Team Control Number

10549

Problem Chosen

A

2020

HiMCM

Summary Sheet

Summer jobs have become an increasingly popular topic among high school students these days. They provide invaluable enrichment opportunities to make money and gain experience and skills. However, due to the outbreak of coronavirus, it is more challenging for high school students to identify the most appropriate summer jobs for them. To help high school students find the best summer job options effectively and efficiently, it is important to examine the variables associated with summer jobs.

We first identify related variables and collect data useful to our model. We consider 22 different summer jobs and consider variables including income, company size, risks, comfort, and skills to be gained. These are often the strongest motivators for high school students seeking summer job opportunities, so incorporating them is essential.

For the overall model, we adopt an Analytic Hierarchy Process to evaluate the weights or relative importance of each variable for each individual. The process is based on the user's subjective evaluation of their personal preferences. In order to address this subjectivity, we also introduce the Entropy Method to analyze the summer job options. Combining the weights we calculate from both models, we then obtain the overall weighting used for each individual. These weights are then applied on the summer job options we have. However, considering students usually want to have an array of open options to choose from, we utilize K-Means clustering rather than simply returning the one single “best” job. Then, the best cluster of job options is output to our user.

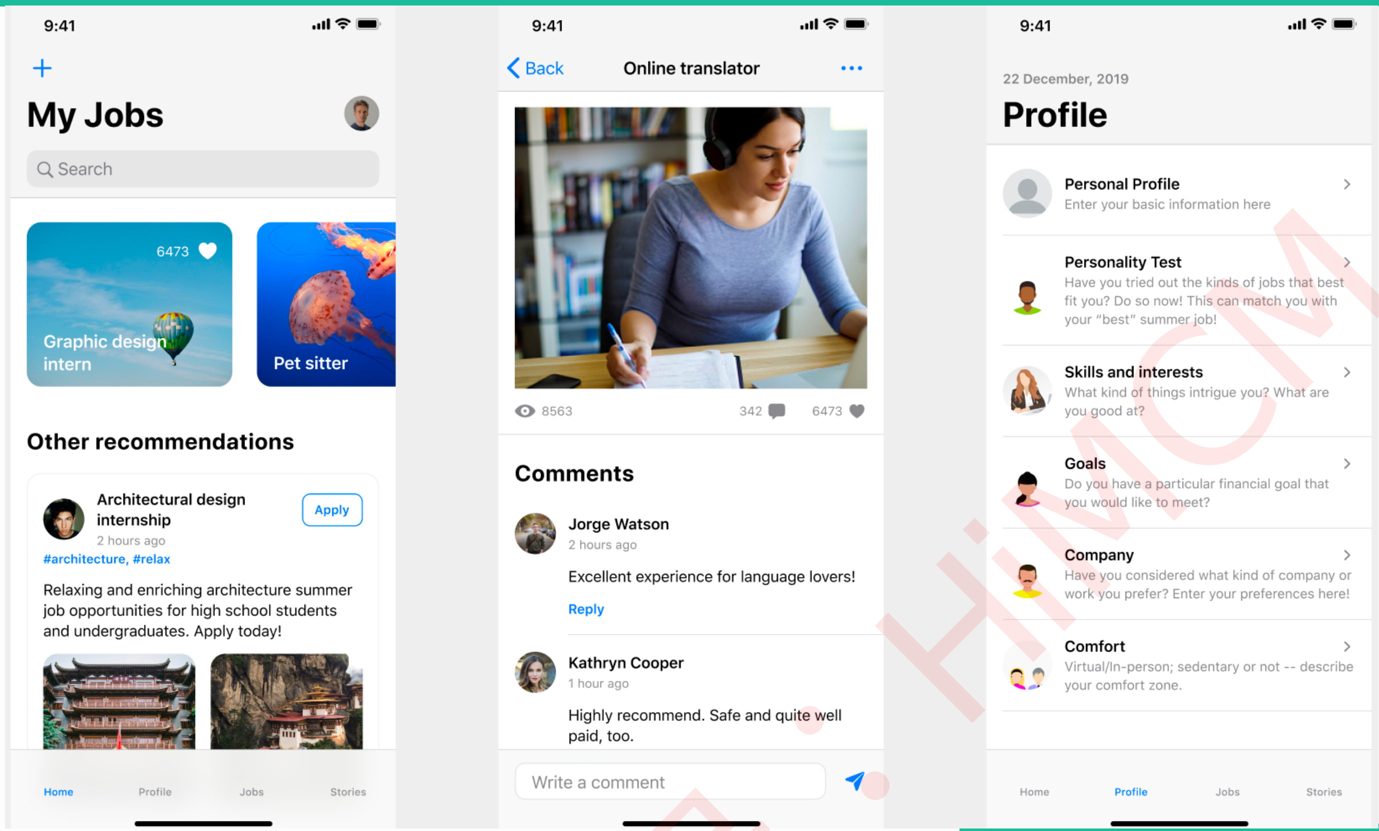
In regards to the fictional characters, we develop 10 different characters that we believe are highly representative of the US high school student population. Application of our developed model onto these fictional characters indicates that our model is fairly effective in identifying the most appropriate summer jobs for students.

For the optimized model, we recognize that high school students are very likely to input inconsistent matrices, which will result in ineffective evaluation of the model. If this occurs, it would be tedious to force students to re enter their information until the result becomes consistent. Consequently, we optimize our model by adjusting the values stored in the original matrix that is unsatisfactory. Then, a more consistent comparison matrix that still demonstrates overall relations among the variables can be used for our model. While our model is in no way perfect, we believe it is a comprehensive system that takes sufficient variables into account. Furthermore, we have combined the AHP model, Entropy system, and our optimized model to develop an overall fairly realistic, effective, and objective mathematical model.

Keywords: Summer jobs; Analytic Hierarchy Process method; Entropy Weight method; K-means Clustering; Optimized model

NOVEMBER 17, 2020

GET A JOB!



HOW TO FIND THE BEST SUMMER JOB?

Have you ever felt frustrated from finding the best summer job for you? Well, you are not alone. Now due to the COVID-19 outbreak, it becomes even more difficult for high school students to find suitable summer jobs.

Thankfully, you don't have to stress about that now! Our research constructs a straightforward model for high school students to use. Our model allows you to simply enter your personal information and goals for the job, and it will quickly return the best summer job options for you! Our model incorporates jobs' income, comfort, skill level, company size, and risks, as well as your personal situations and preferences, to make the best summer job recommendation. Furthermore, our app will offer you an opportunity to view other students' comments on the summer jobs, allowing you to further decide on which job would be the best for you personally.

What are you waiting for? Come and find yourself a job today!

SUMMER JOBS RECOMMENDER

BEAUTIFUL LAYOUT, STRAIGHTFORWARD MECHANISM, AND EFFECTIVE RECOMMENDATIONS

According to our research, teen summer job has declined in the US. This model will provide personalized job recommendations for high schoolers.

