Team 10
Mariam Germanyan, Taline Mardirossian,
Aleksandra Ma, Riya Prahlad, Ayesha Yusuf
Mentor: James Hodson

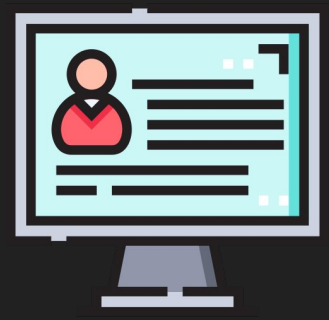
Employment Dynamics

Bay Area

Objective

Does obtaining a degree from **UC Berkeley** lead to better outcomes in the long run?

Dataset

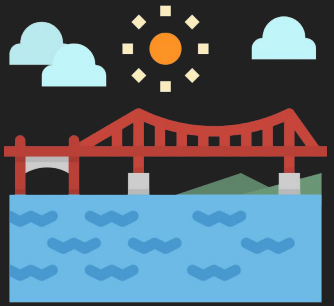


Third Party Sources

LinkedIn, Indeed, etc.

Anonymous

Identification Removed



Bay Area

Current & Previous

Data Schema

ID: unique ID character

Birth year: where available, Default: 2001

Gender flag: 1=female, 2=male, 0=unknown

Skillset1: primary skillset deduced from self-reported skills

Skillset1 weight: the extent to which primary skillset is representative of overall self-reported skills

Skillset2: secondary skillset deduced from self-reported skills

Skillset2 weight: the extent to which secondary skillset is representative of overall self-reported skills

Education: (highest degree attained): 4=bachelor's; 5=master's other than MBA; 6=MBA; 7=doctorate (PhD/JD/MD)

Elite Institution: True=education record from an elite (top 100) institution; False=not education record from a non-elite institution

Start date: start date of the edu/employment record

Flag: Is the start month valid?

End date: end date of the edu/employment record

Flag: Is the end month valid?

Length: length of time spent at the given job (days)

Role: self-reported job role, followed by the normalized job title

Department: self-reported job role, followed by associated department

Company name: Company name if employment record,

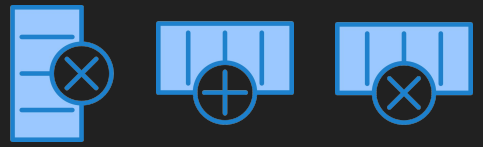
School name if education record

Industry (NAICS): Industry Serial

Education flag: True = education record; False = employment record

Degree type: 0=unknown; 1=high school; 2=vocational degree; 3=associate's; 4=bachelor's; 5=master's other than MBA; 6=MBA; 7=doctorate (PhD/JD/MD);

Filtered Data:



- berk_edu:

  table with rows of all education levels for people who obtained at least one degree from UC Berkeley

- non_berk_edu:

  table with rows of edu levels for people who did NOT obtain any degree from UC Berkeley

- tech_berk:

  table of all jobs that Berkeley graduates got in the tech field

- tech_non_berk:

  table of all jobs in tech industry for non-Berkeley-graduates

Approaches

Promotions

Measure how Berkeley graduates do versus graduates from other schools by the number of promotions they get in their career

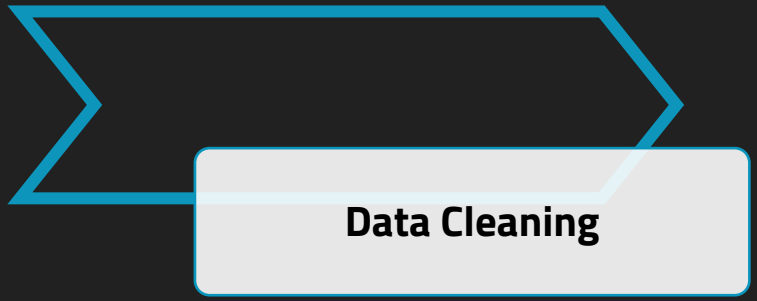
Nearest Neighbor Matching

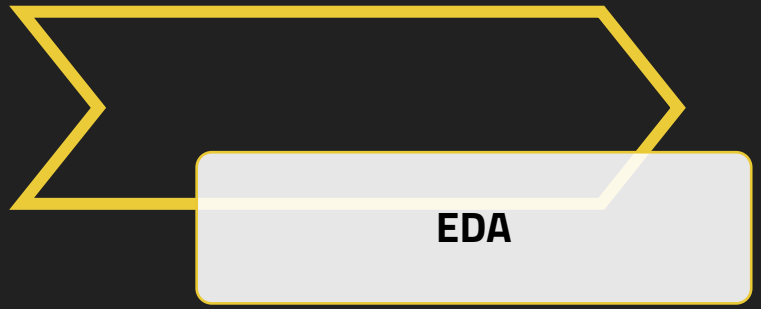
Match a Berkeley-graduate Bay Area worker in tech with the most similar non-Berkeley graduate Bay Area worker in tech

Prediction

Predict the time between graduation and first job in tech based on degree type and skillset weight.

First Steps





Noisy Data + Cleaning Features

Sometimes good features that we want to do analysis on are noisy

Ex. Didn't want education data when looking at jobs, filtered job titles that had keywords related to tech

