Analysis & Display of Data Science Job Posting Data

Harker Russell, Jake Kim, Pratyush Dwivedi,
Ruihao Wei

Motivation and Objective

Motivation:

Uncertainty in preparation to reach career goals

Objective:

Provide Insight to University Students

Identify Trends in:

- Skill Preferences in Hiring
- Compensation by Industry & Company
- Job Availability by Position w/in Field

Motivation and Objective

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Motivation and Objective

- **Analyze** data for:
  - **Most Common Skill/Attribute terms**
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Motivation and Objective

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - **Most Common Positions**
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Motivation and Objective

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Motivation and Objective

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - **Is Salary ~ Job Satisfaction (Rating)?**
  - Is Salary ~ Hiring Frequency?

Motivation and Objective

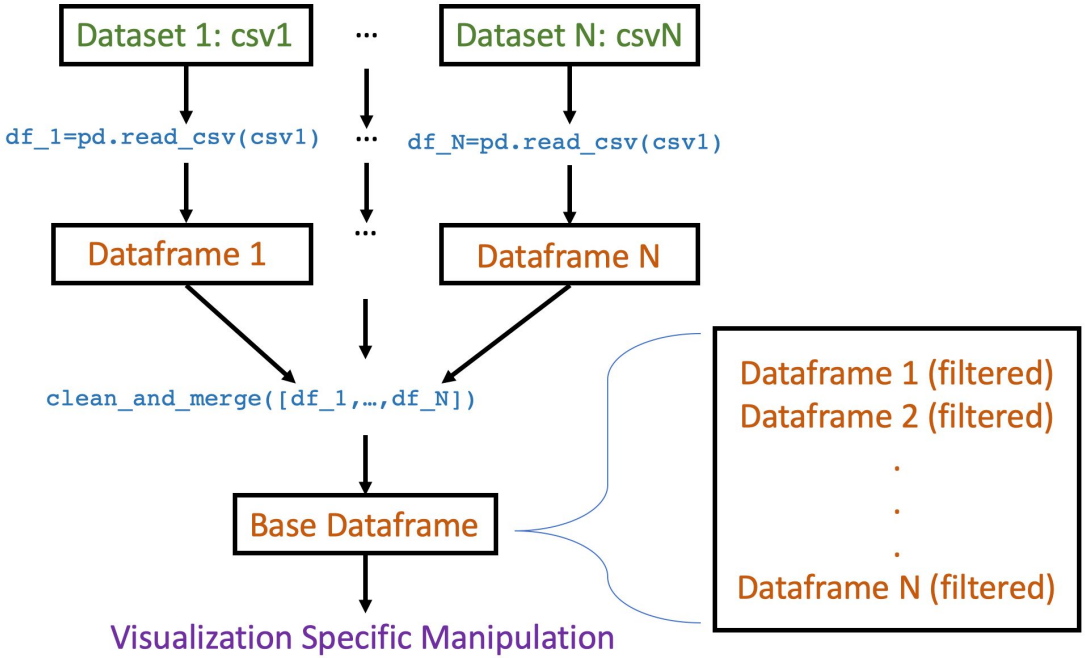
- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - **Is Salary ~ Hiring Frequency?**

Methodology

- **Clean Data**
- **Format Data** with Pandas Dataframes:
- **Analyze** as framed by questions.
- **Visualize** with appropriate plots

Methodology Workflow: Data formatting

- **Import:** CSV to Dataframe
- **Clean:** Removed spurious entries
- **Format:**
  - Concatenated cleaned dataframes
  - Used concatenated dataframe as ‘base’



Methodology Workflow: Visualizations

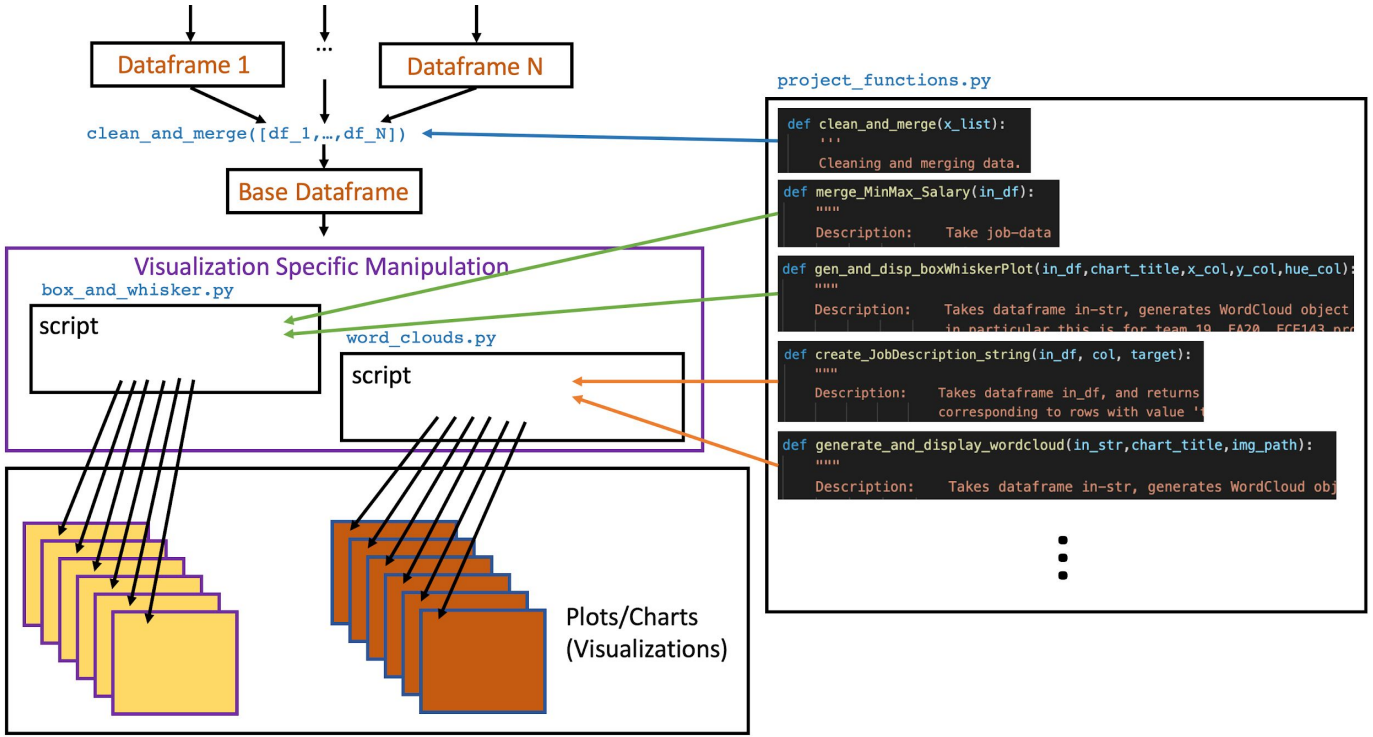
Did our best to stay **modular:**

All functionality used more than once?

Imported from file with common functions.

Propagates features/fixes across all instances of use.

Makes code more readable.