TABLE OF CONTENTS

1. Introduction

1.1 Background

In recent years, summer jobs have become a hot topic among high school students. From restaurant waiter and babysitter to data analyst and research assistant, summer job opportunities encompass a very wide spectrum. Furthermore, not only do they serve as sources of income for high school students, but they also enrich students with invaluable skills and experiences that will be vital to their future major and career. However, as global pandemic COVID-19 continues to roar throughout the world, it becomes progressively more challenging for high schoolers to efficiently and accurately identify the best summer jobs for them in the year to come. Pew Research Center suggests that teen summer jobs become less common in the US [5]. Due to the pandemic, it is also more difficult for students to seek out professional support about finding the most suitable job for them. As such, helping high school students look for the best job for them becomes our problem to solve. Therefore, this paper analyzes the relevant models such as AHP and optimizes them in order to develop a summer job evaluation system that recommends the most suitable jobs to the user.

As high school students often lack a comprehensive and objective understanding of their own needs, we also adopt an optimized model that will address the potential inconsistency in students' input comparison matrix. Considering the subjectivity problem, we will also apply more objective methods like Entropy Method by taking the entropy of information into account.

1.2 Problem Restatement

The essential question is to develop a model to suggest optimal summer job options to high school students. The question can be broken down into the following subquestions:

Question 1: We need to identify all the variables that high school students today would consider when looking for a summer job. We also need to collect data on common summer jobs about the variables that high school students will likely consider.

Question 2: Using the variables identified in question 1, we need to develop a recommendation system that assigns different weights to different variables based on individual abilities, preferences, and situations. This model should be applicable to any high school student and can return the most appropriate job option.

Question 3: Use Entropy Method to offset the subjectivity problem that arises from question 2. Make objective ranking of the jobs and then combine the two results to form a better, more objective model.

Question 4: We need to develop ten fictional characters that are representative of the entire high school student population. Discuss the effectiveness of our model when it is applied on each fictional person.

2. Assumptions and Variables

2.1 Assumptions and Justifications

Assumption 1: The United States is our primary region of concern. Since high school summer jobs are relatively very common in the United States, there is a vast source of data for us to utilize. This makes it easier to test our proposed model.

Assumption 2: All jobs of the same kind have the same salary, work time, and company size. While it is true that there are variations in salary, work time, and company size even for the same type of job, it is safe to neglect that variation for the purpose of modeling.

Assumption 3: High school students have perfect knowledge about their summer plan. We assume high school students have perfect knowledge about their plan in summer. This means that they are well aware of the recreation activities and social events they will have in summer. This allows them to better plan their summer job.

Assumption 4: High school students get paid equally and fairly in summer jobs. Because there is limited data on salaries paid specifically for high school students, we must assume high school students receive the same payment as the amount we found online. It is reasonable to make this assumption, since salaries received from summer jobs generally do not deviate significantly from the average salary we identify.

2.2 Variables Chart

The table below defines all the variables we will use throughout this paper:

Symbol	Definition
C	Comparison matrix
CI	Consistency index
RI	Random index
CR	Consistency ratio
X_{ij}	Judgement matrix of job i and evaluation index j
N_{ij}	Judgement matrix of job and evaluation after normalization
p_{ij}	Weight of the i-th plan of j-th index
e_j	Entropy of j-th index
g_j	Redundancy
w	Weight of the index
B	Regression coefficient for the model combining AHP and Entropy
s_i	Final score of job i calculated using weights from AHP model and Entropy Method
SSE	Sum of the squared errors
k	Number of clusters
G_i	Cluster in K-means clustering
d	Sample point in G cluster
m_i	Center of mass of G cluster
T, U	N-dimensional vectors for matrix
$\overline{T}$	Normalized eigenvector of the normalized matrix
r	Consistency factor of T relative to T
v	Deviation from T to T

3. Models Overview

3.1 Analytic Hierarchy Process

As mentioned earlier, identifying variables that high schoolers consider is the first step. To make the model effective, these variables will be arranged in a hierarchical structure. We employ the Analytic Hierarchy Process (AHP) to evaluate the importance of each variable to the user. The advantage of AHP is that it can display complicated selection factors in simple concepts of hierarchy, which can be accepted easily by a decision-maker [6]. We have considered a wide range of variables, using subjective opinions. By using the AHP model, we are able to rank the weight of each variable for individual users in order to identify the best job for them. All job options will be tested against the basic model, and an index will be calculated for each job.

3.2 Entropy Method

However, the AHP model itself is susceptible to subjectivity. In conjunction with AHP, therefore, we also apply the Entropy Method to address the subjective aspect of our model. Using the quantitative variables we have collected about summer job options, including hourly rate, work time, and company size, we calculate a weight for each subset by deriving their entropy. We can then combine the summer job lists results from both AHP and Entropy Method to obtain a better subset of suitable jobs for the user.

The flow chart below shows a clearer logic:

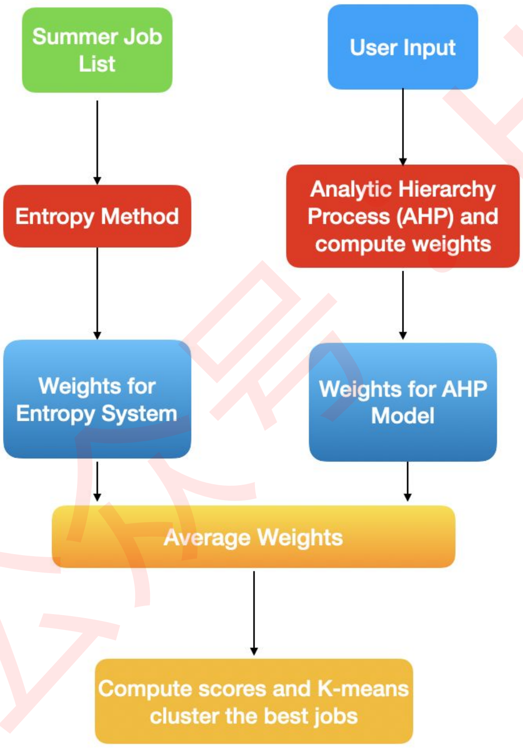


Figure 1: Logic flow for our basic model

3.3 Optimized Model

Considering that high school students usually lack a comprehensive and objective understanding of their true needs and situations, we must assume that the comparison matrix directly given by them would be very likely inconsistent. We thus develop an optimized model for modifying the given inconsistent matrix to be consistent (while maintaining the original information). This section will be discussed toward the end as part of optimization of our model.

4. The Basic Model for Considered Variables (Questions 1 and 2)

The most basic model for evaluating the best summer job is a function between a summer job's utility and other independent variables, including the job's income, comfort, type of job, company environment, relevance to intended major/career, age and ability requirements, and risks. In order to evaluate the significance of each, we adopt the Analytic Hierarchy Process (AHP) to derive the weight for them. In order to analyze our model, we first outline and explain the variables we use for the study, then we explain the AHP model, and finally apply it on fictional characters for testing.

4.1 Deterministic Variables

Certain variables in jobs predetermine whether high school students may participate in the job. These are the basic requirements for the job, and they must be met as prerequisites. Because we are primarily concerned with high school students, one important consideration is the age requirement for the job. If a student is too young to undertake some cumbersome physical activities, they cannot choose the job. Another possible deterministic variable is ability prerequisite. However, our team decides to put that as a probabilistic variable, since in our opinion, in the age of high school, students should do something that can challenge themselves, which a skill-required job is more valuable than jobs with no ability requirement.

