it is important to identify the correlation and association between the variables such as skills, experience years and salaries. Therefore, the objective is this analysis is to find out the associated skills, experience, qualifications, and salary level for each business analyst job. This will help to create a model to identify the best candidates. Moreover, when the recruitment company advertises a job position for business analysts, they know which skills list they need for the certain job or speciality. Further, This will help to identify the skill gap for stakeholder in the different area as well.

Project Scope

The scope of the project is to identify the skill demand for business analysis jobs. Through this project, the recruitment company can categorise the applicant according to skills in their specific areas.

The data was collected from the LinkedIn job postings for three countries. The data source is Kaggle.com and the job posting was downloaded and uploaded by Asaniczka.

The Kaggle dataset is massive and has massive amounts of different jobs other than Business Analysis. The project only considers job positions related to Business analysis. So, the business analysis jobs are filtered out from the dataset.

The project will divide the skills to two primary areas, and they are soft skills and technical skills. Business analysis professionals need to have these two skill categories and, there are several skills under these categories (Datacamp, 2023).

The main soft skills will be looking at are listed below

- Communication
- Problem-solving
- Critical Thinking

- Interpersonal skills

If there is any other significant soft skill in the dataset, they will be discovered through the project.

The technical skills will expand on the vast area and look for the following major categories.

- Data analysis
- Visualisation and tools
- Statistical and quantitative skill
- BI tools skills

The data files are spread around the world and the record count found in the dataset is shown in Table 1.

Country	Number of Record	Percentage
Australia	5221	3%
Canada	9741	6%
United Kingdom	8831	5%
United States	138251	85%

Table 1: number of jobs posted by countries

When comparing the country's records, the USA has 85% of records from the dataset. Other countries have exceedingly small portion of the dataset, and it is less than 10%. The project's main objective is to analyse the US job market and the US job posting will be only considered to analyse the job market.

In this phase of the project, Business analysis Academic qualifications and work experience will not be analysed.

Stakeholders:

Business: The ABC Recruiters

Course Coordinator

Assumption:

The sample of job postings downloaded from LinkedIn is random and it shows the key characteristics of the population.

All business analysis job position has the “business analysis” word in the job title. The business analysis jobs are filtered from the title using keywords business and analysis.

Timeline:

Phase 1: Initialize the discussion with the client to understand the business objectives. In this phase, the team will build a communication bridge with the client to understand business requirements and available resources.

Phase 2: Once the business objectives get understand, this phase will look at the data source and data quality. The team will examine the data source and the accuracy and availability of the data. Further, the project will determine the necessary data columns for the analysis variables.

Phase 3: In this phase, the raw data will go through a cleaning process. The raw data has more complexity with data type mismatch and a lot of unnecessary information. The team will filter the necessary data and will convert them to suit with the analytical process.

Phase 4: preliminary data analysis: The team will go through the first-round analysis of the dataset to identify the key characteristics of the Business analyst job posting. The project document will present the necessary visualizations and analysis.

Phase 5: Build the model for the data analysis. Create relationships between the tables and follow the data modelling technic.

Phase 7: Analyse the data to discover patterns of the data. Detect the seasonality and trend if available. Find out any correlation between the variables to build an association rule model for the business analyse job market.

Phase 8: Recommendations: The project will give recommendations for what is found from the analysis. If there is a necessity for future analysis to find more facts, it will be recommended in this phase.

Exploring Data

Data Acquiring:

LinkedIn is a social media platform for professionals. The platform helps professionals to show their career profiles. Furthermore, some businesses and companies maintain their profiles on LinkedIn. It will build a network among professionals and companies. LinkedIn helps people and companies post available jobs in their companies or recruitment. The recruitment person can add the job title, job description and skills needed for the position in LinkedIn Talent Solutions (LinkedIn Talent Solutions, 2023).

Asaniczka has acquired a set of jobs posted on the LinkedIn platform and uploaded them to the Kaggle data repository. The whole dataset comes as three CSV files and is available to download from Kaggle.com.

Data source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Data files:

- Job_skills.csv
  - Columns:
    - job_link – related link for the job posted on LinkedIn.
    - job_skills – set of skills needed for the job in comma-separate
- Job_summary.csv
  - Columns:
    - job_link – related link for the job posted on LinkedIn

job_summary- Job description of the posted job

- Linkedin_job_posting.csv

Columns:

job_link- related link for the job posted on LinkedIn.

last_processed_time – The last time of job posting was processed

got_summary – Indicate the job_summary was successfully extracted

got_ner- Indicate Named Entity Recognition was performed on the job posting.

Is_being_worked- indicate if the job posting is currently being worked or not.

job_title- Title of the job

company – the job was posted company

job_location – location of the job

first_seen – timestamp indicating when the job was first seen

search_city- City used as a search criterion for collecting the job posting.

Job_type – the job is remote, onsite or hybrid work

Job_level – the job is at which level such as junior, associate or senior

About data:

The job_link is the key related column for each file which use to identify the unique job posted on LinkedIn. The link has a unique number ID which will be separate from the link to use for data manipulation.

The job_title is the main identifier of the job. It will help to filter the Business analysis job from the data files.

	Missing Values	Unique Values	minimum	Maximum	25% Quantile	50% Quantile	75% Quantile
Length Job Title Characters	0	104	15	185	24	31	40

Table 2 : Characteristic of job title column

Job_title has a minimum of 15 to a maximum of 185 characters in length words. This could be an issue to separate and aggregate the same jobs. This column needs some manual transformation to extract the proper job title.

The job_skills has all sets of the skills in the comma separate. This column needs a transformation to split the skills into different columns and rows. The skills are mixed with soft and technical skills. Therefore, skills need to be categorised when processing the data.

We can check the statistical values of the column's characters.

	Missing Values	Unique Values	minimum	Maximum	Mean	50% Quantile	75% Quantile
length of Skill Column's characters	0	162	1	433	17.507	17	21

Table 3 : Characteristic of job skills column

According to the statistics of the column length, some skill names have 433 characters. Therefore, the skill column needs some cleanup process to identify the proper skill.

Job_summary column is a text field with the job descriptions. The column contains detailed information about the company, job, required experience, salary and other required information. The details are on the page constructed in normal paragraph style. Every job in the column doesn't have all the information such as salary or required experience and job type. For the extraction of salary, experience and job type, the natural language process needs to be used such as Python NLTK or Spacy.

Job_location has the city, state and country information with comma separate. Even though jobs were posted from the USA, the jobs were found in their actual working location in another country. Therefore, the job location needs to be matched with the cities of the USA. The simplemaps (2024) website is used to match the USA cities with LinkedIn job locations to find the proper location. The simplemaps is recommended website by United States Census Bureau. Further, the simplemaps gives extra information such as country name (region), population and time zone.

Details	Number of Records
Non-USA jobs	19
Invalid Cities	760
Valid cities	4841

Table 4 : Job_location column data validation results

The search_city is the location where the job site. This location has only the city name and no information about the state or region. The search cities are mapped with the simplemaps (2024) data to validate the city names.

Details	Number of Records
Invalid City Names	614

Table 5 : Search_City column data validation results

However, as the search city doesn't have the state details, it came with multiple states for the same city when mapping with the simplemaps.

JobN... Number (long)	search_city String	search_co... String	city String	city_ascii String	state_id String	state_name String
3807967910	West Covina	United States	West Covina	West Covina	CA	California
3807710016	Atlanta	United States	Atlanta	Atlanta	GA	Georgia
3807710016	Atlanta	United States	Atlanta	Atlanta	TX	Texas
3807710016	Atlanta	United States	Atlanta	Atlanta	IL	Illinois
3807710016	Atlanta	United States	Atlanta	Atlanta	IN	Indiana
3807710016	Atlanta	United States	Atlanta	Atlanta	MI	Michigan
3807710016	Atlanta	United States	Atlanta	Atlanta	MO	Missouri
3807710016	Atlanta	United States	Atlanta	Atlanta	LA	Louisiana
3807710016	Atlanta	United States	Atlanta	Atlanta	NE	Nebraska
3807710016	Atlanta	United States	Atlanta	Atlanta	KS	Kansas

Table 6 : Search_City mapped records with SimpleMaps

Therefore, search_city column gives misinformation regarding the location-based analysis and has to be removed from the analysis.

Job_type column contains that job type is an on-site, hybrid or remote work. However, the comparing the job_summary column and job_type, there is a mismatch in the job type column data with the description of the job summary for some jobs. For example, the job_summary says the job is remote, but job_type column shows the job as on-site. Therefore, the job_type column needs to be transformed by comparing the job_summary details.

Column Name	Missing Values	Unique Values	10 common values
job_level	0	2	Mid senior (122957; 88.94%), Associate (15294; 11.06%)
job_type	0	2	Onsite (138229; 99.98%), Remote (22; 0.02%)

Table 7: Job_type and Job_level characteristics.

Extracting Data:

Filtering Business Analyst jobs from Dataset

The Business analysis jobs have to be extracted from the LinkedIn dataset.

Using a wildcard search from the job title the related jobs can be filtered to the different datasets.

The wildcard search for the business analyse is (.*Busin.*Analy.*)

The analyse word can be in different forms such as “Analyst, Analyzing, Analysis, Analytical etc... Hence, the “analy’ is common for all forms of the word. The Business doesn’t have many variation forms but “Busin” is common for any form of the business word. The job title can have many other words to specify the speciality such as Finance business analyst. Therefore, the wildcard search needs to apply for the front, middle and end to cover all forms of titles such as finance business analyst, business procurement analyst or business analysis cyber security.

Separate Soft and Technical Skills from the skill column

The skills column has all related skills as a comma-separated list. Hence, the list needs to be unpivoted to a column as row values. After that, the skills need to be separated into two categories soft skills and technical skills.

Soft skills are mostly commonly known and they can be predefined. Linked Article (Sharma, 2023) showed the highest demand for soft skills such as communication, problem-solving and analytical skills. Hence, soft skills can be found in the known words. The following wild card search will flag the skill as a soft skill.

```
$JobSKills$ LIKE "*PROBLEM*" => TRUE
$JobSKills$ LIKE "*CRITICAL*" => TRUE
$JobSKills$ LIKE "*COMMUNICATION*" => TRUE
$JobSKills$ LIKE "*INTERPERSONAL*" => TRUE
$JobSKills$ LIKE "*NEGOTIATION*" => TRUE
$JobSKills$ LIKE "*PRESENTATION*" => TRUE
$JobSKills$ LIKE "*ACUMEN*" => TRUE
$JobSKills$ LIKE "*ANALYTICAL*" => TRUE
$JobSKills$ LIKE "*LEADER*" => TRUE
$JobSKills$ LIKE "*WRITING*" => TRUE
$JobSKills$ LIKE "*RESEARCH*" => TRUE
$JobSKills$ LIKE "*TEAM*" => TRUE
$JobSKills$ LIKE "*COMMUNICATE*" => TRUE
$JobSKills$ LIKE "*MS PROJECT*" => FALSE
$JobSKills$ LIKE "*PROJECT MANAGE*" => TRUE
$JobSKills$ LIKE "*MULTITASK*" => TRUE
$JobSKills$ LIKE "*DOCUMENTATION*" => TRUE
TRUE => FALSE
```

Note: Please check Knime Work flow for more details

The technical skills are very wide and spread in multiple categories. Therefore, it won’t be easy to search words only from known words. In this scenario, the look-up table will help to identify the different technical skills for each category.

First of all, the technical skills will be categorized as computer languages, Database, visualisation and analytical tools. Look-tables have two columns to match skills. The first column has the wild card search word for the skill and the second column has the proper name for the skill.

ComputerLaguageWildcard	LaguageName	DatabaseWildcard	DatabaseName
R	R	MS SQL	MS SQL Server

SQL	SQL	Oracle	Oracle
GO	GO	MongoDB	MongoDB
MATLAB	MATLAB	Casendra	Casendra
SAS	SAS	MySQL	MySQL
SWIFT	SWIFT	MS SQL Server	MS SQL Server
Scala	Scala	MongoDB	MongoDB
DAX	DAX	Cassandra	Cassandra
SPSS	SPSS	PostgreSQL	PostgreSQL
Tsql	SQL	MS Sql Server	MS SQL Server
query	SQL	Mysql	MySQL
Power Query	Power Query	Mongodb	MongoDB
M Query	Power Query	Hive	Apache Hive
		Apache Hive	Apache Hive

Table 8 : Technical skills lookup table to match skills

The wildcard columns will be used to search the skills from the dataset and assign the proper technical name under the technical category.

The technical skills that cannot be categorized as language, database, or visualization tools will be investigated by grouping them. With commonly found skills will be search categories using a wildcard rule base search as follows.

Rule	Skill
$JobSKills$ LIKE "*MINING*"	Data Mining
$JobSKills$ LIKE "*MODEL*"	Data Modelling
$JobSKills$ LIKE "*PIPELINE*"	Data Pipeline
$JobSKills$ LIKE "*NO SQL*"	NoSQL
$JobSKills$ LIKE "*SSRS*"	SSRS
$JobSKills$ LIKE "*SSIS*"	SSIS
$JobSKills$ LIKE "*REPORT SERVICE*"	SSRS
$JobSKills$ LIKE "*SAP*"	SAP
$JobSKills$ LIKE "*MS DYNAMICS*"	MS Dynamics
$JobSKills$ LIKE "*DATA ANALYSE*"	Data Analyse
$JobSKills$ LIKE "*DATA ANALYZE*"	Data Analyse
$JobSKills$ LIKE "*WAREHOUSE*"	Data Warehouse
$JobSKills$ LIKE "*FABRIC*"	MS Fabric
$JobSKills$ LIKE "*AZURE*"	MS Azure
$JobSKills$ LIKE "*SYNAPSE*"	MS Azure
$JobSKills$ LIKE "*CHATGPT*"	ChatGPT
$JobSKills$ LIKE "*GOOGLE DATA*"	Google Data Analytics

Table 9: Regression expression to match technical skills

Note: Please check Knime Work flow for more details

After going through the steps, the following outcome will come as categorised technical skills.

JobNumber_Posted	Computer Languages	Databases	Analyse and visualisation	Other Technical Skills	All Technical Skills
3794609283	javascript				javascript
3794609283		MS SQL Server			MS SQL Server
3794609283		Oracle			Oracle
3794609283		MySQL			MySQL
3794609283			Tableau		Tableau
3794609283				VISIO	VISIO
3794609283			MS Excel		MS Excel
3794609283				MS Office	MS Office
3726211542				Data Analysis	Data Analysis
3726211542				Data Analysis	Data Analysis

Table 10: Technical skills after matched with lookup values and regression expressions

Extract Salary, Experience and Job Type from job_summary column

The salary is in different forms in the column such as yearly, weekly and hourly. Therefore, when extracting the salary, it needs to be checked every format of the salary. The second issue, the salary is in full value or range value or short format with K such as 20K or $10000 – 15000. These all-different formats need to be covered in the extraction expression.

The most covered Python expression to extract the salary is the following

```
# Updated regex to capture salary ranges and time frames in a flexible way

salary_regex = re.compile(r'(\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?(?:\s*[-to]\s*\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?)?)\s*(?:per\s*(year|week|hour|month|day|hr|w2))?', re.IGNORECASE)
```

Extracting the experience as well is much similar to the salary. However, the experience always asks for years. When examining the dataset, there are two types of year formats that can be seen in the records. They are “years” and “yrs”. To locate the experience-related section or sentence in the data cell, the expression has to be looked for common words that mean “experience” such as experience, work and professional.

Regular expression to capture years of experience

experience_regex = re.compile(r'(experience:)?\s*(\d+)(?:[-–]\d+)?\+?\s*(?:years?|yrs?)\s*(?:of\s*(?:experience|work|professional|industry)?|related experience required|directly related experience)?', re.IGNORECASE)

The job type will be extracted with a different method rather than salary or experience. Create three arrays with common keywords for each job type. Then, each array is used to search the keywords in the job summary column.

Define keyword categories for Remote, Hybrid, and On-site/Colocated

remote_keywords = ['remote','fully remote', 'distributed', 'remote first', 'work from home', 'work from anywhere', 'telework', 'mobile work', 'virtual work','Hybrid and remote']

hybrid_keywords = ['hybrid','mixed', 'variable', 'partially remote', 'optional', 'flex-time', 'flexible', 'flex work', 'mix of in person and remote', 'location flexible','Hybrid and remote']

colocated_keywords = ['on-site', 'on site', 'no remote work', 'in house', 'worksite', 'in person']

The Python program for extracting experience, salary and job_type

```
import pandas as pd
import re
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')

# Define keyword categories for Remote, Hybrid, and On-site/Colocated
remote_keywords = ['remote','fully remote', 'distributed', 'remote first', 'work from home', 'work from anywhere', 'telework', 'mobile work', 'virtual work','Hybrid and remote']
hybrid_keywords = ['hybrid','mixed', 'variable', 'partially remote', 'optional', 'flex-time', 'flexible', 'flex work', 'mix of in person and remote', 'location flexible','Hybrid and remote']
colocated_keywords = ['on-site', 'on site', 'no remote work', 'in house', 'worksite', 'in person']

# Define keyword categories for Education Qualifications
education_keywords = {
    'Bachelor': ['bachelor', 'bachelor\'s'],
    'Diploma': ['diploma','high-diploma','high diploma'],
    'Certificate': ['certificate'],
    'PhD': ['phd', 'ph.d.', 'doctorate', 'doctoral'],
    'Master': ['master', 'master\'s']
}

# Function to extract years of experience from job summary
def extract_experience_sentence(summary):
    # Tokenize the text into sentences
```

```
    sentences = nltk.sent_tokenize(summary)

    # Regular expression to capture years of experience
    experience_regex = re.compile(r'(experience:)?\s*(\d+)(?:[-–]\d+)?\+?\s*(?:years?|yrs?)'
                          r'\s*(?:of\s*(?:experience|work|professional|industry)?|related     experience
required|directly related experience)?',
        re.IGNORECASE)


    # Loop through each sentence and return the first one that matches the regex
    for sentence in sentences:
        if experience_regex.search(sentence):
            return sentence  # Return the sentence containing the experience details

    return None  # Return None if no match is found
def extract_years(sentence):
    if sentence:
        # Regular expression to extract the number of years
        match = re.search(r'\bfor\s+over\s+\d+\s+years\b', sentence, re.IGNORECASE)
        if match:
            return int(match.group(2))  # Return the extracted number of years
    return None  # Return None if no match is found

# Function to extract the salary sentence
def extract_salary_sentence(summary):
    sentences = nltk.sent_tokenize(summary)
    # Updated regex to capture salary ranges and time frames in a flexible way
    salary_regex = re.compile(
                                                r'(\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?(?:\s*[-
to]\s*\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?)?)'
        r'\s*(?:per\s*(year|week|hour|month|day|hr|w2))?',
        re.IGNORECASE
    )
    for sentence in sentences:
        if salary_regex.search(sentence):
            return sentence  # Return the sentence containing the salary details
    return None  # Return None if no salary is found

def extract_salary_info(sentence):
    if sentence:
        salary_regex = re.compile(
                                                r'(\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?(?:\s*[-
to]\s*\$\d{1,3}(?:,\d{3})*(?:\.\d{2})?(?:k|K)?)?)'
            r'\s*(?:per\s*(year|week|hour|month|day|hr|w2))?',
            re.IGNORECASE
        )
```

```
        match = salary_regex.search(sentence)
        if match:
            salary_value = match.group(1)  # Extract salary range or single value
            salary_time_frame = match.group(2) if match.group(2) else 'Unknown'  # Extract time
frame, if any
            return salary_value, salary_time_frame
    return None, None  # Return None if no salary is found

# Specify a different encoding (e.g., ISO-8859-1 or Windows-1252)
csv_file = r'G:\My Drive\UniSA data analyse\Advanced BI and Analytics\Python\Jobs with
Summary.csv'
df = pd.read_csv(csv_file, encoding='ISO-8859-1')  # or encoding='Windows-1252'

# Apply the extraction function to the 'summary' column to create a new 'experience' column
df['experience_sentence'] = df['job_summary'].apply(lambda x:
extract_experience_sentence(str(x)))

# Apply the function to extract years from the 'experience_sentence' column
df['experience_years'] = df['experience_sentence'].apply(lambda x: extract_years(str(x)))

# Apply the extraction function to find the sentence containing the salary information
df['salary_sentence'] = df['job_summary'].apply(lambda x: extract_salary_sentence(str(x)))

# Apply the function to extract salary and time frame from the 'summary' column
df[['salary_value', 'salary_time_frame']] = df['salary_sentence'].apply(lambda x:
pd.Series(extract_salary_info(str(x))))

# Function to categorize based on keywords
def categorize_work_type(summary):
    summary = summary.lower()  # Convert to lowercase for case-insensitive matching

    # Check for remote keywords
    for keyword in remote_keywords:
        if re.search(r'\b' + re.escape(keyword) + r'\b', summary):
            return 'Remote'

    # Check for hybrid keywords
    for keyword in hybrid_keywords:
        if re.search(r'\b' + re.escape(keyword) + r'\b', summary):
            return 'Hybrid'

    # Check for colocated/on-site keywords
    for keyword in colocated_keywords:
        if re.search(r'\b' + re.escape(keyword) + r'\b', summary):
            return 'On-site'
```

```
    return 'Unknown'  # Default category if no keywords match

# Function to identify education qualifications
def extract_education_qualification(summary):
    summary = summary.lower()  # Convert to lowercase for case-insensitive matching
    found_qualifications = []
    for qualification, keywords in education_keywords.items():
        for keyword in keywords:
            if re.search(r'\b' + re.escape(keyword) + r'\b', summary):
                found_qualifications.append(qualification)
                break  # No need to check other keywords for this qualification
    return ', '.join(found_qualifications) if found_qualifications else 'None'

# Apply the categorize_work_type function to the 'summary' column to create a new 'category' column
df['job_type_summary'] = df['job_summary'].apply(lambda x: categorize_work_type(str(x)))

# Apply the extract_education_qualification function to the 'summary' column to create a new 'education' column
df['Qualification'] = df['job_summary'].apply(lambda x: extract_education_qualification(str(x)))

# Display the updated DataFrame with the new 'experience' column
print(df[['job_summary', 'experience_sentence', 'experience_years','salary_sentence', 'salary_value', 'salary_time_frame', 'job_type_summary','Qualification']])

# Save the updated DataFrame to a new CSV file
df.to_csv('jobs_with_experience.csv', index=False)
```

The result will be like the table 3

experience_years	salary_value	salary_time_frame	job_type_summary
30			Remote
7			Hybrid
3			On-site
			Hybrid
4			Remote
2			Unknown
9			Unknown
5			Remote
5	$90,000	Unknown	Unknown
	$51,900.00 - $74,200.00	year	Remote

5	$90,000 - $110,000	Unknown	Hybrid
8	$76,200.00 - $114,300.00	Unknown	Unknown
8			Hybrid
	$140,000-$160,000	Unknown	Unknown
4	$78,670-$150,330	Unknown	Unknown
8			Unknown
1			Unknown
2			Hybrid
5			Hybrid
4	$126,480.00 - $163,680.00	Unknown	Unknown
5			Unknown
5			Hybrid
100	$120,000	Unknown	Remote

Table 11: Extracted experience, salary, salary type and job type results from the summary column

The Python program would be able to extract the keywords for experience, salary and job type. However, the result has several mixes of values. The experience has 30 and 100 years, there are salaries with a range and salary types are unknown and several unknown job types. Hence, the extracted result needs some manual cleaning and transformation processing to get the workable values.

- If the salary is in a range, it is calculated to get the middle value.
- If the salary is at an hourly rate, it is calculated by assuming normal working hours as 7.5. The equation is hourly salary * 7.5 * 5 * 52
- If the salary is at a weekly rate, it is calculated by multiplying 52
- If the salary is at a monthly rate, it is calculated by multiplying 12

Entities (Dimensions):

The dataset shows different entities when analysing the business analysis jobs. Each job has a set of skills and skills can be divided into soft skills and technical skills. The identified entities are listed in the following.

Soft skills – personal skills to carry the work duties such as communication, problem-solving and teamwork
Technical skills – The specialised knowledge or expertise for a special skill or set of skills to perform the job.
Cities – The job position related work city
State – The job position-related work state
Speciality – In which speciality area the job is looking
Industries – For which industry this job need
Qualification – The highest qualification requested for the job
Experience – the number of years of experience needed for the job.

Salary Years- Posted salary per year
Job Type- Whether the job is onsite, remote, or hybrid.

Relations:
The dimensions of jobs have a relation to the job and other dimensions. The one entity can be related to other entities in different ways such as one (1) to many (*) or many (*) to many (*).

Job $\xrightarrow[1 \qquad\qquad *]{}$ Soft Skills

Job $\xrightarrow[1 \qquad\qquad *]{}$ Technical Skills

City $\xrightarrow[1 \qquad\qquad *]{}$ Jobs

State $\xrightarrow[1 \qquad\qquad *]{}$ Jobs

Speciality $\xrightarrow[1 \qquad\qquad *]{}$ Skills

Industry $\xrightarrow[1 \qquad\qquad *]{}$ Jobs

Industry $\xrightarrow[1 \qquad\qquad *]{}$ Skills

Speciality $\xrightarrow[1 \qquad\qquad *]{}$ Jobs

Job $\xrightarrow[1 \qquad\qquad 1]{}$ Salary

Job $\xrightarrow[1 \qquad\qquad 1]{}$ Experience

Job $\xrightarrow[1 \qquad\qquad 1]{}$ Job Type

Semantic Data Model

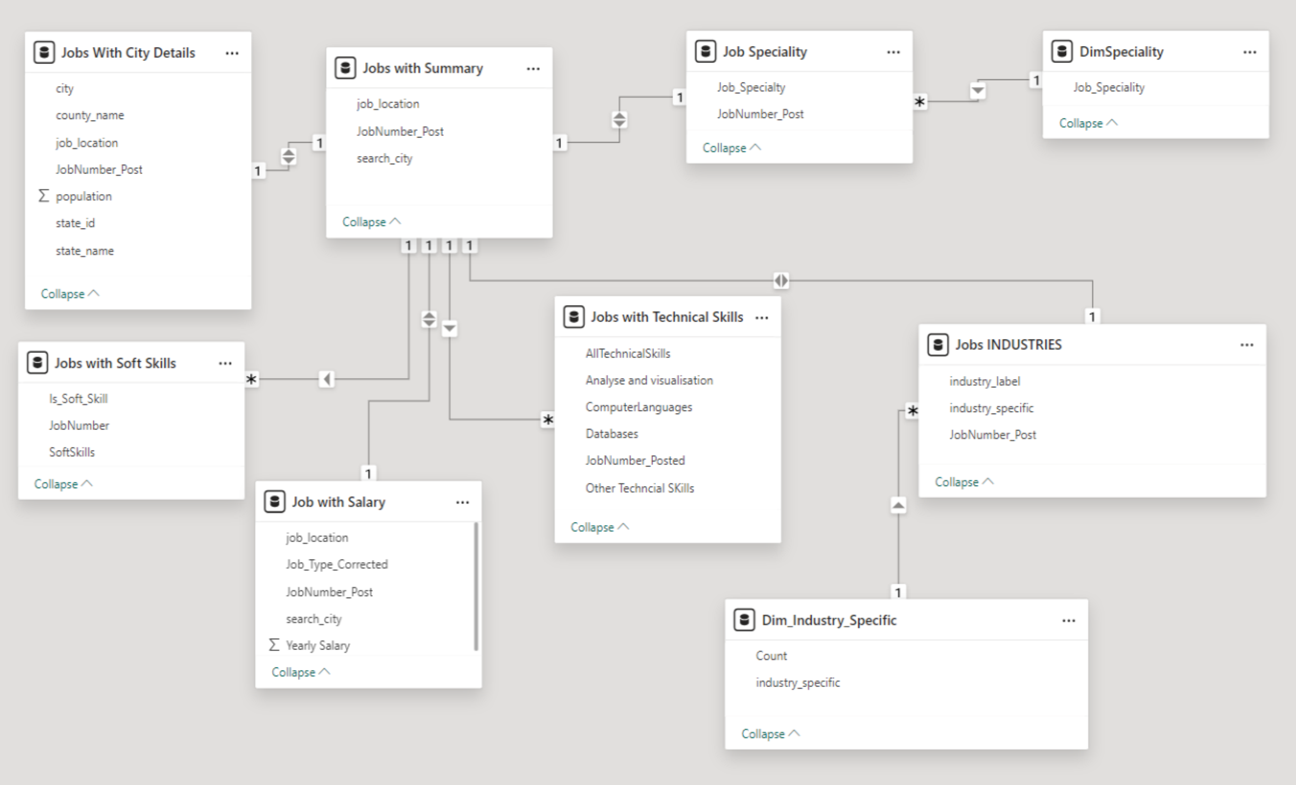


Figure 1: The semantic model of the Business Analysis Skill Analysis

Data Analysis

Each business analyst job has a list of skills requested by the company or recruiters. Knowing how these skills spread around the business analysts' jobs is very important. First, the project aims to determine how many skills need to be recruited per job. Then, there is knowledge about how skills are distributed in the dataset. The next part is identifying the key skills in the data files. Once the key skills are determined, need to find out how the skills affect each job speciality. This way skills will be examined throughout the project.

After examining the data on business analyst jobs, the project aims to find out the association's rules for the skills, experience, qualifications, and salary with a job speciality. Th. The rule will help to find the best skills set for each speciality and industry for applicants. Further, the model will tell how many years of experience are expected for each specific speciality with the qualifications.

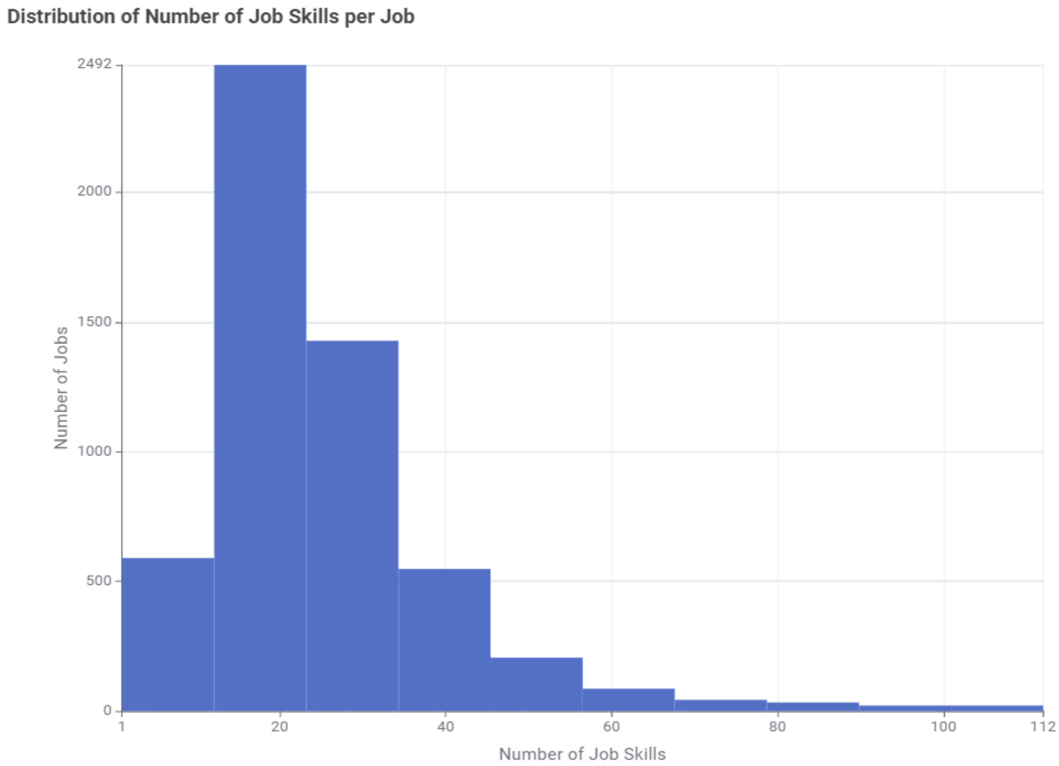


Figure 2: Number of Skills per job Histogram

Business Analyst professional needs various skills to get a job in the market. There are two types of skills, and they are technical skills and soft skills. The above histogram shows the distribution of the required skills requested by each job. The distribution is right-skewed and has a long tail and there is one significant mode which is 21 to 22 skills, and the distribution has three or four modes. It seems like some various factors influence the number of skills for the job. There are some outliers in the dataset which can be seen with the boxplot.

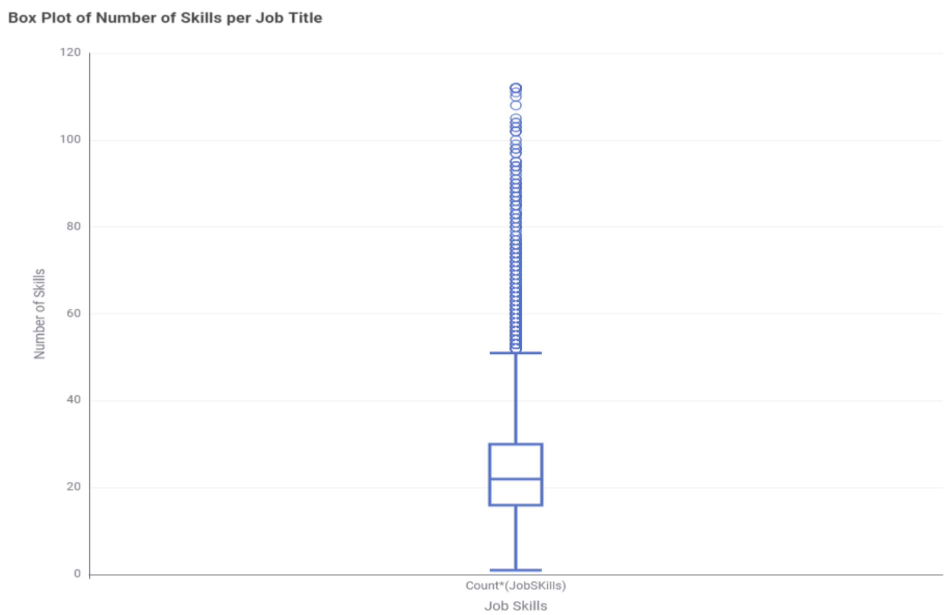


Figure 3: Boxplot of number of skills per job

Figure 3 shows the boxplot of the number of skills listed per job. According to the boxplot, the maximum number of skills that should be in a job is around 51. Hence, some jobs have a high number of skills, and they are outliers.

The skills are divided into two main sections to analyse in the deep. The two main sections are technical skills and soft skills. Further, the skills depend on the speciality of the job. Therefore, the skills can be analysed with each specialty.

Figure 4 shows the box plot of the number of skills per job after soft skills are separated from the skills list. Still, there are several outliers shown in the soft skill category. Therefore, the soft skills have to go through another treatment to reduce the outliers.

