Data Analytics

The Best Place to Work.

Objective:

Provide a snapshot of the current job market as it relates to the Data Science Field in the US job Market.

- Where are the most opportunities?
- Who is hiring?
- Where can I utilize my skills? Python?
- Where are the most competitive salaries offered?
- How much have salaries increased over the past year?
- Where is the best place to pursue a career?

Hypothesis:

The best place to pursue a career in data science and data analytics is New York City.

Methodology:

- ✓ Using an API, gather current live job postings

- ✓ Categorize listings into Data Science or Data Analyst roles

- ✓ Aggregate outside datasets for salaries, cost of living, and other analytics

- ✓ Summarize and analyze the findings

- ✓ Rank each location based on criteria

How do we define "Best"?

Lowest Score Wins!!!

Ranking System

1. Job Openings per City
2. Python Skills by Job Postings
3. Salary by City
4. YoY Salary Increase by City
5. Cost of Living by State

DS Salary		DS Increase YoY		DS Open Positions		COL Ranking		Python Required	
salary/K	point	%	point	# jobs	point	costIndex	point	# jobs	point
150-159	1	8-8.99	1	80-89	1	80-89	1	100-119	1
140-149	2	7-7.99	2	70-79	2	90-99	2	80-99	2
130-139	3	6-6.99	3	60-69	3	100-109	3	60-79	3
120-129	4	5-5.99	4	50-59	4	110-119	4	40-59	4
110-119	5	4-4.99	5	40-49	5	120-129	5	20-39	5
100-109	6	3-3.99	5	30-39	6	130-139	6	0-19	6
				20-29	7	140-149	7		
				10-19	8	150-159	8		
				0-9	9				

DA Salary		DA Increase YoY		DA Open Positions		COL Ranking		Python Required	
salary/K	point	%	point	# jobs	point	costIndex	point	# jobs	point
90-99	1	7-7.99	1	70-79	1	80-89	1	100-119	1
80-89	2	6-6.99	2	60-69	2	90-99	2	80-99	2
70-79	3	5-5.99	3	50-59	3	100-109	3	60-79	3
60-69	4	4-4.99	4	40-49	4	110-119	4	40-59	4
		3-3.99	5	30-39	5	120-129	5	20-39	5
				20-29	6	130-139	6	0-19	6
				10-19	7	140-149	7		
				0-9	8	150-159	8		

tdc

Our Data Sources

Source	Type	URL	Purpose
themuse	API (2) (6.20.20)	https://www.themuse.com/	Job & Company Search Tool / Career & Industry Guidance
G API node	API	https://developers.google.com/apis-explorer	Google Maps
glassdoor	CSV	https://www.glassdoor.com/research/job-market-report-historical/	Mean Salary Statistics
World Population Review	CSV	https://worldpopulationreview.com/states/cost-of-living-index-by-state/	Cost of Living Index

Dependencies

- Python
- Pandas
- Numpy
- Pandas

- Wx
- Sys
- Time
- Beautiful Soup

- Requests
- Matplotlib
- Gmaps
- Gmaps.dataset

Data Collection

- Two API calls to themuse.com
  - Job Search Endpoint
    - Four parameters passed
    - Output to CSV
  - Company Search Endpoint
    - Four parameters passed
    - Output to CSV

- API call to Google Maps
- Consolidation of 11 CSV files from 2 sites
- Created Data Collection tool

Data Wrangling

- Merged results from API calls to one DataFrame
- Parse Job skills
  - Parse job skill from HTML
    - Beautiful Soup

- Clean Data
  - Drop unwanted cities
  - Replace Job title with Simple Name
  - Split Location into City and State
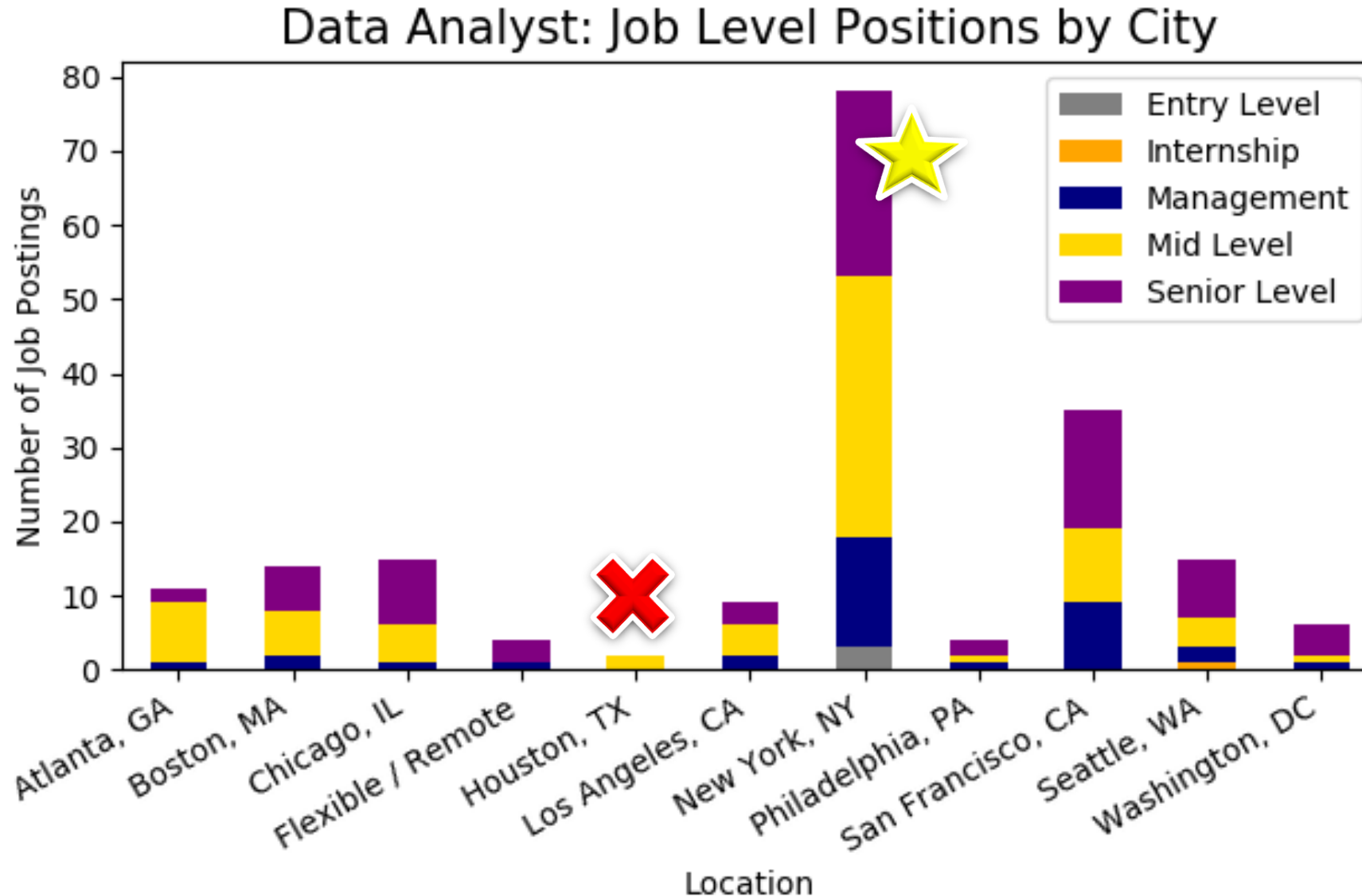- Combine 10 datasets to one DataFrame

