**Weekly Update**

Week1: Sept.28, 2020 - Oct.2, 2020

Project: Greenwich

Group: Gannett Peak

Members: Isaac Choi, Matt Ko, Binqi Shen, Congda Xu

This past week, our group mainly focused on the overall preparation for our project and we had the opportunity to invite Mr. Sparrow, the CEO of Greenwich.HR, along with Borchuluun to our first kickoff meeting.

Prior to our meeting, we read through the ‘Database Table Spec’ and pinpointed the primary keys  and  linkages among different tables. Thanks to the well-structured database tables, we were also able to identify several key job-related input vectors (e.g. salary,  role\_primary, duration\_days, etc.) that we can possibly use in our future model. In addition to the job-related metrics, we searched for an online API (<https://www.alphavantage.co/documentation/#dailyadj>) in order to get reliable financial metrics including stock prices, sales, revenues, etc., and we confirmed with Mr. Sparrow that we can use outside API sources.

During our kickoff meeting on Sept. 30, 2020, we were able to raise our questions and concerns to Mr. Sparrow, who kindly provided us informative and essential answers. Some key points discussed are summarized below for future references:

* Initial data (Greenwich data source): job listings in the United States (70%); identified through proprietary technology that scans about 1.5 million webpages and 800,000 organizations; data is typically not clean (intentional)
  + They keep the unclean data because many of Greenwich’s clients want to see it, and Greenwich adds additional fields that has the normalized information
* Trading strategy should be limited to long and short. (No financial derivatives)
* Cut the dataat January 31 2020 to set aside the Covid factor
* Timeline to be tested on: If we are measuring 6 months, we stop 6 months in advance and use that as our test set. (Note: stop on January 31)
* We need to discern which is the leading and lagging indicator for job postings. (Initial hypothesis is that all are leading)
  + (For some companies) company lists an opening that could be **lagging** if the job is posted as a result of an approval that was previously understood.
  + For more companies, it is a **leading** indicator, because the job listing will not result in a hire for 2 to 4 months that anyone outside the company can track. We should be looking at how far in advance they are looking ahead, like 1 and 5 months. We can identify a segment of companies that our job listing metrics (several and with a pay metric), where we get a solid reading a few days to 90 days out, maybe different segments have a 30 day to 60 day and maybe blend it together.
* Portfolio selection (risk versus return): want more insights of companies that have the HIGHEST (85% to 90%) correlation to the inputs and stock performance.
* Future meeting plans:  more frequently in the beginning
  + Mr. Sparrow’s available schedule: 8:15-11:00 Tues., Wed., Thur.
  + Mr. Sparrow suggested us to contact him directly to set up meetings
* Final Deliverables:
  + Trading strategy must be **validated** by faculty.
  + An article: They want to communicate the value of their dataset to their clients.
  + No preference on languages and packages to visualize data and to execute strategies

The initial meeting further strengthened our understanding of the project data and helped us formulate future plans. Next week, we plan to decide on a recurring meeting schedule and make appointments with Mr. Sparrow in advance. We think the major task for next week is to get the data prepared for future modeling. The first thing to do is  to upload the part\_00000 files we downloaded from the database to PostgreSQLserver and seek ways to combine/link tables together. After that, we will do an initial screening of the dataset and try to filter out the data that are suitable for our analysis. For example, we will select the data from the appropriate time range (pre-COVID data), and select the columns that we think are useful features for training model input after our brainstorming. Another task to do is to download the finance data that we found on external sources. If we have time, we might try to connect the finance data with our HR data and do some data cleaning tasks.