Table of Contents

[Initial Analysis 2](#_Toc476345756)

[Drawbacks of the 2017- 3-Tier System 3](#_Toc476345757)

[Determining Optimal Number of Clusters based on Elbow Method 3](#_Toc476345758)

[University Classification using K means with 4 clusters 4](#_Toc476345759)

[Impact on Resource Utilization 5](#_Toc476345760)

[Conclusion 5](#_Toc476345761)

[Appendix 6](#_Toc476345762)

[Feature Engineering 6](#_Toc476345763)

[Tier -Definitions 6](#_Toc476345764)

[Exhibit 1 – Drawbacks of 3 Tier system 7](#_Toc476345765)

[Exhibit 2 – Visual Validation of 4 Tier System 8](#_Toc476345766)

[Exhibit 3 – Descriptive Validation of 4 Tier System 9](#_Toc476345767)

[Exhibit 4 – Correlation of Variables with Acceptance 10](#_Toc476345768)

[Exhibit 5- R Code 10](#_Toc476345769)

## Initial Analysis

The TFA applicant data provided a platform to analyze the 3-tier university classification system put in place this year. The initial hypothesis that a combination of key parameters would play a crucial role to influence the end objective proved to be true (Please refer [Exhibit4](#_Exhibit_4_–) – Correlation among variables). These parameters have been grouped under 2 headings*- Awareness and Potential*. We call it the Awareness-Potential Classification model (AP model). The factors which define AP model based on high correlation with Acceptance Rate have been shown below:

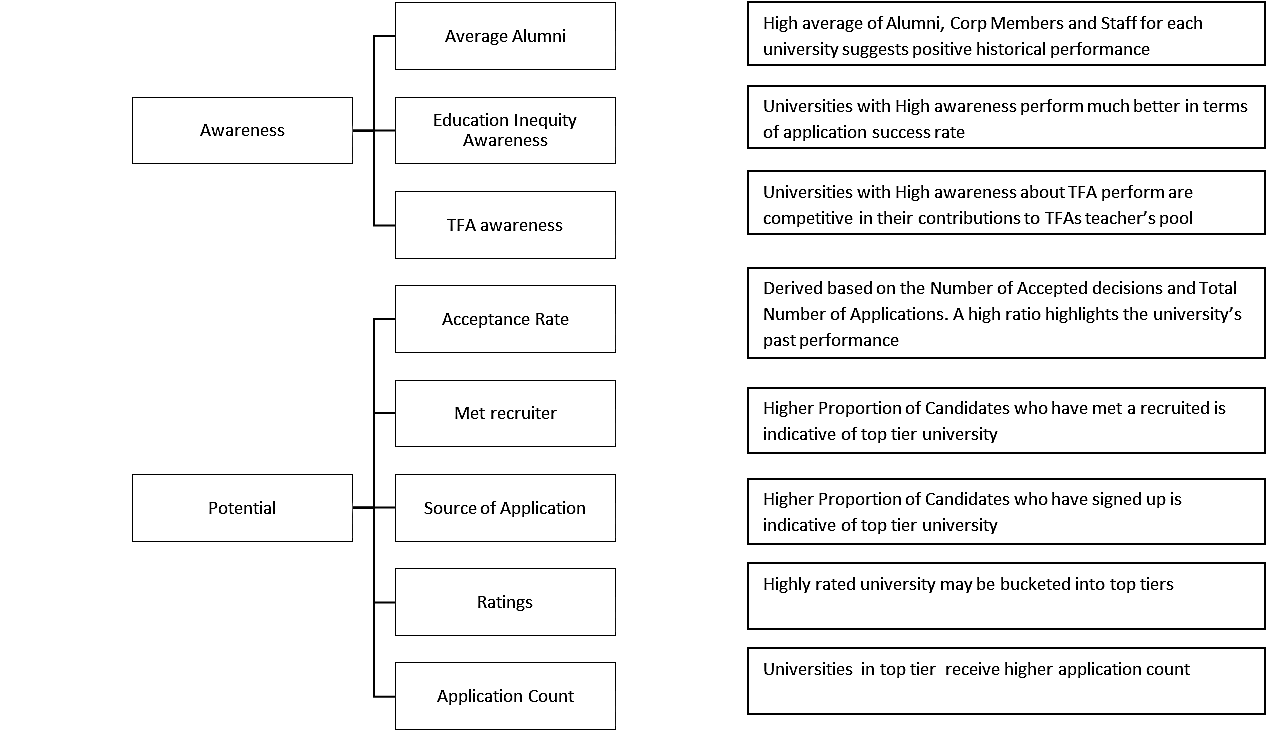


Fig 1

## Drawbacks of the 2017- 3-Tier System

* The means of the density distribution of the ‘Number of Applications’ and ‘Acceptance’ for the 3-tier system fall within a very narrow range (Please refer [Exhibit 1](#_Exhibit_1_–) for the distribution charts). This implies that the recruiting efforts across 3 tiers is resulting in the same throughput from each tier which should not be the case as more effort is being spent in top tiers (Tier 1 Universities has one dedicated recruiter assigned)
* Under the Awareness-Performance(AP) model parameters, the current allocation of universities into 3-tiers do not shown any specific patterns. For instance, there are schools in tier 1 which have a combination of below average values for the AP-model parameters and are potential candidates for lower tiers. Such misclassifications could prove costly to TFA as there is a cost factor associated in allocating resource to universities under these tiers

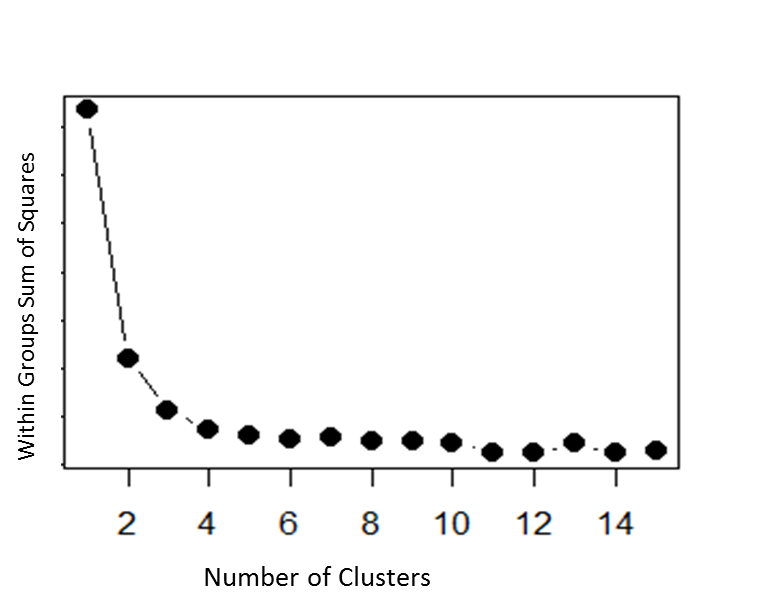
|  |  |  |  |
| --- | --- | --- | --- |
| Tier | Total Number of Universities | Number of Universities Misclassified | Misclassification Rate |
| 1 | 79 | 11 | 11/79 |
| 2 | 119 | 48 | 48/119 |

* The stratification of key parameters in the data-set do not rationalize the 3-tier system. There are 4 distinct levels – Very High, High, Medium, and Low for each of the parameters grouped under the AP-model which points towards a 4-tier system. To confirm this hypothesis, ‘Elbow Method’ has been used to arrive at the optimal number of clusters based on the observed difference in the within-cluster dissimilarity

## Determining Optimal Number of Clusters based on Elbow Method

The optimal number of clusters have been arrived by using K-means algorithm. The idea of the elbow method is to run k-means clustering on the given TFA dataset for suitable values of cluster and for each value calculate the sum of squared errors (SSE). The value post which the SSE shows minimal decrease becomes the optimal cluster for K means.

From figure below, it can be observed that after 4 clusters, within-cluster dissimilarity difference is not substantial. Consequently, we can say that the optimal number of clusters to be used is 4.



Group variance doesn't decrease significantly after 4 clusters implying that 4 is the optimal number of clusters for the given data set

Fig 2

## University Classification using K means with 4 clusters

TFA University Classification is typical case of unsupervised learning. Hence, we have used K-means which is one of the unsupervised machine learning algorithms to classify the universities into optimal number of 4 clusters derived above using the elbow method. Below is the result of classification obtained after fitting the K-means algorithm:

#### 4 Tier Classification vs 3 Tier Classification

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Tier 1** | **Tier 2** | **Tier 3** | **Tier 4** | **Total** |
| **Campus: 1** | 17 | 30 | 21 | 11 | 79 |
| **Campus: 2** |  | 12 | 59 | 47 | 118 |
| **Campus: 3** |  |  | 7 | 87 | 94 |
| **Grand Total** | 17 | 42 | 87 | 145 | 291 |

Fig 3

For detailed classification at University Level-Please refer [Exhibit 3](#_Exhibit_3_–)

