14th International Conference on Innovative Computing, Information and Control (ICICIC2019)
August 26-29, 2019, Seoul, Korea

Mining World Talent Indicators among OECD Economies Based on Gini Index

Dian-Fu Chang[1] and Chia-Chi Chen[2,*]

[1]Graduate Institute of Educational Policy and Leadership, Tamkang University, Taiwan
[1]140626@mail.tku.edu.tw
[2]Doctoral Program of Educational Leadership and Technology Management, Tamkang University
Taiwan
[2]sophiabv03@gmail.com
* Corresponding author

ABSTRACT. *This study selected OECD 35 economics as the research target to tackle the equality with their World Talent survey data. The data with three different catalogues from 2014 to 2018 were transformed by Gini index to realize the distribution change among the economics. For further comparison among the indicators, this study proposes cluster analysis with Minitab to determine the optimal clusters for interpretation. The result reveals Gini index can be used to transfer the classical indicator data with equality reasons properly. The dendrogram reveals two clusters among the economics by using Ward linkage and Euclidean distance. The findings provide meaningful information for further interpretation of the data set.*

Keywords: Cluster analysis, Data mining, Gini index, OECD, World Talent Ranking

1. Introduction. This study focuses on the data set collected by the International Institute for Management Development (IMD). The data have been applied widely to realize the countries' performance by selected three categories- investment and development, appeal and readiness to publish annual World Talent Report. The report also called the IMD World Competitiveness Yearbook covers 63 countries in 2018. The selected three factors assess how countries perform in the specific areas. Various knowledge from detailed data such as the education, apprenticeships, workplace training, language skills, cost of living, quality of life, remuneration and tax rates have been included. The data response to the IMD executive opinion survey is used to provide the ranking. Even though the ranking system has been accepted by public generally, the report is hard to tell how wide the discrepancy existed among the countries. In this sense, this study takes the IMD's data set as an example to tackle the meanings of their discrepancy. This study applied the notions of data mining to explore the issue. Data mining has been attracting a significant amount of research, industry, and media attention recently. Data mining is also becoming a major tool for analyzing th large amounts of data, by using a data warehouse and analyzing web data or enables organizations calculated decision by assembling accumulating, analyzing and accessing corporate data [1-2]. Data mining can be used to identify inconsistencies in the composition and integration of data that comes from different sources of information [3]. It is also considering the process of discovering meaning correlations, patterns, and trends by shifting through amounts of data stored in repositories [1]. This technique is also being a convergence of multiple fields for various applications, including not limited to bioinformatics, medical informatics, consumer profiling, intrusion detection, security, web mining, etc. [1-3].

Cluster analysis is a popular statistical method and clusters are formed such that objects in the same cluster are very similar, and objects in different clusters are very distinct. Cluster evaluation determines the optimal number of clusters for the data using different evaluation criteria in diverse settings and could apply to decide the level or scale of clustering that is most appropriate for the data application [4].

According to the IMD World Talent Ranking Reports 2014 to 2018, the trends show Western Europe leads, Eastern Europe lags and North America gives strong performance [6-10]. In the latest 2018 report, it also reveals Switzerland once again confirms its role as an important global talent center. Several European countries fall within the 25 most competitive with respect to talent. As the other ranking countries, such as Belgium (11th), Cyprus (15th), Portugal (17th), Ireland (21st), United Kingdom (23rd), and France (25th), these countries just exactly refer to the Organization for Economic Co-operation and Development (OECD) economies are named. This study selected OECD economies as the research target [11]. There are 35 member countries in OECD from North and South America to Europe and Asia-Pacific. This study focuses on the 35 countries, which usually called rich countries in the world, to detect the equality of their completion in the stock of talents. Specifically, the purposes of this study are as follows:

a. to realize the patterns of world talent competitive among the OECD economies;

b. to compare the Gini index result and cluster result to interpret the OECD economies' World Talent competitive to check their similarities.

Given these purposes, the structure of this paper is as follows: Section 1 calculates the Gini index. Section 2 displays the result of cluster analysis. Finally, the conclusions are displayed.

