

Web Scraping on Job Market



FALL19' Data Analytics
Final Project



Data Analyst University of California San Francisco ★★★★★ 418 reviews San Francisco, CA The Data Analyst works on research data reporting assignments that are of moderate diversity in scope. Exercises judgment within generally defined practices and... 15 days ago save job more...
Research Data Analyst University of California San Francisco ★★★★★ 418 reviews San Francisco, CA The Research Data Analyst requires significant independent work on the front end of the study design to ensure the right information and data collection tools... 1 day ago save job more...
Data Analyst Google ★★★★★ 3,194 reviews San Francisco, CA 94105 (Financial District area) 5 years of experience using a quantitative and qualitative approach to measure the effectiveness of programs, process improvements, and efforts for... 3 days ago save job more...
People Data Analyst Square ★★★★★ 83 reviews San Francisco, CA 94103 (South Of Market) Square's People Analytics team strives to make data accessible and useful so that guide better and more informed decisions. 1 day ago save job more...

Data Sources

All job data was scraped from [Indeed.com](#)





Data Scraping

- SiteMap and Web Scraper from “Indeed”



Data Cleaning

- Data Filtering
- Data Transformation
- Parse Data



Data Visualization

- Visualization using pandas, matplotlib and Seaborn



Analysis

- Interpretation and Comparison

Data Scraping

Tools and Techniques used:

- SiteMap JSON
 - Layout of the web-content on the page.
- Web Scraper
 - Google Chrome Development Tools extension.

```
{"_id": "jobs", "startUrl": ["https://www.indeed.com/jobs?q=Data+Analyst&l=California#"], "selectors": [{"id": "listings", "type": "SelectorElement", "parentSelectors": ["_root", "next"], "selector": "div.jobsearch-SerpJobCard:nth-of-type(n)", "multiple": true, "delay": 0}, {"id": "link", "type": "SelectorLink", "parentSelectors": ["listings"], "selector": "a.jobtitle", "multiple": false, "delay": 0}, {"id": "job_title", "type": "SelectorText", "parentSelectors": ["link"], "selector": "h3.icl-u-xs-mb--xs", "multiple": false, "regex": "", "delay": 0}, {"id": "company", "type": "SelectorText", "parentSelectors": ["link"], "selector": "div.icl-u-lg-mr--sm:nth-of-type(1)", "multiple": false, "regex": "", "delay": 0}, {"id": "rating", "type": "SelectorElementAttribute", "parentSelectors": ["link"], "selector": "a.icl-Ratings-starsCountWrapper", "multiple": false, "extractAttribute": "aria-label", "delay": 0}, {"id": "next", "type": "SelectorLink", "parentSelectors": ["_root", "next"], "selector": "div.pagination a:last-of-type", "multiple": false, "delay": 0}, {"id": "reviews", "type": "SelectorText", "parentSelectors": ["link"], "selector": "a.icl-Ratings-starsCountWrapper div.icl-Ratings-count", "multiple": false, "regex": "[^reviews]+", "delay": 0}]}]
```

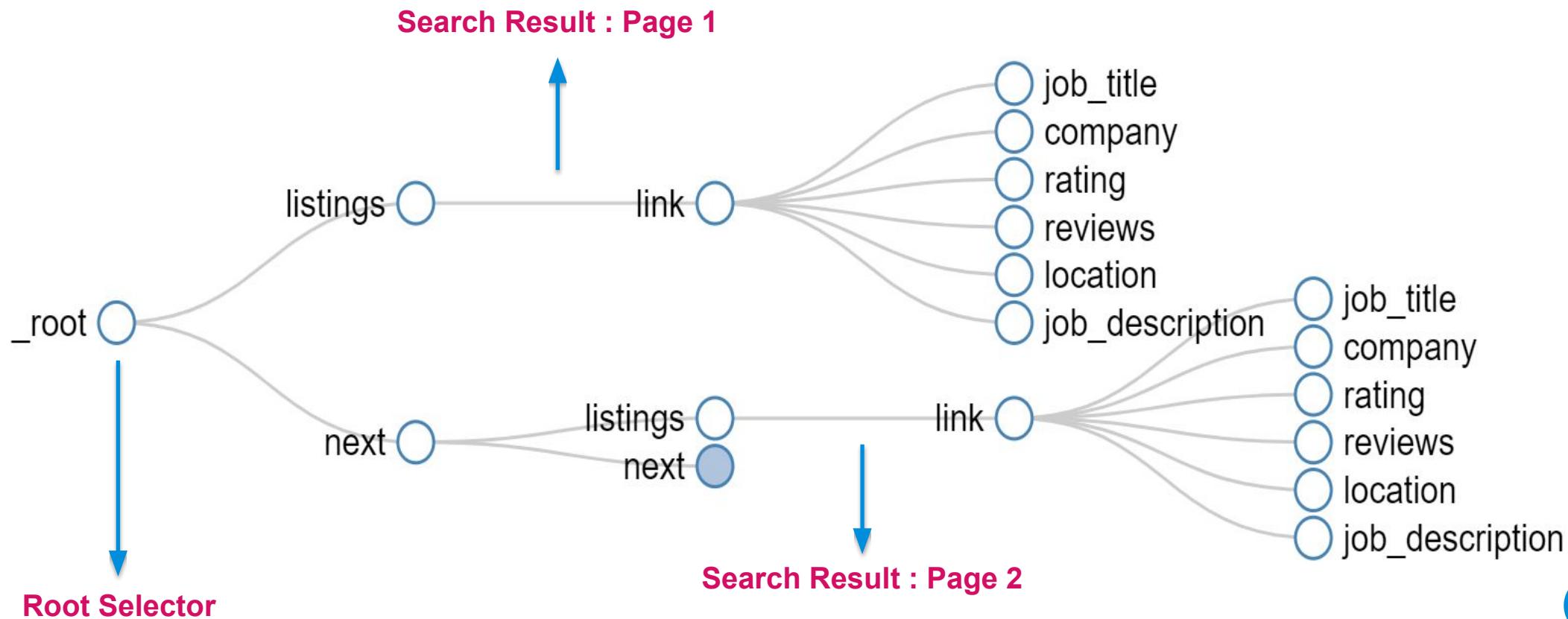
```
{"_id": "jobs", "startUrl": ["https://www.indeed.com/jobs?q=Data+Analyst&l=California#"],  
"selectors": [  
    {  
        "id": "listings",  
        "type": "SelectorElement",  
        "parentSelectors": ["_root", "next"],  
        "selector": "div.jobsearch-SerpJobCard:nth-of-type(n)",  
        "multiple": true, "delay": 0  
    }  
]
```

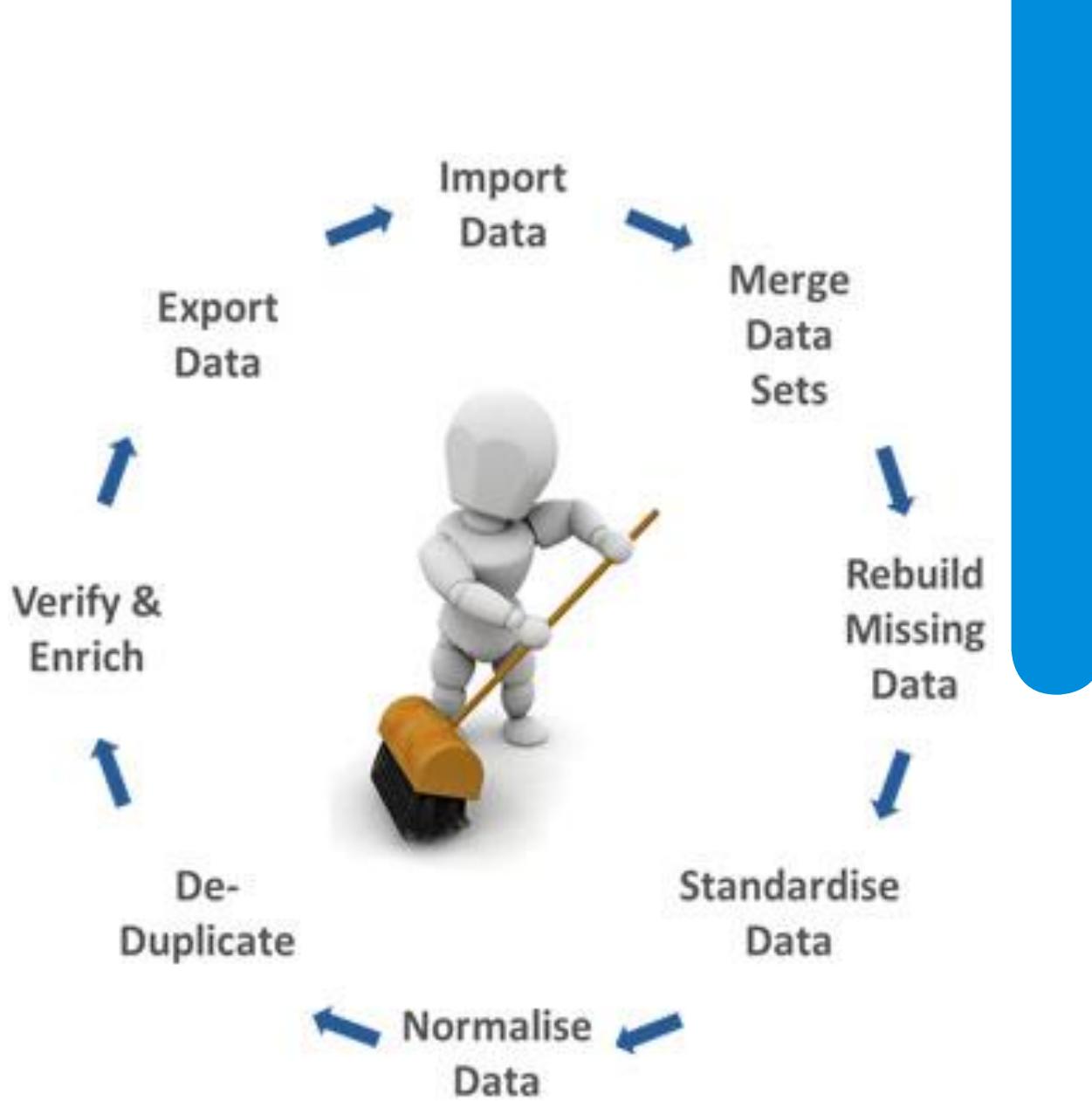


Multiple selections True/False

Matches selector with the nth element of the page.

Scraping Flow





Data Cleaning

Procedures:

- Treating blank cells (Fill blank cells with “0” or “NA”)
- Extra Spaces (Trim Function)
- Remove Duplicates
- Parsing Data using Text to Column
- Sorting Data
- Delete formatting