Other skills have a high volume of outliers still. Other skills can be filtered with specific technical skills to identify common skills.

Please check Appendix the Knime workflow for the procedure of data manipulation and filtering.

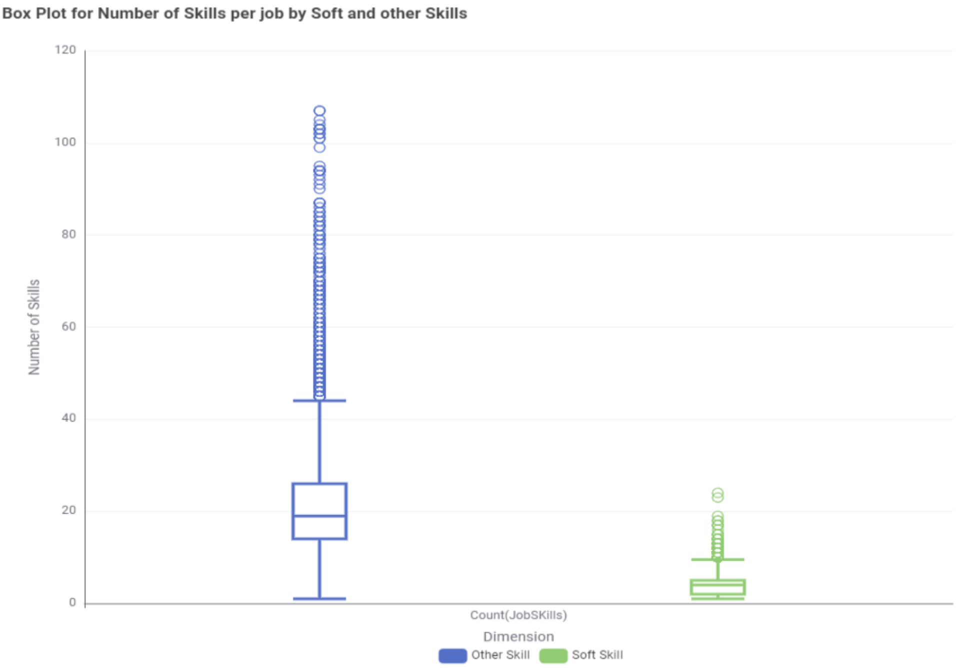


Figure 4: Boxplot of the number of skills per job by soft skills and other skills

First, the soft skills is analysed as it is pivotal to performming a business analyst job. The business analyst needs to communicate with different stakeholders to understand the business objectives. After understanding the business objectives, the analysis plan will need to be documented and presented to the stakeholders. Therefore, communication, documentation and presentation are also very important in Business analysis jobs to perform the analysis of the business (Arvind A. 2020).

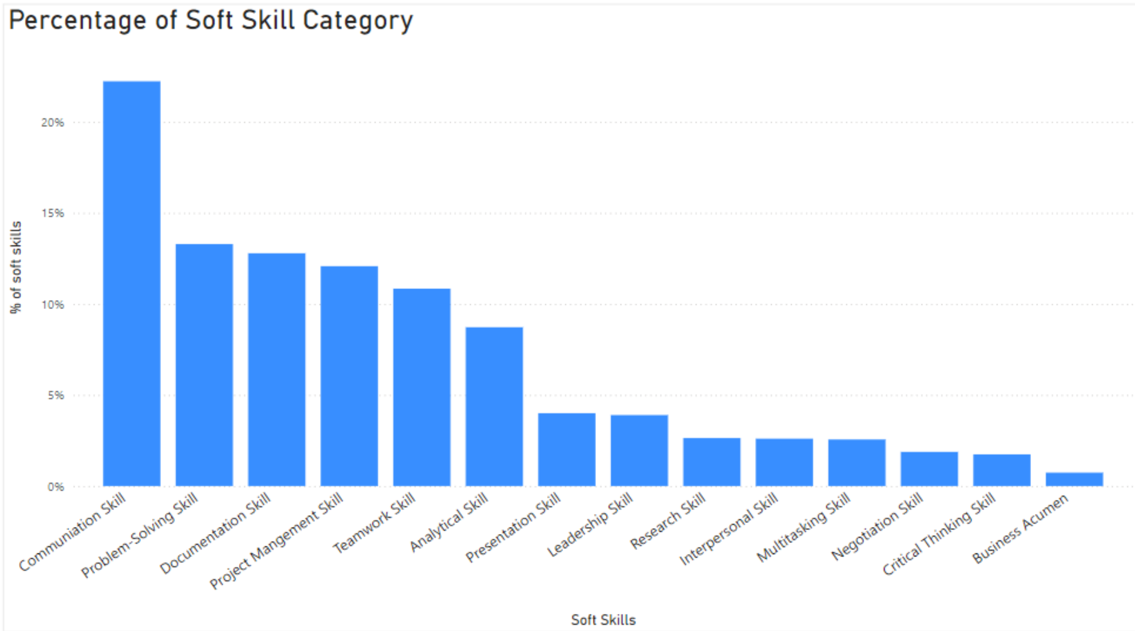


Figure 5 : Percentage of each soft skill request from the job posting

According to Figure 05, communication skills show a significantly high percentage (more than 20%). This is because almost every job has asked for communication skills has a requirement. Problem-solving skill is the second most requested skill but there is not much difference with document and project management skills. Therefore, the four most important soft skills are communication, problem-solving, documentation, and project management. Teamwork and analytical skills are also relatively higher than the rest. Each business analyst's speciality has different requirements. So, it would be better to check how these top six soft skills are requested in each speciality.

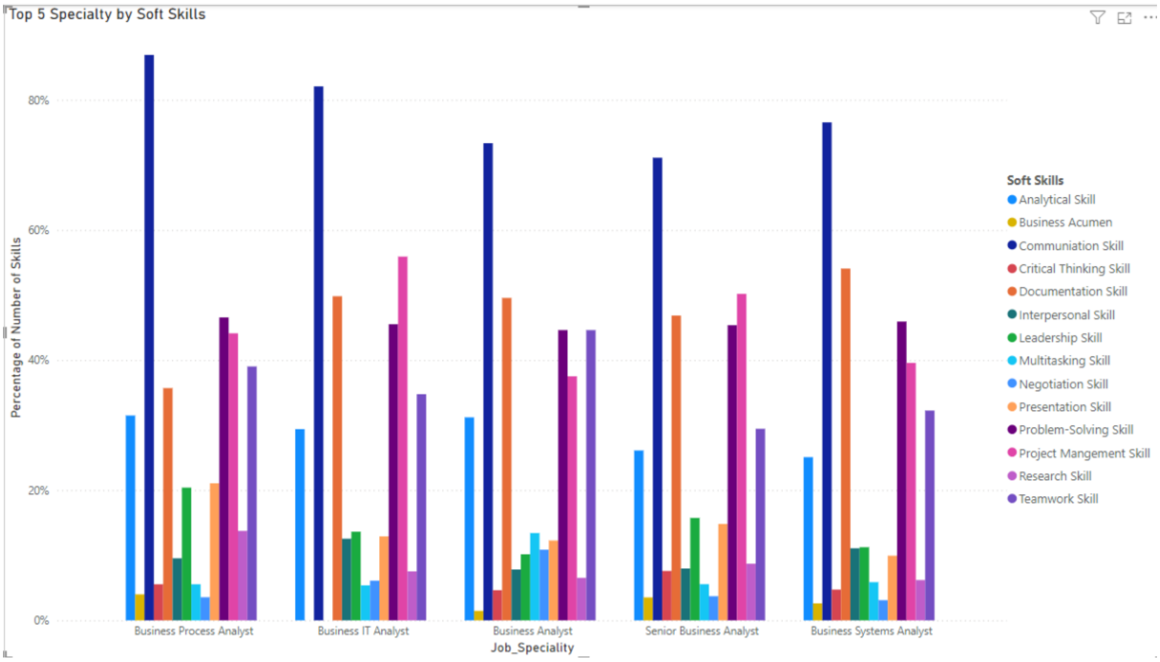


Figure 6: Percentage of each soft skill request by job speciality category

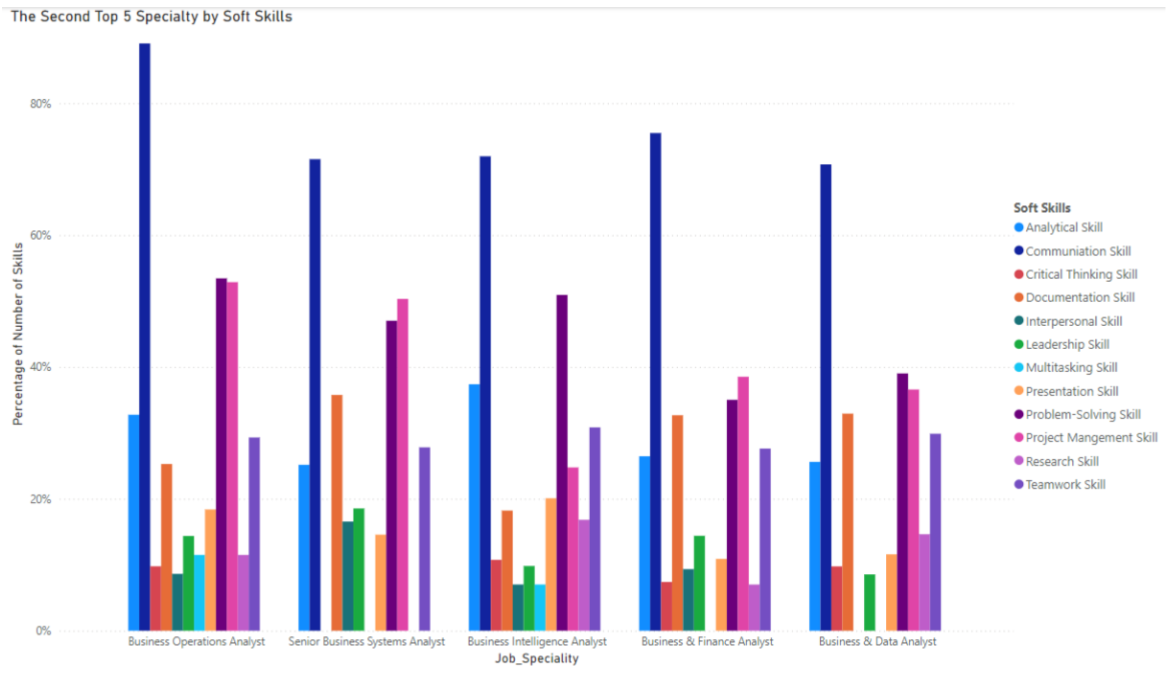


Figure 7: Percentage of each soft skill request by job speciality category

According to Figure 6 and 7, when comparing the top 10 specialties, communication skill is the highest requested skill among every skill. The problem-solving skill shows as the second highest requested skill among the specialties. Documentation, project management and teamwork skills are in third place in different specialties. The senior specialties show a high demand for Leadership skills as well.

Technical Skills by category

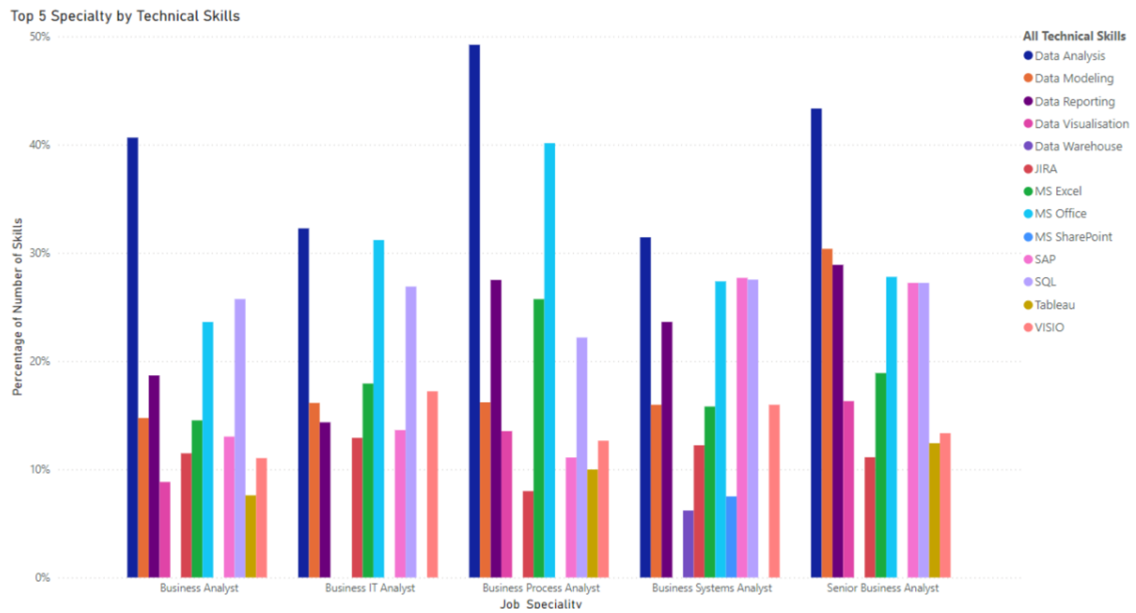


Figure 8: Percentage of each technical skill request by job speciality category

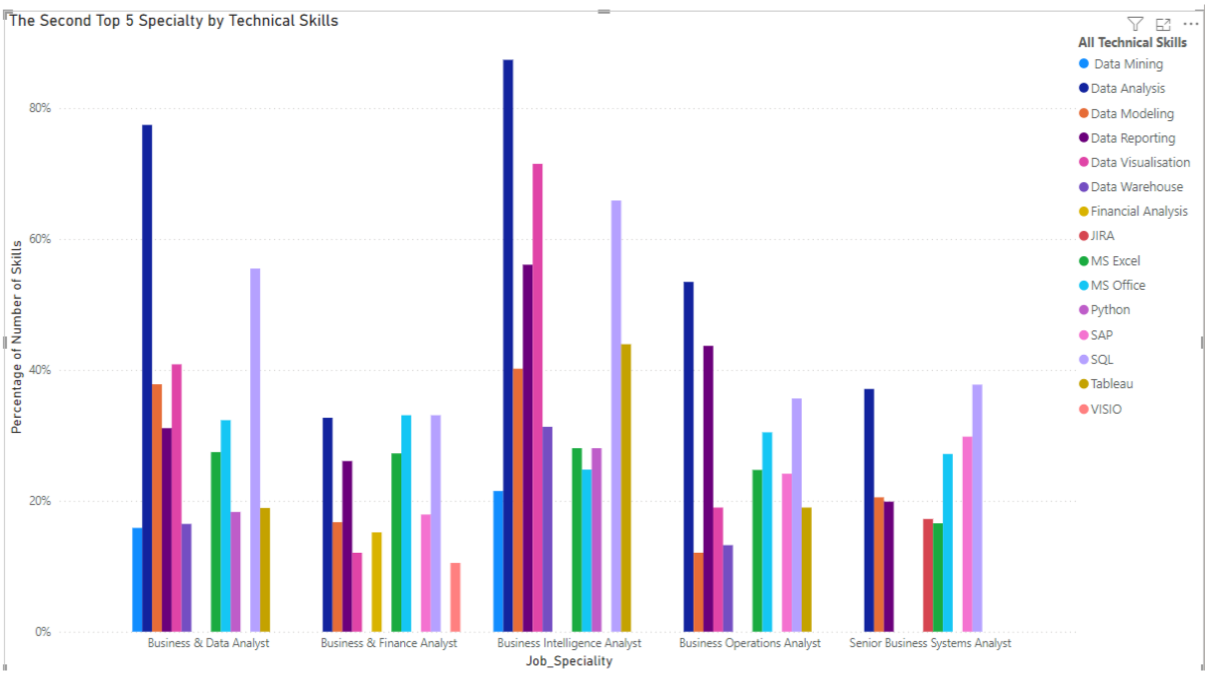


Figure 9: Percentage of each technical skill request by job speciality category

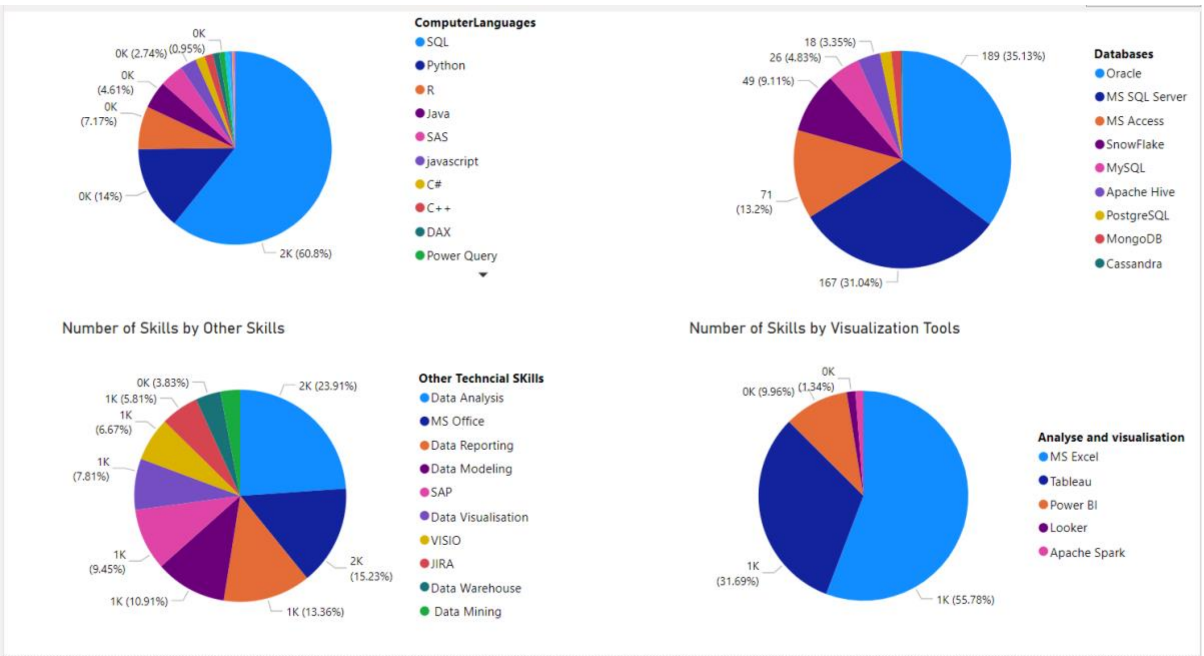


Figure 10: Demand for technical skills in each speciality category.

When comparing the technical skills, every specialty has expected data analysis skills as the highest required skill. As a technical skill, SQL has a high demand in the most of specialties. Business intelligence and senior business intelligence analysts show a high demand for data visualization and data reporting skills. Further, MS Excel and MS Office are also in demand in posted jobs.

Further considering the technical skills and tools for Business analysis jobs, they can be divided into several categories according to the speciality. Applicants are capable of computer language skills, database skills, visualization tools and analytic concept skills (TechCanvas, 2024). Figure 10 shows how each skill has been demanded in each speciality. SQL is the leading demand for business analyst jobs with 60% portion of the computer language skills. While the SQL demand comes to the Oracle and MS SQL server have high demand on the database speciality. MS Excel is the leader in visualization and analytical tools speciality passing Tableau and Power BI. In other skills, Data analysis came as the highest requested skill for the business analyst. MS Office came to second place however, this needs to be investigated as MS Office is a collection of programs such as MS Excel, and MS PowerPoint. This result is a misleading demand for technical skills.

Skill requirements for the Speciality

When considering the technical skills and soft skills bar charts in Figures 6,7,8, and 9, there is a significant difference can be seen in the speciality. Therefore, there is a hypothesis that some speciality needs special technical skills or soft skills than others. Hence, the assumption of the null hypothesis is no difference in the mean of the number of skills required for the speciality.

First, check the normality of the number of jobs to decide which test carries the hypothesis test for technical skills and soft skills.

The UNIVARIATE Procedure
Variable: Count_jobs

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.355664	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.356932	Pr > D	<0.0100
Cramer-von Mises	W-Sq	29.62123	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	145.4848	Pr > A-Sq	<0.0050

Table 12: Normality test result of the number of jobs per technical skill in each speciality.

The UNIVARIATE Procedure
Variable: count_jobs

Tests for Normality				
Test	Statistic		p Value	
Shapiro-Wilk	W	0.381062	Pr < W	<0.0001
Kolmogorov-Smirnov	D	0.348226	Pr > D	<0.0100
Cramer-von Mises	W-Sq	11.21321	Pr > W-Sq	<0.0050
Anderson-Darling	A-Sq	55.98238	Pr > A-Sq	<0.0050

Table 13: Normality test result of the number of jobs per soft skill in each speciality.

According to Tables 12 and 13, the values are not in normal distribution as all p-values are less than 0.05. Therefore, the assumption of normality has to be rejected. Hence the values don't follow the normality, Kruskal-Wallis test has to be used to check the means differences.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
143.0703	23	<.0001

Table 14: Kruskal-Wallis Test result for technical skills

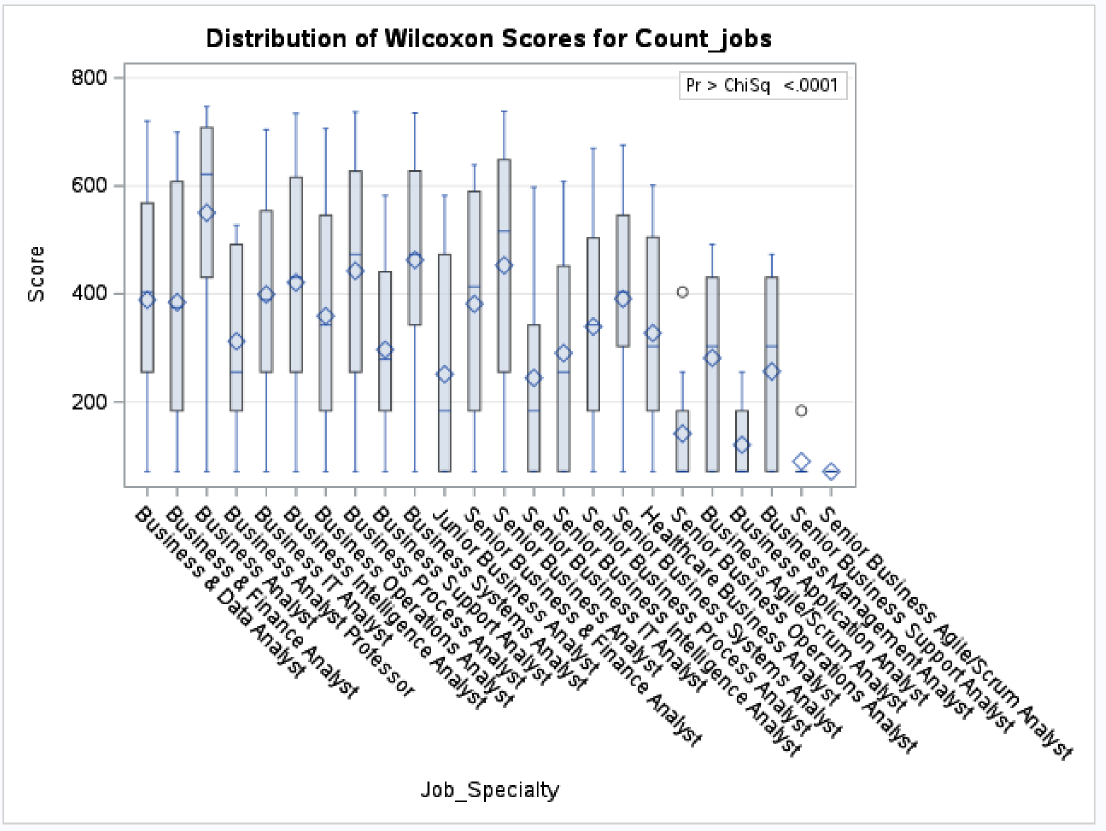


Figure 11: Boxplot of comparison of job specialties Kruskal-Wallis Test result for technical skills

The Kruskal-Wallis test gave chi-square = 143.0703 (data freedom = 23) with a p-value is less than 0.001. Hence, if the p-value is less than 0.05, we can reject the assumption of the null hypothesis of equal means. This means the specialities have different technical skill requirements. Figure 11, the boxplot shows the same results as the boxes are not aligned with the same range.

Kruskal-Wallis Test		
Chi-Square	DF	Pr > ChiSq
175.0076	23	<.0001

Table 15: Kruskal-Wallis Test result for soft skills

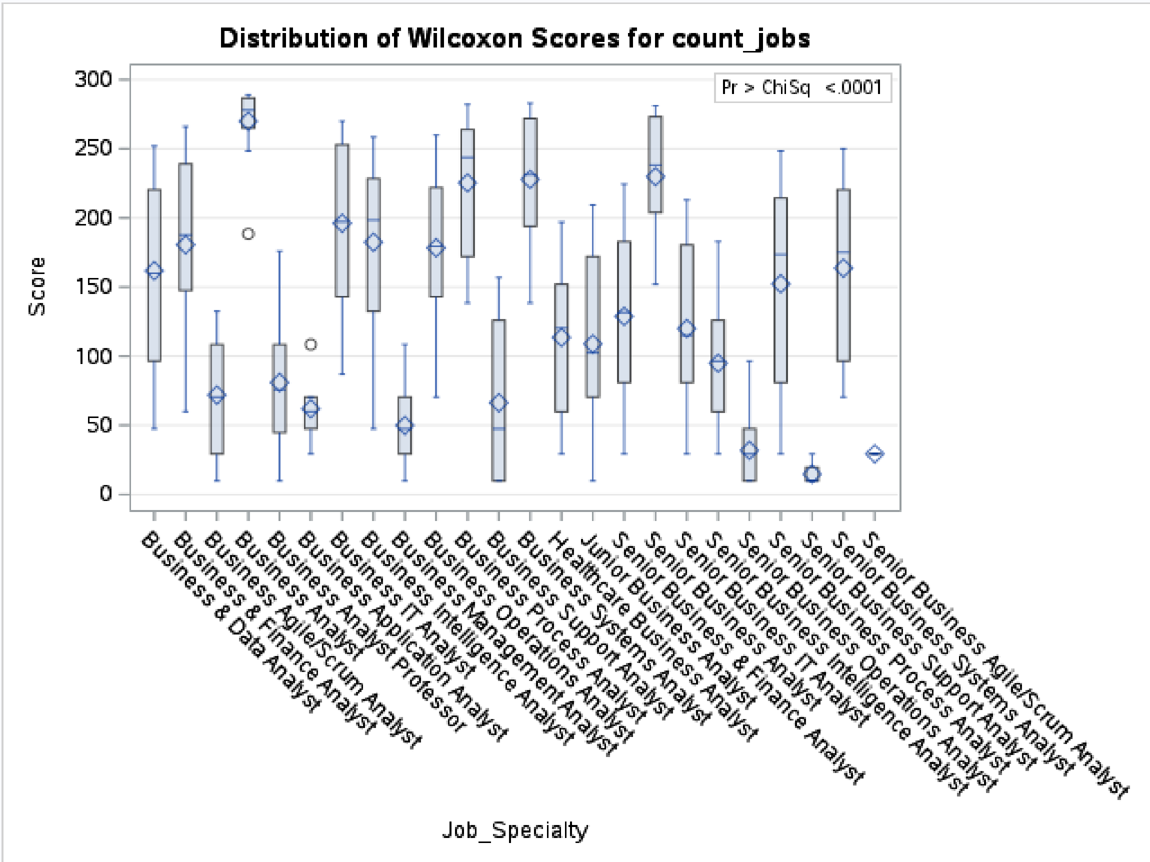


Figure 12: Boxplot of comparison of job specialties Kruskal-Wallis Test result for soft skills

According to Table 15 and Figure 12, soft skills also show the same characteristics. The p-value is less than 0.0001 for the Kruskal-Wallis test and the null hypothesis is rejected for the assumption for equal means. The speciality has different means for each soft skill.

Checking the association of skills in each job specialty.

Considering all facts, each job specialty needs to have some special skills in technical and soft skills. Therefore, we can find how the skills are associated with each other for the specialities. LinkedIn job posts have given the skills requirements for each job.

Job Number	Skills Required
3110767987	[Data Visualisation]
3120115540	[Data Analysis]
3148221520	[JIRA, SAP]
3198529598	[Data Analysis, SQL, Python, Tableau, MS Excel, Data Analysis]
3275602758	[SQL, Data Warehouse, Data Analysis]
3282258671	[Data Reporting, Data Modelling, Data Reporting]
3324133236	[SQL, Data Warehouse]
3441455649	[Data Analysis, Data Reporting, Data Visualisation, Tableau, SQL, SQL, DAX, SSRS, SSIS]
3441460823	[SAP]

Table 16: Skills requested for each job

As shown in Table 16, each job has a set of skills from 1 to many. Some skills are listed in the many jobs and some skills are listed in one or two jobs.

The association rules are built according to the transaction of each job.

(Data analysis) à (SQL)

(SQL) à (Data Warehouse)

(Data analysis, SQL) à (Tableau)

To derive the association rules, the minimum support needs to be set up.

The minimum support = $\frac{Number\ of\ times\ the\ items\ are\ bought\ together}{number\ of\ transactions} * 100$

Considering all facts, the minimum support is set to 10% as the items need to be listed 10 times per 100 transactions.

The next fact is the confidence level of the association rule. That tells how confident skills come when selecting another skill. The confidence level is set to 50% as it will drop all less confident association rules.

The Knime workflow will be used to develop the association rules for job skills.

Note: The complete workflow is in the Appendix

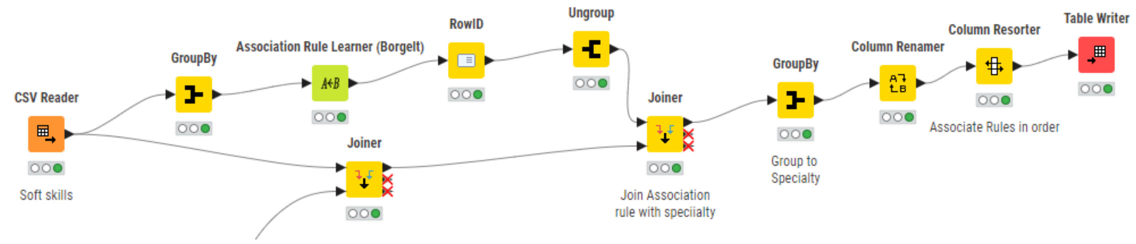


Figure 14: Association rule learner for soft skills

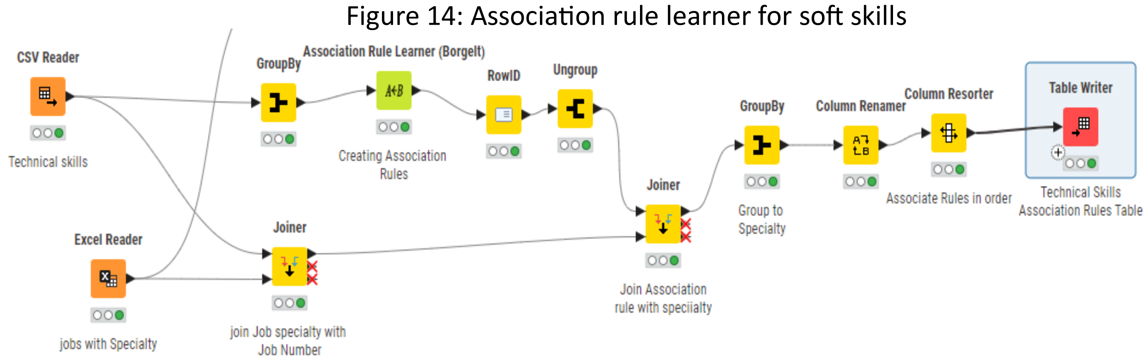


Figure 13: Association rule learner for technical skills

According to Table 2, skills' medians are statistically significant. This means some skills have a higher demand than other skills.

With this result, the association rule can be built for the skills. The minimum support level is set for 5.0.

The following table shows a sample of association rules built for the job skills.

Antecedent Technical Skill	Rule Support	Rule Confident%	Recommended Technical Skills	Recommended Job Speciality
Data Analysis, Python	231	88.5	SQL	Business & Finance Analyst
Data Analysis	231	88.5	SQL	Business Agile/Scrum Analyst
Data Analysis, Python	231	88.5	SQL	Business Analyst
Python, Data Analysis	231	88.5	SQL	Business Analyst Professor
Data Analysis	231	88.5	SQL	Business Application Analyst
Python, Data Analysis	231	88.5	SQL	Business IT Analyst

Data Analysis, Python	231	88.5	SQL	Business Intelligence Analyst
Data Analysis	231	88.5	SQL	Business Management Analyst
Data Analysis, Python	231	88.5	SQL	Business Operations Analyst
Python, Data Analysis	231	88.5	SQL	Business Process Analyst
Data Analysis, Python	231	88.5	SQL	Business Support Analyst
Data Analysis, Python	231	88.5	SQL	Business Systems Analyst
Python, Data Analysis	231	88.5	SQL	Senior Business Operations Analyst
Data Analysis, Python	231	88.5	SQL	Senior Business Process Analyst
Data Analysis, Python	231	88.5	SQL	Senior Business Systems Analyst
Python	328	85.9	SQL	Business & Data Analyst
Python	328	85.9	SQL	Business & Finance Analyst
Python	328	85.9	SQL	Business Analyst
Python	328	85.9	SQL	Business Analyst Professor
Python	328	85.9	SQL	Business IT Analyst
Python	328	85.9	SQL	Business Intelligence Analyst
Python	328	85.9	SQL	Business Operations Analyst
Python	328	85.9	SQL	Business Process Analyst
Tableau, MS Excel	148	55.8	Data Visualisation	Senior Business Intelligence Analyst

MS Excel	148	55.8	Data Visualisation	Senior Business Operations Analyst
MS Excel, Tableau	148	55.8	Data Visualisation	Senior Business Process Analyst
MS Excel, Tableau	148	55.8	Data Visualisation	Senior Business Systems Analyst
Data Analysis, Python	144	55.2	Tableau	Business & Data Analyst
Data Analysis, Python	144	55.2	Tableau	Business & Finance Analyst
Data Analysis	144	55.2	Tableau	Business Agile/Scrum Analyst
Data Analysis, Python	144	55.2	Tableau	Business Analyst
Python, Data Analysis	144	55.2	Tableau	Business Analyst Professor
Data Analysis	144	55.2	Tableau	Business Application Analyst
Python, Data Analysis	144	55.2	Tableau	Business IT Analyst
Data Analysis, Python	144	55.2	Tableau	Business Intelligence Analyst
Data Analysis	144	55.2	Tableau	Business Management Analyst
Data Analysis, Python	144	55.2	Tableau	Business Operations Analyst

Table 17: Association rules table sample

Testing Association rules

The association rules can be tested using the Figure 15 Knime workflow.