Different Subsets of Data

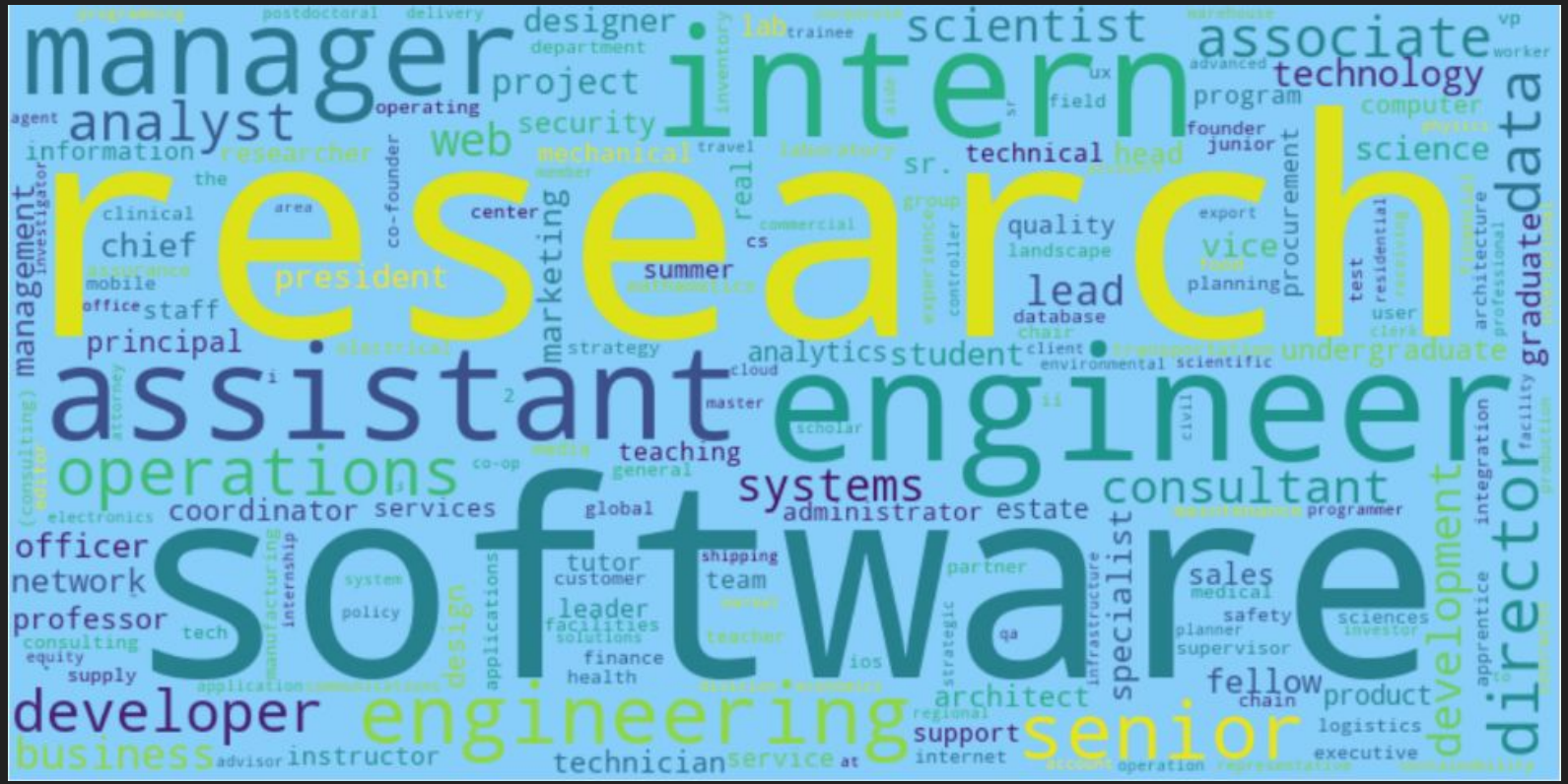
Berkeley grads vs Non-Berkeley grads (ex. Grads from Stanford, UC Davis, etc.)