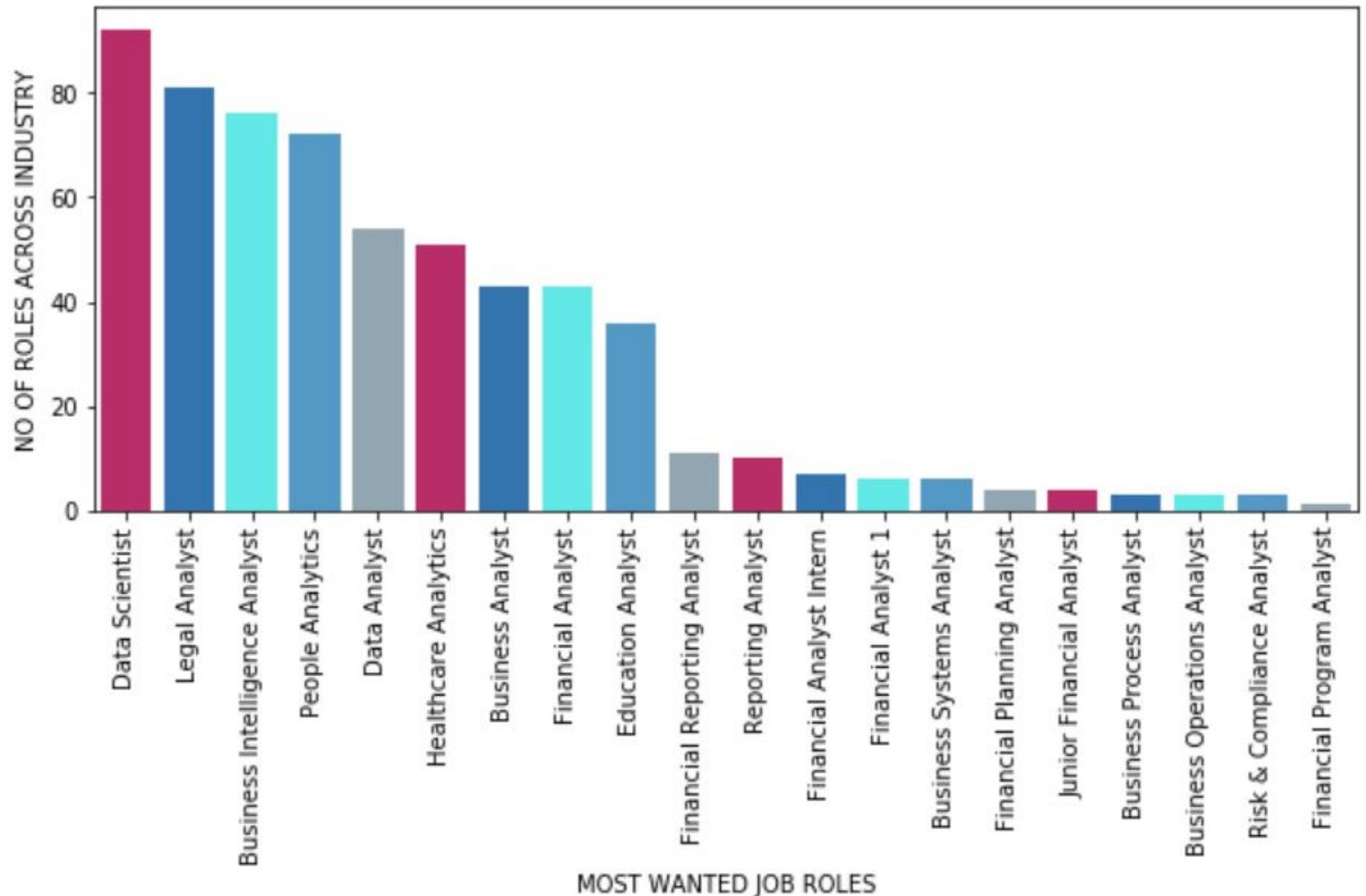
Data Set

Job Title	Company	Location_City	Location_Stat	
Business Analyst	San Francisco Unified School District	San Francisco	CA	The SFUSD Department of Technology seeks a talented Senior Business Analyst
Business Analyst	Havenly	Denver	CO	At Havenly, we believe that everyone deserves a beautiful home they love. Fou
Business Analyst	Oracle	Denver	CO	IT Strategy Planning Business Analyst Intern - Oracle Veteran Internship Program
Business Analyst	IBM	Southbury	CT	IntroductionThe Office of the Chief Information Officer (CIO) owns IBM's IT stra
Business Analyst	Federal Reserve Board of Governors	Washington	DC	The Quality Assurance Analyst intern will work in the Application Development
Business Analyst	Amtrak	Washington	DC	Your success is a train ride away.Amtrak connects businesses and communities
Business Analyst	NextEra Energy	Juno Beach	FL	Requisition ID: 35219Our reliability is one of the best in the nation, and we're v
Business Analyst	InnerWorkings	Chicago	IL	BUSINESS ANALYSTThe Business Analyst performs analysis to ensure internal cli
Business Analyst	Centeva	Chicago	IL	**On-site in Hines, IL, Non-Telecommuting Position**Job SummaryWe're lookin
Business Analyst	Hubbell Incorporated	Rolling Meadows	IL	Corporate OverviewHubbell Incorporated was founded in 1888 and has grown i
Business Analyst	HighTower Advisors	Chicago	IL	The Revenue System Business Analyst is responsible for collecting business req
Business Analyst	MedPro Group	Fort Wayne	IN	Meet MedPro Group. We're an industry-leading Berkshire Hathaway insurance
Business Analyst	NTT DATA Services	Frankfort	KY	Req ID: 63419At NTT DATA Services, we know that with the right people on boar
Business Analyst	Zumbiel Packaging	Hebron	KY	Business AnalystPosition Status: SalaryCompany Overview:Zumbiel Packaging is
Business Analyst	Cyrious Software	Baton Rouge	LA	Cyrious Software Business Analyst are technical experts who guide customers t
Business Analyst	Warner Bros Careers	Needham	MA	Company OverviewWarner Bros. has been entertaining audiences for more tha
Business Analyst	The TJX Companies, Inc.	Marlborough	MA	Success is always in style at TJX, a Fortune 100 company and the leading off-pr
Business Analyst	Cumberland Farms, Inc.	Westborough	MA	Position Summary:The Data Engineer / Business Analyst is responsible for design
Business Analyst	Cumberland Farms, Inc.	Westborough	MA	Position Summary:With a spirit of 'Never Settling', the IT Business Analyst (ITBA