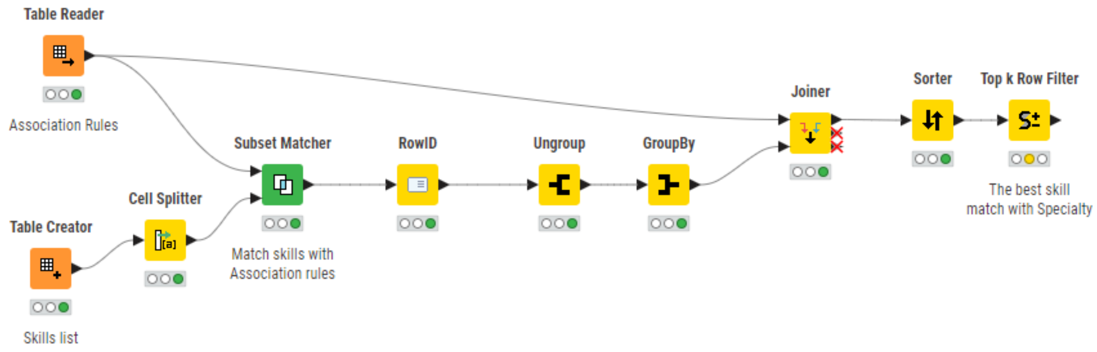


Figure 15: Test Association rule to match the best specialty

The association rule is tested with these skills Python "MS Excel" "Data Analysis" SQL to match the best specialty.

Antecedent Skills	Rule support	Rule Confident	Rule Lift %	Recommended Skills	Recommended Specialty
[SQL, MS Excel]	332	69.2	142.18	Data Analysis	Business & Finance Analyst
[MS Excel, SQL]	332	69.2	142.18	Data Analysis	Business & Data Analyst
[SQL, MS Excel]	332	69.2	142.18	Data Analysis	Business Application Analyst
[Data Analysis]	337	60.7	174.68	SQL	Business Agile/Scrum Analyst
[Data Analysis]	337	60.7	174.68	SQL	Business Application Analyst

Table 17: Association rules match skills and specialties

Reflection

The analysis of the job skills shows a very high association with Data analysis and SQL for the technical skills. It is necessary to know SQL, Data Analysis and Data Visualization to work in the Business Analysis field for any speciality. LinkedIn article (Shubham A. 2023) says SQL needs to extract and manipulate large datasets to understand large datasets and analyse them. Further, SQL always helps to aggregate and summarize data to easily understand patterns and trends. Another point is sometimes data needs to be merged and joined from different tables or sources. SQL is the best way to get joined products from different sources.

Data visualization is another main skill for business analysts. Without knowing the data visualisation, it is not easy to communicate information with the clients. The visualization helps to identify and isolate the factors related to the client. Further, customers can see the trends or patterns in their data to identify where went wrong or success (Hashemi-Pour, Brush, Burns 2024).

As the data analysis and data visualization show high demand and association in skills, MS Excel and Tableau are also in slightly high demand in technical skills. In the Medium article, Sapra (2023) express that it is vital to know MS Excel and Tableau applications to data clean, transform, analyse and visualize.

The Knime association rules learner does not give a good result in finding highly associated skills for each speciality. The initial analysis shows there is a significant difference in means of skills in speciality. Therefore, this analysis can further drill down to more categories to find missing connections. It would need some decomposition to databases, languages and visualization tools for the technical skills.

Data Quality

Completeness:

Number of Records Match with Business Analyst:

LinkedIn_job_posting = 1348454

Job_summary = 1297332, missing = 1348454- 1297332 = 51,122, percentage of missing data: 4%

Job_skills = 1296381, missing = 1348454- 1296381 = 52,073, percentage of missing data: 4%

Completeness Score = (96 + 96)/2 = 96%

Impact: very small

Accuracy:

The data has been collected from the LinkedIn social network. The data has been managed by LinkedIn. Therefore, as the data was entered into the job-creating portal, the data was accurate to use for the project.

Job_location has some city names that didn't match with the simplemap data. The job_location data has some location for small town and different countries.

Description	Number of Records
Non-USA jobs	19
Invalid Cities	760
Valid cities	4841

valid data = 4841

Accuracy score = (4841/ 5621) * 100 = 86%

Impact: Minimum

Consistency:

Kaggle data repository stored the data files in CSV format. According to the Kaggle (Asaniczka , 2024), all the columns' data type has been defined. There are no mismatch fields in the data files.

Impact: None

Timeline:

The timeline of the data is between 12 January 2024 to 22 January 2024. According to Asaniczka (2024), the data has been updated in March. Hence, the data collected very recently, and they are most suitable to use for the analysis.

Impact: None

Validity:

The Kaggle data repository has defined the data validity and there is no mismatch in the data. It confirmed that the available data is 100% valid (Asaniczka , 2024). However, the skills, job title, cities columns showed data is in an irregular format. EG: problem solving skill has been entered in diverse ways such as problem-solving, managing problem or solving business problems.

The search_city column has the city name without the state where the city belongs. When the city match with the simplemap data, the search_city gives multiple states for the same city. Therefore, the search_city data doesn't have validity for the data analysis.

Impact: High**,** Need heavy transformation to avoid or fix invalid entries.

Uniqueness:

The dataset has 4 duplicate data records.

Total records = 6702, percentage of duplicate: less than 1%

Uniqueness score = 100%

Impact: minimum

References

Arvind, A, (2020 February). https://www.linkedin.com/pulse/business-analysis-soft-skills-new-hard-arvind-arcot/

Currency: The article was written by Arvind A who is IIBA Australia's Board Director. The article was published in 2020 February. It is 2 years back and the concept and theory are still valid for today.

Relevance: The article shows how important of the soft skills for data analysis
Authority: LinkedIn is the hosting website for the article, LinkedIn keep creditability of the article published.
Accuracy: The publisher is award-winning person. Therefore, the contain should be accurate according to his experience and knowledge.
Purpose: Understanding the most important soft skills

Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Bothma, J. (2023, November). https://www.datacamp.com/blog/top-business-analyst-skills

Currency: The article was written by Joleen who is a data science consultant & writer at statistically relevant. The article was published in 2023 November and it is a very recent article

Relevance: The article helps to understand soft and technical skills. The skills mentioned in the article helped to filter from the dataset.

Authority: Datacamp is the hosting website for the article, and it is famous for teaching companies the skills they need to do with data in the real world

Accuracy: The article information has been cross-checked with the Simplilearn.com website article from Duggal.

Purpose: Identify the soft and technical skills from the dataset file.

BrainStation, (2024). https://brainstation.io/career-guides/are-business-analysts-in-high-demand

Hashemi-Pour, C., Brush, K., Burns, E., (2024). https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization#:~:text=Why%20is%20data%20visualization%20important,to%20be%20understandable%20by%20anyone.

Rathod, R. (2024, April 02). https://medium.com/@reshmarathod0000/case-study-job-market-data-analysis-using-power-bi-efa555cacb80

Sapra, B. (2023, May 1). https://medium.com/analysts-corner/do-you-need-to-excel-in-data-analytics-before-moving-into-a-ba-role-981b64902991

Sharma, S. (2023, October 09). https://www.linkedin.com/pulse/essential-skills-seek-your-career-business-analyst/

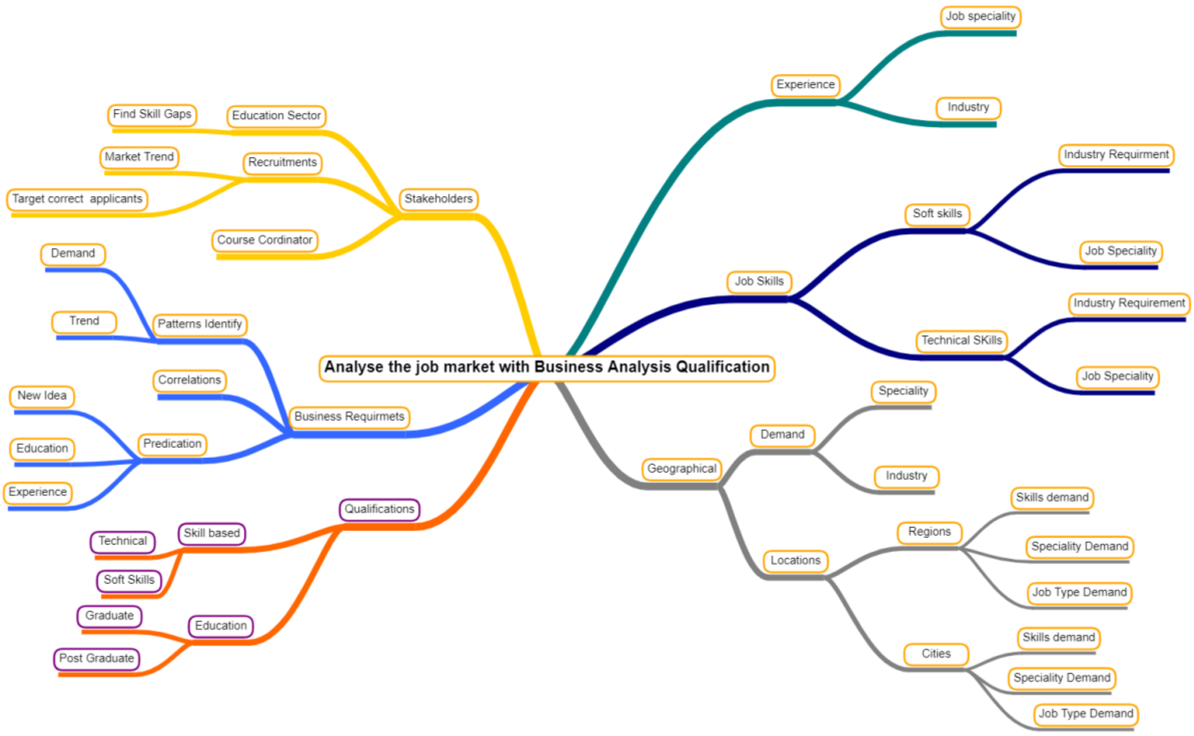
Shubham, A. (2023, March 01). https://www.linkedin.com/pulse/sql-queries-know-business-analyst-shubham-a/

TechCanvass, (2024, March 12). https://businessanalyst.techcanvass.com/what-technical-skills-a-business-analyst-should-know/

William, tc, (2023, November 14). https://medium.com/@william.tc/measuring-and-reporting-data-quality-4137e1d5aec5

Appendices

Mind map



Data Dictionary

Job_link

Definition: The URL link of the data downloaded from LinkedIn. The link contains the title and unique job number.
Format: String
Comment: This column is the primary key to all three data files.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

JobNumber_Posted

Definition: The Unique number from the job_link extracted as a separate value.
Format: String
Comment: Data must be unique and derived from the job_link.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_title

Definition: The title of the job listed in the LinkedIn.
Format: String
Comment: The job_title is used to filter business analyse jobs from the title.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_skills

Definition: The list of skills related to the job_title.
Format: String
Comment: Skills are listed as comma-separate
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Is_Soft_Skill

Definition: The skill defined as a soft skill or not.
Format: Boolean
Comment: The skills have been identified as soft skills with predefined values such as communication, problem-solving and analytical skills. This column is a flagged to separate soft skills from other skills
Code: 1 – True (Soft skill)
0- – False (non-soft skill)
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

SoftSkills

Definition: The soft skills category.
Format: String
Comment: The soft skills are derived from the **Job_skills** column with predefined categories.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

ComputerLanguges

Definition: The technical skills identified as a computer language such as SQL, Python, R. Non-computer language skills showed as null
Format: String
Comment: The skills are derived from **Job_skills** column with predefined categories.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Databases

Definition: The technical skills identified as databases such as Oracle, MS SQL server and SnowFlake.

Format: String

Comment: The skills are derived from **Job_skills** column with predefined categories.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

AnalyseAndVisualTools

Definition: The technical skills identified as AnalyseAndVisualTool such as Tableau, Power BI or MS Excel.

Format: String

Comment: The skills are derived from **Job_skills** column with predefined categories.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

OtherTechnicalSkills

Definition: Technical skills cannot be categorised as languages, databases or visualisation tools.

Format: String

Comment: The skills are derived from **Job_skills** column with predefined categories.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

AllTechnicalSkills

Definition: languages, databases, visualisation tools and OtherTechnicalSkills are combined.

Format: String

Comment: The skills are derived from **ComputerLanguges, Databases, AnalyseAndVisualTools, and OtherTechnicalSkills** columns after joining.

Job_Speciality

Definition: The posted job relevant speciality such as finance, cyber security or system analysis.

Format: String

Comment: Manually categorised each job according to the job title and search area.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_summary

Definition: The summary of the job listed in the LinkedIn.
Format: String
Comment: The experience has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Experience_years

Definition: How many years of experience need for the job.
Format: Number
Comment: The experience has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

YearlySalary

Definition: The yearly salary for the listed job.
Format: currency
Comment: The salary has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type

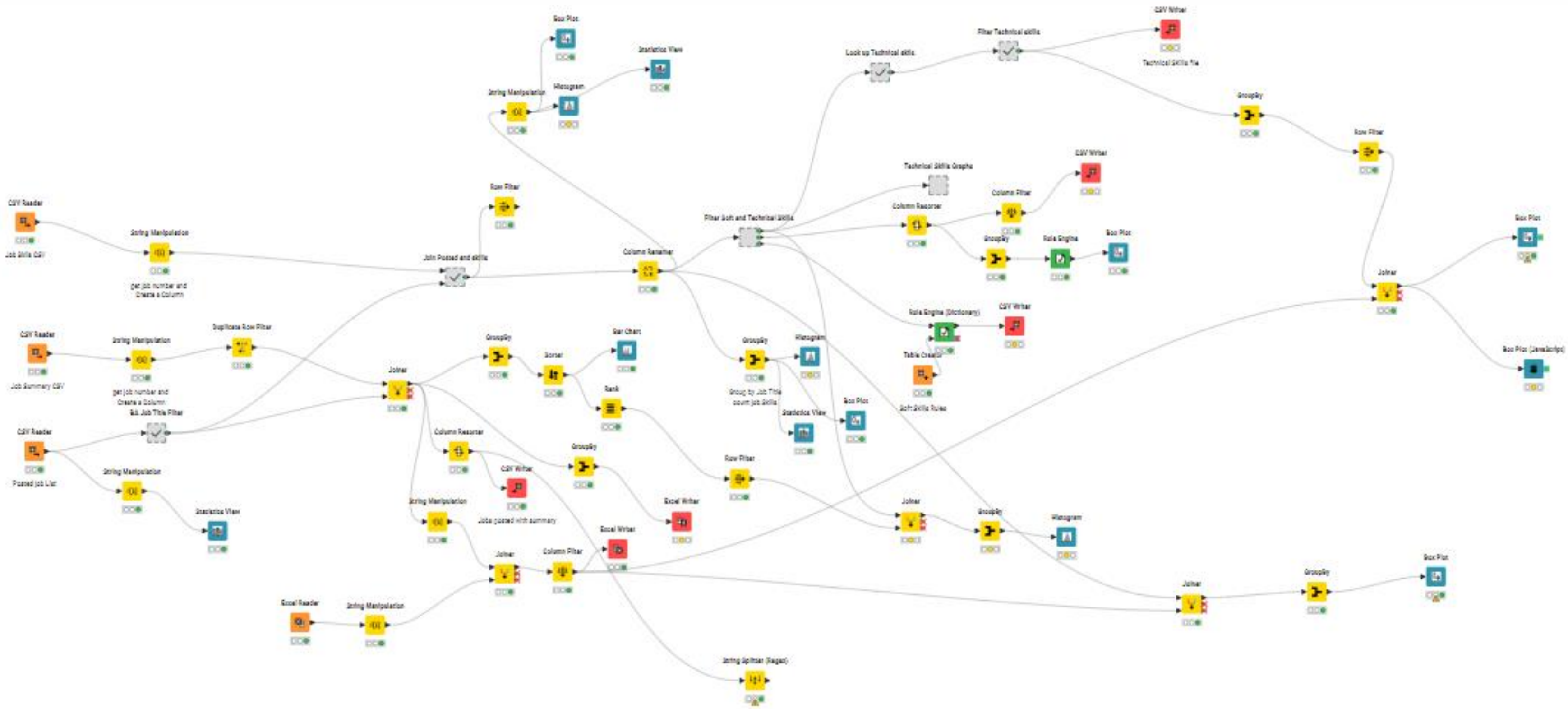
Definition: Type of the job such as remote or on-site
Format: String
Comment: The job type as remote, one-site or hybrid
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type_summary

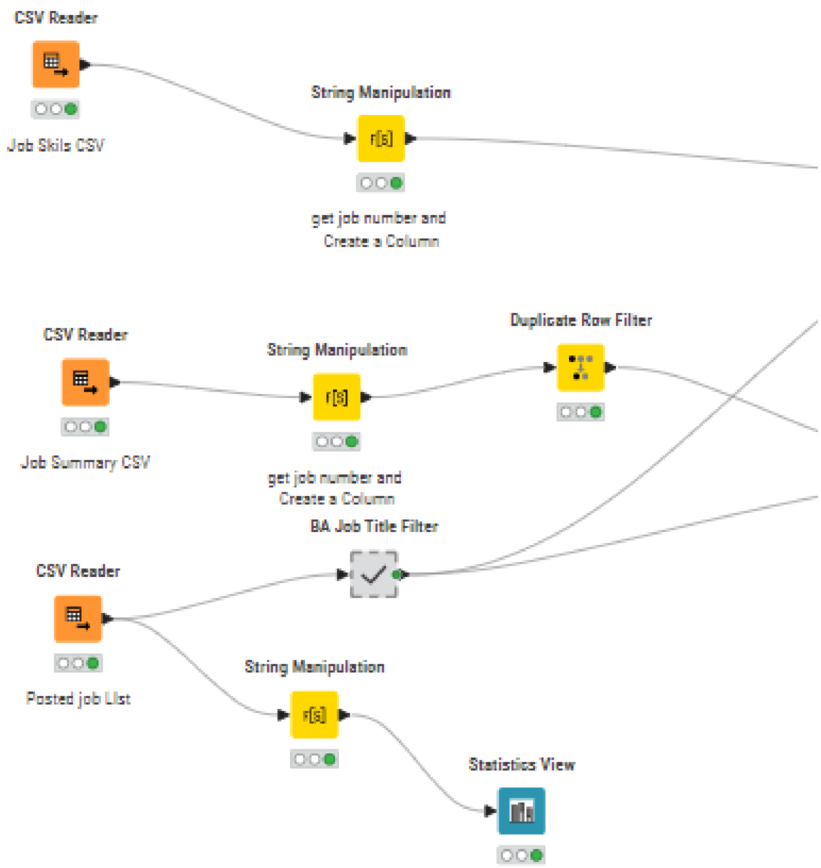
Definition: Type of the job such as remote or on-site
Format: String
Comment: The job type has been extracted from the job_summary column
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_type_corrected

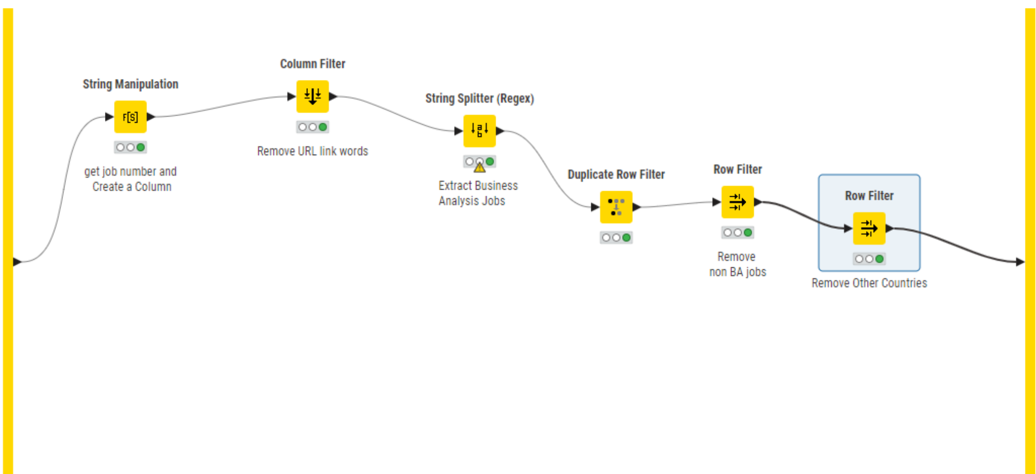
Definition: corrected job type to use for the analyse
Format: String
Comment: comparing the job_type_summary and job_type column derived the correct job type.
Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024



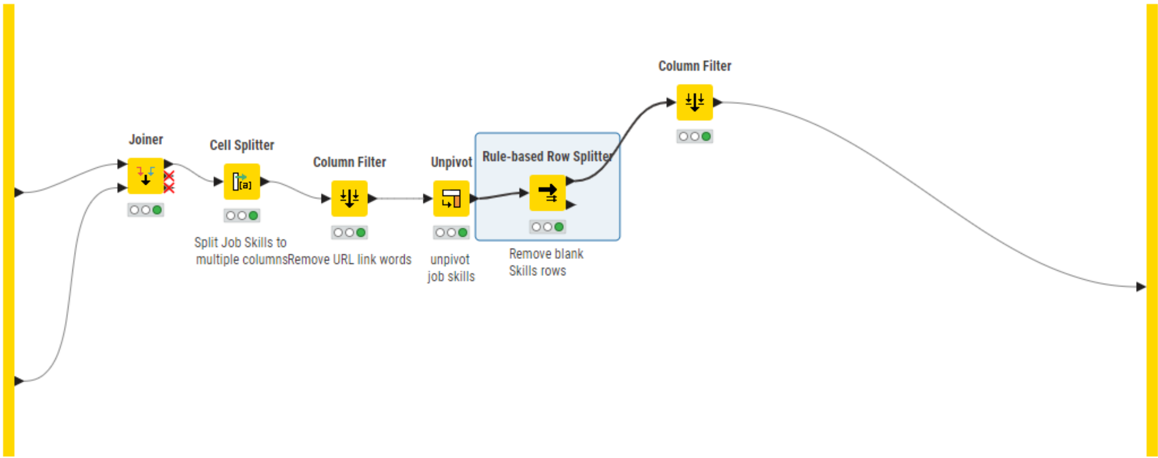
Entire workflow data load, extract, transform and store.



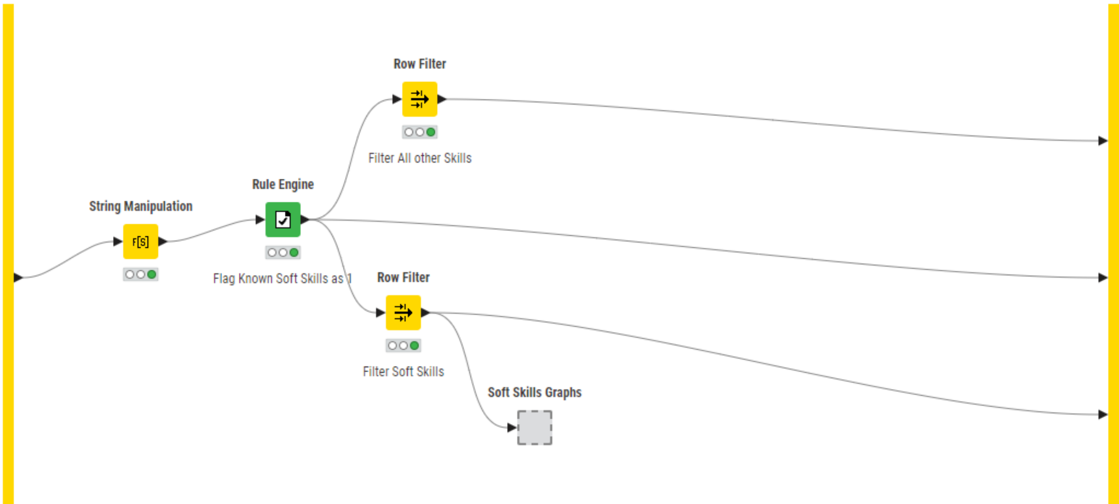
Load data from csv filese



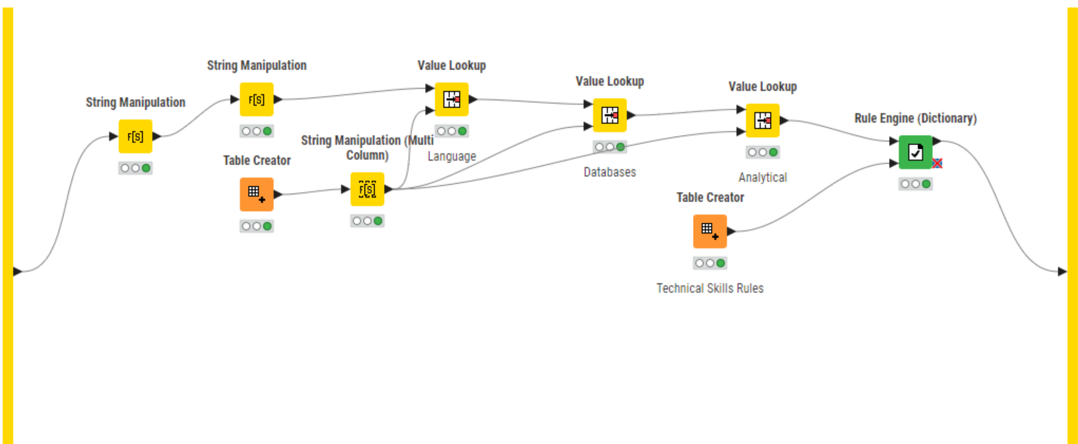
Extract Business Analysis job with regular expression: (.*Busin.*Analy.*)



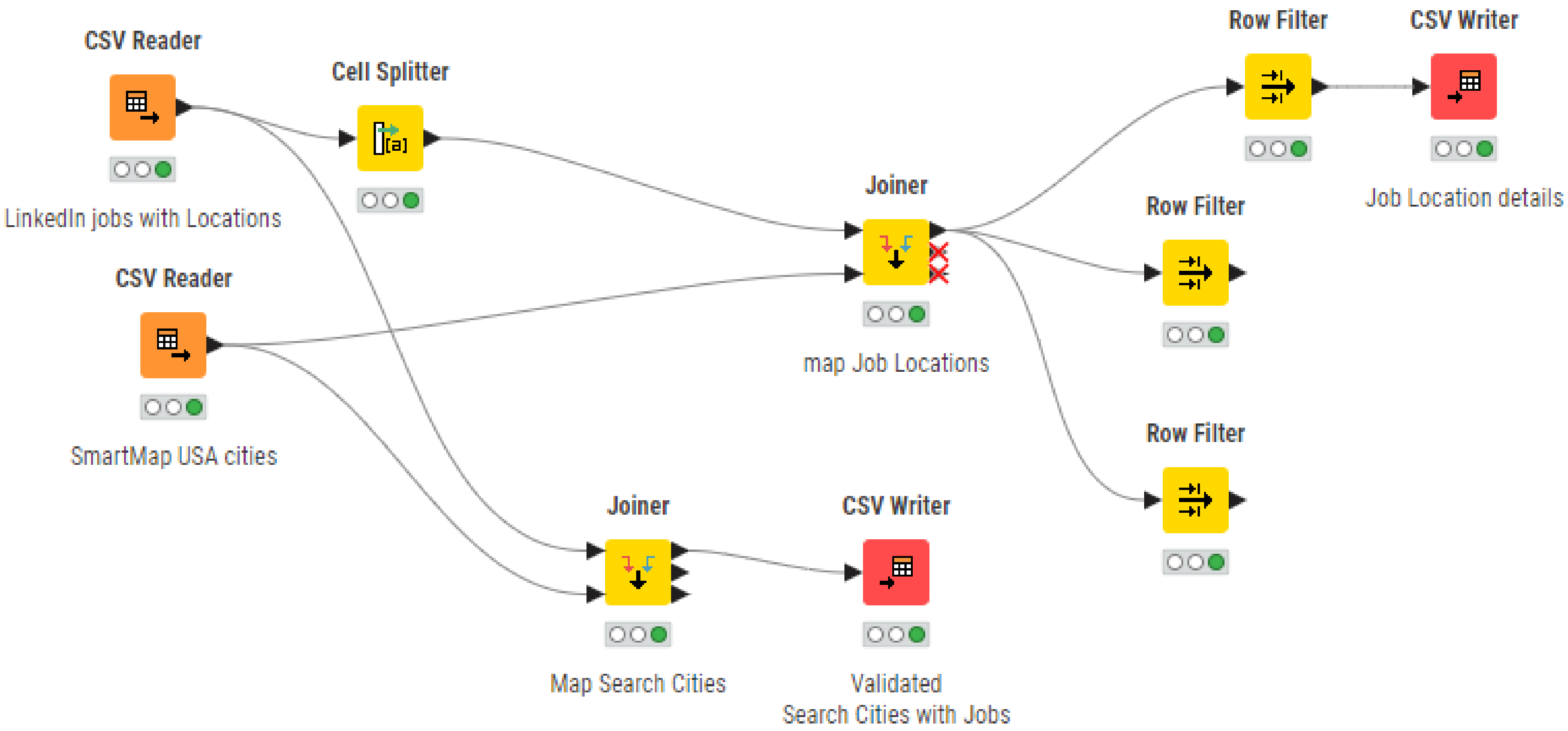
Transform skills to a column with rows



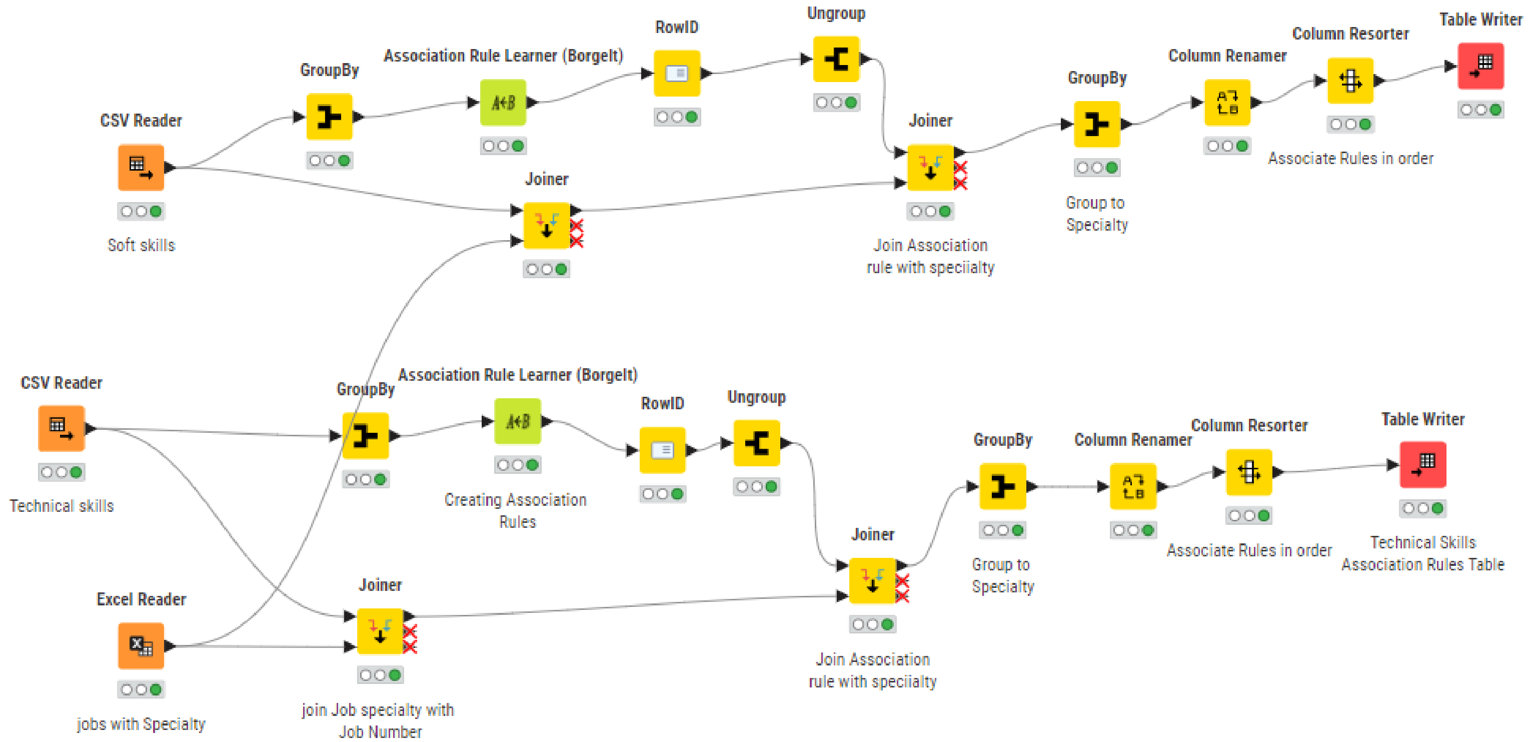
Flag soft skills with rule engine



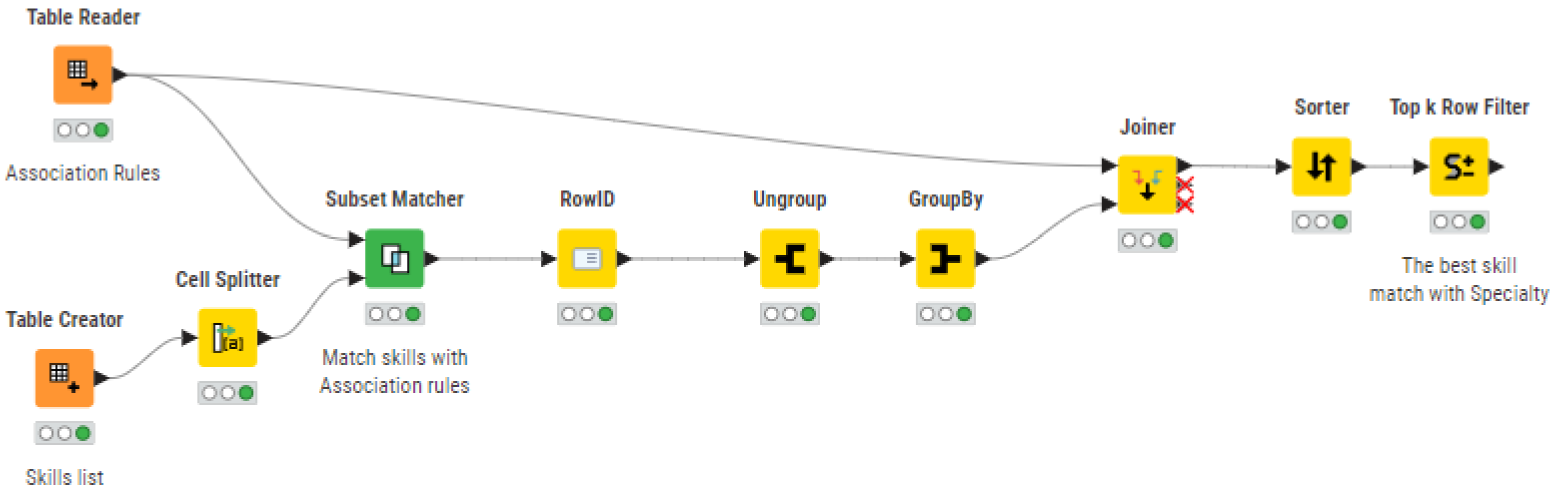
Technical skills lookup for each technical speciality



Map smartmap data with the LinkedIn job location



Association rule workflow for Job skills



Association Rules Testing

SAS codes:

```
/* Check Normality of tech skills count */

proc univariate data=work.tech_skills normaltest;
	var count_jobs;

	ods select Testsfornormality momentum;
run;

/* Non parametric test for tech  skills count */
proc npar1way data=work.tech_skills dscf wilcoxon;
	class job_specialty;
	var count_jobs;
run;

/* ANOVA test for the skills count by speciality */
proc glm data=work.soft_skills;
	class job_specialty softskills;
	model count_jobs = job_specialty|softskills /solution SS3;
	means  job_specialty|softskills /hovtest welch tukey;
	lsmeans job_specialty|softskills /adjust=tukey pdiff;
	/*means LarkOwl*Gender  /slice=Gender;
	lsmeans LarkOwl*Gender  /slice=LarkOwl;
	format gender genderF.;*/
run;

/* Check Normality of soft skills count */

proc univariate data=work.soft_skills normaltest;
	var count_jobs;

	ods select Testsfornormality momentum;
run;

/* Non parametric test for Soft  skills count */
proc npar1way data=work.soft_skills dscf wilcoxon;
	class job_specialty;
	var count_jobs;
run;
```

Mark Ushan Piyasena - Piyum001

Assignment 1

1. Mind map

Figure 1 Mind map

2. Data Dictionary

1. Jobs with Summary

1. search_city: This fluctuation catches the American city with business analysts' employment. It points up employment hotspots and locations-based job availability trends.