Male Workers vs Female Workers

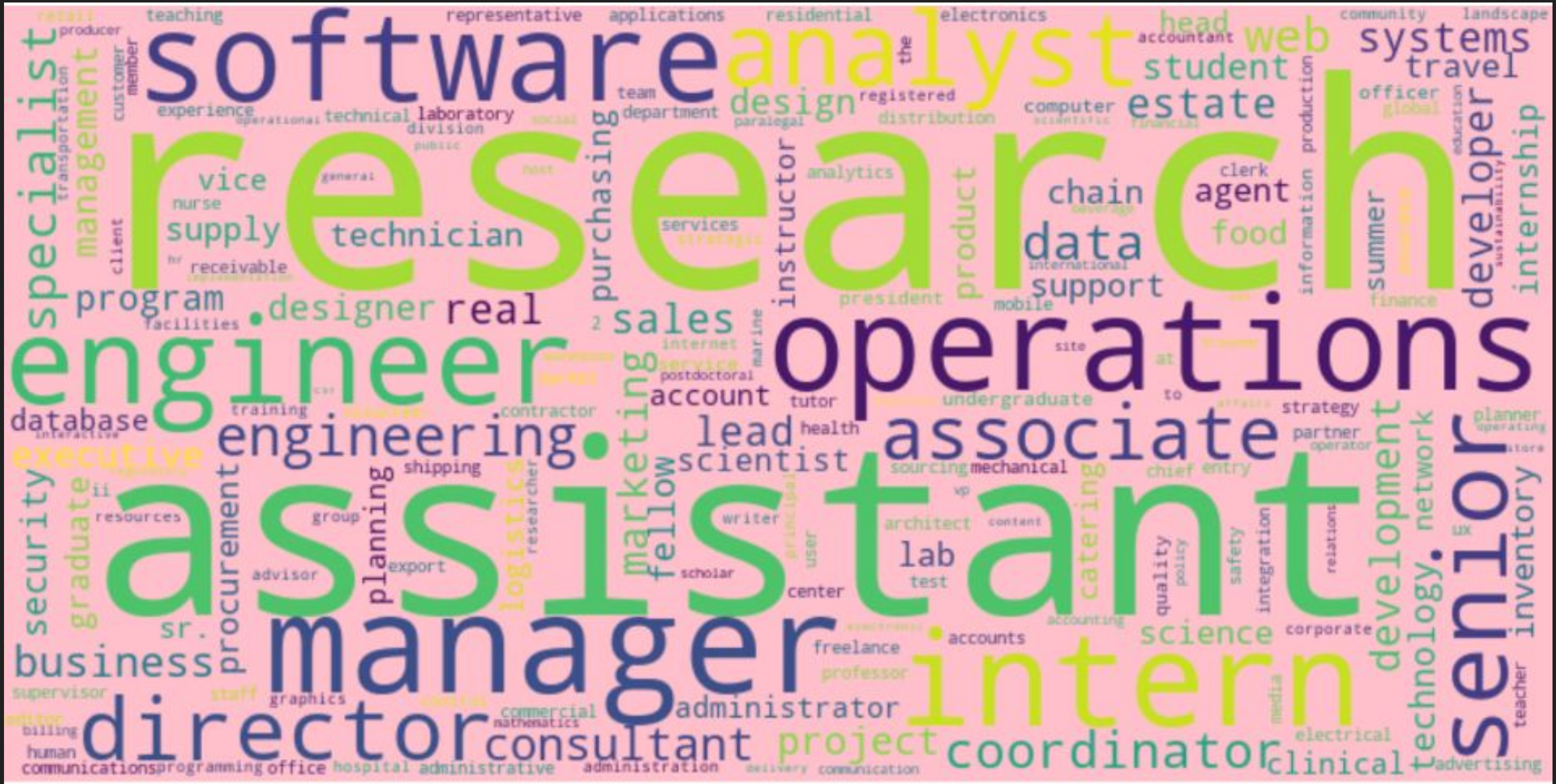
Different Degree Types

Word Cloud for Berkeley-Grad Females in Tech

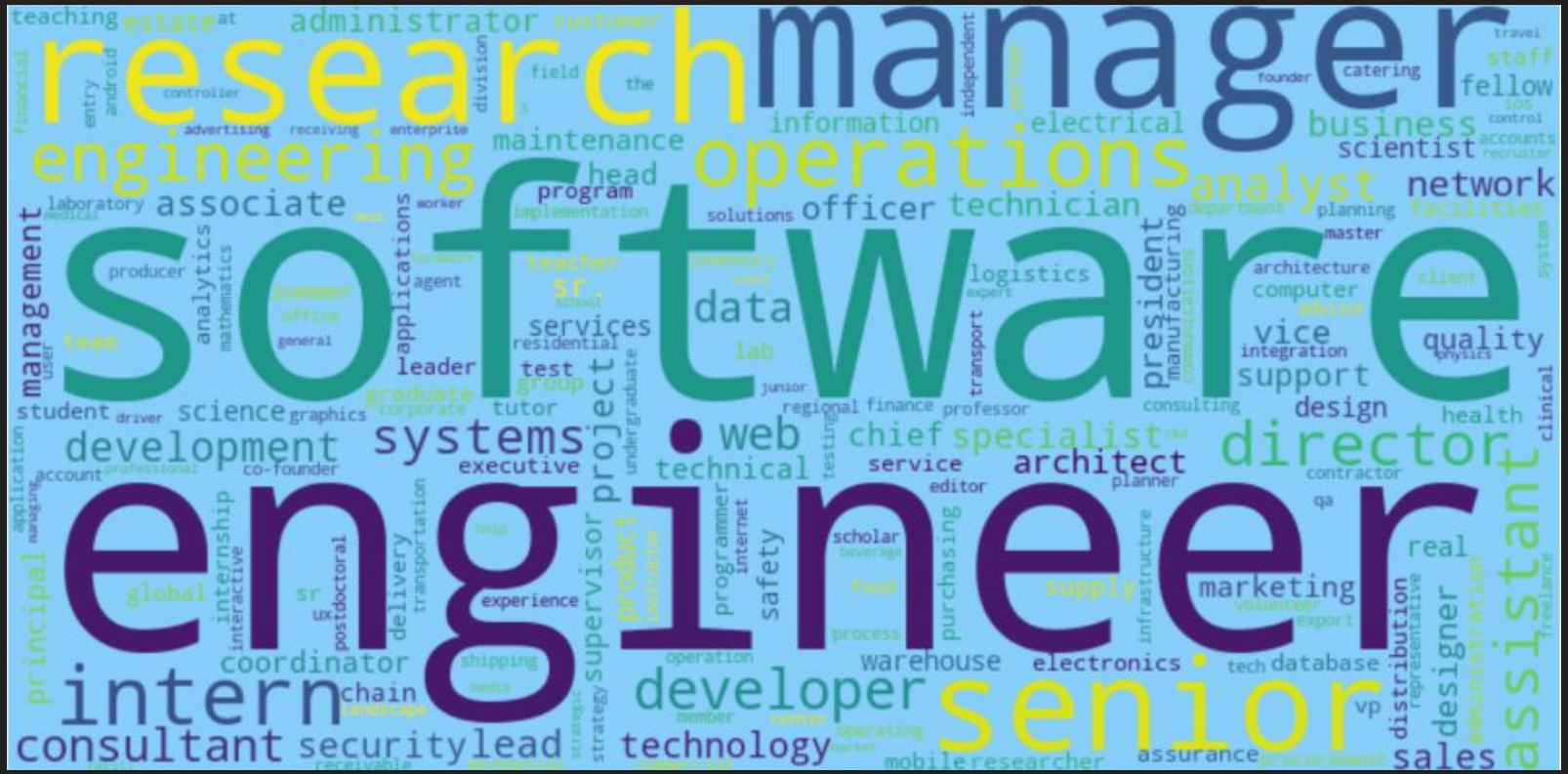
Word Cloud for Berkeley-Grad Males in Tech