2. Method. In this study, we assumed once the data been mining, the patterns with trends and the main impacted factors could be explored much through among the OECD economies. In the method section, we addressed how the Gini indices have been used to analyze 100 percent of the IMD World Talent Ranking Report overall scores, then go through the other three impact factors to determine the inequality of the world talent competitiveness. This study conducted cluster analysis for the 35 OECD economies to deepen the meanings of the world talent competitive results. The details of the methods will be demonstrated in the following sections.

2.1 Definition of World Talent data set. The World Talent Ranking report is an annual survey that put together by the IMD. IMD examined 63 countries using the surveys with over 6,000 company executives as well as economic data [6-10]. All the participated countries were given an overall score based on how they define talent competitiveness into the three main factors, namely investment and development, appeal, and readiness. First, investment and development factor measures the resources dedicated to homegrown talent. The second factor is measured how well a country does in attracting foreign talents and retain local talents. The readiness factor shows the quality of skills and jobs available in a country. The methodology of the World Talent Ranking has shown well defined in the Talent Competitiveness with three factors independently. It contains the same weight in the overall consolidation of results in terms of 33.33 % for each (total is 3x33.33%=100%). The countries' rankings consist of hard and soft data, for example, competitiveness refers to the hard data with the statistics from international regional sources, while the soft data based on the international panel of experts and executive opinion survey. The 2018 World Talent Ranking result has been presented in Table 1

TABLE 1. The World Talent Ranking 2018 overall and each factor scores for 35 OECD economies

	Country(code)	Overall Score	Investment& Development Factor	Appeal Factor	Readiness Factor
1	Switzerland(CHE)	100	83.93	100	90.9
2	Denmark(DNK)	91.97	97.96	75.34	77.42
3	Norway(NOR)	86.37	85.97	72.95	75.02
4	Austria(AUT)	86.1	91.76	71.92	69.44
5	Netherlands(NLD)	85.25	70.1	74.85	85.63
6	Canada(CAN)	84.5	65.32	80.31	82.7
7	Finland(FIN)	83	82.45	63.27	78.09
8	Sweden(SWE)	82.45	76.77	74.94	70.45
9	Luxembourg(LUX)	81.63	66.81	78.68	74.2
10	Germany(DEU)	81.11	75.09	75.67	67.39
11	Belgium(BEL)	80.54	77.67	67.28	71.49
12	United States(USA)	79.22	62.22	83.4	66.87
13	Australia(AUS)	78.57	62.63	65.26	82.63
14	Iceland(ISL)	77.21	72.89	64.6	68.95
15	Portugal(PRT)	76.76	78.35	59.83	66.92
16	Israel(ISR)	75.86	70.66	61.39	70.34
17	New Zealand(NZL)	74.12	58	66.95	72.22
18	Ireland(IRL)	73.93	49.72	73.3	73.6
19	United Kingdom(GBR)	72.63	55.98	66.92	69.8
20	France(FRA)	70.85	63.61	62.97	60.78
21	Estonia(EST)	67.92	69.85	54.22	54.5
22	Japan(JPN)	64.95	63.26	59.83	46.59
23	Slovenia(SVN)	64.69	62.51	46.63	59.75
24	Spain(ESP)	63.34	56.52	61.2	47.13
25	Italy(ITA)	62.42	57.6	52.2	52.29
26	Korea(KOR)	62.32	63.78	46.71	51.29
27	Latvia(LVA)	61.67	73.18	41.46	45.19
28	Czech Republic(CZE)	61.02	55.54	52.7	49.65
29	Poland(POL)	60.81	63.21	46.91	47.15
30	Chile(CHL)	55.07	30.88	58	51.14
31	Greece(GRC)	54.98	59.98	38.53	41.25
32	Hungary(HUN)	47.76	54.22	30.15	33.74
33	Turkey(TUR)	45.94	28.49	43.16	40.99
34	Slovak Republic(SVK)	39.63	40.36	31.9	21.43
35	Mexico(MEX)	38.86	11.97	46.62	32.81

