*Corresponding Author's Email: bolukbas@yildiz.edu.tr
Proceedings of the International Conference on Research in Business, Management and Economics
Vol. 1, Issue. 2, 2024, pp. 28-49
DOI: https://doi.org/10.33422/icrbme.v1i2.920

ISSN: 3030-0525 online

Evaluating the Competencies of Incubation Firms by Using PCA with K-means Clustering Analysis

Ufuk Bölükbaş*, Mustafa Sönmez, Duygu Tüylü, and Ali Fuat Güneri
Department of Industrial Engineering, Yildiz Technical University, 34349, Besiktas, Istanbul, Türkiye

Abstract

In this research, a survey consisting of 116 questions, 95 technical and 21 general, was prepared in order to analyse the entrepreneurial ecosystem as a result of the literature research on entrepreneurship. All questions were directed to the entrepreneurial firm officials participating in the research from different regions in Türkiye, and a statistical database was created from answers received. The survey consists of three sections: personal information of entrepreneur, information about entrepreneurial firm, and technical analysis section where the entrepreneur firm is evaluated. Within the scope of the research, a survey was applied to 304 incubation firms. Related dimensions in the questionnaire are used to evaluate the performance of the entrepreneurial firms based on the clusters. Cluster analysis is applied to the research data to evaluate and group the entrepreneur firms. The general entrepreneur profile in the research is obtained with basic statistical analyses and descriptive statistics; incubation firms are grouped according to their performances in terms of the determined variables with Principal Component Analysis (PCA) and K-Means Cluster Analysis. The aim of the study is to statistically evaluate the competency levels and performance of entrepreneurial firms by comparing them in different clusters. All performance results are evaluated and firms are analysed based on the entrepreneurial ecosystems. The study purposes to analyse entrepreneurial firms by using K-means clustering with the PCA method. In the research, the competency evaluations of entrepreneurs are interpreted depending on the performance clusters. As a result of the analysis, performance clusters were evaluated over 8 clusters.

Keywords: Entrepreneurship, Entrepreneurial Firms, K-Means Clustering, Principal Component Analysis

1. Introduction

Incubation centers are centers that provide support and resources to entrepreneurs so that they can develop both their businesses and new product ideas (Olufunke, et. al., 2020). Entrepreneur who starts a new venture faces many challenges. At this point, incubations provide entrepreneurs with both financial and moral support, bringing newly established businesses to life, allowing them to survive in their early stages, and helping them build a solid foundation for sustainable growth (Deyanova, et. al., 2022).

Despite the importance of incubation firms, grouping and clustering firms within themselves presents several challenges. The clustering result changes as the number of cluster parameters changes, so the main difficulty of cluster analysis is that the number of clusters or model parameters is rarely known and must be determined before clustering (Kodinariya&Makwana, 2013).

An entrepreneurship research provides significant evidence that entrepreneurship contributes positively to regional and national economic growth (Jha and Pande, 2024). Rai et al. (2025) evaluate that educational institutions in the context of India are seen as engines of growth in the knowledge-based economy, and their importance to economic growth and social development has been emphasized over the past five decades. Technology transfer, patenting and commercial activities are evaluated as the third role of educational institutions, along with research and teaching, in what is called a paradigm shift towards entrepreneurial universities.

Ashraf et al. (2024) emphasize that entrepreneurship and the blossoming of entrepreneurial ventures is necessary for economic sustainability. The study provides actionable insights for teachers and policymakers to design more effective entrepreneurial education programs that develop business skills and intentions in students. The article aims to explore the effects of entrepreneurial education on entrepreneurial intentions and examine the mediating effect of entrepreneurial mindset and entrepreneurial alertness on this relationship in graduate students.

The uncertainty inherent in the entrepreneurial ecosystem is affected by factors such as resource constraints, intense market competition and scaling difficulties, which explains the high failure rates (Font-Cot, et. al., 2025). Despite the economic and social benefits offered by successful startups in the entrepreneurial ecosystem, the failure rates of startups are quite high on a global scale. In this context, systematic reviews conducted to analyze the factors determining startup success reveal the effects of personal, organizational and environmental factors on the sustainability of startups. Findings in the literature reveal that startup success should be addressed from a multidimensional perspective and that not all factors are equally effective. In this context, entrepreneurs and policy makers need to evaluate success factors with a holistic approach (Argaw and Liu, 2024).

In entrepreneurial ecosystems, connections with similar people (homophily) and collaborations with different people (heterophily) are important factors affecting startup formation rates. The study shows that entrepreneurial ecosystems that connect with different actors create more startups. It emphasizes that closed ecosystems based only on familiar circles can limit entrepreneurship and that ecosystems should be more open and diverse (Prokop and Thompson, 2023).

The K-means method is the most well-known and used clustering method that al-lows data clustering within predefined variables (Sinaga&Yang, 2020). Principal Component Analysis (PCA) is a multivariate technique that analyses a data table in which observations are described by quantitative dependent variables that are related to each other. Its aim is to extract important information from the table, represent it as a new set of orthogonal variables called principal components, and show the similarity pattern of observations and variables as points on maps (Abdi&Williams, 2010).

While K-means is a probabilistic method, PCA performs the process by applying a series of linear matrix transformations to the data. In this process, PCA allows significant advantages to be obtained by reducing the dimensionality by removing the de-pendency structure between the variables in the data set. PCA also increases the performance of the model and provides more clarity and easy understanding with visualization opportunities.

The aim of this study is to cluster the incubation firms in Türkiye and to reveal the relationships between the clusters. For this purpose, a comprehensive field study was conducted for 305 incubation firms in 2024 year through a survey to collect data on various dimensions. One of the firms participating in the survey was removed from the data set because its responses were found to be inconsistent. Therefore, the number of firms decreased to 304.

The rest of this paper is structured as follows. Section 2 provides an overview of the performance dimensions used to assess the competency of incubation firms. Section 3 introduces the mathematical methodologies and their steps. Section 4 presents the application of K-Means with PCA method. Finally, discussions and conclusions are explained.

2. Literature Review on Performance Dimensions

Incubation centers and also known as incubators are structures that provide support to new ventures for a certain period of time through the common services they offer, encouraging the establishment and growth of Small and Medium-Sized Enterprises (SMEs). These centers are designed to support the development of ventures and provide them with the resources they need (Udell, 1990). Incubators share costs and reduce overhead by offering a "one-stop shopping" approach, thus significantly increasing the survival and growth prospects of new ventures (Busulwa et al., 2020). Many of the core assets and sources of competitive advantage of incubators consist of intangible elements and affect business performance and strategic outcomes (Crammond, 2024).

The dimensions used in the performance and competency analysis are based on expert judgement and literature review. The related studies are demonstrated in Table 1.

