**GROWTH VS. TURNOVER: AN INDUSTRIAL ANALYSIS OF RISK REWARD RATIOS**

**Introduction**

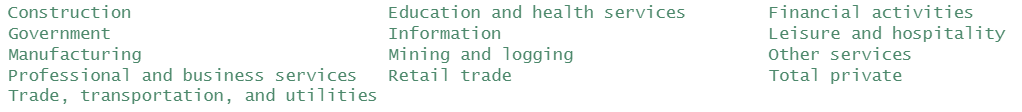
In today’s latest of posts – we examine what it means to be part of a given industry. Which industries are hiring, and which ones are experiencing churn, and how the two are related. As one would expect, just as at the micro level, reward risk ratios are closely related, so too can be said at the macro level.

As someone situated within a company proud of its I/O psychologists, many of them hailing from the GMU department, I’ve had the chance to experience what it means to be part of an organization built by I/O psychologists. As a result, this has me to think in terms of personnel as our greatest resource.

**Data**

Then again, as someone trained as an economist, I cannot help but lean towards looking at labor as an input to production. For today’s analysis, we lean on the BLS data collected in conjunction with the Census Bureau.

After combining multiple datasets, we are left with 3055 observations, on the following 13 industries:

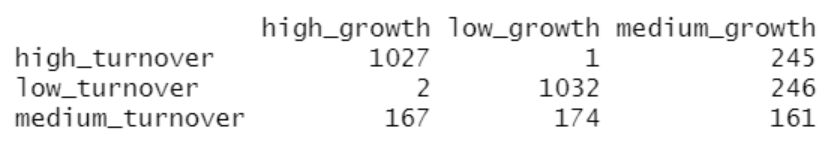


Variables include job openings, hiring, earnings, quits, layoffs, and total separations. The ultimate goal of the analysis is to draw a line between turnover and growth. In order to do so, we code composite variables comprising of the above listed variables. Our first categorical variable is turnover: (quits, layoffs, and total separations) made up of the 3 component variables. The next is growth: (job openings, hiring, earnings), composed also of 3 variables. The unit of analysis is the industry in a given month, for a time period spanning the year 2000 to the present.

The two data sources used are the Job Openings and Labor Turnover Survey (JOLTS) and the Consumer Expenditure Survey (CES), both released monthly by the BLS. Values are all in levels, seasonally adjusted, with the earnings values scaled per 1000 people.

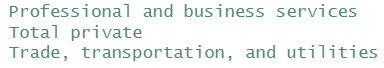
**Summary Stats**

We find from a preliminary examination, that the high growth-high turnover industries are the more communication-based ones. Whereas the low growth-low turnover industries are less communications based. A basic cross-tabulation yields the following results:

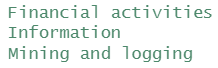


A further drilling down into high growth-high turnover, and low growth-low turnover industries yields the following results:

* High Growth, high turnover industries:

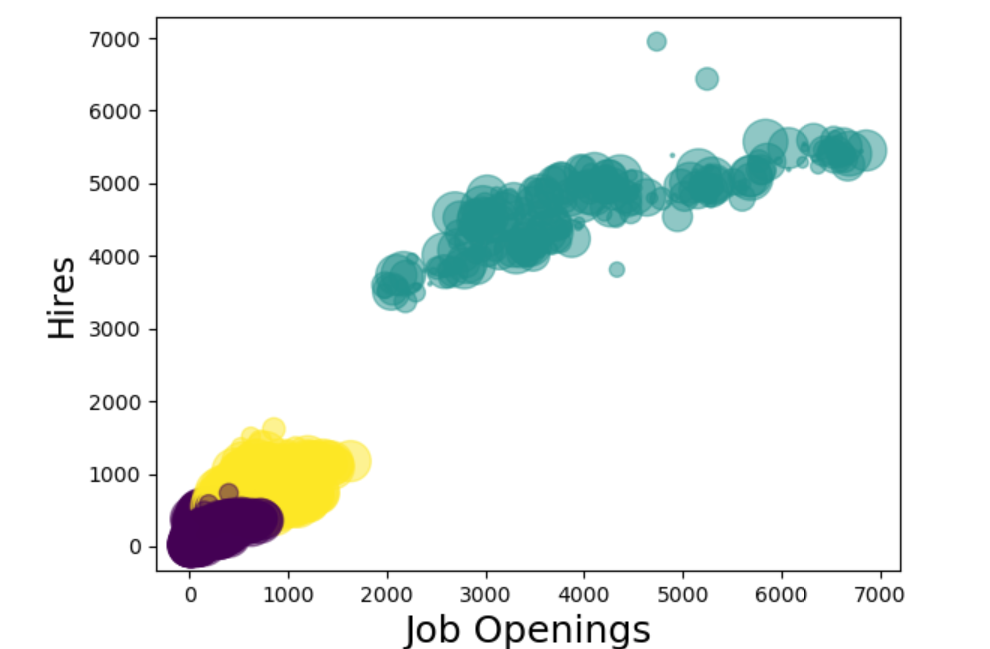


* Low Growth, low turnover industries:

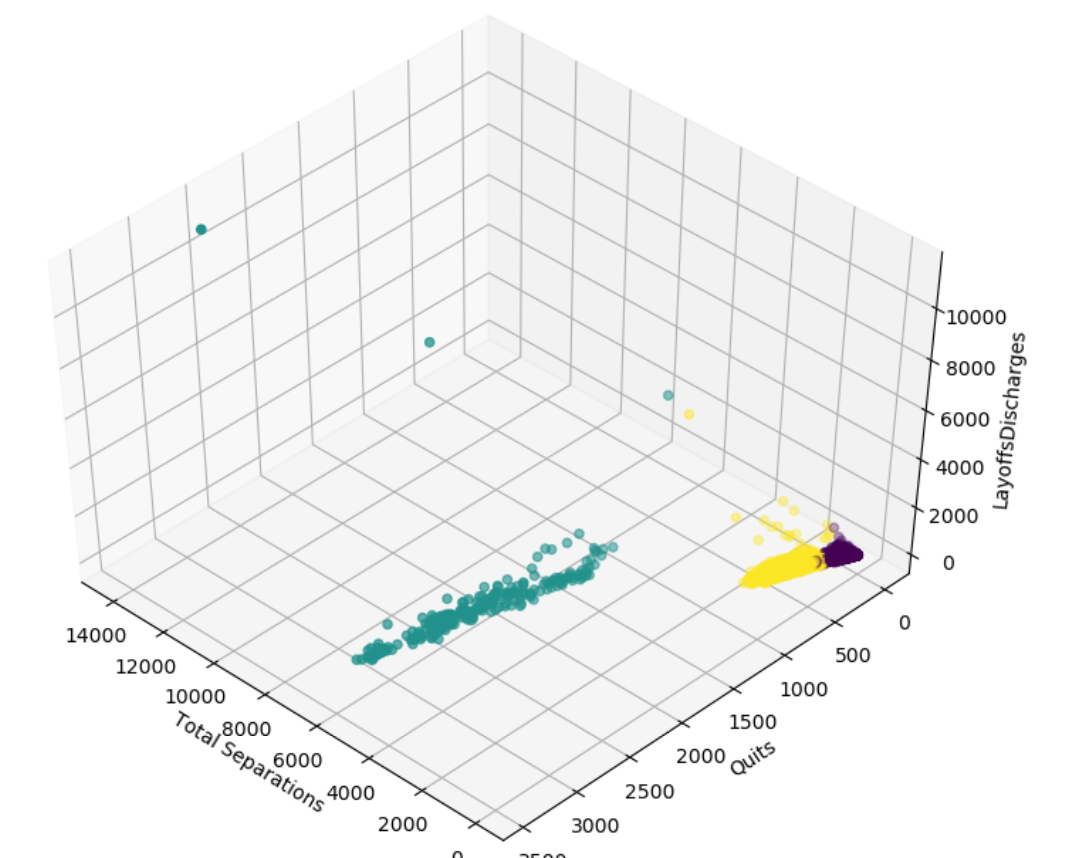


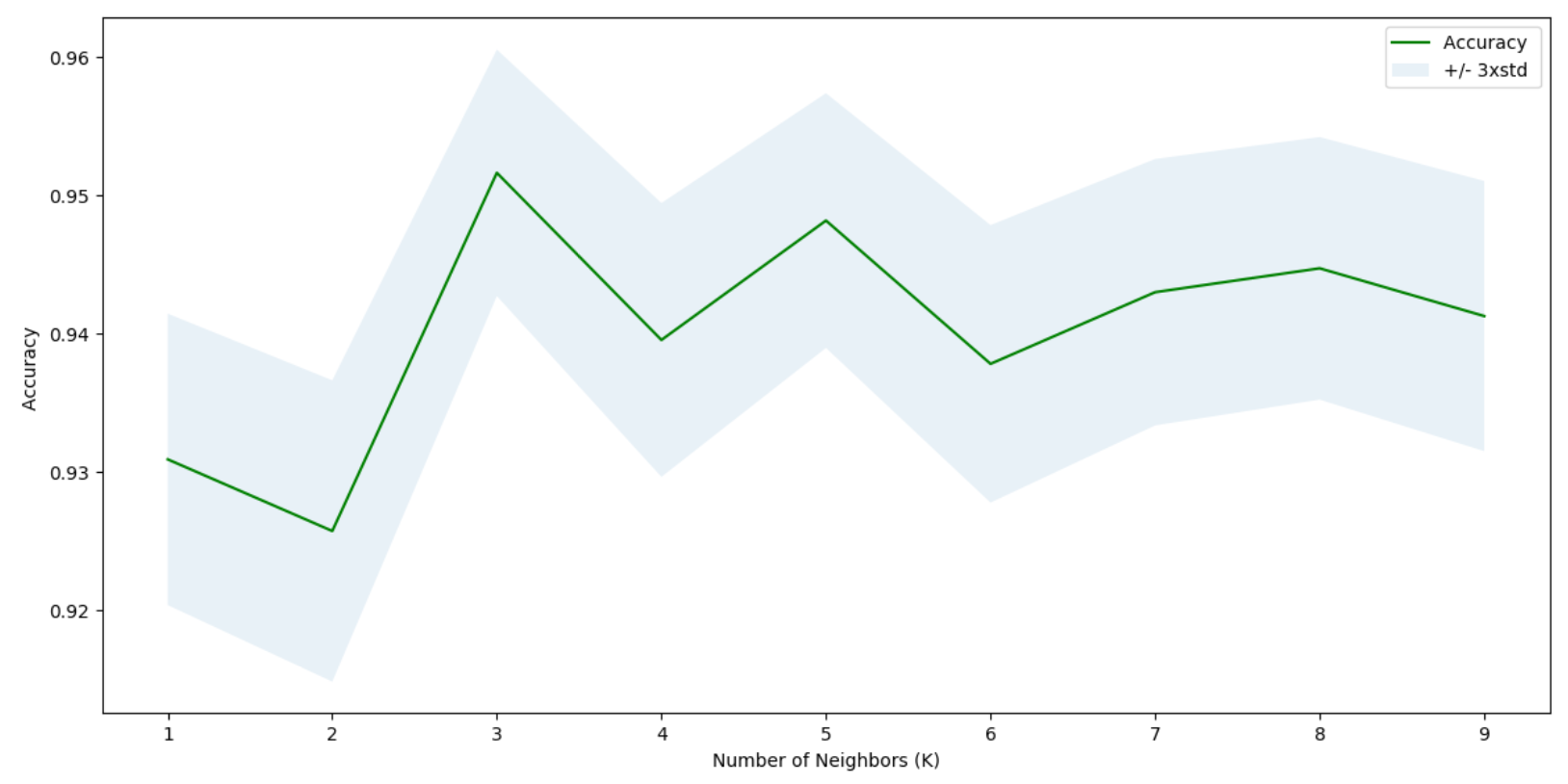
This is not very surprising given that technically based industries typically require higher barriers to entry. Whether or not they are established industries, or growing as a whole, they will not rank as high growth when compared with business services, or trade and transportation.