Data Analysis:
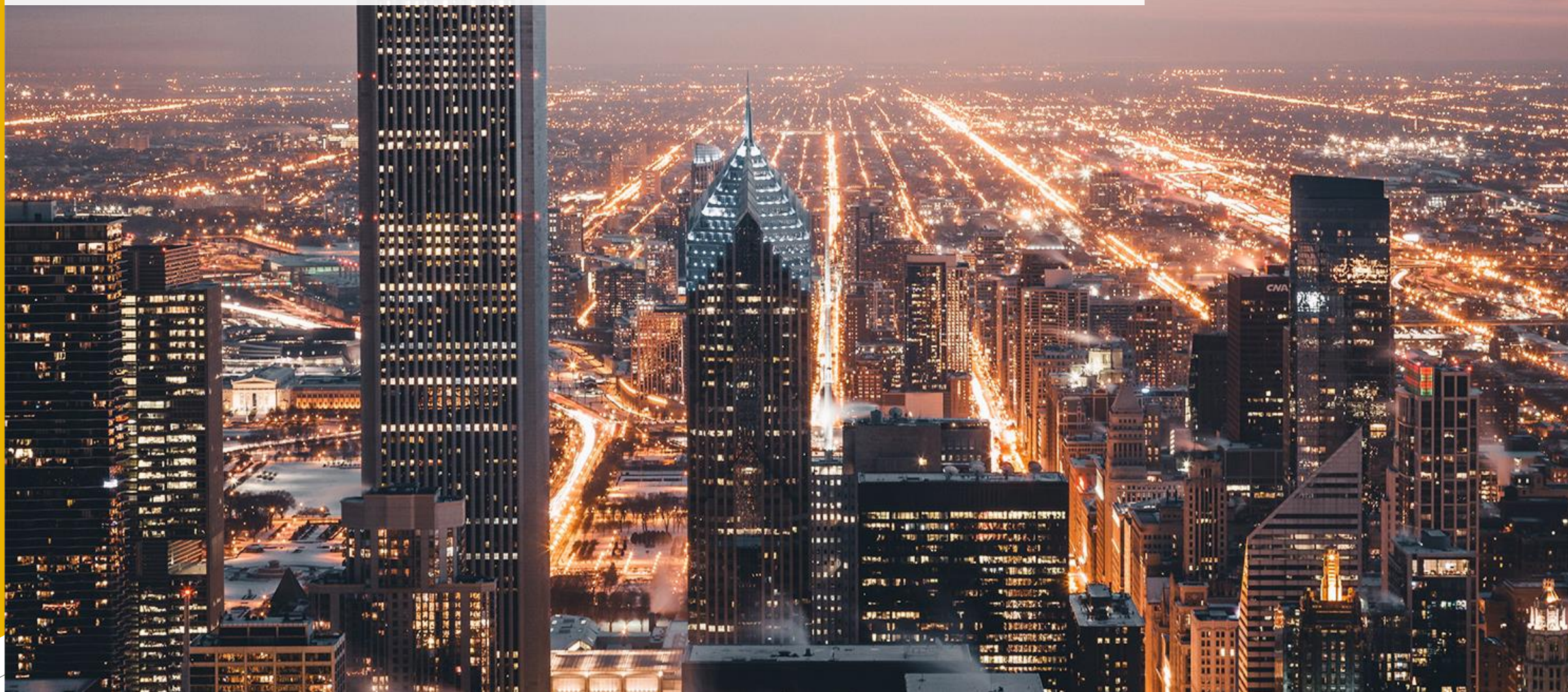
How many job openings does each city have for Data Science?

