Data Analysis - Data Cleaning and Analysis - Data Analyst Job Listings Documentation

Introduction

In the wake of the global pandemic, which brought about widespread job losses and economic turbulence, I thought to build a project about the "Data Analyst Job Listings." Crafted by picklesueat, this meticulously compiled dataset contains a repository of over 2000 data analyst job listings, providing a glimmer of hope for individuals embarking on their job search journey.

In a world increasingly reliant on data-driven decision-making, data analysts have become indispensable, playing a crucial role in deciphering complex datasets and guiding organizations toward informed choices. The "Data Analyst Job Listings Documentation" dataset serves as a bridge between job seekers and employers, offering opportunities amid adversity.

Data Summary

This dataset is a valuable resource, offering a diverse range of features such as comprehensive job descriptions, required qualifications, detailed company profiles, and geographic placement. These attributes are tailored to assist both aspiring data analysts and those seeking to transition within the field.

Job Title: This column contains the job titles for various data analyst positions. It provides information about the specific roles being offered.
Salary Estimate: This column indicates the estimated salary range for each job listing. It provides insights into the potential earnings associated with the positions.
Job Description: This column includes detailed job descriptions, outlining the responsibilities and requirements for each data analyst role. It offers a comprehensive overview of the job expectations.
Rating: This column displays ratings associated with the hiring companies or organizations. It reflects the company's reputation or performance as rated by employees or other sources.
Company Name: This column lists the names of the companies or organizations offering the data analyst positions. It identifies the potential employers.
Location: This column specifies the location of the job listings, indicating where the positions are based. It allows job seekers to consider geographic preferences.
Headquarters: This column provides information about the headquarters location of the hiring companies. It may be relevant for those interested in the company's geographic presence.
Size: This column indicates the size of the companies in terms of the number of employees. It offers insights into the scale of the hiring organizations.
Founded: This column mentions the year in which the hiring companies were founded. It provides historical context about the organizations.
Type of Ownership: This column describes the ownership structure or status of the hiring companies, such as whether they are nonprofit organizations, publicly traded companies, privately owned, etc.
Industry: This column specifies the industry to which the hiring companies belong. It categorizes the companies based on their primary business activities.
Sector: This column further categorizes the hiring companies into sectors, offering

additional information about their industry classification.
Revenue: This column provides details about the revenue of the hiring companies, typically in terms of millions or billions of US dollars. It gives insights into the financial scale of the organizations.
Competitors: This column may list competitors of the hiring companies, providing information about the competitive landscape within the industry.
Easy Apply: This column indicates whether the job listing offers an easy application process, potentially simplifying the application procedure for candidates.

Data Preprocessing

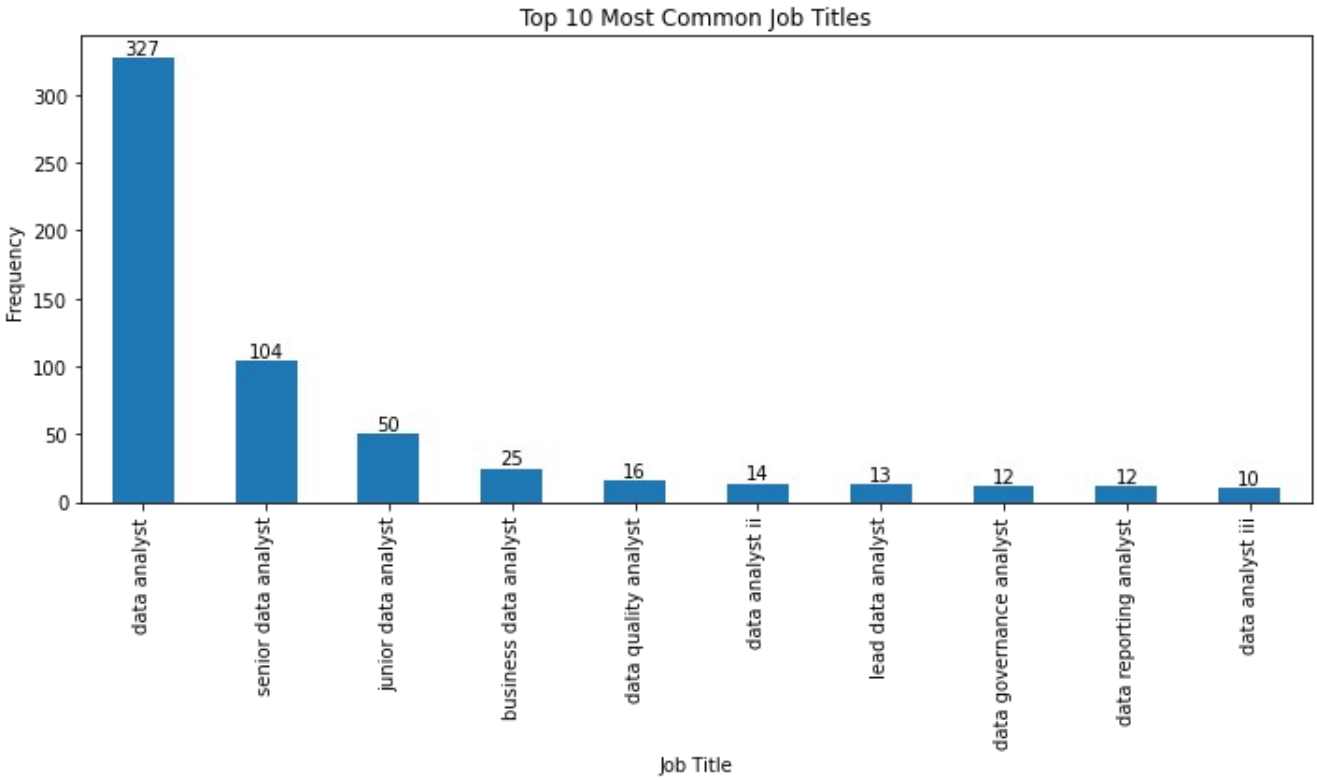
I dropped the "Unnamed: 0" and "Competitors" columns from the dataset as they didn't contain relevant information for my analysis.
Next, I checked for and removed any rows with missing or null values to ensure the dataset's quality and completeness.
I also examined the dataset for duplicate records and eliminated them to ensure that each job listing was unique and free from redundancy.
I standardized the representation of missing or invalid values by replacing various forms of -1, -1.0, and '-1' with NaN (Not-a-Number) values throughout the dataset.
To simplify text analysis and comparisons, I converted all the text data to lowercase, ensuring uniformity.
I divided the "Location" column into two separate columns, "City" and "State," providing more detailed geographic information for each job listing.
To simplify company names, I removed any text that appeared after double quotation marks ("") in the "Company Name" column.
For the "Easy Apply" column, I converted the values to boolean True or False, with -1 mapped to False, making it easier to understand easy application options.
I derived two new columns, "Maximum Salary" and "Minimum Salary," from the "Salary Estimate" column to facilitate salary analysis.
I analyzed job descriptions in the "Job Description" attribute to identify and extract specific skills mentioned, allowing for categorization or searching based on required skills.
I standardized job titles by grouping similar titles together; for example, "Senior Data Analyst" and "Sr. Data Analyst" were treated as equivalent.
Finally, I dropped any remaining rows with missing values after the previous preprocessing steps to ensure a clean and organized dataset ready for analysis.