Word Cloud for Non-Berkeley-Grad Females in Tech



Word Cloud for Non-Berkeley-Grad Males in Tech



Main Takeaway from Word Clouds

Male

More emphasis on software roles

Female

More emphasis on assistant roles

Berkeley

- Research stands out more
- More Internship positions

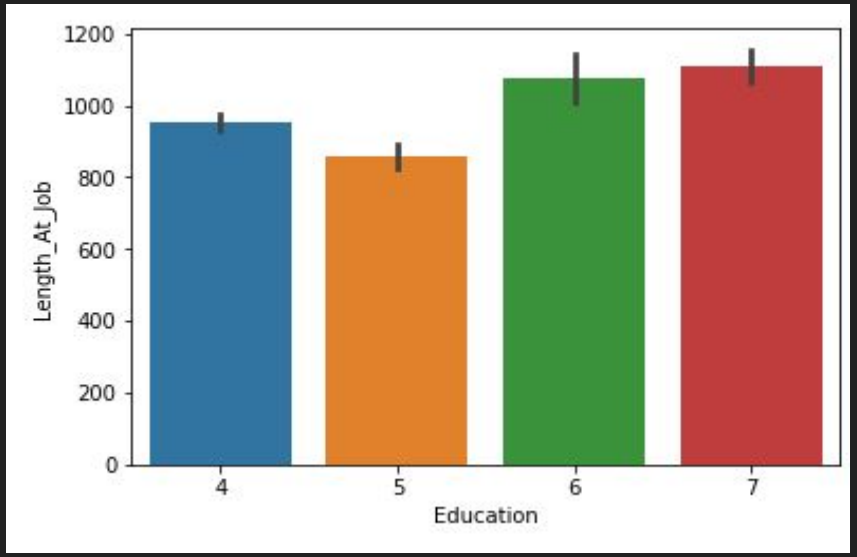
Non-Berkeley

More manager positions

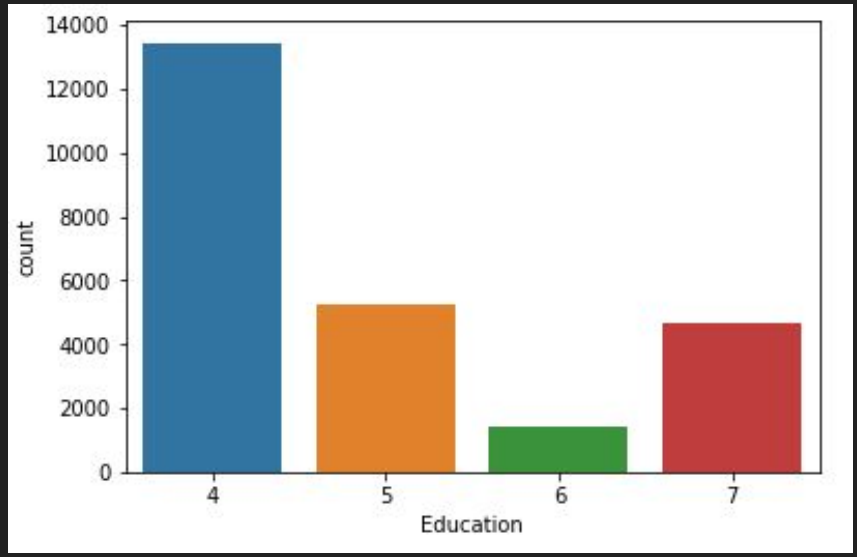
Disclaimer: Different lengths of each in the datasets:

- Berkeley females: 3345 records
- Berkeley males: 5762 records
- Non-Berkeley females: 2466 records
- Non-Berkeley males: 5622 records

Education Analysis for Tech



Relative average length at tech job* for highest education achieved

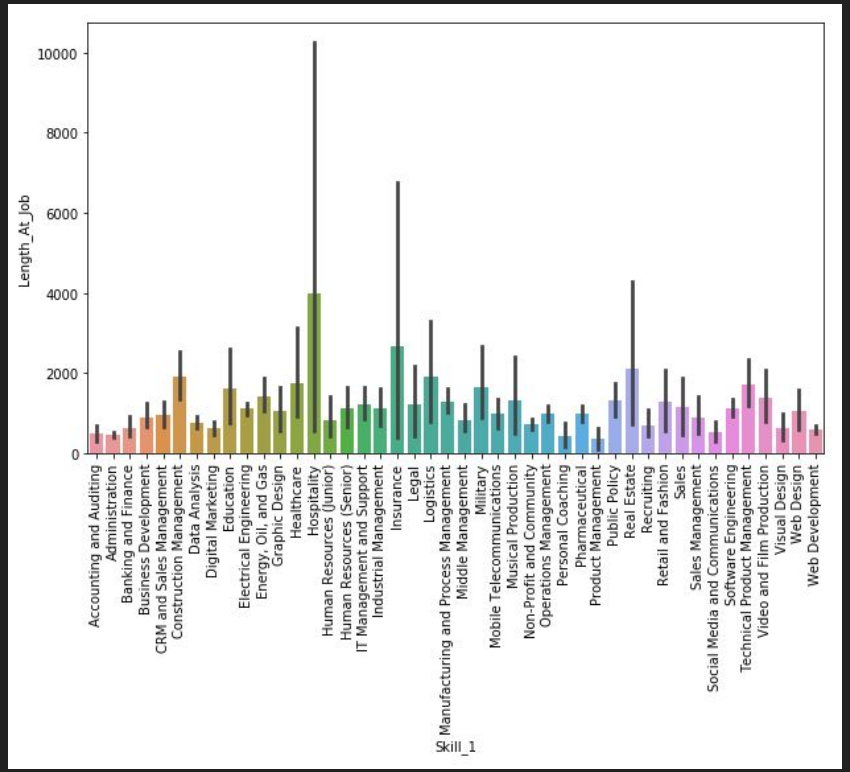


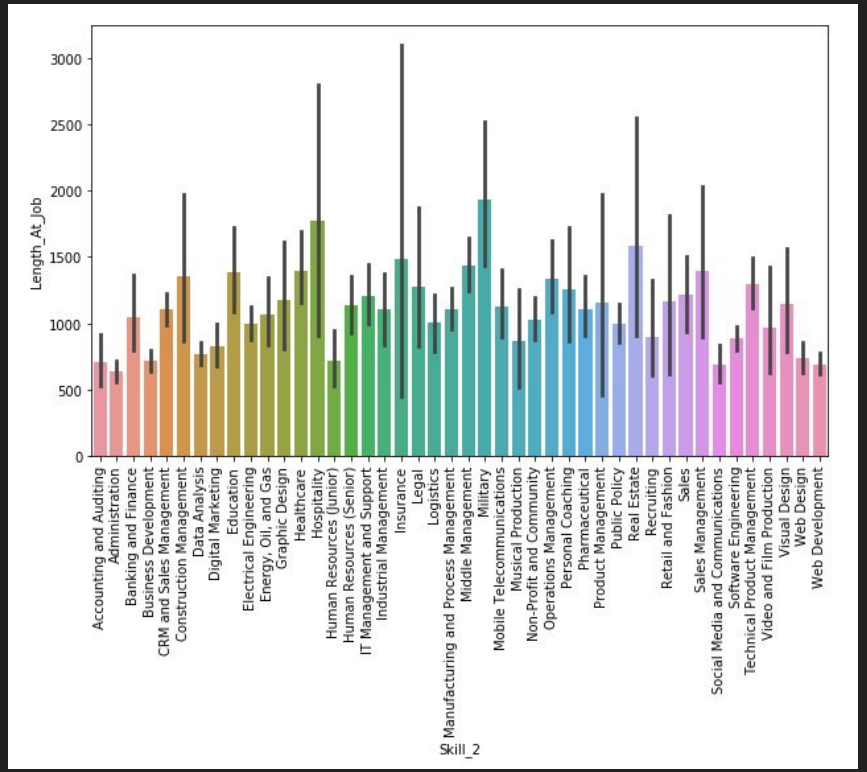
Count of highest education achieved

4: bachelor 5: masters 6: mba 7: doctorate

* Jobs in tech are defined through certain filters on department, industry, and job title