Interestingly, even mining and logging is ranked among the low growth-low turnover industries, possibly due to all the permits required to establish mining quarries, or logging outposts, and the technical expertise, plus risk associated with being a logger. In essence, a technically based profession amongst the blue-collar professions.



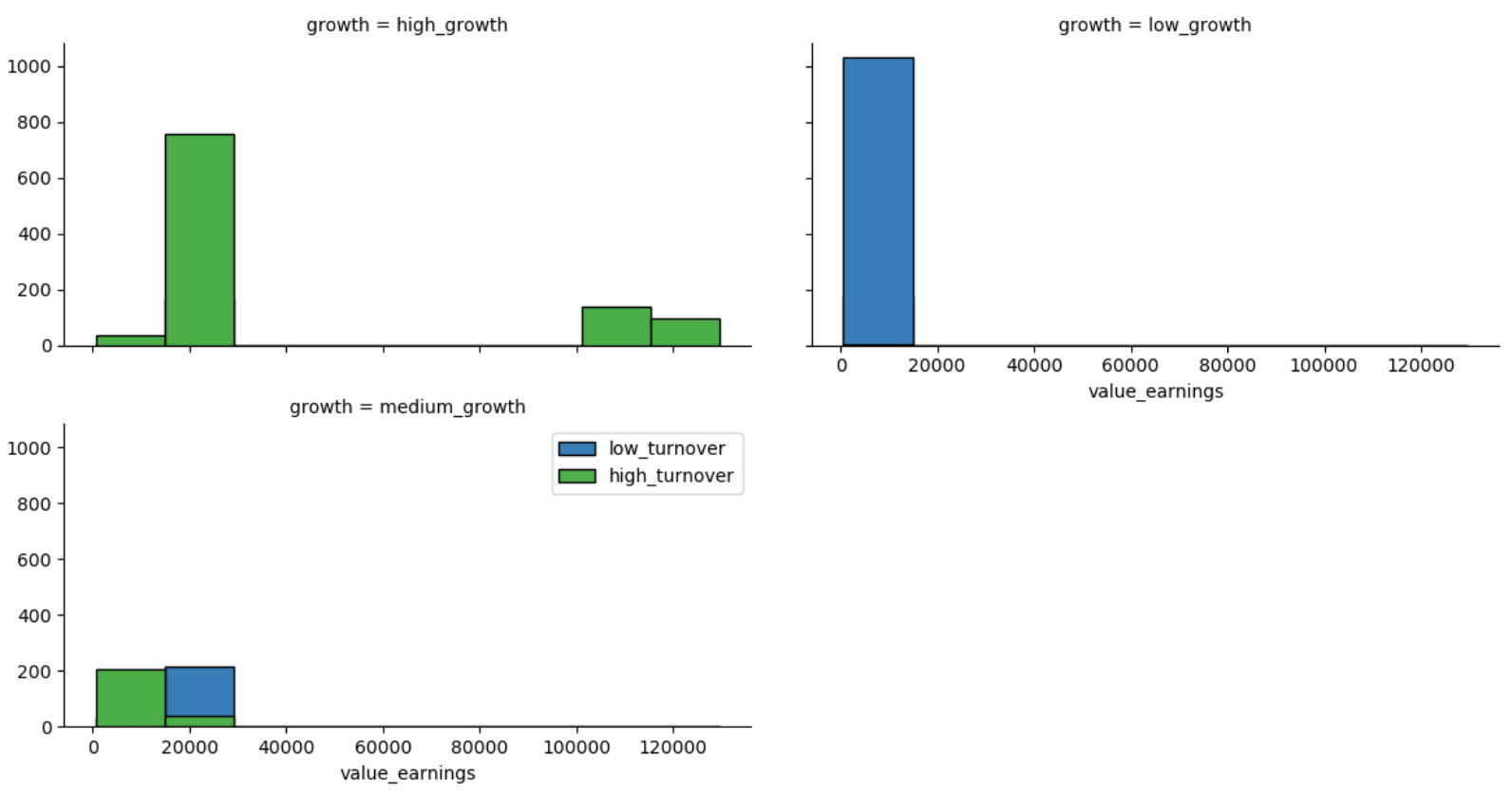
The cluster analysis above, mapping the number of job openings to hires while far from surprising in its positive relationship is nonetheless interesting and worth noting. At the lower ends of the spectrum, we see more tightly bunched relationships between openings and hiring, possibly due to the tighter fit of hiring than at the upper end of the spectrum – wherein opening are filled with hiring, but due to the looser fit, also result in higher turnover.