These preprocessing steps have helped clean and structure the dataset, making it more suitable for analysis and modeling. The result is a cleaner, more organized, and more informative dataset that can yield valuable insights when exploring job listings for data analyst positions.

Question 1: What are the most common job titles in the dataset?

Data Analyst: With 327 occurrences, the "Data Analyst" job title is the most common in the dataset. This suggests a significant demand for professionals who specialize in analyzing and interpreting data to provide insights and support decision-making processes. It indicates that organizations across various industries value individuals who can extract meaningful information from data.
Senior Data Analyst: The "Senior Data Analyst" job title appears 104 times in the

dataset. This indicates that there is a notable demand for experienced data analysts who possess advanced skills and knowledge in data analysis techniques, statistical modeling, and data visualization. Organizations may require senior data analysts to lead projects, mentor junior analysts, and provide strategic insights based on data analysis.
Junior Data Analyst: The presence of 50 listings for "Junior Data Analyst" suggests that there are opportunities for entry-level professionals in the field of data analysis. These roles may be suitable for recent graduates or individuals with limited work experience in data analysis. Organizations may offer training and mentorship to develop the skills and expertise of junior data analysts.
Business Data Analyst: The "Business Data Analyst" job title appears 25 times in the dataset. This indicates a specific focus on data analysis within a business context. Business data analysts typically work closely with stakeholders from various departments to analyze data related to business operations, market trends, and customer behavior. Their role involves translating data insights into actionable recommendations for improving business performance.
Data Quality Analyst: With 16 occurrences, the "Data Quality Analyst" job title highlights the importance of ensuring data accuracy, consistency, and reliability. These professionals are responsible for assessing and improving the quality of data by identifying and resolving data issues, implementing data governance practices, and ensuring compliance with data standards and regulations.
And so on



Question 2: Which sectors and industries have the highest number of job listings?

The analysis reveals the sectors and industries that showcase the highest number of job listings, providing valuable insights into the areas of robust employment

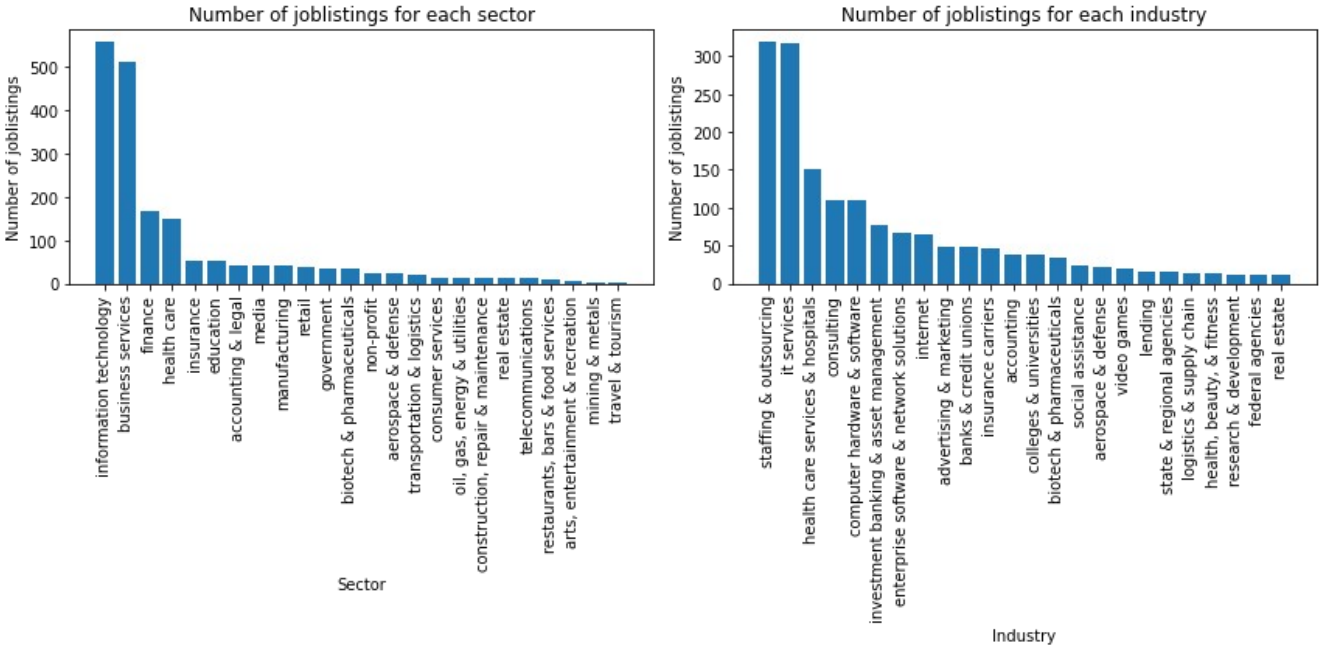
opportunities. These sectors and industries attract a significant demand for professionals with diverse skill sets and expertise. Let's explore the top sectors and industries with the highest number of job listings:

Sectors:

Information Technology: With an impressive 558 job listings, the information technology sector stands out as the most thriving, reflecting the continuous growth and demand for technology professionals.
Business Services: Following closely with 511 job listings, the business services sector presents ample opportunities across various domains, offering a wide range of professional services to support organizational operations.
Finance: Boasting 166 job listings, the finance sector demonstrates a strong presence, indicating a demand for professionals skilled in finance, banking, and investment.
Health Care: With 150 job listings, the health care sector presents promising prospects, showcasing a multitude of opportunities in hospitals and health care services.
Insurance: The insurance sector claims its spot on the list with 51 job listings, offering career possibilities in insurance-related roles and functions.

Industries:

Staffing & Outsourcing: Dominating the list with 319 job listings, the staffing and outsourcing industry provides abundant job opportunities for individuals interested in recruitment and workforce management.
IT Services: In a close second with 317 job listings, the IT services industry exhibits its vitality, highlighting the persistent demand for professionals providing a wide array of technology-related services.
Health Care Services & Hospitals: With 150 job listings, the health care services and hospitals industry ensures a steady stream of job opportunities within health care facilities and service providers.
Consulting: The consulting industry secures its place on the list with 110 job listings, attracting professionals who can offer expert advice and guidance across various domains.
Computer Hardware & Software: Sharing the same number of job listings as consulting, this industry signifies the demand for professionals skilled in computer hardware and software development, powering technological advancements.
These findings shed light on the sectors and industries that exhibit a higher concentration of job listings, indicating areas of growth and potential career paths for individuals seeking employment opportunities.



Question 3: Which industries and sectors have the highest average salaries?

When considering career opportunities, understanding the industries and sectors that offer higher average salaries is essential. This analysis unveils the industries and sectors that provide professionals with the potential for attractive financial rewards. Let's explore the industries and sectors with the highest average salaries:

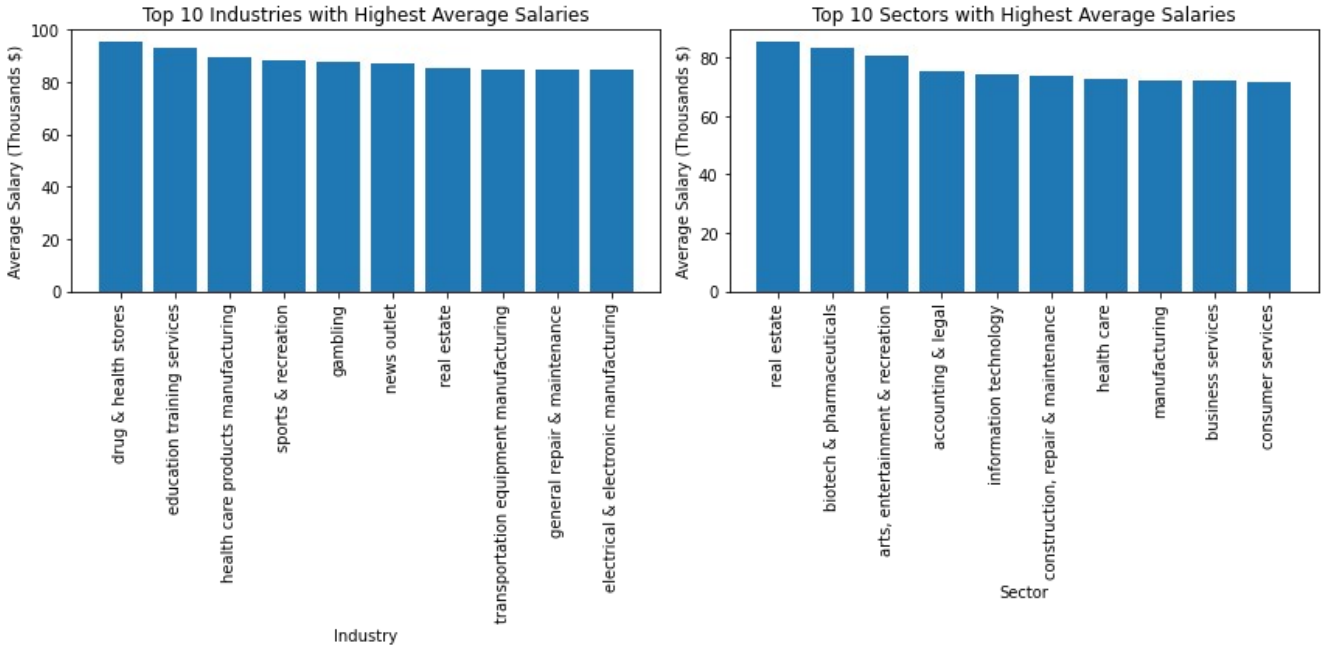
Industries:

Drug & Health Stores: Emerging as the top-paying industry, professionals in drug & health stores enjoy an average salary of 95,250. This industry encompasses businesses specializing in pharmaceuticals and health-related products, recognizing the value of expertise in this field.
Education Training Services: Following closely with an average salary of 92,833, the education training services industry emphasizes the significance placed on qualified educators and trainers who play a vital role in shaping future generations.
Health Care Products Manufacturing: With an average salary of 89,800, the health care products manufacturing industry acknowledges the value of professionals involved in producing essential health care products.
Sports & Recreation: Professionals in the sports & recreation industry command an average salary of 88,166, reflecting the financial rewards associated with careers in this dynamic and engaging field.
Gambling: Offering an average salary of 88,000, the gambling industry recognizes the importance of skilled professionals within the gaming and betting sectors.