Table 1. Overview of dimensions and related references

Dimensions	References
Customer	(Ashraf et. al., 2024), (Argaw and Liu, 2024), (Júnior et. al., 2022), (Karambakuwa & Bayat, 2022), (Pugliese et. al., 2021), (Cubero et. al., 2021), (Rijnsoever & Eveleens, 2021), (Ratten, 2020), (Pearce & Pearce, 2020), (Solano et. al., 2020), (Innocenti & Zampi, 2019), (Rannikko et. al., 2019), (Amankwah-Amoah et. al., 2019), (Bacalan et. al., 2019), (Le Trinh, 2019), (Dai et. al., 2018), (Geissdoerfer et. al., 2018), (Rompho, 2018), (Seo et. al., 2018)
Technology	(Rai et. al., 2025), (Argaw and Liu, 2024), (Karambakuwa & Bayat, 2022), (Quero et. al., 2022), (Júnior et. al., 2022), (Cheah & Ho 2021), (Pugliese et. al., 2021), (Thukral, 2021), (Ratten, 2020), (Pearce & Pearce, 2020), (Solano et. al., 2020), (Amankwah-Amoah et. al., 2019), (Bertoni et. al., 2019), (Bacalan et. al., 2019), (Le Trinh, 2019), (Innocenti & Zampi, 2019), (Rannikko et. al., 2019), (Dai et. al., 2018), (Geissdoerfer et. al., 2018), (Rompho, 2018), (Seo et. al., 2018)
Research and Development	(Rai et. al., 2025), (Argaw and Liu, 2024), (Quero et. al., 2022), (Júnior et. al., 2022), (Cheah & Ho 2021), (Pugliese et. al., 2021), (Cubero et. al., 2021), (Luo et. al., 2020), (Matricano, 2020), (Ratten, 2020), (Solano et. al., 2020), (Amankwah-Amoah et. al., 2019), (Innocenti & Zampi, 2019), (Rannikko et. al., 2019), (Bacalan et. al., 2019), (Le Trinh, 2019), (Rompho, 2018)
Competition	(Font-Cot, et. al., 2025), (Argaw and Liu, 2024), (Júnior et. al., 2022), (Cheah & Ho 2021), (Pugliese et. al., 2021), (Rijnsoever & Eveleens, 2021), (Innocenti & Zampi, 2019), (Rannikko et. al., 2019), (Amankwah-Amoah et. al., 2019), (Bertoni et. al., 2019), (Bacalan et. al., 2019), (Seo et. al., 2018), (Dai et. al., 2018), (Geissdoerfer et. al., 2018)
Investment	(Rai et. al., 2025), (Argaw and Liu, 2024), (Júnior et. al., 2022), (Pugliese et. al., 2021), (Matricano, 2020), (Pearce & Pearce, 2020), (Luo et. al., 2020), (Solano et. al., 2020), (Amankwah-Amoah et. al., 2019), (Bertoni et. al., 2019), (Bacalan et. al., 2019), (Innocenti & Zampi, 2019), (Le Trinh, 2019)
Marketing	(Rai et. al., 2025), (Ashraf et. al., 2024), (Argaw and Liu, 2024), (Estep et. al., 2021), (Pugliese et. al., 2021), (Cubero et. al., 2021), (Pearce & Pearce, 2020), (Ratten, 2020), (Solano et. al., 2020), (Seo et. al., 2018)
Environment and Sustainability	(Rai et. al., 2025), (Ashraf et. al., 2024), (Jha and Pande, 2024), (Argaw and Liu, 2024), (Amankwah-Amoah et. al., 2019), (Dai et. al., 2018), (Geissdoerfer et. al., 2018)
Human Resources	(Rai et. al., 2025), (Ashraf et. al., 2024), (Prokop and Thompson, 2023), (Júnior et. al., 2022), (Karambakuwa & Bayat, 2022), (Cheah & Ho 2021), (Pugliese et. al., 2021), (Thukral, 2021), (Matricano, 2020), (Staniewski & Awruk, 2019), (Bacalan et. al., 2019), (Seo et. al., 2018), (Dai et. al., 2018)
Commercialisation	(Rai et. al., 2025), (Prokop and Thompson, 2023), (Quero et. al., 2022), (Júnior et. al., 2022), (Karambakuwa & Bayat, 2022), (Cheah & Ho 2021), (Pugliese et. al., 2021), (Cubero et. al., 2021), (Pearce & Pearce, 2020), (Solano et. al., 2020), (Innocenti & Zampi, 2019), (Rannikko et. al., 2019), (Bertoni et. al., 2019), (Bacalan et. al., 2019), (Dai et. al., 2018), (Geissdoerfer et. al., 2018), (Rompho, 2018)

The number of R&D Projects, Research & Development dimension; the total number of national and international patents and the number of grants, commercialisation dimension is examined.

With the customer dimension, the necessity for a firm to focus on customer demands and needs by prioritising customer-oriented innovation and building trust is examined. In the technological dimension, the importance given to technological infrastructure through the use of functional websites, computer-aided design and business intelligence applications are evaluated. The R&D dimension reflects the degree of consideration of stakeholder co-operation, product quality and innovation activities. Competition dimension reflects competitor analysis, favourable price strategies and strategies to develop solutions to social problems.

In the investment dimension, factors such as government incentives and openness to risky investments are given importance, while in marketing, reaching the target audience and creating brand value with innovative strategies are taken into consideration. In the environment and sustainability dimension, energy efficiency and recycling are considered as elements to be focussed on. In the human resources dimension, factors such as team motivation and effective communication are analysed. In the commercialisation dimension, market testing of prototype products, customer relationship management and commercialisation of minimum viable products are considered.

The analysis data is based on 12 dimensions as Customer, Technology, Research & Development (R&D), Competition, Investment, Marketing, Environment and Sustainability, Human Resources (HR), Commercialisation, The number of R&D Projects, The total number of national&international patents and The number of grants.

2.1 Descriptive Statistics

Within the scope of the research, the database information is obtained from 304 incubation firms in Türkiye. The top three provinces that contributed the most to the field research were Istanbul, Ankara and Van, respectively. The information regarding to the field of work or sectors, duration of activity, education level, gender, age and type of entrepreneurial firms that responded to the survey is explained in this section.