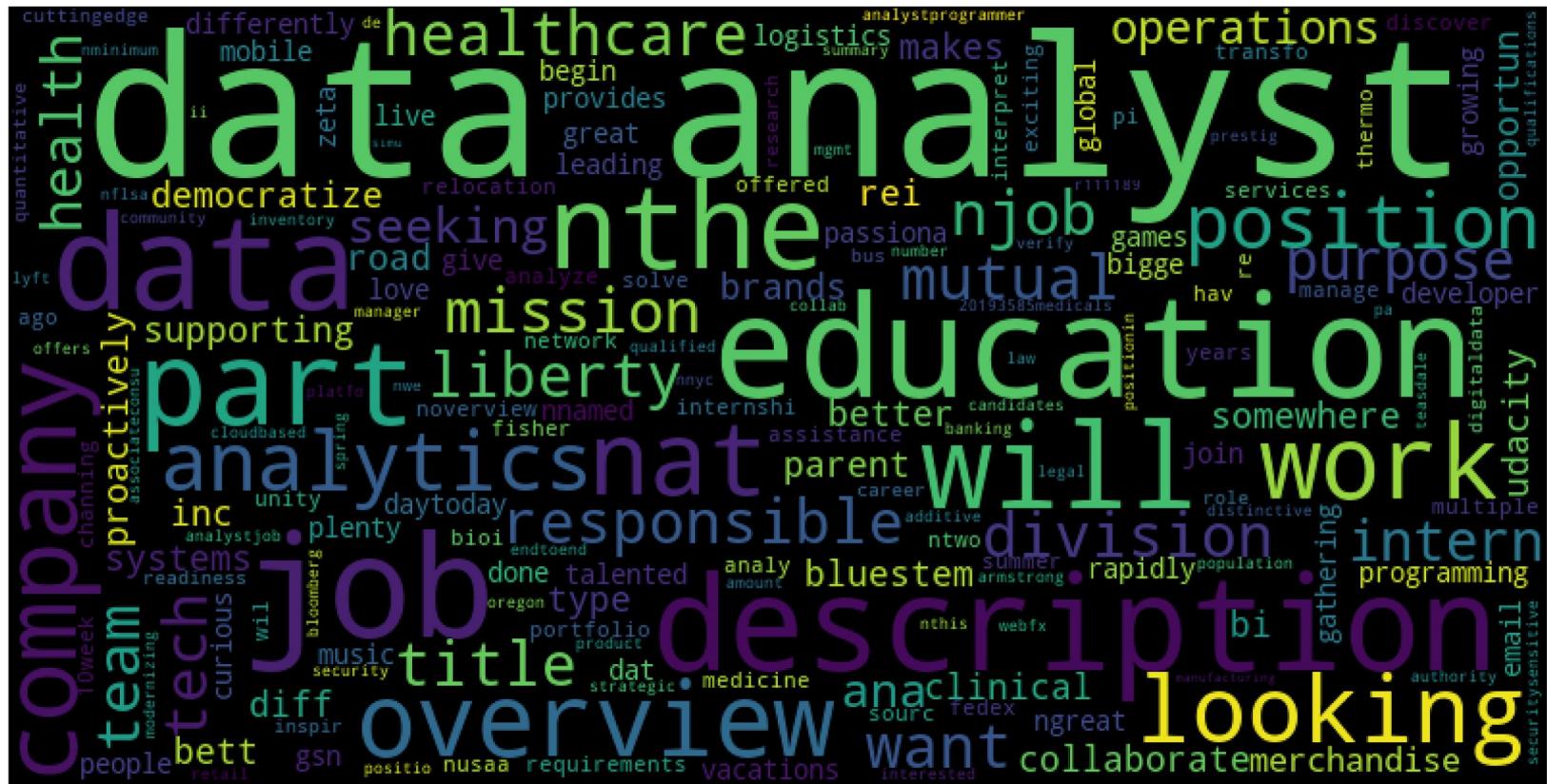
Title-wise Analysis

Job Roles across Industry

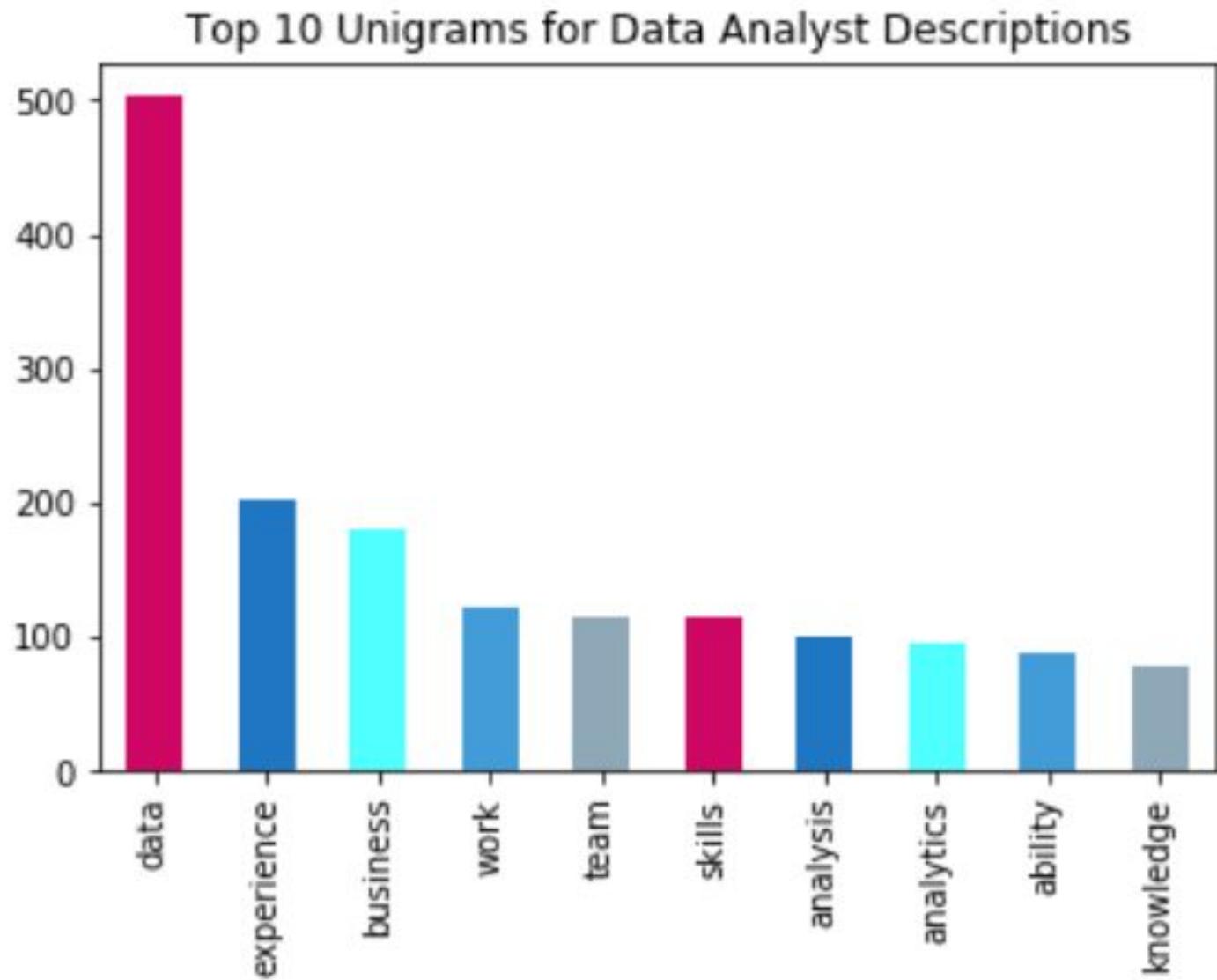


Word Cloud – Data Analyst

Word Cloud of Data Analyst Descriptions

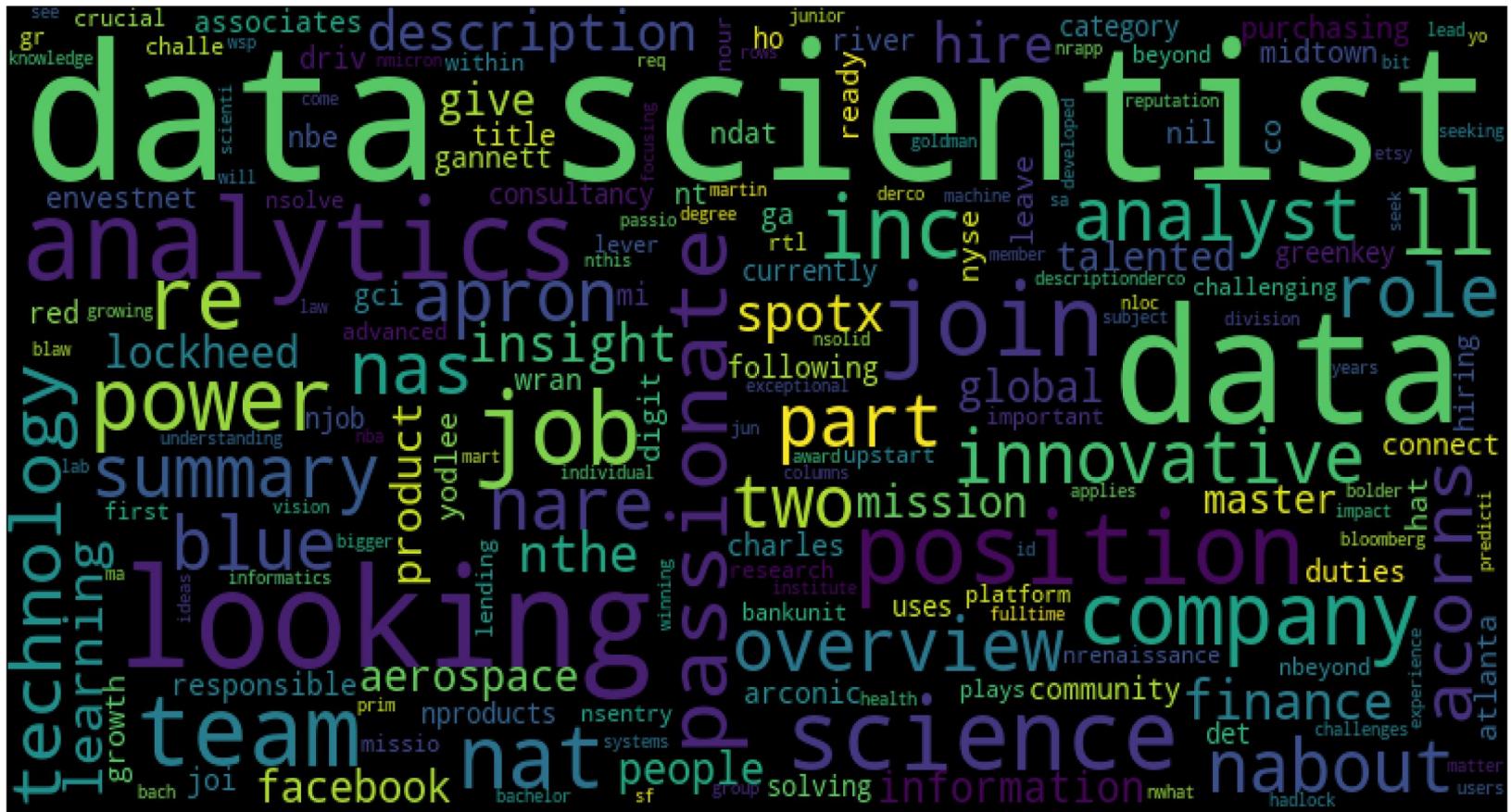


Word Ingram – Data Analyst

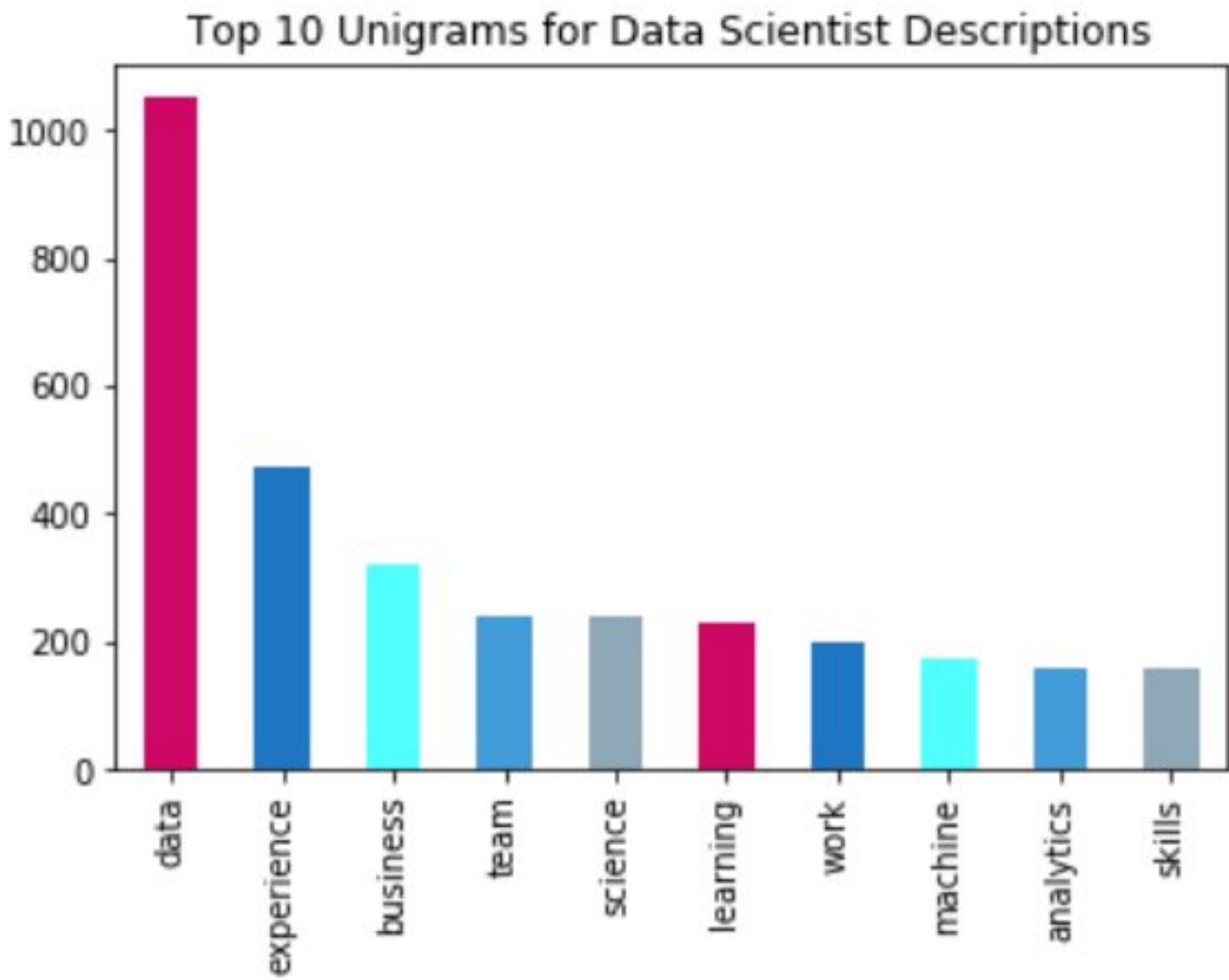


Word Cloud – Data Scientist

Word Cloud of Data Scientist Descriptions



Word Ingram – Data Scientist

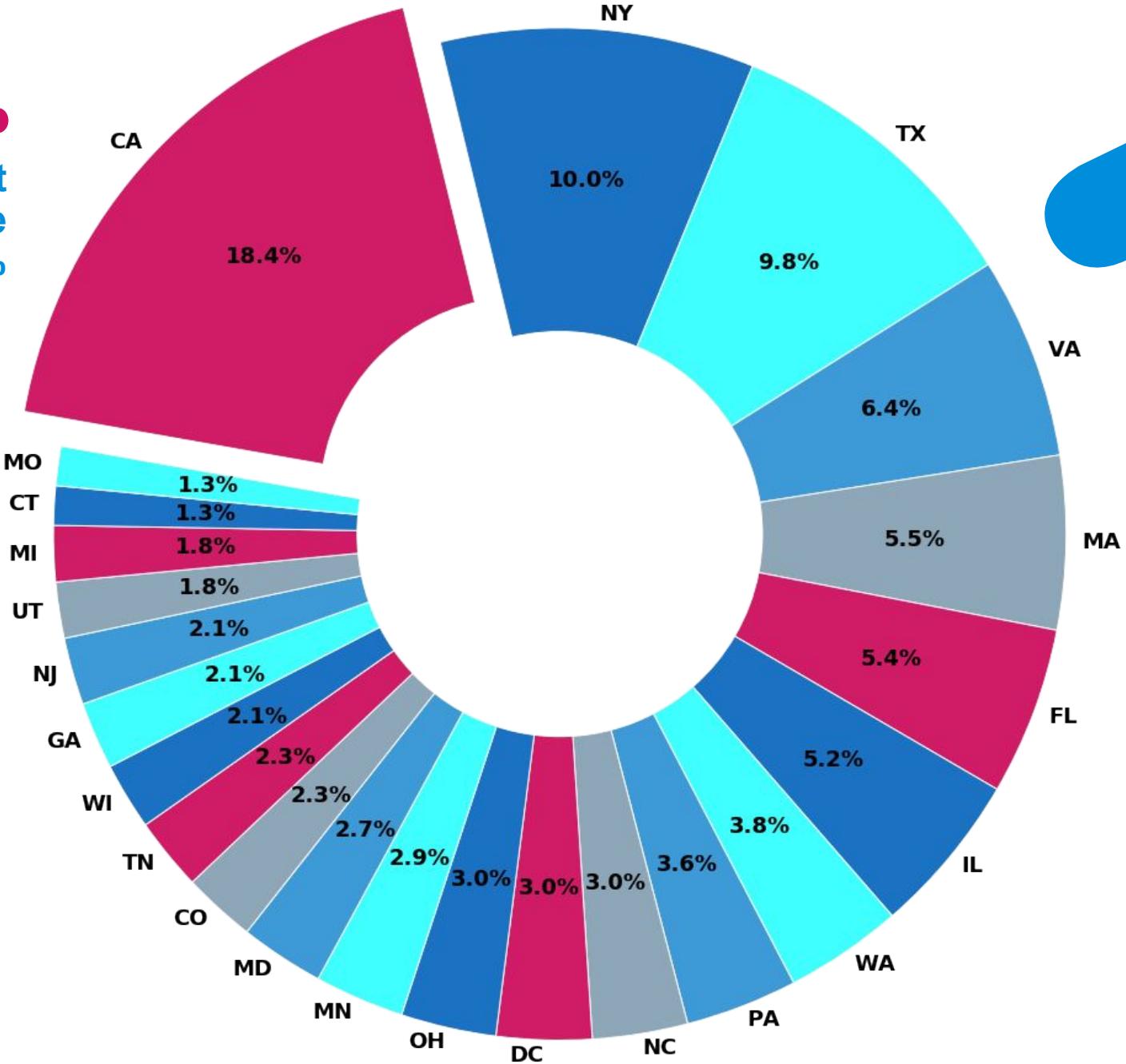




Location-wise Analysis

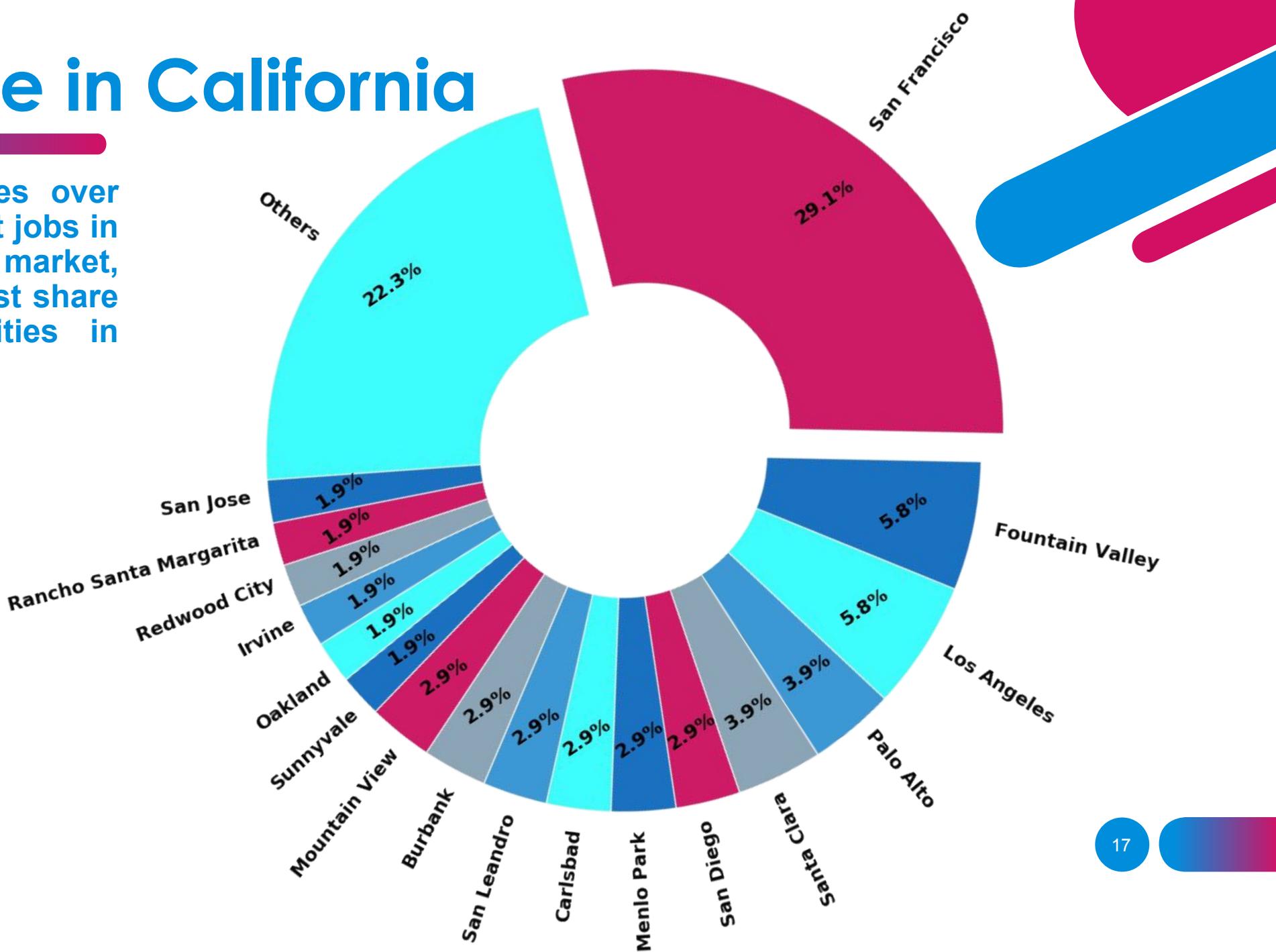
State Share

California has the largest share among the all the states, taking over 18.4% of the analyst job market.