Sectors:

Real Estate: Topping the list with an average salary of 85,227, the real estate sector provides lucrative opportunities for professionals involved in property management, sales, and development.

Biotech & Pharmaceuticals: With an average salary of 83,106, the biotech and pharmaceuticals sector showcases competitive remuneration for professionals engaged in research, development, and manufacturing within this innovative industry.
Arts, Entertainment & Recreation: Boasting an average salary of 80,643, the arts, entertainment, and recreation sector celebrates the financial rewards associated with careers in creativity, entertainment, and recreational activities.
Accounting & Legal: With an average salary of 75,221, the accounting and legal sector offers attractive compensation for professionals specializing in finance, accounting, and legal roles.
Information Technology: Recognized for its importance, the information technology sector provides professionals with an average salary of 74,302, reflecting the high demand for skilled individuals in various technology-related roles.
These findings shed light on the industries and sectors that offer higher average salaries, presenting promising paths for professionals seeking financially rewarding career opportunities.

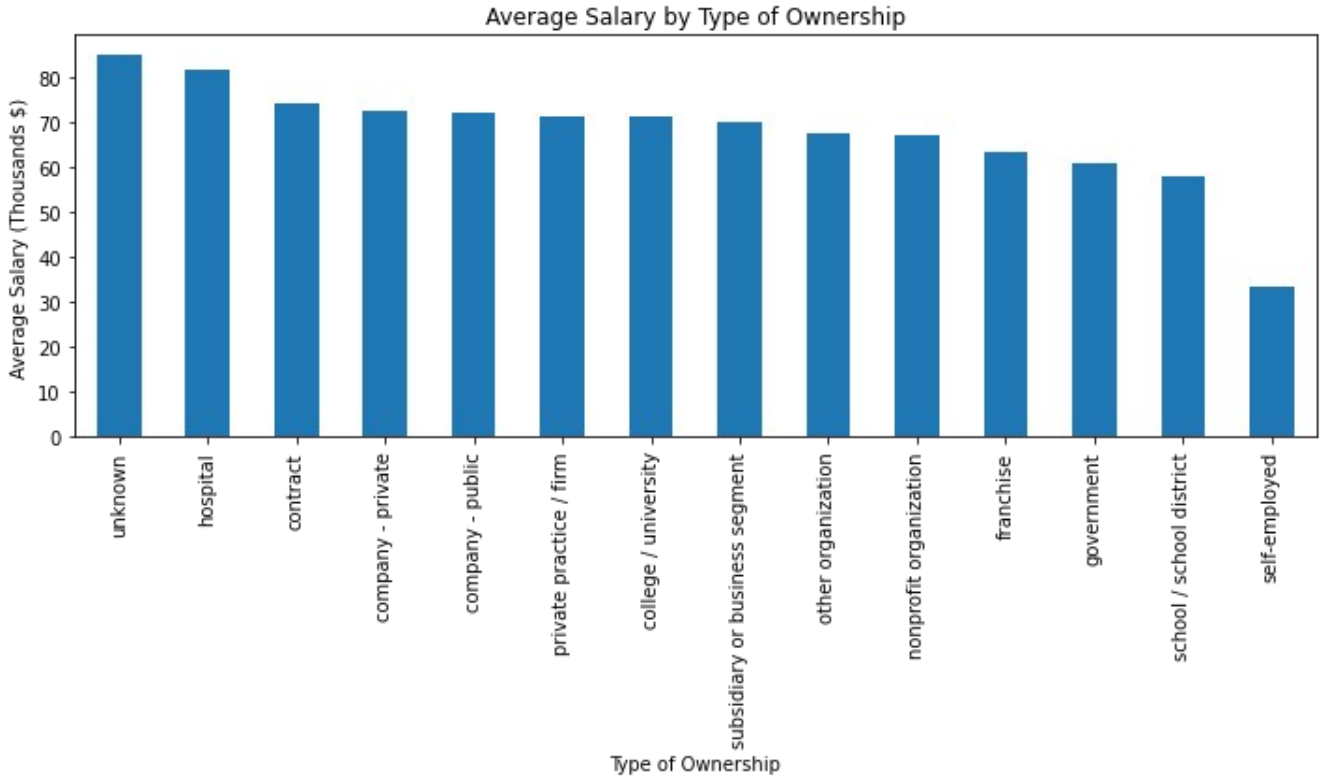


Question 4: How does the type of ownership (e.g., public, private, non-profit) relate to the average salary?

The average salary varies across different types of ownership, providing insights into the potential salary range based on the ownership structure.