Dataset

- Data Science job listings scraped from glassdoor June 2020

- **4 Sets** as .csv files
  - Texas, New York, Washington, & SF Bay Area

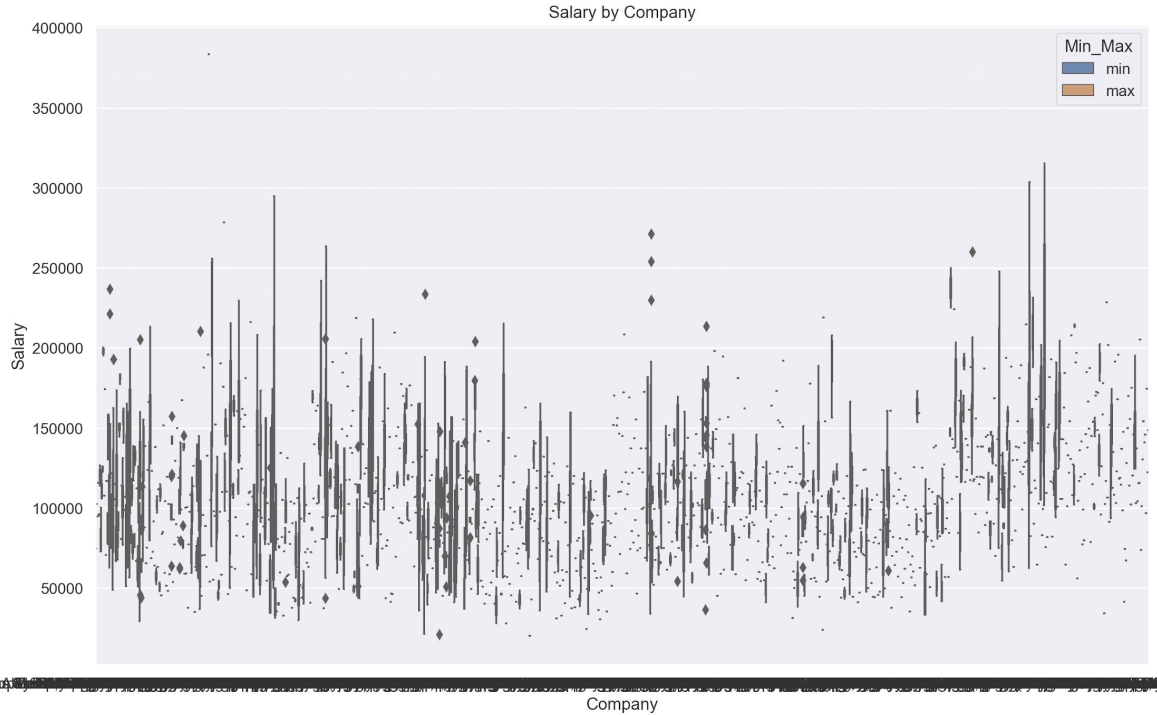
- **12 Useful Fields**
  - e.g. Job Description, Job Title, Salary

https://www.kaggle.com/atharvap329/glassdoor-data-science-job-data

Dataset: Issues

State	City	Min_Salary	Max_Salary
NY	New York	-1	-1

ABA Therapist	Kids Learning Loft Applied Behavic
Construction Project Manager	The LiRo Group
Diesel Mechanic	Kingdom Associates
Plumber	Plumbing for life Inc.
OT/ICS Systems Engineer	Forescout Technologies Inc.

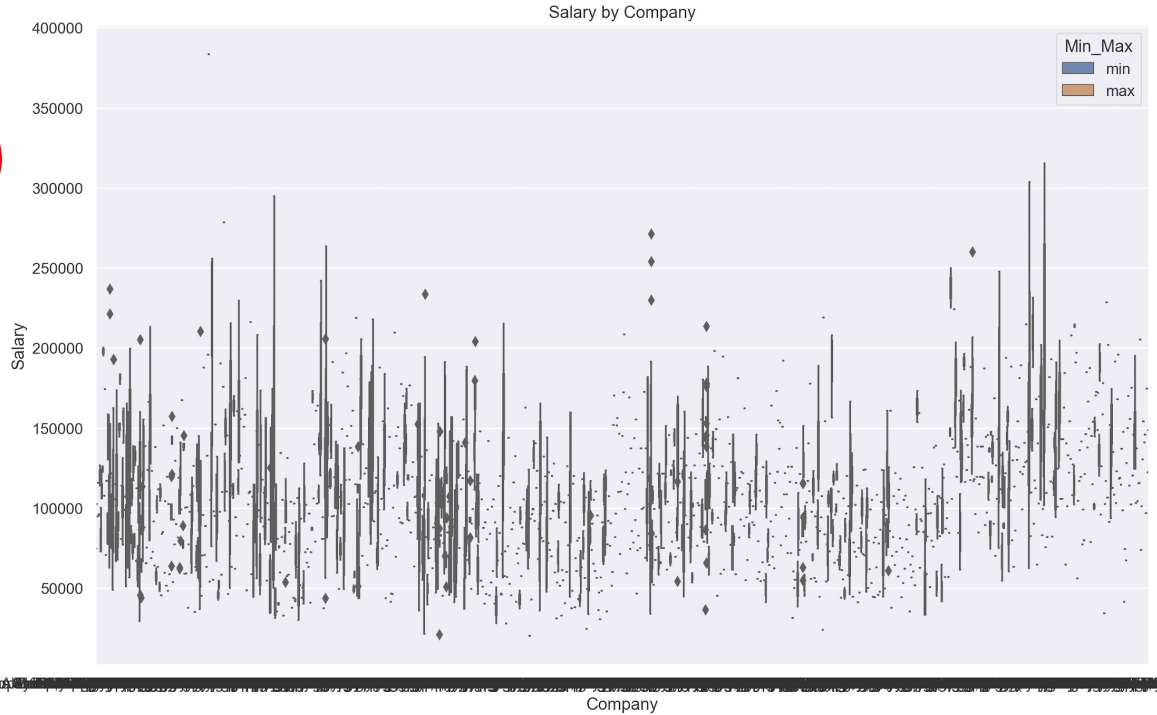


Dataset: Issues

Odd Salary

State	City	Min_Salary	Max_Salary
NY	New York	-1	-1

ABA Therapist	Kids Learning Loft Applied Behavic
Construction Project Manager	The LiRo Group
Diesel Mechanic	Kingdom Associates
Plumber	Plumbing for life Inc.
OT/ICS Systems Engineer	Forescout Technologies Inc.



