It is no surprise that tech has become the dominant industry in many cities across the world. With this boom comes the potential for startups raise money, grow, and potentially be acquired or go public. Using data from Crunchbase, I want to investigate the likelihood of a company being acquired versus continuing to run independently. In addition, I’d like to predict how the number of acquisitions per year will look in the future. Data for this project is available by requesting basic access to the Crunchbase API and downloading the appropriate csv files. The biggest challenge I will face is missing data. Crunchbase data is crowd sourced, which means companies and people enter information on themselves. At first glance there are over 700,000 companies listed with entries dating back to the year 1900 however this number will decrease significantly once I perform feature engineering and remove null values. Since this will be a binary classification problem and I have labeled data, I will be able to create a supervised an unsupervised model. Based on the performance of each I will be able to determine which performed better at solving this problem. For the time series analysis, I will use a long-short term memory network which is a type or recurrent neural network.

Results of this project could be useful for venture capital firms and entrepreneurs because they would be able to cash out their investments in a startup. Features that could be useful for this project include but are not limited to dollars raised, industry, number of funding rounds, and number of employees. I will break this project into two phases. Phase 1 will include gathering the data and performing exploratory data analysis and Phase 2 will be modeling and reaching a conclusion.