2. **TotalJobCount**: This variable lists every job opening for city business analysts. Combining job announcements from several cities reveals those with great business analyst demand.

3. **SoftSkills**: Many fields of work demand soft skills including communication, cooperation, and problem-solving. Using job descriptions helps to ensure uniformity in the collection.

4. **AllTechnicalSkills**: Skills in Tableau, MS Excel, data modeling, SQL relevant to occupations abound in this variable. These result from job descriptions, much as soft skills do.

5. **TotalSoftSkillsJobCount**: These show the soft skill-needed job openings for every city. Combining job ads stressing important soft skill requirements helps one to find cities with such demand.

6. **TotalTechnicalSkillsJobCount**: Here are the technical job announcements for every city. Like soft skills, it highlights cities where technological knowledge is much needed.

7. **company**: Companies hiring business analysts can be found on this page. It considers sites and hiring companies.

8. **TotalJobCountPerCompany**: These are total employment available by company in every city. It reveals which businesses most use Business Analyzes, so enabling research on employment patterns particular to every business.

9. **job_title**: List business analysts, management analysts, etc., depending on this variable employment advertising. This guides which employment titles would be appealing to particular cities.

10. **search_position**: This variable in the dataset looks at job openings. Classifications of employment tasks help to distribute market demand among positions.

11. **job_level**: This mark associates, mid-senior, etc. position. It looks at how work is impacted by geography, titles, and skills.

12. **JobSkills**: This variable comprises of soft and technical skills pertinent to employment. Combining relevant job advertising skill sets demonstrates the skills valued at different career levels and job titles.

2. Purpose of Variables in Visualizations

I. **Overall Job Distribution by City**: Searches TotalJobCount and Search City for positions in big cities as business analysts.

II. **Soft Skills and Technical Skills Distribution**: Using search_city, SoftSkills, AllTechnicalSkills, TotalSoftSkills Job Count, and Total Technical Skilled Job Count, present job searchers and organizations regional skill needs.

III. **Company and Position Analysis**: Title for the job, search position, TotalJobCountPerCompany, company shows which businesses employ more and which job openings are more common in different cities. This information may be helpful for job seekers who target certain firms and organizations benchmarking their hiring requirements.

IV. **Job and Skill Distribution Using Box Plots**: Finds city job and skill distribution consistency and stability using search_city, company, job_title, SoftSkills, AllTechnicalSkills, and TotalJobCount. Knowing market dynamics helps one to identify threats and opportunities.

V. **Flow of Job Levels, Skills, Titles, and Locations**: Job levels, skills, title, and search city view displayed by Job_level, JobSkills, Job_title, and Search_city. This career path and regional trend expression assists both strategic career planning and talent acquisition.

These comprehensive definitions and goals show how methodically structured data analysis uses every variable to deliver job seekers, businesses, and teachers' relevant information.

3. Relations in the Data

The related information in the infographics emphasizes the employment opportunities for business analysts. These are some significant interconnections among the several criteria:

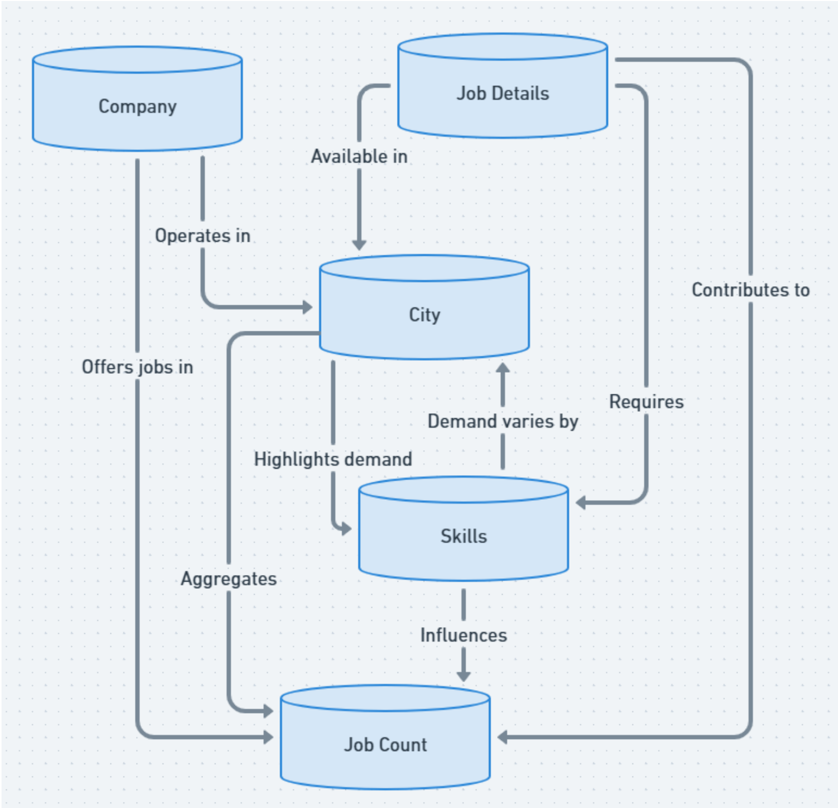


Figure 2 the relationships between the key elements in the Business Analyst job market data

1. **City (search_city) and Job Count (TotalJobCount)**:

Relation: Search_city and TotalJobCount are connected as job announcements combine city job counts. The link identified cities with most business analyst demand.

Usage: This link links "Overall Job Distribution by City" visuals' job hotspots.

2. **City (search_city) and Skills (SoftSkills and AllTechnicalSkills)**:

Relation: Job advertisements differ in soft and technical capacity based on the city. Search_city reflects local skill need by all technical ability as well as soft skills.

Usage: Through visualizations such as "Soft Skills and Technical Skills Distribution," this relationship shows regional skill requirements, therefore helping job seekers to pinpoint highly sought-after abilities in particular cities.

3. **Company (company), City (search_city), and Job Count (TotalJobCountPerCompany)**:

Relation: TotalJobCountPerCompany records city-based company job offers. This site allows you to locate leaders in regional job markets.

Usage: Visuality Job seekers can apply to highly sought-after organizations by knowing which businesses have the most job ads in each city using "Company and Position Analysis".

4. **Job Title (job_title), Position (search_position), and Job Count (TotalJobCount)**:

Relation: Whereas the job title and search position list displays, TotalJobCount shows city role frequency. This link enables one to estimate the popularity and demand of titles and roles.

Usage: Helping job searchers find employment, the "Top Job Positions by City" graphic displays the most sought-after positions across cities.

5. **Job Level (job_level), Skills (JobSkills), and Job Title (job_title)**:

Relation: There are several degrees of job skills. Changes in title and skill characterize career progress.

Usage: Strategic career advice is given via the Sankey diagram "Flow of Job Levels, Skills, Titles, and Locations" which displays employment advancement and necessary abilities.

6. **Skills (SoftSkills, AllTechnicalSkills) and Job Distribution (TotalSoftSkillsJobCount, TotalTechnicalSkillsJobCount)**:

Relation: Every technical ability and soft talent meets business needs, which affects the employment numbers in cities. Technical skills count; soft skills count; blend demand here.

Usage: Visualizations of skill distributions among cities help stakeholders see the worth of various skills in different marketplaces.

Summary

The data presents a full picture of the business analyst job market by showing relationships between employment locations, companies, job titles, and relevant skills. These conversations reveal crucial information that will enable companies, educators, and job seekers to have an impact. These links offer targeted visualizations of employment market dynamics and stakeholder strategies.

3. Results of Exploring the Data

1. Overview of Data Exploration

Introduction

Examining the data helped one to understand American city business analyst job trends. This study searched LinkedIn job listings for companies, teachers, and job seekers in order to identify regional trends, skill demand, and employment availability. The top 20 American cities with the most job ads were under analysis. This enables focused study in major geographic areas of demand for business analysts.

Data Filtering and Scope Definition

Emphasizing relevancy, the dataset focused on US job advertising. Original worldwide employment announcements decreased to American opportunities. More research revealed the top 20 cities generating employment. This criterion focused on the labor market in big cities, where business and industry dominant and Business Analyses are much sought after. This study shows that these cities may present the finest opportunities for job seekers.

Examination of Key Variables

looked at titles, regions, soft and technical skills, job levels. These attributes were supposed to be pertinent for the job scene.:

Job Level: Experience directs job classification as Associate, Mid Senior, or Senior. Job levels mirror city need for experienced and entry-level employment.

Job Title: Job titles run from Management Analyst to Business Analyst to Data Analyst. Which employment titles are most in demand could direct job seekers into a line of work.

Job Location: The location variable notes state and job posting cities. This allowed one to map job opportunities and business analyst demand cities.

Skills: Tableau, data modelling, SQL removed soft skills such communication and problem-solving from one other. Examining these competencies indicated talents valued by companies worldwide for business analysts.

Data Cleaning and Standardization

To ensure analytical correctness, data was meticulously standardized and cleansed. Duplicates were removed to stop slant of data. Crucially, also, was managing missing values, which can skew results. Depending on the degree of missing data, data imputation or elimination was addressed.

Geographic criterion and title first took front stage. Job titles were standardized using name standards ("Business Analyst" vs. "BA") to maintain dataset consistency. Standardizing city and state names and removing misspellings standardized employment sites. This enabled location-based sorting and evaluation of employment advertising.

Conclusion

The numbers reveal interesting U.S. city business analyst job markets. The analysis revealed notable trends and patterns by way of key elements and rigorous data cleaning and standardizing. This study will provide for locales suitable for Business Analyst-friendly distribution, skill demands, and employment level distributions. Negotiators of the evolving corporate analytics employment market have to be aware of this.

2. Key Visualizations and Findings

1. **Visualization 1: Overall Job Distribution by City.**

Description

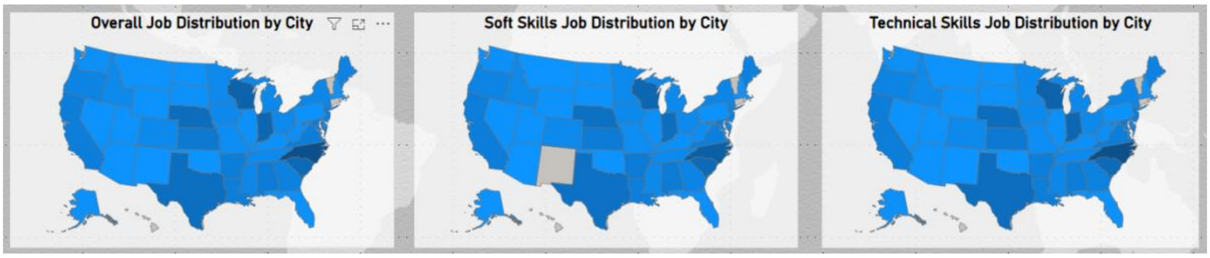


Figure 3 Overall Job Distribution by City

The choropleth map shows American city business analysts' jobs. The map displays job numbers in blue, with darker tones signifying more listings. This graphic tool indicates expanding areas where Business Analyses could quickly and organically uncover prospects.

Why It Was Done

This graphic shows job availability by business analyst demand hotspots and area. This let firms, businesses, and job seekers find hotspots for Business Analyses. The map identifies highly promising cities, thereby streamlining market strategy, job search, and relocation.

This map helps businesses, schools, and job seekers locate rivals' recruiting locations and modify their courses to meet the demand of the market. It clearly shows where to draw talent or motivate pupils to seek employment.

Result

The map "Overall Job Distribution by City" shows Montpelier and Columbus to have the most Business Analyst jobs. Montpelier is smaller, yet it has more job advertising than IT hotspots. The deep blue map colors of these sites suggest strong local demand for business analysts, suggesting they could be building centers.

Among larger, more developed cities with significant employment numbers, Montpelier and Columbus stand out for showing job diversity. As distant work grows more prevalent, local economic policies, new firms, or decentralization in the employment market could bring about this transformation.

Interpretation

The "Overall Job Distribution by City" infographic pushes one to look outside of New York, San Francisco, and Chicago. Top contenders Montpelier and Columbus offer great opportunities away from the competitiveness and expensive living of larger cities. Starting to take front stage are flexibility and remote or relocation work.

Montpelier is small, but its large workforce suggests a focused industry or regional shift towards data-driven business strategies. Montpelier can show a concentration of companies with high business analytics capacity, thereby increasing demand for business analysts. State or municipal government tech company incentives could also help to explain the job growth.

The findings direct efforts in corporate talent acquisition. Companies may find growing IT centers intriguing since growing IT centers may have less talent competition than more established ones. Local business analysts could reduce the demand for relocation incentives or major remote work for the corporations in these sectors.

Schools around these hotspots should modify their Business Analyst courses to match demand and prepare graduates for local jobs. Integration of industry and education can increase graduate employability and regional job generation.

The shocking "Overall Job Distribution by City" map of big cities with considerable Business Analyst demand compels study of the optimum locations. A dynamic approach to business analytics job hunting and recruiting would benefit growing cities like Montpelier and Columbus.

2. visualization 2: Soft Skills and Technical Skills Distribution

Description

Various primary renderings of the "Soft Skills and Technical Skills Distribution" visualization show U.S. city soft and technical skill demand. Visual assistance is:

1. **Figure 4: Top Cities for Soft Skills Demand**: On a bar chart showing soft skill job ads, Montpelier, Atlanta, and Santa Clara placed top.

Figure 4 Top Cities for Soft Skills Demand

2. **Figure 5: Soft Skills Distribution by City and Skill Type**: City: soft skill development in communication, teamwork, and problem-solving treasure map Every city values soft talents, as the treemap reveals.

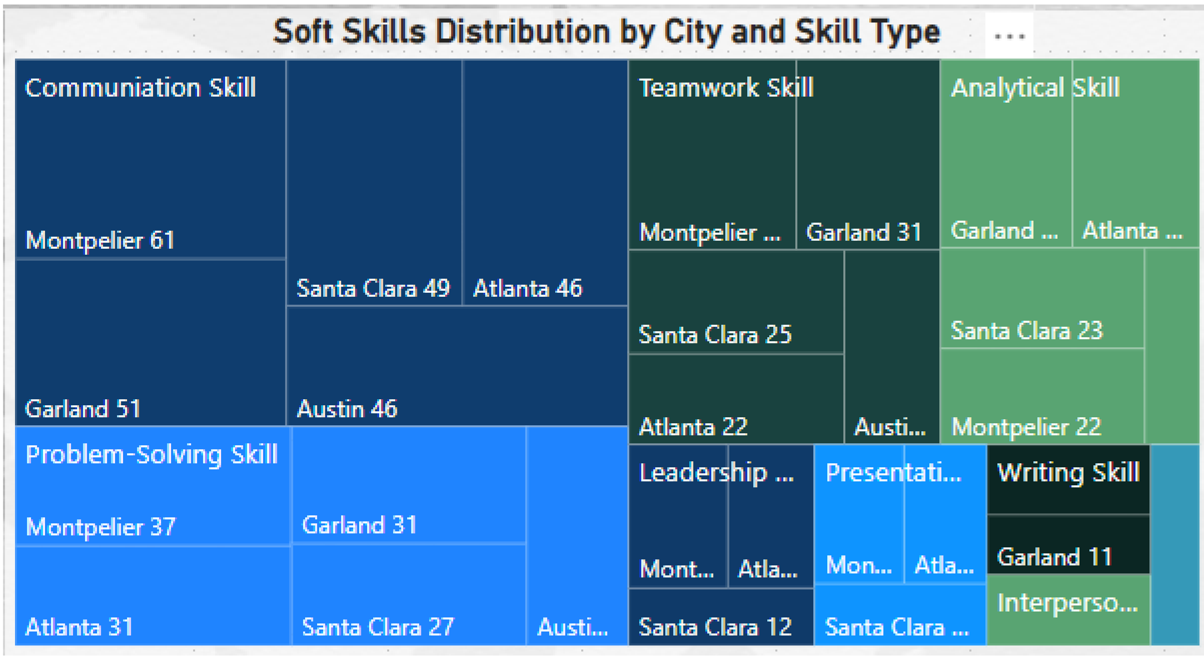


Figure 5 Soft Skills Distribution by City and Skill Type

3. **Figure 6: Top Cities for Technical Skills Demand**: Technical skill job announcements for Santa Clara, Garland, and Summit displayed on a bar chart.

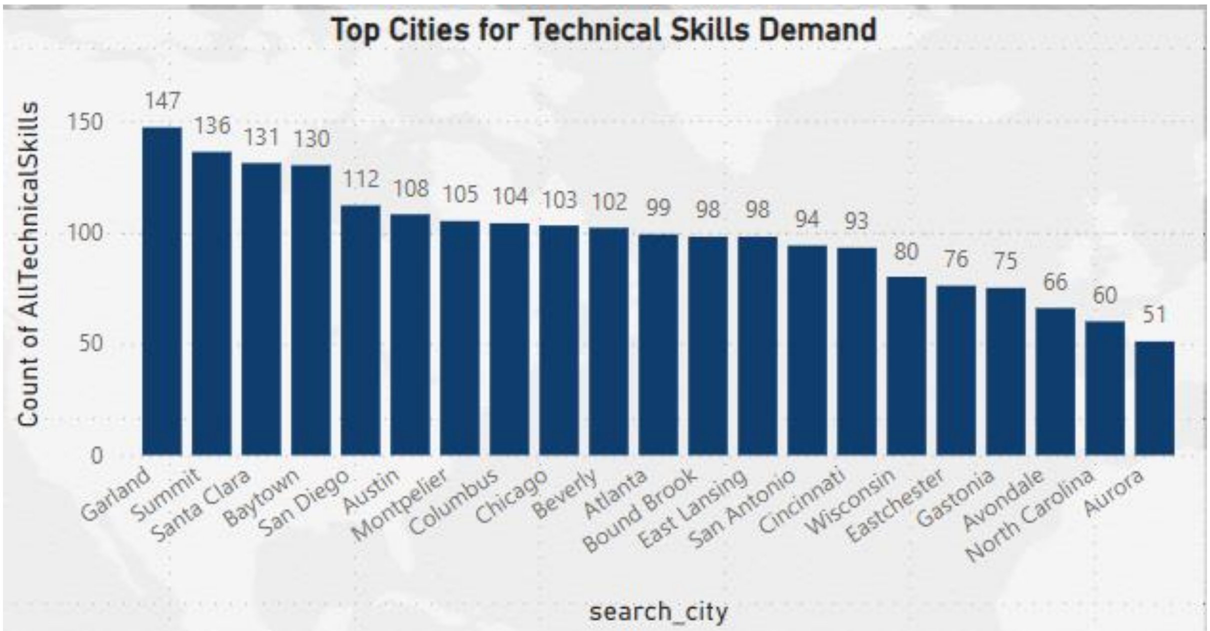


Figure 6 Top Cities for Technical Skills Demand

4. **Figure 7: Technical Skills Distribution by City and Skill Type**: a treemap showing Python application, data modelling, city SQL tools like Tableau. Every city shown here prioritizes technology.

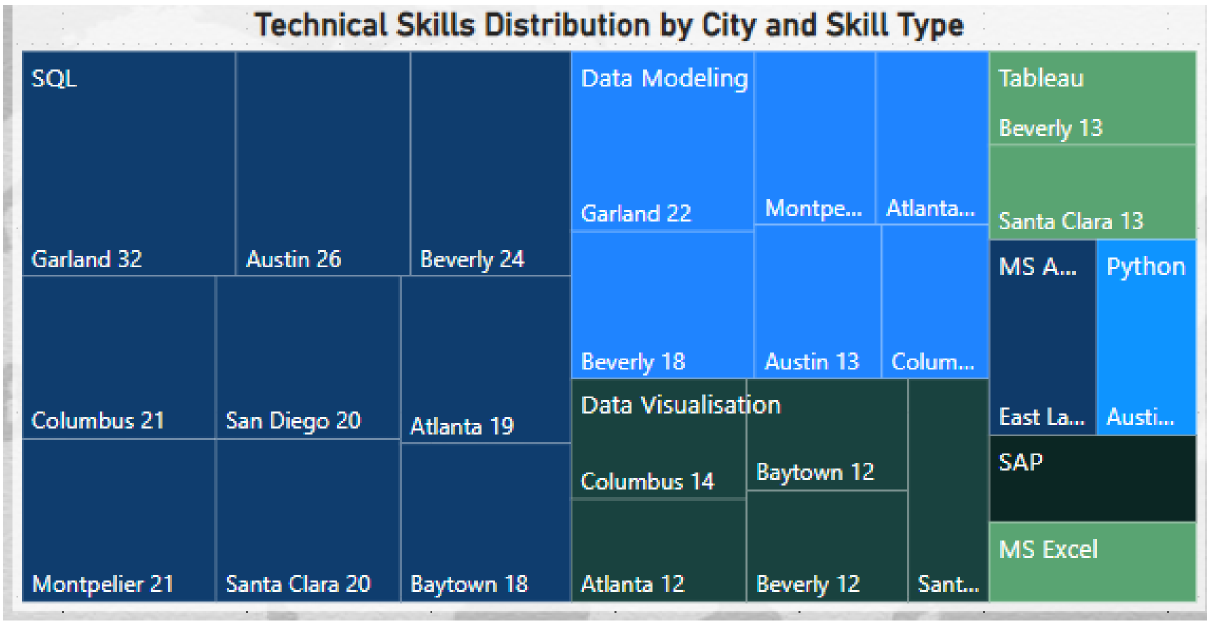


Figure 7 Technical Skills Distribution by City and Skill Type

Why It Was Done

These infographics highlight geographic differences in the skills organizations want. Treemaps and bar charts showed the different business analyst degrees in different cities. This information helps businesses and job seekers to identify shortages and surpluses in the chosen spheres of activity.

Result

Many American cities stress communication, problem-solving, and teamwork, according figures 2 and 3 of the "Soft Skills and Technical Skills Distribution" visualizations. Figures 4 and 5 feature professionals in visualization, data modelling, and SQL drawing from Garland, Santa Clara, and San Diego.

Interpretation

The pictures stressing the need of soft and technical skill balance in "Soft Skills and Technical Skills Distribution" show Business analysts have to be highly analytical and good communicator since they are needed everywhere in the sector and in different cities. Demand for technical skills linked to data indicates growing interest on data-driven decision-making in business analytics.

These skills enable job seekers to thrive in quite demanding employment settings. Demand for these skills is always present in tech hubs as well as in developing towns, thereby creating more possibilities. Knowing these trends would help businesses choose outstanding cities and boost employment. These discoveries help educational institutions to adapt their curricula to the needs of the market and equip graduates for employment.

3. Visualization 3: Company and Position Analysis

Description

The "Company and Position Analysis" visual aid, bar charts and treemaps, present U.S. city employment openings by business and position. These infographics show companies hiring, job seekers, businesses, and academic institutions exhibiting fields most in demand.

1. **Figure 8: Top Companies by City**: On a bar chart, leading business analyst job centres Montpelier, Garland, and Atlanta.

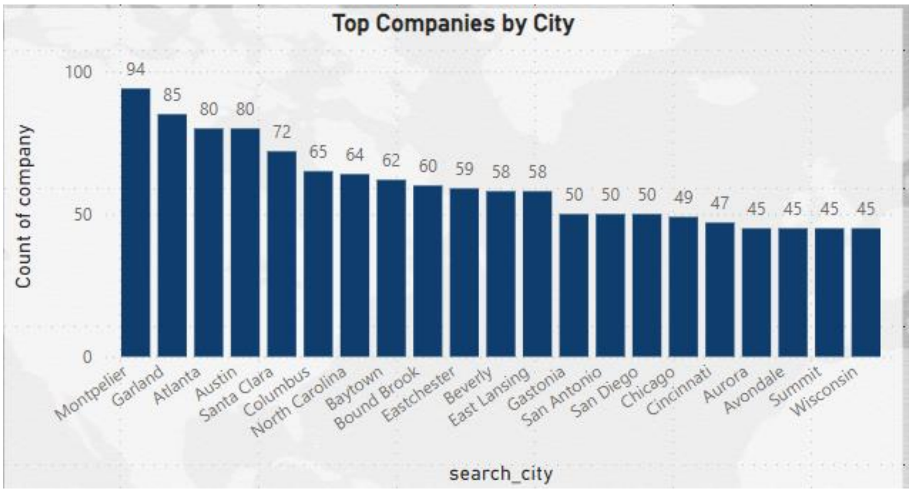


Figure 8 Top Companies by City

2. **Figure 9: Job Distribution by Company and City**: This treemap shows corporate job distribution over cities. For Santa Clara, Montpelier, and East Lansing, dice, prohires, and winmax rule.

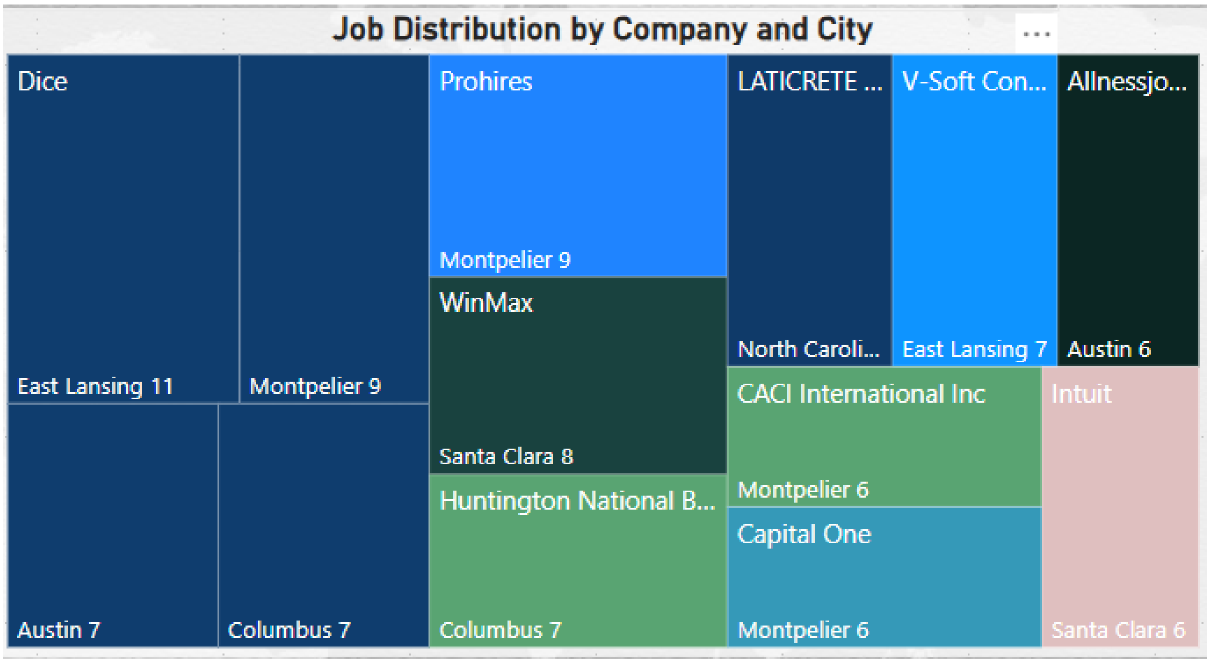


Figure 9 Job Distribution by Company and City

3. **Figure 10: Top Job Positions by City**: Among business and management analysts, a bar chart shows Montpelier, Garland, and Atlanta ranking highest.

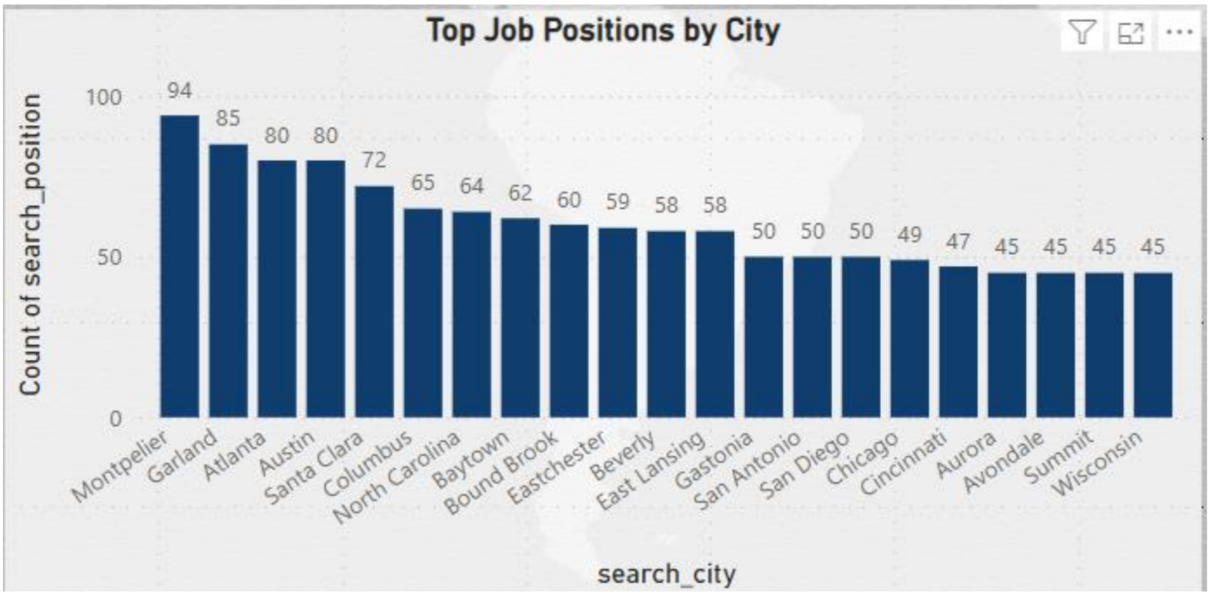


Figure 10 Top Job Positions by City

4. **Figure 11: Job Distribution by Position and City**: a city treemap showing the most regularly occurring employment in the disciplines of Management Analyst, Job Analyst, and Consultant Economist.

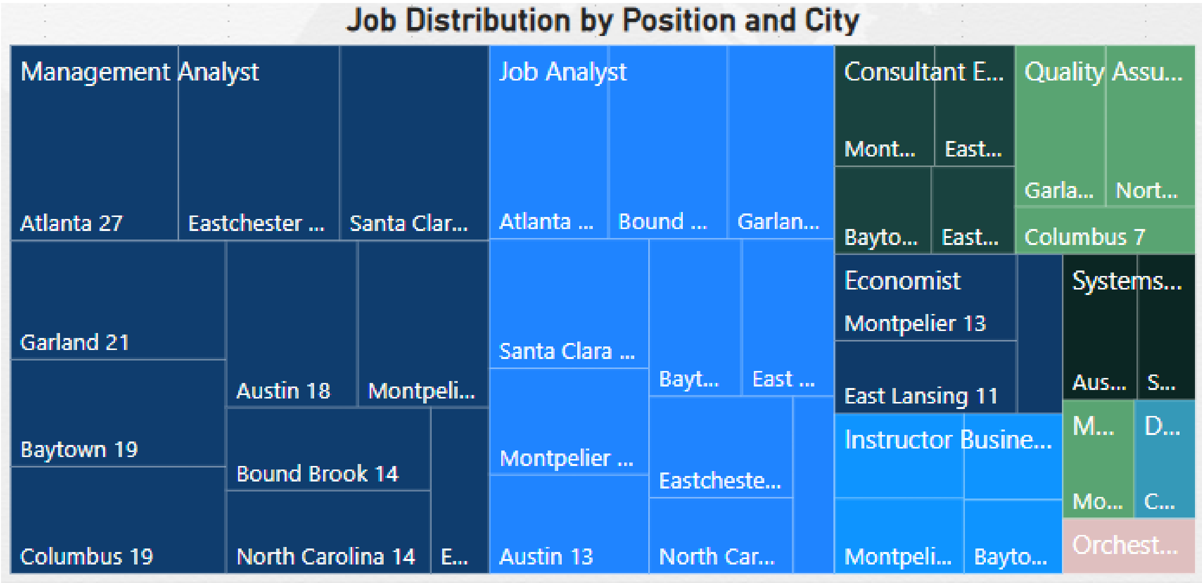


Figure 11 Job Distribution by Position and City

Why It Was Done

These infographics show which companies pay more for which jobs are most sought for. Big businesses and key job titles were evaluated in order to present a full picture of Business Analyst employment. Both companies assessing their hiring policies against competitors and job seekers targeting at specific companies or positions depend on this knowledge.

Result

Figures 6 and 7, "Company and Position Analysis," show Dice and Prophies had the most job openings. Among the notable East Lansing and Montpelier companies were these ones Figure 8 shows in numerous cities Management Analyst and Job Analyst as the most in-demand occupations. Focused in certain cities, jobs like consultant economist and teacher business analyst show local demand for specialized knowledge (Figure 9's treemap).