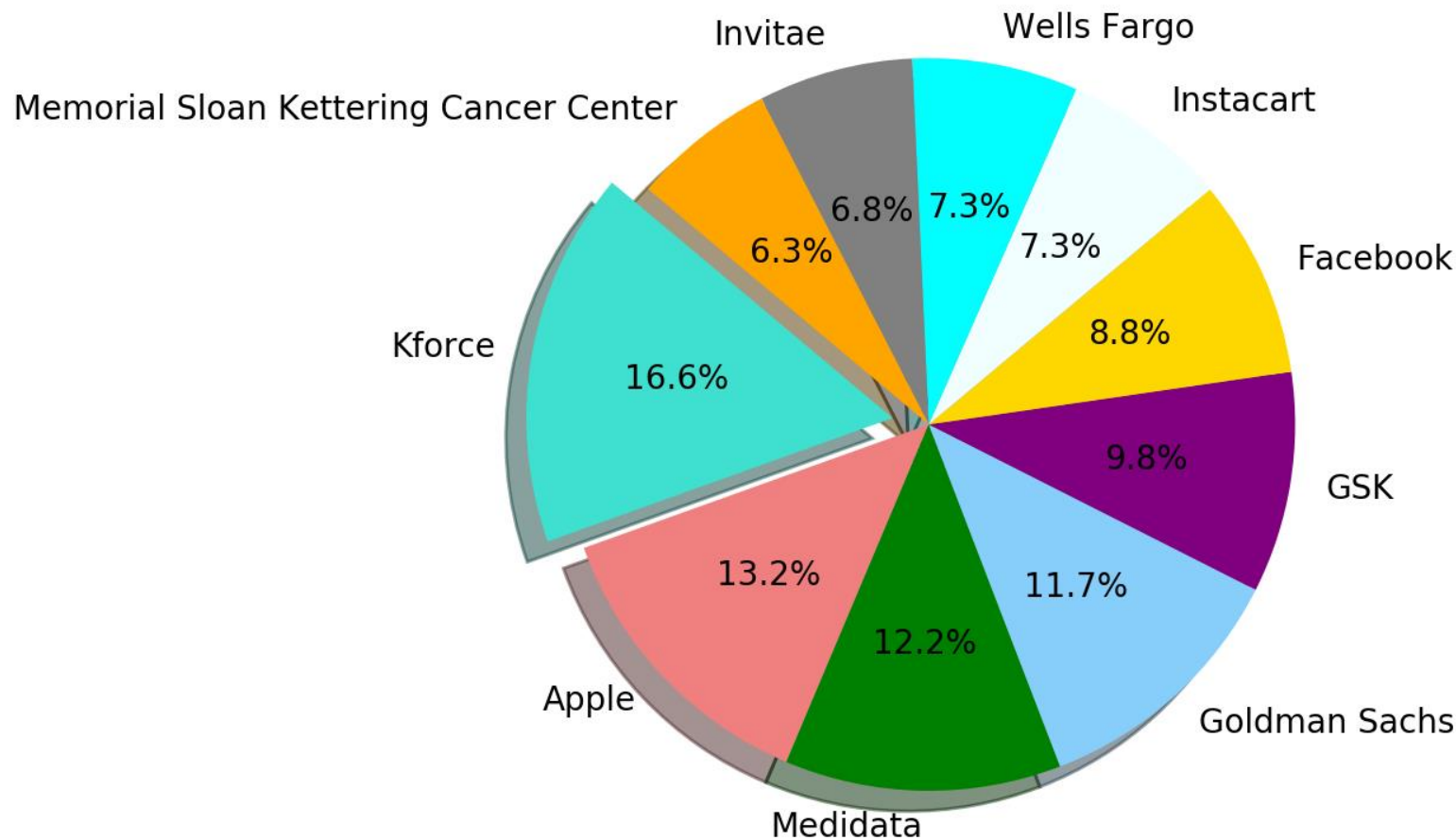Data Scientists: Job Level Positions by City
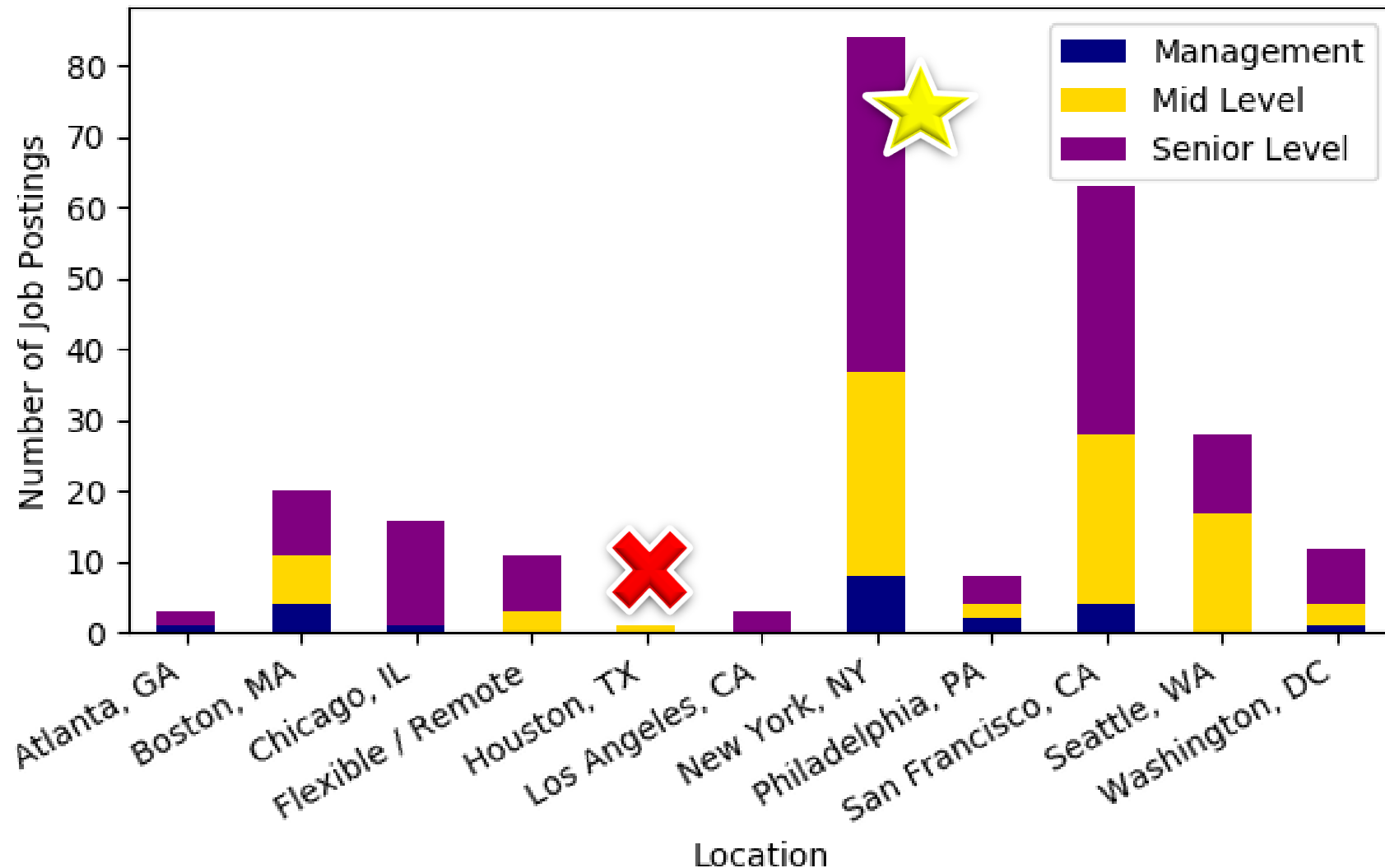
How many job openings does each city have for Data Analyst?