4.2 Probabilistic Variables

In this section, we consider variables that do not predetermine the types of jobs that students can choose. They are among the most important variables for high school students based on secondary sources [7] and in our perspective as high school students.

- **Income**

For many high school students, the income of a job is a crucial factor. Students who confront financial difficulties at home especially need jobs that offer higher salaries. The total income earned through working at a job, M, can be defined as

$$M = R \times T$$

where R is the hourly rate (in dollars per hour) and T is the total work time (in hours). We collect data on our summer job options' income using a variety of sources online, including Investopedia, Indeed, and Zipecruiter. Note that work time could also be a deterministic variable, as there is an upper limit for the amount of time available for students.

- **Comfort and Type of Job**

Especially during COVID-19, the nature of the job (i.e. virtual or in-person) could be a critical factor for selecting a summer job. We consider three independent probabilistic variables under this category, including virtual or not, sedentary or not, and the location of the job. The first variable depends on the local pandemic situation as well as individual preference for safety. The other two variables may depend on the student's personality and economic needs.

- **Company characteristics**

Company's features are factors usually considered by high school students as well. One consideration is the size of the company, which heavily influences student's decisions, as some students may prefer individual work over teamwork. We vaguely define the size of the company, N, as the number of faculty employed.

- **Risks**

High school students likely want to minimize the danger associated with any summer job. One consideration may be the prevalence of COVID-19 in the region. According to Yonatan Grad, an epidemiologist at Harvard T.H. Chan School of Public Health,"The total incidence of [coronavirus infection] through 2025 will depend crucially on this duration of immunity"[8] Since this novel disease is unpredictable, we must take this into account as we are aiming at identifying the best summer job for students in 2021. As risks are often poorly defined, we adopt a Risk Score, RS, on a scale of 1-5 to assess the possibility for employees to encounter danger. One on the scale represents the lowest level of risks, which corresponds to jobs that are sedentary, indoor, or virtual. By contrast, works that require outdoor experience and first-aid experience are more likely to attain scale three. For all jobs, we would like to minimize RS. Risks of COVID-19 are also considered.

- **Skills**

High school students also intend to gain skills and experiences through summer jobs. Consequently, they want summer jobs that are relevant to their intended major or career. We define a skill index on a scale of 1-5 to determine how challenging the job is in terms of its ability requirement. For students looking for more advanced experience, jobs with higher ability index will fit them.

More compactly, the following is a schematic of the variables this study considers.

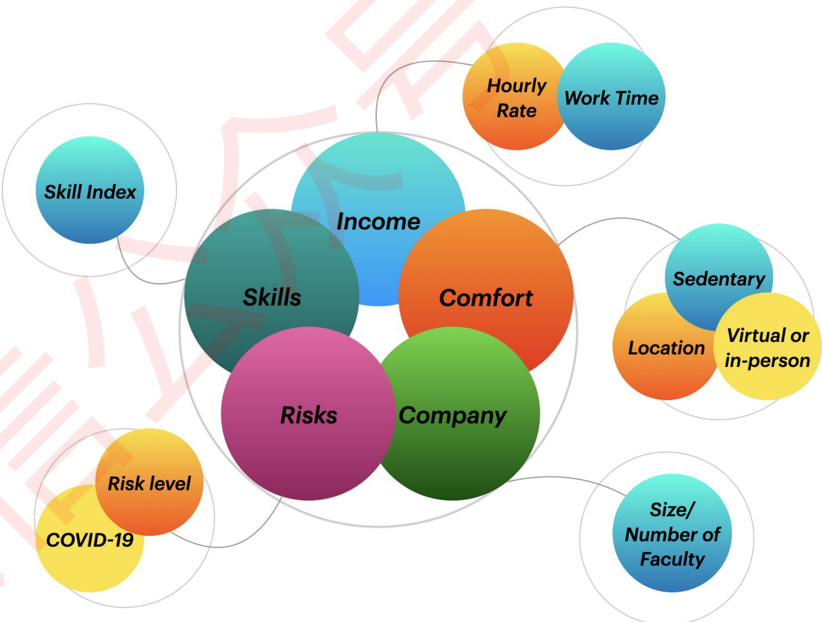
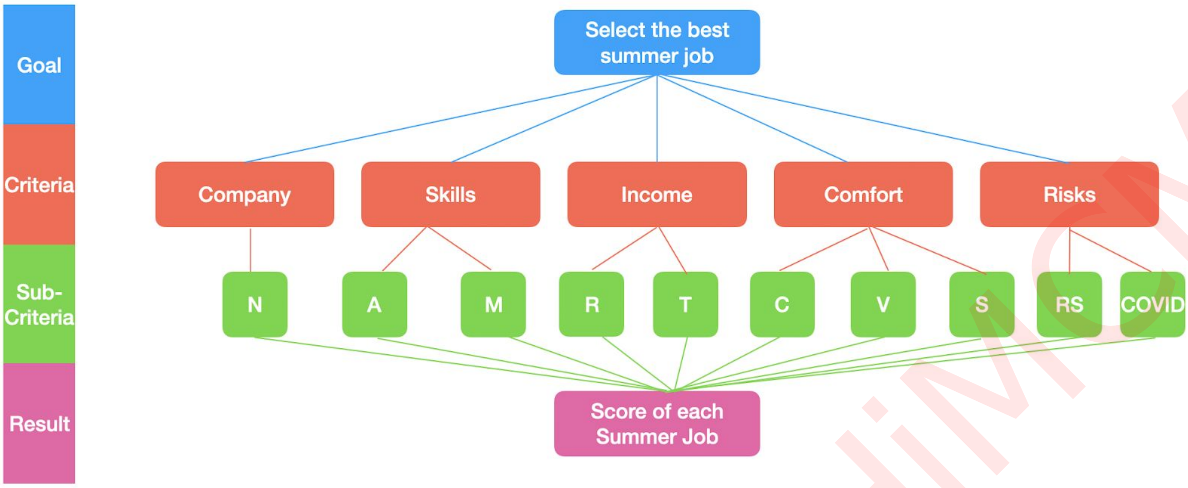


Figure 2: An overview of the variables this study considers

4.3 AHP Model

Having established the variables we need to consider, the following subsection will attempt to solve for the weights of each variable. This allows us to assess the relative importance of each variable, thereby letting us recommend the best summer job for each student user. In order to achieve this, the AHP model is an appropriate method. First developed by Thomas Saaty in the 1970s, the AHP model deals with a hierarchy of variables and can calculate the weights of each variable. A hierarchy structure of our variables is summarized below.

Figure 3: Hierarchy structure of variables



Variables Definition

Symbol	Definition
R	Hourly Rate of the summer job
T	Work time of the summer job
C1	Location of the company
V	Work completed virtually or in person
S	Work completed sedentary or not
N	Number of faculty/colleagues of the company
M	Relevance of the job to the student's future major
A	Ability requirement of the summer job
RS	Risk scale of the summer job
C2	Possibility of infecting COVID-19

AHP model effectively structures variables into a hierarchy. It is a way of assigning variables different weights based on the accompanied 1-9 scale table, which compares the importance of different factors. It determines the relative significance of each variable.