Skills vs. Job Length for People in Tech



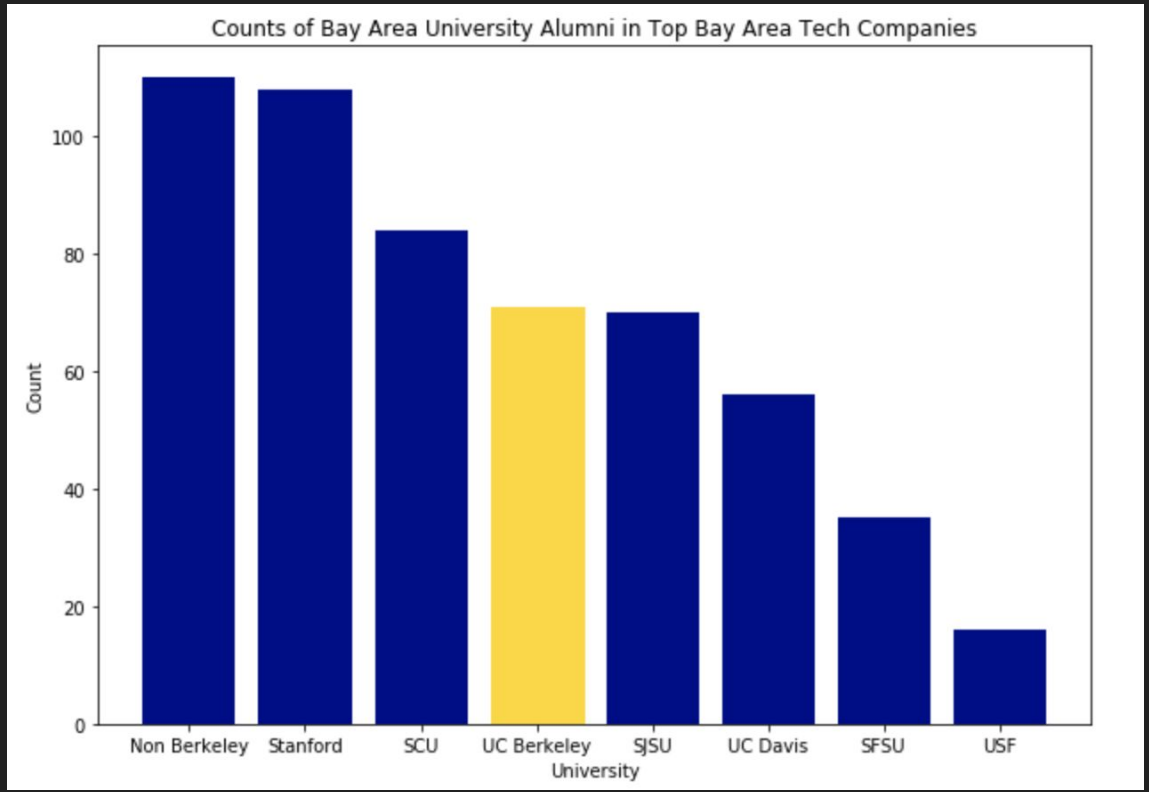


Procedures for Nearest Neighbor Matching Model

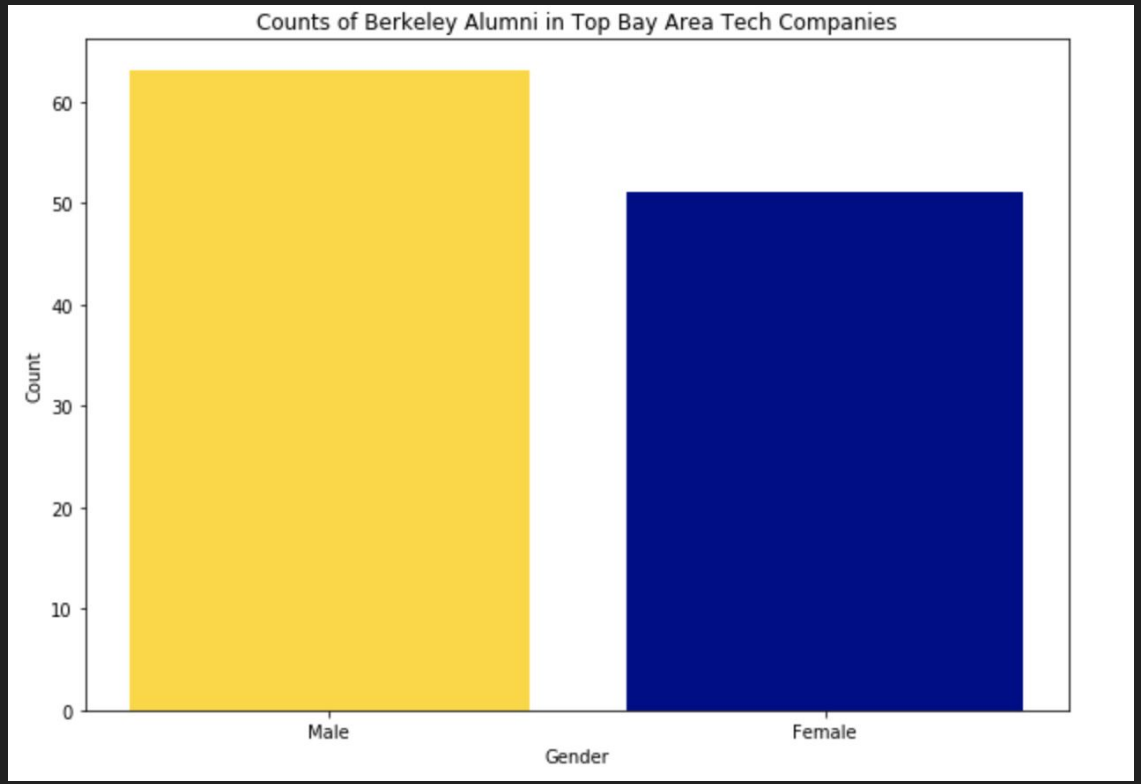
1) For each person in the Berkeley Tech dataset, we “scored” everyone in the Non-Berkeley Tech dataset based on their similarities based on:
    a) Skills and Skill Weights *
    b) Degree Type (Bachelors vs. MBA vs. PhD) *
    c) Elite University
    d) Major
2) Each Berkeley-Tech person was matched with the most similar Non-Berkeley Tech person, using samples of each dataset.