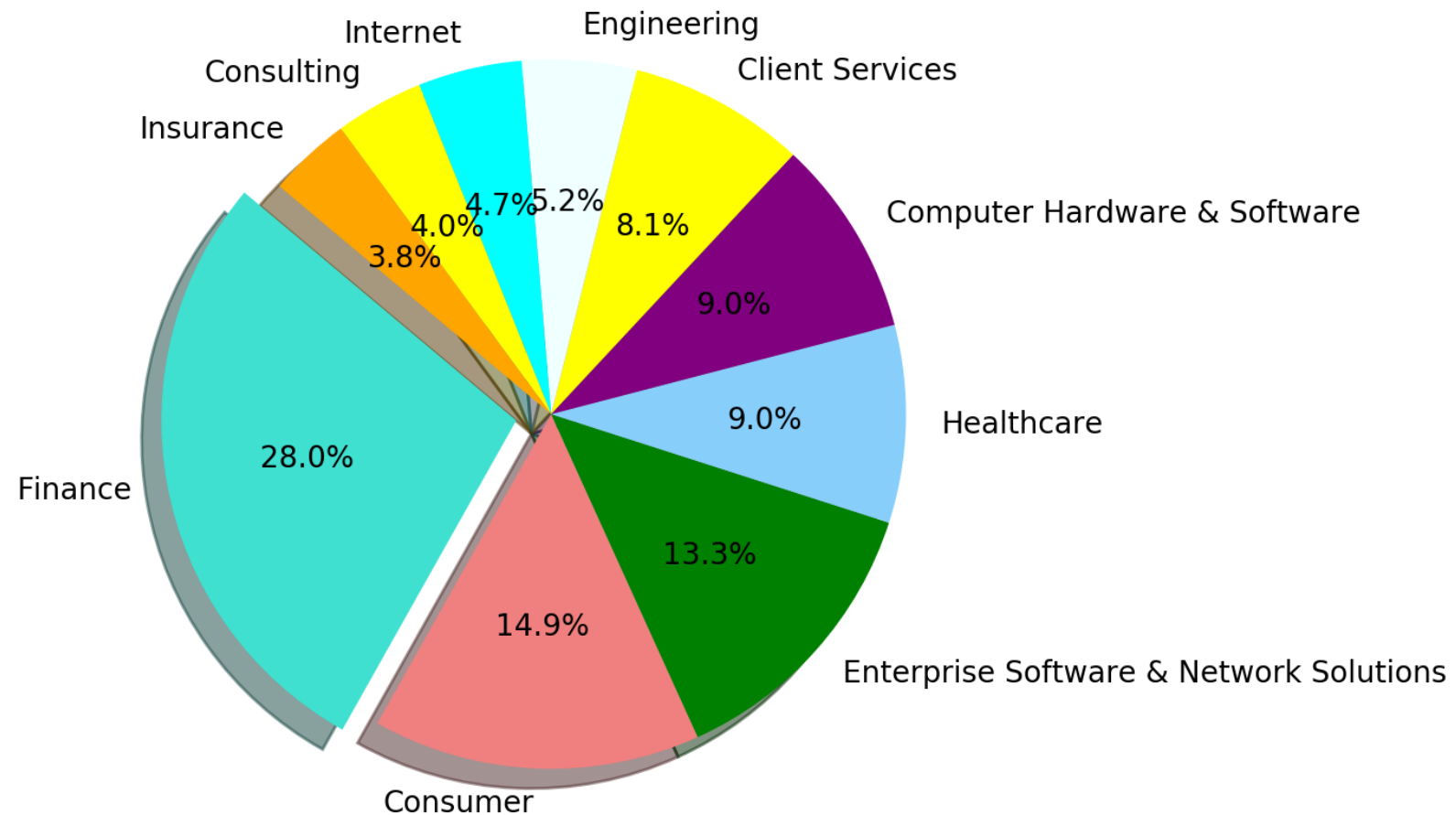# Top 10 Companies Posting Jobs in the Data Science Field

Top 10 Companies Posting Jobs in the Data Science Field

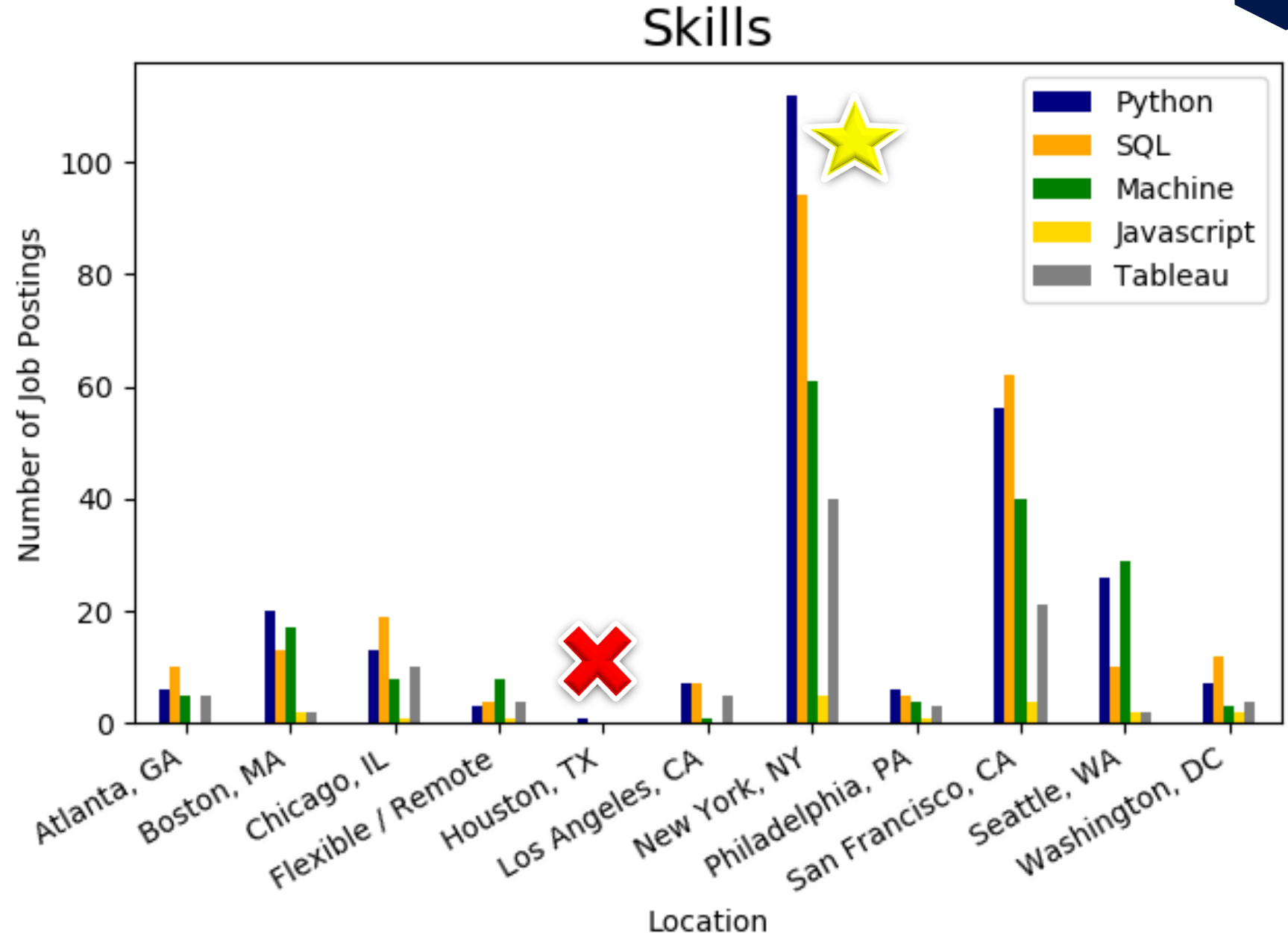# Top 10 Industries Hiring in the Data Science Field