Interpretation

Looking at "Company and Position Analysis" benefits various interested parties:

For Job Seekers: Knowing top firms and most sought-after positions helps job seekers focus. Aiming for well-known companies like Dice and Prophies or popular professions like Business Analyst and Management Analyst will help job seekers find suitable work.

For Companies: These disclosures help companies to match their recruiting needs to competitors. By knowing which companies control job markets in various cities, companies will be able to draw in superior staff. Businesses should investigate less competitive markets or offer unique value to attract customers.

For Educational Institutions: These discoveries will help educational institutions to modify their courses to meet the needs of big companies, therefore ensuring that graduates have the required credentials and skills. Knowing popular career routes and hiring firms helps institutions educate their graduates for the workforce.

The "Company and Position Analysis" visualizations provide job market activities of business analysts, therefore providing recruiting and career strategy data. Focusing on data specific to businesses and roles helps find and grab prospects.

4. Visualization 4: Job and Skill Distribution Using Box Plots

Description

"Job and Skill Distribution Using Box Plots" presents U.S. city job statistics by employer, position, and skill type using multiple box plots. Analyzing job availability range, median, and outliers helps one to be aware of employment market stability and volatility in different places.

1. **Figure 12: Job Count Distribution by City and Company**: a box map with Montpelier, Columbus, and Bound Brook's corporate job counts. This perspective highlights cities with more job opportunities and outlays.

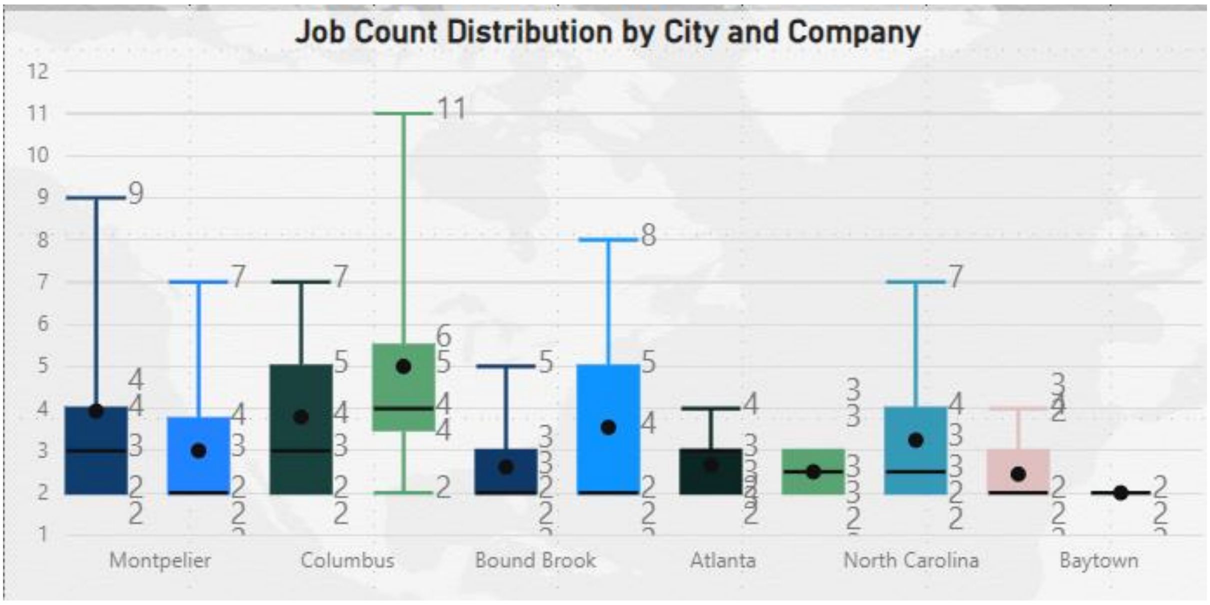


Figure 12 Job Count Distribution by City and Company

2. **Figure 13: Job Count Distribution by City and Position**: Box maps for Montpelier, Austin, and Eastchester displaying job counts by position. Demand for business analysts and management analysts is consistent in many different sectors, as this data reveals.

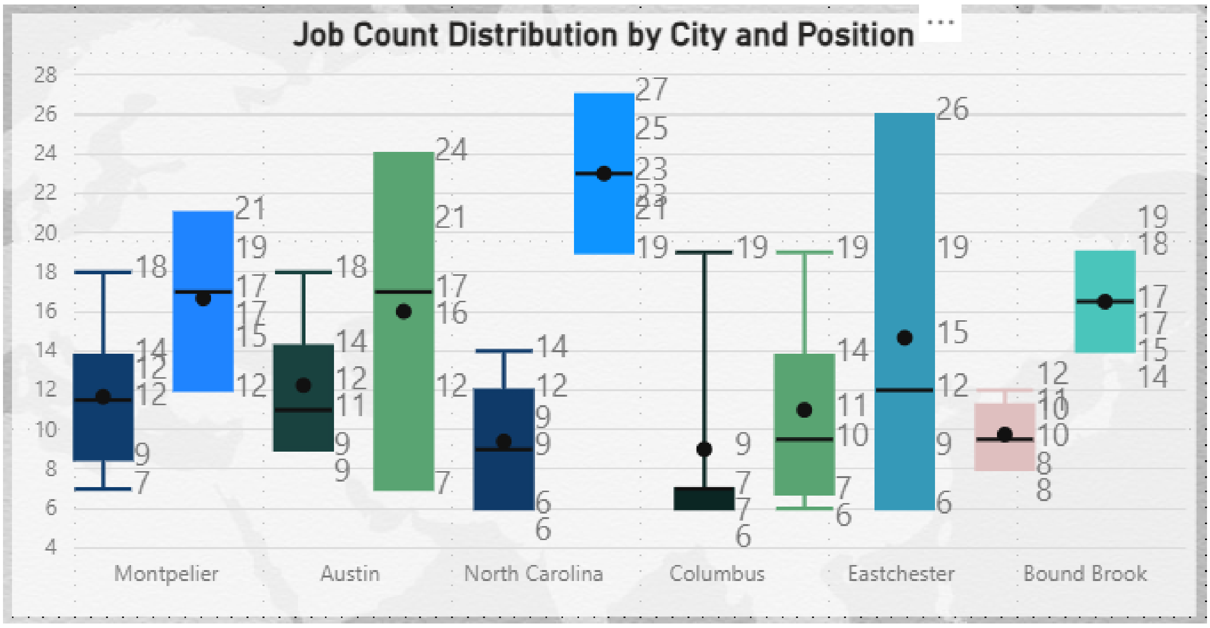


Figure 13 Count Distribution by City and Position

3. **Figure 14: Soft Skills Job Count Distribution by City and Skill Type**: Box map of soft skill employment by city and skill type includes communication and teambuilding. The differences in Montpelier, Atlanta, and Garland talent point to changing corporate needs.

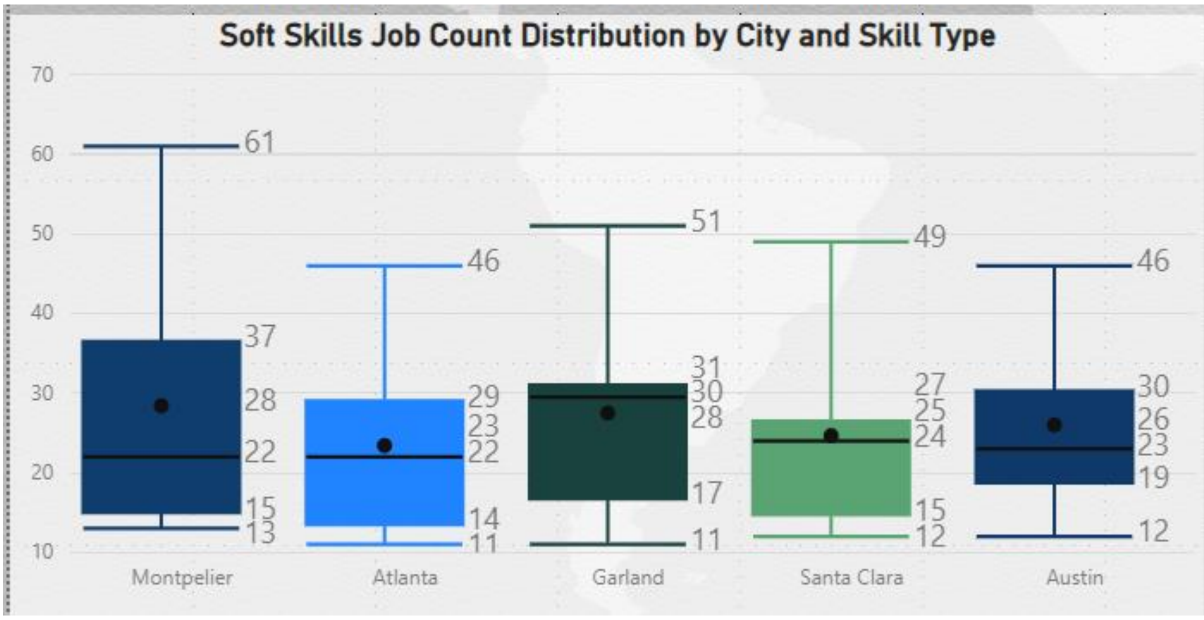


Figure 14 Soft Skills Job Count Distribution by City and Skill Type

4. **Figure 15: Job Skills Count Distribution by City and Skill Type**: Box maps of Montpelier, Atlanta, and Garland job counts by skill—technical and analytical as well. Reflecting

changing job markets, this graphic shows maximum variability and demand for skill sets.

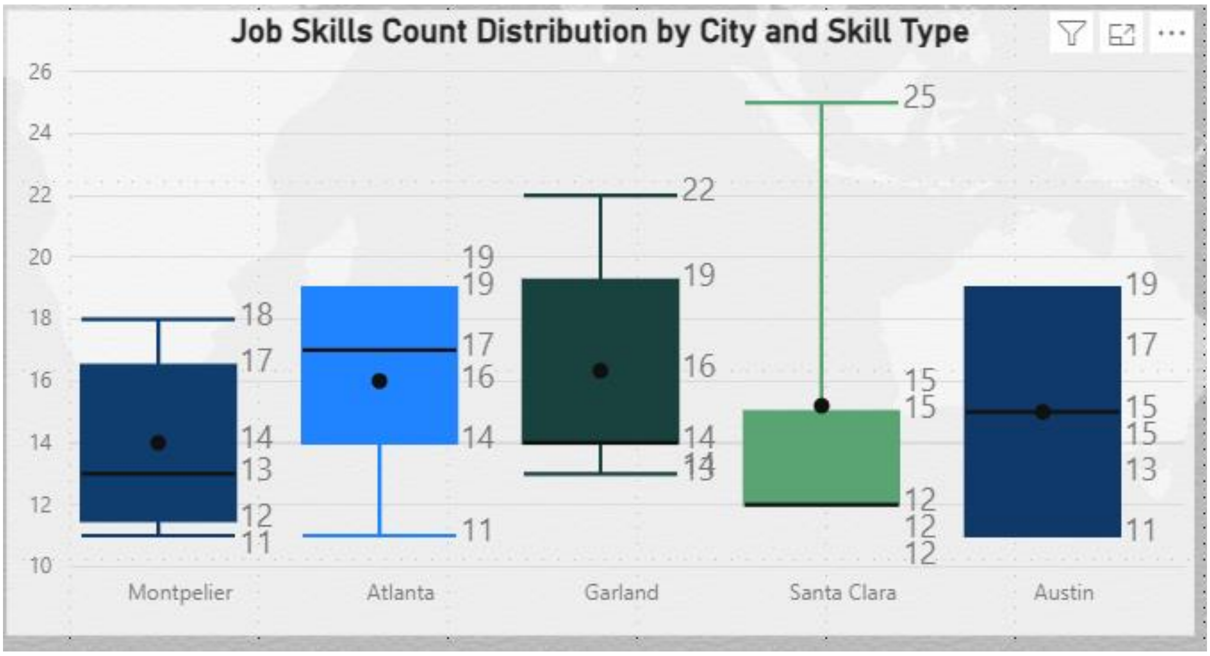


Figure 15 Job Skills Count Distribution by City and Skill Type

Why It Was Done

Cities' job counts by company, position, skill type was compared using box graphs. Availability of either steady or chaotic labor could affect the strategy of businesses and job seekers. Analyzing the range and outliers of these distributions helps local job market hazards and opportunities to be found by stakeholders.

Result

Visualizations of "Job and Skill Distribution Using Box Plots" exposed notable variances in job count for specific towns and skill types (Figures 10–13). Position-specific job counts mean Columbus and Eastchester have somewhat distinct job availability. Garland and Santa Clara showed notable variations in soft and technical skill demand, suggesting that both economies had evolving skill needs.

Interpretation

The "Job and Skill Distribution Using Box Plots" suggest that cities with high employment numbers could expose job seekers to irregular options and hazards. Different job availability could affect the strategy used by job seekers. Columbus can go through significant downturns even with so many opportunities; she has many businesses and employment.

Those looking for jobs could be able to manage unstable employment markets by being flexible and considering many cities. Understanding the fluctuations in the job market helps businesses decide what to hire and project.

Emphasizing the importance of adaptability and market awareness in employment and career choices, the "Job and Skill Distribution Using Box Plots" visualizes show the job market.

5. Visualization 5: Flow of Job Levels, Skills, Titles, and Locations (Sankey Diagram)

Description

Views perspective of view Sankey graphs let "Flow of Job Levels, Skills, Titles, and Locations" display job levels, skills, titles, and locations. This picture illustrates how employment levels—that of an associate, mid-senior—translate into specific competencies linked with titles and positions. These paths, as the Sankey graphic demonstrates, highlight business analysis career development and skill needs.

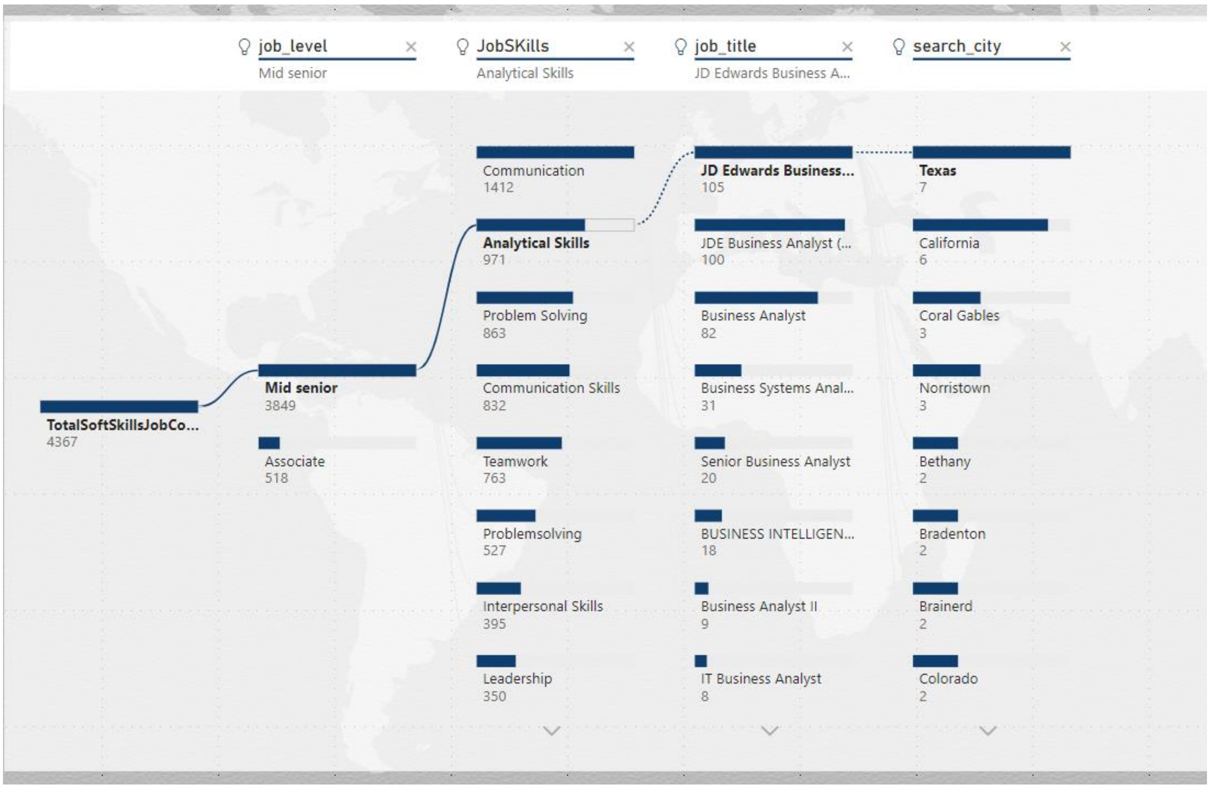


Figure 16 Flow of Job Levels, Skills, Titles, and Locations (Sankey Diagram)

Job Levels: Associate and mid-senior roles in this Sankey diagram match sites, titles, and skill sets. It shows how degrees and professional pathways shift in response to analytical capabilities and communication.

Skills: analytical job title; communication; ability to solve problems; This graphic shows skills of Senior Business Analyst and Business Analyst.

Job Titles: Talents align in this graph for positions including JD Edwards Business Analyst, Business Systems Analyst, and IT Business Analyst. Higher level employment requires enhanced flow stresses improving capacities.

Locations: Third phase links job titles to U.S. states including Texas, California, and Colorado as well as city names. It reveals where several jobs are most common.

Why It Was Done

The Sankey diagram illustrates the links in the employment market between job levels, skills, titles, and locations of business analyst employment. In order to provide career advice, we underlined how strengthening various talents might lead to specific job titles and opportunities in particular sectors. By helping teachers, businesses, and job seekers to evaluate how abilities affect employment outcomes, this visualization helps to identify development and emphasis areas.

Result

Skill-job level links ran routinely throughout titles and sites in the Sankey graphic "Flow of Job Levels, Skills, Titles, and Locations." Figures mid-senior positions—which ranged from Senior Business Analyst to JD Edwards Business Analyst—often called for analytical skills. Figure demonstrates how job market dynamics and industry concentrated these positions in Texas and California.

Interpretation

The Sankey diagram "Flow of Job Levels, Skills, Titles, and Locations" illustrates business analysis career growth.:

Career Progression: The interaction among job levels, competencies, and titles highlights the need of skill development in professional advancement. Mid-senior and specialty employment might come from good analytical skills and communication.

Geographical Trends: Map of job titles and locations reveals the concentration of various roles in particular localities. Those seeking these talents could benefit from concentrating on technological hotspots like Texas and California since they provide analytical and technical jobs.

Strategic Skill Development: Understanding of how employment levels, skills, and locations relate benefits career planners. Analyzing problems and approaching challenges can enable you to develop both on travel and in your work.

Sankey diagrams for "Flow of Job Levels, Skills, Titles, and Locations" indicate locations, job duties, and skills. Actionable insights help to match markets for recruiting, career planning, and educational initiatives.

3. Reflections on the Results

1. General Insights

A market study revealed for business analysts' significant cities and strengths. Montpelier and Columbus are looking for business analysts. This emphasizes the significance of having a complete set since soft and technical skills are required.

Needed skills and city demand are listed below:

Table 1 Needed skills and city demand

City	Top Skills Required	Demand Level
Montpelier	Communication, Data Modeling, SQL	High
Columbus	Analytical Skills, Problem-Solving	High
Austin	Teamwork, Data Visualization	Moderate
Santa Clara	Leadership, Technical Writing	Moderate
Atlanta	Presentation Skills, Power BI	Moderate

Key Insights from the Data:

1. City-Specific Demand: The great demand for Montpelier and Columbus business analysts suggests possible expansion in the industry. Often disregarded in favor of bigger cities, these little towns offer decent jobs in less saturated markets.
2. Skill Balance: Businesses seek soft as well as technological skills. Top skills were always communication and data modelling, which emphasizes the great value of technical expertise and precise presentation of difficult material.
3. Interpersonal and Technical Competencies: Business analysts have to be adaptive in SQL, Power BI, leadership, and problem-solving. Their capacity to balance numerous roles qualifies them as important in many different fields.

Recommendations for Job Seekers and Stakeholders:

For Job Seekers: Choose soft and technical talents to work in different cities. More opportunities will be opened by continuous training as well as Power BI and SQL certificates.

For Employers: Montpelier and Columbus have low skill competitiveness but big demand, hence hire there. This approach can boost hiring and acquisition of talent.

For Educators: Curricula have to contain soft and technical abilities. Simulations and projects let students be more ready for different employment environments.

This data demonstrates the variations in business analyst employment. Making industry-relevant decisions requires companies, job seekers, and educators to grasp these trends.

2. **Surprising Findings**

Data show Montpelier has a lot of jobs for business analysts. While some say New York and San Francisco control this field, other cities also generate employment. The apparently broader labor market than expected allows professionals to grow outside of IT hubs.

Key Findings on City Prominence:

Table 2 Key Findings on City Prominence

City	Expected Prominence	Actual Demand	Key Skills in Demand
Montpelier	Low	High	Communication, Analytical Skills
Columbus	Moderate	High	Problem-Solving, Data Modeling
New York	High	Moderate	Business Analytics, Technical Writing
San Francisco	High	Moderate	Data Visualization, Leadership
Santa Clara	Moderate	High	Technical Skills, Teamwork

Analysis and Implications:

1. **Unexpected Prominence of Smaller Cities:** Here surprise came from Montpelier, a city not known for business analysts. This suggests that smaller towns could have competitive job markets and growing companies with value for corporate analytics.
2. **Competitive Advantage in Smaller Markets:** Perhaps more than smaller cities, big cities could compete. Job seekers could so look at occupations in fields with less competition but high demand.
3. **Strategic Considerations for Job Seekers:** Job seekers should go outside traditional sites given this dispersion. Smaller cities could provide better chances for professional growth away from the saturation of large metropolis.

Recommendations:

Job Seekers: Apply for high demand Montpelier and Columbus jobs with low competition. This strategy can help to create career options and skill-matched employment prospects.

Employers: Less saturated are smaller city talent pools. In fields where business analytics is in great demand, businesses might consider remote or satellite operations.

Policy Makers and Urban Planners: Regional development strategies should be guided by knowledge of the employment market potential of smaller towns in order to attract outstanding experts.

Conclusion: The research challenges the notion that business analysts operate just in big cities. Seeing the growing value of smaller cities helps interested parties locate opportunities, personnel, and make regional development projects possible.

3. Implications for Stakeholders

Business analyst roles affect firms, job seekers, and education as well. Data analysis reveals strategies designed for each category aiming to improve outcomes.

1. For Job Seekers

The findings highlight the acquisition of soft and technical competencies. Learning new skills can increase job possibilities in non-traditional IT industries. A demand-based table highlights areas of important ability and location:

Table 3 important ability and location

Key Skills	Recommended Cities	Rationale
Communication, Teamwork	Montpelier, Santa Clara	High demand for soft skills in smaller cities
Analytical, Data Modeling	Columbus, Austin	Strong technical skill requirements
Problem-Solving, SQL	Garland, Baytown	Demand for balanced skill sets

Recommendations for Job Seekers:

Explore Beyond Traditional Hubs: Montpelier and Columbus both require business analysts, even if there isn't much rivalry.

Skill Development: Learning data modelling and communication are two absolutely essential universal abilities.

2. For Employers

Data guides one to maximize hiring. Recruiting remotely or from smaller towns helps to reduce skills shortages and recruiting costs. The following compiles methods of recruitment:

Table 4 compiles methods of recruitment

Strategy	Potential Benefits	Target Locations
Remote Hiring	Access to a broader talent pool, reduced overhead	Smaller cities with high demand
Regional Expansion of Recruitment	Tapping into untapped markets, increased diversity	Montpelier, Columbus
Focus on Skill-Based Hiring	Greater alignment with job requirements, better retention	Cities with specific skill demands

Recommendations for Employers:

Expand Recruitment Scope: Consider Montpelier potential, where less competitive but plenty of skill resides.

Emphasize Remote Work Opportunities: This approach increases agility, knowledge, and variety.

3. For Educators

Educational institutions have to provide Business Analyses ready for changes in the job environment. Training should thus cover both soft and technical abilities since both are equally vital. The curriculum has evolved seen in the table below:

Table 5 The curriculum has evolved

Curriculum Focus	Skill Sets to Include	Educational Outcomes
Balanced Soft and Technical	Communication, Analytical, Problem-Solving	Preparedness for diverse job requirements
Real-World Application	Project-Based Learning, Case Studies	Enhanced practical understanding and skills
Continuous Skill Updates	Emerging Tools and Technologies	Keeping pace with market trends

Recommendations for Educators:

Develop Balanced Curricula: Integrate soft skills such SQL, data modelling, and communication in courses.

Emphasize Practical Learning: Make use of case studies and employment market-relevant projects.

Conclusion

Those applying this method have to be adaptable and strategically focus on education, hiring, and skill development. Matching strategy with data helps employers, job seekers, and teachers negotiating the evolving Business Analyst job market.

4. Next Steps

Focusing on specific issues in further business analyst job market evaluations helps to provide more comprehensive insights and better strategic recommendations for stakeholders. Suggested are the following phases of study and areas of interest:

1. Explore Time Trends in Skill Demand

Job market trends in skill demand could show either developing or declining talents over time. The study can enable interested parties to project future demands and change their plans of action.

Table 6 Explore Time Trends in Skill Demand

Analysis Focus	Objective	Potential Outcome
Time Series Analysis	Track changes in demand for specific skills over time	Identification of emerging skills
Seasonal Demand Patterns	Examine seasonal fluctuations in job postings	Insight into cyclical hiring trends
Skill Evolution	Analyze the rise or decline of particular skills	Strategic focus on future skill development

2. Investigate Salary Data Correlations

Pay and employment demand assist one to grasp the dynamics of the market. Job advertising and salary data expose how differently businesses and sectors value skills and employment.

Table 7 Investigate Salary Data Correlations

Analysis Focus	Objective	Potential Outcome

Salary vs. Skill Demand	Correlate job demand with salary ranges	Insights into how skill demand impacts compensation
Regional Salary Analysis	Compare salaries for similar roles across different cities	Identification of high-paying regions
Role-Specific Salary Trends	Investigate salary variations by job title	Understanding the financial incentives of specific roles

3. Expand Geographic and Sectoral Analysis

Seeing different sites or companies helps to define niche markets. This benefits people seeking specialism or diversification.

Table 8 Expand Geographic and Sectoral Analysis

Analysis Focus	Objective	Potential Outcome
Broader Geographic Scope	Include more cities and states in the analysis	Discovery of new regional hubs
Sector-Specific Analysis	Focus on industries with high demand for Business Analysts	Tailored recommendations for sector-specific strategies
Cross-Industry Comparisons	Compare demand across different sectors	Identification of sectors with the most robust opportunities

4. Evaluate Skill Gaps and Training Needs

Spotting skill gaps between job criteria and the workforce helps teachers and trainers to respond to market needs.

Table 9 Evaluate Skill Gaps and Training Needs

Analysis Focus	Objective	Potential Outcome
Skill Gap Analysis	Identify discrepancies between required and available skills	Recommendations for targeted education and training
Training Program Effectiveness	Evaluate the impact of existing training programs	Insights into which programs best align with market needs
Future Skill Forecasting	Predict which skills will become critical in the near future	Proactive curriculum adjustments

Conclusion and Strategic Focus

These next phases provide more focused and extensive research to enable Business Analyses to meet the rapidly changing job environment. Data on temporal patterns, wage statistics, geography, and talent shortages allow stakeholders to create future-ready decisions. Using this strategy, companies, job seekers, and educators might respond to changes and grab business analysis prospects.

4. Data Quality Analysis

1. Dataset-Level Data Quality Analysis

The dataset was examined in completeness, consistency, accuracy, and timeliness. Table below shows the outcomes:

Table 10 Dataset-Level Data Quality Analysis

Dimension	Issue Description	Quantitative Measure	Potential Fixes	Impact on Validity/Trustworthiness
Completeness	Missing data in JobNumber and job_summary	2.65% missing in both variables	Imputation or exclusion of missing records	Missing data can lead to incomplete or biased analysis.
Accuracy	Inconsistent job titles (e.g., variations like "BA" vs "Business Analyst")	Inconsistent labels for job titles	Standardize job titles using controlled vocabulary	Inaccurate classifications can misrepresent demand.
Consistency	Unique value checks across categorical fields show variability	Varying counts of unique values (e.g., job_title: 2741 unique values)	Standardize entries, merge duplicates	Inconsistencies can lead to errors in aggregation.
Timeliness	Outdated job postings with some records older than expected	0 entries with dates outside the range	Filter records based on a relevant timeframe	Outdated data can mislead current market analysis.

2. Variable-Level Data Quality Analysis

The table below shows varying-level data quality issues including missing data, consistency, correctness, and timeliness checks together with solutions.

Table 11 Variable-Level Data Quality Analysis

Variable	Data Quality Issue	Missing %	Consistency Check	Accuracy Check	Timeliness Check	Fixes and Recommendations
JobNumber	Missing data	2.65%	Unique values: 149	No future dates detected	No unreasonable dates	Impute missing values or exclude records
job_title	Variations in title naming conventions	0%	High variability: 2741 unique values	Standardize job titles	No specific issues found	Standardize and merge similar titles
job_location	Inconsistent naming (e.g., abbreviations)	0%	1117 unique values	Cross-referencing needed	No specific issues found	Use consistent formats and verify against standards
job_summary	Missing data	2.65%	Consistent entries	No specific accuracy issues	No specific issues found	Impute or exclude missing data
search_city	Inconsistent formatting	0%	527 unique values	Some entries lack specificity	No specific issues found	Standardize city names, merge duplicates
search_country	Consistent entries	0%	Only "United States"	No specific accuracy issues	No specific issues found	Maintain consistency
search_position	High variability	0%	216 unique values	Verify role definitions	No specific issues found	Standardize naming conventions
job_level	Limited variability	0%	Only 2 unique values	Check for level definitions	No specific issues found	Validate job level classification

job_type	Consistent entries	0%	Only 2 unique values	Validate against definitions	No specific issues found	Maintain current standards
first_seen	Timeliness issues with outdated entries	0%	Consistent formats	Verify date accuracy	All dates within reasonable range	Validate date entries and exclude outdated records

3. Impact of Data Quality on Results

Reliability of analytical results depends on data quality. Inaccurate data can skew assessments and understanding. The way data quality—completeness, consistency, accuracy, and timeliness—affects analytical results is examined in this section.

1. Completeness

It is said to be complete if the dataset contains all relevant information. Missing job numbers and job-summary data undermine the analysis. Lack of data points could lead to erroneous results and poor study. Missing data does not follow a random distribution; hence trends could influence the outcomes. This study suggests that data conclusions could skew the opinions of stakeholders.

2. Consistency

Data consistency involves consistent entries. Geographic names or job titles ("BA" instead of "Business Analyst") can lead to aggregations issues. Misclassifications can confuse groups or comparisons of like objects, therefore impeding comparative research. Stakeholders use these insights to make strategic decisions—such as detecting significant hiring trends or regional demand for particular skills—which influences business decisions and plans could lead to incorrect or partial interpretations of conflicting facts.

3. Accuracy

Data consistency conforms data entries to reality. Clearly, poor job titles can skew employment demand. Inaccurate job name collecting could change market trends and skew work obligations. Distortion can cause stakeholders to undervalue target job categories or skill demand, therefore affecting recruitment, training, and strategic planning.

4. Timeliness

Data timeliness is what we use right now. Older databases have an influence on market trends. Using obsolete data, the study could skew the employment market. This mismatch can mislead businesses, job seekers, and educators implementing these concepts to direct decisions. Older data could reveal demand for out-of-demand skills, waste of training or hiring campaigns.

4. Summary of Recommendations

One can improve data analysis validity in several respects. First data has to be cleaned and standardized. Standardizing variable forms and dataset homogeneity reduces inconsistent, missing, and erroneous data. Second, imputation—that is, removal of records with notable gaps—helps to preserve analytic integrity for variables with missing data. Finally, changing the dataset and deleting obsolete information will ensure that the research fairly depicts market realities.

Assignment 2

1. Background and Motivation

Politicians, companies, and job seekers all have to be familiar with the American employment scene. Declining to 7.4 million in October 2024, the lowest since January 2021, job adverts allude to a hiring slump. But government and healthcare have increased roles, reflecting a complex and shifting job situation.

Salary trends, job distribution, and skill demand all reveal insights. This knowledge raises employability and helps job searchers by aligning abilities to market needs. Businesses can adjust their hiring policies and identify areas lacking skills. Those that understand economic situation could focus on workforce development.

Recent research show that 37% of the top 20 skills sought for the average American job have changed since 2016. World Economic Forum's "Future of Jobs Report 2023" looks at changing employment rolls and skill development.

Examining employment distribution, skill demand, and salary changes carefully can help stakeholders make smart decisions and match the changing U.S. labor market.

2. Material for Analysis

This section addresses the degree, quality, source, preparation of the analytic dataset, etc.

1. Dataset Description and Scope

This paper covers a spectrum of U.S. business analyst roles. The material addresses jobs, places, skills, and wage ranges. These tools let one look at business analyst wage trends, skill in demand, and job posting dispersion.

The scope of this dataset includes:

Job Types: This can search by level entry-level, mid-senior, senior, and name criteria.

Locations: Job advertising from several American cities expose regional job availability and patterns.

Skill Requirements: Complete technical and soft skills satisfy the needs of companies' analysts.

Salary Information: Different job roles expose different compensation trends in salary ranges.

2. Data Quality Issues

Although a lot of data is available, the preliminary study found some data quality problems. This might impact on the accuracy of insight; hence data preparation and cleansing were carried out:

Missing Values: Many entries lacked salary and skill criteria, which affects research on wages and skill-demand.

Duplicates: The dataset included many duplicate records, most likely from job advertising. Duplicates were removed to avoid bias and give every job posting unique character.

Data Granularity Limitations: Although lacking specifics, the dataset provides broad knowledge on business analysts. Rare industry-specific jobs and job function breakdowns could influence studies of certain labor roles.

3. Data Source and Collection Period

Providing a whole picture of current skills and compensation trends, the dataset comprises 2024 business analyst job listings from professional platforms and aggregated job boards.

4. Data Cleaning and Preparation Steps

Several approaches of data preparation were used to increase analytical dependability and data quality.:

1. Handling Missing Values: For missing wage or skill data, imputation or exclusion preserved data integrity.
2. Duplicate Removal: Special IDs helped to lower overrepresented job announcements and redundant data.
3. Standardization: Standardizing occupational titles and skill sets—that is, "BA" for "Business Analyst—"made finding and ranking related occupations easier."

This dataset is finally ideal for looking at U.S. city employment distribution, skill requirements, and wage trends. The appendices' data dictionary and quality check summary clarify the dataset's constraints and framework.

3. Analytic Approach

This part covers analyses of systematic employment market data. This visualized, used thorough research, and prepared data. Analytical phases enlarged the dataset, enhanced data integrity, and provided insights satisfying analytical goals.