* These features were weighted more heavily in determining scores.

Using the Nearest Neighbor Approach



Counts of Berkeley Male & Female Alumni in Top Tech Companies



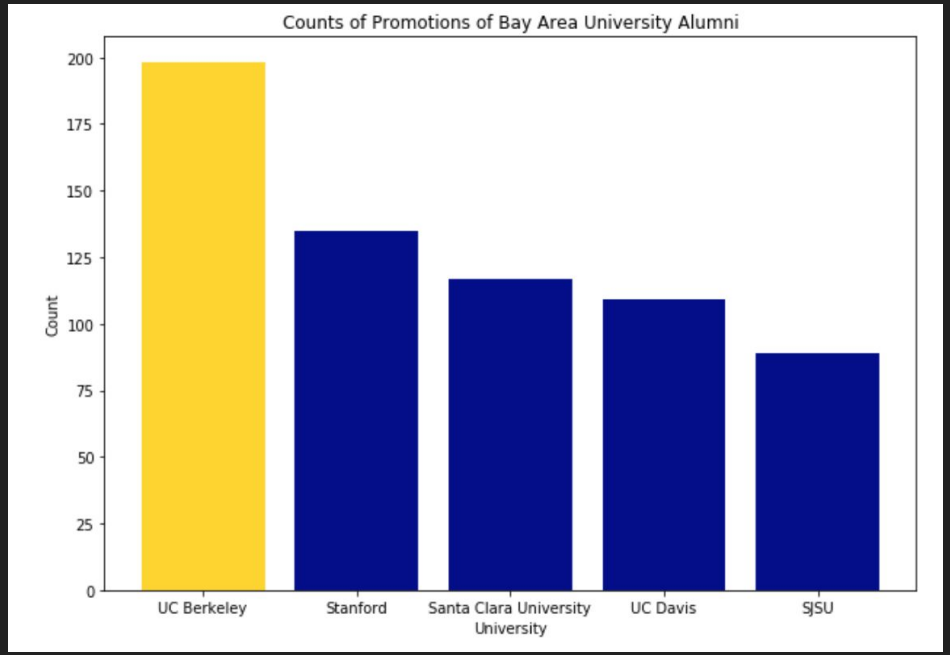
Process for Determining Promotions

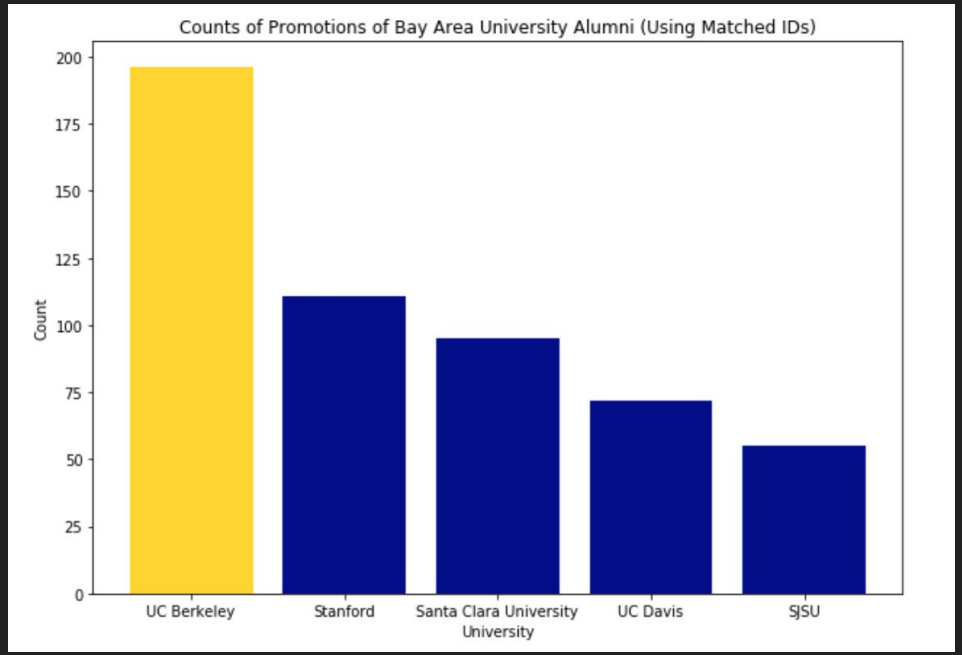
Using the Berkeley Tech dataset and generating tech datasets for other schools:

1. For each unique ID, count the total number of companies one has worked for
2. Count the number of unique companies within total number of companies
3. Subtract number of unique companies from the total number of companies to get number of promotions

*Promotion = if someone has the same company listed multiple times in their job history, we assume they moved positions in an upward direction

Non-Matched vs. Matched Promotions





Procedures for Predicting Time Between Last Education and First Tech Job (Post Grad)

1) Test, Validation, and Train Split on the Data
2) Separate between Berkeley and Non-Berkeley students into two tables: an education table and a job experience table (total four subtables).
3) Extract the end date of the last education record and the start date of the first job in tech* after the last date of education in order to compute the duration of commiting to a job after graduating.
4) Build a linear regression model to predict how long it will take to find a job given what type of degree you are obtaining and the self-identifying skills weights