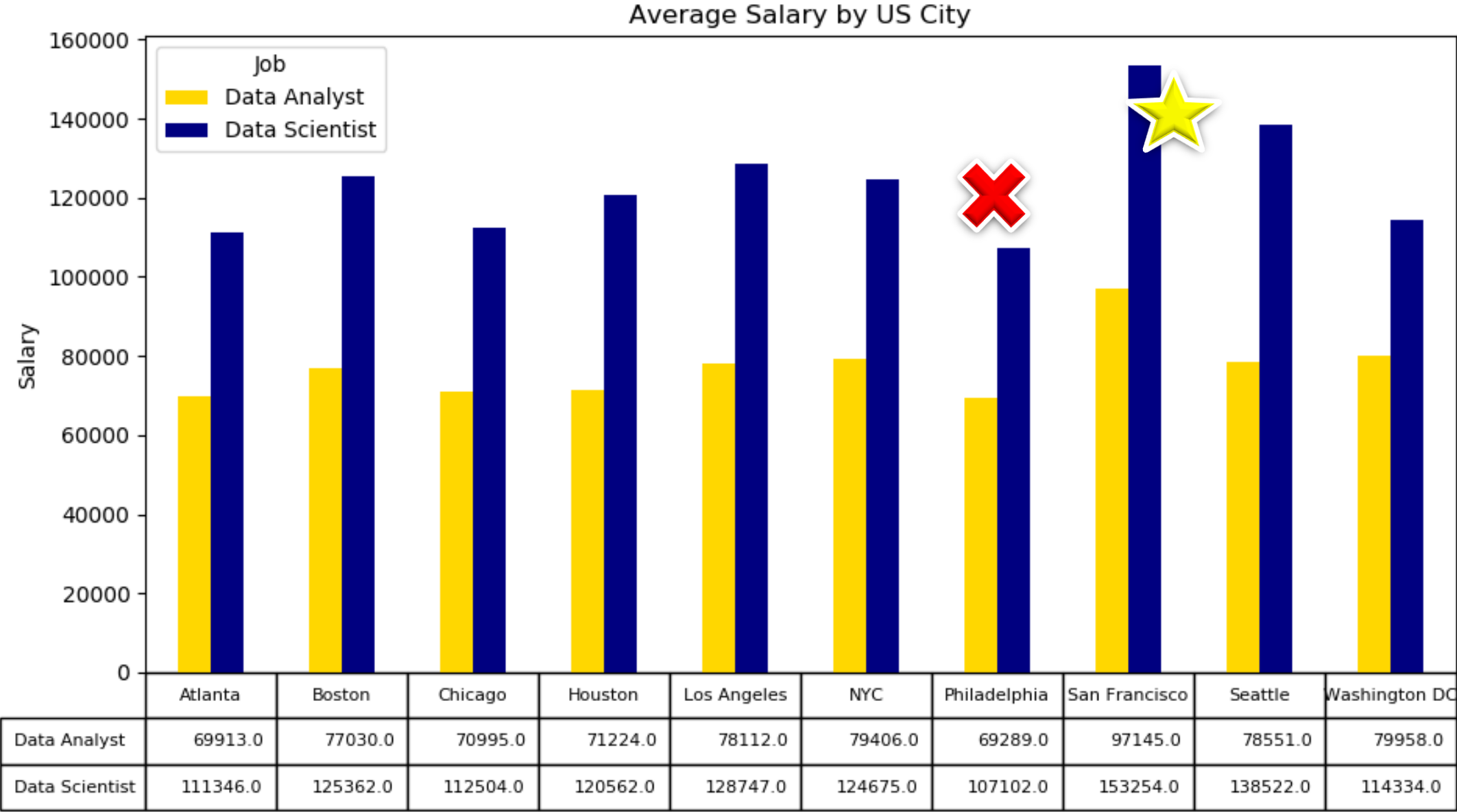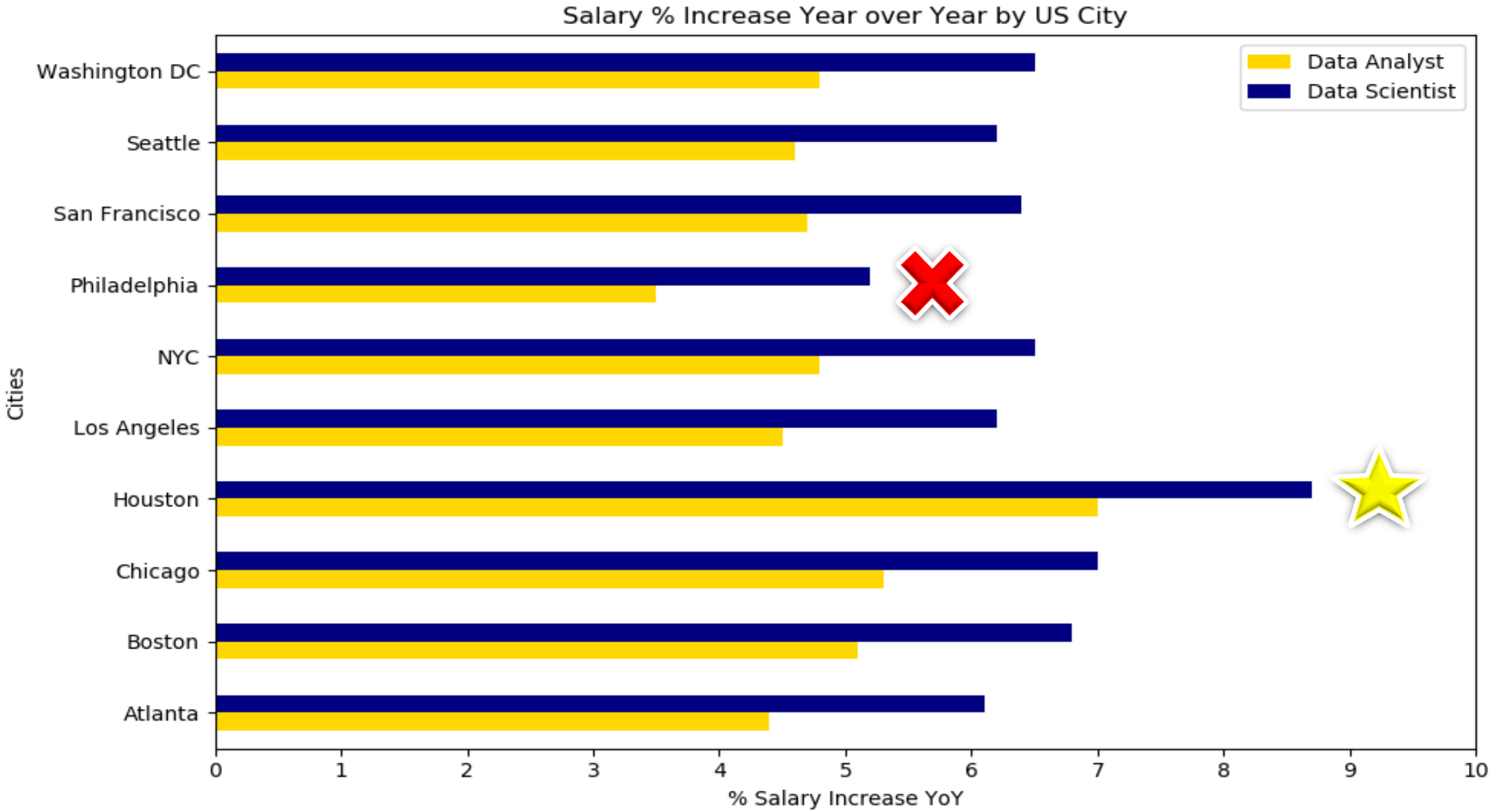Top 10 Industries Hiring in the Data Science Field