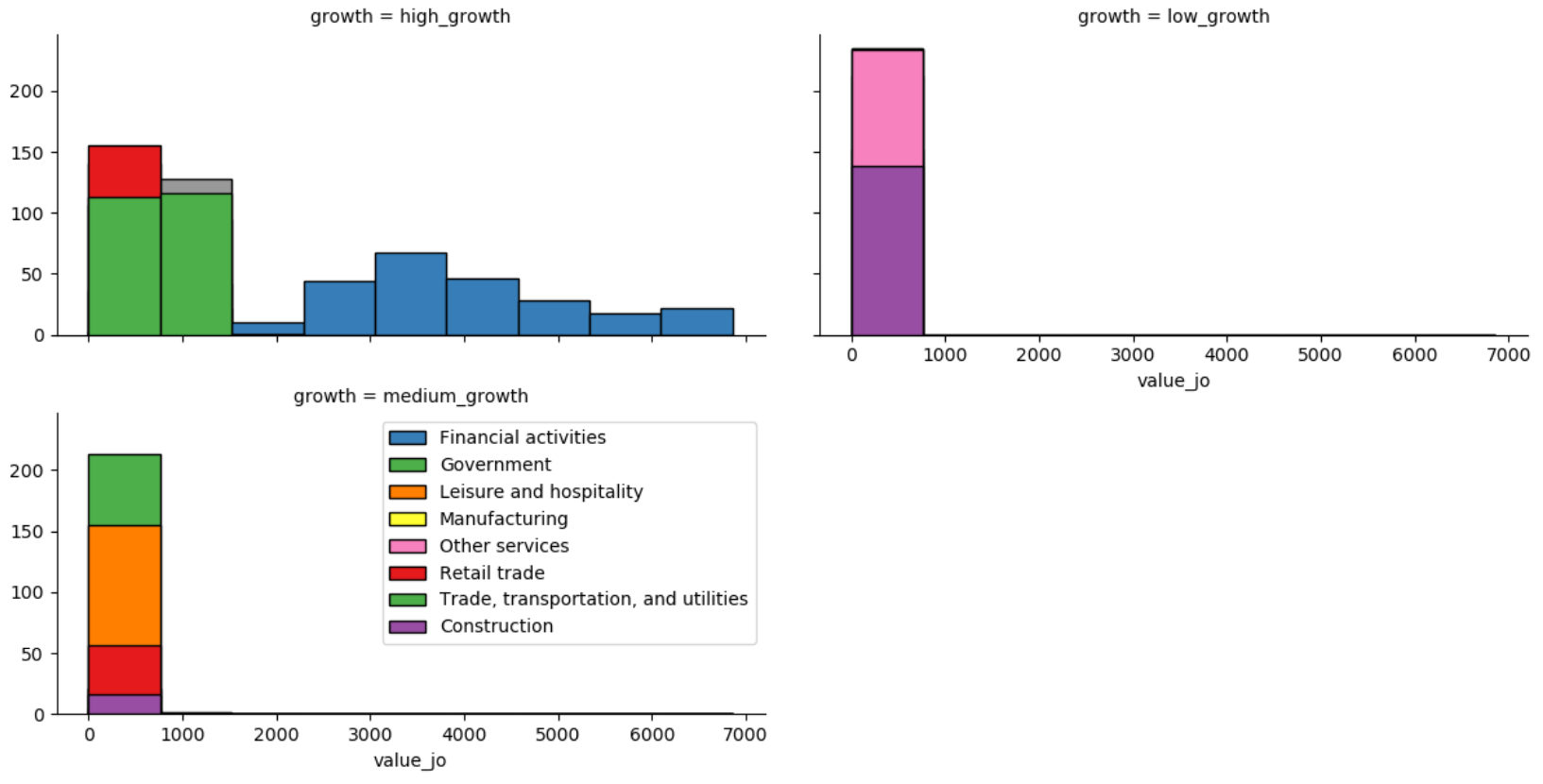
In performing the cluster analysis above, the decision to select 3 clusters was informed by the above analysis, that scores the performance of the clustering algorithm for various selections of values of K. Here, all values of k are less than 10, with k=3 emerging as the global optimum value of k.