2.2 Transformation of Gini indices. The Gini Index also called Gini coefficient or Gini ratio. Originally, it is a valid index for measuring the extent of income inequality. The value of the Gini index varies from 0 (representing perfect income equality) to 1 (representing perfect income inequality) [12]. The index was calculated as a ration of the areas on Lorenz curve and using the Lorenz Curve to measure the Gini index [12-10]. For example, if the area between the line of perfect equality and Lorenze curve is A, and the area under the Lorenze curve is B. The basic formula of the Gini index will be A/(A+B). Meanwhile, since A+B=0.5, the Gini index, G=2A=1-2B. For a discrete probability function *f*(*y*), let y_i=1 to *n*, it denotes the points with nonzero probabilities which is indexed in increasing order (y_i<y_{i+1}). When we calculated Gini index, it can be integrated to the following format:

$$G = 1 - \frac{\Sigma_{i=1}^{n} f(y_i)(s_{i-1}+s_i)}{s_n}$$

Where, we sort the *y* variable values in an increasing sense and present $\Sigma_{i=1}^{n}(y_i)$, with *i* =1 and s_i=$\Sigma_{i=1}^{n}(y_i)$, which represents the sum of the first ordered y variable values.

Usually, the accepted standard was defined as Gini ≤0.2 represents absolute equality, 0.2 < Gini ≤0.3 means low inequality in the study, 0.3 < Gini ≤0.4 indicates medium inequality, and 0.4 < Gini ≤0.5 means high inequality. When the value larger than 0.5 means very high inequality [12-15]. Moreover, a Gini index above 0.4 is often seen as a crucial point to justify the inequality. Inequality above this level is frequently associated with political instability or growing social tensions.

2.3 Logic of cluster analysis. Cluster analysis also called segmentation analysis or taxonomy analysis for partitions sample data into groups or clusters. The related analyses can be found in a variety of fields, such as psychology and other social science, biology, statistics, pattern recognition, information retrieval machine learning, and data mining [16-17]. Clusters are calculated by similarity or distance of objects with different characteristics. Previous studies have provided various examples for conducting cluster analysis [16-17]. This study follows the following steps: First, selecting the data; Then, hierarchical clustering with Minitab statistic package to determine the clusters. Basic cluster algorithms are as follows:

- Select k point as initial centroids,
- Repeat,
- From *k* clusters by assigning each point to its closest centroids,
- Re-compute the centroids of each cluster,
- Until centroids do not change.

Typically, hierarchical clustering groups data over a variety of scales by creating a cluster tree or dendrogram. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. The dendrogram function plots the cluster tree. Based on the dendrogram, which displays both the cluster and sub-cluster relationships and the order in which the clusters were merged or spilt [16-17]. Based on the information, this study decides the level or scale of clustering that is most appropriate for the data application. The Ward method considers the minimum variability as the criterion for merging to form the within-cluster sum of squares is minimized. It indicates that the similarity within the group is high. The Ward method was used to transform the data according to the following format:

$$d_{A.B} = n_A \|\bar{x}_A - \bar{\bar{x}}\|^2 + \; n_B \|\bar{x}_B - \bar{\bar{x}}\|^2$$

Where, $d_{A.B}$ denotes the calculated distance between A and B; n_A and n_B refer to number of variables in cluster A and B. Assuming that $\overline{x_A}$ and $\bar{x}_B$ represent the indicators vector for talents, $\overline{x_A}$ and $\bar{x}_B$ in cluster A and B, and $\bar{\bar{x}}$ the centroid of cluster A or B, in other words, which will calculate the minimum distance squared of $\|\bar{x}_A - \bar{\bar{x}}\|^2$ and $\|\bar{x}_B - \bar{\bar{x}}\|^2$.

3. Results.

3.1 Explanation of Gini indices. Using 2014-2018 the World Talent Ranking report data, we found the Gini indices with overall Talent scores have a little declining from 0.219 in 2014 to 0.131 in 2018. However, the result reflects the overall Gini index is located in absolute equality level, see Figure 1. Furthermore, checking the specific three factors, the results reveal the range of investment and development factor's Gini indices is from 0.161 to 0.160; The range of appeal factor's Gini indices is from 0.187 to 0.167; The range of readiness factor's Gini indices is from 0.218 to 0.167 during these periods. All the Gini indices are located in the absolute equality level.