Skills

- Python is the most desirable/required skill across all cities

- SQL is the second most desirables/required skill in NYC but not all cities

- Machine learning and Tableau are present in all cities

13

Which US city has highest salary?

Average Salary by US City

	Atlanta	Boston	Chicago	Houston	Los Angeles	NYC	Philadelphia	San Francisco	Seattle	Washington DC
Data Analyst	69913.0	77030.0	70995.0	71224.0	78112.0	79406.0	69289.0	97145.0	78551.0	79958.0
Data Scientist	111346.0	125362.0	112504.0	120562.0	128747.0	124675.0	107102.0	153254.0	138522.0	114334.0

Which US city has highest salary increase over the last 12 months?

Salary % Increase Year over Year by US City

Which US State has the lowest cost of living?

findings
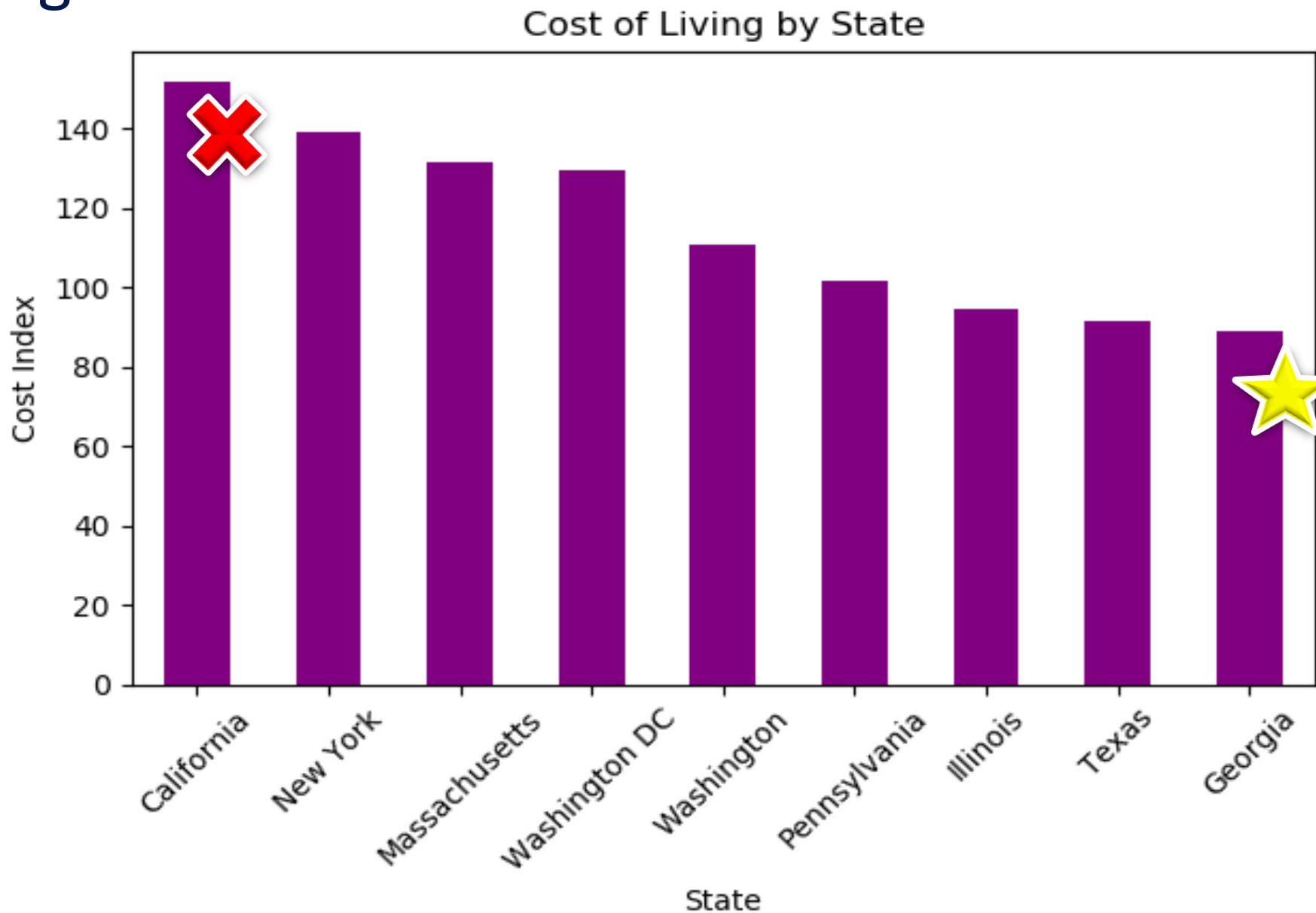
Cost of Living by State

Hypothesis:

The best place to pursue a career in data science and data analytics is New York City.

Summary – Ranking