Unknown: The highest average salary is observed for job listings with an unknown type of ownership, with an average salary of 85,300. This category may include companies that do not disclose their ownership structure or have a unique ownership arrangement.
Hospital: Job listings associated with hospitals show the second-highest average salary at 81,710. Hospitals are typically large organizations operating in the healthcare sector and are known for offering competitive salaries to attract and retain skilled professionals.
Contract: Job listings categorized as contract positions have an average salary of 74,250. Contract roles often offer specialized expertise on a project basis and may

command higher compensation due to the short-term nature of the employment.
Company - Private: The average salary for job listings in privately-owned companies is 72,730. Private companies vary in size and industry, and their compensation packages can be influenced by factors such as financial performance, industry norms, and the company's growth stage.
Company - Public: Job listings in publicly-owned companies have an average salary of 72,390. Public companies are often subject to market pressures and regulatory requirements, which may influence their compensation practices.
Private Practice / Firm: Job listings associated with private practices or firms have an average salary of 71,610. This category may include professions like law, medicine, consulting, or other specialized services that operate as independent practices or small partnerships.
College / University: The average salary for job listings in the college/university sector is 71,360. Academic institutions often have structured salary scales based on education levels, experience, and faculty ranks.
Subsidiary or Business Segment: Job listings related to subsidiaries or business segments have an average salary of 70,300. These positions are typically part of larger organizations, and salary structures may be influenced by the parent company's policies and industry dynamics.
Other Organization: Job listings associated with other types of organizations have an average salary of 67,860. This category may include diverse organizations such as professional associations, research institutes, or cultural institutions.
Nonprofit Organization: Nonprofit organizations offer an average salary of 67,210. While salaries in the nonprofit sector can vary, they are often influenced by factors such as the organization's funding sources, mission-driven work, and budget constraints.



Question 5: Are there any cities or states that have a higher concentration of job listings?

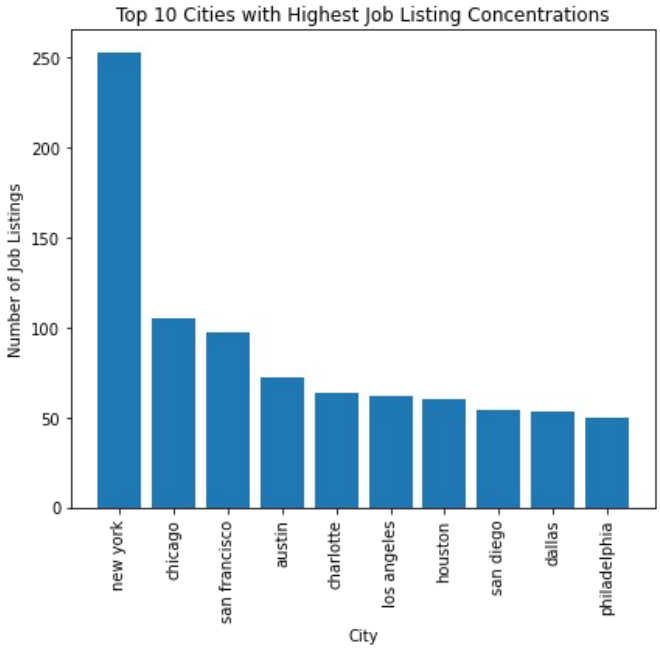
Exploring the distribution of job listings across different cities and states can provide valuable insights into the areas with higher concentrations of employment opportunities.

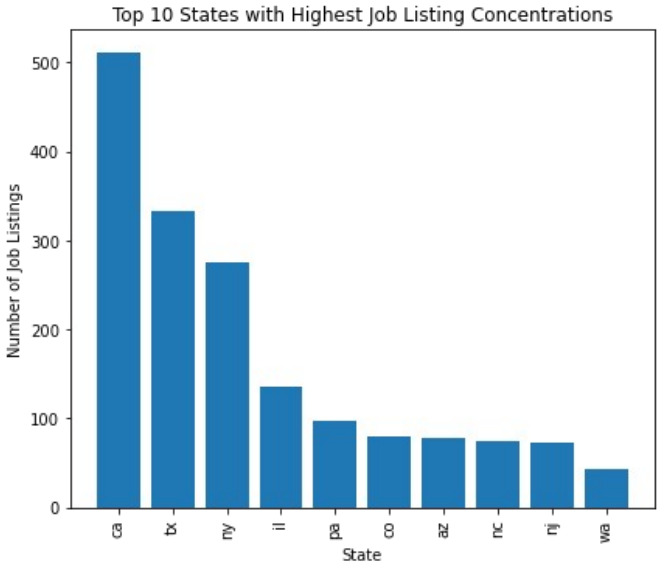
Cities:

New York: Topping the list is New York, with a significantly higher concentration of job listings at 253. This indicates a robust job market in various industries within the city, attracting a large number of employers and job seekers.
Chicago: With 105 job listings, Chicago holds the second position, showcasing a considerable concentration of employment opportunities in the city.
San Francisco: Known for its thriving tech and innovation scene, San Francisco follows closely with 97 job listings, indicating a high concentration of opportunities in this hub of technological advancements.
Austin: The city of Austin demonstrates a notable presence with 72 job listings, highlighting its emergence as a vibrant and growing job market.
Charlotte: Charlotte showcases a concentration of 64 job listings, suggesting a strong job market in this city.

States:

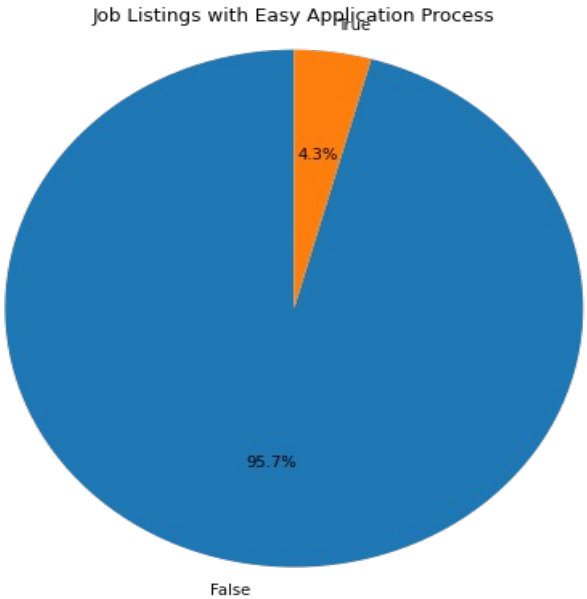
California (CA): The state of California exhibits the highest concentration of job listings, with a staggering 511 positions. This reflects the state's diverse economy, including technology, entertainment, healthcare, and other thriving industries.
Texas (TX): Texas follows with 332 job listings, indicating a substantial job market spread across various cities within the state.
New York (NY): New York state ranks third with 276 job listings, highlighting the presence of diverse industries and employment opportunities.
Illinois (IL): With 136 job listings, Illinois showcases a significant concentration of employment opportunities, primarily driven by Chicago, a major economic center in the state.
Pennsylvania (PA): Pennsylvania has 97 job listings, suggesting a concentration of job opportunities in this state.
These results indicate that certain cities and states have a higher concentration of job listings, indicating a stronger job market and a potentially larger pool of employment opportunities in these locations.





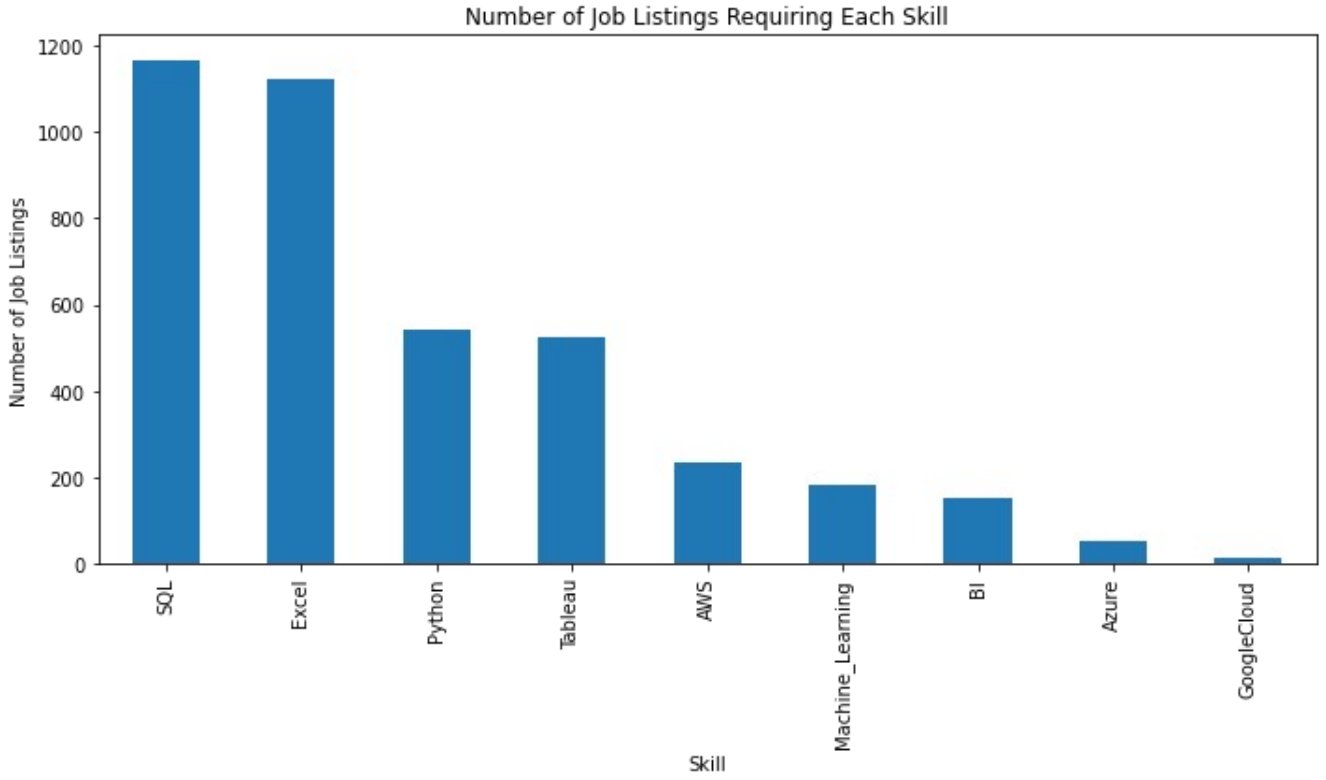
Question 6: How many job listings offer an easy application process?

Out of the total job listings analyzed, 80 of them, which corresponds to approximately 4.3% of the listings, offer an easy application process. On the other hand, the majority of job listings (1,779 or 95.7%) do not provide a streamlined or simplified application process.
This finding suggests that while a small portion of job listings prioritize ease of application, the majority of listings may require additional steps or qualifications for applicants. Job seekers should be prepared to navigate a more traditional application process for the majority of opportunities in this dataset.



Question 7: What is the most required skill for a data analyst?

Among the listed skills, SQL and Excel emerge as the two most required skills for data analyst roles. SQL is the most commonly required skill, appearing in 1166 job listings, followed closely by Excel, which is required in 1123 listings. These findings highlight the importance of data querying and manipulation skills (SQL) as well as proficiency in spreadsheet tools (Excel) for data analysis roles.
Python is the next most frequently required skill, appearing in 542 job listings. Python's popularity in the data analysis field stems from its versatility, ease of use, and extensive libraries for data manipulation and analysis.
Tableau is also a sought-after skill for data analysts, appearing in 526 job listings. Tableau is a powerful data visualization tool that allows analysts to present insights and findings in a visually appealing and interactive manner.
Other skills that are in demand, although to a lesser extent, include AWS (Amazon Web Services) mentioned in 234 job listings, Machine Learning in 181 listings, BI (Business Intelligence) in 151 listings, Azure in 54 listings, and GoogleCloud in 13 listings.