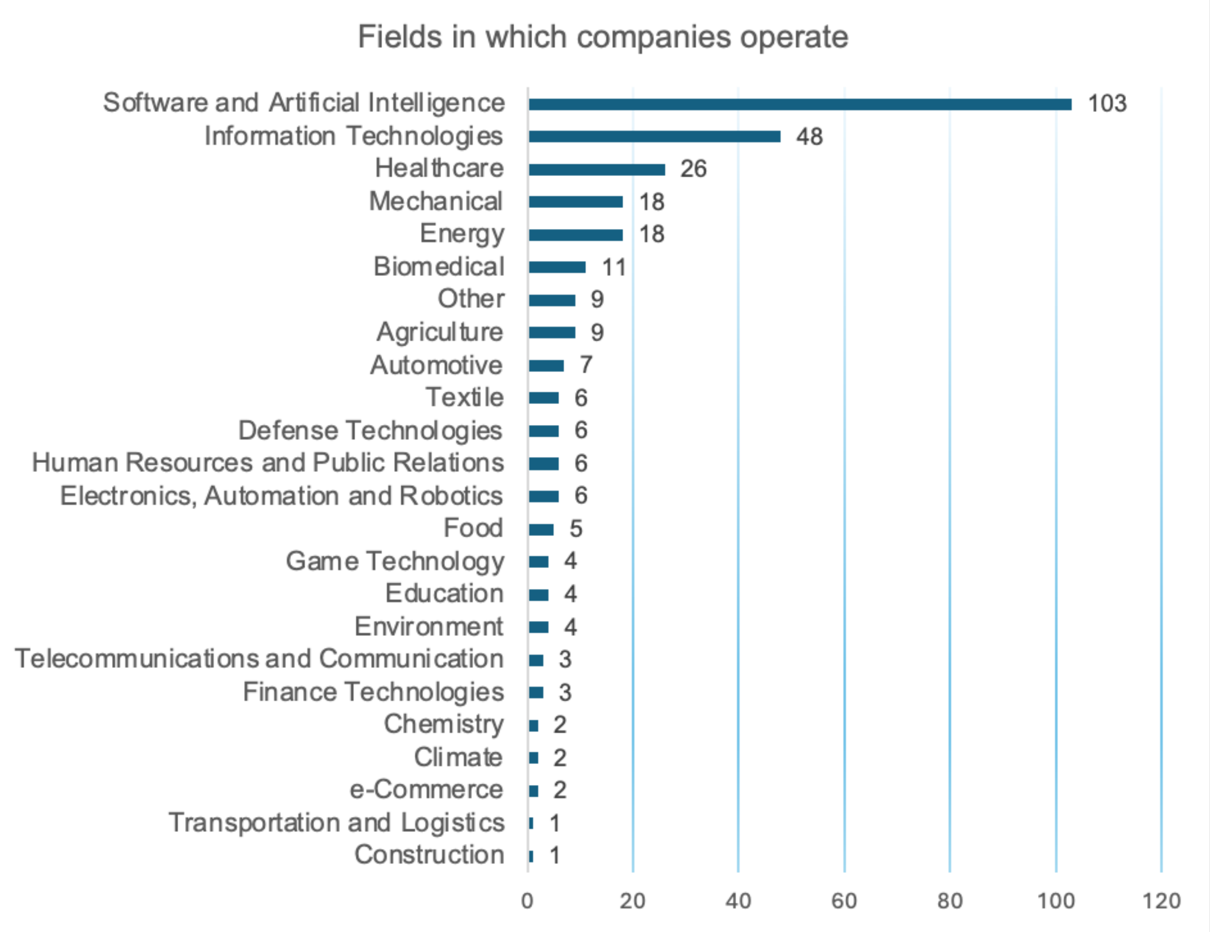

Figure 1. Fields in which incubation firms operate

The areas in which incubation firms operate are shown in Figure 1. The trend is seen to be in the “Software and Artificial Intelligence” area. The software and artificial intelligence areas are followed by information technologies and health, respectively.

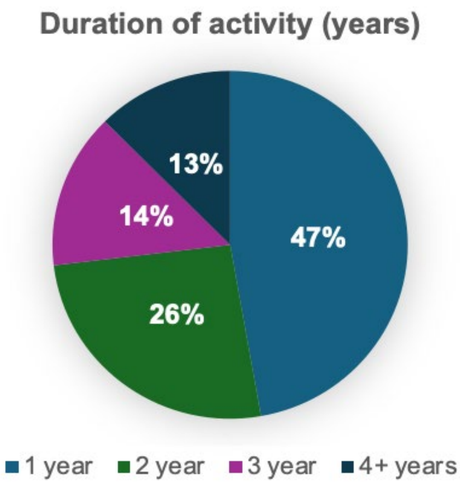

Figure 2: Duration of activity (years)

The operating periods of the firms are given in Figure 2. Totally, 144 incubation firms that have been operating for 1 year have the largest share in the sample. This number constitutes 47.21% of the incubation firms participating in the survey.

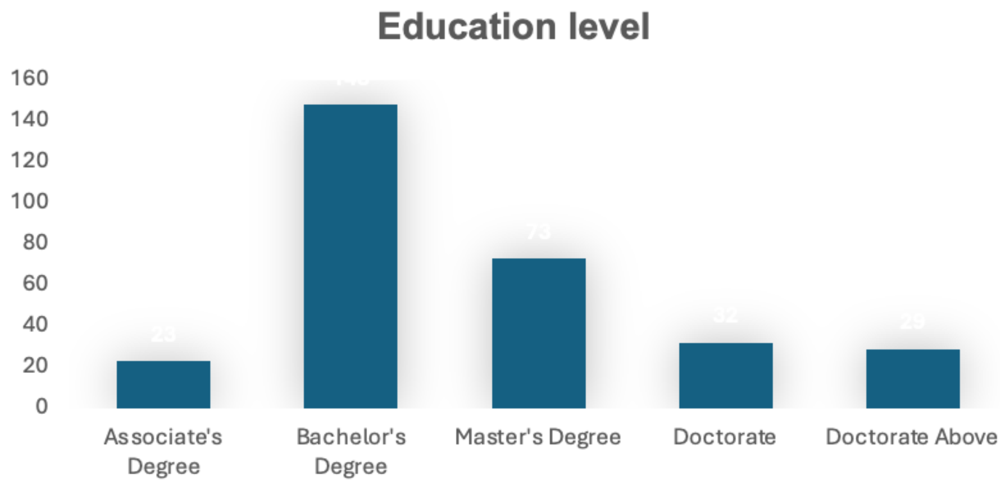

Figure 3. Education level

The educational backgrounds of entrepreneurial firm representatives/officials are shown in Figure 3. It is seen that the level of education is concentrated at the Bachelor's and Master's levels.

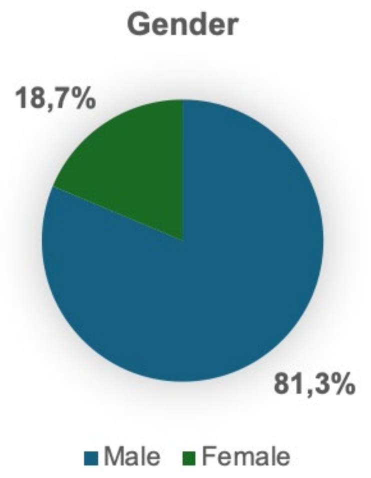

Figure 4. Gender distribution of participants

The genders of people working in authorized positions in incubation firms are given in Figure 4. According to the data here, 81.3% are male and 18.7% are female.