Cities	DS Final Ranking	Rank COL	Rank DS Salary	Rank DS Salary YOY	Rank DS Posts	Rank Python	Final Score
NYC	1	6	4	3	1	1	15
San Francisco	2	8	1	3	3	4	19
Seattle	3.5	4	3	3	7	5	22
Houston	3.5	2	4	1	9	6	22
Chicago	5	2	5	2	8	6	23
Atlanta	6	1	5	3	9	6	24
Boston	7	6	4	3	7	5	25
Washington DC	8	5	5	3	8	6	27
Philadelphia	9	3	6	4	9	6	28
Los Angeles	10	8	4	3	9	6	30

Cities	DA Final Ranking	Rank COL	Rank DA Salary	Rank DA Salary YOY	Rank DA Posts	Rank Python	Final Score
NYC	1	6	3	4	1	1	15
Houston	2	2	3	1	8	6	20
Chicago	3	2	3	3	7	6	21
San Francisco	4.5	8	1	4	5	4	22
Atlanta	4.5	1	4	4	7	6	22
Seattle	6	4	3	4	7	5	23
Boston	7	6	3	3	7	5	24
Washington DC	8.5	5	3	4	8	6	26
Philadelphia	8.5	3	4	5	8	6	26
Los Angeles	10	8	3	4	8	6	29

Summary – Data Scientists
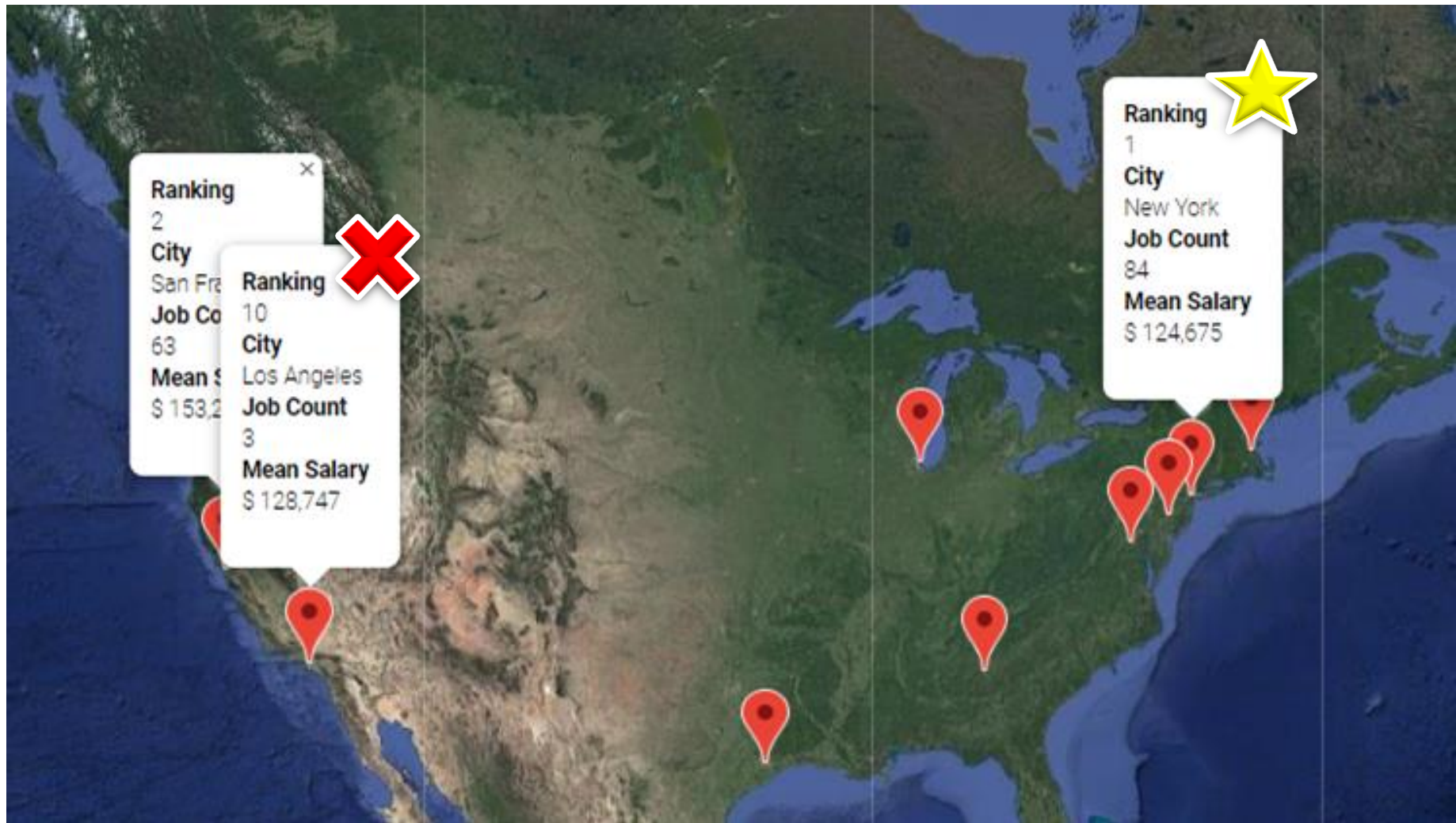
Cities	DS Final Ranking
NYC	1
San Francisco	2
Seattle	3.5
Houston	3.5
Chicago	5
Atlanta	6
Boston	7
Washington DC	8
Philadelphia	9
Los Angeles	10