The 4 Tier system results in a major reshuffle at tier 1 and tier 2 levels as shown in the above table:

* Tier 1 with 79 universities has been recategorized in Tier 1 and Tier 2(Please refer [appendix](#_Tier_-Definitions) for new tier definitions and [Exhibit3](#_Exhibit_3_–)- for reshuffling of university across the 4 tiers)
* A new tier in tier 2 has been created to account for universities which may or may not need special full time attention always during the year. Hence, a recruiter from this tier can handle one or two universities
* The classification of universities into 4 tiers ensures that mean of the density distribution of Number of Applications and Acceptance across 4 tiers are well separated. (Please refer [Exhibit2](#_Exhibit_2_–)). In-addition, the top tiers have higher values which implies increased throughput in terms of quality (Acceptance Rate) and quantity (Number of Applications)

### Impact on Resource Utilization

In terms of Resource Utilization, it can be observed that the available resources have been spread across the 4 tiers. This has the following effect:

* Recruiters can focus on a narrower range of universities especially for the first 2 tiers which can positively impact the quality and quantum of applications

#### Resource Requirement based on number of schools in a tier

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Tier 1** | **Tier 2** | **Tier 3** | **Tier 4** | **Total** |
| **3 Tier** | 79 | 39 | 14 | N/A | 132 |
| **4 Tier** | 17 | 36 | 29 | 50 | 132 |

Assumption: Total number of resources has been assumed to be 132 which has been derived from dataset

Fig 4

## Conclusion

TFA’s goal to provide all deserving candidates an opportunity can be achieved by:

* Creating a structure which enables TFA personnel to reach out to the right universities
* Creating a mechanism which enables TFA personnel to maximize their efforts by reaching out to apt candidates

#### Road map to Improved Efficiency with 4 tier system

* The 4-tier system provides an intricate grouping of universities to minimize variation of recruitment parameters within each tier ([Exhibit 3](#_Exhibit_3_–))
* It allows detailed and structured recruitment better customized for each tier
* Continuous monitoring will not require extensive effort
* Mid-cycle or inter-cycle reclassification of specific universities may be attempted without disturbing the existing structure or needing extensive transition efforts.

## Appendix

### Feature Engineering

The following variables have been derived from the given data set to be used in K means classification algorithm:

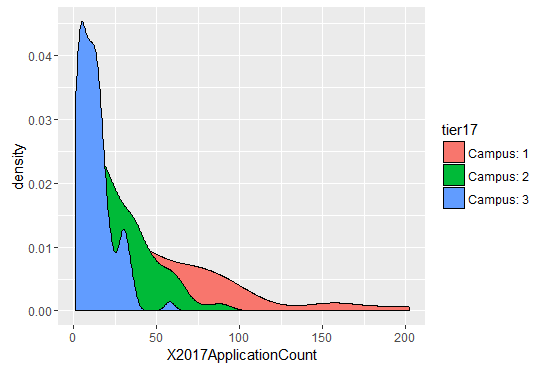
|  |  |
| --- | --- |
| X17AppliedNum | Total Number of applications from each university for 2017 |
| X16AppliedNum | Total Number of applications from each university for 2016 |
| X17accceptnum: | Total Number of candidates accepted from each university for 2017 |
| X16accceptnum | Total Number of candidates accepted from each university for 2016 |
| X17withdrawnum | Total Number of candidates who withdrew application from each university in ‘17 |
| X16withdrawnum | Total Number of candidates who withdrew application from each university in ’16 |
| source. Sourced | Percentage of candidates sourced by recruiters from each university |
| metrt. Y | Percentage of candidates for each university who have met the recruiter |
| gpa36.Y | Percentage of candidates from each university who have a GPA of 3.6 and above |
| Average Alumni | The average value of Alumni, Corporate Members, and Staff at TFA |

### Tier -Definitions

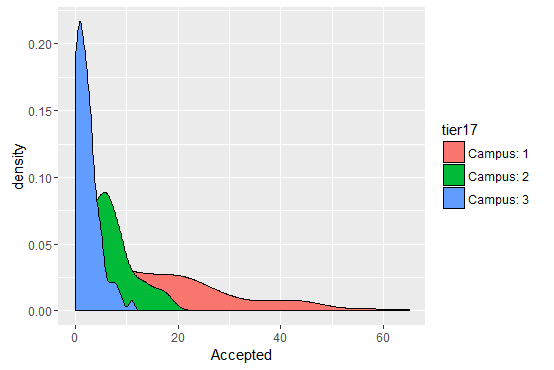
* Tier 1: Same as 2017-Tier 1. Each university has its own recruiter
* Tier 2: In tier 2 each recruiter will have a portfolio of 1-2 schools. This tier would give flexibility to the recruiter to handle 1 or 2 more important from either
  + Existing pool of Tier 1 schools which may not need his full attention Or
  + Existing pool of Tier 2 schools which may need additional attention
* Tier 3: Same as 2017-Tier 2 - each recruiter will have a portfolio of 2-3 schools
* Tier 4: Each recruiter will have a portfolio of more than 3-5 schools

