

## **Example Cluster Analysis Report:**

### **Cluster Analysis:**

Developing a flight risk score for individual teachers is one-way human capital analytics can help the organization manage its turnover problem. Another possibility is to organize teachers in groups based on basic factors. One way to group teachers is through cluster analysis.

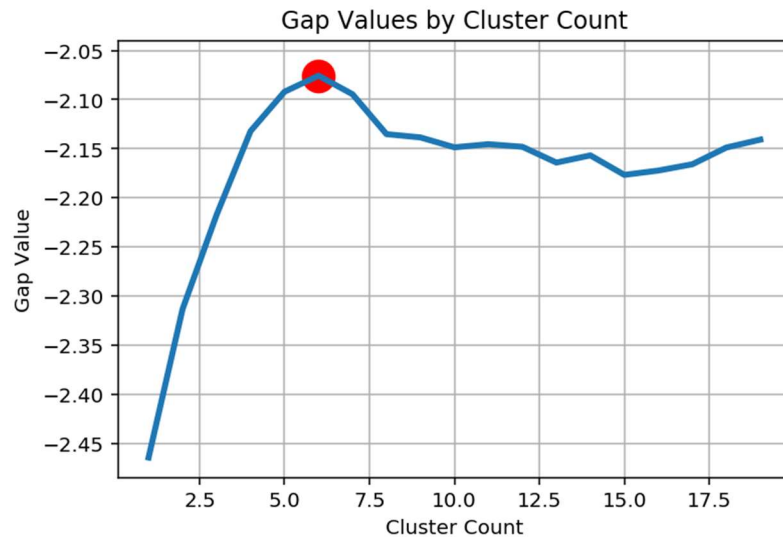
Cluster analysis is a different form of modeling than the modeling we've done in predicting teacher turnover. In the teacher turnover predictive model, we trained the model to predict an outcome we already knew (turnover). Training toward a known outcome is called supervised learning. Cluster analysis is unsupervised learning. We aren't trying to train the model to some known group membership; rather, the cluster algorithms attempt to discover group membership from the structure of the data. This is called unsupervised learning. Clustering algorithms differ in their approach, but generally try to place data into similar groups where the intra-group similarity is greater than the between group similarity.

When conducting a cluster analysis, there are typically no more than three key decisions one must make:

(1). *What data features to include in the clustering analysis.* In our case, we'll be using some basic information we know about teachers early in the school year: prior evaluation data (if available), current evaluation data (fall nine box), current engagement (fall employee satisfaction), and years teaching with the organization. These measures were selected based on their availability and their potential relevance to management decisions.

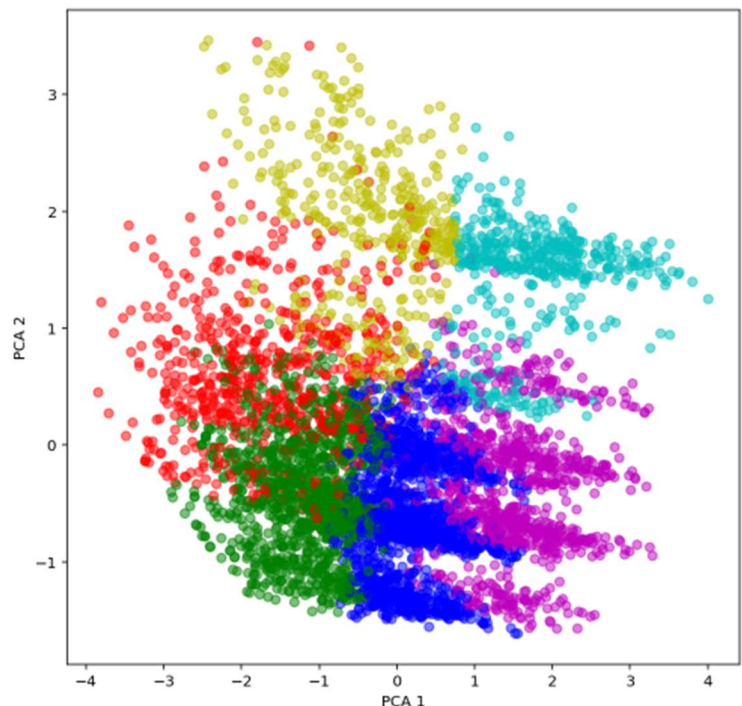
(2). *What clustering algorithm to use.* There are a wide variety of clustering algorithms (k-means, mean-shift, DBSCAN, Agglomerative Hierarchical Clustering, etc.). Each offer their advantages and a deep discussion exceeds the scope of this document. For this analysis, we leveraged k-means as it is computationally efficient and a good general-purpose clustering algorithm. K-means uses an iterative algorithm that seeks the center (centroid) of clusters that minimizes the distance of elements to the cluster. K-means is not a complicated algorithm to implement, but it does require knowing the number of clusters *a priori*.

