

Deliverable 0

Group 1 of BCNC

Group Member: Chenyu Cao, Jiawen Chu, Johanne Antoine, Fangxu Zhou, Zhechao Chen

I. Description of project

There is a growing large number of Asians that live in Boston, but yet the research towards them is limited and insufficient. The motivation of this project is to understand more about the living conditions of Asians in America. This project will mainly focus on the employment of the Asian community in Massachusetts. We think this could largely reflect how Asian communities fit and develop in the local society.

The research will be based on some relevant datasets including the PPP data which is provided by the client. Through this project we want to understand what kind of job is most appealing to Asians (and how their employment is related with their education). Moreover, the shift of their work and salary in recent years.

II. Goals of our team

- Collect datasets in the scope of employment, retrieve data that is relative to Asians, clear up and combine these data so that they can be more available to the public, for example the nonprofits, organizers, policy makers, etc.
- By analysing the data collected, try to find out, if there is any, relations between the employment and Asian people
- By analysing the data collected, try to find the future trend of employment of Asian people

III. Limitations and potential risks

We have difficulty finding datasets related to employment of Asians in Boston. We did find some, but those were not free of charge. We are for now still struggling on finding proper datasets. We've tried many common sites to locate appropriate datasets that can be used for our project, like "Google Dataset Search", "Kaggle", "Data.gov". Some of them are not public, some of them are not free, some lack attributes that are essential for our research. We may need advice and help to find more datasets that will be useful.

• Limitations:

- (If combined with education: SY19-20 Participation Data (public) is limited with the enrolling number of Asian students in high school, we can not get more specific information and datasets for university students, which can reflect more about the choices of their study direction and interests.)
- The PPP data is more about the loans related to business owners instead of jobs, and we are unable to extract the features(such as job titles, employees' info) that we need.
- Most of the samples in IPUMS' are missing the city attribute, which can help us to locate the samples in Boston.

- Hard to distinguish which type of Asian(e.g. Chinese/Vietnamese/Korean) of the samples because most of the data only include the general race and ethnicity.
- **Potential risks of achieving project goal:**
 - The datasets suggested are unavailable
 - The collection of datasets involves privacy or even legal issues
 - Unable to find sufficient datasets to support the research
 - The leakage of datasets in the process of handling them
 - Fail to preprocess the datasets leads to the extra cost of additional features

IV. Datasets plan to collect

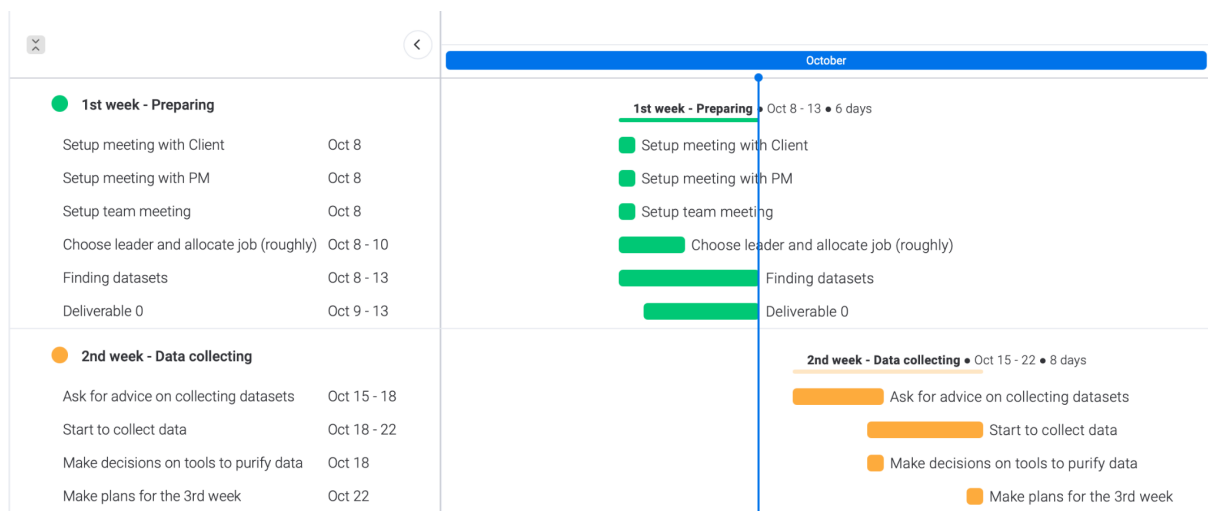
(As have been mentioned, we are still working on finding proper datasets. Those listed below are two of which are usable among the datasets)

- [PPP](#) data
- [IPUMS](#) data with selected attributes

V. Questions plan to answer

- Are there any relations between the employment and education of Asians?(cross out if we should only focus on the employment rather than education-employment)
- What are the popular career choices and the trend within Asians?
- What is the salary distribution of Asians?
- Are there any problems with the living conditions of Asians reflected by these datasets?
- What is the main difference between Asians and Whites in these aspects?
- Which career choices in Asians have the highest earnings?
- Are there differences between different groups of Asians?

VI. Scrum(Using [Monday](#))



1st week - Preparing						Owner	Subte...	Status	Priority	Timeline
1st week - Preparing	Setup meeting with Client							Done	High	Oct 8
	Setup meeting with PM							Done	High	Oct 8
	Setup team meeting							Done	High	Oct 8
	Choose leader and allocate job (roughly)							Done	Medium	Oct 8 - 10
	Finding datasets							Stuck	High	Oct 8 - 13
	Deliverable 0							Done	High	Oct 9 - 13
	Ask for advice on datasets							Future steps	High	-
+ Add										
										Oct 8 - 13
2nd week - Data collecting						Owner	Subte...	Status	Priority	Timeline
2nd week - Data collecting	Ask for advice on collecting datasets							Future steps	High	Oct 15 - 18
	Start to collect data							Future steps	High	Oct 18 - 22
	Make decisions on tools to purify data							Future steps	Medium	Oct 18
	Make plans for the 3rd week							Future steps	Medium	Oct 22
+ Add										
										Oct 15 - 22

VII. Logistics

- Tools we use to communicate and hold meetings: Slack, Zoom
- Tools we use for Kanban and Scrum: Monday (web application)