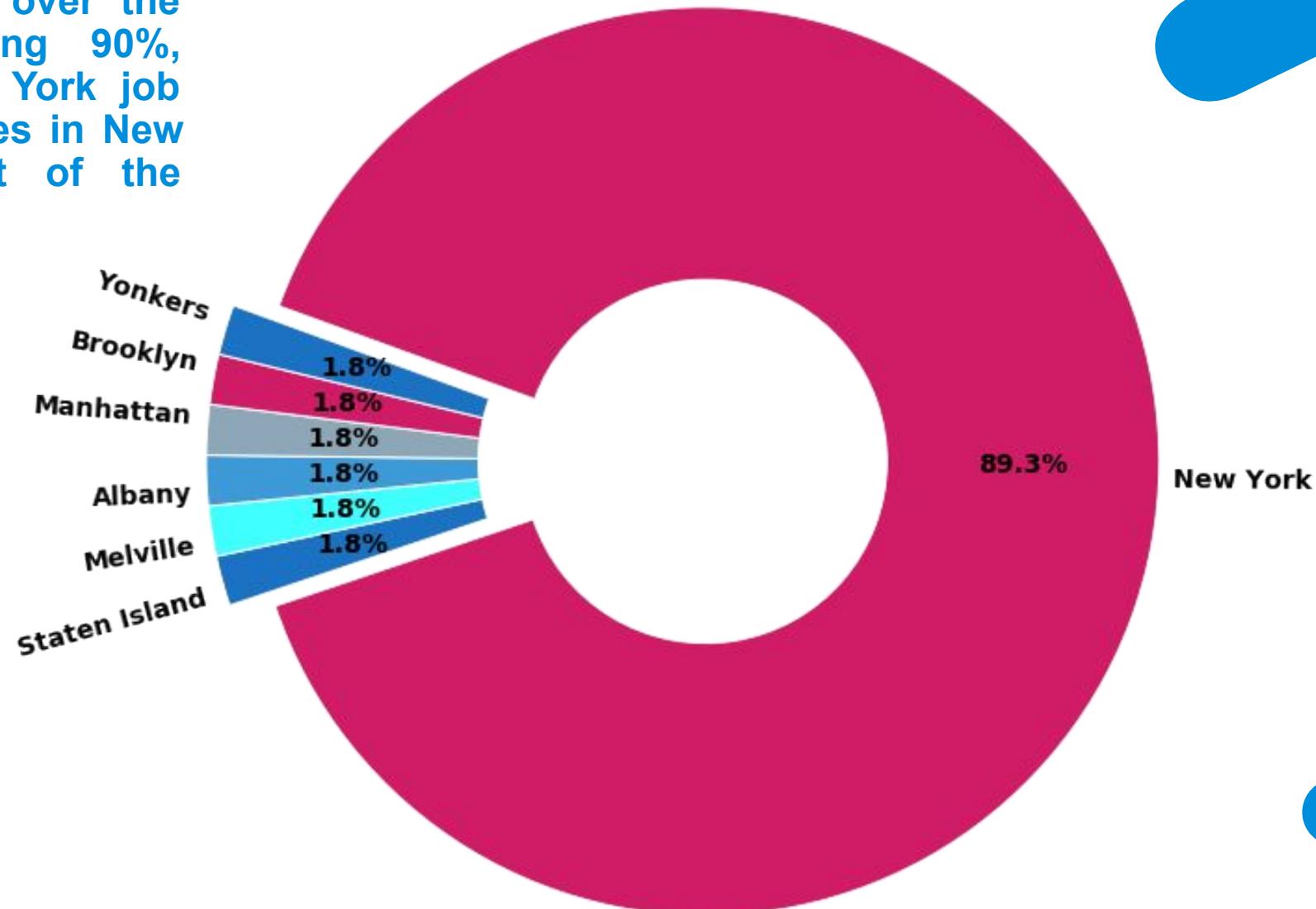
City Share in California

San Francisco takes over 29.1% of the analyst jobs in California job market, becoming the largest share among all the cities in California



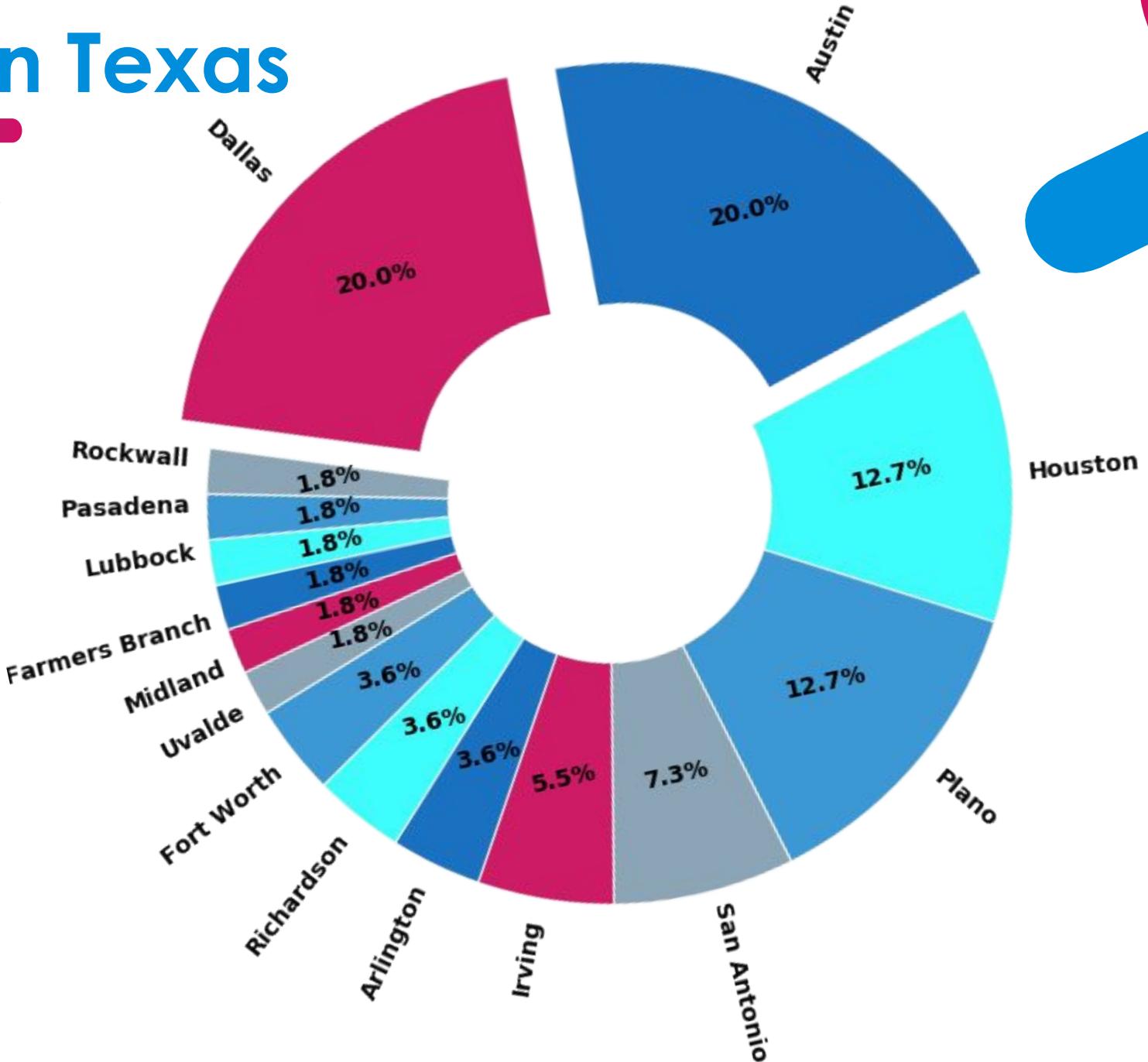
City Share in New York

New York City takes over the majority, almost being 90%, analyst roles in New York job market, and other cities in New York share the rest of the analyst jobs equally.



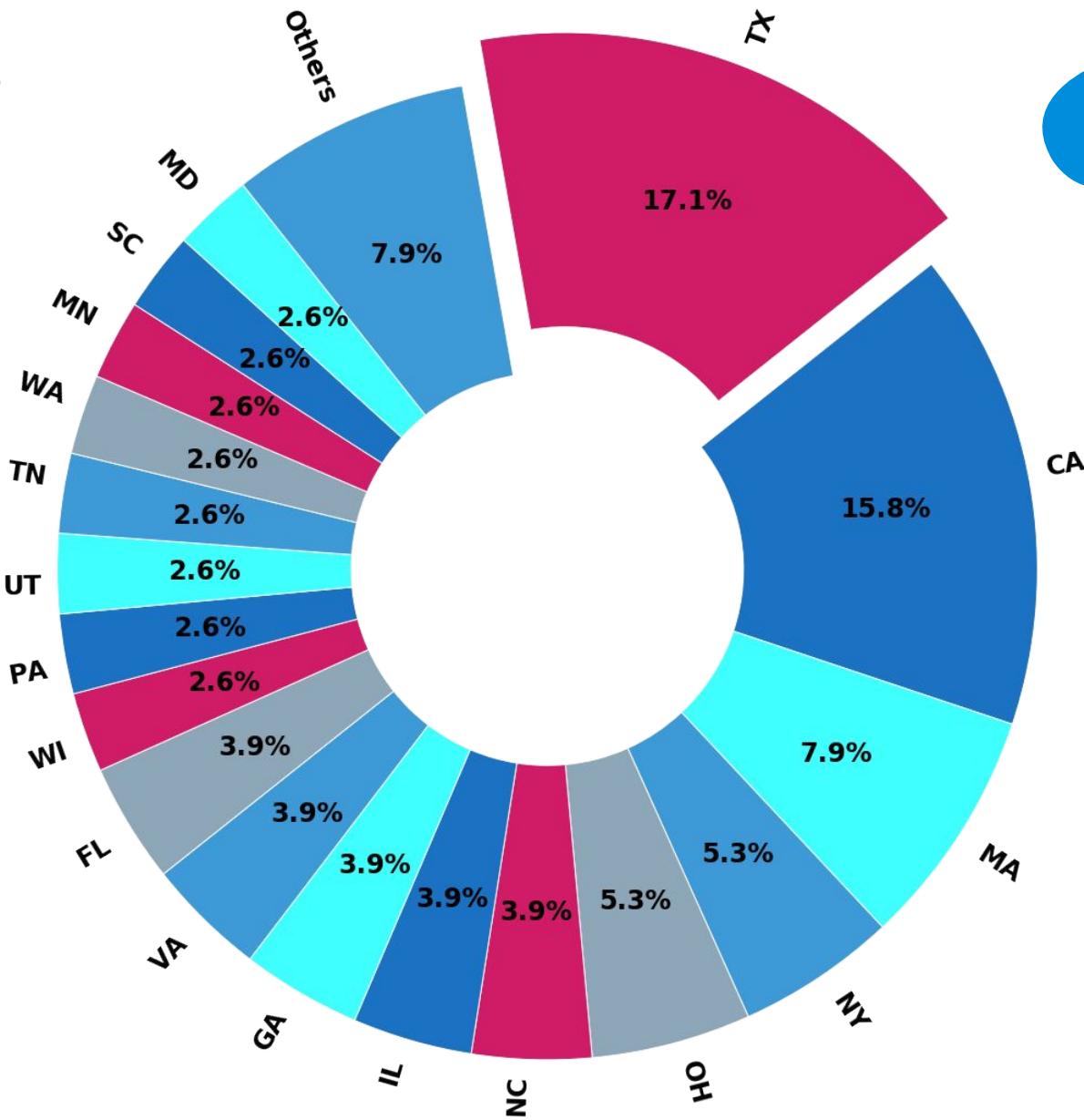
City Share in Texas

Each of Dallas and Austin takes over 20% of the analyst jobs in Texas job market, and each of Houston and Plano take over 12.7% of the analyst roles in Texas job market. These four cities take over most of the analyst jobs in Texas.



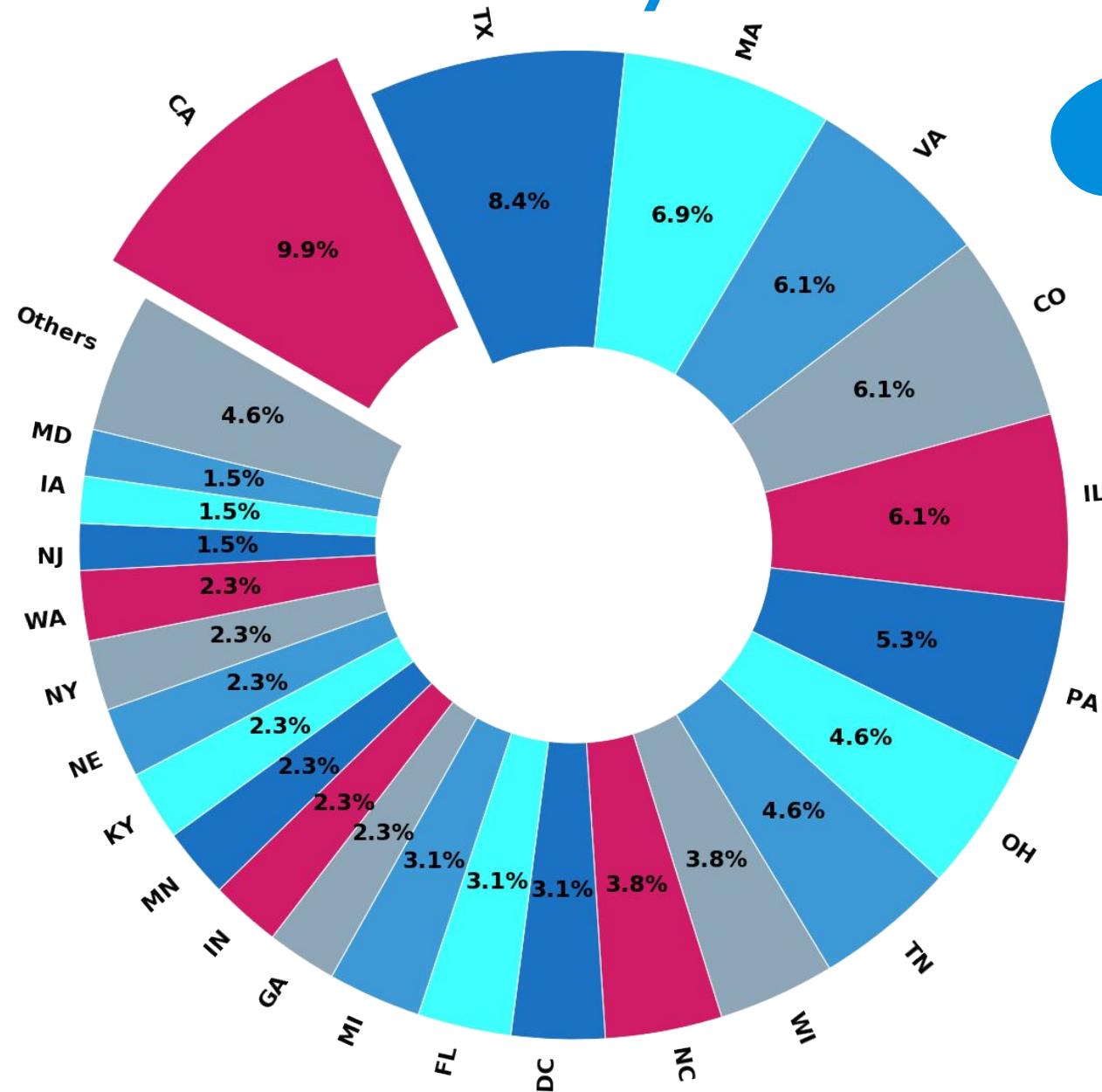
Which State Needs Financial Analysts The Most?

