Optional- Provide a brief description of any prior research experience.

As described in my cover letter and resume, I had prior experiences on two data science projects along with a computer science one at Grinnell. However, I believe that my current project to explore the NYC Subway GTFS dataset along with some others for my Data Science Education Practicum is akin to fellowship projects done at the Lab. Below is some of my initial research questions based on survey of similar projects done at Boston and Chicago.

- Where and when are subway delays happening? Have such occurrences changed over the last 3 years (within the time frame of available data)?
- Do the delays coincide with high volume of people and/or rush hour? Does severe weather play a role in delays? Can we explore other effects of weather on the subway system?
- Is there any pattern if we overlay social-economic data of neighborhoods and capital spending throughout the years on subway delays and other incidents?

As earnest as the questions sound, there may be limitations based on the available data. We may have cause to request the MTA to release more data that can help us answer questions to both keep our public officials accountable and provide recommendations for improvement where we can.

Optional - Describe your background in programming

As an undergraduate computer science major, I was exposed to a slew of programming languages from C and Java to Scheme. I also learnt SQL with MySQL database. After college, I took upon myself to learn Ruby On Rails. At Columbia University, I picked up Python and now, I would say that I am most familiar and proficient in R and STATA. With R, my advisor and I have set sights on developing a package to handle asynchronous data within the next two semesters.

Optional - Please describe your background (courses taken, research projects, etc.) in Statistics and Mathematics.

By summer of 2017, I would have finished Advanced Modeling (I) with topics in discriminant analysis, multivariate analysis of variance, canonical variate analysis, exploratory factor analysis and multidimensional scaling. Finishing multilevel models, I would be able to handle nested data and its complex correlational structure. Statistical Inference will provide me the skills to carry out model selection over a range of models, knowledge and practice of Bayesian Inference along with substantial simulation programming. Missing Data is going to culminate in applying multiple imputation method. Last but not least, the Applied Statistical Modeling and Inference class would expose me to E-M algorithm, Newton-Raphson and Monte Carlo techniques.