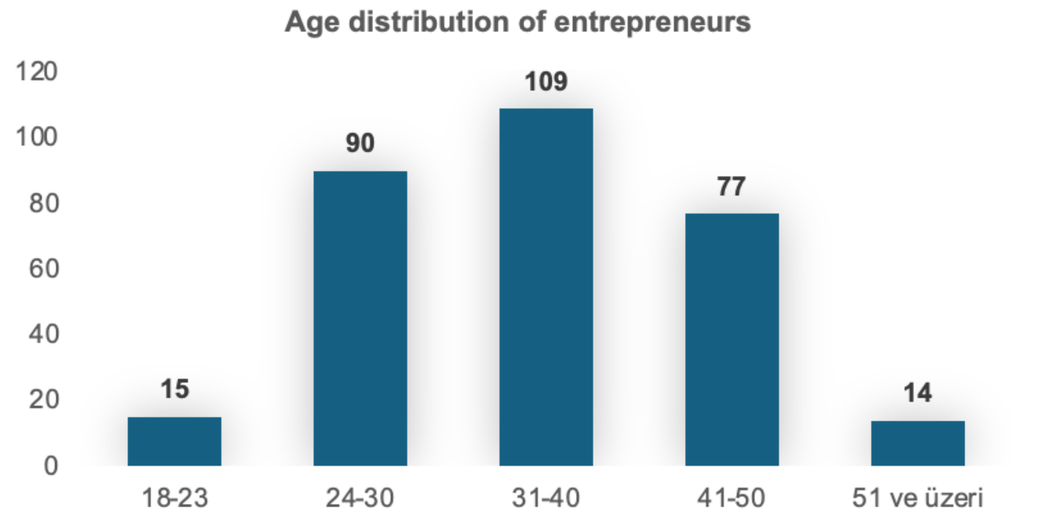

Figure 5. Age distribution of entrepreneurs

The age distribution of people in positions of authority in entrepreneurial firms is given in Figure 5. It is seen that the number of entrepreneurs between the ages of 24-40 constitutes 65.1% of the total number of entrepreneurs.

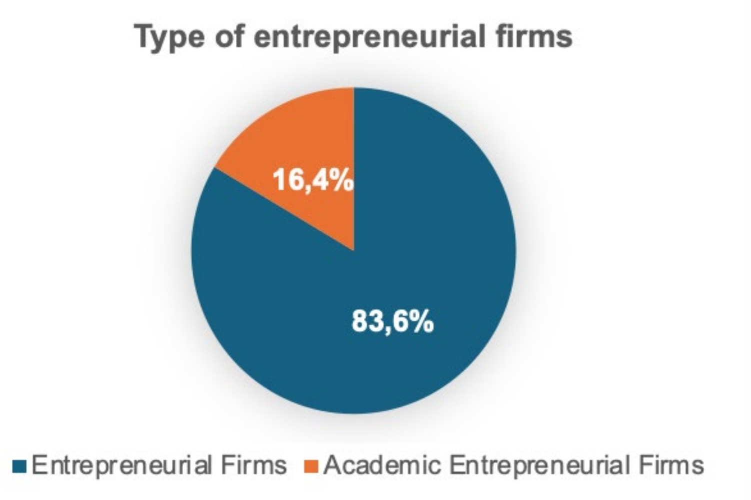

Figure 6. Type of entrepreneurial firms

Figure 6 shows the type of entrepreneurial firms. It is seen that 16.4% of the enterprises are Academic Incubation Firms. It is seen that the non-academic incubation firm rate is 83.6%.

Clustering analysis is performed in addition to these 9 performance dimensions, by adding the number of R&D projects carried out by the firm, the total number of national and international registered patents owned by the firm and the number of grants (TÜBİTAK-The Scientific and Technological Research Council of Türkiye, KOSGEB-Small and Medium-sized Enterprises Development Organization of Turkey, Angel Investor Networks, Ministry of Industry and Technology, Türkiye Technology Development Foundation, Development Agency, European Union, United Nations and OECD Supports etc.). These dimensions have been evaluated in the study as tools that play a strong role in the analysis. The evaluation of firms allows them to be divided into different clusters.

Table 2. Dimensions and properties for clustering

No	Dimensions	Variables	Usage	Status	Score range
1	The number of R&D Projects	A1	Beneficial	Single	$0 \leq X \leq 5$
2	The total number of National and International patents	A2	Beneficial	Total	$0 \leq X \leq 4$
3	The number of Grant	A3	Beneficial	Total	$0 \leq X \leq 4$
4	Customer	A4	Beneficial	Average	$1 \leq X \leq 5$
5	Technology	A5	Beneficial	Average	$1 \leq X \leq 5$
6	Research & Development	A6	Beneficial	Average	$1 \leq X \leq 5$
7	Competition	A7	Beneficial	Average	$1 \leq X \leq 5$
8	Investment	A8	Beneficial	Average	$1 \leq X \leq 5$
9	Marketing	A9	Beneficial	Average	$1 \leq X \leq 5$
10	Environment & Sustainability	A10	Beneficial	Average	$1 \leq X \leq 5$
11	Human Resources	A11	Beneficial	Average	$1 \leq X \leq 5$
12	Commercialisation	A12	Beneficial	Average	$1 \leq X \leq 5$

In the study, metric and non-metric questions are prepared for the structured questionnaire. At the same time, the number of R&D projects completed by the firms, the total number of national and international patents and the number of grants utilised by the firm are also included in the evaluation. For metric variables, the scales 1-5 and for non-metric variables, the scales 0-4 and 0-5 are used in the study. The maximum value for R&D projects of 5 and above is transformed into 5, and the maximum value for the total number of national and international patents and grants of 4 and above is transformed into 4. In this way, all data are transformed the number. Data regarding the survey questions are explained in Table 2. Variables are named from A1 to A12.

3. Mathematical Approaches

The main purpose of this article is to cluster firms over 12 dimensions by adding the number of completed R&D projects, the total number of national and international patents and the number of grants used in addition to the nine basic dimensions. In the study, 304 data groups consisting of 12 dimensions are analysed using k-means clustering with PCA method in the Python programming language.

3.1 Cluster Analysis with K-Means

Cluster analysis investigates the number of clusters by examining data groups with different structures. In cluster analysis, the aim is for the data within the same cluster to have the same structure, and the data between clusters to have different structures. This distinction is made by taking into account the similarities and differences of the data.

K-means is a popular and efficient clustering algorithm that divides data points into k clusters based on their similarity. However, it has limitations such as sensitivity to initial cluster centers, getting stuck in local optima, and assumptions about cluster shapes. The algorithm initially determines k centers and assigns data points to these centers, then calculates the mean of each cluster and repeats this process until convergence (Ikotun et al., 2023; Hastie et al., 2009).

In the case of hierarchical clustering, a distance or similarity matrix must be created between all pairs of observations. In case of working with large datasets (n > 250), all possible distances are calculated. Instead of hierarchical clustering, the k-means technique is considered a more suitable method since it does not require the calculation of all distances (Bolukbas & Guneri, 2018).

The number of clusters in the K-means technique needs to be predetermined or known by the researcher or expert. A clustering technique called the K-means method splits data into k-clusters and tries to minimize the sum of squares inside each group. The strategy aims to reduce similarities across clusters and increase similarities within clusters. Each observation is assigned to the closest cluster based on its distance from the cluster centers after the cluster centers have been identified. The following is a list of the fundamental computing steps in the K-means method:

In Equation (1), the Euclidean distance formula is employed to delineate disparate clusters utilising disparate coordinates, which are designated as values of p and q. These values are represented as follows: p = (p1, p2, ..., pn) and q = (q1, q2, ..., qn).

$$\sqrt{\sum_{i=1}^{n}(p_i-q_i)^2} = \sqrt{\sum_{i=1}^{n}(p_1-q_1)^2+(p_2-q_2)^2+\cdots(p_n-q_n)^2} \qquad (1)$$

i = 1, 2, . . . , n and i is the number of observations as the number of firms, k is the number of clusters.