Texas needs Financial analyst the most, secondly it's California.



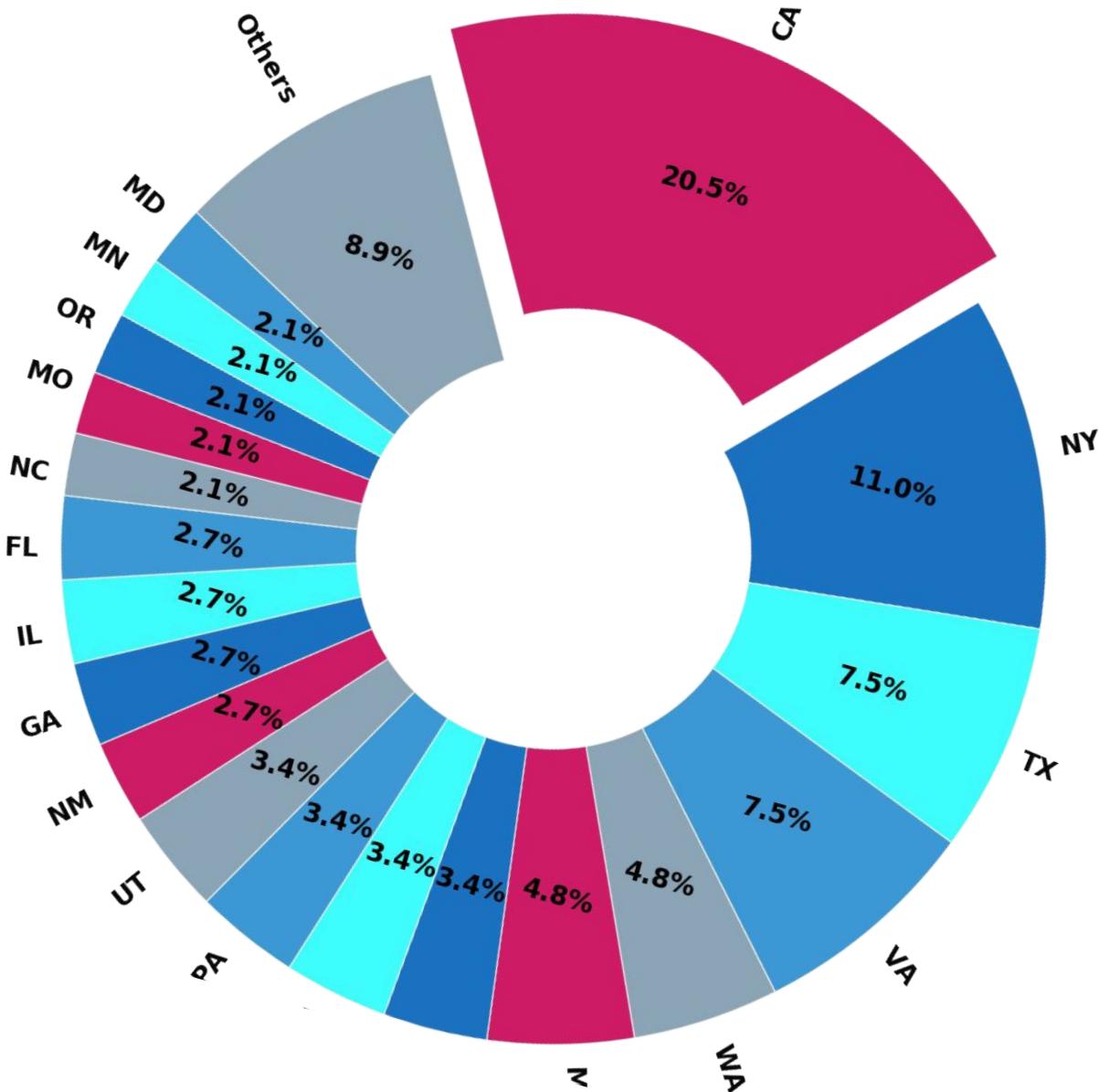
Which State Needs Business Analysts The Most?

California needs Business Analysts the most, taking over 10% Business analyst roles, secondly it's Texas, with a share of 8.4%. And Massachusetts, Virginia, Colorado and Illinois take over almost the same share, between 6% and 7%, in the Business analyst job market.



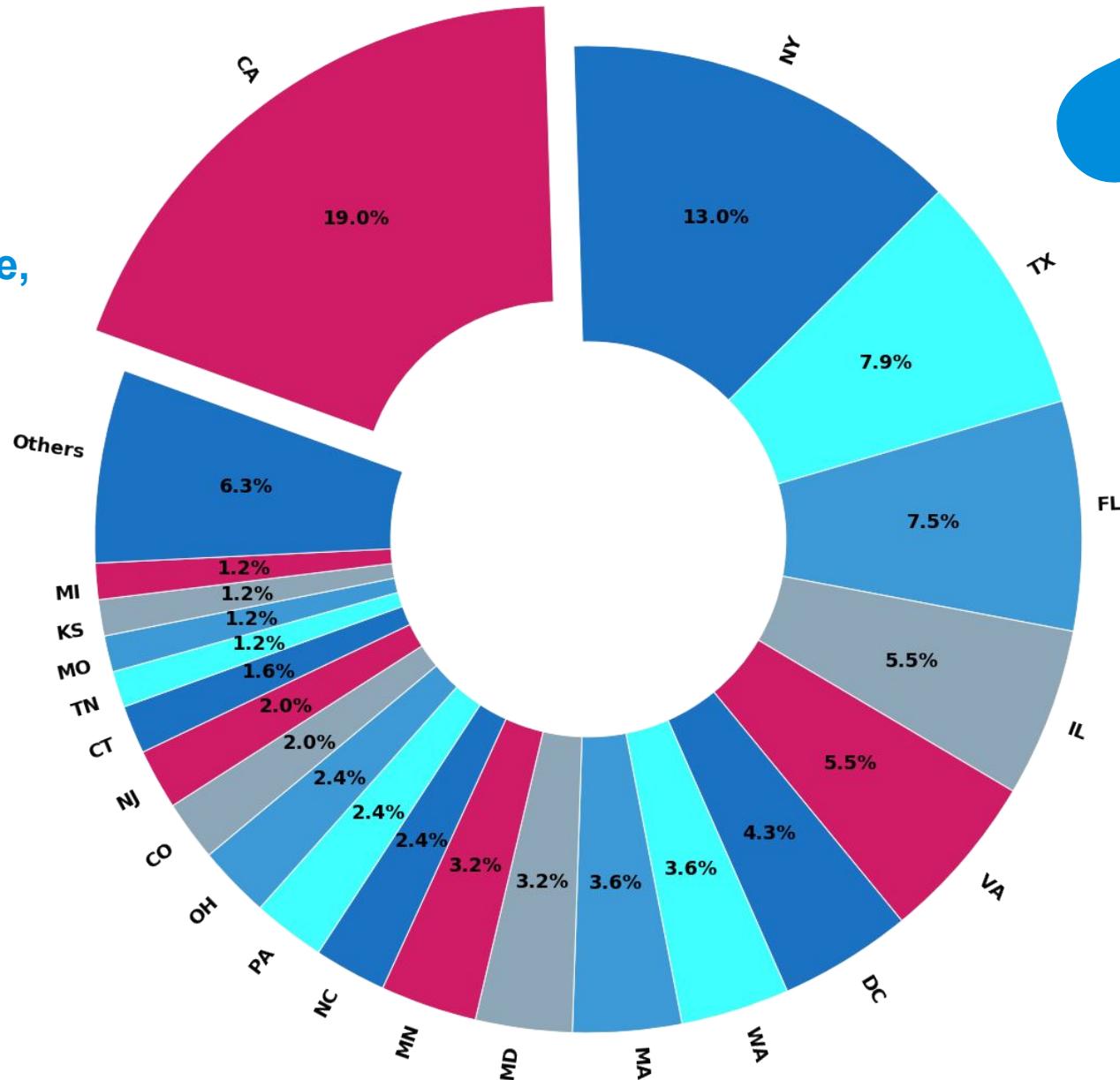
Which State Needs Data Analysts The Most?

California needs data analysts and data scientists the most, taking over about 20% market share, which makes sense because Silicon Valley is a core of technology industry.



How Does Analyst Roles in Other Domains Distribute?

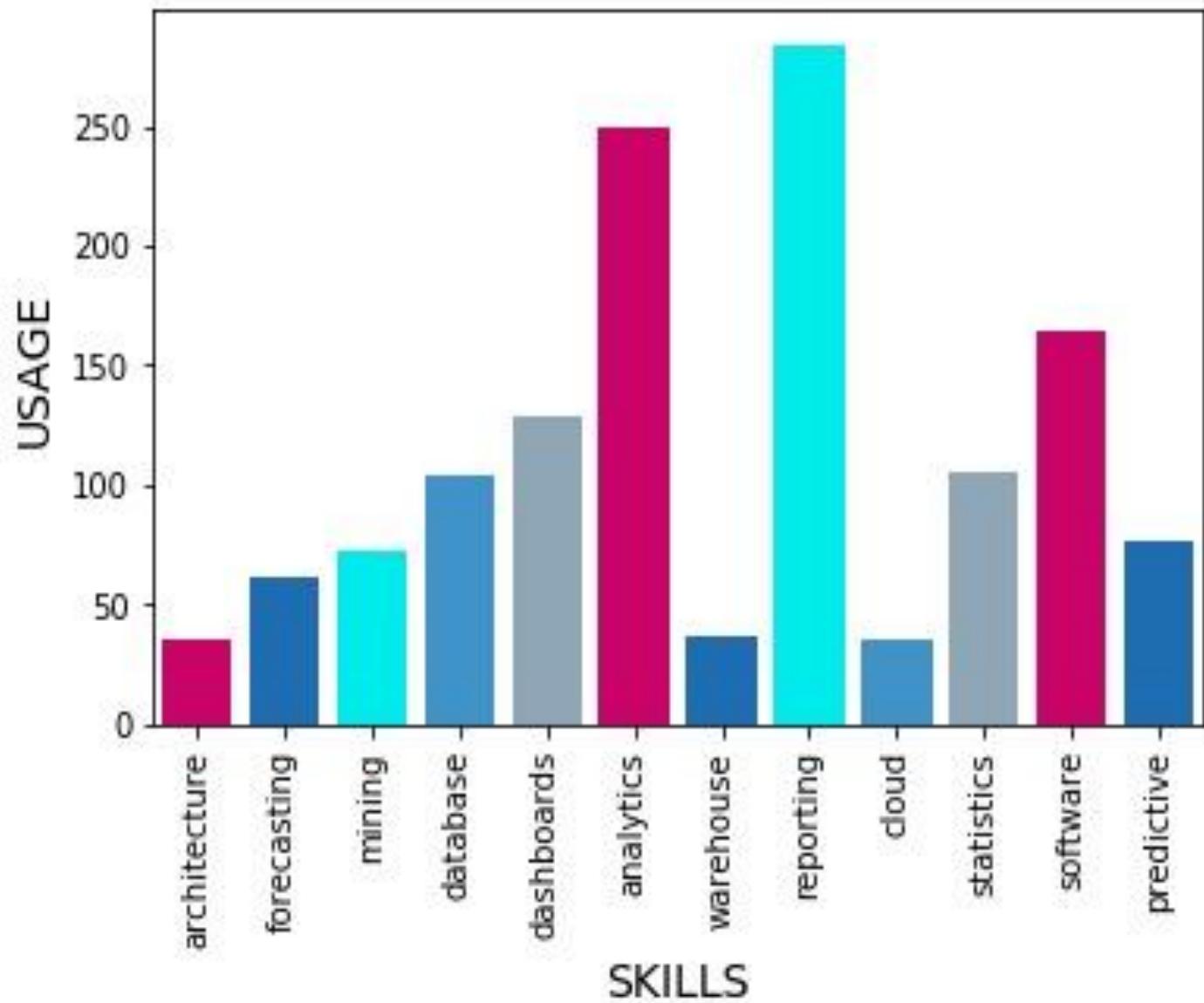
This chart shows the location distribution of analyst roles in health, human resource, legal and other domains. California, apparently, has the largest share, 19%.