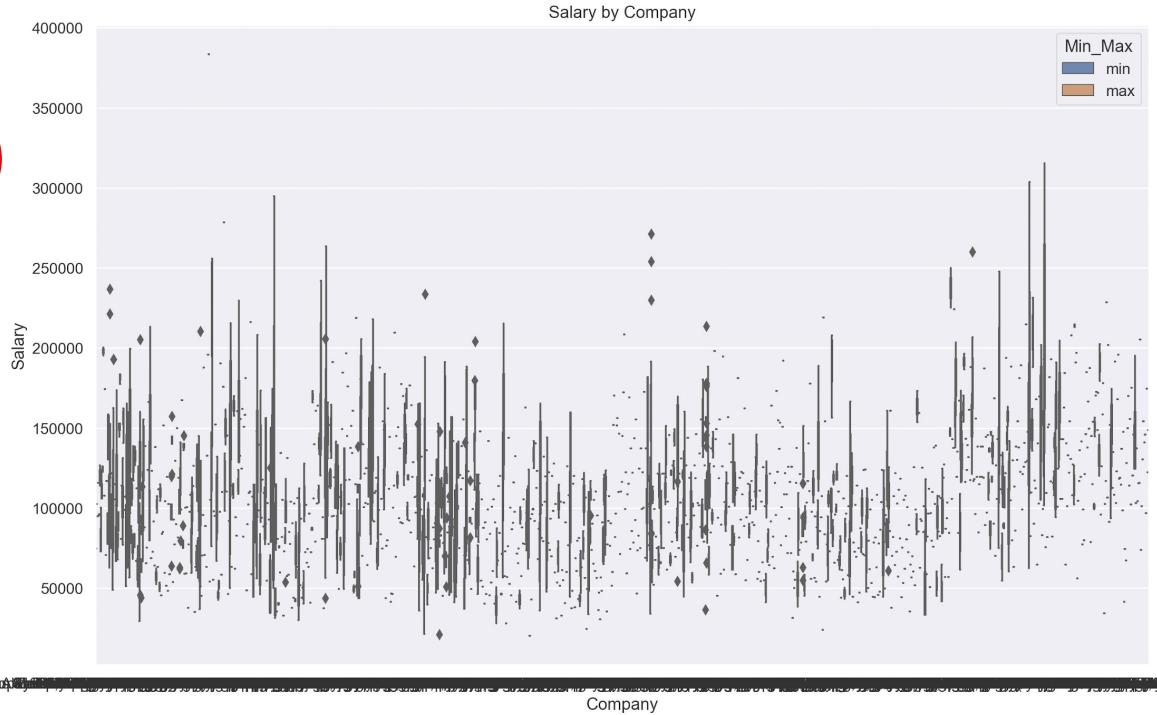
Dataset: Issues

Odd Salary

State	City	Min_Salary	Max_Salary
NY	New York	-1	-1

Irrelevant

ABA Therapist	Kids Learning Loft Applied Behavic
Construction Project Manager	The LiRo Group
Diesel Mechanic	Kingdom Associates
Plumber	Plumbing for life Inc.
OT/ICS Systems Engineer	Forescout Technologies Inc.



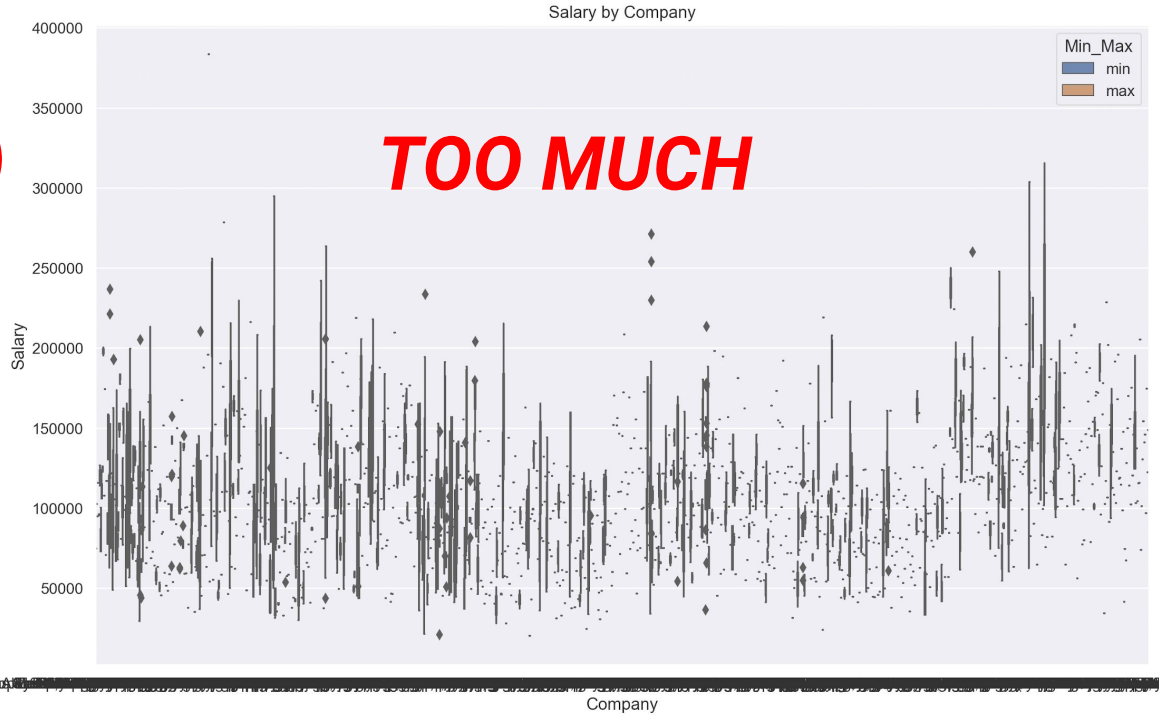
Dataset: Issues

Odd Salary

State	City	Min_Salary	Max_Salary
NY	New York	-1	-1

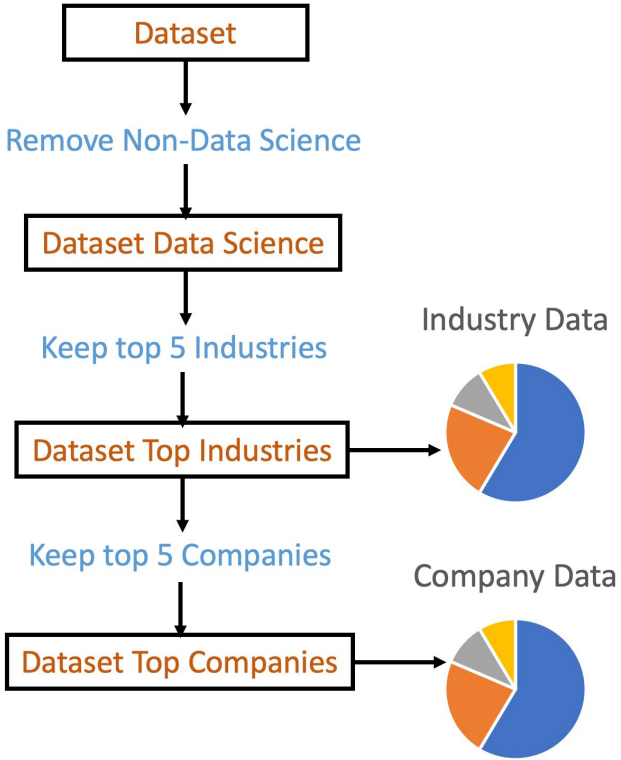
Irrelevant

ABA Therapist	Kids Learning Loft Applied Behavic
Construction Project Manager	The LiRo Group
Diesel Mechanic	Kingdom Associates
Plumber	Plumbing for life Inc.
OT/ICS Systems Engineer	Forescout Technologies Inc.



Dataset: Issues

- **Filtered** by number of posts
- More intuitive plots



Findings

Job Descriptions by Industry

- **Analyze** data for:
  - **Most Common Skill/Attribute terms**
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Word Clouds (Job Descriptions by Industry)

In General:

Experience is huge:

Skills, degree

Data Analysis is prevalent

Various coding languages:

Python, js, matlab

Word Clouds (Job Descriptions by Company)

In General:
Experience is huge:
Skills, degree
Data analysis is prevalent
Various coding languages:
Python, js, matlab

Interesting
Interpersonal / ‘soft skills’ (i.e. communication, team, presentation) seem to be very desirable.

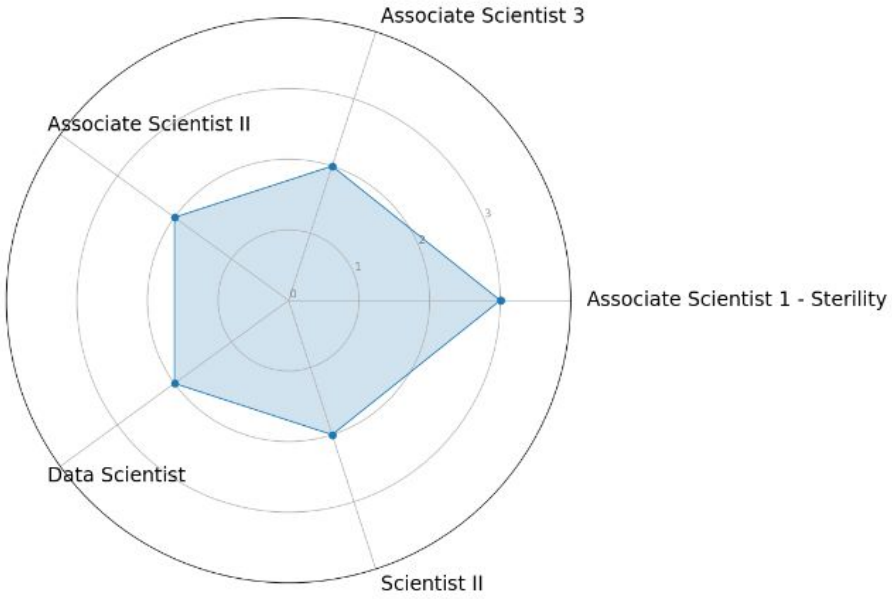
Job Profiles by Industry

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - **Most Common Positions**
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