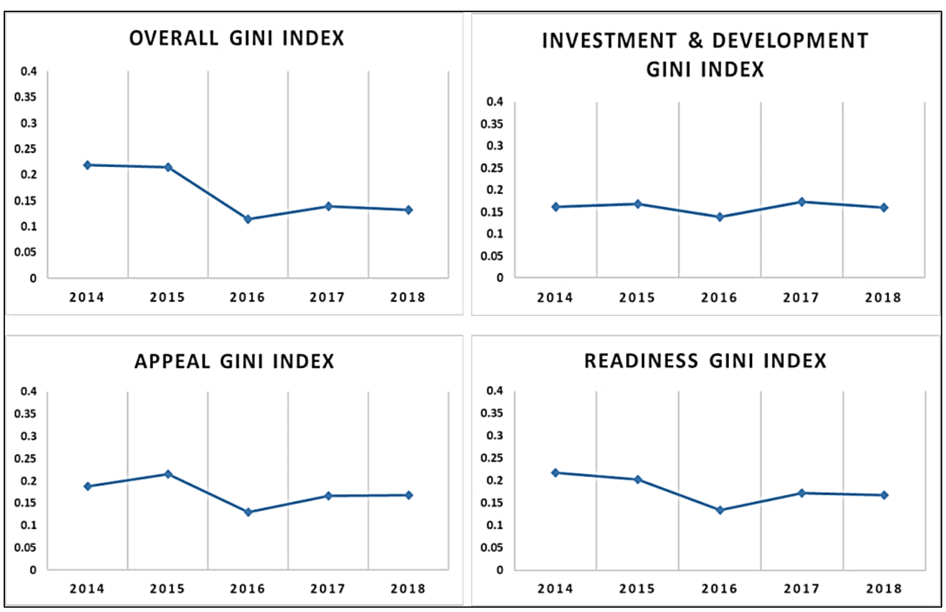

FIGURE 1. The World Talent overall's and each factor's Gini indices

The Lorenz curve also presents a signal for checking the equality of the World Talent competitive among the economics. Totally, even though the change is very limited, the result displays the difference. The Lorenz curves reflect on the equality of different years have shown in Figure 2. As to the World Talent competitive among the economics from

2014 to 2018, trend as the Figure 2, where the GINI indices are slightly decreasing in 2016 among all three kinds of factors: investment and development, appeal, and readiness. However, the overall Lorenz curves reflect as the Gini indices are relatively equality level.

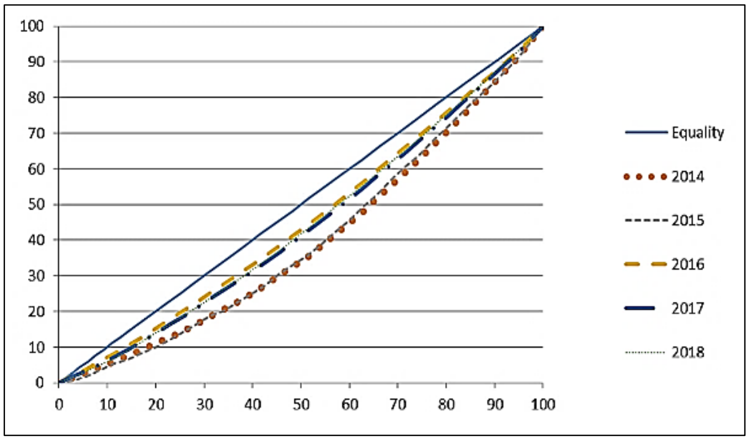

FIGURE 2. Lorenz curves reflect on the equality of different years

3.2 Findings of cluster analysis. The key results of cluster analysis include the similarity and distance values, the dendrogram, and the final partition. The appropriate number of clusters in the World Talent Competitiveness in OECD 35 economics was determined by computing the Euclidean distance with Ward linkage. This study found the cluster analysis provides useful information for examining the case data's distances. The lower the distance level, the closer the observations are in each cluster. The dendrogram with two clusters drew by Ward linkage and Euclidean distance has a relatively high similarity level and a relatively low distance level, see Figure 3. Minitab also provides different colors with the groups to identify.