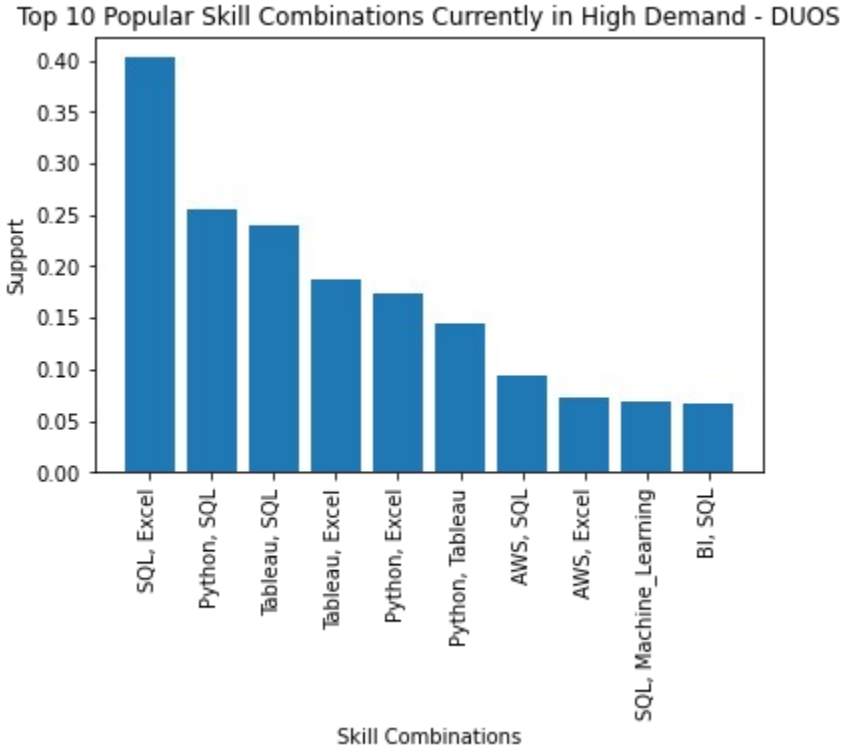
Question 8: What is the most required group of skills for a data analyst?

By examining the skill combinations in demand for data analyst roles, we can gain insights into the most required group of skills that employers seek in candidates. Let's explore the popular skill combinations, both in pairs and trios, to identify the skill sets that are highly valued for data analyst positions.

Popular skill combinations in duos:

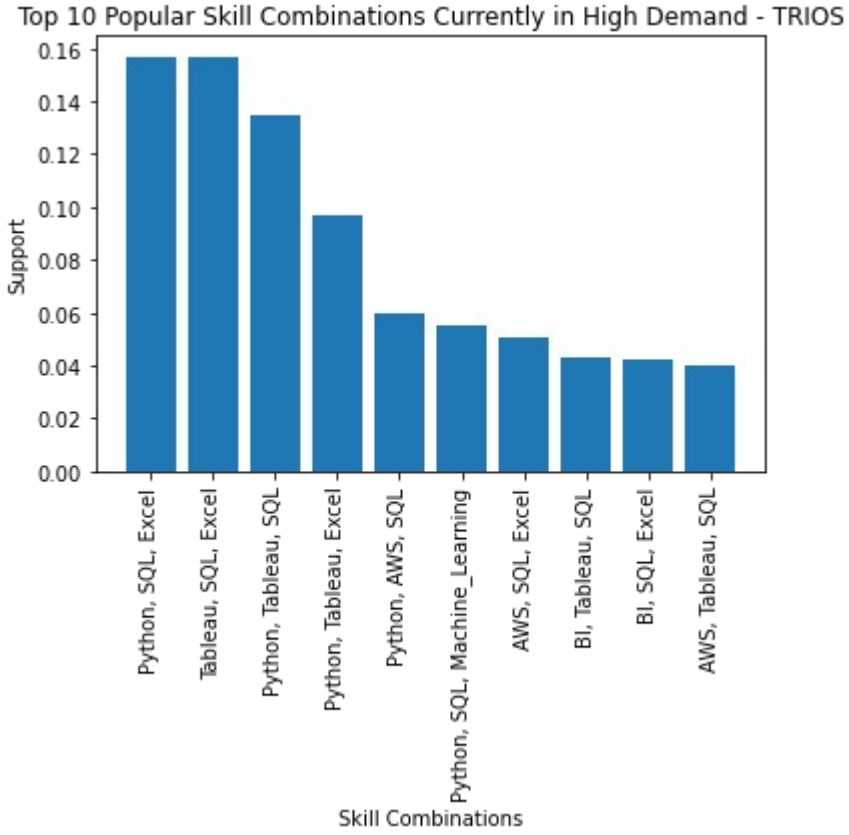
The most sought-after duo is SQL and Excel, appearing in 40.2% of the skill combinations. This indicates that proficiency in SQL for data querying and Excel for

data analysis and manipulation are highly desired skills.
The combination of Python and SQL is also in high demand, appearing in 25.6% of the skill combinations. This suggests that proficiency in both programming (Python) and database querying (SQL) is valued for data analyst roles.
Tableau and SQL form another popular duo, appearing in 24.0% of the skill combinations. This suggests that expertise in data visualization (Tableau) alongside data querying (SQL) is desirable.