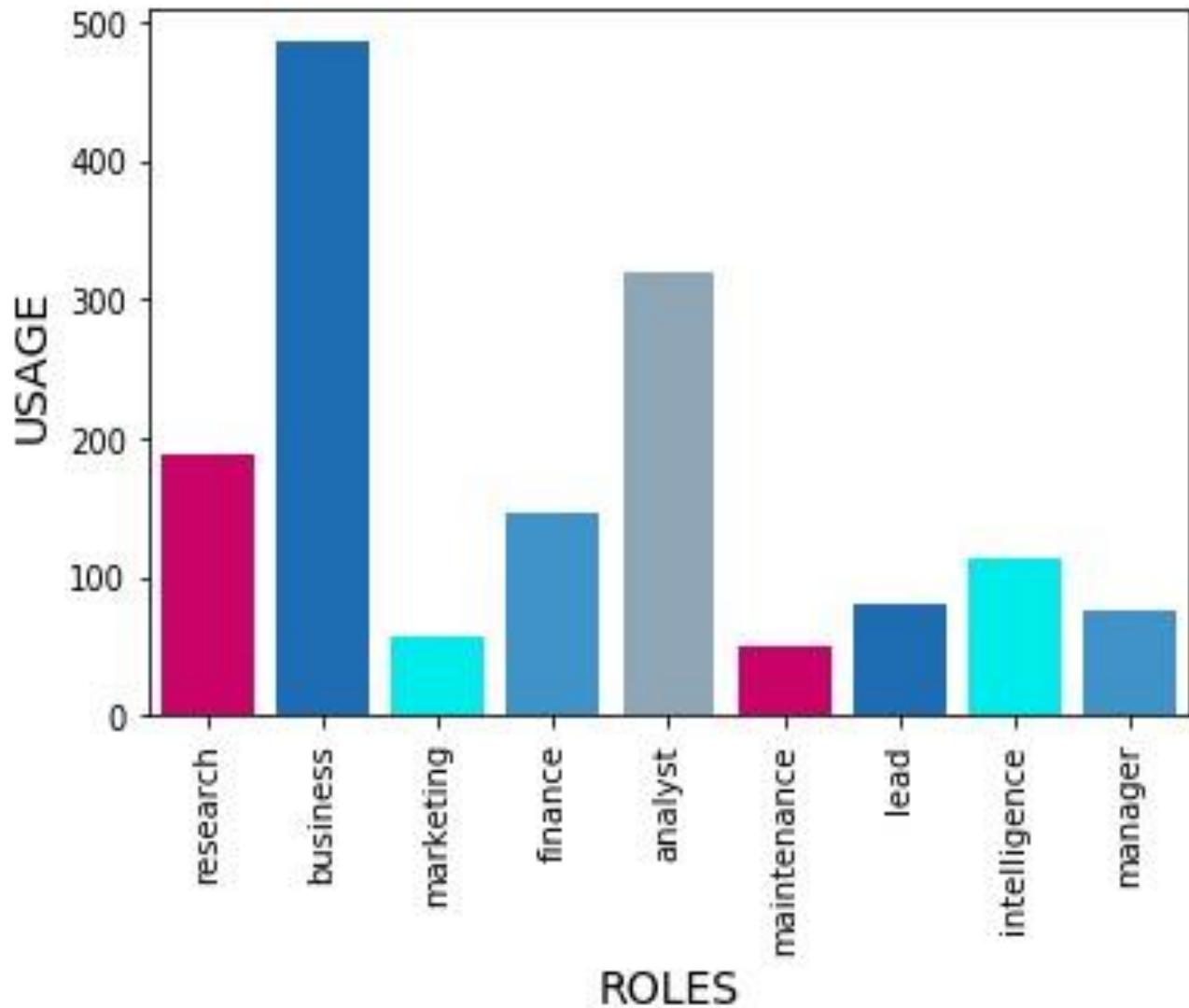
Description-wise Analysis

```
    if _operation == "MIRROR_X":  
        mirror_mod.use_x = True  
        mirror_mod.use_y = False  
        mirror_mod.use_z = False  
  
    elif _operation == "MIRROR_Y":  
        mirror_mod.use_x = False  
        mirror_mod.use_y = True  
        mirror_mod.use_z = False  
  
    elif _operation == "MIRROR_Z":  
        mirror_mod.use_x = False  
        mirror_mod.use_y = False  
        mirror_mod.use_z = True  
  
#selection at the end -add back the deselected mirror modifier object  
modifier_ob.select= 1  
mirror_ob.select=1  
 bpy.context.scene.objects.active = modifier_ob  
print("Selected" + str(modifier_ob)) # modifier ob is the active ob  
 mirror_ob.select = 0  
bone = bpy.context.selected_objects[0]  
bone.data.mirror_type = 'CLIPPING'  
if bone.select: bone.select = False  
else: bone.select = True  
#print(bone.select) verify the state of bone select
```

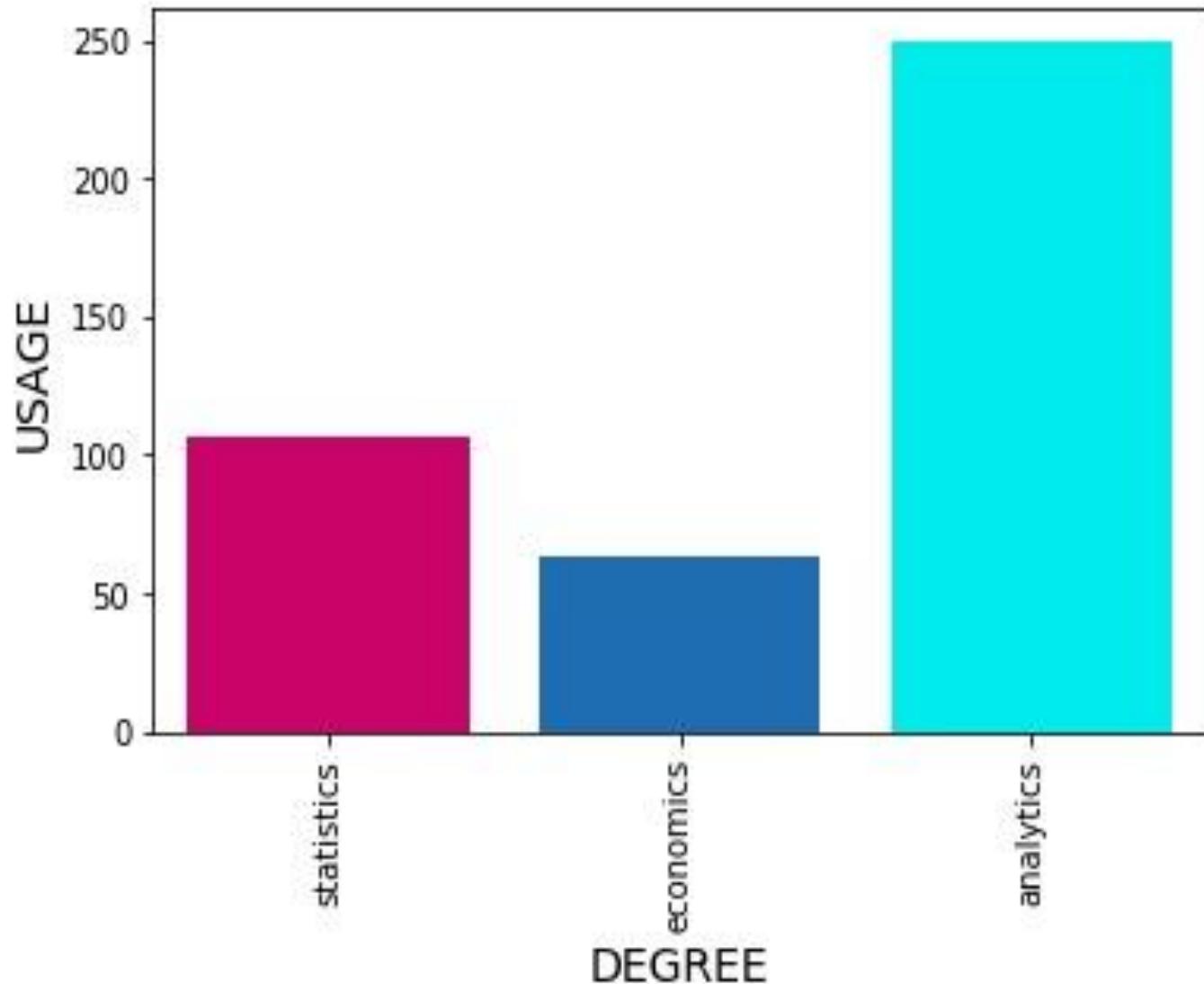
Skills (Using Keyword Filter)



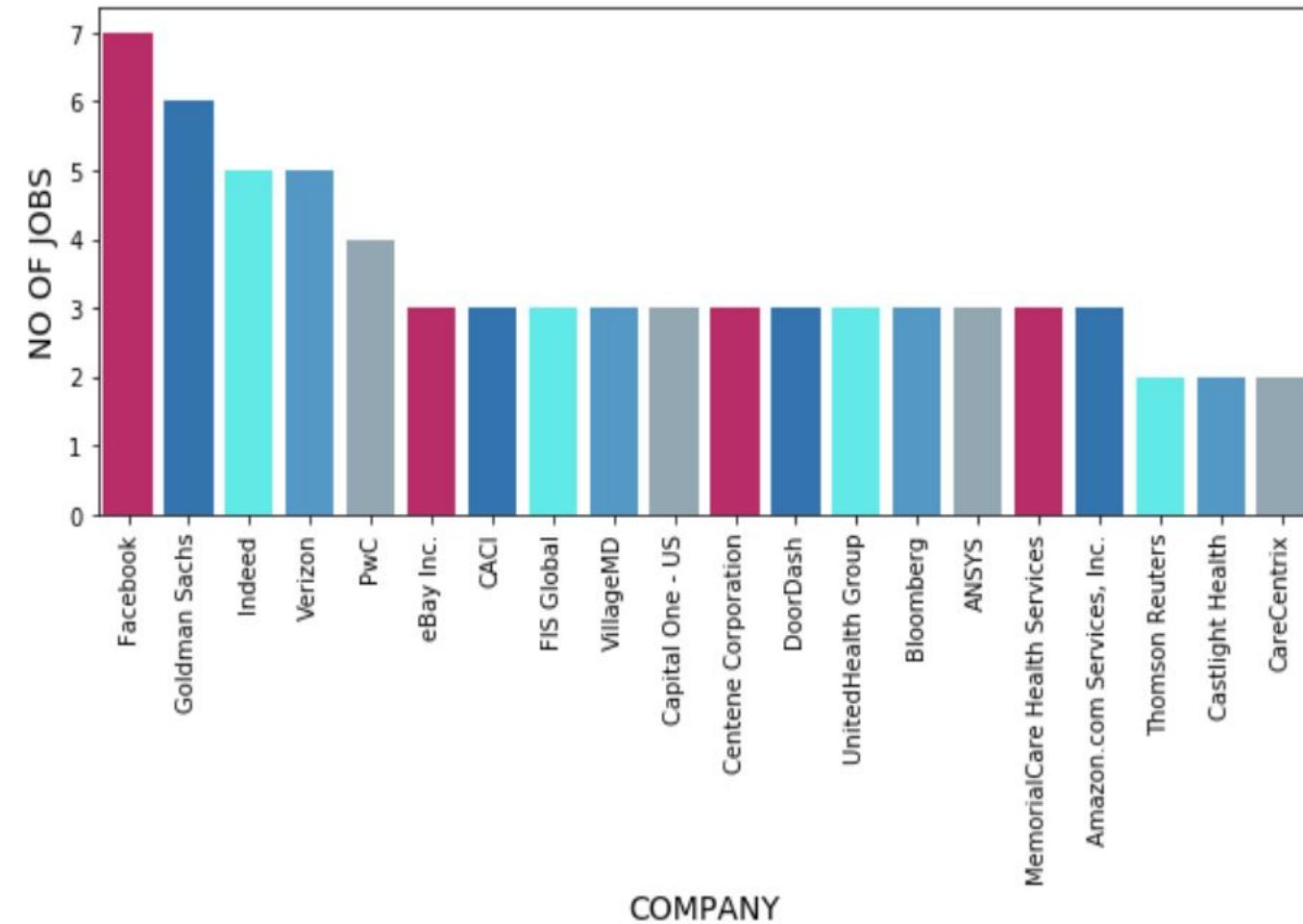
Roles (Using Keyword Filter)



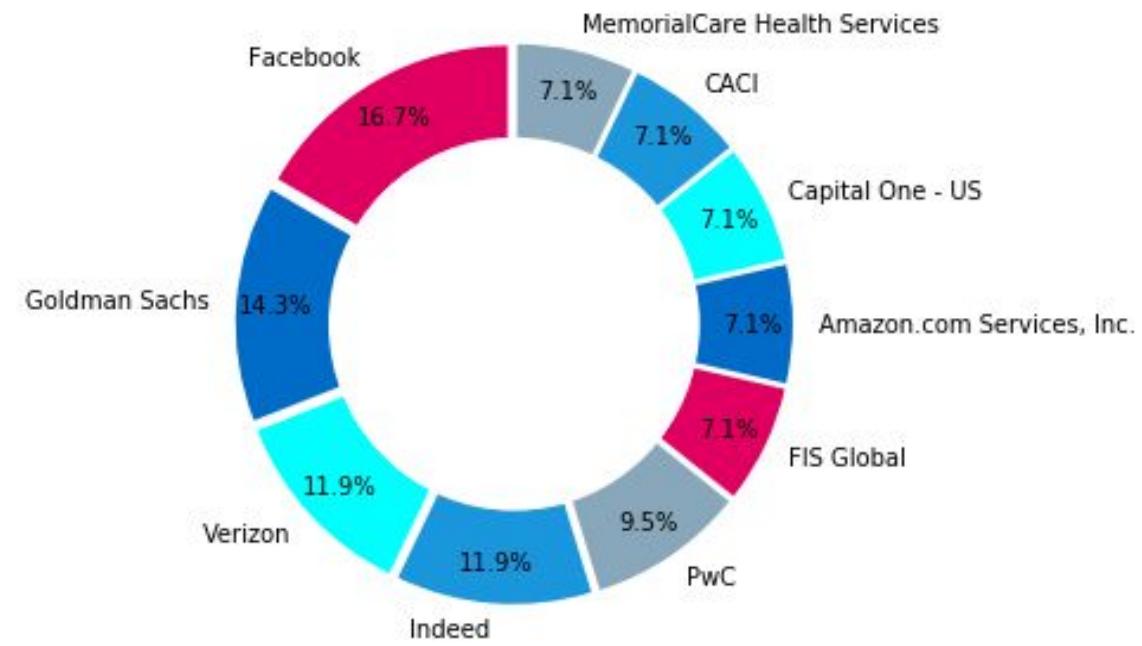
Degree (Using Keyword Filter)



