**Topic Proposal for the 2nd Meeting**

**Duo Zhou**

**Preface**

I have a strong feeling that many groups will try to investigate the Covid-19 case. I personally suggest that we try to tackle some other projects. My topic choice is the analysis of Chicago public school progress reports. I also did a little bit of research on Covid-19.

**Project Design**

**Education**

The Chicago Public Data site has public school progress reports from SY11-12 to SY18-19.

<https://data.cityofchicago.org/browse?q=progress%20report&sortBy=relevance>

We can also find Consensus (income level, ethnicity and many other types) data by zip codes from the following sites

<http://zipatlas.com/us/il/chicago.htm>

These data need scraping using Python3.

We can do a student performance analysis by ethnicity, income and population education level.

**Covid-19**

We can start with investigating the basic data for covid-19 from a geographic perspective, which is the most common way of looking at issue (everyone is doing this, since that is the most common form for covid19 data). I do not think we need to run the data with predictive models.

There are multiple sources that will provide such data

<https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data>

<https://github.com/CSSEGISandData/COVID-19>

But the more interesting question should be how people are affected by Covid-19 in terms of economic background and ethnicity.

Since New York and California are the two most diverse states in terms of economic classes and ethnicity and the only states (As far as I know) published covid-19 data by ethnicity, age groups and income level. We should do an analysis on covid-19 numbers by ethnicity and income levels.

<https://github.com/nychealth/coronavirus-data>

<https://github.com/datadesk/california-coronavirus-data>

We can also scrape additional consensus data from the following site using Python3 to expand our analysis

<http://zipatlas.com/us/>

All of the data above have relatively small size and easy to manipulate, we may get a low sore on the project due to the simplistic nature of our project data.

**Movies**

If you want to challenge yourselves, we can do a movie data analysis, the data sets are very big and requires additional tools to process.

Data Tables:

IMDB

<https://www.imdb.com/interfaces/>

Web Scraping from

Fandango

<https://www.fandango.com/20001_movietimes?mode=general&q=20001>

TMDb

<https://developers.themoviedb.org/3/getting-started/introduction>

We can even further challenge ourselves to by apply the following ML package to do movie recommendations.

<https://data-flair.training/blogs/data-science-r-movie-recommendation/>