The k-means algorithm commences with k randomly generated cluster centroids. Each point within a cluster is assigned to the cluster with the nearest centroid. The centroid values are obtained by calculating the mean values of all points. The calculations continue until the invariant/constant values of the cluster are obtained as a steady state.

K number of observations are selected as object values and k is the number of clusters. The midpoints/centres of each cluster are calculated by Equation (2) below:

$$M_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_{ik} \qquad (2)$$

The centre values of the objects M1, M2, ..., Mk are called 'cluster centres' (Gersho & Gray, 2012).

The squared error formula is used to minimise the sum of squares within groups and is shown in Equation (3) (Linde, et. al., 1980).

$$e_i^2 = \sum_{i=1}^{n_k} (x_{ik-M_k})^2 \qquad (3)$$

The recalculation of cluster centres occurs after the allocation of each observation to a cluster, with the newly determined cluster centres then informing the creation of subsequent assignments for the observations. This process is repeated until a lack of discernible change in the cluster centres is detected. The utilisation of a class of techniques known as cluster analysis, predicated on a predetermined set of factors, facilitates the organisation of cases into groups that are heterogeneous amongst and relatively homogeneous within each group. These groups are designated clusters. The experimental nature of this approach stems from its inability to differentiate between dependent and independent variables (Hagen, et. al., 2012).

3.2 Principal Component Analysis

Large datasets are becoming increasingly common in many disciplines. Various techniques have been developed to reduce the dimensions while preserving the information in such datasets. Among these methods, principal component analysis (PCA) is one of the oldest and most widely used approaches with the goal of preserving maximum 'variability' while reducing the data dimensionality (Jolliffe & Cadima, 2016).

Principal Component Analysis (PCA) is a dimensionality reduction process. It is a frequently preferred method in cases where there are many variables, and the variables are related to each other (Heckler, 1996).

PCA, in addition to being a widely used and flexible data analysis tool in its standard form, also stands out with its adaptations developed for various situations and data types in different disciplines. (Collins, et. al., 2001).

The most important point in PCA analysis is to look at the data obtained from the right angle and to allow for better visibility of the details (Jolliffe & Cadima, 2016). This method selects a new coordinate system for the dataset and places the one with the largest variance on the first axis, then places the one with the second largest variance on the second axis, and so on (Jang, et. al., 1997). The PCA method is used to extract important information from the data, to compress the data set while preserving important information, and to provide easy explanation of the data set (Abdi & Williams, 2010). Basically, PCA has the following:

Data Preparation: Non-metric data has been digitized. Data has been preprepared.

Calculate the Covariance Matrix: The covariance matrix is calculated using the prepared data. The covariance matrix measures the relationship between data features.

Calculation of Eigenvalues: Eigenvalues are calculated from the covariance matrix. Eigenvalues help determine the importance of different components in the data set.

Choosing Principal Components and Forming a Dimension Vector: The components with the highest eigenvalues are selected and a dimension vector is created using them. This provides a reduced-dimensional representation of the data.

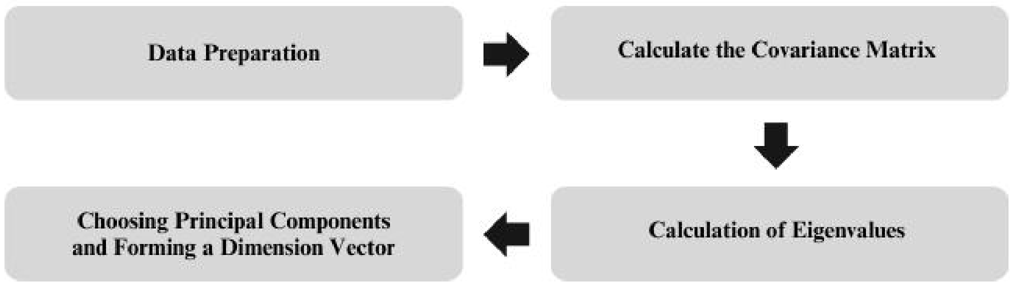

Figure 7. PCA flow

The basic flow of the PCA method is shown in Figure 7 as four steps. Each stage in the data analysis process is discussed in detail.

The performance of the K-means clustering algorithm is usually evaluated by the Silhouette Score and Davies-Bouldin Index. The Silhouette Score takes values between -1 and +1, where values close to 1 indicate good clustering, values close to 0 indicate overlapping clusters, and negative values indicate incorrect clustering (Shahapure & Nicholas, 2020). For the Davies-Bouldin Index, values close to 0 indicate better clustering, and values between 0 and 1 are generally considered good clustering (Petrovic, 2006).

4. Application and Results

A dataset of 304 datasets consisting of 12 dimensions are analysed using PCA method in the Python programming language and k-means clustering. A sample part of the data set analyzed in this study is shown in Table 3.

Table 3. A sample of dataset

Firms	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
1	2.00	0.00	1.00	4.44	3.75	4.62	4.33	3.44	4.33	5.00	3.25	3.56
2	2.00	0.00	2.00	4.22	3.50	3.85	4.17	3.89	3.11	4.33	4.75	3.78
3	3.00	0.00	1.00	4.00	4.50	4.31	4.67	3.78	4.00	5.00	2.75	4.22
4	1.00	0.00	1.00	4.11	3.50	4.08	4.33	4.00	3.33	4.00	3.75	3.89
5	1.00	2.00	2.00	3.89	3.75	3.69	3.50	3.33	3.78	4.00	4.00	3.33
…	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**	**…**
304	3.00	0.00	2.00	4.89	4.25	5.00	5.00	4.44	4.56	5.00	4.50	4.44

Eigenvalue is a vector whose direction does not change with a certain linear transformation. At the same time, when the linear transformation is applied, the eigenvector is scaled by a constant factor. Eigenvalue values above 1 are considered suitable for studies in the literature (Verbano, et. al., 2015). As seen in Table 4, there are three (3) components with eigenvalues above 1. For this study, the number of components have accepted as three (3, which provides approximately 71% explainability.

Table 4. Eigenvalue valuables

Component	Eigenvalue	Variance Ratio	Cumulative Variance Ratio
1	3.085	32.189	32.189
2	2.707	28.242	60.431
3	**1.005**	**10.482**	**70.914**
4	0.929	9.697	80.610
5	0.630	6.576	87.187
6	0.276	2.877	90.064
7	0.261	2.725	92.789
8	0.202	2.112	94.901
9	0.165	1.722	96.623
10	0.156	1.631	98.254
11	0.095	0.987	99.242
12	0.073	0.758	100.000

Figure 8 shows how much of the total data variance is explained by the main components as a result of the PCA analysis. Within the scope of the analysis, three (3) components are selected, representing approximately 71% of the cumulative variance of all dimensions.