Number of Jobs - (Company Wise)



Here is the pie-chart of top 10 companies:



Add-on: Big-4 Accounting Firms Analyst

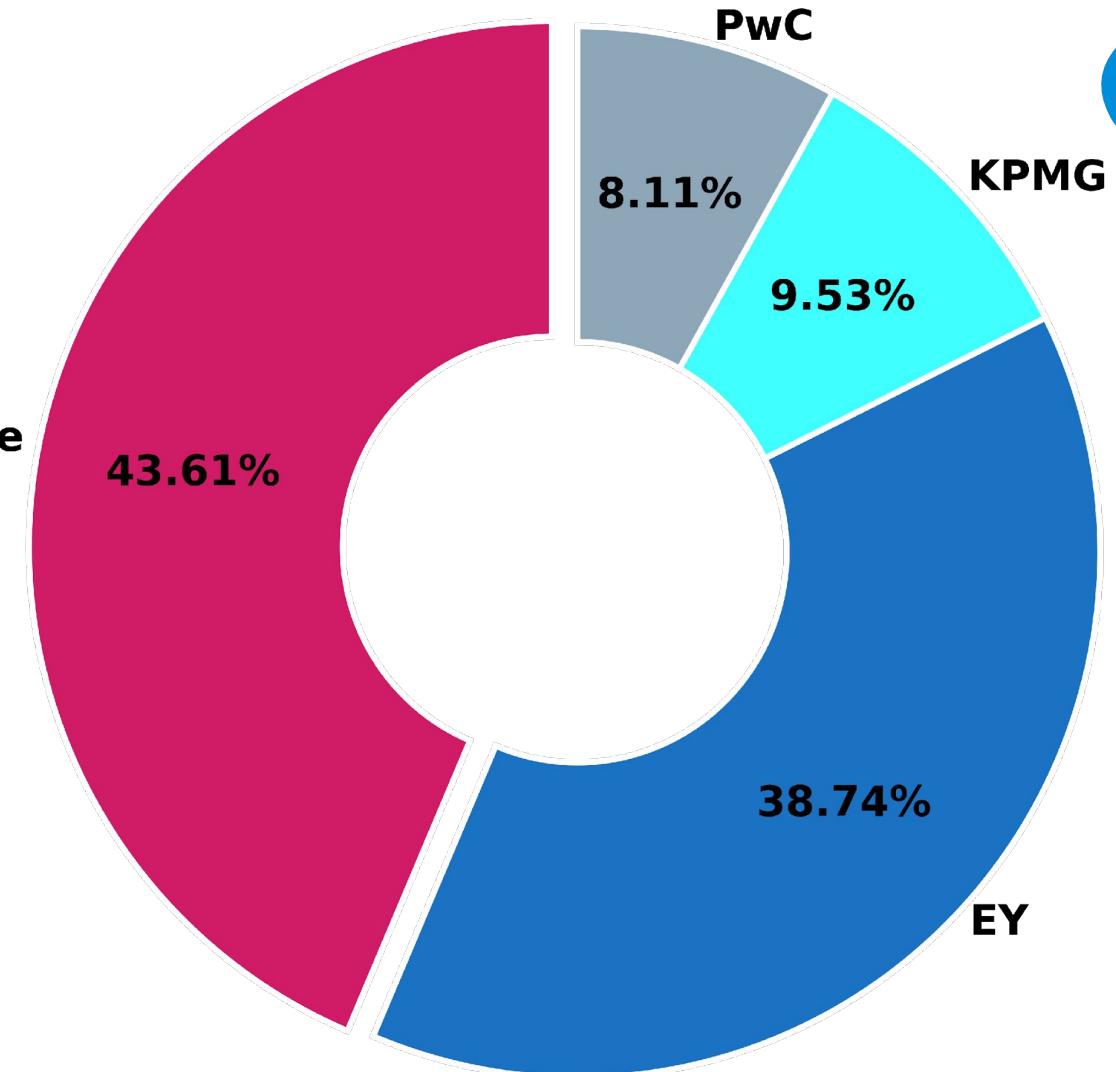
In this part, we will be analysing 4 datasets scraped from Indeed based on the Big 4 accounting companies: Deloitte Touche Tohmatsu Limited ([Deloitte](#)), Ernst & Young ([EY](#)), KPMG International Cooperative ([KPMG](#)), and PricewaterhouseCoopers ([PwC](#)).

Company Share

Deloitte has the largest share among the Big-4 Account Firms, taking over 40% of the job market of analysts.

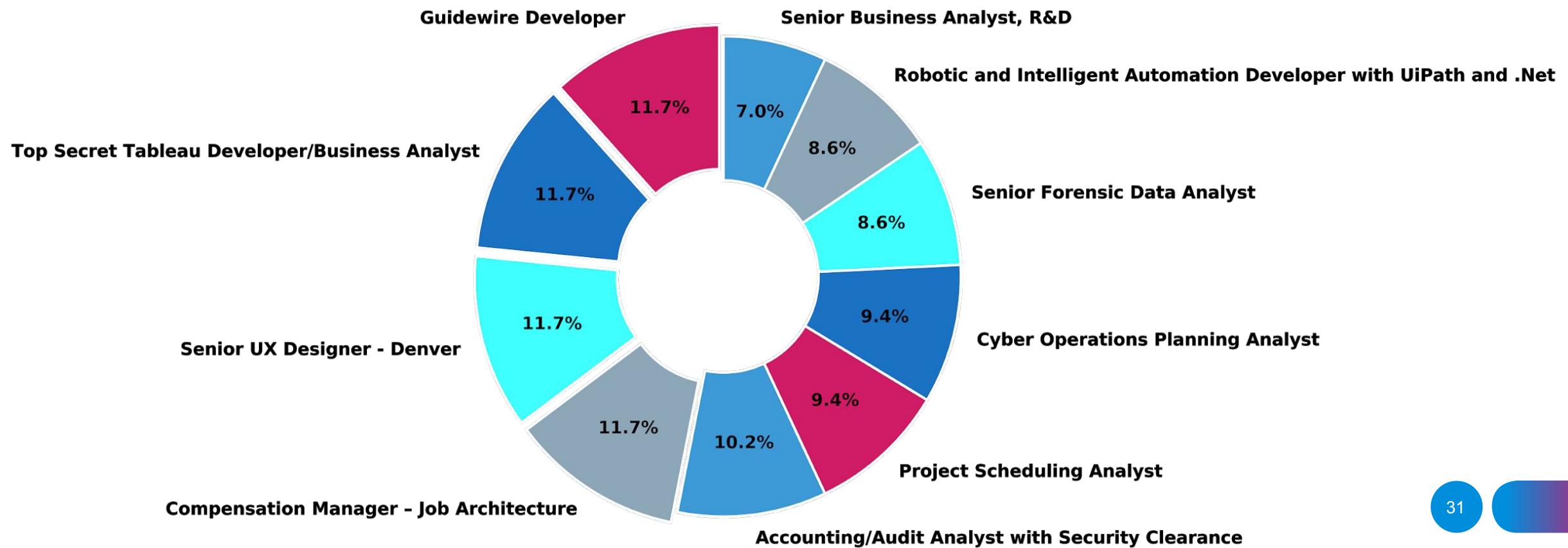


Deloitte.



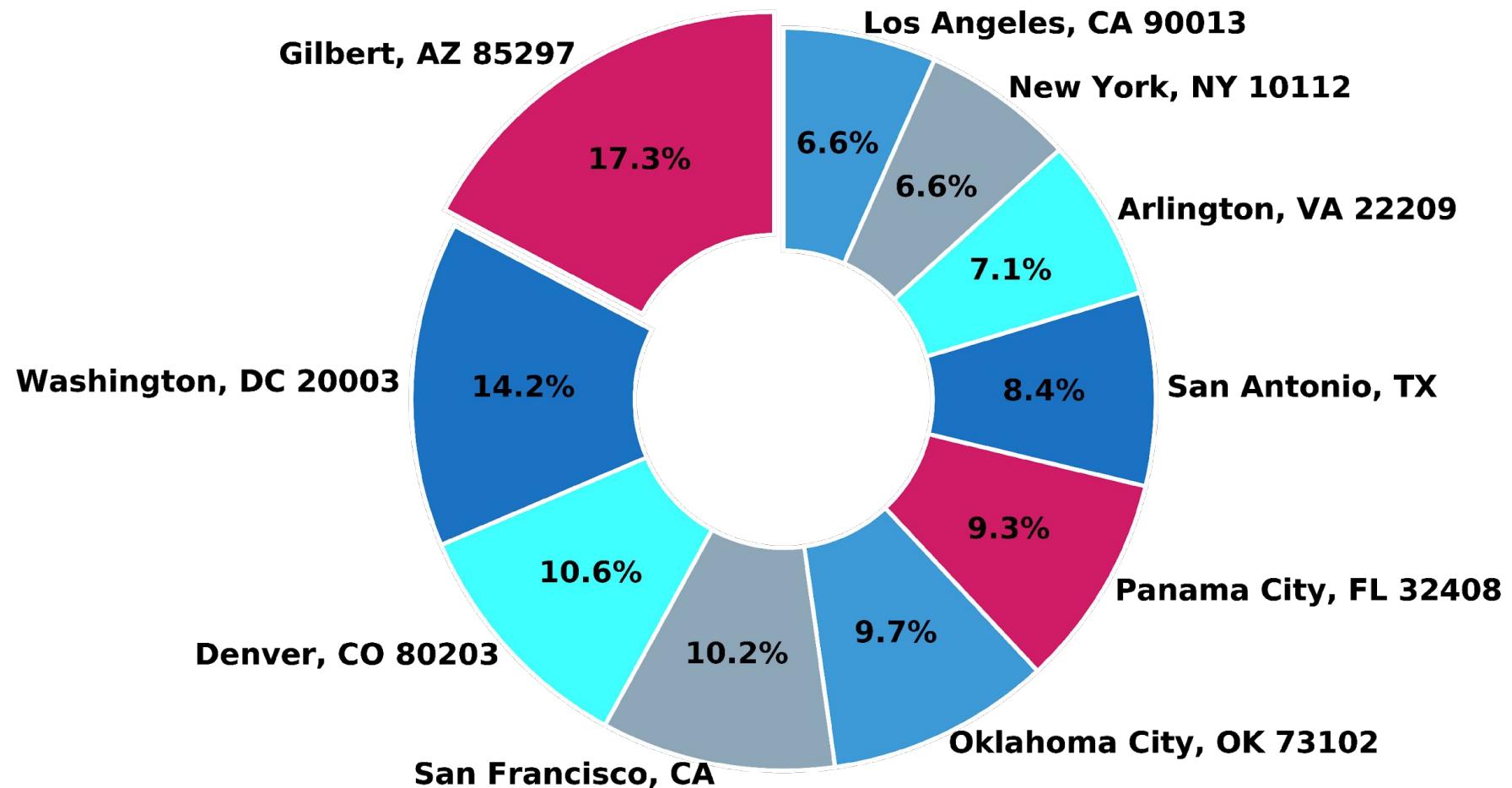
Most Listed Job Titles

Guidewire Developer, Tableau Developer/Business Analyst, Senior UX Designer, and Compensation Manager are the 4 most hot-handed job in the Big-4 accounting market.