1-9 Scale Table

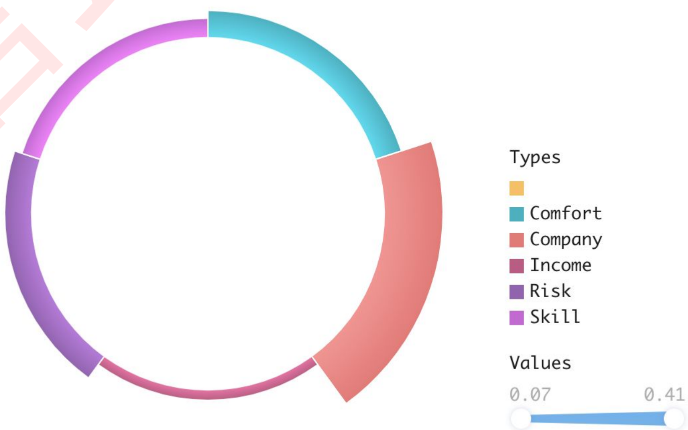
Scale	Definition
1	factor i is equally important as factor j
3	factor i is slightly more important than factor j
5	factor i is apparently more important than factor j
7	factor i is strongly more important than factor j
9	factor i is extremely more important than factor j
2, 4, 6, 8	the intermediate values
Reciprocal	importance scale between factor j and factor i

Based on the 1-9 scale table, it is possible to establish a comparison matrix for the variables. A comparison matrix is a square matrix of the following form. For each entry, C_{ij} represents the importance of C_i relative to C_j, as measured by the above scale table.

$$C=\begin{bmatrix} C & C_1 & C_2 & C_3 & \cdots & C_n \\ C_1 & C_{11} & C_{12} & C_{13} & \cdots & C_{1n} \\ C_2 & C_{21} & C_{22} & C_{23} & \cdots & C_{2n} \\ C_3 & C_{31} & C_{32} & C_{33} & \cdots & C_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ C_n & C_{n1} & C_{n2} & C_{n3} & \cdots & C_{nn} \end{bmatrix} \quad (3.1)$$

To evaluate the efficiency of the model, we construct a comparison matrix for a user whose primary concerns are comfort and skills gained rather than money. For this hypothetical example, we consider a high school student, Christopher, who values company and more than the income. The student's primary goal for the job is to expand his experience working in a large company. We accordingly compute her weights for the main criteria, as shown below:

Figure 4: Rose diagram of weights for criteria



We also compute weights for sub-criteria. The results are summarized below using the preference of Christopher:

Table 1. Comparison matrix and weights

Destination Layer	Criterion Layer	Weight	Sub-Criterion Layer	Weight	Final Weight
The most ideal summer job for high school students	Income	0.0735	R	0.667	0.0490
			T	0.333	0.0245
	Comfort	0.1903	C1	0.0836	0.0159
			V	0.4443	0.0845
			S	0.4721	0.0898
	Company	0.4115	N	1	0.4115
	Skill	0.1356	M	0.8	0.1084
			A	0.2	0.0271
	Risk	0.1892	RS	0.5	0.0946
			C2	0.5	0.0946

Furthermore, to confirm our assumption that the inputs are consistent and hence to verify that our fictional character is realistic, we consider the consistency of an n*n matrix with the greatest eigenvalue λ_{max}, which can be expressed as the following:

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{3.2}$$

After calculating CI, we further introduce RI and CR to determine if the level of inconsistency stays within the tolerance range of 0.1.

Table 2. Random consistency index

n	3	4	5	6	7	8
RI	0.58	0.9	1.12	1.24	1.32	1.41

$$CR = \frac{CI}{RI} < 0.1 \tag{3.3}$$

Applying these formulas on our fictional matrix using Matlab, we confirm that CR is always much lower than 0.1, meaning that our comparison matrix indeed is realistic and consistent.

4.4 Objective Weight System using Entropy Weight Method (Question 3)

Instead of solely depending upon the AHP method that gives us weight based on the inclination of each high schooler, we also need to consider the degree of dispersion of the evaluation index that will in turn give us a more objective decision upon the optimal jobs for the high school students. Therefore, we adapted another evaluation method called Entropy Weight Method (EWM) that will calculate the information entropy of each index and derive the weight.

Based on the summer jobs i, and the evaluation index j, we can build a judgement matrix: (n is the total number of jobs, and m is the number of index)

$$\left(\mathrm{X}_{ij}\right)_{nm} \quad (i=1,2,3\cdots n,\ j=1,2,3\cdots m) \tag{4.1}$$

In our model, we classify the indexes into two categories:

1) Increasing (As it increases, the job is better: eg. salary)
2) Decreasing(As it decreases, the job is better: eg. work time)

Each index will have a different unit of measure. Thus, it is necessary to first have a normalization of the jobs, and get a matrix that will calculate the entropy. The calculations of increasing and decreasing indexes are given in the equation below:

$$N_{ij}=0.998\frac{X_{ij}-min\left\{X_{1j},\ldots,X_{nj}\right\}}{max\left\{X_{1j},\ldots,X_{nj}\right\}-min\left\{X_{1j},\ldots,X_{nj}\right\}}+0.002 \quad \text{for increasing index} \tag{4.2}$$

$$N_{ij}=0.998\frac{max\left\{X_{1j},\ldots,X_{nj}\right\}-X_{ij}}{max\left\{X_{1j},\ldots,X_{nj}\right\}-min\left\{X_{1j},\ldots,X_{nj}\right\}}+0.002 \quad \text{for decreasing index} \tag{4.3}$$

After the normalization, we can calculate the weight p of the i-th plan under the j-th index through following formula:

$$p_{ij}=\frac{N_{ij}}{\sum_{i=1}^{n}N_{ij}}(j=1,2,\ldots,m) \tag{4.4}$$

Then the following can be used to derive the entropy of index j:

$$e_j=-k\sum_{i=1}^{n}p_{ij}\ln p_{ij} \quad where\ k=\frac{1}{\ln(n)},\ e_j\geq 0 \tag{4.5}$$

Based on the entropy, we can get the weight for each index

$$w_j=\frac{g_j}{\sum_{j=1}^{m}g_j} \tag{4.6}$$

where g_j is the redundancy:

$$g_j=1-e_j \tag{4.7}$$

Using the summer jobs list we define in the appendix, we can derive the weight of each evaluation index as the following:

Table 3. Weights derived from EWM

R	T	C1	V	S	N	M	A	RS	C2
0.0386	0.0677	0.1179	0.0918	0.3159	0.0363	0.1651	0.0678	0.0650	0.0339

4.5 Combining AHP and Entropy Method

As the AHP process and Entropy Method are both utilized, it is important to properly combine the two models to effectively reduce the subjectivity of the AHP model. In order to bolster the accuracy and reliability of our overall model, therefore, we use the below formula to minimize the deviation between the AHP model and the entropy system:

$$B = min \sum_{j=1}^{n} \left(\theta_1 w_{ahp} - \theta_2 w_{entr} \right)^2$$

$$s.t \ \ \theta_1 + \theta_2 = 1 \tag{5.1}$$

where w_{ahp} denotes the weight for the AHP model and w_{entr} denotes weight for the EWM. θ_1 and θ_2 are constants for taking deviation into account. Together using ordinary least squares, we could get the regression coefficient B that showed the ratio of w_{ahpj} to w_{entr}. Therefore, it is easy to show that by adding w_{ahp} and w_{entr} based on their proportion (u1&u2), we can get the final optimized weight w_f.

$$u1 = \frac{1}{1+B} \ , \ u2 = \frac{B}{1+B}$$

$$w_f = u1 \cdot w_{ahp} + u2 \cdot w_{entr} \tag{5.2}$$

As a result, we can get the final weight of each index based on both Christopher's personal preference and the objective evaluation of the index.