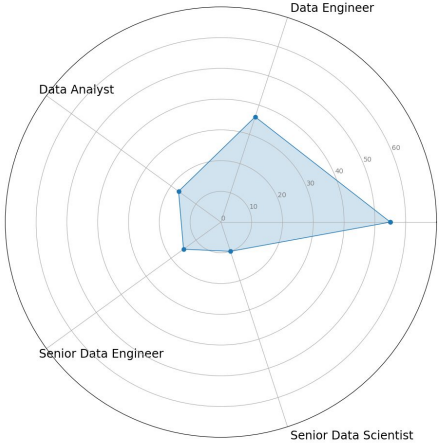
Job Titles by Industry

Job profiles for Biotech industry were unique w.r.t others.

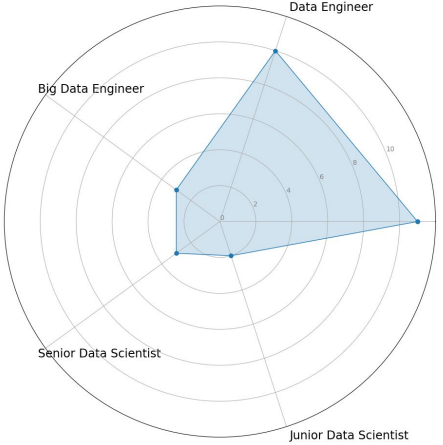
Most were scientists: performing analysis on lab data (for drug development, disease analysis etc.)



Most demanded job profiles in Biotech & Pharmaceuticals



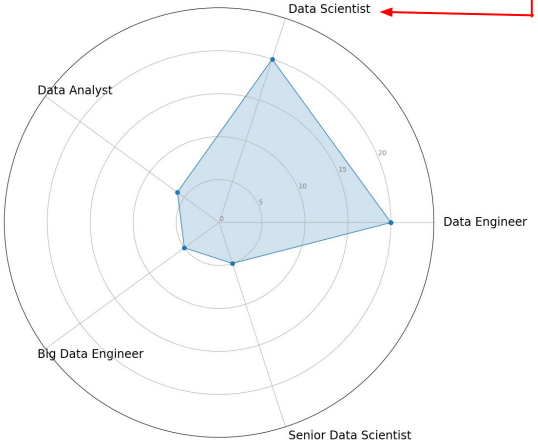
Most demanded job profiles in Information Technology



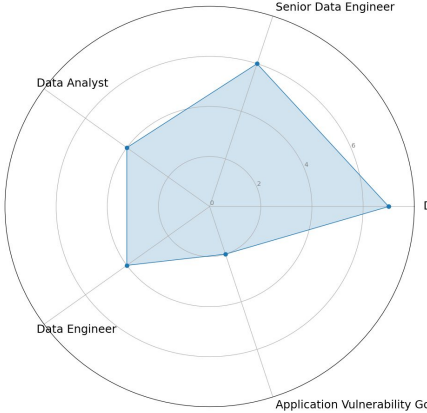
Most demanded job profiles in Aerospace & Defense

Most common position??

Data Scientist



Most demanded job profiles in Business Services



Most demanded job profiles in Finance

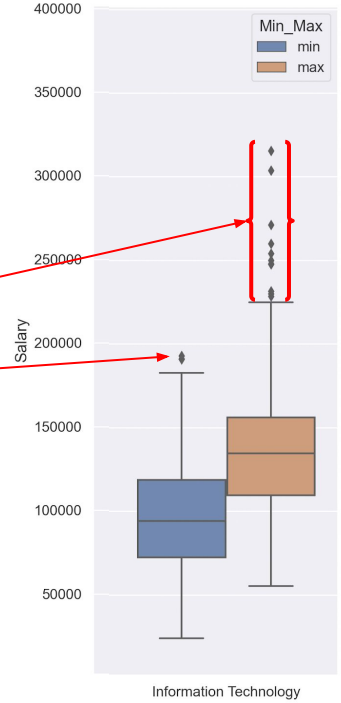
Max & Min Salary by Industry

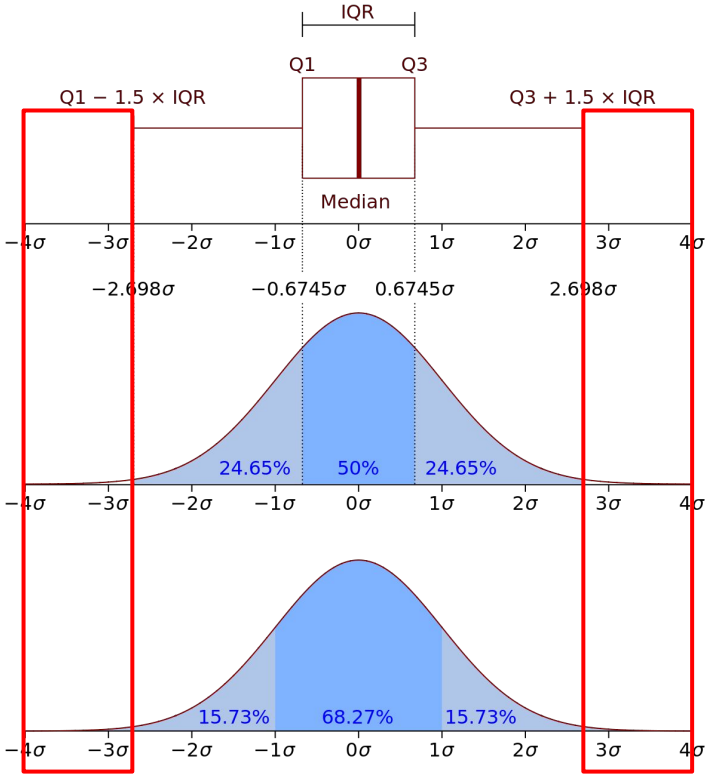
- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Quick Explanation: Box and Whisker Plots

- **Diamonds** are outliers
- **Whiskers** show Max and Min non-outliers
- **Box** represents mean
- **Centerline** of box represents median