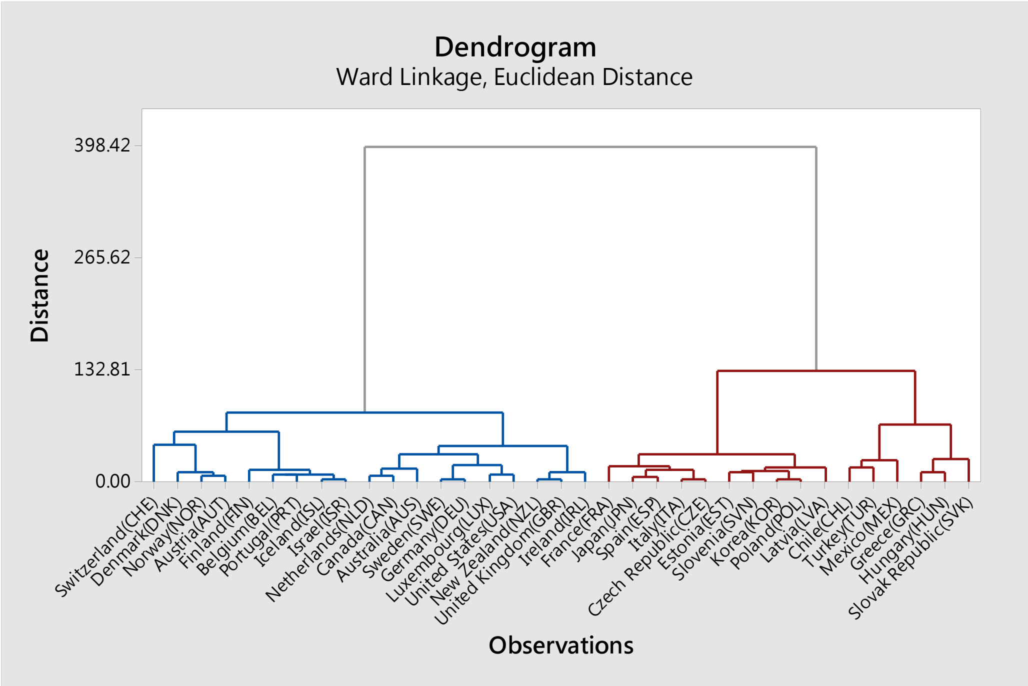

FIGURE 3. Two clusters for OECD economies with their World Talent Competitive 2018

This dendrogram was created using a final partition of two clusters; the cluster1 (far left) is composed of 19 observations (Switzerland, Denmark, Norway, Austria, Finland, Belgium, Portugal, Ice Land, Israel, Netherlands, Canada, Australia, Sweden, Germany, Luxembourg, United States, New Zealand, United Kingdom and Ireland). The cluster 2 is composed of 16 observations (France, Japan, Spain, Italy, Czech Republic, Estonia, Slovenia, Korea, Poland, Latvia, Chile, Turkey, Mexico, Greece, Hungary, and Slovak Republic). Cluster analysis for the IMD World Talent Ranking indicators demonstrates the two different groups among the economics which is indicated the cluster 1 is grouping with higher World Talent competitive than the cluster 2. After the dendrogram is determined the final groupings, this study displays the final partition in Table 2 and Table 3 which show the characteristics of each cluster. These two cluster centroids can be seen as representing the average observation within a cluster across all the variables in the analysis.