Table 4. Weights derived from EWM

R	T	C1	V	S	N	M	A	RS	C2
0.0455	0.0391	0.0505	0.0870	0.1665	0.2842	0.1276	0.0409	0.0845	0.0740

4.6 Results and Analysis

Based on Table 1, we derive the weights corresponding to the illustrative example. We now define the index score for each summer job according to the following formula:

$$f(x) = w_1 R + w_2 T + w_3 C1 + w_4 V + w_5 S + w_6 N + w_7 M + w_8 A + w_9 RS + w_{10} C2$$

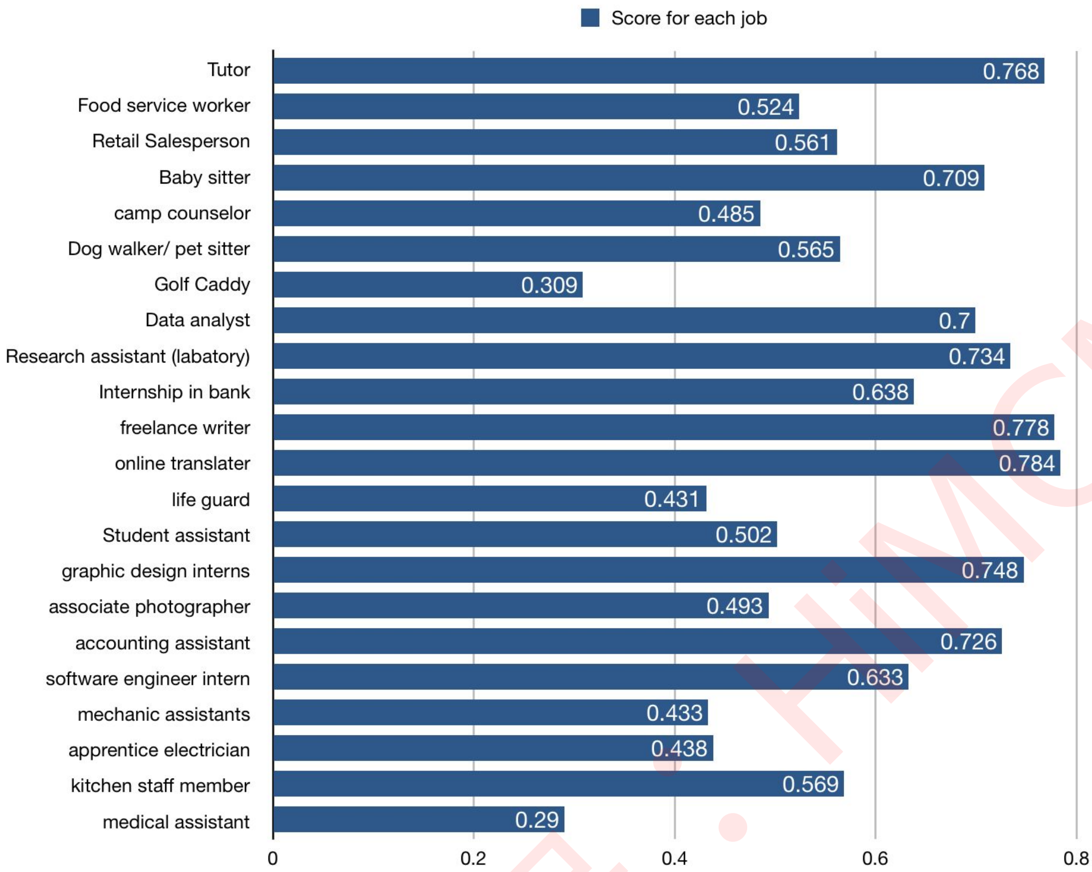
or in a more condense way:

$$s_i = \sum_{j=1}^{m} w_j N_{ij} \tag{6.1}$$

where w_j is the final weight for indexes j, and N_{ij} is the normalized value of indexes j for job i.

Accordingly, the higher the final score is, the better the summer job is for the individual. For our summer job list (see Appendix B), we compute that the job with the highest index score is an online translator, which has an index score of 0.784. The values of all summer jobs calculated from our model for Christopher can be seen in the graph below:

Figure 5. Job Scores for Christopher



4.7 K-Means Clustering

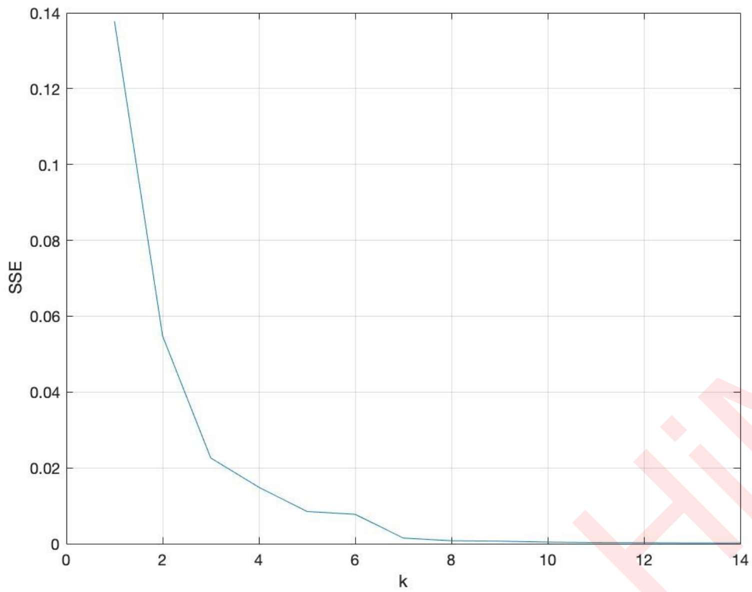
Oftentimes, we want more than the one single best job. Rather, we want a cluster of possible options. After using the AHP model and Entropy Weight method to assess the ideal summer job for a student, we use K-means clustering to further determine the exact classification of the student's ideal summer job. The advantages of K-means are its simplicity of implementation, scalability, speed of convergence and adaptability to sparse data [9]. Using the Euclidean Distance formula as the criteria for comparing similarity and distance between value K and each point, we apply SSE (sum of squared errors) to determine the exact value of K.

$$d(a,b)=\sqrt{(a_1-b_1)^2+(a_2-b_2)^2}$$

$$SSE=\sum_{i=1}^{k}\sum_{d\in G_i}\left(\left|d-m_i\right|\right)^2 \tag{7.1}$$

When k, the number of clusters, increases, the clusters will be divided into more precise groups that have higher clustering level. Thus SSE will gradually decrease. Furthermore, when k is less than the real optimal number of clusters, SSE decreases substantially since an increase in k will significantly increase the aggregation of the clusters. However, when k reaches the true cluster number, the return of the aggregation level of clusters will become rapidly smaller. Hence the decline of SSE is dramatically reduced, moving closer toward constant as higher k values. Below is the graph of SSE versus k using the final scores of summer jobs previously presented:

Figure 6. Relationship between SSE and K-values



From the graph above, it is evident that SSE decreases with the increases in k value, and the change of gradient is the greatest from k=2 to k=3, thus we choose k=3 as the final number of clusters. After applying the K-means clustering, we get the recommended list of Christopher's summer list, which is the cluster that has the highest center. Finally, we use the deterministic variables mentioned earlier to eliminate undesirable jobs which in this case is the age requirement. Following is the final outcome of recommended jobs for Christopher:

Figure 7. Scatter-plot for clusters of Christopher's summer jobs

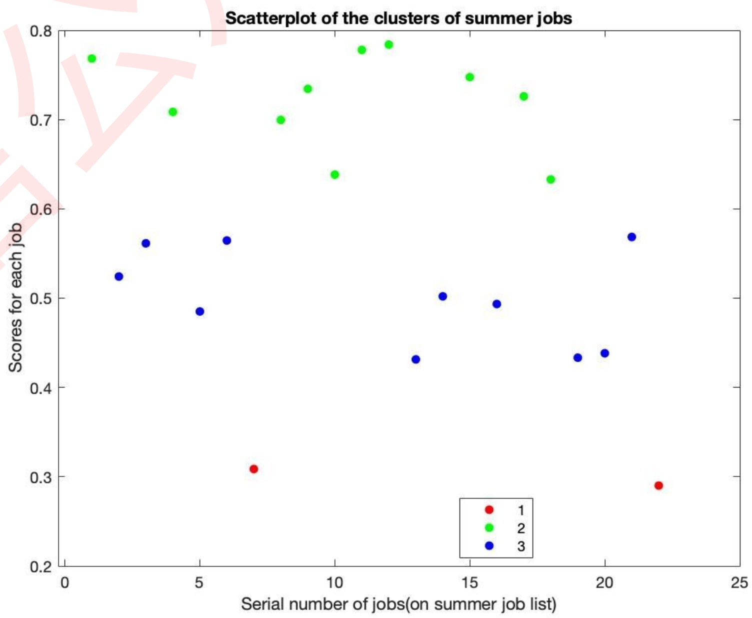


Table 5. Best summer job cluster for Christopher

Christopher's recommended job list	
Tutor	freelance writer
Baby sitter	online translater
Data analyst	graphic design interns
Research assistant (labatory)	accounting assistant
Internship in bank	software engineer intern

Indeed, this job list supports our initial hypothesis about Christopher, who values experience and company more than income. Jobs such as data analyst, research assistant, and internship in a bank are the kinds of jobs that Christopher values, as they are often associated with larger companies and more advanced skills. This in turn validates our model.

5. Testing the Models (Question 3)

In the previous section, we explain our development of basic models and their quick application on one fictional individual Christopher. In this section, nine other representative fictional characters are considered and tested against the previously developed mathematical models. While it is impossible to cover the entire high school student population demographic and preference using only ten characters, we develop our subjects to be as consistent and diverse as possible, allowing us to better examine the usefulness of our model. We first provide our rationale and descriptions for the fictional characters and then explain the model's efficacy.

5.1 Fictional Characters Development

Our overarching rationale for the fictional characters is to make them as representative as possible. High school students have different socioeconomic backgrounds, abilities, ages, and personal goals for doing summer jobs. In developing our fictional characters' profiles, we consider how different students may place different relative importance on various variables. The comparison matrices and weightings for each variable for our fictional characters are summarized by Appendix A. Below we provide a basic profile for each fictional character (numbers in the parentheses indicate their age).

1. Anna (16) is a high school student interested in making more money. She doesn't really care about the type of company she will be working for. She does pay attention to her safety and comfort during the job, in addition to income.
2. Betty (17) is a high school junior whose primary concern is comfort. Her second largest focus is income. Nevertheless, she doesn't mind how skillful or large her job/company would be.
3. Christopher (Sample case described in the previous sections)(18) has a keen focus on company. He is not particularly interested in the income or skill that he will gain; it is the experience that appeals to him.

4. David (15) pays a great deal to risks. Given the unprecedented circumstances of COVID-19, his top priorities are safety and comfort. He does not care about how much he will earn or what company he works at -- safety comes first!
5. Erik (18) is an advanced high schooler interested in advancing his skills. At the same time, he would also like to make more money for himself. As a hard-working student, his considerations for risks and comforts are minimal.
6. Fiona (17) also focuses on the company she would work at. But she also wants to earn sufficient money for her summer fun time. Fiona isn't really worried about the skills she may gain; her emphasis goes to company, income, and comfort.
7. Gary (16) is a high school student who focuses on the job's risks and safety. His secondary consideration is the size of the firm. And frankly, he doesn't care too much about other things he will gain. He just wants a safe job for the summer.
8. Henry (18) prioritizes income and company. In fact, due to his poor socioeconomic background, his emphasis on income is the highest among all our fictional characters. He is relatively unconcerned about the risks associated with the job, as long as he can earn money for his family.
9. Iris (15) is a high school student who focuses on skill and comfort of the job. For her, income isn't essential. Gaining skills from a big company is what she wants.
10. Jacky (17) would love to make more money over the summer. His primary consideration is thus income. He values his secondary considerations, risks and skills, equally. While Jacky wants to maximize his income, however, he doesn't care too much about the size of his company.

5.2 Results and Analysis

Similar to how we apply the AHP model, entropy system, and deterministic variables on the example character Christopher, we now apply the same models on these characters to test the results. The results are summarized below:

Table 6. Best summer job clusters for all fictional characters

Christopher(18)	**Anna(16)**	**Betty(17)**	**David(15)**	**Erik(18)**
Tutor	Tutor	Tutor	Tutor	Data analyst
Baby sitter	freelance writer	Data analyst	Baby sitter	Research assistant (labatory)
Data analyst	online translater	Research assistant (labatory)	online translater	graphic design interns
Research assistant (labatory)	graphic design interns	Internship in bank		
Internship in bank	software engineer intern	freelance writer		
freelance writer		online translater		
online translater		graphic design interns		
graphic design interns		accounting assistant		
accounting assistant		software engineer intern		
software engineer intern				
Fiona(17)	**Gary(16)**	**Henry(18)**	**Iris(15)**	**Jacky(17)**
Tutor	Tutor	Research assistant (labatory)	Tutor	Research assistant (labatory)
Research assistant (labatory)	Baby sitter	freelance writer		Internship in bank
freelance writer	freelance writer	online translater		graphic design interns
online translater	online translater			software engineer intern
	graphic design interns			

It is interesting to notice that there is a huge similarity between these ten fictional characters. For most of them, the jobs in the recommending clusters are tending to have a relatively high ability index. This is because we consider the ability requirement as part of our evaluation indexes which jobs that have a high ability requirement would much likely to appear in the final recommendation. In addition, we can conclude that for jobs that appear several times like "tutor," "online translator," and "freelance writer," they should become the top-list summer jobs for a typical high school student.

6. The Optimized Model

The optimization of our above model focuses on the input information from our high school student user. Given that our target audience are high school students, it is unrealistic to expect them to compute their comparison matrix and input it completely accurately. It is very likely that the resulting input may be inconsistent; this will force the user to reenter the information, causing inconvenience.

In order to address comparison matrices that do not meet the consistency criterion, we will modify these to make them consistent based on the original information given.