1. Data Preprocessing

Data organization guarantees correctness and consistency. This phase is comprised in:

Handling Missing Values: Depending on how the data influenced the study, missing income and skill requirements numbers were imputed or filtered out.

Removing Duplicates: Duplicate job advertising was removed to avoid overrepresentation. IDs filtered the info for me.

Standardizing Fields: Simplified grouping and aggregation made feasible by consistent job titles, city names, and skill names across entries Standardizing " BA" became "Business Analyst."

2. Data Filtering and Grouping

Data was cleaned and normalized; next, employment market factors shaped filtering and classification of it:

City-Based Grouping: City employment ads allow one to evaluate local demand, salary, and skill sets. That provides geographic research on the employment market for business analysts.

Skill and Qualification Segmentation: To identify which skills are most in demand for different occupations, job advertising was categorized by skill needs—technical vs. soft—and qualifications—Bachelor's, Master's, Ph.D.

Salary Banding: Range of salaries helped to pinpoint job level pay trends and geography.

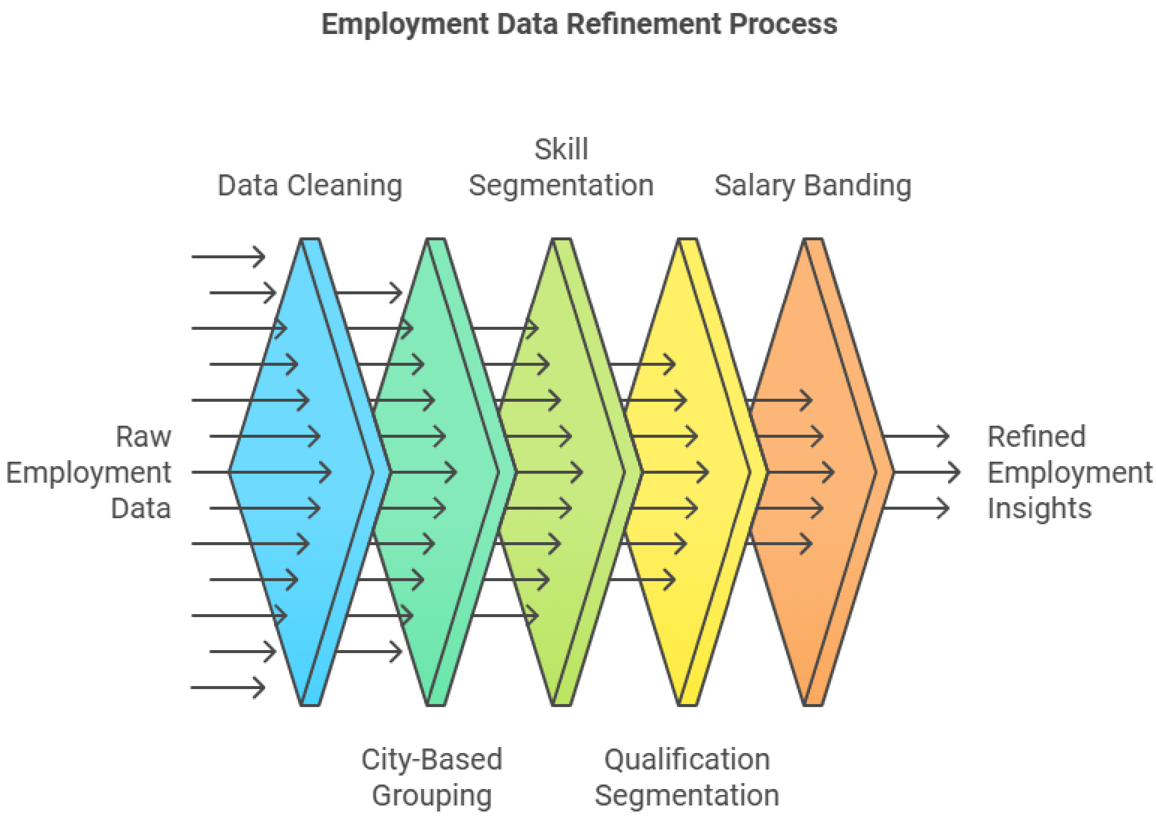


Figure 17 Employment Data Refinement Workflow

3. Analysis and Visualization Tools

Many projects visualized and examined data:

Microsoft Power BI: straightforward and unambiguous Power BI displays revealed city, skill, and income patterns.

Microsoft Excel: Excel helped to examine and summarize data before loading into Power BI.

4. Flow Diagram of Analytic Approach

Analyzing moves in a flowchart.:

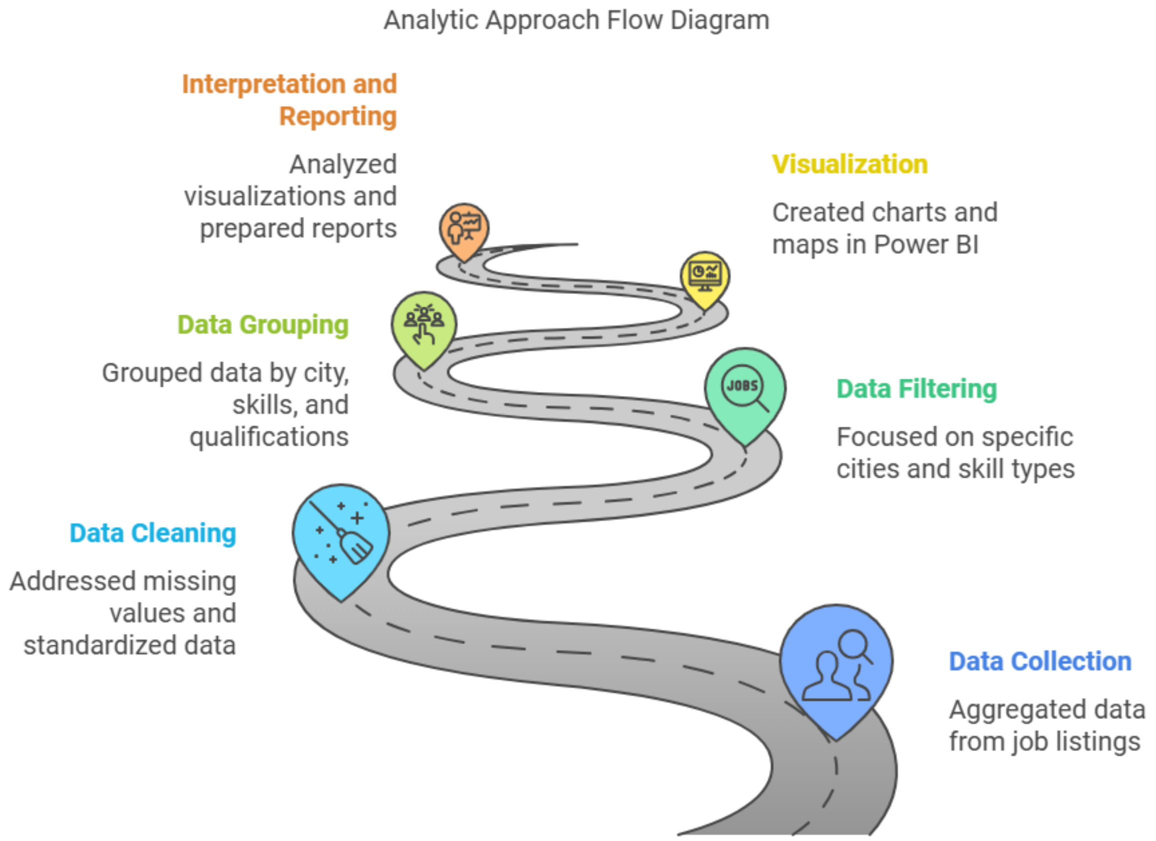


Figure 18 Flow Diagram of Analytic Approach

1. Data Collection: Aggregated data from job listings.
2. Data Cleaning: Eliminated redundancies; standardized titles and competencies.
3. Data Filtering: Designed communities and skills especially for business analysts.
4. Data Grouping: Grouped data by city, skills, qualifications, and salary ranges.
5. Visualization: Created charts and maps in Power BI to visualize patterns and insights.
6. Interpretation and Reporting: Review and record outcomes with images.

This all-encompassing approach produced detailed study on employment market statistics. A flow chart shows these analytical stages in the appendix.

4. Findings

This section covers the main conclusions of data analysis using charts and explanations to accentuate significant trends and insights. Every picture presents income by area, skill level, distribution of qualifications, and trends in the business analyst employment.

1. Job Distribution by City and Salary

This section includes American cities with the highest yearly earnings based on pay scales and job distribution. Information on employment distribution and pay can be exposed to stakeholders' location-based salary concentration trends and highly compensated work opportunities.

Key Findings

A handful of locations routinely rank well in annual pay. This information helps businesses to find talent, job seekers to optimize salary, and politicians to direct economic development.

1. Top 20 Cities by Total Yearly Salary

Chart Location: With this bar chart, show cities with the highest cumulative incomes top this section.

Figure 19 Top 20 Cities by Total Yearly Salary

Description: The bar chart "Top 20 Cities by Total Yearly Salary" ranks the cities with highest total pay. Summit pays $2.37 million; Santa Clara $1.05 million; San Diego $0.94 million. These figures suggest that businesses seeking local talent and job seekers choose places with more highly paid employment.

Insights: Reflecting a spectrum of high-paying job areas, technology centers and less-traveled sites, top cities by total annual income are Santa Clara and San Diego reflect high income California; conversely, Concord and Gastonia offer competitive pay in particular industries.

2. Geographical Distribution of Yearly Salary

Chart Location: Underneath the bar chart on this map chart is a geographic U.S. pay distribution display.

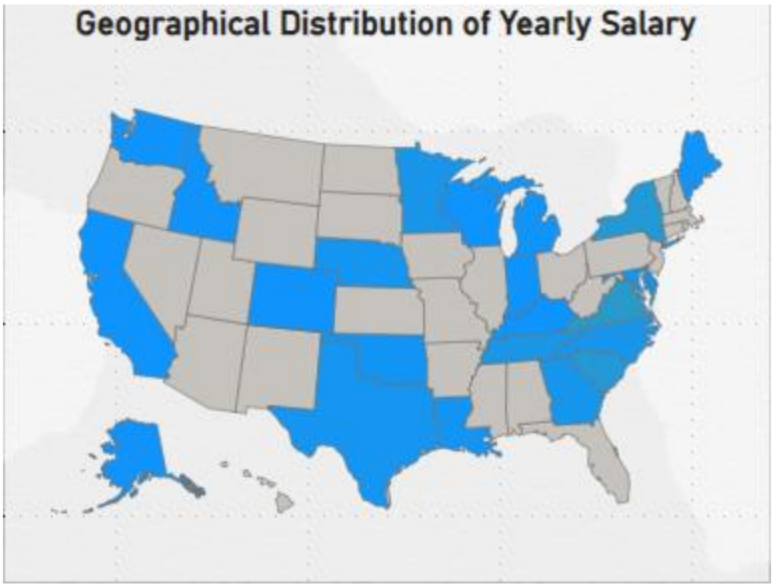


Figure 20 Geographical Distribution of Yearly Salary

Description: Blue states on the map chart, "Geographical Distribution of Yearly Salary," show high-salary areas based on total annual wages all over the United States. On this map are high paying job clusters in Washington, New York, California, and Texas.

Insights: On this map are shown high-salary employment markets east and west coastwise. More high-paying cities in California and Texas than in the Midwest and South, taken together the distribution leads to high earnings choosing specific states.

To help governments, companies, and job seekers in making wise decisions, these images depict American city employment and income distribution.

2. Highest Salary by City

Investigated in this section it is the city with the highest personal income in the dataset in order to understand regional salary differences and causes of income spikes. Finding this top-paying city helps us to grasp the U.S. economic situation and high-salary work environment.

Key Findings

The highest personal income Everett has is $346,000. This could entail changes in the cost of living, competitive demand for specific talents in this city, or industrial strengths.

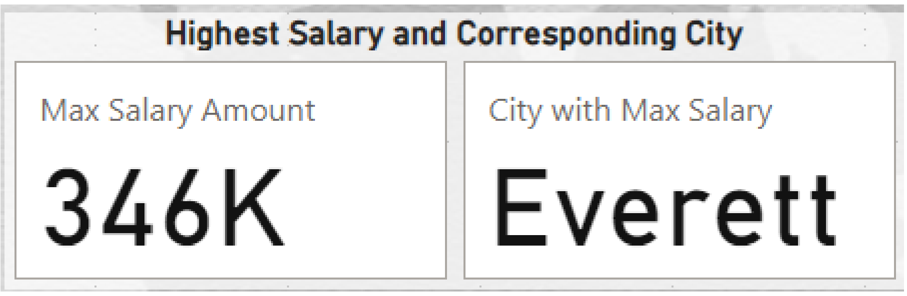


Figure 21 Highest Salary by City

1. Max Salary Amount

Visual Location: Use this card graphic at the top left of this part to quickly display the dataset's maximum salary.

Description: The highest pay in the dataset on a Max Pay Amount card is 346K. It is especially important since it determines the highest income in the American cities under analysis.

Insights: The statistics show highly paid or specialized jobs needing extensive knowledge, credentials, or particular skills. Top earners find Everett enticing because of IT, healthcare, and leadership; these factors help to explain the great pay.

2. City with Max Salary

Visual Location: Match this card image with the "Max Salary Amount" card to show viewers the city with highest salary.

Description: Everett benefits most from the city pay card. This helps to orient the income statistics since highly paid jobs are regionally concentrated.

Insights: Local industrial needs or competitive wages in specific industries could help to explain Everett's high pay. It could also reflect the city's economy, high living prices, or great demand for knowledge, employers are ready to pay extra to attract competent workers.

3. Job Distribution by Skills (Technical and Soft Skills)

Content

This section looks at city technical and soft skill needs to develop regional priorities. Certain sectors could emphasize technical knowledge above soft abilities like communication and teamwork. These variations reflect the demands for regional talent, which guides businesses assessing job seekers' skill level as well as their own.

Charts to Include

1. Map Chart: “Technical Skills Job Distribution by City”

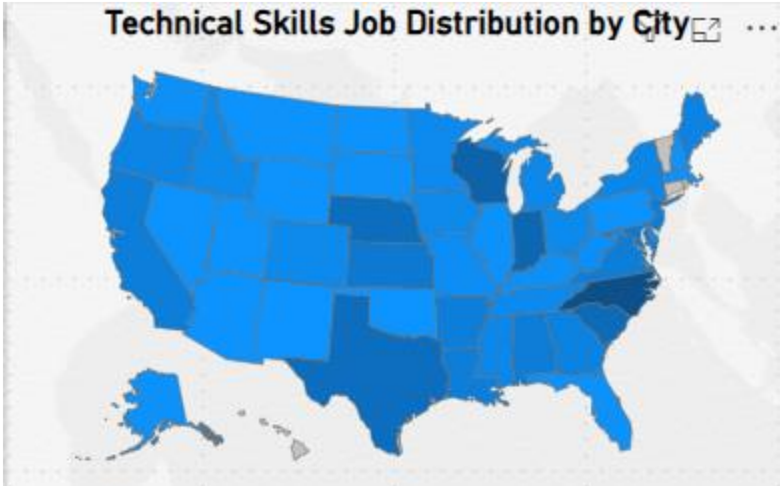


Figure 22 Technical Skills Job Distribution by City

Location in Report: Beginning the part on talents, this map picture will show demand for technical abilities.

Description: The "Technical Skills Job Distribution by City" map reveals American city technical skill needs. Technical employment on darker toned maps is higher. Given their tech companies, Austin and Seattle might find increased demand.

2. Map Chart: “Soft Skills Job Distribution by City”

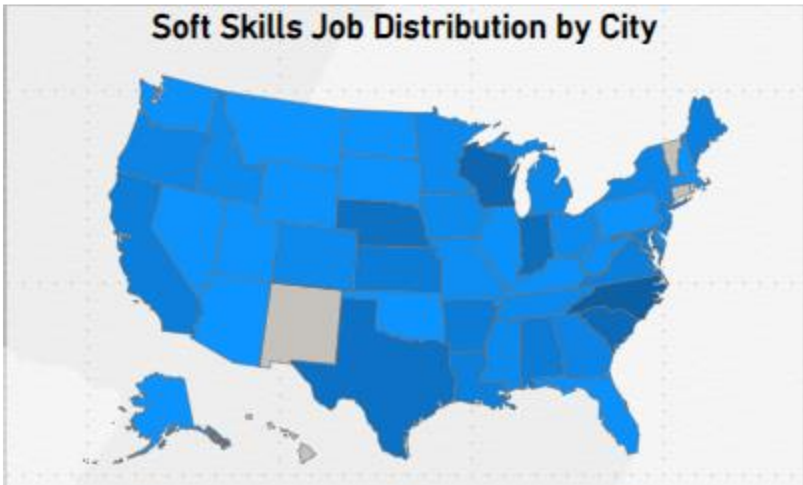


Figure 23 Soft Skills Job Distribution by City

Location in Report: Following the distribution map for technical skills, this map visual lets one compare technical and soft skill demand.

Description: For every city, the "Soft Skills Job Distribution by City" map provides soft skill job vacancies. Darker tones indicate soft skill demand, much as with the technical skill map. This image shows where in customer-facing positions soft skills like communication and problem-solving are more crucial.

These numbers could assist companies in determining if cities value soft and technical skills as well as assist job seekers.

4. Comparison of Job Types Across Cities

Content

Emphasizing their frequency, this section compares cities onsite, distant, and hybrid work modes. Job categories could show patterns in onsite roles in areas like manufacturing or healthcare or in remote work preferences.

Charts to Include

1. Donut Chart: "Comparison of Job Types Across Cities"

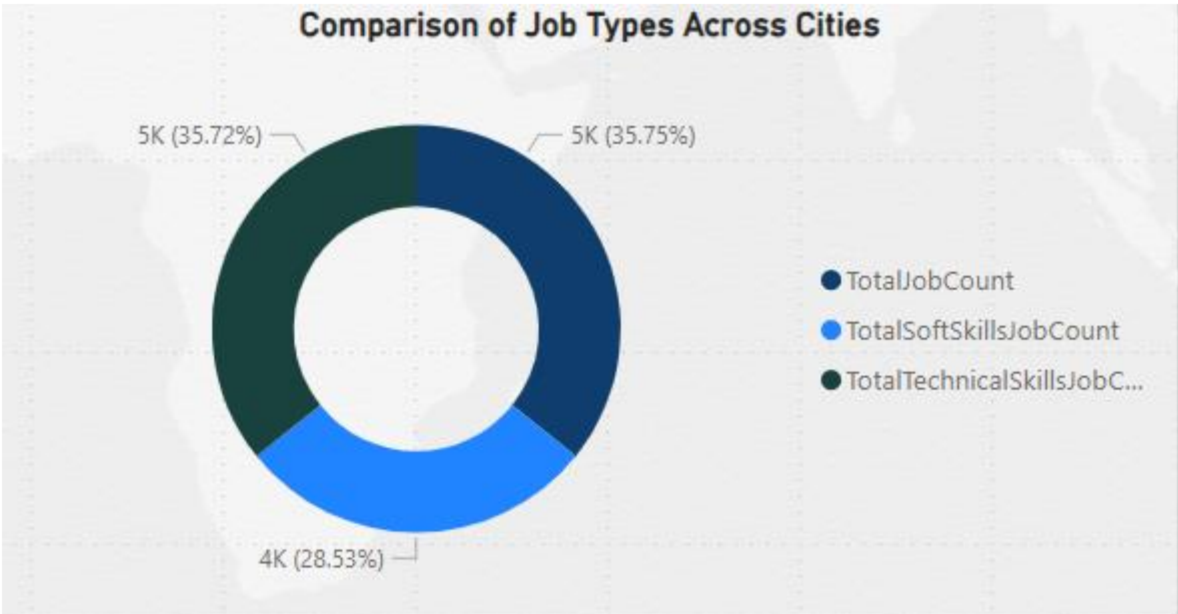


Figure 24 Comparison of Job Types Across Cities

Location in Report: This chart on the distribution of city employment type should fall under "Findings".

Description: The "Comparison of Job Types Across Cities" donut graphic displays U.S. city onsite, remote, and hybrid job ratios. Every cut of a donut depicts a different type of market job. Technology-oriented communities could have more remote jobs than those concentrated on industry or logistics.

Cities shown on this chart would rather employ remote workers than onsite ones. This kind of method helps one to understand how labor expectations and employment patterns are affected by regional demands and economic factors.

5. Skill Demand by Qualification Level

Content

Industry trends, regional distribution, and talent need by level of qualification are covered in this part. Data filtering under qualifying criteria indicates industry preferences, regional patterns, and jobs connected to educational degrees. To determine educational worth, politicians, companies, and job seekers as well as others need this knowledge.

1. No Qualification Requirement (None)

Figure 25 "None" Qualification Level

Description for "None" Qualification Level

With the highest annual "None" incomes were Santa Clara, Malibu, and Everett. These places could offer highly paid jobs employing expertise and skills without degrees. Everett has the best non-degree salaries; hence many employments could be advantageous. The map reveals concentrated opportunities in Virginia, Texas, and California, therefore demonstrating the significance of practical knowledge even in the absence of official credentials.

2. Certificate Requirement

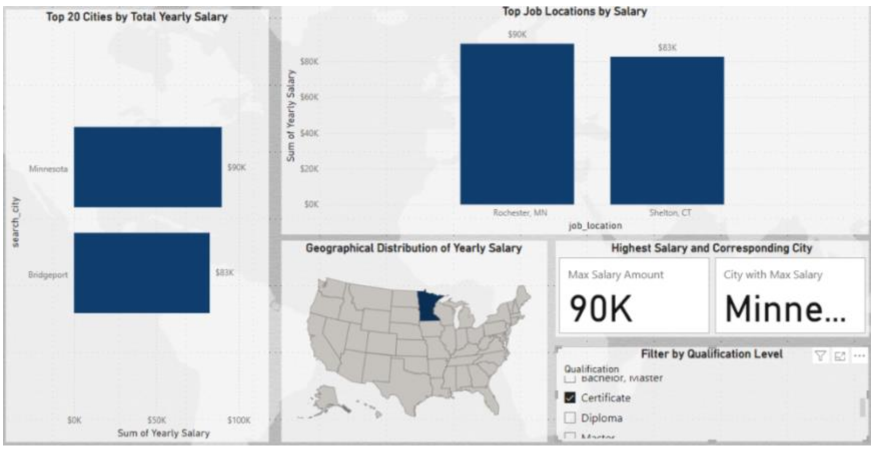


Figure 26 "Certificate" Qualification Level

Description for "Certificate" Qualification Level

Bridgeport paid $90K annually in certificate-required salary, following Minnesota. Rochester's individual pay ties for highest are According to the map, Minnesota prefers certified talents since chances are concentrated in that state. This indicates demand for specialized work when degrees are not required for job qualifications and certificate competency meets them.

3. Diploma Requirement

Figure 27 "Diploma" Qualification Level

Description for "Diploma" Qualification Level

Norfolk claims the highest diploma-required pay at $0.15 million. Durham's top salary, $145,000, indicates demand for degrees holders. On the "Top Job Locations by Salary" list Virginia Beach, VA with $151K ranks top. Raleigh gives 151K. Concentrated in North Carolina and Idaho, diploma-equal employment exhibit expertise.

4. bachelor's degree Requirement

Figure 28 "Bachelor's" Qualification Level

Description for "Bachelor's" Qualification Level

For bachelor's degree positions, Summit pays $1.8M; Avondale and Concord follow at $0.5M. Following Seattle, WA's bachelor-qualified positions paying $0.91M, the "Top Job Locations by Salary" graphic shows Nashville, TN and Bellevue, WA at $0.74M. Washington and Tennessee most require bachelor-qualified positions, according to the geographical distribution chart. Up to $195K, the "Highest Salary and Corresponding City" card ranks Columbus first in bachelor-qualified professional high-paying jobs. This data demonstrates a spectrum of highly paid U.S. bachelor's employment.

5. master's degree Requirement

Figure 29 "Master's" Qualification Level

Description for "Master's" Qualification Level

Summit claims the highest master's-level annual pay with $0.53M; San Juan Capistrano and Layton follow with $0.28M and $0.26M. Under the "Top Job Locations by Salary" category, Bellevue, WA master-qualified positions pay $0.34M; Los Angeles, CA follows at $0.30M. Master's-level job hotspots on the map fall in Texas, Washington, California, and Based on the "Highest Salary and Corresponding City" card, Boulder claims the highest paying master's-qualified employment at $173K. According to this report, many American cities' high-paying employment demand master's degrees.

6. PhD Requirement

Figure 30 "PhD" Qualification Level

Description for "PhD" Qualification Level

In North Carolina, a job requiring a PhD pays $0.13 million year; in Tarrytown, it pays $0.10 million. The "Top Job Locations by Salary" graphic shows Charlotte, NC paying PhD-qualified workers $126K; Queens, NY follows at $99K. According to the geographical distribution map, North Carolina features a concentration of highly paid PhD jobs. The "Highest Salary and Corresponding City" card shows North Carolina to have the highest PhD compensation. This analysis shows that while Charlotte, NC has excellent earning potential, PhD-level employment are concentrated in less places.

7. Overall Interpretation Summary

Degree level employment demand exposes regional education needs and industry-specific needs. These insights help job seekers—based on their credentials—find openings in highly sought-after areas. This data enables businesses and legislators to adapt workforce development for specific local job markets.

5. Interpretations

Here this examine qualification level and job type statistics in order to identify the reasons behind U.S. employment market trends and their implications for businesses and job seekers.

1. High-Salary Cities

The numbers show Santa Clara, Summit, and Everett pay high overall annual incomes across levels. This can be the outcome of some areas' very profitable sectors or main industries. Santa Clara lives in Silicon Valley, where demand for IT experts is high and pay rises following this.

Paying top dollar for outstanding individuals, aerospace and financial services could affect Everett and Summit. Targeting these cities could offer profitable employment, even if businesses may have to pay competitively to retain qualified staff.

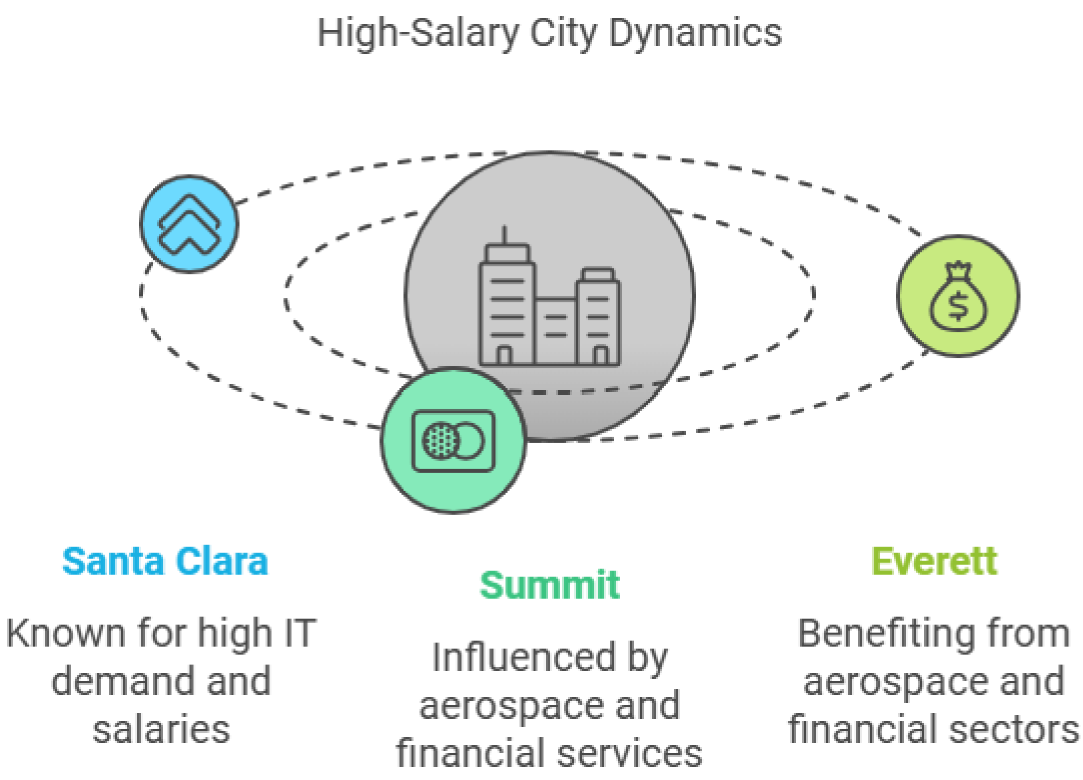


Figure 31 High-Salary City Dynamics Overview

2. Regional Skill Demands

The demand for regional abilities is somewhat different. Seattle and New York, centers for banking and IT, call for tech knowledge. Businesses in these specialized cities call upon engineers, data analysts, and software programmers. Businesses that depend on soft abilities frequently prioritize interpersonal communication, customer service, and healthcare among other things. Job seekers could find it helpful to match their qualifications to the leading industries in these areas, and companies could use these particular talents to boost hiring.

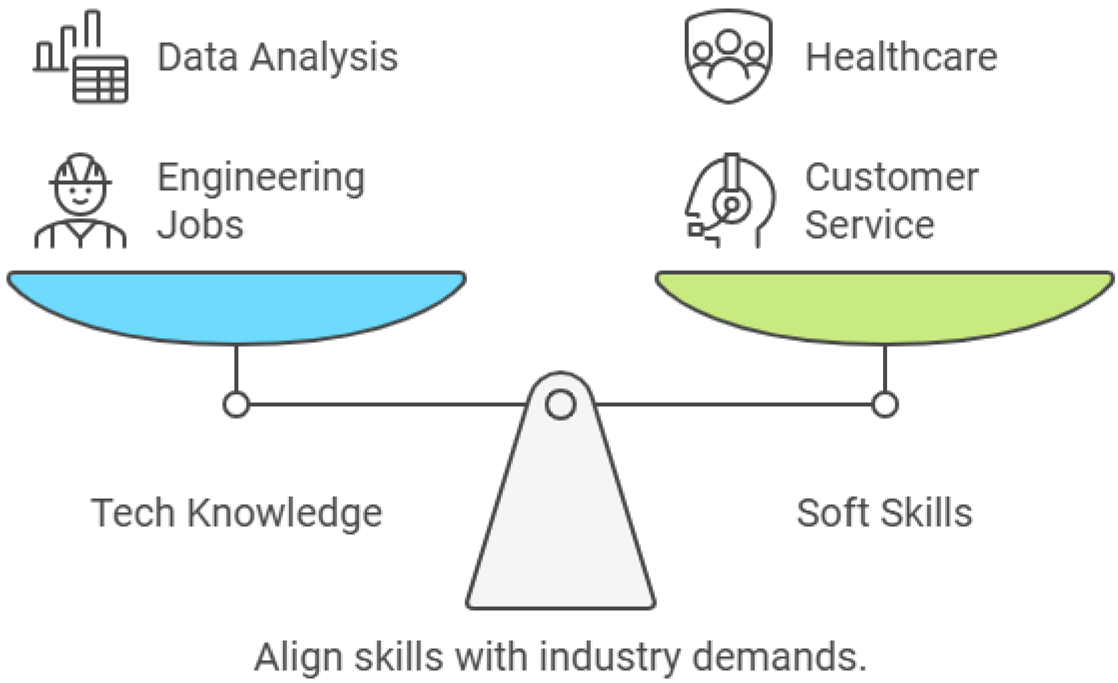


Figure 32 Regional Skill Demand Balance: Tech vs. Soft Skills

3. Job Type Distribution: Onsite vs. Remote

The analysis reveals unusual onsite and remote employment patterns specific to many locales. Big cities like San Francisco and New York have more on-site businesses including banking, manufacturing, and healthcare since corporate offices and businesses demanding physical presence are there. Austin and Raleigh are creating digital task-oriented innovation and business centers and offering remote work. Particularly in IT, remote employment could be a response to conflicting working patterns aimed at attracting expertise.

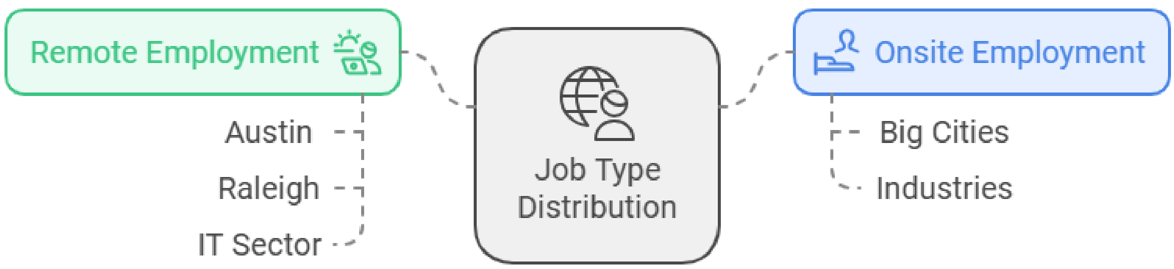


Figure 33 Job Type Distribution: Remote vs. Onsite Employment Trends

4. Impact of Qualification Levels on Salary and Demand

Based on what this found, Masters and PhDs make more and live in less cities. Master's and PhD applicants in Boulder and North Carolina pay greatly for research, academics, and innovative technologies. This tendency shows that advanced degree holders should choose countries where their knowledge is recognized, even as competitive work conditions may drive corporations to offer extra incentives to recruit top personnel. High-paying locations attract top candidates for very specialized vocations.