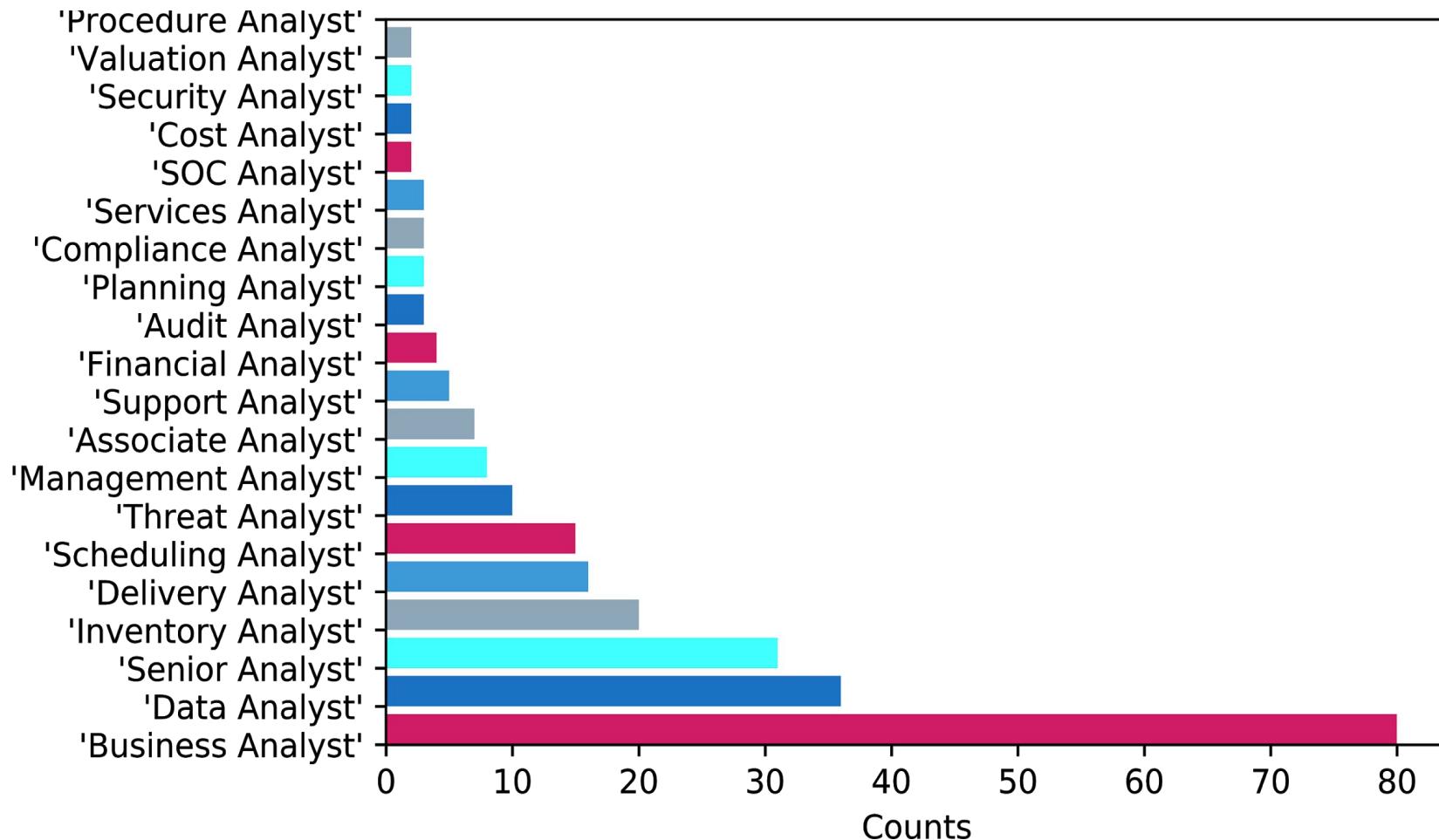
Most Frequent Locations

Gilbert, AZ has the most job opportunities, followed by Washington, DC. There are only 2 cities in CA in the top-10 tire, which is San Francisco, ranking at the 4th, and Los Angeles, raking at the 10th.



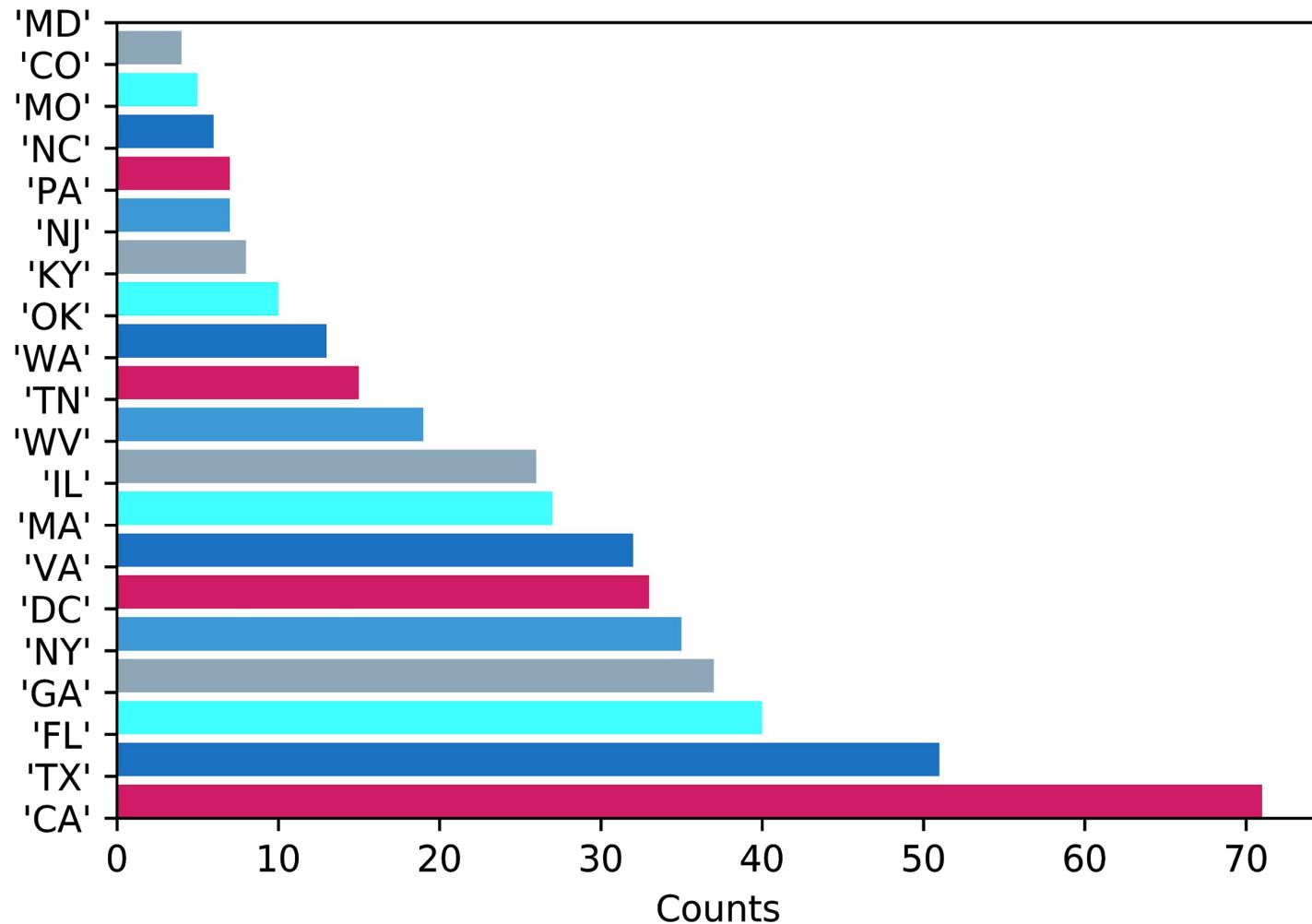
Most Wanted Analyst Jobs

Using the Re method, the word pairs of 'XXX Analyst' were filtered out. The below chart shows that the highest 3 occurrence jobs were Business Analyst, Data Analyst, and Senior Analyst.



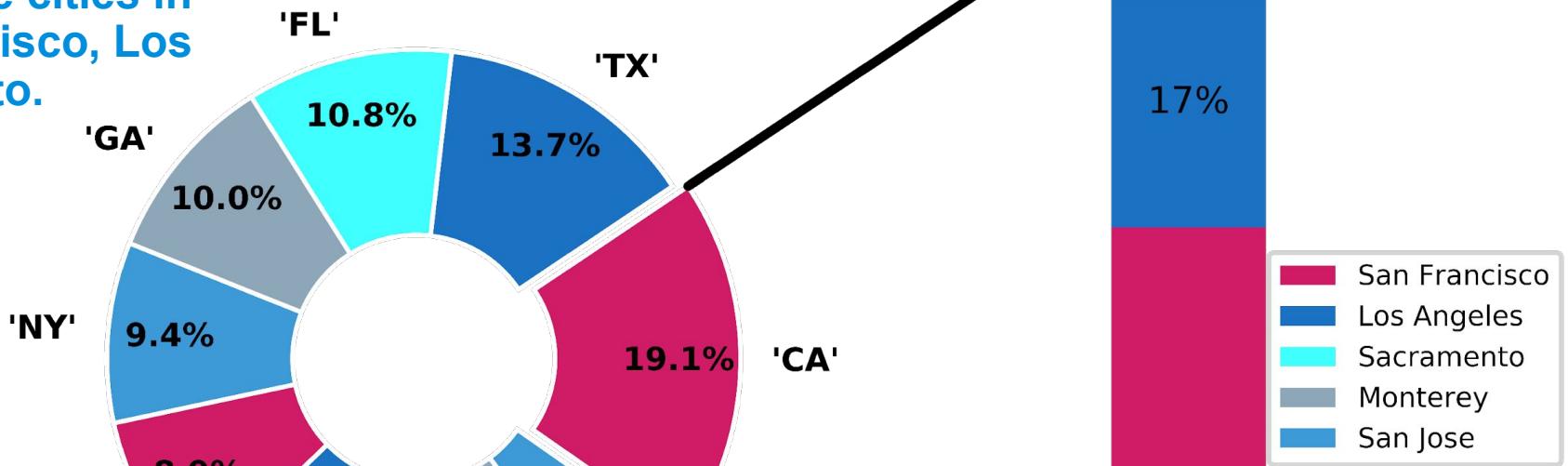
Most Common States

Using the Re method, the states were filtered out. The below chart shows that the highest 3 occurrence states are California, Texas, and Florida.



Most Common Cities in CA

Using the Re method, the cities were filtered out. The below chart shows that the highest 3 occurrence cities in California are San Francisco, Los Angeles, and Sacramento.



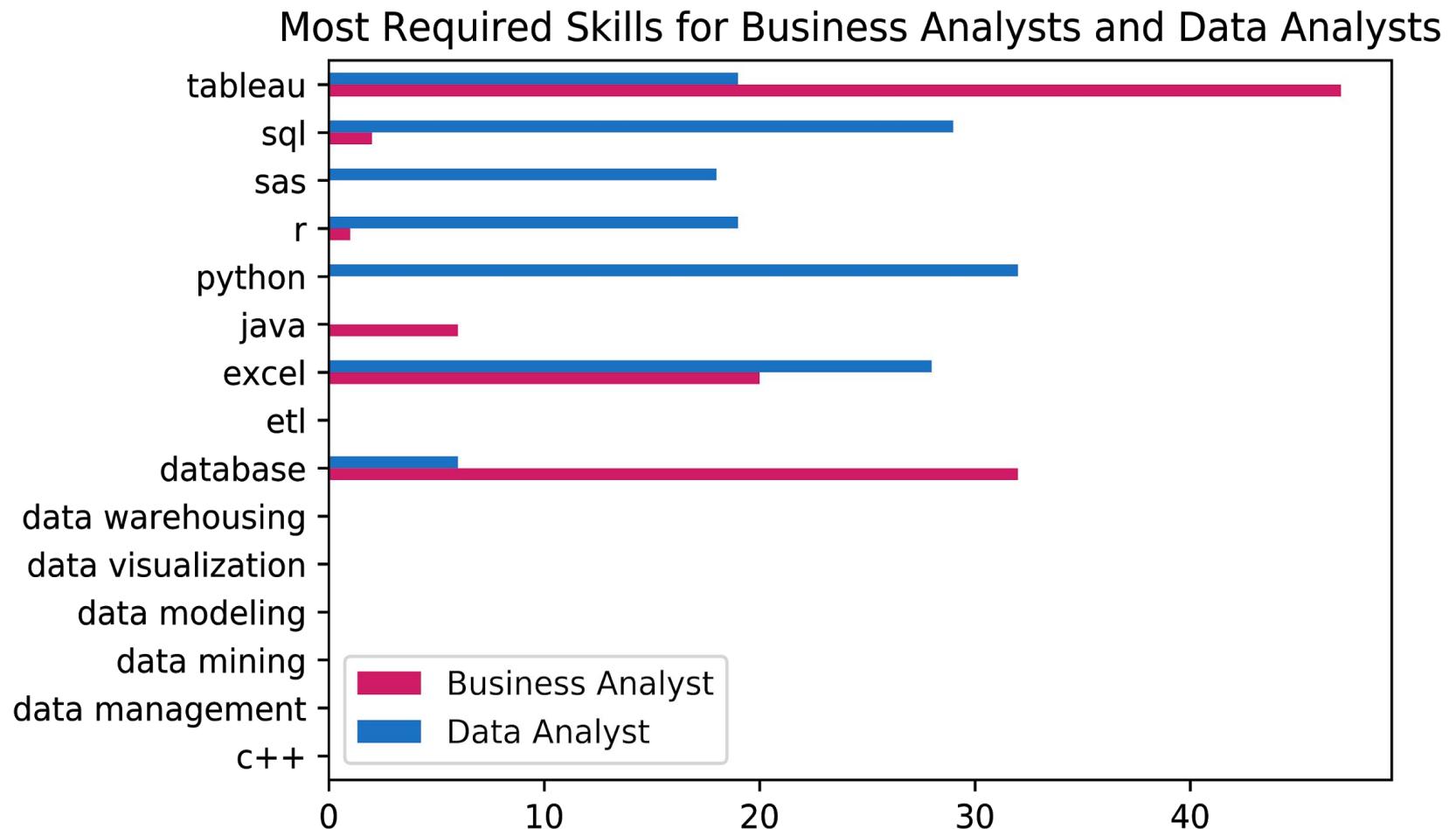
Job Descriptions Word Cloud

Using the Natural Language package in Python, the relatively higher occurrence words in all job descriptions include ‘deloitte’, ‘years’, ‘experiences’, ‘religion’, and ‘work’.



Most Required Skills

Using the Re method, some selected keywords were filtered out. The below chart shows the most required skills for the 2 most required analyst jobs: Business Analyst and Data Analyst.



THANK YOU!

Any questions?

Please don't hesitate to ask!