(3). *How many clusters to create (situational).* As mentioned above, some algorithms require that we determine the number of clusters before the algorithm can run. In the case of k-means, the algorithm needs to know the number of clusters so it can set the number of centers it will optimize to. How does one determine the right number of clusters? If there isn't a business (or academic) reason to set the number of clusters to a specific value, then we can use one of several graphical techniques to empirically determine the optimal number of clusters. One option is to measure the "gap value" of each cluster set and determine which cluster set has the highest gap value, the distance between the clusters and a random uniform distribution of points (basically, data with no clustering whatsoever). What follows is a graph of gap values for various numbers of clusters, using the features we selected under k-means.



The graph makes it clear that the optimal number of clusters for this problem is six. We can use this information to run the k-means algorithm and sort our teachers into one of six clusters based on their prior evaluation data (if available), current evaluation data (fall nine box), current engagement (fall employee satisfaction), and years teaching with the organization.

Having run k-means for six clusters, we can examine the general orientation of the clusters by plotting the data according to its principal components. Principal components analysis is a dimension reduction technique that allows us to represent this multidimensional in two dimensions. Our six clusters are displayed to the right, give a general sense to the cohesion of the clusters, and where there might be potential overlap. It appears that k-means has navigated the shape of the data and identified reasonable clusters with which to work.



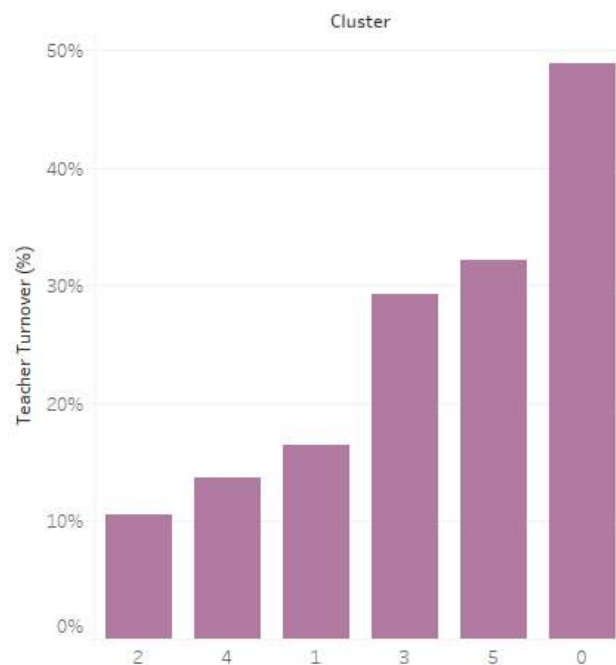
The first question to further verify the viability of our clusters is to examine group membership distribution. Before we analyze the attributes of clusters, we should first verify that each clusters has a meaningful number of teachers. Looking at the 2018-19 teacher set, we have adequate representation in each cluster. Our smallest cluster (0) has around 7% of the organization's

teachers in it. Assuming 3500 teachers in the system, this still represents around 250 teachers.



