

# **Initial Exploration Document**

## **Group Members:**

**Lih Yun Chang**

**Rohit Govindan**

**Ziyu Li**

**Ran Meng**

**Lu Wang**

## **1. Principal Questions**

1.1 To what extent are company layoff events driven by a need to change or redistribute worker skill sets?

1.2 Do company layoffs affect people who possess a specific skill-set more so than others? If so, does having a wider range of skill sets lower your chances of being laid off?

1.3 Analyze timelines of different Bay Area companies belonging to different industries, find whether there is a skillset that is more immune to lay-off events than others across industries.

1.4 Lifetime of a specific skill: Analyze past data and predict when a particular skill set will become obsolete.

## **2. Literature Review**

2.1 <http://reports.weforum.org/future-of-jobs-2016/skills-stability/>

Title: The 2016 Future of Jobs Report, Chapter 1, Skills Stability

Organisation: World Economic Forum

Summary: This 2016 report explores the shortening of shelf-life of employees' skillset. It studies the drivers of this change, from demographic and socio-economic to technological. It also gives a breakdown of skill stability in various industries such as infocomm and finance. It also has a mapping of different skill sets to their relevance in various industries.

2.2 [http://www3.weforum.org/docs/WEF\\_Future\\_of\\_Jobs\\_2018.pdf](http://www3.weforum.org/docs/WEF_Future_of_Jobs_2018.pdf)

Title: The 2018 Future of Jobs Report

Organisation: World Economic Forum

Summary: This 2018 report re-emphasizes the growing skills instability, and predicts and "average shift of 42% in required workforce skills over the 2018–2022 period." It also gives a breakdown of the projected take-up of technologies (and hence required skillsets) by 2022. It also predicts that redundant roles will decrease from 31% in 2018 to 21% in 2022, and gives examples of what roles can be considered redundant. It gives a breakdown both by industry and by country on the technology adoption, and the increasing/decreasing roles.

2.3 <https://www.bls.gov/news.release/tenure.nr0.htm>

Title: Employee Tenure Summary

Organisation: Bureau of Labor Statistics

Summary: This 2018 report summarizes the average tenure of employees based on gender, age etc. Of particular interest is the information of the breakdown based on occupation types, e.g. engineering versus service industry.

2.4 [https://assets.pewresearch.org/wp-content/uploads/sites/3/2016/10/ST\\_2016.10.06\\_Future-of-Work\\_FINAL4.pdf](https://assets.pewresearch.org/wp-content/uploads/sites/3/2016/10/ST_2016.10.06_Future-of-Work_FINAL4.pdf)

Title: The State of American Jobs

Organisation: PEW Research Center

Summary: This 2016 report gives findings from a survey regarding the outlook of jobs in USA. In particular it captures the surveyees sentiments towards skills and training. For example, “(54%) of adults who are currently in the labor force say that it will be essential for them to get training and develop news skills throughout their work life in order to keep up with changes in the workplace.” The report also gives a breakdown of these sentiments by industries.

2.5 [www.hays-index.com/wp-content/uploads/2017/09/Hays-GSI-2017-Report.pdf](http://www.hays-index.com/wp-content/uploads/2017/09/Hays-GSI-2017-Report.pdf)

Title: Skills and Demand and Tomorrow’s Workforce

Organisation: Hays

Summary: This 2017 report, Hays assessed USA’s key skills in demand are:

- Software Developers
- Cyber security professionals
- Construction Estimators
- Clinical Research Associates
- Big Data Professionals

2.6 <https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Future%20of%20Organizations/Skill%20shift%20Automation%20and%20the%20future%20of%20the%20workforce/MGI-Skill-Shift-Automation-and-future-of-the-workforce-May-2018.ashx>

Title: Skill Shift: Automation and the future of the workforce

Organization: Mckinsey Global Institute

Summary: This paper outlines the issues of workplace skill sets mismatch. Many companies across the world complain that it is difficult to find the candidates that are good fit to their positions. On the other hand, employees often report that they are under-qualified or overqualified for their jobs. Clearly skill shortages and mismatch are common problems and they have negative consequences on the economy.

2.7 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.198.4858&rep=rep1&type=pdf>

Title: When do jobs become obsolete, and when does it matter?

Summary: By interviewing on tertiary education graduates researchers conclude that almost a third of skills obtained in tertiary education becomes obsolete within a seven-year time frame. Skill obsolescence is closely related to rapid changes at workplace.

2.8 <https://www.linkedin.com/jobs/blog/linkedin-workforce-report-august-2018-san-francisco-ca>

Title: Linkedin Workforce Report, San Francisco Bay Area.