Outliers

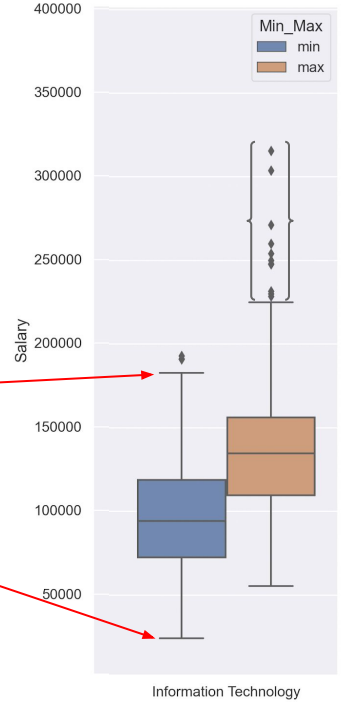


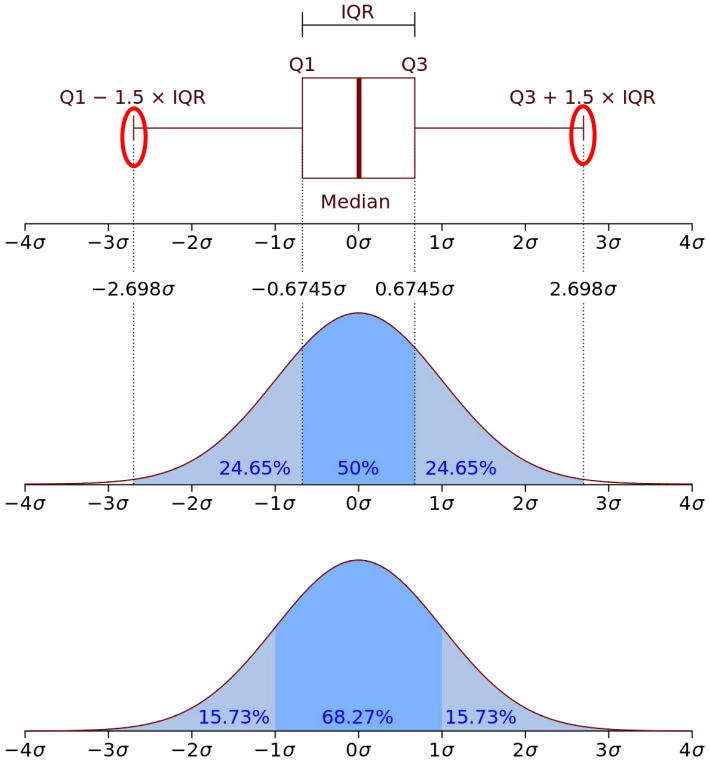


Quick Explanation: Box and Whisker Plots

- **Diamonds** are outliers
- **Whiskers** show Max and Min non-outliers
- **Box** represents mean
- **Centerline** of box represents median

High & Lows

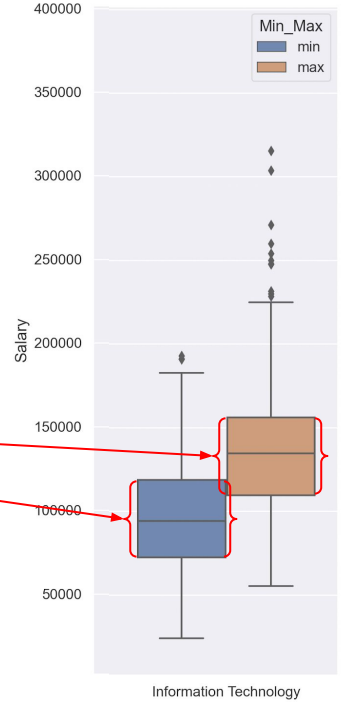


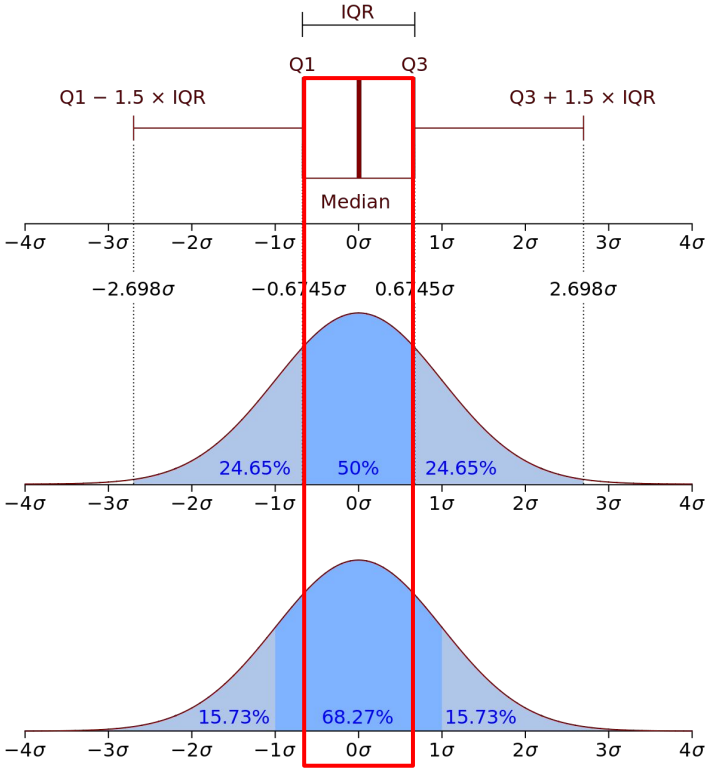


Quick Explanation: Box and Whisker Plots

- **Diamonds** are outliers
- **Whiskers** show Max and Min non-outliers
- **Box** represents mean
- **Centerline** of box represents median

Mean

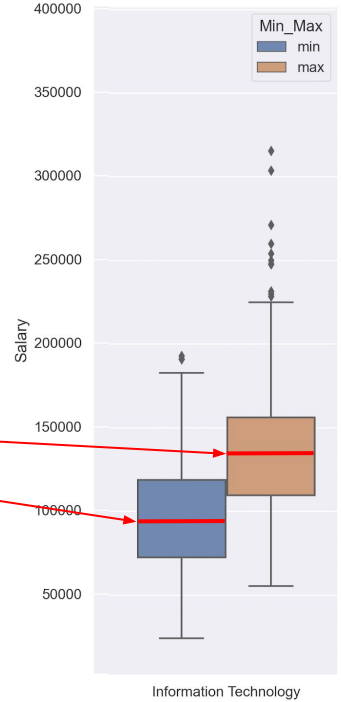


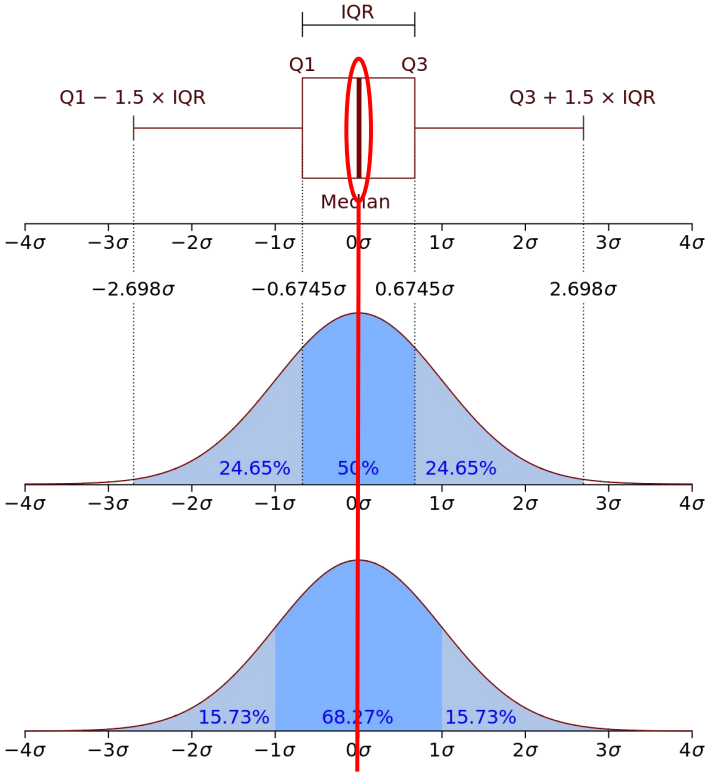


Quick Explanation: Box and Whisker Plots

- **Diamonds** are outliers
- **Whiskers** show Max and Min non-outliers
- **Box** represents mean
- **Centerline** of box represents median