6.1 Model Construction

We will first introduce the algorithm of our optimized model. Define the cosine value of the angle between two n-dimensional vectors, $T=\left(a_1,a_2,\cdots,a_n\right), U=\left(b_1,b_2,\cdots b_n\right)$, as the following:

$$\cos\theta=\frac{\langle T,U\rangle}{|T||U|}=\frac{\sum_{i=1}^{n}a_ib_i}{\sqrt{\left(\sum_{i=1}^{n}a_i^2\right)\left(\sum_{i=1}^{n}b_i^2\right)}} \tag{1.1}$$

where $\langle T,U\rangle$ denotes the inner product of the two vectors T and U. Therefore, any two vectors correspond to one unique angle. Let A be some comparison matrix, such that its normalized matrix $\overline{T}=\left(T_1,T_2,\ldots,T_n\right)$ is , where $T_i, i=1,2,\ldots,n$ are n-dimensional vectors. Therefore, the the normalized eigenvector of matrix C can be represented as:

$$\overline{T}=\frac{1}{n}\sum_{i=1}^{n}T_i \tag{1.2}$$

If C is completely identical, then by definition $\overline{T}=T_i, i=1,2,\ldots,n$. If C is not completely identical, then each T_i can be related to $\overline{T}$ with a similarity proportion, which can be measured by the angle between the vectors. Since inconsistency in the comparison matrix is mostly caused by stochastic disturbance, so T_i distributes around $\overline{T}$, the center. If we consider $\overline{T}$ as the ideal estimation value, then we define θ_i for $i=1,2,\cdots n$ as the angle between T_i and $\overline{T}$.

Subsequently, we define the following quantity:

$$r_i=\cos\theta_i=\frac{\langle T_i,U\rangle}{|T_i||U|},\ v_i=\sqrt{1-r_i^2} \tag{1.3}$$

where r_i is the consistency factor of T_i relative to $\overline{T}$; v_i is the deviation of T_i from $\overline{T}$. Hence, when $v_i < v_j$, would T_i have a better consistency than T_j. When a comparison matrix fails to pass the consistency test, therefore, the terms with higher deviation need to be adjusted first.

Essentially, the adjustment method and algorithmic steps can be summarized as below:

1. Normalize all comparison matrices that are not satisfactory or consistent.
2. Compute the consistency factor r_i and deviation v_i between each column vector and the normalized eigenvector.
3. Based on the column vector with the smallest deviation, adjust the proportion of reciprocal elements in the two column vectors with the largest deviations.
4. The process reiterates and tests the new matrix's consistency until the result is satisfactory enough.

6.2 Application of Optimized Model

In this subsection, we apply the above optimized model on a simplified hypothetical case to test its effectiveness. Consider the following matrix C

$$C = \begin{bmatrix} 1 & 3 & 5 \\ \frac{1}{3} & 1 & \frac{1}{2} \\ \frac{1}{5} & 2 & 1 \end{bmatrix}$$

It has a CR=0.1407>0.1, failing to pass the consistency test. Utilizing the aforementioned method for optimization, we normalize the matrix to get

$$C = \begin{bmatrix} 0.652 & 0.500 & 0.769 \\ 0.217 & 0.167 & 0.077 \\ 0.130 & 0.333 & 0.154 \end{bmatrix}$$

We can then compute the following:

$$T_1 = \begin{bmatrix} 0.652 \\ 0.217 \\ 0.130 \end{bmatrix}, T_2 = \begin{bmatrix} 0.500 \\ 0.167 \\ 0.333 \end{bmatrix}, T_3 = \begin{bmatrix} 0.769 \\ 0.077 \\ 0.154 \end{bmatrix}, \overline{T} = \begin{bmatrix} 0.640 \\ 0.154 \\ 0.206 \end{bmatrix}$$

Based on these vectors, we calculate the deviation and the consistency factors:

$$r_1 = 0.990, r_2 = 0.961, r_3 = 0.982$$

$$v_1 = 0.141, v_2 = 0.277, v_3 = 0.189$$

Since $v_1 < v_3 < v_2$, T_1 has the best consistency, while T_2 has the worst consistency. Using T_1 as our caliber for adjustment, we can then derive the following adjusted matrix:

$$C = \begin{bmatrix} 1 & 3 & 5 \\ \frac{1}{3} & 1 & 1 \\ \frac{1}{5} & 1 & 1 \end{bmatrix}$$

This gives a CR=0.025055<0.1, which now passes the consistency criterion. Therefore, by applying this optimized model that is capable of automatically improving the matrix's consistency, it is evident that we could significantly reduce the need for users to re enter their information.

7. Evaluation of Model

In conclusion, our model incorporates the AHP model, entropy method, K-Means clustering, and an optimized model to help recommend the best summer jobs to high school students. Some of our model's overall strengths and weaknesses are summarized below:

7.1 Stengths

- While the problem does not provide any data, our model incorporates real life data we have collected from over 20 different jobs. This makes the model realistic.
- Our model is fairly comprehensive, as it encompasses five categories of variables.
- Our fictional characters are quite representative of the US high school population.
- We utilize Entropy Method to offset the subjectivity in our AHP model, ensuring the objective results to be objective.
- We consider the high probability that high school students may enter information that is inconsistent; in response, we develop an optimized model that can make the matrix consistent that's still based on original information.

7.2 Weaknesses

- The AHP model still has subjectivity that may negatively impact our model.
- We did not consider all the summer job options available to high school students.
- Our model only generalizes to the United States; we also did not consider geographical differences at a smaller regional scale for each fictional individual.
- For indexes that can either be increasing or decreasing based upon each student's reference, we only provide one possibility for our indexes. For instance, in our model, we only define the size of a company as an increasing index, but for some people, they think a small company is better which corresponds to a decreasing index.

8. References

[1] Investopedia. https://www.investopedia.com/.
[2] Indeed. https://www.indeed.com/about.
[3] Zipecruiter. https://www.ziprecruiter.com/.
[4] Payscale. https://www.payscale.com/.

[5] Desilver, Drew. "In the U.S., teen summer jobs aren't what they used to be." https://www.pewresearch.org/fact-tank/2019/06/27/teen-summer-jobs-in-us/.
[6] Sari, B., Sen, T., & Kilic, S. E. (2007). Ahp model for the selection of partner companies in virtual enterprises. The International Journal of Advanced Manufacturing Technology, 38(3-4), 367-376. doi:10.1007/s00170-007-1097-6
[7] Lash, M., Wolcott, G., & Kahn, P. (1997). When Your Teenager is Injured: Preparing for Work and Adulthood. Research and Training Center in Rehabilitation and Childhood Trauma.
[8] Kissler, S., Tedijanto, C., Goldstein, E., Grad, Y., & Lipsitch, M. (2020, May 22). Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Retrieved November 16, 2020, from https://science.sciencemag.org/content/368/6493/860
[9] International Journal of Computer Science and Information Security, IJCSIS, Vol. 7, No. 1, pp. 292-295, January 2010, USA

9. Appendix

9.1 Appendix A

Comparison matrices for the 10 fictional characters (sorted in alphabetical order):