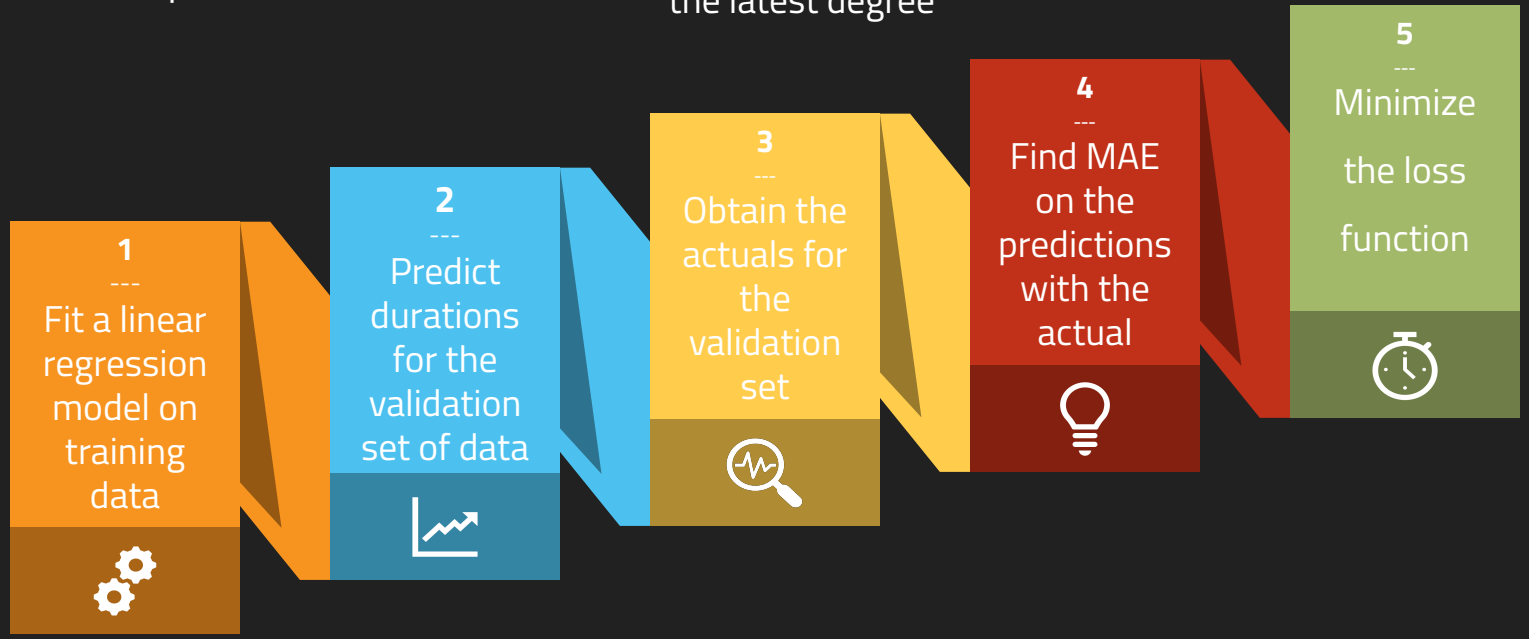
Architecture of Predicting Duration

Linear Regression Model

X: Degree Level

Self-reported Skills

Y: Time taken before finding the first job and completing the latest degree



Comparing Berkeley vs. Non-Berkeley duration between last education and first tech job

```
np.mean(final_berkeley['Duration'])
```

```
1752.0987377279102
```

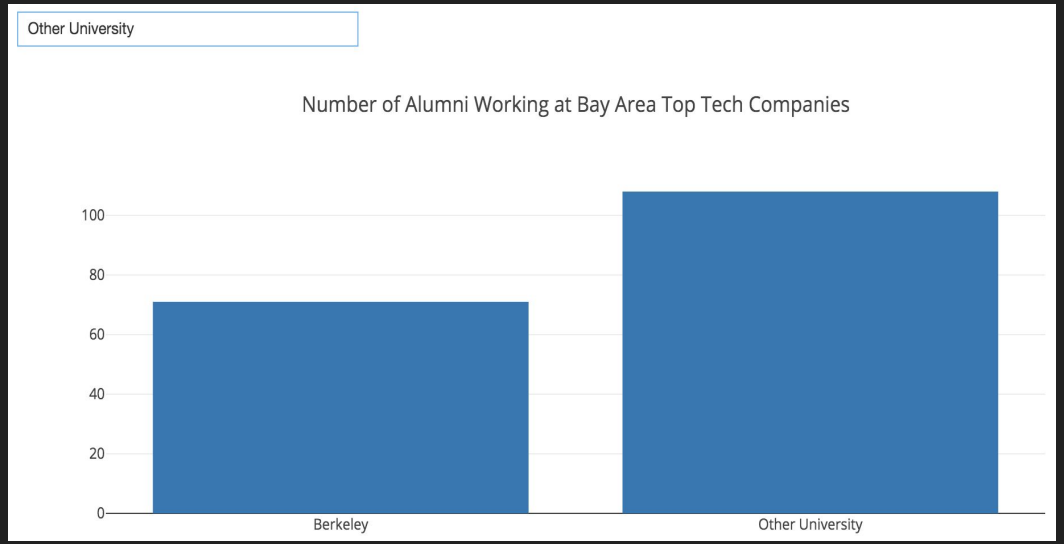
```
np.mean(final_non_berkeley['Duration'])
```

```
1850.1097744360902
```

Taking the averages of our calculations for duration, we found that on average, people with a degree from Berkeley find jobs in relevant fields within the Bay Area faster than those not graduating from Berkeley.