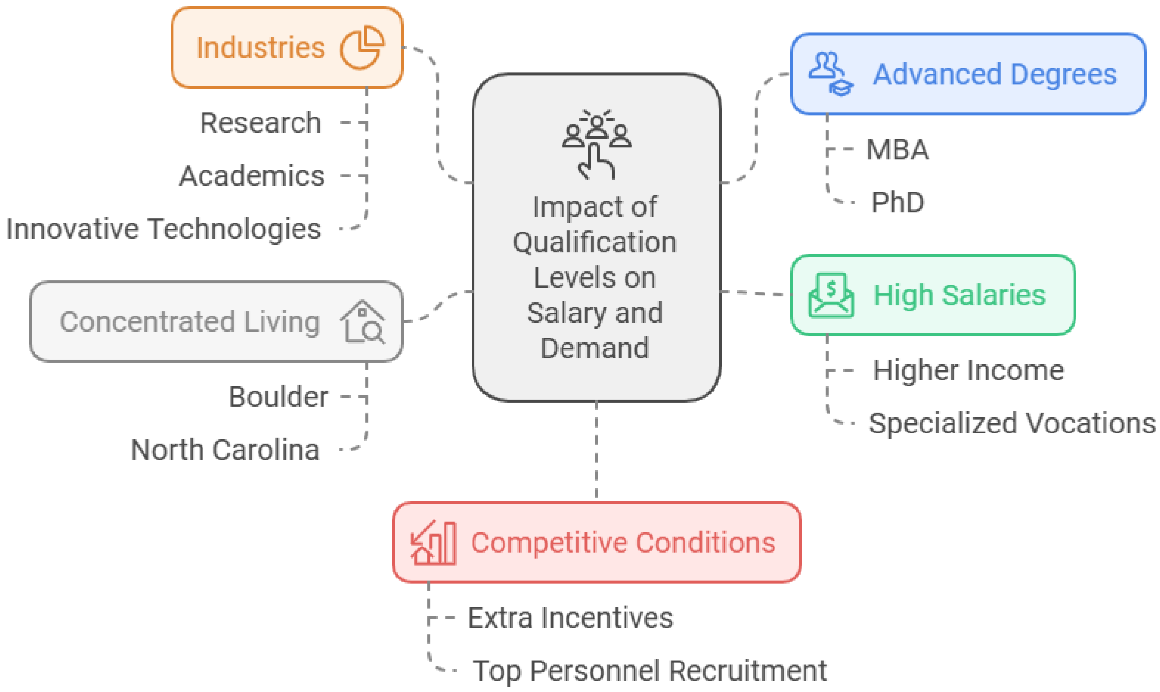


Figure 34 Impact of Qualification Levels on Salary and Job Demand

5. Influence of Qualifications on Job Opportunities by City

Filtering according to credentials reveals employment opportunities. One first of course:

Certificate-level Many cities translate more jobs into more entry-level opportunities.

Bachelor's degree Since they bring business, technological, and financial knowledge, jobs largely control Seattle and New York.

Master's and PhD Level employment are concentrated in specific cities due to specialization. PhD positions are concentrated in Durham and Charlotte, probably due to research and higher education.

Figure 35 Job Opportunities by Qualification Level and City

These changes will help companies and job hunters. Knowing qualification demand will enable job searchers to choose where their credentials are valued. Knowing regional qualifying requirements allows businesses to select candidates with the required background for the local business.

Summary

These figures highlight some crucial features of the American job market:

Regional specialization: City with industry clusters affect compensation and demand of talent.

Qualification-driven demand: Higher degrees pay more in specialist positions alone.

Shift in job types: In business services and technology, onsite and remote work mirror changing job requirements.

This data helps job seekers restrict their search by industry and degree of qualification and helps businesses to establish competitive plans to draw and retain talent in their local areas.

These dynamics enable stakeholders to make decisions important to the U.S. employment market.

6. Recommendations

This study of employment distribution, salary trends, and qualifying requirements can help job searchers, businesses, and governments all over many American cities:

1. For Job Seekers

Target Cities with High Demand for Specific Skills: Targeting skill-demanding cities will assist job seekers. The "Technical Skills Job Distribution by City" graphic displays techies either working remotely in tech hotspots or cities with more job openings.

Leverage Qualifications for Higher Salaries: Candidates for master's and PhD degrees should search for employment in areas where their particular work would be adequately compensated. Based on qualifying tables, North Carolina (PhDs) and Boulder (Master's) pay more but lure highly qualified people.

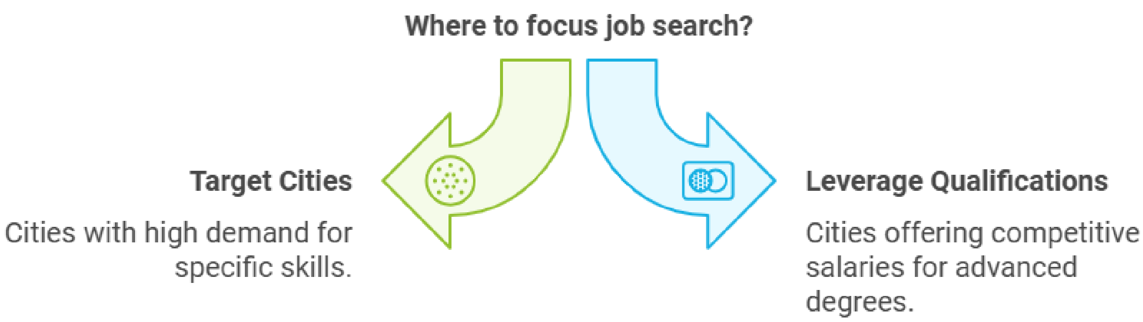


Figure 36 Job Search Strategy for High-Demand Skills and Advanced Qualifications

2. For Employers

Offer Competitive Salaries in High-Salary Cities: Businesses seeking excellent candidates should consider municipal pay scales. New York, Seattle, and Everett have high pay expectations, according to "Top Job Locations by Salary". These sites might maintain competent staff paid fairly.

Expand Remote Opportunities: Given the growing trend toward remote work, especially in tech—businesses should offer remote choices for non-physical professions. This strategy can attract from low-demand positions, therefore augmenting the pool of talent.

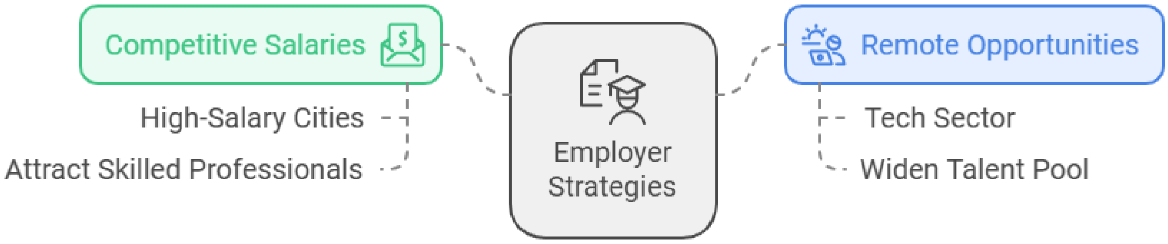


Figure 37 Employer Strategies: Competitive Salaries and Remote Opportunities

3. For Policymakers

Address Regional Skill Shortages Through Educational Programs: Policymakers in areas lacking some skills should back educational and career programs. Local industry-specific training encourages the closing of skill gaps by means of motivation. Local school cooperation and STEM projects could assist places needing high tech and analytical ability.

Encourage Regional Development for High-Qualification Roles: In specialized research, a field of significant demand with few employment possibilities, regional development incentives could be applied. Research institutes and innovation clusters could provide jobs in numerous sectors, therefore balancing the demand for qualified personnel

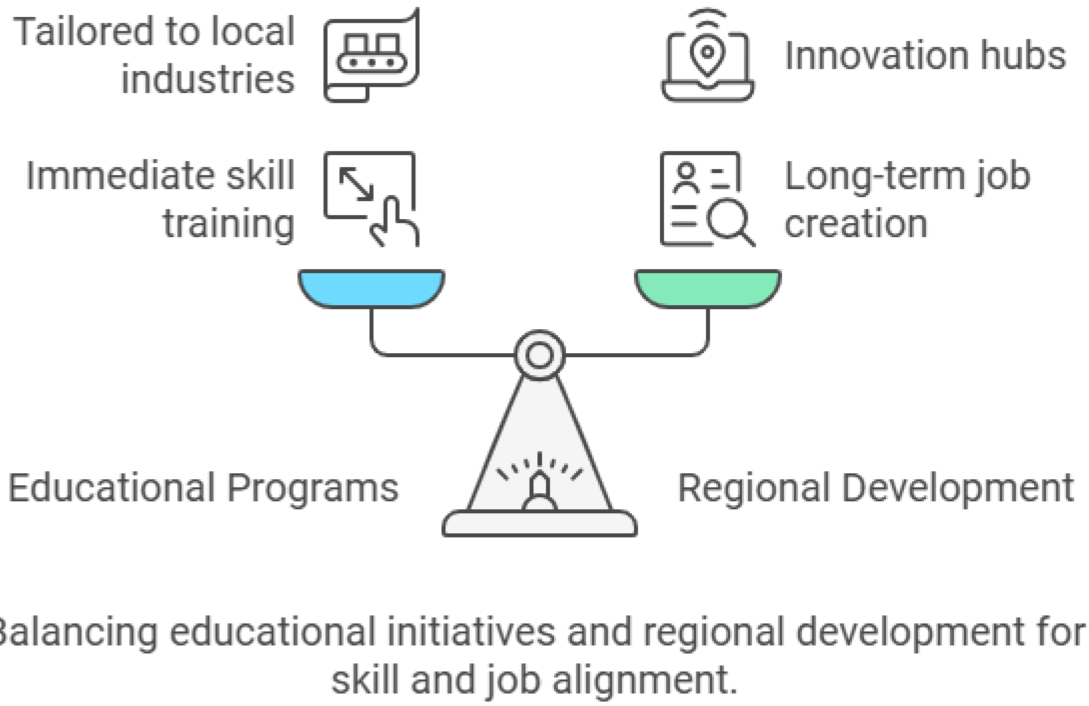


Figure 38 Policymaker Strategies: Balancing Educational Programs and Regional Development

By aligning their employment market trends with those recorded in this study, following these recommendations will help stakeholders maximize job prospects, talent acquisition, and workforce development all around the United States. These suggestions are supported in the employment and wage distribution charts by the priorities of the stakeholder groups.

Bibliography

1. Coursera. (2024). *What Is a Business Analyst? 2024 Career Guide*. Retrieved from Coursera. This guide provides an overview of the growing demand for Business Analysts, highlighting the skills required and the industries where these roles are most in demand.

2. Harvard Business Analytics Program. (2024). *9 Skills Every Business Analytics Professional Needs*. Retrieved from Harvard Business Analytics Program. This article discusses the core skills necessary for Business Analysts, emphasizing the importance of both technical and soft skills in the field.

3. IIBA®. (2024). *Business Analyst Career Road Map*. Retrieved from IIBA. This resource outlines the career opportunities for Business Analysts and the emerging roles within business architecture and intelligence that are in high demand globally.

4. MIT Sloan. (2024). *Explore the Program | Master of Business Analytics*. Retrieved from MIT Sloan. This program description emphasizes the value of business analytics skills in solving real-world problems, particularly in finance and technology sectors.

5. IIBA®. (2024). *What is Business Analysis? | Global Standard for Business Analysis Certification*. Retrieved from IIBA. This source defines business analysis and highlights the role of Business Analysts in driving organizational change through data-driven solutions.

6. IIBA®. (2024). *Business Analysis Global Standards of Practice*. Retrieved from IIBA. This document provides a comprehensive overview of the standards and best practices for business analysis, including the key skills and methodologies that professionals need to succeed.

7. Penn, University of. (2024). *Business Analytics Specialization [5 courses]*. Retrieved from Coursera. This course series from the University of Pennsylvania covers essential skills for business analytics, such as data-driven decision-making and performance management.

8. CareerExplorer. (2024). *The Job Market for Business Analysts in the United States*. Retrieved from CareerExplorer. This source provides insights into the job market for Business Analysts in the U.S., including job growth projections and regional demand.

9. BrainStation. (2024). *Are Business Analysts in High Demand?* Retrieved from BrainStation. This article discusses the increasing demand for Business Analysts due to digital transformation and the expanding role of data analytics in business strategy.

10. U.S. Bureau of Labor Statistics. (2024). *Occupational Outlook for Business Analysts*. Retrieved from BLS. This government source provides statistical data on the employment outlook for Business Analysts, including projected growth rates and salary information by region.

11. Bureau of Labor Statistics. (2024, November 1). Current Employment Statistics Highlights October 2024. Retrieved from https://www.bls.gov/web/empsit/ceshighlights.pdf

12. Robert Half. (2025). 2025 Salary Guide. Retrieved from https://www.roberthalf.com/us/en/insights/salary-guide

13. Payscale. (2024, October 15). Payscale Index- Quarterly Wage Trends for The United States. Retrieved from https://www.payscale.com/payscale-index

14. U.S. Bureau of Labor Statistics. (2024, September 6). Employment Projections: 2023-2033 Summary. Retrieved from https://www.bls.gov/news.release/ecopro.nr0.htm

15. U.S. Bureau of Labor Statistics. (2024, September 6). Employment Projections. Retrieved from https://www.bls.gov/emp/

16. U.S. Department of Labor. (n.d.). Wages Earnings and Benefits. Retrieved from https://www.dol.gov/general/topic/statistics/wagesearnings

17. Statista. (2024, July 1). Employment in the U.S. - statistics & facts. Retrieved from https://www.statista.com/topics/771/employment/

18. Willis Towers Watson. (2024, January). 2024 employee pay trends. Retrieved from https://www.wtwco.com/en-us/insights/2024/01/2024-employee-pay-trends

19. Glassdoor Economic Research. (n.d.). United States Job Market Report. Retrieved from https://www.glassdoor.com/blog/job-market-report-united-states/

20. U.S. Bureau of Labor Statistics. (2024, May). May 2023 National Occupational Employment and Wage Estimates. Retrieved from https://www.bls.gov/oes/current/oes_nat.htm

21. U.S. Bureau of Labor Statistics. (n.d.). Job Openings and Labor Turnover Survey (JOLTS). Retrieved from https://www.bls.gov/jlt/

22. Robert Half. (2024, August). July 2024 Jobs Report: Employers Add 114,000 Jobs. Retrieved from https://www.roberthalf.com/us/en/insights/research/july-2024-jobs-report-employers-add-114000-jobs

23. Robert Half. (2025). 2025 Hiring and Market Outlook. Retrieved from https://www.roberthalf.com/us/en/insights/salary-guide/market-outlook

24. U.S. Bureau of Labor Statistics. (n.d.). U.S. Bureau of Labor Statistics. Retrieved from https://www.bls.gov/

25. Argaam. (2024, November 1). US job growth slows sharply in October, unemployment rate steady at 4.1%. Retrieved from https://www.argaam.com/en/article/articledetail/id/1765448

26. CNN Business. (2024, May 3). US job market shows signs of cooling, with just 175,000 jobs added in April. Retrieved from https://www.cnn.com/business/live-news/april-jobs-report-markets-05-03-24/index.html

27. arketWatch. (2024, September 1). Want to make more than $150,000? Job hopping may be your best bet - but it's harder than it used to be. Retrieved from https://www.marketwatch.com/story/want-to-make-more-than-150-000-job-hopping-may-be-your-best-bet-but-its-harder-than-it-used-to-be-0b57fb40

28. Associated Press. (2024, November 1). US employers added just 12,000 jobs last month as hurricanes and strikes sharply reduce payrolls. Retrieved from https://apnews.com/article/dcf54cd3b40f01ca78bccf86cd8e13d5

29. Investopedia. (2024, October 29). Hiring Likely Took Big Hit In October. Retrieved from https://www.investopedia.com/hiring-likely-took-big-hit-in-october-8735573

30. The Wall Street Journal. (2024, November 1). Jobs Report Today: U.S. Added 12,000 Jobs in October as Storms Hit Hiring. Retrieved from https://www.wsj.com/livecoverage/stock-market-today-jobs-dow-sp500-nasdaq-live-11-1-2024

31. The Wall Street Journal. (2024, September 1). Bosses Are Finding Ways to Pay Workers Less. Retrieved from https://www.wsj.com/lifestyle/careers/salary-workers-pay-cuts-2024-54101d66

32. Barron's. (2024, October 28). Why It's So Hard to Find a Job in This Strong Labor Market. Retrieved from https://www.barrons.com/articles/jobs-hiring-work-payrolls-d5b004e9

33. Financial Times. (2024, September 30). US pay rises forecast to shrink as inflation and worker demand ease. Retrieved from https://www.ft.com/content/8ce6e640-7938-4bb7-9c77-26d9804d9a10

34. Reuters. (2024, November 1). US labor market staggers after blow from hurricanes, strikes. Retrieved from https://www.reuters.com/markets/us/us-job-growth-slows-sharply-october-unemployment-rate-unchanged-41-2024-11-01/

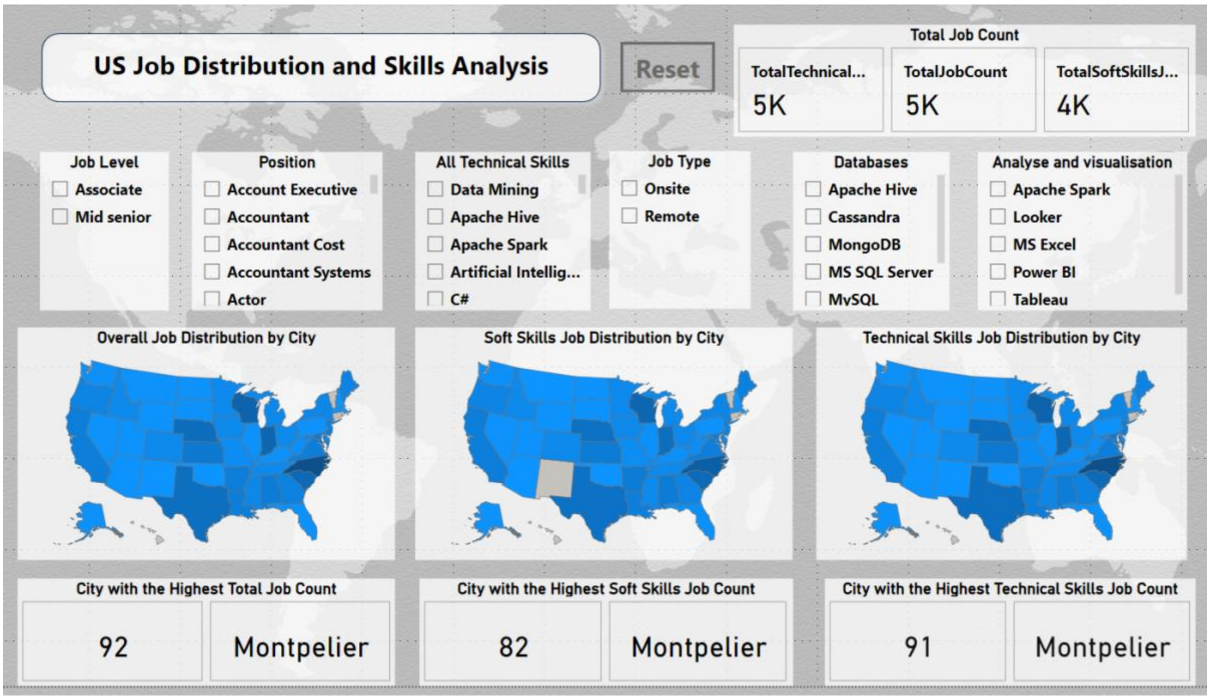
35. Reuters. (2024, November 1). Hurricanes, strikes seen denting US job growth as Americans prepare to vote. Retrieved from https://www.reuters.com/markets/us/hurricanes-strikes-seen-denting-us-job-growth-americans-prepare-vote-2024-11-01/

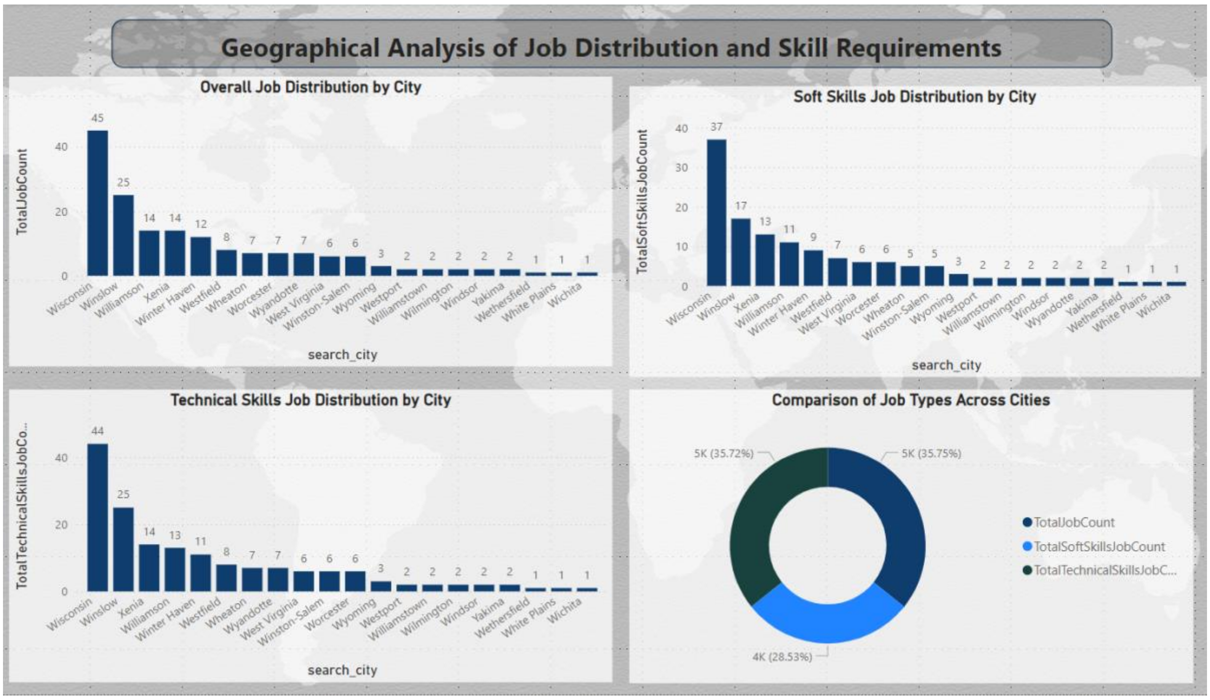
36. FXStreet. (2024, November 1). US job creation collapses but Fed to “overlook” data, Trump victory raises doubts on December rate cut. Retrieved from

https://www.fxstreet.com/analysis/us-job-creation-collapses-but-fed-to-overlook-data-trump-victory-raises-doubts-on-december-rate-cut-202411011350

37. FXEmpire. (2024, October 31). US Dollar Index Analysis: Gold Drops $50 as Jobs Report and Fed Rate Cut Uncertainty Loom. Retrieved from https://www.fxempire.com/forecasts/article/us-dollar-index-analysis-gold-drops-50-as

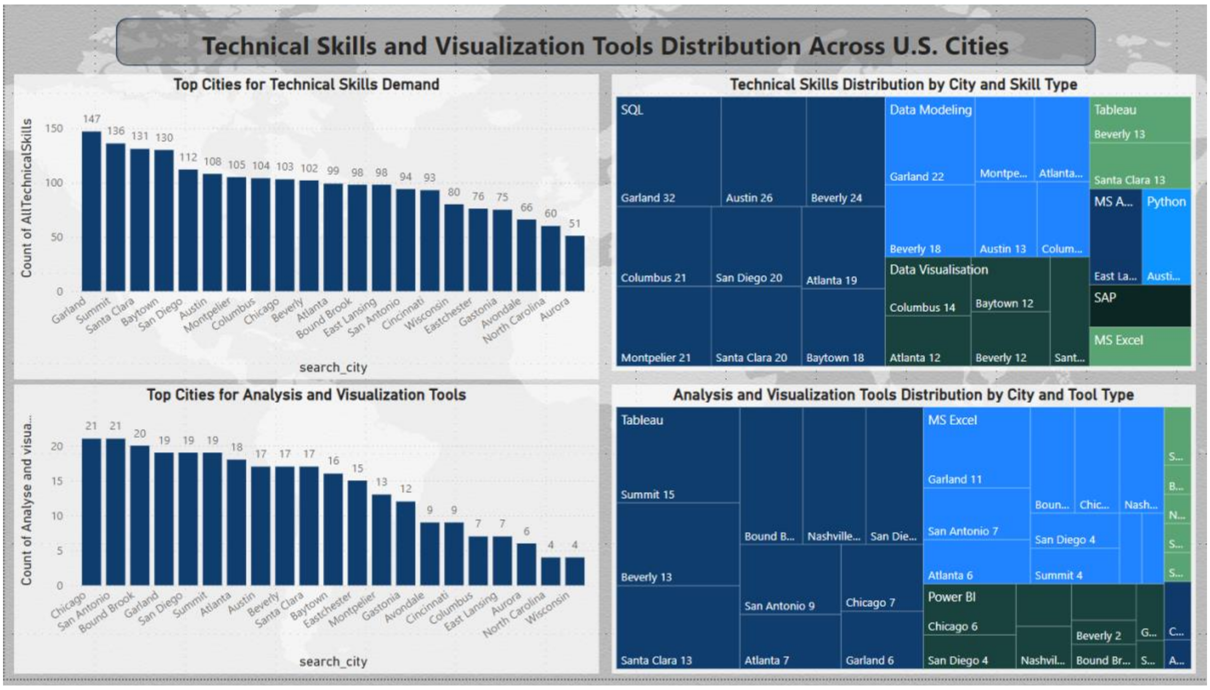
Appendix

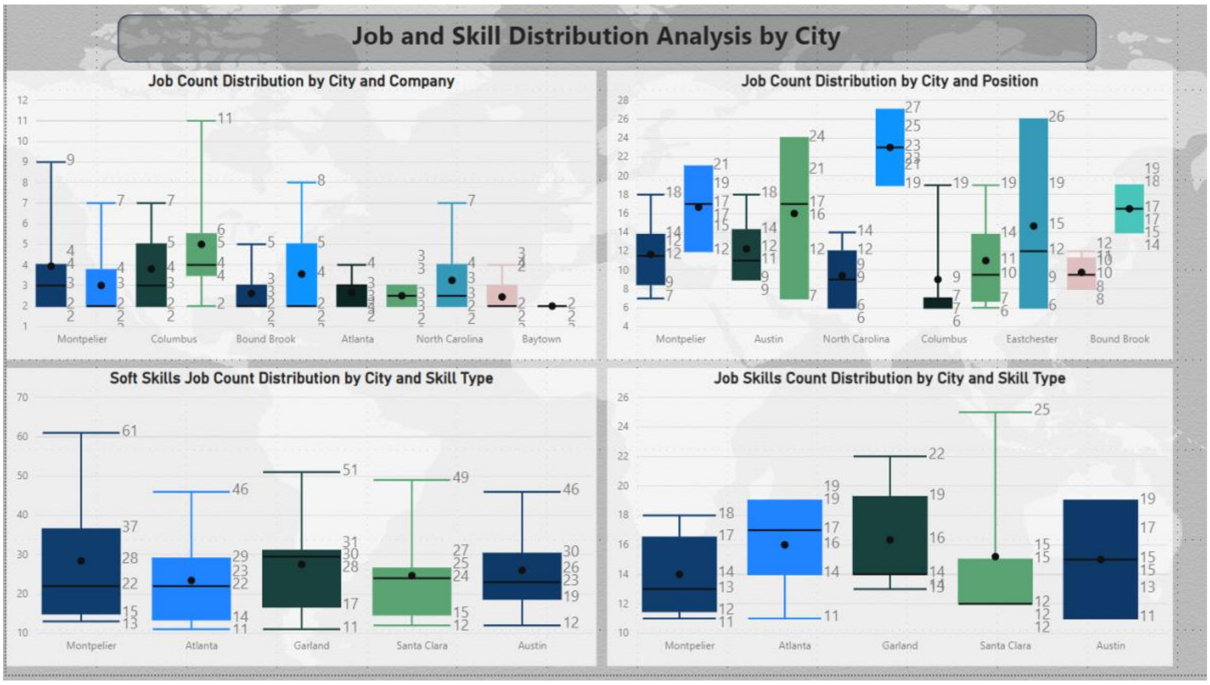


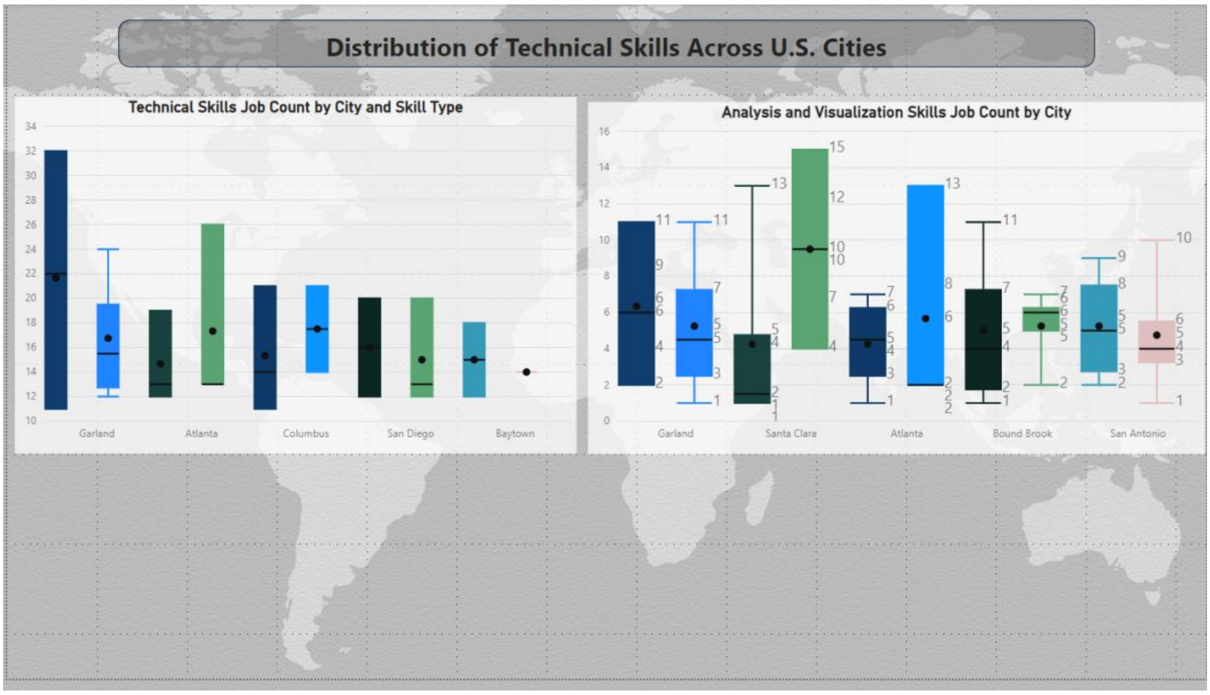

Geographical Analysis of Job Distribution and Skill Requirements
Overall Job Distribution by City
TotalJobCount
search_city
Soft Skills Job Distribution by City
TotalSoftSkillsJobCount
search_city
Technical Skills Job Distribution by City
TotalTechnicalSkillsJobCo...
search_city
Comparison of Job Types Across Cities
5K (35.72%)
5K (35.75%)
4K (28.53%)
TotalJobCount
TotalSoftSkillsJobCount
TotalTechnicalSkillsJobC...