**Methods**

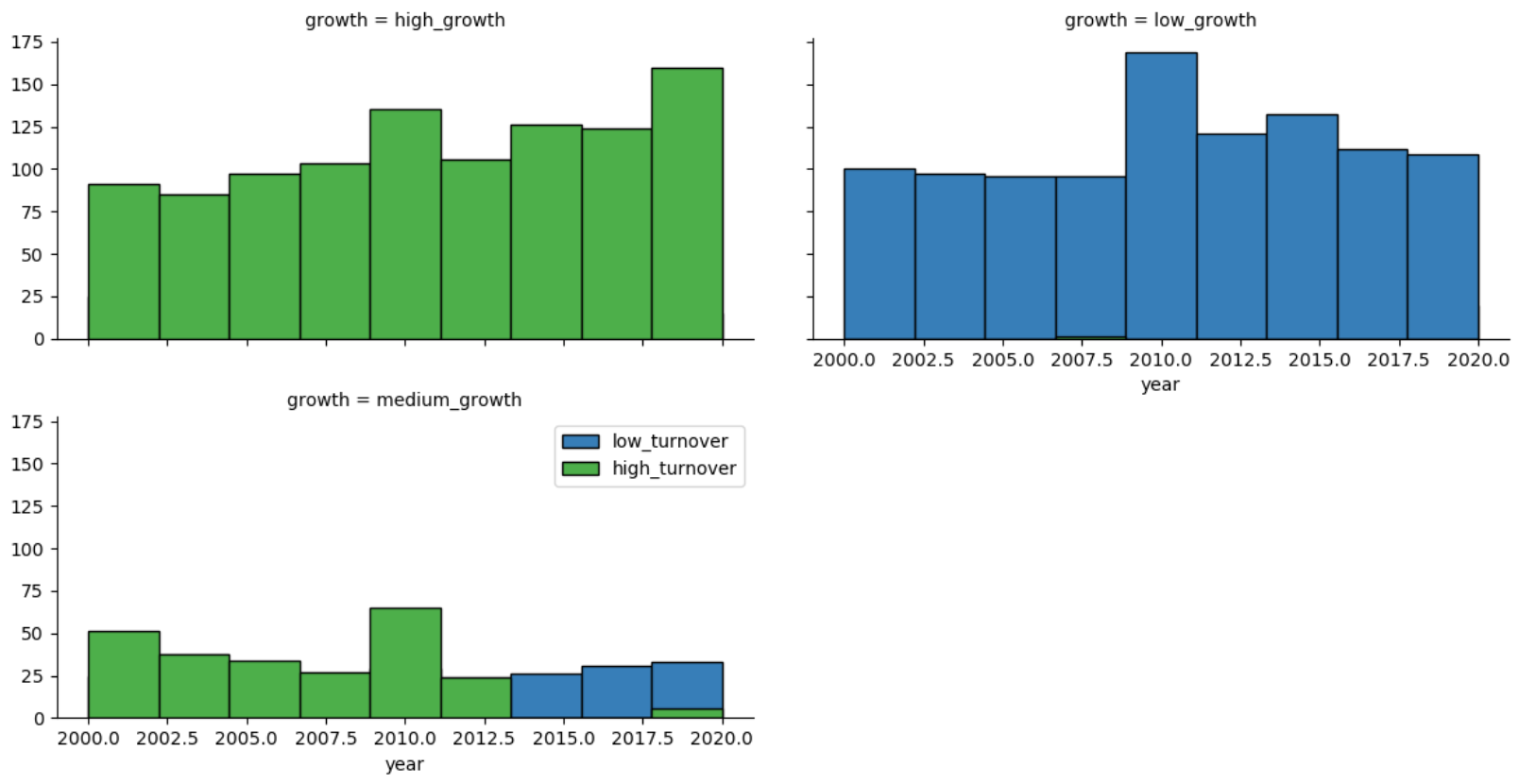
From glancing at the data in a few preliminary charts, we can make a few inferences before moving on to a deeper dive into predictive modeling. The following chart maps high growth vs high turnover industries by their earning potential. What we find is that while low growth-low turnover industries almost strictly have lower earnings, high growth-high turnover industries are bunched at the lower earning end of the spectrum, with a few outsized outliers at the upper end. Read differently, this graph would imply that if someone’s skill level dictated that they were in a lower skilled position, they would be better off situating themselves in the low growth industries due to their low turnover. Put differently, the reward to risk ratio of high growth-high turnover industries is only salient for very select professions. 