Intended User Interface

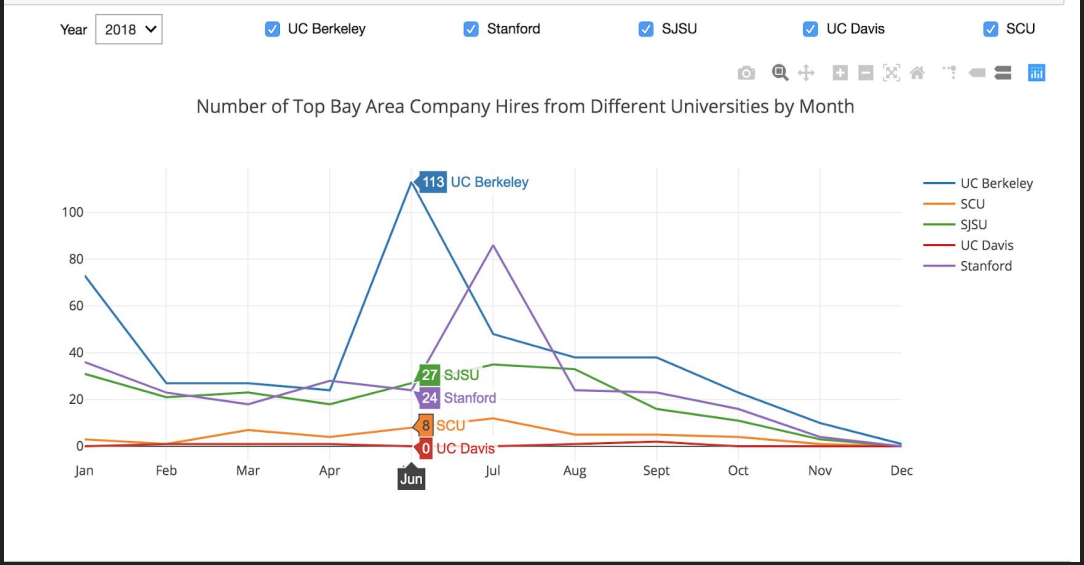
JUPYTER WIDGETS USER INTERFACE



- Load in the manipulated data from CSVs.
- Interactive widgets for the User to see the different visualizations for features.

Intended User Interface

JUPYTER WIDGETS USER INTERFACE



- User able to select different Universities and Year to compare statistics on

Learning Path

Real World Data

Can be messy
Needs lots of cleaning
In-depth understanding of what columns represent/how each can be beneficial for analysis

Logic of Implementation

Closely working with our mentor

Constantly improving our logic

Having multiple iterations of our code

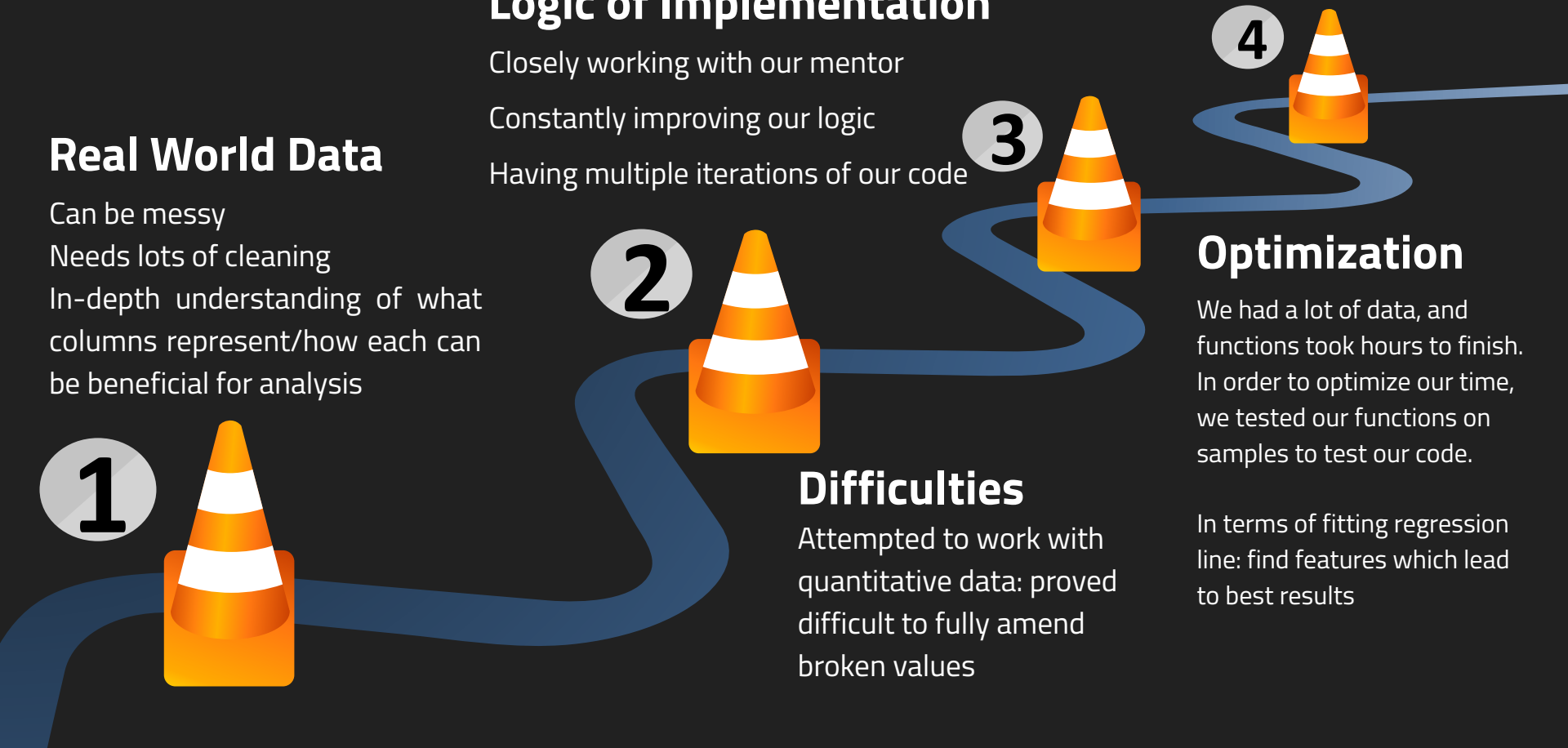
Difficulties

Attempted to work with quantitative data: proved difficult to fully amend broken values

Optimization

We had a lot of data, and functions took hours to finish. In order to optimize our time, we tested our functions on samples to test our code.

In terms of fitting regression line: find features which lead to best results



Potential Areas to Improve On

- Additional factors to take into account:
  - *Matching algorithm doesn't consider age
  - Predicting Duration: First tech job doesn't occur until ~1972
  - *Promotions: Assuming multiple jobs at same company = promotion (for a single person)
  - *User Input: Hard to consider multiple names of single university
  - Need to incorporate more quantitative features in Regression Model
  - *Skills are self-reported
  - Do specific analysis comparing UC Berkeley and another Bay Area University for duration analysis

Does obtaining a degree from UC Berkeley lead to better outcomes in the long run?

Conclusion

It seems that it isn't necessarily easier or harder for UC Berkeley grads to get into top companies.

Berkeley grads find jobs in the Bay Area 6% faster than other graduates.

Berkeley grads achieve more promotions within their companies!

Thank You!

Link to GitHub repo : https://github.com/rprahlad1/ai4good