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.3049	0.2091
T	0.0436	0.0523
C1	0.1359	0.1294
V	0.0322	0.0536
S	0.1146	0.1870
N	0.0846	0.0672
M	0.0563	0.0955
A	0.0563	0.0605
RS	0.0571	0.0599
C2	0.1144	0.0854

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.0886	0.0675
T	0.1775	0.1313
C1	0.0585	0.0835
V	0.2154	0.1633
S	0.1587	0.2249
N	0.0526	0.0457
M	0.1325	0.1462
A	0.0221	0.0414
RS	0.0188	0.0382
C2	0.0753	0.0579

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.0693	0.0599
T	0.0173	0.0327
C1	0.0372	0.0619
V	0.2162	0.1781
S	0.0224	0.1123
N	0.0513	0.0467
M	0.0997	0.1197
A	0.0498	0.0553
RS	0.0624	0.0632
C2	0.3743	0.2701

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.1921	0.1453
T	0.0640	0.0651
C1	0.0140	0.0457
V	0.0280	0.0475
S	0.0280	0.1159
N	0.1571	0.1202
M	0.0874	0.1111
A	0.3497	0.2637
RS	0.0598	0.0614
C2	0.0199	0.0242

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.2267	0.1787
T	0.0283	0.0383
C1	0.0275	0.0506
V	0.0794	0.0825
S	0.0191	0.0948
N	0.4403	0.3373
M	0.0503	0.0796
A	0.0072	0.0227
RS	0.0303	0.0391
C2	0.0908	0.0763

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.0167	0.0227
T	0.1340	0.1159
C1	0.0551	0.0722
V	0.0101	0.0324
S	0.0083	0.0921
N	0.2424	0.1862
M	0.0629	0.0908
A	0.0157	0.0299
RS	0.2273	0.1830
C2	0.2273	0.1746

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.3823	0.2851
T	0.0425	0.0496
C1	0.0863	0.0953
V	0.0241	0.0432
S	0.0456	0.1221
N	0.2687	0.2030
M	0.0316	0.0694
A	0.0633	0.0646
RS	0.0062	0.0228
C2	0.0494	0.0450

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.0206	0.0271
T	0.0206	0.0375
C1	0.0240	0.0578
V	0.1256	0.1134
S	0.1097	0.1838
N	0.1231	0.0919
M	0.0465	0.0892
A	0.3254	0.2328
RS	0.0409	0.0495
C2	0.1637	0.1170

Sub-Criterion Layer	AHP Weight	Final Weight
R	0.1087	0.0872
T	0.4349	0.3225
C1	0.0105	0.0434
V	0.0552	0.0664
S	0.0482	0.1302
N	0.0748	0.0630
M	0.0167	0.0622
A	0.1171	0.1020
RS	0.0149	0.0302
C2	0.1189	0.0929

9.2 Appendix B

Summer Job List (22 sample jobs) for EWM and scoring analysis

Job/position	number of faculty	Ability index(scal	Relevance	Hourly rate ($USD	Work Time(hr/day)	Distance(1-5)	Indoor/outc	Sedentary	Risk level(1-5 basec	Covid	age requirement	Ability requirement	Citation
Tutor	1	4	5	15	2	2	1	0	1	1	14	Depends on the subject	investopedia
Food service worker	12	1	2	9.43	4	2	1	1	3	2	16	Communication	indeed
Retail Salesperson	15	2	1	12.6	4	2	1	1	2	3	16	Communication(may also relates to sp	indeed
Baby sitter	1	3	1	12.5	3	1	1	1	1	1	14	Patience, nurturing skills	investopedia
camp counselor	6	4	3	9.16	8	4	0	1	3	4	18	Camp experience, communication, lea	indeed
Dog walker/ pet sitter	1	2	2	15.65	1	4	0	1	2	2	14	petting exp, maybe animal first aid ski	indeed
Golf Caddy	40	3	3	30	8	5	0	1	4	2	16	knows about the rules of golf	nytimes
Data analyst	12	5	5	22.82	4	1	1	0	1	2	17	cs skills, comm	zipecruiter
Research assistant (labat	10	5	4	22	8	3	1	0	3	2	17	comm, specific subject knowledge, cs	zipecruiter
Internship in bank	26	4	4	0	7	3	1	0	1	2	17	Math, relecant economic knowledge	none
freelance writer	1	3	4	23.9	3	1	1	0	1	2	14	writing skill	payscale
online translater	1	2	4	26.06	2	1	1	0	1	1	14	foreign language	indeed
life guard	3	2	3	9.5	3	2	0	1	3	4	15	requires liscense, swimming	investopedia
Student assistant	8	1	5	11.24	3	3	1	1	1	3	16	comm	indeed
graphic design interns	10	5	4	12.92	5	2	1	0	1	2	16	adobe skills, design(web, logo, etc) sk	indeed
associate photographer	7	5	4	12.55	3	5	0	1	2	3	16	photography, editing	indeed
accounting assistant	8	3	5	16.4	5	3	1	0	1	2	17	accounting(data entry, etc.) skills, bas	indeed
software engineer intern	25	4	5	25.19	6	2	1	0	1	2	17	programming lang	indeed
mechanic assistants	24	4	4	14.96	6	2	1	1	3	2	17	mechanic exp(repairing, diagnostic, et	indeed
apprentice electrician	17	5	5	17.61	7	2	1	1	4	2	17	electrical exp, comm, construction exp	indeed
kitchen staff member	8	3	2	12.05	4	3	1	1	3	3	16	kitchen exp(knife, cooking, serving, sa	indeed
medical assistant	45	5	4	15.46	3	4	1	1	2	5	17	medicine knowledge, comm, clinical e	indeed
							I=1	y=0					
							O=0	n=1					

9.3 Appendix C

Code for K-means clustering:

```
x=[0.768348538949563
0.524220820934521
0.561391156339757
0.708583630045647
0.485091292157426
0.564557638411574
0.308605613248477
0.699563412343942
0.734345019866725
0.638368468717021
0.777964137990791
0.783905630087951
0.431416564593624
0.502010468564320
0.747513370773647
0.493417855810747
0.726007314870837
0.632889371599188
0.433315295888702
0.438273370474133
0.568528746331102
0.290076112273174];%input=scores of job for an individual

SSE=compute(x,15); %determining the best k value
plot(SSE)

y=["Tutor"
    "Food service worker"
    "Retail Salesperson"
    "Baby sitter"
    "camp counselor"
    "Dog walker/ pet sitter"
    "Golf Caddy"
    "Data analyst"
    "Research assistant (labatory)"
    "Internship in bank"
    "freelance writer"
    "online translater"
    "life guard"
    "Student assistant"
    "graphic design interns"
    "associate photographer"
    "accounting assistant"
    "software engineer intern"
    "mechanic assistants"
    "apprentice electrician"
    "kitchen staff member"
    "medical assistant"];%summer job list
[labels,c]=kmeans(x,3); %clustering

a=find(c==max(c));%find the jobs that are in the most desirable cluster
b=find(labels==a);

d=[]; %list of recommended jobs
for i=1:size(b)
d=[d,y(b(i))];
end
d=d'


y1=1:size(y); %graphing scatterplot
gscatter(y1,x,labels);
xlabel("Serial number of jobs(on summer job list)")
ylabel("Scores for each job")
title("Scatterplot of the clusters of David's summer jobs")
hold on
gscatter(c);
```