The following graph offers a little more insight into our previous findings. What we find now is that with the high growth-high turnover industries mapped in the upper left quadrant of the following facet grid, trade transportation and utilities is responsible for a majority of the turnover mapped, while commanding lower earnings. Finance, meanwhile, reflects lower turnover, but is mapped evenly distributed, with higher earnings, towards the upper end of the spectrum.

What this says finally is that fields like finance, where there is high growth and low turnover, may be the sweet spot for a profession- whether at lower ends of the spectrum, or at upper ends – primarily for their growth potential, but also for their low turnover.



Finally, we have the following graph – indicative of the growth of high growth industries over the years, while also taking on more risk, and thereby resulting in higher turnover. The converse is true of the top right group of industries, with growth declining, or increasing at a slowing pace, and additionally experiencing lower overall levels of turnover.

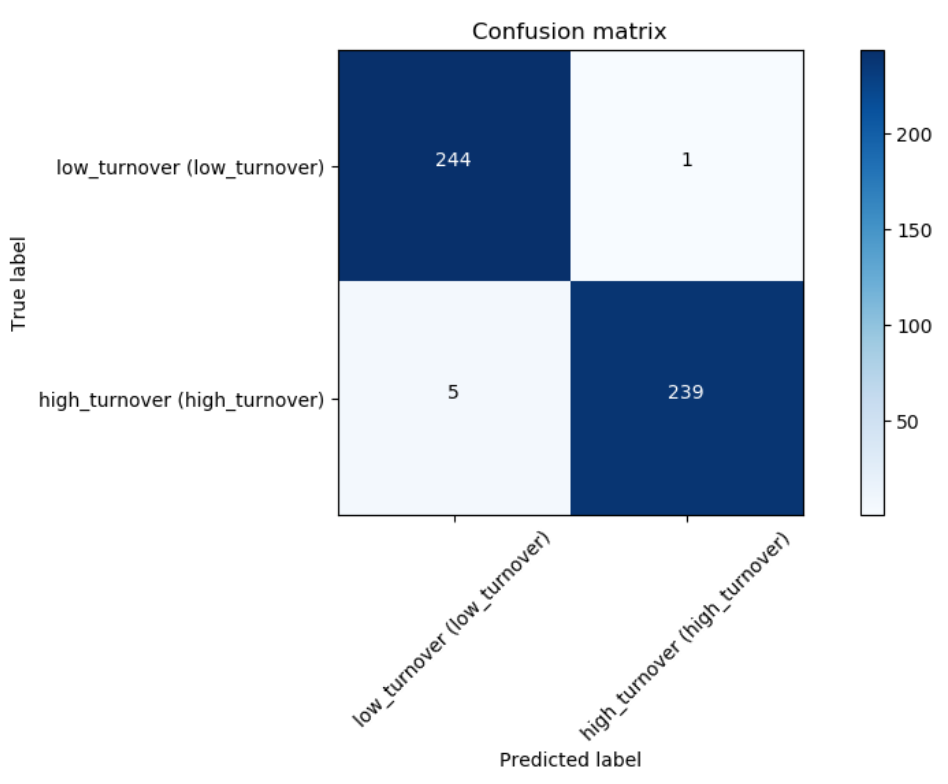


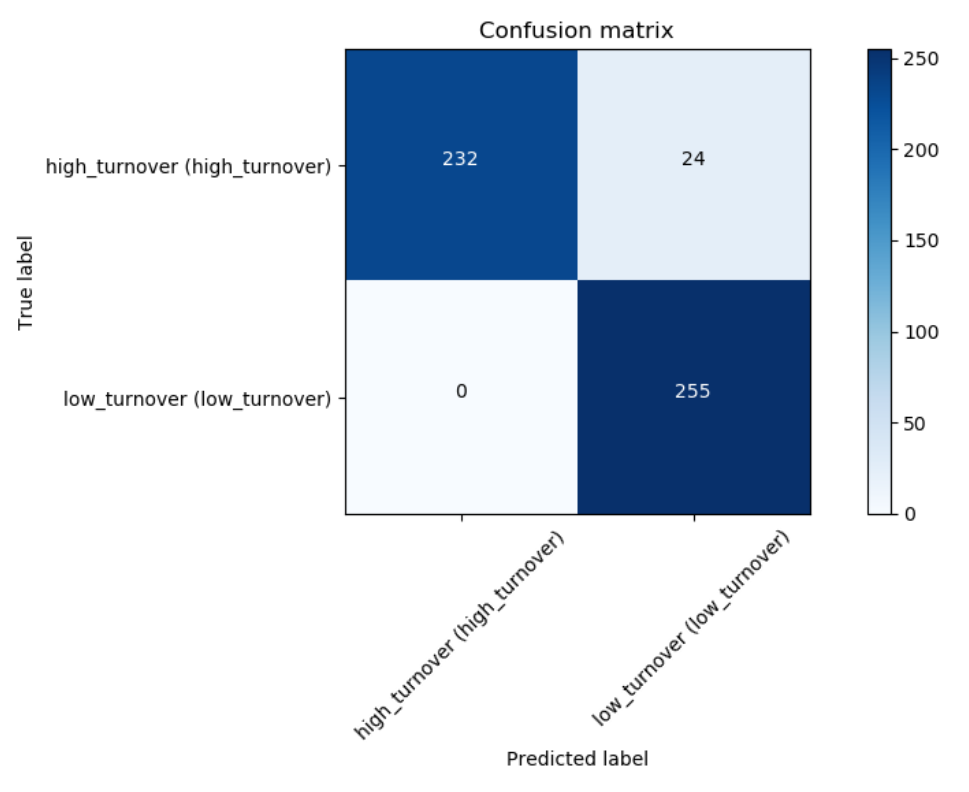
**Models, caveats:**

As mentioned previously, we used a variety of models embedded in the scikitlearn and scipy packages for cluster analysis, and to split and train our models, in order to better predict turnover.

The question then becomes this: given metrics such as earnings, job openings, and levels of hires, what can we say about the turnover in a given industry? Or differently applied, about the turnover in a company?

The following two confusion matrices tell us how two separate models perform at predicting employee turnover when provided feeder data that include metrics of earnings, job openings, hires, industry, and even which half of the year it is when looking at hiring decisions.





**Inferences, key takeaways**

In interpreting these analyses, it is important not to read too much into them. There is always more need for data, and more need for analyses. Yet, preliminary analyses, and dialogue lay the ground for a more productive analysis and dialogue.

I recently heard it said at a divisional meeting of the advanced analytics group, where I sit, that I/O psychologists strive to optimize between high ability and stability in their hires. This is one of the lenses through which they make their hiring decisions – others being indicators such as fair hiring practices, commitment, engagement, personality, culture, motivations, and so on.

And so, by trading off high ability for stability, managers are better able to make decisions that will lead to better employee performance, and commitment – occasionally returning to those models to review their decision-making process, and update it as needed.

This got me thinking, how can I test these hypotheses? I am still learning machine learning and its many applications. Where can I find data to actually deploy models? How often has administrative data been put under machine learning algorithms?

I am always pleasantly surprised by what I find from conducting these analyses – if not by the overarching theme, certainly by their intricacies. Please feel free to comment or criticize below.