Of course, if we are going to use clusters to improve on teacher turnover, it is worth asking: are there differences in teacher turnover by cluster? We can answer this question descriptively by plotting the teacher turnover rate by cluster:

This simple bar chart suggests that there is meaningful variation of turnover rates by cluster. This is what we would hope to find. Given these data, we can analyze the attributes of each



group and build profiles that will help the organization design management interventions for its various groups of employees.

#### Employee Profiles:

The difference between a cluster and profile lies in understanding the “typical” member of the cluster and describing them in relevant terms. In our case, we selected a core set of features and ranked each cluster from Lowest performance (1) to highest performance (6) on that performance measure. Doing so produced the following table:

Cluster	Turnover	eSat	LCI Fall	Evaluation	Terminations	Yrs. Teaching	Mean Rank
0	1	1	1	1	1	1	1.0
3	3	3	2	2	2	2	2.3
5	2	2	4	4	3	4	3.2
1	4	5	3	3	4	3	3.7
2	6	4	6	5	5	6	5.3
4	5	6	5	6	6	5	5.5

The clusters (in the first column) are ordered by their mean rank. From these data, we can infer personas associated with each cluster.

#### Cluster 0: The Overwhelmed

Inexperienced, underskilled, and likely to leave describes the teachers categorized in cluster 0. They score the lowest in every one of the performance measures we considered. They are the most likely to be disengaged, underperforming, have a poor evaluation, leave the organization (nearly 50% turnover), and constitute non-regrettable turnover. Two-thirds of the overwhelmed teachers are in urban schools. This is an unfortunate but consistent finding with national research that schools in urban areas employ a disproportionately high proportion of under-equipped teachers.

Interestingly, this group is also paid closest to market pay:

Cluster	Pay Relative to Market
2	
4	
3	
5	
1	
0	

These data are compatible with the organization’s pay model, which offers the most competitive salaries relative to market to teachers early in their career. The high turnover of cluster 0 (the overwhelmed) may also explain why pay relative to market is not a stronger predictor in the

teacher turnover model, given that this group performs the highest in turnover and pay relative to market.

### **Cluster 3: Will, Not Skill**

Cluster 3 is fairly like cluster 0 (The Overwhelmed). They perform near the bottom in most performance categories and are employed disproportionately in urban schools. The biggest difference between the cluster 3 and 0 is that teachers in cluster 3 are much more engaged (even though their pay is worse relative to market than the Overwhelmed group). They may not do well, but they are happy and engaged. In terms of skill vs. will, they have the will --- the desire to do good work in the classroom, but they lack the skill (both behavioral and technical competencies to perform at a high level).

### **Cluster 5: Skill, Not Will**

Cluster 5 has the second highest turnover and second lowest satisfaction out of the six clusters. These teachers have above average tenure at the organization and above average evaluation scores. They are more experience and more skilled, but yet their attrition rate is relatively high (over 30% in 2018-19). They seem to have the skill but are dissatisfied with their experiences and leave at high rates. Yet, only 2.8% of their departures are classified as “not regrettable.” This group doesn’t have the skill gap that the Overwhelmed cluster does, but they do require some sort of engagement intervention. Part of the puzzle might be their compensation, which is over XXXX lower than market rates on average.

### **Cluster 1: Missionally Minded**

Cluster 1 also over indexes on urban teachers (65% of the cluster 1 teachers work in urban schools), but these urban teachers demonstrate higher competency than the other two clusters (Overwhelmed, Will Not Skill) and better engagement than most clusters (second highest fall employee satisfaction scores). These teachers are the missionally minded who work in predominantly disadvantaged schools and are the core to those schools’ success. Working in historically disadvantaged communities is very challenging, and these teachers – with their performance and persistence --- are key to the organization’s high success rate in urban communities. The good news for the organization is that while their pay is still below market, the Missionally Minded have the second smallest gap to market for all teaching clusters.