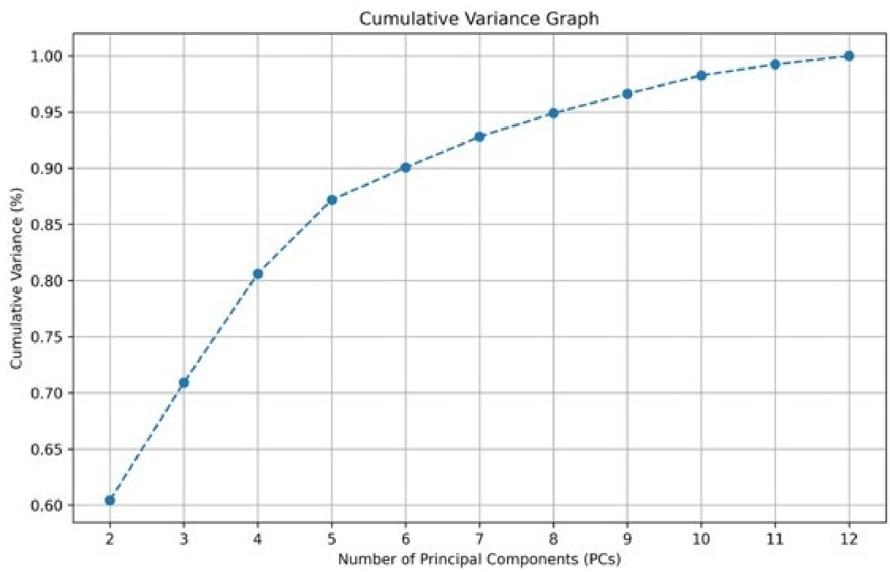

Figure 8. Cumulative variance graph

Although the original data set consists of 12 dimensions, reducing the data to three (3) main components significantly reduces the number of dimensions. This makes data processing more efficient and facilitates the analysis process.

The degree to which each dimension is represented by the components identified as three (3) different compenents is shown in Table 5.

Table 5. Contributions of dimensions to principal components

Dimensions	PC1	PC2	PC3
A1	0.967	-0.232	-0.071
A2	0.065	-0.105	0.782
A3	-0.047	-0.072	0.343
A4	-0.087	-0.284	-0.123
A5	-0.069	-0.303	-0.136
A6	-0.083	-0.296	-0.006
A7	-0.071	-0.293	-0.088
A8	-0.051	-0.270	-0.058
A9	-0.070	-0.309	-0.167
A10	-0.100	-0.485	0.372
A11	-0.087	-0.291	-0.141
A12	-0.091	-0.310	-0.189

The explained variance ratios for the principal components are shown in Figure 9. Component 1 explains the highest variance and explains approximately 30% of the total variance. Component 2 explains a variance ratio close to the first component and explains approximately 27% of the total variance. Component 3 has the lowest variance ratio and explains a little more than 10% of the total variance.

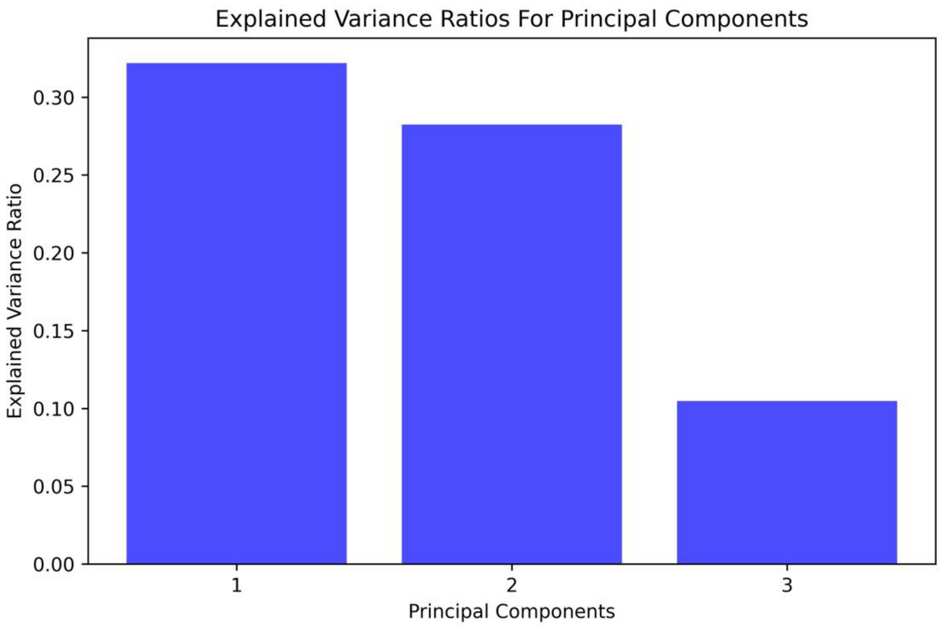

Figure 9. Screen plot of PCA

Elbow method is calculated by the sum of the square of the distances of the points to the cluster centre for each K-value. In this process, a graph is drawn for each K-value, and the elbow point on the graph where the difference between the sums starts to decrease is determined as the most appropriate K-value (Syakur, et. al., 2018).

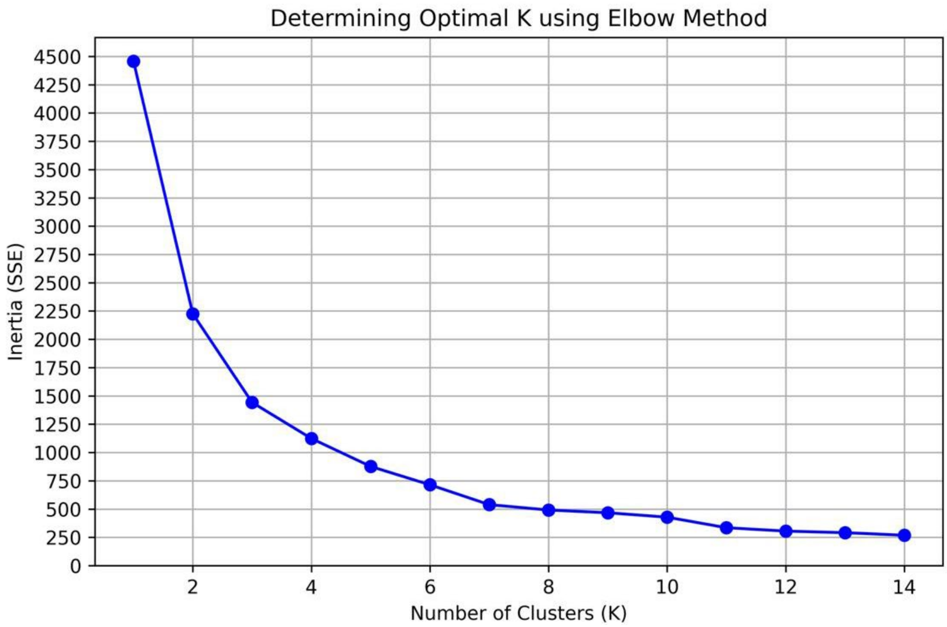

Figure 10. Determining optimal k-using elbow method

Within the scope of the study, the number of cluster is determined as eight (8) according to the Elbow method and the graph used in this process is given in Figure 10.