Median





Max & Min Salary by Industry

Min/Max Salary **by industry**

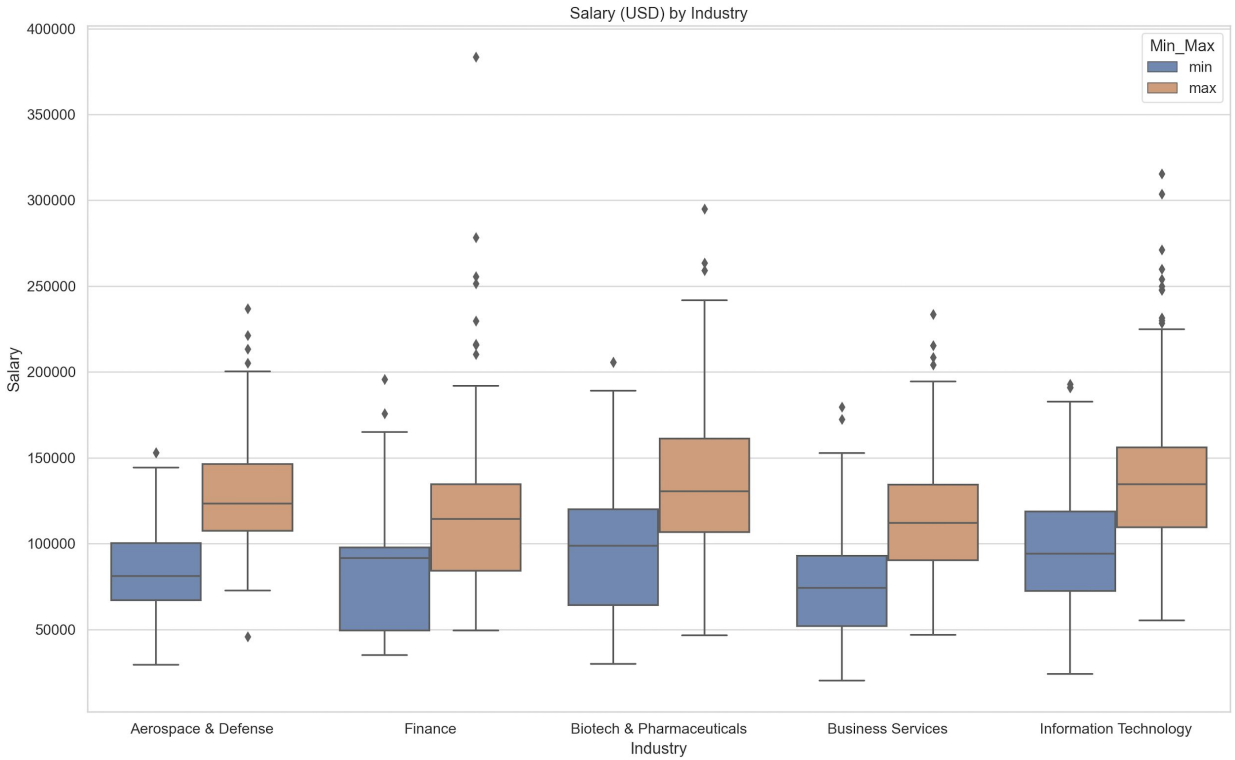
Essentially flat

High Salary:

Biotech & Pharmaceuticals

Low Salary:

Business & Finance



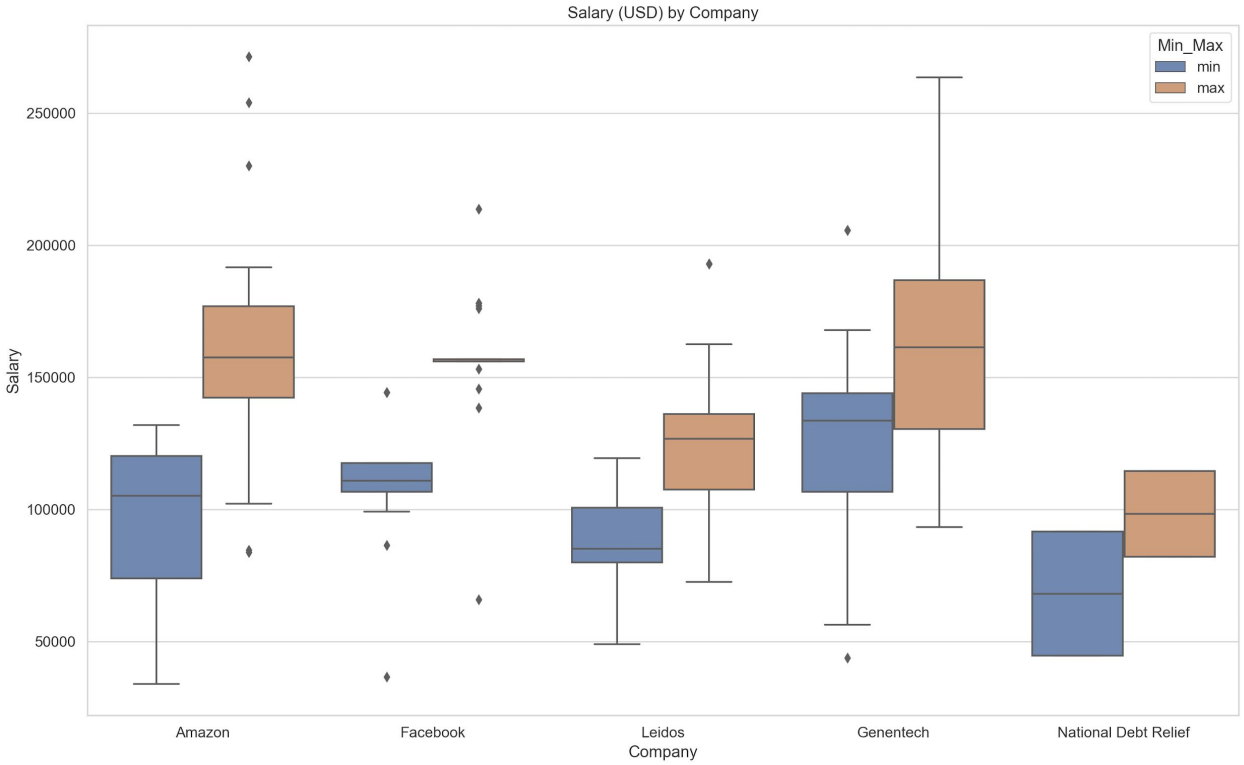
Max & Min Salary by Company

Min/Max Salary **by company**

High Salary:

Genentech
(Biotech & Pharmaceuticals)

Highest median minimum and maximum pay.



Average Salary: Top 3 Companies by Industry

Average Salaries:
Top Companies by Industry:

High Salary:
Biotech & Pharmaceuticals

Low Salary:
Business & Finance

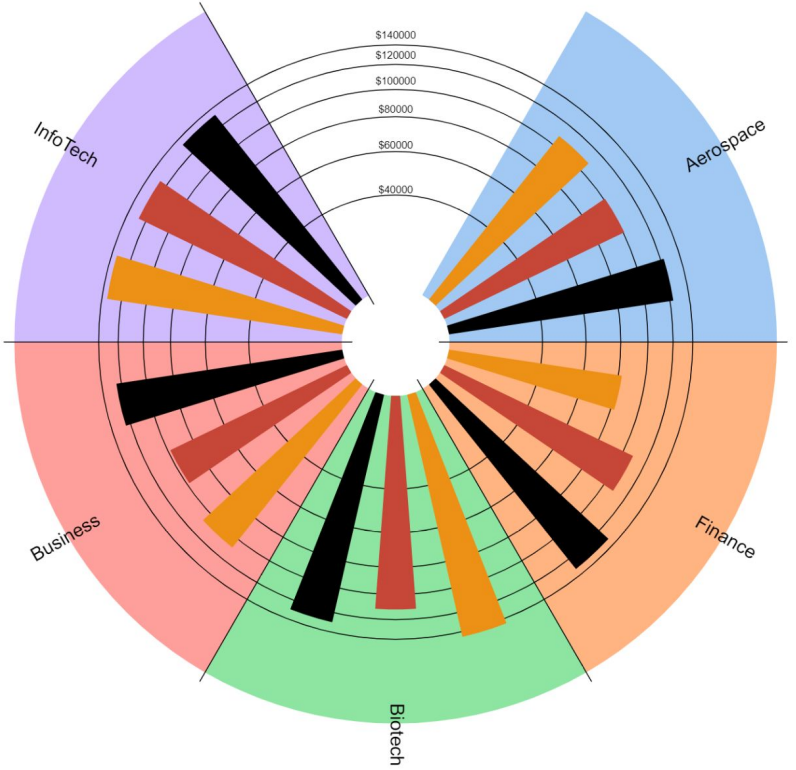
Industry-Aerospace
Leidos: $108342
BAE Systems USA: $100770
The Aerospace Corporation: $124323