Popular skill combinations in trios:

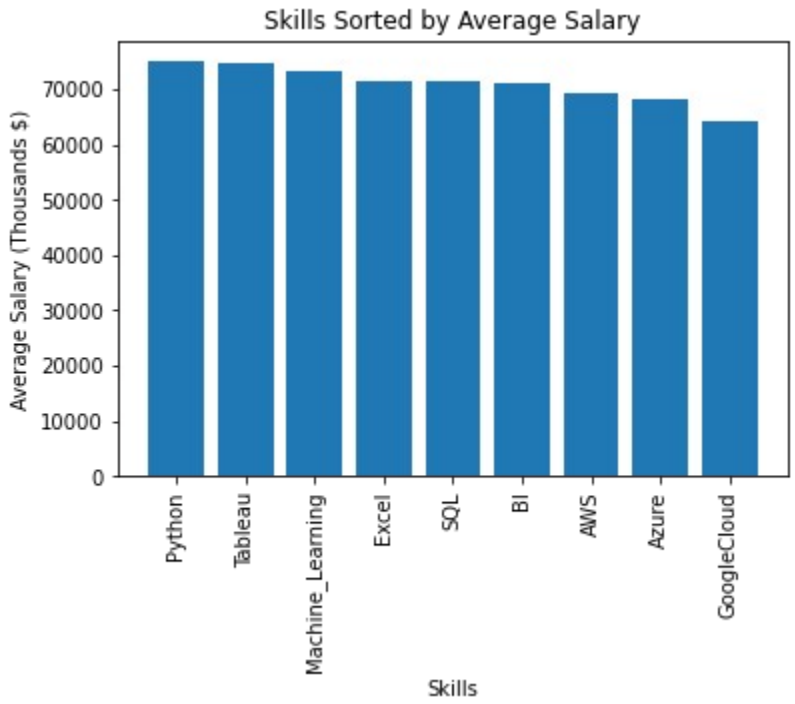
The trio of Python, SQL, and Excel is the most required combination, appearing in 15.7% of the skill combinations. This highlights the importance of having skills in data manipulation and analysis (Python and SQL) as well as data organization and visualization (Excel).
Tableau, SQL, and Excel form another popular trio, appearing in 15.6% of the skill combinations. This indicates the value of expertise in data visualization (Tableau) alongside data querying (SQL) and data analysis (Excel).
Python, Tableau, and SQL appear together in 13.4% of the skill combinations, suggesting that proficiency in these three skills is highly sought after for data analyst roles.
These results indicate that a combination of technical skills such as SQL, Python, and Tableau, along with proficiency in data manipulation (Excel), data visualization, and data analysis, are in high demand for data analyst positions. Job seekers looking to excel in this field should consider acquiring and showcasing these skills to increase their competitiveness in the job market.



Question 9: What is the highest paying skill to have?

In exploring the salary aspect of different skills for data analysts, we can uncover which skill stands out as the highest paying in the field. Let's analyze the average salaries associated with each skill to determine which skill holds the most financial value in the data analyst job market.

Among the listed skills, Python appears to be the highest paying skill for data analysts, with an average salary of 75,007 dollars. Python's versatility and widespread use in data analysis, machine learning, and automation contribute to its high demand and higher salary offerings.
Tableau follows closely, with an average salary of 74,655 dollars. Tableau's proficiency in data visualization and its ability to create interactive dashboards and reports make it a valuable skill in the data analyst field.
Machine Learning is another highly paid skill, with an average salary of 73,348 dollars. As companies increasingly leverage machine learning techniques to derive insights and make data-driven decisions, professionals with expertise in this field are in demand and rewarded accordingly.
Excel and SQL are also among the top paying skills, with average salaries of 71,597 dollars and 71,521 dollars, respectively. Proficiency in Excel for data analysis, reporting, and manipulation, as well as expertise in SQL for database querying and data extraction, continue to be highly valued in data analyst roles.
Other skills such as BI (Business Intelligence), AWS (Amazon Web Services), Azure, and GoogleCloud also command competitive salaries in the data analyst field, with average salary ranges from 64,192 dollars to 71,046 dollars.



Conclusion

In conclusion, the "Data Analyst Job Listings Documentation" project has provided valuable insights into the data analyst job market. As a data analyst, I recognize the significance of these findings in guiding both job seekers and organizations in this ever-evolving landscape.

First and foremost, the dataset's examination revealed that "Data Analyst" is the most common job title, highlighting the strong demand for professionals who can analyze and interpret data to inform decisions. This reflects the increasing reliance on data-driven decision-making across industries.

Furthermore, the dataset unveiled the sectors and industries with the highest number of job listings. Information Technology, Business Services, and Finance emerged as thriving sectors, offering diverse opportunities to data analysts. Understanding these trends can help job seekers tailor their career paths.

Salary analysis based on industries and sectors demonstrated that certain fields offer more competitive compensation. For instance, professionals in Drug & Health Stores, Education Training Services, and Health Care Products Manufacturing tend to enjoy higher average salaries, underlining the financial rewards associated with expertise in these areas.

The dataset also explored how the type of ownership relates to salary, with hospitals and unknown ownership structures offering some of the highest average salaries. This insight helps job seekers make informed decisions based on their salary expectations and preferences.

By delving into the concentration of job listings across cities and states, the dataset identified key hubs like New York and California as hotspots for job opportunities. This knowledge allows job seekers to target locations with a higher density of job listings.

Regarding the ease of application, the majority of job listings do not offer a streamlined application process. While only a small percentage of listings provide an easy application option, job seekers can be prepared for a more traditional application

process.
The analysis of required skills revealed that SQL and Excel are the most commonly sought-after skills for data analysts. Job seekers can prioritize honing these skills to align with industry demands.
Lastly, understanding the highest paying skill is essential for career planning. Python emerged as the highest paying skill, underscoring its versatility and significance in data analysis roles.

In an era where data-driven insights drive success, this dataset offers crucial information for aspiring data analysts and organizations seeking talent. It equips job seekers with knowledge about job titles, sectors, salaries, and skills, empowering them to make informed career choices in the dynamic field of data analysis.