The positionings of the clusters in the PCA method are determined by examining the cluster centers for each cluster. The coordinates of cluster centers are given in Table 6.

Table 6. The coordinates of clusters centres

Clusters	PC1	PC2	PC3
Cluster 1	-1.236	1.125	-0.448
Cluster 2	2.120	-0.732	0.046
Cluster 3	-0.436	0.803	1.042
Cluster 4	1.304	-1.894	2.688
Cluster 5	2.519	0.767	-0.610
Cluster 6	2.990	7.485	0.986
Cluster 7	1.714	-2.538	-0.794
Cluster 8	-1.543	-0.920	0.045

For this study, the visualisations where 12 dimensions are reduced to 3 components are as shown in Figure 11. The figures below illustrate the relationships between the different PCA components. The points on the figure represent the different clusters; the X represents the centre of the clusters. The cluster centres are identified with the help of the mean points of the data. The first graph shows the relationship between PC1 and PC2 components, the second graph shows the distribution of PC1 and PC3 components and the third graph describes the comparison of PC2 and PC3 components.

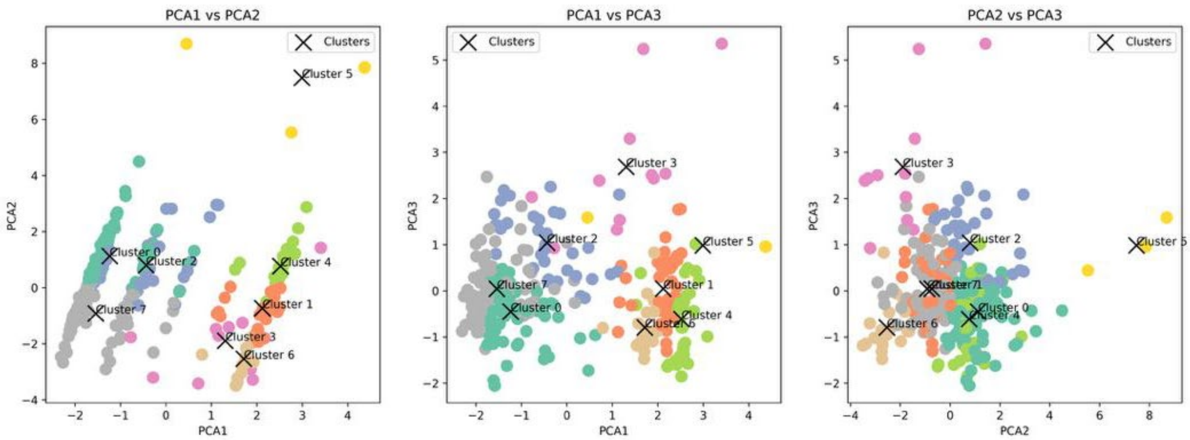

Figure 11. Visualization of clusters according to different axes

Such visualisations are an effective tool for analysing how well clusters are separated and how well each cluster centre fits the data. Together with the PCA analysis, it reveals the main variations in the data and allows a better understanding of the datasets.

The 3D visualization of the cluster analysis results visualized from different axes is given in Figure 12. Here, it is observed that while some clusters have clear distinctions, there may also be cluster transitions between some cluster members.

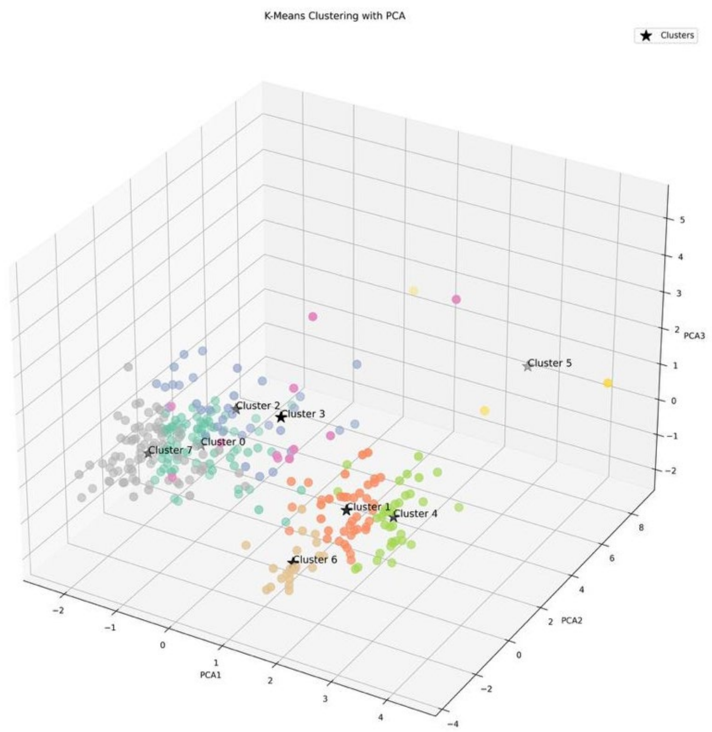

Figure 12. 3D visualisations of the clusters

The distances between the final cluster centres are given in Table 7. While determining the performance ranking of the clusters, the distances between the cluster centres are taken into consideration. In the final cluster centre table, Clusters 2 and 5, which have the smallest value with 1.683 that shows similar properties, so they are located side by side. Clusters 6 and 7 represent samples with different properties and structures, which are the farthest from each other with 10.259.

Table 7. Distances between final cluster centres

Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Cluster 1	0.000	3.867	1.722	5.039	3.775	7.769	4.716	2.126
Cluster 2	3.867	0.000	3.143	2.999	***1.683***	8.315	2.033	3.668
Cluster 3	1.722	3.143	0.000	3.607	3.385	7.509	4.377	2.278
Cluster 4	5.039	2.999	3.607	0.000	4.408	9.680	3.565	4.005
Cluster 5	3.775	***1.683***	3.385	4.408	0.000	6.921	3.406	4.446
Cluster 6	7.769	8.315	7.509	9.680	6.921	0.000	**10.259**	9.596
Cluster 7	4.716	2.033	4.377	3.565	3.406	**10.259**	0.000	3.732
Cluster 8	2.126	3.668	2.278	4.005	4.446	9.596	3.732	0.000

The final cluster centres for the variables and firm performance groups are given in Table 8. According to the evaluations in Table 7, the overall performance rankings of the clusters are as follows; the enterprises in Cluster 4 show the best performance. The enterprises in Cluster 8 rank second, followed by those in Cluster 7 in third place. Enterprises in Cluster 2 are ranked fourth, while those in Cluster 5 are in fifth place. The enterprises in Cluster 3 rank sixth, followed by those in Cluster 1 in seventh place. Finally, enterprises in Cluster 6 show the worst performance, ranking eighth.