Organization: LinkedIn.com

Summary: These monthly reports rank the top skill surpluses and shortages in the Bay Area, Skills like Property Management, Negotiation and Educational Administration are consistently on top of the surplus list, while skills like Business Management, Data Science and Oral

Communication consistently feature on the shortages list. This could give us an idea about the skills that will likely have a longer shelf life due to high demand and low supply.

### **3. Data Catalogue**

Note: The datasets below provide information related to layoffs in different regions/companies/industries and are meant to complement the Bay Area Worker Skill set data provided by James.

3.1 Description: Link to datasets containing the number of WARN notices filed by California-based companies, from 2014-2018, along with when the notices were filed, and when the layoff events would take place. (The Worker Adjustment and Retraining Notification Act requires companies to notify their workers of any mass layoffs at least 60 days before the event occurs).

Use: These data sets can be juxtaposed with the “Bay Area Worker Employment” data set to determine if any major layoff events correspond with an event involving a mass migration of skill-sets, or a period of time when worker skill-sets skewed in a particular direction.

[https://www.edd.ca.gov/jobs\\_and\\_training/Layoff\\_Services\\_WARN.htm](https://www.edd.ca.gov/jobs_and_training/Layoff_Services_WARN.htm)

3.2 Description: Link to multiple datasets that give an overview of the number of mass-layoff/extended mass-layoff events occurring in years 2007-2013. Some of the datasets further categorize layoff events based on the sectors in which they occur(Manufacturing, Consulting, Transportation, Finance etc.).

Use: These data sets can be used to determine the general layoff trends in different industries in the US over the 7 year period,

<https://www.bls.gov/opub/ted/mass-layoff-statistics.htm>

3.3 Description: A dataset that gives us the number of mass layoffs occurring each month in California from 2003-2013.

Use: Similar to 2, this dataset gives us an overview of layoff trends specific to California over an 11 year period.

<https://data.bls.gov/cgi-bin/surveymost>

3.4 Description: This dataset from US. Bureau of Labor Statistics summarizes information of Initial Claimants on Unemployment Insurance by regions from 1995- 2012.

Use: We can filter out the data relevant to Bay Area and learn about the general trend of lay-off events within the last 15 years.

<https://www.bls.gov/mls/cntyicmain.htm#>

3.5 Description: A dataset on what candidates Google would like to hire. Features include but not limited to *location*, *job categories* and *responsibilities*.

Use: Since we're primarily analyzing layoff statistics in Bay Area companies, we can use this dataset to determine the types of skill-sets/jobs that are in-demand in Silicon Valley.

<https://www.kaggle.com/niyamatalmass/google-job-skills>

3.6 Description: Link to data-sets providing number of people employed in different industries in California in each month from 2000-2018(Seasonally Adjusted).

Use: We can filter out data relevant to the Bay Area, and determine if a significant increase/decrease in employment corresponds with major lay-off events, or significant changes in employee skill-set patterns.

<https://www.labormarketinfo.edd.ca.gov/data/employment-by-industry.html>

3.7 Description: Link to data-sets highlighting top 10 occupations with the most job-ads in different districts in California. The data set also provides the top 10 employers with the most job-ads, cities with the most job ads, and the Number of job ads vs number of people unemployed.

Use: We can find data specific to the Bay Area, which will help us determine the most “in-demand” occupations (which, in theory, would be less likely to be subjected to layoffs).

<https://www.labormarketinfo.edd.ca.gov/data/employment-by-industry.html>

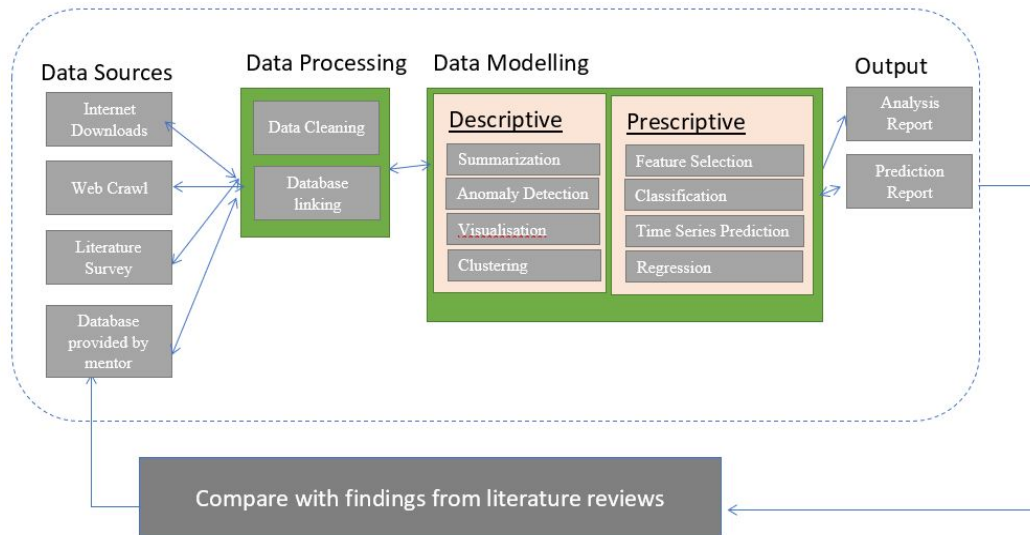
#### **4. Data Cleaning, Linking, and Modeling Methodologies**