### Exhibit 1 – Drawbacks of 3 Tier system

The figure below shows that there is no distinct movement towards the right for the mean value of the density distribution of the Application Count across 3 tiers

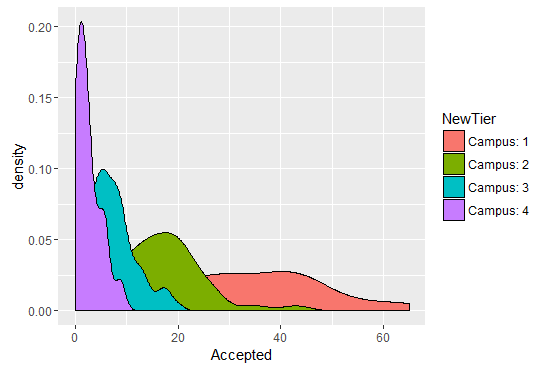


The figure below shows that there is no distinct movement towards the right for the mean value of the density distribution for Acceptance rate across 3 tiers

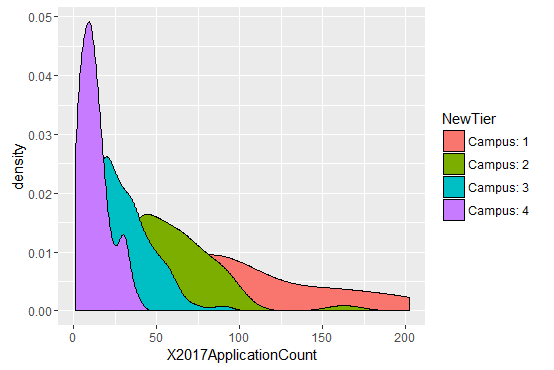


### Exhibit 2 – Visual Validation of 4 Tier System

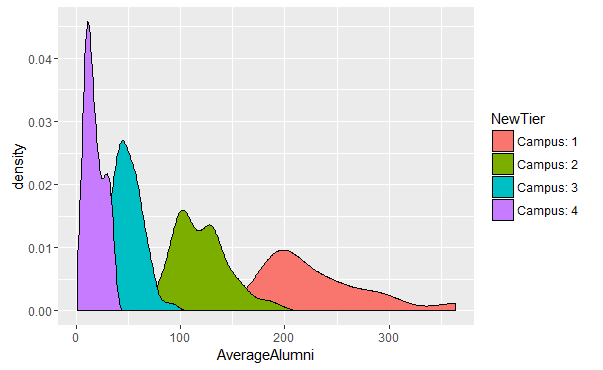
The figure below shows that for the top tiers as the effort per university increases the acceptance rate also increases proportionally



The figure below shows that for the top tiers as the effort per university increases the application rate also increases proportionally



It can be observed that Average Alumni has a higher density a Tier 1 and Tier 2. This validates the classification system as universities which have performed well historically are being misplaced in top 2



### Exhibit 3 – Descriptive Validation of 4 Tier System



It can be observed from the table below that all universities classified into the 4 tiers conforms with the average value of the parameters under Awareness and Potential for that tier



### Exhibit 4 – Correlation of Variables with Acceptance

|  |  |
| --- | --- |
| **Independent Variables** | **Acceptance** |
| University application count '16 | 0.71 |
| Candidates met / University | 0.7 |
| Average Alumni | 0.59 |
| Accepted count '16 | 0.55 |
| Candidates sourced/University | 0.5 |
| University Application count '17 | 0.45 |
| Ratings | 0.4 |
| Average GPA/ University | 0.3 |
| Private University | 0.2 |
| Size | 0.2 |
| Undergraduates/University | 0.4 |

Historical application count, historical accepted count, Sourced candidates/university, Ratings and met candidates/university have high correlation with Accepted Count. Hence, we have used these variables as key parameters for our analysis.

### Exhibit 5- R Code

[Github Link](https://github.com/AnilrajPazhety/Projects)