### **Cluster 2: Competent, Committed, But Critical**

You can walk into any school find a cluster 2 teacher. They are relatively evenly spread across the different kinds of schools that the organization runs. Cluster 2 teachers are competent (high overall evaluations), committed (the lowest turnover rate of any cluster group), but they are not as happy as their performance and tenure would otherwise suggest. These teachers have been with the organization and the longest, and perhaps in that time developed some critical views of certain aspects of school operations or company policy. They do work in schools that tend to

have the best overall culture (high learning culture index scores), but one of the biggest flags in their data is pay.

#### **Cluster 4: It's All Good**

The top performing cluster on average is group 4, which over indexes at suburban schools. For these teachers, "it's all good." They perform very well relative to other clusters with top performance in evaluations, non-regrettable turnover, and overall engagement. Their engagement scores are particularly noteworthy, as over 82% of this cluster are either satisfied or highly satisfied working at their school. They are less experienced than the "Competent, Committed, But Critical Group," but their performance is incrementally better. Perhaps one reason that their satisfaction score is high is that their building culture is very strong (high learning culture index) and their pay --- while still low to market relative to other schools -- is not nearly as low as those in the Competent, Committed, and Critical group.

#### **What can the organization do with these personas?**

The purpose of categorizing the organization's teachers into clusters and then analyzing those clusters so as to build personas around them is to help management design targeted interventions for teachers in those clusters that may be more effective than general interventions for all teachers. The philosophy is similar to customer segmentation in marketing: understand your different customer segments and target marketing efforts that are tailored to be effective for those groups.

In the case of the employee personas, management can consider to what degree it can resource improving the experiences, engagement, and skill of teachers in various segments. In the reality of this business with its fixed funding formulas and limited resource, it is unlikely that the organization will be able to address all the needs represented in the employee personas presented, and in some cases, it should not. Below are some basic recommendations management can consider as it determines next steps.

#### *Cluster 0: The Overwhelmed*

These teachers underskilled and unengaged. A pragmatic assessment needs to be made for those in this group as to their ability and interest in becoming a solid teacher. National data suggest that a high percentage of teachers leave the profession within their first five years (some estimates are 50%). Many people go into teaching because it is familiar and a seemingly safe career choice; however, teaching is an incredibly demanding profession, particularly in underserved communities. The organization's best strategy here may be to focus on improving its pipeline to bring in more teaching talent so it can part ways more quickly with those who lack the skill and drive to be successful in the role.

#### *Cluster 3: Will, Not Skill*

This group wants to be successful but lacks the skills to do so. More professional development and mentoring may be very beneficial to a group that is already generally engaged.

#### Cluster 5: Skill, Not Will

The organization's Human resources should consider conducting focus groups with teachers in the Skill, Not Will group. Evaluation data suggest that these teachers have the skills necessary to be successful but are otherwise unengaged and have an above average risk of turning over. Listening to employee concerns may yield targeted and cost-effective interventions that improve the engagement of this group.

#### *Cluster 1: Missionally Minded*

Celebrate this group. These teachers predominantly work in disadvantaged schools, have consistent performance, and are committed to the work they are doing.

#### *Cluster 2: Competent, Committed, But Critical*

Don't take this group for granted. Yes, they have the longest tenure and lowest threat of leaving the organization. They also have enduring concerns, perhaps related to their compensation, that is driving a relatively high level of dissatisfaction (compared to other clusters, and relative to their performance). While the organization may not be able to fully address the wage gap to market, but managing expectations and addressing concerns in a high value/low cost way will be key for this cluster. Consider doing a conjoint analysis just on this cluster to determine what benefits people value vs. what they cost and search for the highest value-to-cost ratio initiatives. In the past, student loan forgiveness plans have fit that description of being highly valued, but less expensive than raising pay fully to market.

#### *Cluster 4: It's All Good*

This cluster is not your problem. They are highly engaged and effective. Monitor their data as updates become available so as not to miss an inflection point in either performance or engagement.