As with all data analysis work, we first have to clean the data before we perform further work on it. The cleaning of the data has to take into consideration our principal questions. First, for ease of data management, we can drop the data features that are deemed as redundant for the question. We may also need to further process certain fields in the current data to extract the relevant information out. For example, we may need to study the skill-set information that was available in the database, and assess if we need to further aggregate these skill-sets into specific baskets. Care also has to be taken to handle missing data. Depending on the proportion of missing data, we can choose to either drop the missing records, or perform deductive guesses to predict the missing values. We will also have to check the data for coherence, e.g. capitalization, spelling errors.

As we would be using data from multiple data sources (e.g. Bay Area employment records, and labor records), we would like to link these databases. These can be performed determining a key/index and combining the database. A reasonable key/index in this case would be the person in question.

There are many modelling methodologies that we can explore, and these can be broadly categorized as descriptive (i.e. finding patterns in data) and predictive (i.e. predicting outcomes). There are many models for each of these categories, e.g clustering or linear regression. We will have to assess which models are suitable for the problem statement.

A high-level data flow is as shown on the next page:



## 5. Structure of the Final Product

The main goal of our project is to understand the relation between lay-off events and employees' skill set. Since layoff-events have a huge impact on both individuals and society they live in, we hope to provide strong evidence for individuals to develop their skill set, companies to optimize employment structure and government to make employment policy by exploring how employees' skill set affect the laid-off decision.

In order to ensure our project is easily accessed and can be further studied by others, we will indicate the datasets used in our study and post our code on public Git repositories. Also, we will point out the potential direction of future study. We will mainly make our findings accessible to the unemployment and students to help them improve their professional competitiveness.

## 6. Problem Space and Beneficiaries

By Challenger's tally, U.S. companies have cut nearly 495,000 jobs in 2018, which is 28 percent higher than the number of job cuts in 2017. Technology revolution, political activities, and financial crises can all have significant impacts on the labor market. In 1993, IBM's core mainframe business was becoming obsolete with the emergence of the personal computer and the client server, and 60,000 positions were cut. In 2001, the 9/11 terrorist attacks slammed air travel and prompted the job cut of 31,000 of Boeing. In 2008, as the credit crisis took hold of Wall Street, Citigroup announced 50,000 job cuts. Any big event can be a sign of a layoff, and even the resignation of a senior manager can lead to a large-scale layoff in a company.

Companies can cut positions to gain more competitiveness even when all the company's divisions are profitable and growing, or in most cases, they have to layoff employees to survive from a crisis. However, the layoff domino effect will eventually slower economy as well as the

development of the company. Therefore, we did this research and wish to share our insights with all sectors of the society to help decision makers improve the employment structure and make recruitment process more efficient so that layoff can be reduced.

There are three main target audience that we wish to share our deliverables with – companies, educational institutions, and governments:

- Companies

Our deliverables can differentiate the vulnerability of different skill sets to layoffs. In addition, our prediction on the lifetime of a skillset will provide companies with reference and help companies produce strategic recruiting processes. For example, if we expect python is a more promising programming language and can be actively used in workplace for more years than C, then companies would consider adding python instead of C as a job requirement.

- Educational Institutions

Educational institutions prepare people for their future careers, and it is important for education to fill the gap between theoretical knowledge and practical skills required by workplace. Our deliverables address the skill sets that are important in labor market. If we could share with educational institutions our insights, more schools would invest in teaching those skills, and organizations would have a larger amount of qualified candidates.

- Government

“[Science] is an approach to the world, a critical way to understand and explore and engage with the world, and then have the capacity to change that world.” Obama sees opportunities in STEM majors and believes that knowledge in STEM has the power to make change to the world, so in 2010, Obama helped launch Change the Equation, a new non-profit with full-time staff dedicated to mobilizing the business community to improve the quality of STEM education in the United States. With this political decision, STEM resources have been stored in schools and the talent pool has been increased in size. Therefore, government is the most effective and efficient channel to share insights with all sectors of the society. If our insights can be shared with the government and effectively help reduce layoff, it will largely save the unemployment cost, and eventually, tax rate can be lowered so that all taxpayers will benefit from the saved tax payments.