Summary – Data Analyst
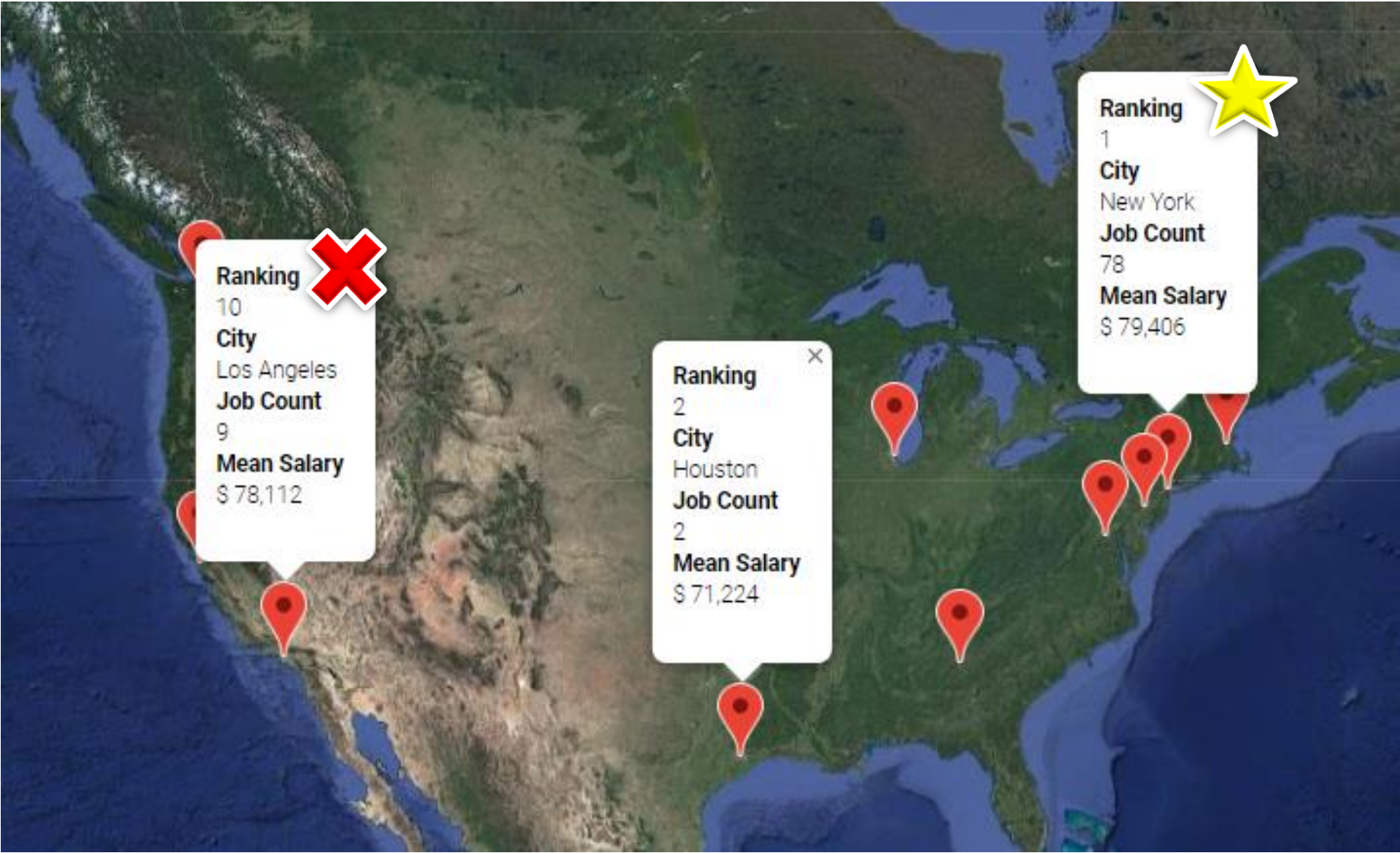
Cities	DA Final Ranking
NYC	1
Houston	2
Chicago	3
San Francisco	4.5
Atlanta	4.5
Seattle	6
Boston	7
Washington DC	8.5
Philadelphia	8.5
Los Angeles	10

Hypothesis:

The best place to pursue a career in data science and data analytics is New York City.

Alternate Conclusion??

Alternate Conclusion??

- Data Science

Cities	Alt Final Ranking
Houston	1.0
Chicago	2.5
Atlanta	2.5
Seattle	4.0
San Francisco	5.0
NYC	7.5
Boston	7.5
Washington DC	7.5
Philadelphia	7.5
Los Angeles	10.0

- Data Analyst

Cities	Alt Final Ranking
Houston	1.0
Chicago	2.0
Atlanta	3.0
Seattle	4.0
Boston	6.0
Washington DC	6.0
Philadelphia	6.0
San Francisco	8.5
NYC	8.5
Los Angeles	10.0

Challenges and Limitations

- Finding open APIs to be used for minimal costs (with limited time)
- Accessing current/relevant data
- Subjective grouping of Analysts vs Data Scientist
- Inconsistent data

Areas of Opportunity/Next Steps

- Use multitude of sources for job data / expand the sampling
- Gather additional information such as reviews and industry ranking

Q & A