Industry-Finance
National Debt Relief: $83201
Capital One: $108667
Chime: $133391

Industry-Biotech
Genentech: $146911
AstraZeneca: $112065
Amgen: $130681

Industry-Business
LMI: $107879
Booz Allen Hamilton: $97921
Guidehouse: $125944

Industry-InfoTech
Forescout Technologies Inc.: $1355
Amazon: $128606
Facebook: $133789



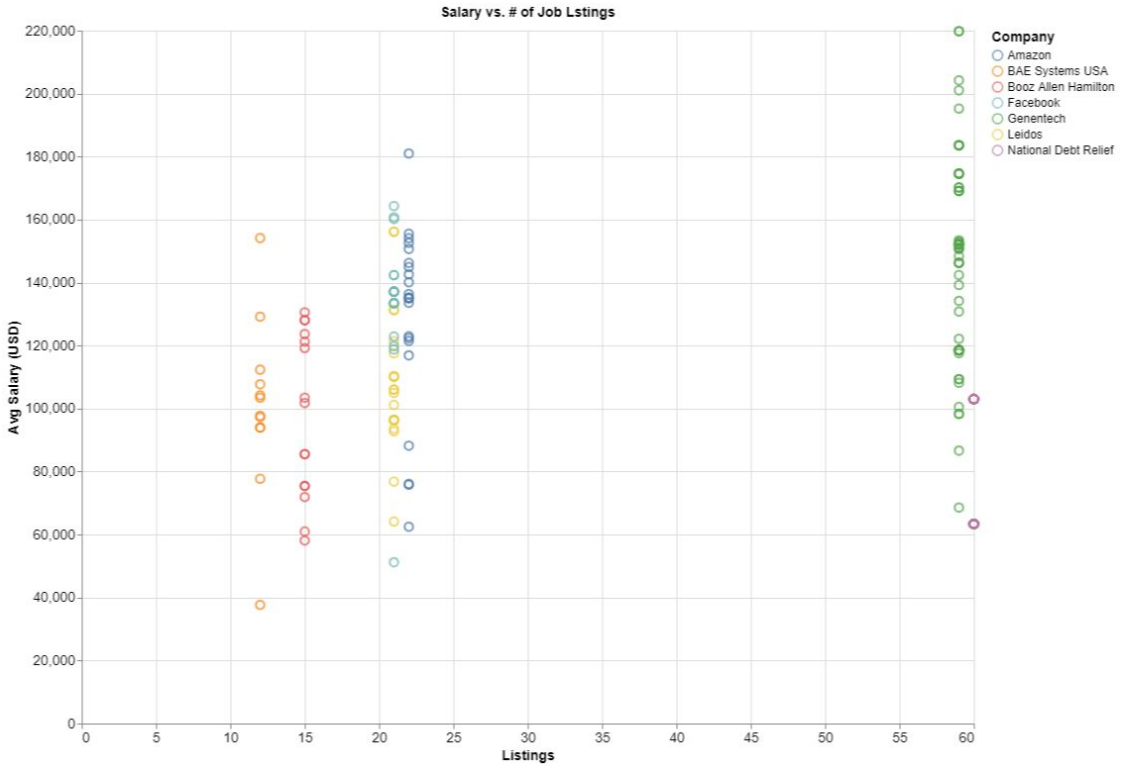
of Job Listings vs Salary

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - **Is Salary ~ Hiring Frequency?**

Salary vs Job Listings

Job Listings vs Salary:

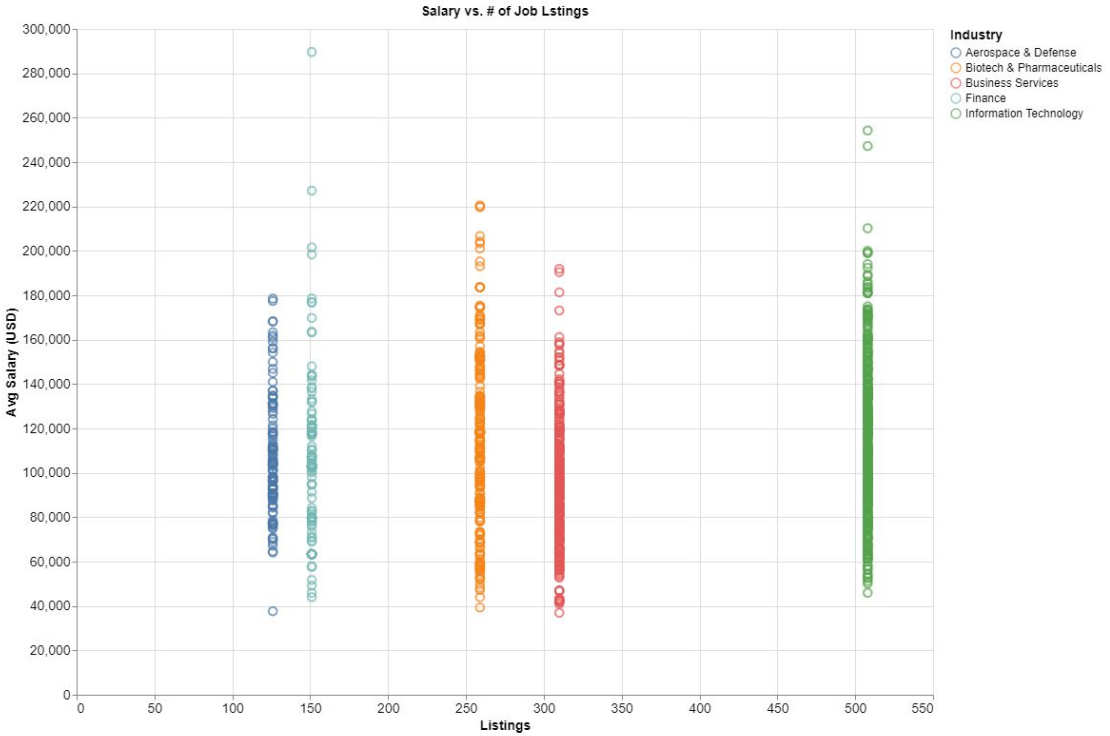
Salary flat as listings rise



Salary vs Job Listings

Job Listings vs Salary

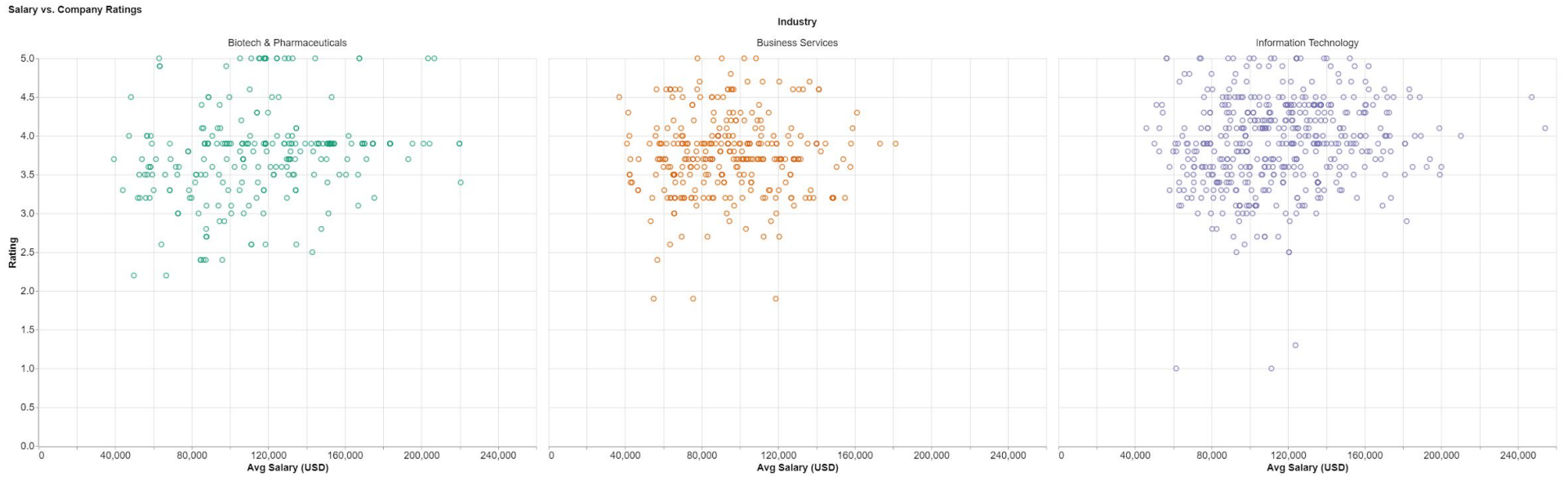
Salary flat as number of listings rise



Salary ~ Rating

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - **Is Salary ~ Job Satisfaction (Rating)?**
  - Is Salary ~ Hiring Frequency?

Salary ~ Rating

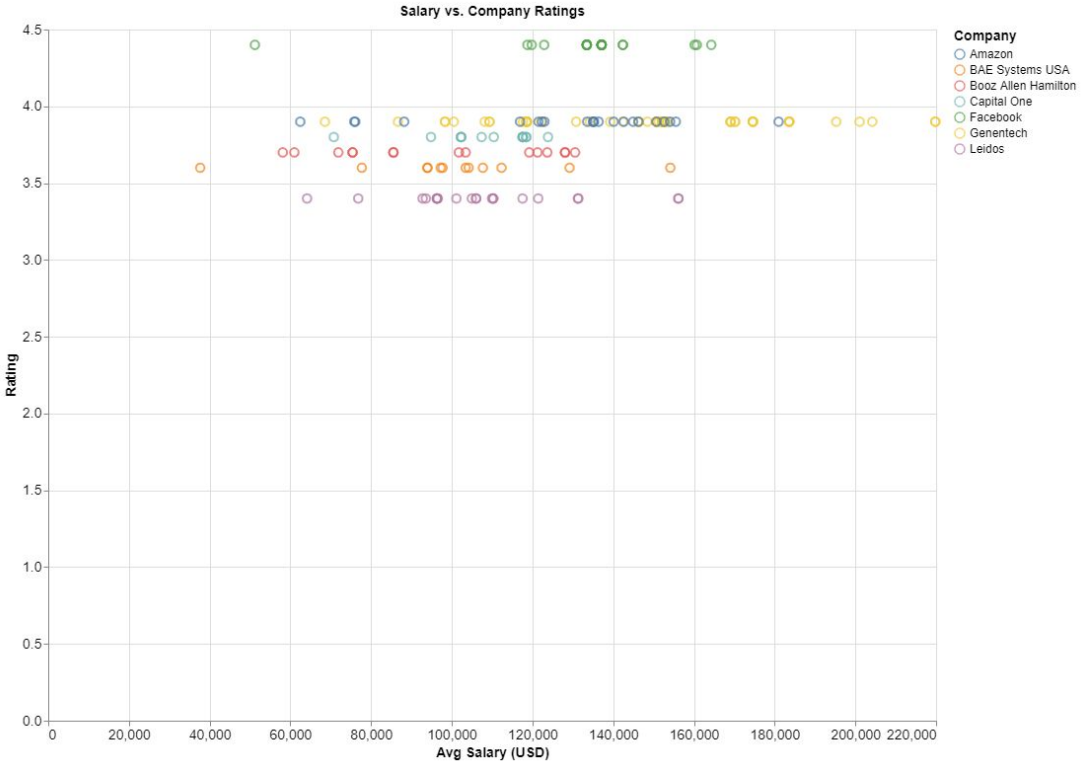


Salary ~ Rating

Hypothesized that there would be **strong correlation** between salary and rating

In reality there was **little to no correlation**

Might indicate that company value and environment is a much better measure of company



Conclusions

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

- Experience & Teamwork
- Title > Industry
- Don’t Sweat Salary!

Conclusions: 3 Takeaways

- **Analyze** data for:
  - **Most Common Skill/Attribute terms**
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

- **Experience & Teamwork**
- Title > Industry
- Don’t Sweat Salary!

Conclusions: 3 Takeaways

- **Analyze** data for:
  - **Most Common Skill/Attribute terms**
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

- **Experience & Teamwork**
  - **Clubs, Hackathons**
- Title > Industry
- Don’t Sweat Salary!

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - **Most Common Positions**
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

- Experience & Teamwork
  - Clubs, Hackathons

- **Title > Industry**

- Don’t Sweat Salary!

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

- Experience & Teamwork
  - Clubs, Hackathons

- Title > Industry

- **Don’t Sweat Salary!**

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - **Is Salary ~ Job Satisfaction (Rating)?**
  - Is Salary ~ Hiring Frequency?

- Experience & Teamwork
  - Clubs, Hackathons

- Title > Industry

- **Don't Sweat Salary!**
  - **Benefits and Culture**

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - **Is Salary ~ Job Satisfaction (Rating)?**
  - Is Salary ~ Hiring Frequency?

- Experience & Teamwork
  - Clubs, Hackathons

- Title > Industry

- **Don't Sweat Salary!**
  - **Benefits and Culture**
  - **$85 - $125k**

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - **Salary by Industry & Company**
  - **Is Salary ~ Job Satisfaction (Rating)?**
  - **Is Salary ~ Hiring Frequency?**

- Experience & Teamwork
  - Clubs, Hackathons

- Title > Industry

- **Don't Sweat Salary!**
  - **Benefits and Culture**
  - **$85 - $125k**
  - **~Not 'Rare'**

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Any Questions?

- Experience & Teamwork
  - Clubs, Hackathons
- Title > Industry
- Don’t Sweat Salary!
  - Benefits & Culture
  - $85k - $125k
  - ~Not ‘Rare’

Conclusions: 3 Takeaways

- A

rms

ting)?

Any C

- Experience & Teamwork
  - Clubs, Hackathons
- Title > Industry
- Don’t Sweat Salary!
  - Benefits & Culture
  - $85k - $125k
  - ~Not ‘Rare’

Conclusions: 3 Takeaways

- **Analyze** data for:
  - Most Common Skill/Attribute terms
  - Most Common Positions
  - Salary by Industry & Company
  - Is Salary ~ Job Satisfaction (Rating)?
  - Is Salary ~ Hiring Frequency?

Any Questions?

- Experience & Teamwork
  - Clubs, Hackathons
- Title > Industry
- Don’t Sweat Salary!
  - Benefits & Culture
  - $85k - $125k
  - ~Not ‘Rare’