TABLE 2. Final partition of cluster analysis OECD/ the IMD World Talent Competitiveness 2018

Clusters	Number of observations	Within cluster sum of squares	Average distance from centroid	Maximum distance from centroid
Cluster1	19	5231.63	15.1887	33.9436
Cluster2	16	7261.98	19.1427	43.5396

TABLE 3. Cluster centroids for OECD/ the IMD World Talent Competitiveness 2018

Variables	Cluster1	Cluster2	Grand centroid
Investment & Development Factor	72.856	53.435	63.978
Appeal Factor	72.466	48.324	61.430
Readiness Factor	74.424	45.980	61.421

4. Conclusions. This study aims to tackle the IMD World Talent Ranking report data with Gini index among OECD economies from 2014 to 2018. The trends of Gini indices with different indicator data show relative equality in the research target. Basically, the Gini index transformation provides an alternative way to review the ranking data. The criteria for justifying the Gini indices with equality or inequality are reliable and realizable. This is a convenient tool using to monitor development for long-term or short-term purposes. In addition, the result of cluster analysis for the IMD World Talent Ranking indicators demonstrates the two different groups among the economics. This study provides an alternative to interpret the ranking data. The finding can provide an alert signal for the economics within the low competitive group. For further studies, we perceived that the IMD World Talent Competitiveness data is one of databank based on the survey of economics. With the data mining notions, it can be extended widely for reinventing the meaning of data.

REFERENCES

[1] D. Larose and C. Larose, *Discovering Knowledge in Data. An Introduction to Data Mining*, John Wiley & Sons, Hoboken, 2014.

[2] Y. A. Alsultanny, Labor market forecasting by using data mining. *Procedia Computer Science*, vol.*18*, 1700-1709, 2013.

[3] A. Rodrigues *et al*, Enhancing contest specifications for dependable adaptive system: A data mining approach, *Information and Software Technology*, vol.112, pp115-131, 2019.
[4] D. F. Chang and C. C. Chen, Cluster analysis for student performance in PISA2015 among OECD economies, *ICIC Express Letters Part B: Applications*, vol.9, no11, pp1139-1146, 2018.
[5] M. Kashif, M. Aslam, A. H. Al-Marshadi and C. Jun, Capability indices for non-normal distribution using Gini's mean difference as measure of variability, *IEEE Access*, vol.4, pp.7322-7330, 2016.
[6] IMD World Talent Ranking 2018 - IMD business school, 2018, https://www.imd.org/globalassets/wcc/docs/release-2018/talent-2018/talent_ranking_2018.pdf.
[7] IMD World Talent Ranking 2017 - IMD business school, 2017, https://www.imd.org/globalassets/wcc/docs/talent-ranking/talent_ranking_2017_web.pdf
[8] IMD World Talent Ranking 2016 - IMD business school, 2016, https://www.imd.org/globalassets/wcc/docs/talent_2016_web.pdf
[9] IMD World Talent Ranking 2015 - IMD business school, 2015, http:/www.whygeneva.ch/en/imd-world-talent-report-2015
[10] IMD World Talent Ranking 2014 - IMD business school, 2014, https://www.talentcorp.com.my/clients/TalentCorp_2016_7A6571AE-D9D0-4175-B35D-99EC514F2D24/contentms/img/publication/IMD_World_Talent_Report_2014.pdf
[11] OECD, Members and partners, 2018, http://www.oecd.org/about/membersandpartners/
[12] M. O. Lorenz, *Methods of Measuring the Concentration of Wealth. Publications of the American Statistical Association*, vol.9, no.70, pp.209-219, 1905.
[13] A. Ultsch and J. Lotsch, A data science based standardized Gini index as a Lorenz dominance preserving measure of the inequality of distributions, *PloS one*, vol.12, no.8, 2017.
[14] Y. Lin, Q. Zhang, W. Chen and L. Ling, The social income inequality, social integration and health status of internal migrants in China, *International Journal for Equity in Health*, vol.16, no.1, pp.139-150, 2017.
[15] P. Giudici and E. Raffinetti, A Gini concentration quality measure for ordinal variables, Notebooks of the Department of Economics, Statistics and Law, University of Pavia, Statistical Series, (1), 2011.
[16] M. J. Brusco, R. Singh, J. D. Cradit and D. Steinley, Cluster analysis in empirical OM research: Survey and recommendations, *International Journal of Operations & Production Management*, vol.37, no.3, pp.300-320, 2017, https://doi.org/10.1108/IJOPM-08-2015-0493.
[17] R. C. Tryon, *Cluster analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality*, Edwards Brothers, Ann Arbor, MC, 1939.