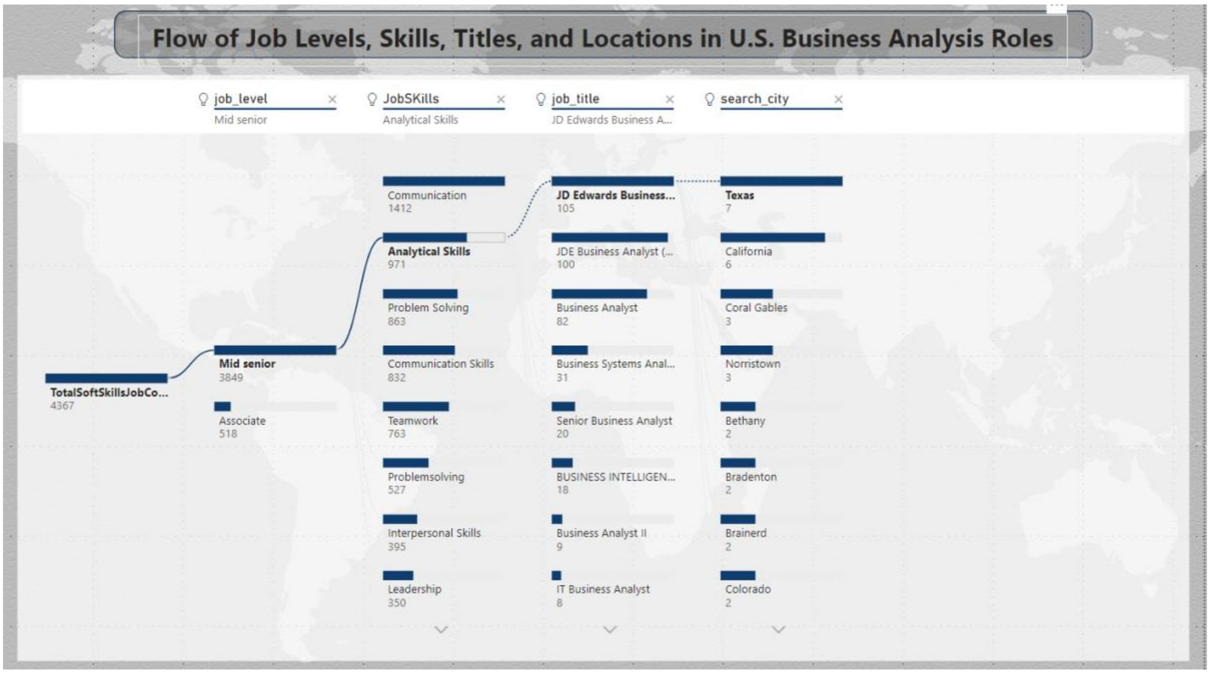

Company and Position Analysis Across U.S. Cities
Top Companies by City
Count of company
search_city
Job Distribution by Company and City
Top Job Positions by City
Count of search_position
search_city
Job Distribution by Position and City


Soft Skills and Job Skills Distribution Across U.S. Cities
Top Cities for Soft Skills Demand
Count of SoftSkills
search_city
Soft Skills Distribution by City and Skill Type
Top Cities for Job Skills Demand
Count of JobSkills
search_city
Job Skills Distribution by City and Skill Type

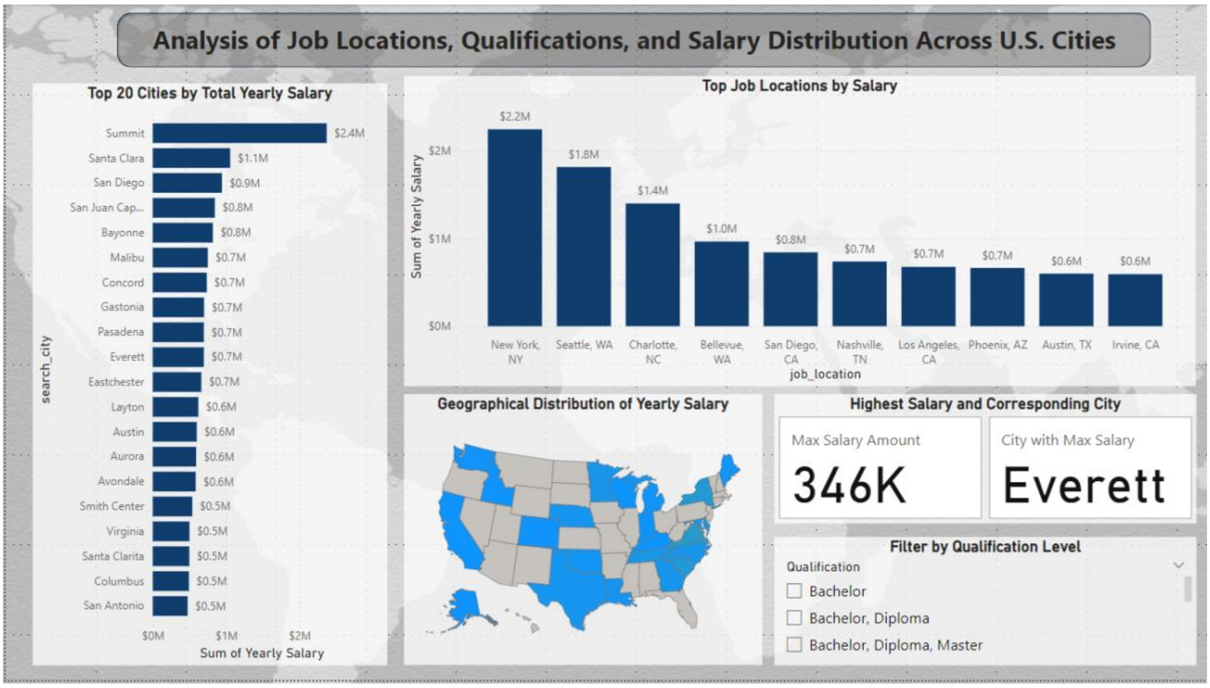

Technical Skills and Visualization Tools Distribution Across U.S. Cities
Top Cities for Technical Skills Demand
Count of AllTechnicalSkills
search_city
Technical Skills Distribution by City and Skill Type
Top Cities for Analysis and Visualization Tools
search_city
Analysis and Visualization Tools Distribution by City and Tool Type


Job and Skill Distribution Analysis by City
Job Count Distribution by City and Company
Montpelier
Columbus
Bound Brook
Atlanta
North Carolina
Baytown
Job Count Distribution by City and Position
Montpelier
Austin
North Carolina
Columbus
Eastchester
Bound Brook
Soft Skills Job Count Distribution by City and Skill Type
Montpelier
Atlanta
Garland
Santa Clara
Austin
Job Skills Count Distribution by City and Skill Type
Montpelier
Atlanta
Garland
Santa Clara
Austin


Distribution of Technical Skills Across U.S. Cities
Technical Skills Job Count by City and Skill Type
Garland
Atlanta
Columbus
San Diego
Baytown
Analysis and Visualization Skills Job Count by City
Garland
Santa Clara
Atlanta
Bound Brook
San Antonio



Flow of Job Levels, Skills, Titles, and Locations in U.S. Business Analysis Roles
job_level
Mid senior
JobSKills
Analytical Skills
job_title
JD Edwards Business A...
search_city
TotalSoftSkillsJobCo...
4367
Mid senior
3849
Associate
518
Communication
1412
Analytical Skills
971
Problem Solving
863
Communication Skills
832
Teamwork
763
Problemsolving
527
Interpersonal Skills
395
Leadership
350
JD Edwards Business...
105
JDE Business Analyst (...
100
Business Analyst
82
Business Systems Anal...
31
Senior Business Analyst
20
BUSINESS INTELLIGEN...
18
Business Analyst II
9
IT Business Analyst
8
Texas
7
California
6
Coral Gables
3
Norristown
3
Bethany
2
Bradenton
2
Brainerd
2
Colorado
2



Analysis of Job Locations, Qualifications, and Salary Distribution Across U.S. Cities
Top 20 Cities by Total Yearly Salary
Summit $2.4M
Santa Clara $1.1M
San Diego $0.9M
San Juan Cap... $0.8M
Bayonne $0.8M
Malibu $0.7M
Concord $0.7M
Gastonia $0.7M
Pasadena $0.7M
Everett $0.7M
Eastchester $0.7M
Layton $0.6M
Austin $0.6M
Aurora $0.6M
Avondale $0.6M
Smith Center $0.5M
Virginia $0.5M
Santa Clarita $0.5M
Columbus $0.5M
San Antonio $0.5M
search_city
$0M $1M $2M
Sum of Yearly Salary
Top Job Locations by Salary
Sum of Yearly Salary
$0M $1M $2M
$2.2M $1.8M $1.4M $1.0M $0.8M $0.7M $0.7M $0.7M $0.6M $0.6M
New York, NY
Seattle, WA
Charlotte, NC
Bellevue, WA
San Diego, CA
Nashville, TN
Los Angeles, CA
Phoenix, AZ
Austin, TX
Irvine, CA
job_location
Geographical Distribution of Yearly Salary
Highest Salary and Corresponding City
Max Salary Amount
346K
City with Max Salary
Everett
Filter by Qualification Level
Qualification
Bachelor
Bachelor, Diploma
Bachelor, Diploma, Master

Project Scope

This project examines and compares salary offerings for Business Analyst roles across various sectors and regions within the United States. Using a LinkedIn-sourced dataset comprising three files, this analysis involves both text-based and tabular data with attributes such as job title, industry, salary, and location.

Given the predominantly text-based structure of one file, certain data manipulations were required to enable deeper analysis. The objective is to evaluate salary variations by industry and region, offering insights into job market trends and compensation patterns for Business Analysts.

Data Exploration

In the initial phase of data exploration, we are examining three LinkedIn datasets that provide a comprehensive view of roles. The first dataset contains information on roles and their associated skills, offering insight into the specific competencies required for various positions. The second dataset includes metadata related to the roles, such as location (city and country), seniority, and employment type (part-time or full-time), which allows for a deeper analysis of role distribution across geographic and organizational levels. Lastly, the third dataset provides text descriptions for each role, which can be analysed for common keywords or themes. This exploration will help uncover patterns, trends, and relationships across the datasets, such as the correlation between location and skill requirements or the prevalence of certain role descriptions in particular regions.

The skills file was ingested with Knime and manipulated to remove the job_id and linked that with each skill. The job_skill column are values separated via the use of commas, hence is relatively easy to split in columns. And the unpivoting the same is crucial. As well, by extracting the job_id we can now link all other files.

The file summary has key columns. Job_tittle is how we can filter to search for Business Analyst roles. For this string manipulation and splits are needed. Columns such as search_country allow for a better understanding of from which countries is this data being scrapped. Job_level after linked the file with Skills, can be used to identify the relationship between seniority and skills or qualifications required.

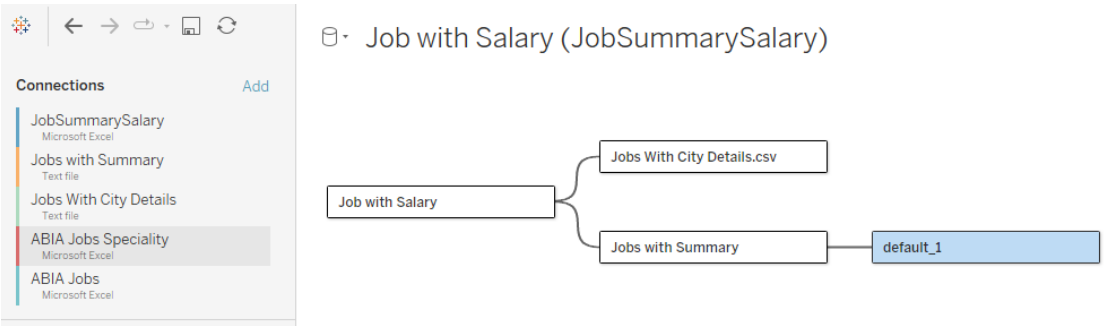
The file job_postings file is long text, basically how the role is listed on LinkedIn. And therefore, not analysed further at this stage.

Data modelling

In our analysis, we divided the dataset into soft and technical skills categories, using rule-based classification to assign job postings to specific industries. Skills were then weighted according to their frequency and demand within each industry. For processing and visualization, we used KNIME, Python, and Tableau, with Excel supporting data normalization and skill frequency comparisons.

To enable effective salary analysis by location and industry, we created columns to capture minimum, maximum, and average salary figures, representing each role's salary range. This structure allows filtering by Location to uncover geographic salary trends and by Industry to compare salary differences across various sectors. With this model, our goal is to highlight patterns in salary offerings to support a comprehensive, data-driven analysis of Business Analyst salaries by both region and industry.

Data semantic model



Analysis

Skills gap analysis

The initial analysis aims to understand the distribution between formal qualifications vs industry experience and most common skills, both technical and soft. Data exploration comparing roles requiring degrees, industry recognised certificates and none of the aforementioned was implemented to understand the market needs.

Graph for in-demand technical skills:

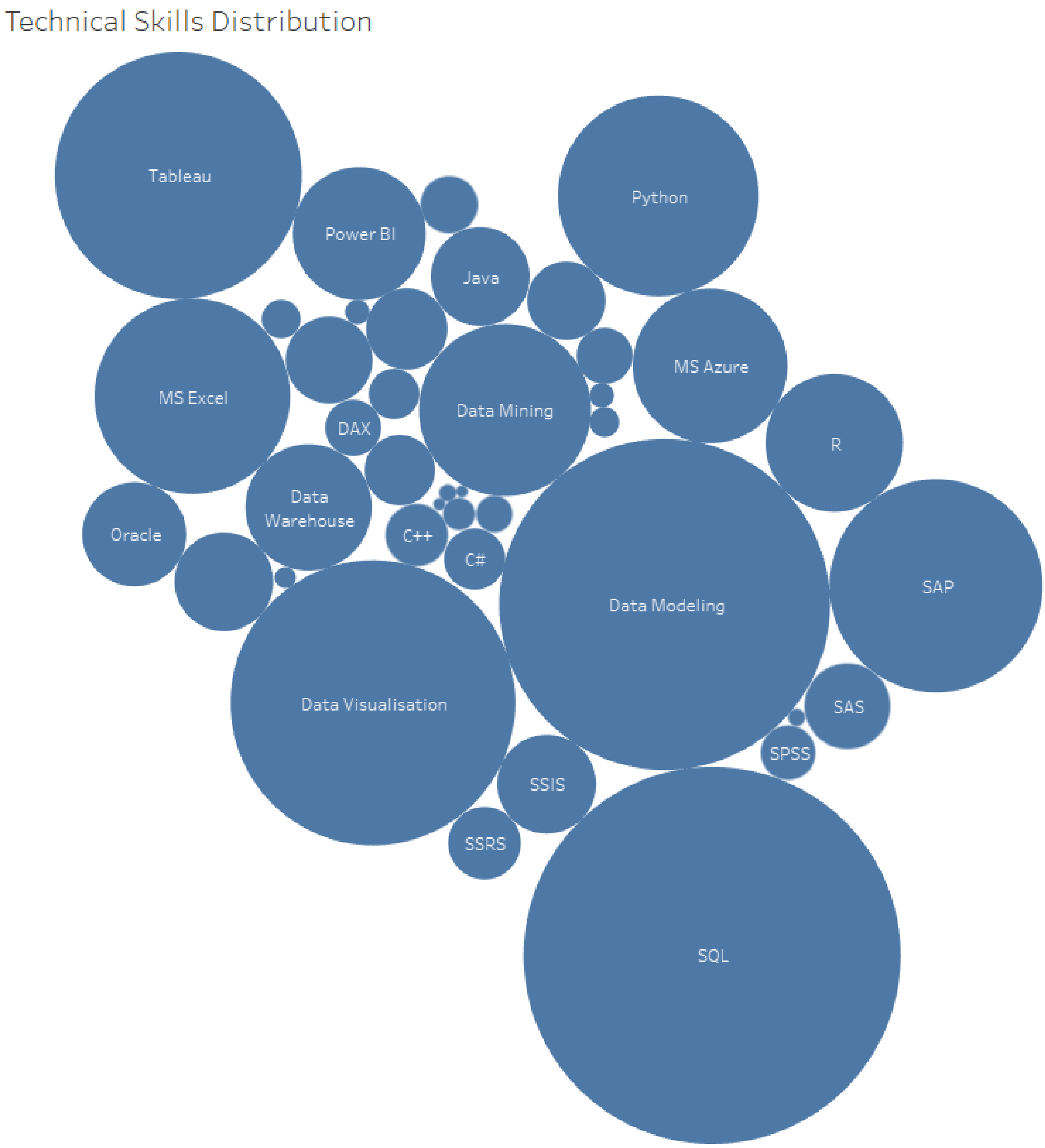


Figure 1

The graph clearly shows that data manipulation and visualization is the most desirable technical skillset. This aligns with the most in-demand skills as per The Institute of Business Analysis (IIBA, 2021).

Being the top 5 as per Figure 2:

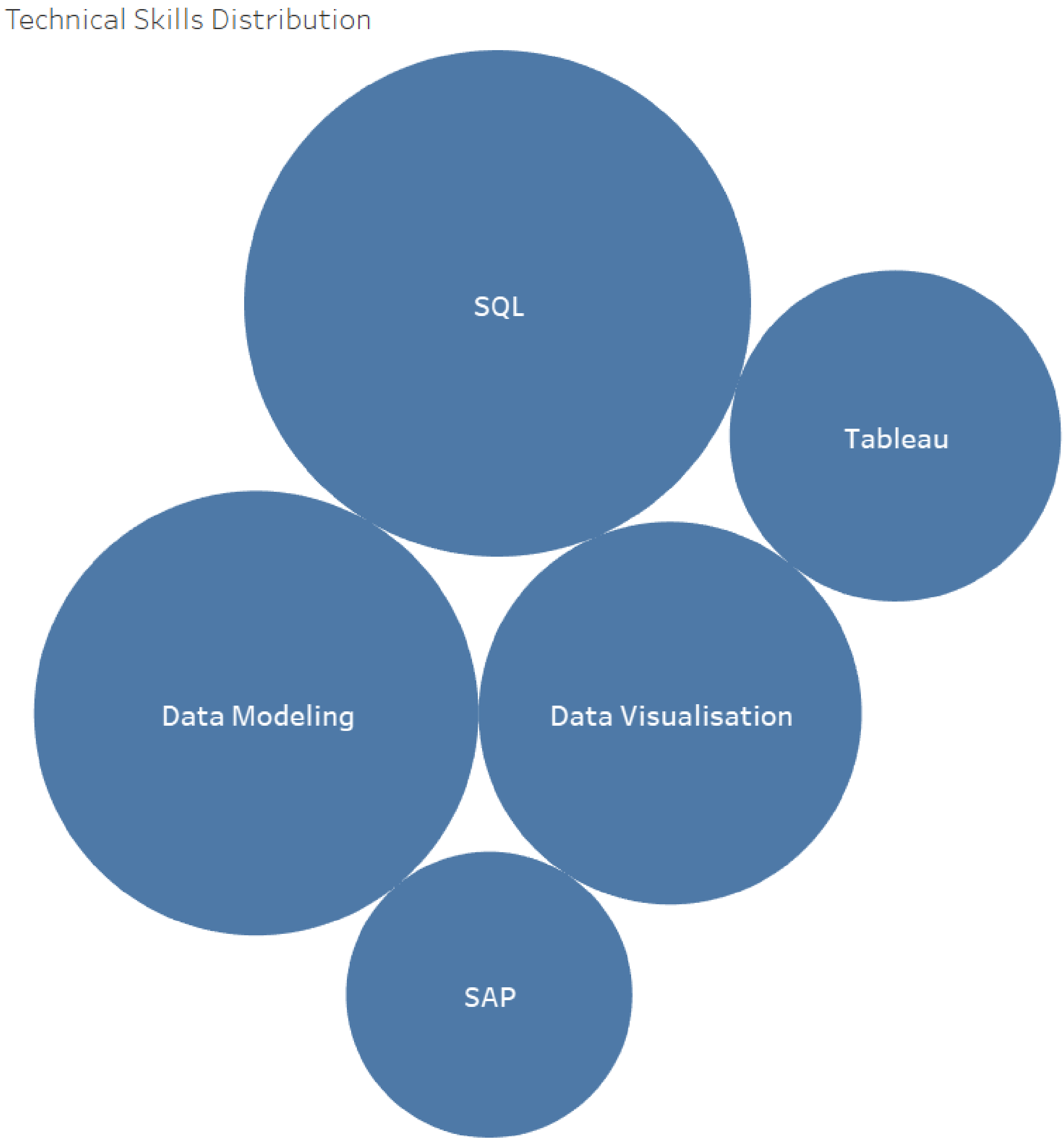


Figure 2

Which consequently aligns with the mentioned most in-demand technical skills on Coursera (Coursera, 2024)

From the pool of jobs advertised on LinkedIn only 702 out of 3.453 required a Bachelors' degree as shown in Figure 3. Which suggests that University overall is not crucial for a Business Analyst role.

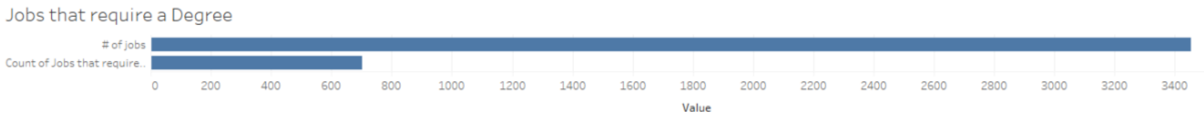


Figure 3

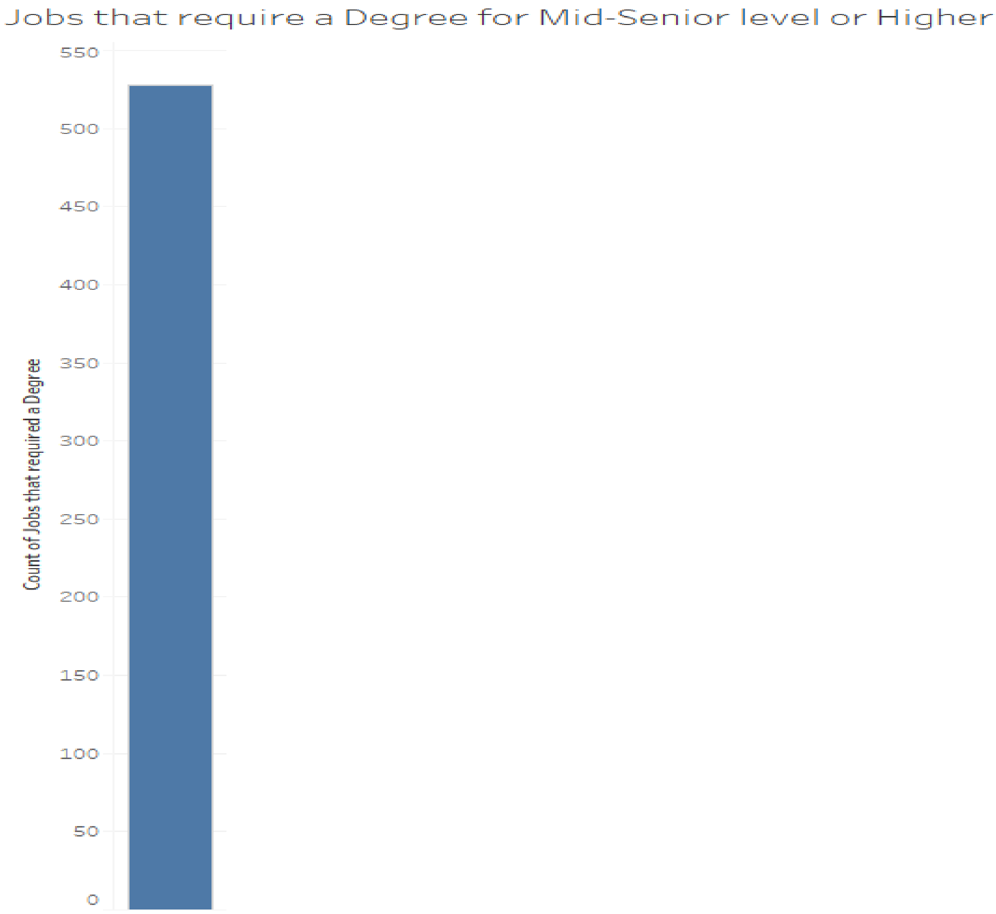


Figure 4

However, Figure 4 suggests that for Mid-Senior or higher-level roles, a degree is relevant. As out of the approximately 750 roles that required one, 530 are for the above-mentioned levels.

The pool of soft skills is as per Figure 5:


All soft skills
Project Mangement Skill
Problem-Solving Skill
Leadership Skill
Interpersonal Skill
Documentation Skill
Critical Thinking
Presentation Skill
Analytical Skill
Negotiation Skill
Multitasking Skill
Communiation Skill
Teamwork Skill
Research Skill

Figure 5

Being the top 5 as per Figure 6:

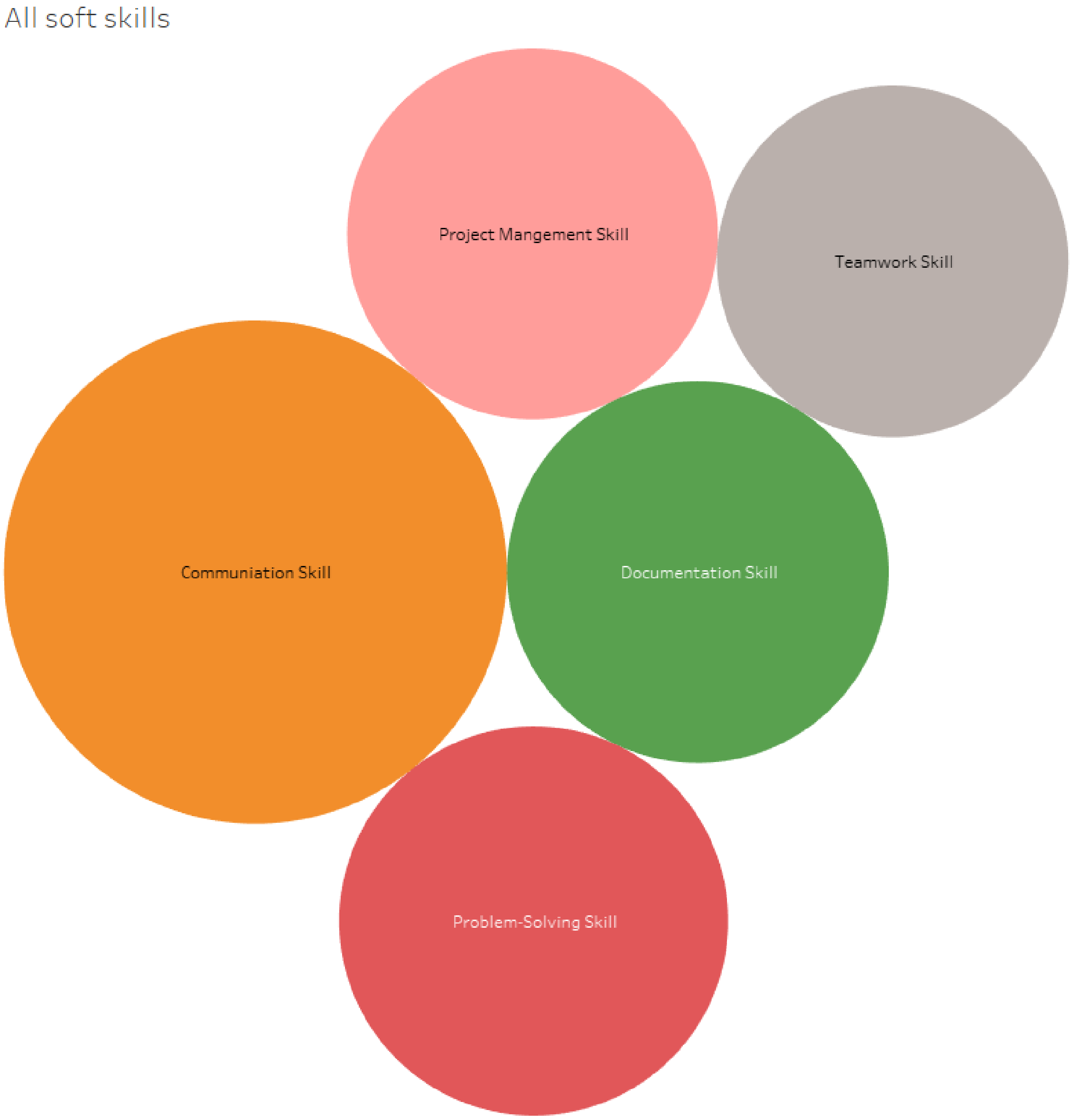


Figure 6

Based on the presented graphs we can infer that although the Business Analyst role aims to be technical, around 20% required formal higher education. Approximately 30% required an industry recognised certification.

As per Figure 7 the entire market for BA with PhD level is comprised of 6 available roles. Of which two are require for roles in the Education industry as Professors.

PhD Level Qualifications

Job Specialty	Search City	PhD	PhD, Master
Business & Finance Analyst	Austin		■
Business Analyst Professor	Aurora	■	
	Boulder	■	
Business Process Analyst	North Carolina	■	
Healthcare Business Anal..	Levittown		■
Senior Business Analyst	Tarrytown		■

Figure 7

Salary and geographical analysis

The salary and geographical analysis indicate that the top 10 cities in terms of available roles are as per Figure 8.

Job Location	
Austin, TX	8
Bellevue, WA	8
Charlotte, NC	14
Denver, CO	9
Irvine, CA	9
Los Angeles, CA	8
Manhattan, NY	9
New York, NY	30
San Diego, CA	14
Seattle, WA	18

Figure 8

New York is clearly the market with the most roles available, however from a population perspective Denver is the city with more roles per capita as per Figure 9 and Los Angeles is the most competitive market with 1 role every 479,364 habitants.

City	Roles Available	Population	Roles per capita
Austin	8	1027845	128,481
Bellevue	8	152347	19,043
Charlotte	14	910204	65,015
Denver	9	75665	8,407
Irvine	9	314621	34,958
Los Angeles	8	3834915	479,364
Manhattan	9	1645867	182,874
New York	30	8335897	277,863
San Diego	14	1428042	102,003
Seattle	18	771888	42,883

Figure 9

Salary ranges

Figure 10 shows that the range within the top 10 locations is between 95,000 thousand and 110,000 thousand a year. Which is slightly above the lowest end of the range for the Other cities however, the top end of the range is slightly below. With Other cities being closest to 120,000 thousand.

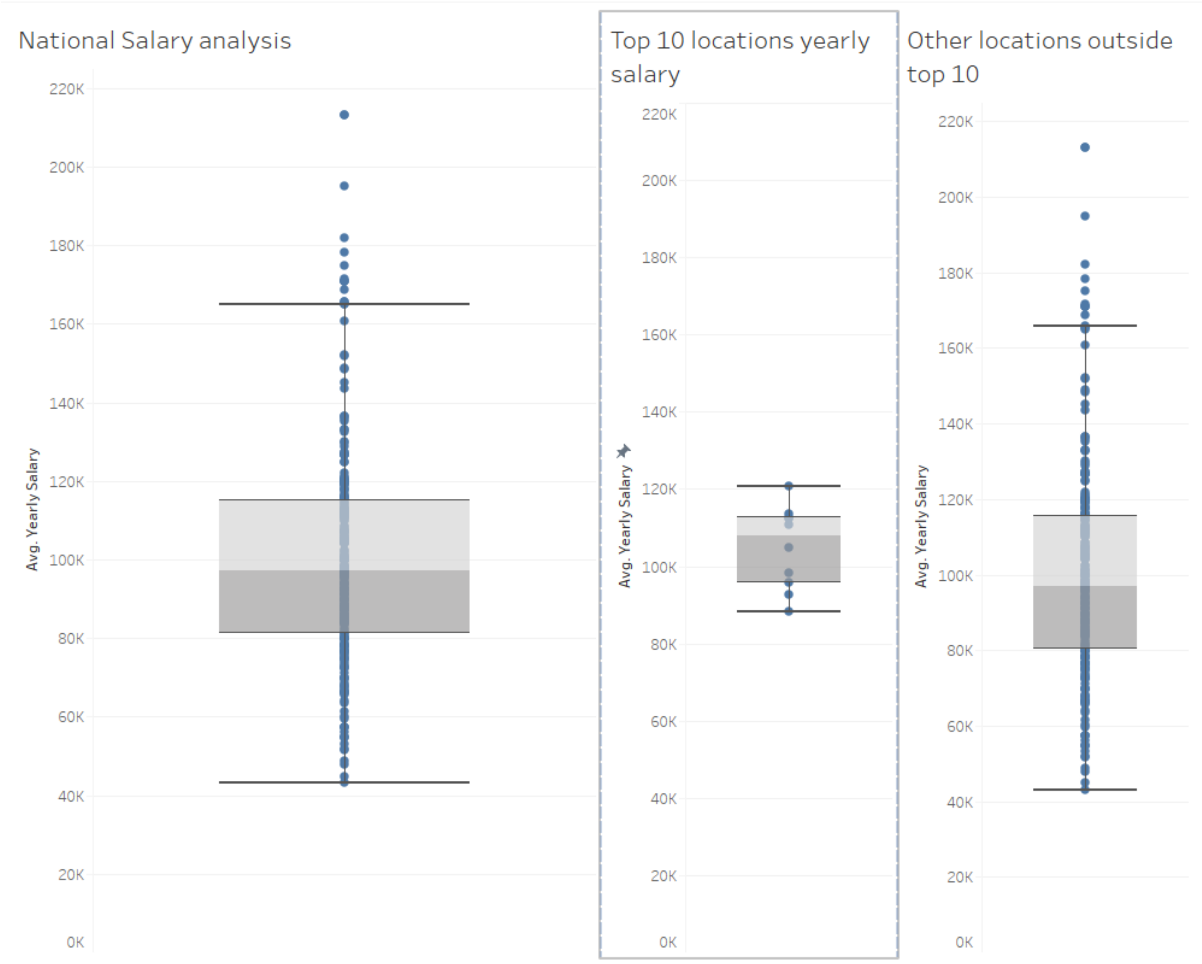


Figure 10

Other cities show a higher spread between salaries, suggesting that finding a qualified BA is harder outside of the metropolis and a higher salary becomes a decisive factor.

Conclusion

This paper examines salary trends for Business Analyst roles across various U.S. locations and industries, drawing on LinkedIn job postings and data from Kaggle. The findings provide several insights into the Business Analyst job market:

Location-Based Salary Trends:

- Regional Differences: Salary levels vary widely by location, with states like California and New York offering average salaries above $120,000. These higher figures may be due to increased demand or higher living costs in these areas.
- Demand-Driven Wages: States with a high volume of job listings, such as California and Texas, report elevated average salaries, indicating a relationship between demand and compensation. However, some areas show an oversupply of jobs with lower pay, suggesting market saturation may drive wages down in certain regions.

Industry-Specific Salary Insights:

- Top-Paying Sectors: Industries including Commercial Equipment & Repairs, Media, and Food & Beverage provide some of the highest average salaries for Business Analysts, likely due to the need for specific expertise within these fields.
- Demand in IT and Outsourcing: The IT and Outsourcing sectors exhibit strong demand for Business Analysts, reflecting broader trends in technological advancement and the expansion of service industries.
- Salary Ranges: While many industries offer average salaries in the $60,000 to $90,000 range, certain specialized fields, particularly finance and technology, report figures exceeding $120,000.

Additional Insights from Whisker Plots:

Whisker plot analysis suggests that Business Analysts may benefit from seeking roles outside the top 10 highest-paying locations. Salaries in other regions often present greater potential for competitive compensation when accounting for the cost of living and job demand.

In conclusion, this analysis sheds light on salary expectations for Business Analysts across the U.S. Higher-paying locations typically reflect high demand, though other factors, such as specialization and local cost structures, also influence compensation. These findings can help both job seekers and employers align salary benchmarks within the Business Analyst job market.

Data Quality Analysis

Completeness

Number of Records:

- LinkedIn Job Posting: 1,348,454
- Job Summary: 1,297,332 (Missing = 51,122, 4%)
- Job Skills: 1,296,381 (Missing = 52,073, 4%)

Pattern of Missing Data: The missing records are spread evenly across both job summaries and job skills, indicating that about 4% of the job postings are incomplete in both these fields.

Impact: While 96% completeness is a strong score, the missing 4% could still affect insights into certain roles or skills. Further investigation is needed to see if critical job postings (e.g., for high-demand roles like Business Analyst) are disproportionately affected.

Conclusion: With a combined completeness score of 96%, the data is largely reliable, though further attention to specific gaps (e.g., job skills) might be warranted, especially if the missing data shows any patterns across certain roles or regions.

Accuracy

The files have been populated using a web scrapper on LinkedIn. Job_skills and Job_summary are well structure and semi-standardised, hence are in conditions for analysis and can be used for the project. The file likedin_job_postings however, is text based and has symbols mixed within the words that create a difficulty for analysis. As well with different formats for the text among listings.

Consistency

All three files are in csv format. Job_summary is the best formatted one, all metadata is broken into columns with relevant types. Job_skills is only two columns with the URL in one and all skills associated with it on the other. And these are separated with a comma.

Uniqueness

The files have a job id on the URL for each role. Hence is possible to join the files via it.

Relevance

On the Job_summary file there are some irrelevant columns for the analysis. Examples are 'is_being_worked' and 'got_ner'.

Timeliness

The data is current, there is a column in the Job_summary file with the timestamp of when the record was last processed, and the oldest one is from '2024-01-19'.

Data Dictionary:

Job_title

Definition: The title of the job listed in the LinkedIn.

Format: String

Comment: The job_title was standarised and is used to determine the most common ones both overall and segregated by region.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_link

Definition: The source link from which the role was scrapped.

Format: String

Comment: This is the original link from which the role was source. User to link roles with skills.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_skills

Definition: raw version of the skills set required for each role.

Format: String

Comment: From the job_skills file. Unfiltered and unstandardised. Use to create the JobSkills column.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

JobSkills

Definition: A standarised version of the skills list.

Format: String

Comment: The list from one of the first workable files, standarised. This allowed a first glance of the most common technical and none technical skills.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Search_country

Definition: The country where the role is advertised.

Format: String

Comment: This extract of the country from the location of the role allows the user to understand from which countries is the data being sourced.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Search_city

Definition: The city where the role is advertised.

Format: String

Comment: This extract of the city from the location of the role allows the user to understand from which cities is the data being sourced. Which is then used to identify hotspots.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

Job_level

Definition: The seniority of the role

Format: String

Comment: This column is use to determine the relationship between seniority and skills set and qualifications required.

Source: Asaniczka, (2024). https://www.kaggle.com/datasets/asaniczka/1-3m-linkedin-jobs-and-skills-2024

References

Coursera, 2024. *Coursera.* [Online]
Available at: https://www.coursera.org/articles/what-does-a-business-analyst-do-and-how-to-become-one
[Accessed 14 September 2024].

IIBA, 2021. *International Institute of Busines Analysis.* [Online]
Available at: https://www.iiba.org/business-analysis-blogs/in-demand-business-analysis-skills-for-2021/
[Accessed 14 September 2024].

Appendix

Mind Map