Table 8. Final cluster centres for the variables

Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8
Firm count	81	40	34	11	35	4	21	78
A1	1.210	4.875	1.971	4.091	4.943	3.750	4.952	1.372
A2	0.074	0.525	1.118	3.818	0.143	0.000	0.095	0.308
A3	0.481	0.775	1.147	1.364	0.314	0.250	0.476	0.910
A4	4.085	4.183	3.915	4.525	3.898	1.306	4.762	4.528
A5	3.975	4.163	3.632	4.545	3.729	1.188	4.762	4.378
A6	3.787	3.952	3.817	4.406	3.593	1.288	4.641	4.339
A7	3.765	3.996	3.735	4.227	3.743	1.250	4.651	4.361
A8	3.255	3.517	3.232	3.707	3.171	1.250	4.153	3.829
A9	3.808	4.061	3.621	4.374	3.679	1.250	4.704	4.358
A10	2.905	4.067	3.706	4.394	2.724	1.167	4.651	4.538
A11	4.130	4.131	3.794	4.455	3.829	1.250	4.821	4.397
A12	3.763	3.836	3.484	4.293	3.521	1.167	4.672	4.316
Levels	*Seventh*	*Forth*	*Sixth*	***First***	*Fifth*	***Eighth***	*Second*	*Third*

Cluster 4 is the best performing cluster in terms of the number of completed R&D projects, the total number of national and international patents and the number of grants received. Cluster 6 is the worst performing cluster in terms of the total number of national and international patents and the number of grants received.

5. Discussion

In the study, differences were observed between private, public and R&D-focused universities, as well as significant differences were found between private and state-supported incubation centers. It is evaluated that these differences are due to financing structures and management strategies. While structures with more flexible decision-making mechanisms can direct investments towards the entrepreneurship and innovation ecosystem more quickly, this process progresses more slowly in organizations subject to bureaucratic processes. In particular, it is observed that incubation centers that contribute more to the development of technology transfer and innovative initiatives grow faster. It is evaluated that these differences have a direct impact on the support processes of entrepreneurs and the effectiveness of incubation centers.

Although there is no significant difference between the first and second ranked clusters, it is observed that these clusters are strong in terms of the number of R&D projects carried out. In addition, it is seen that these clusters attach importance to management and economic processes and have developed in human resources and commercialization. On the other hand, weak clusters consist of incubation firms that are inadequate in the areas of human resources, commercialization, marketing and finance.

In our analysis, K=8 appears to be the optimal choice, as after this point, a stabilization of inertia values is observed and the marginal benefit of additional clusters decreases significantly. This observation is in line with the basic principles of the K-means algorithm and recent research findings. The significant inertia drop in the range from K=5 to K=8 indicates that more detailed structural features in the dataset need to be captured, while the stabilization after K=8 confirms that the optimal number of clusters has been reached and further splitting is unnecessary.

The Silhouette Score was calculated as approximately 0.318, which indicates that the clusters have a medium level of separation quality. While a higher value is expected for an ideal clustering, this score can be considered an acceptable result depending on the data structure. On the other hand, the Davies-Bouldin Index calculated as 0.998 indicates that the separation between clusters and the similarity within clusters are at a good level. This shows that the data is well separated, but there may be some complexity in their internal structure.

Business incubators provide support capability, mentorship and financial assistance to mitigate market risks and optimize start-up sustainability and performance in the market. Incubators offer a unique platform for sustainable development by creating an environment where universities, corporate sponsors, governments, and society intersect. Governments can benefit from the economic growth, job creation, and entrepreneurial ecosystems fostered by these incubators. Lastly, society as a whole experiences the broader impacts of these innovations, such as technological advancements, social solutions, and a more robust economy. This collaborative model emphasizes the importance of synergy between the public and private sectors, reinforcing how educational institutions can serve as catalysts for broader societal change. Universities, technoparks, policymakers and other related learning institutions should design their entrepreneurship, mentoring and acceleration programs to inform researchers and students to picking up entrepreneurial cues from their environment in the ecosystem.

6. Conclusion

In this study, it is aimed to evaluate the clusters obtained as a result of k-means cluster analysis performed with PCA and analyse their performance. Within the scope of the analysis, distances between clusters and various performance indicators of each cluster have taken into consideration, and the cluster analysis results of incubation firms in certain dimensions are evaluated. The analysis are carried out on various metrics such as the number of R&D projects of the firms, patents, grants, customer satisfaction, technology, R&D, competition, investment, marketing, environment and sustainability, human resources and commercialisation. In the study, the similarities and differences of the clusters are evaluated by taking into account the proximity or distance of the clusters to each other. As a result of the study;

Cluster 4; is especially strong in patent and grant areas and shows **the best performance** in general. This cluster is a very strong cluster with its high number of R&D projects and patents.

Cluster 7; is in a leading position in various performance indicators and exhibits good performance. With high R&D project numbers and generally high averages, this cluster shows a strong performance. When the cluster average values of the last 9 dimensions (A4-A12) are investigated, which are metric variables, the highest performance is presented in cluster 7.

Cluster 8; shows good performance in various areas. Despite the low number of R&D projects, this cluster is observed to exhibit good performance with high other metrics.

Cluster 2; has high values in general performance indicators. With high R&D project numbers and generally high averages, this cluster shows a good performance.

Cluster 5; shows average performance in various indicators. Despite the high number of R&D projects, this cluster cannot sufficiently evaluate its potential with low patent and grant numbers.

Cluster 3; shows medium performance. With low R&D project numbers, this cluster is weak in terms of innovation potential but medium in terms of patent numbers.

Cluster 1; shows good performance in some areas despite the low number of R&D projects. With low R&D project numbers and patent numbers, this cluster is weak in terms of innovation and growth potential.

Cluster 6; It generally exhibits **the lowest performance** and has a structure that is quite different from the other clusters. With very low values in all metrics, this cluster has the weakest performance. Based on the cluster average values of the last 9 dimensions (A4-A12), the lowest performance is seen in this cluster .

The expert evaluation is very important to analyse the Table 7. It is also of critical importance for incubation firms to have high scores based on the number of R&D projects, grants and patents dimensions. Therefore, Cluster 4 ranks first in the general ranking with its superior performance and total score in these metrics. Cluster 6, on the other hand, has a structure that is quite different from the other clusters and shows the lowest performance. This analysis provides an important basis for evaluating the performance and potential of incubation firms. In terms of the number of R&D projects, Cluster 7 (4.952) and Cluster 5 (4.943) have the highest values. In terms of patents, Cluster 4 (3.818) is clearly ahead. In terms of grants, Cluster 4 (1.364) has the highest grant value. These three metrics are critical in determining the innovation and financial support capacity of clusters.

The study provides field research and evaluation of important parameters of entrepreneurial firms. Original, technical, objective and subjective analyses were conducted thanks to the perspectives of incubators and entrepreneurs. The view that entrepreneurs with good and bad performance have similar characteristics was supported.

Acknowledgment

This study was funded by The Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Grant Number 122G192).

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459. https://doi.org/10.1002/wics.101

Adomako, S., Amankwah-Amoah, J., Danso, A., Konadu, R., & Owusu-Agyei, S. (2019). Environmental sustainability orientation and performance of family and nonfamily firms. *Business Strategy and the Environment*, *28*(6), 1250-1259. https://doi.org/10.1002/bse.2314

Argaw, Y. M., & Liu, Y. (2024). The Pathway to Startup Success: A Comprehensive Systematic Review of Critical Factors and the Future Research Agenda in Developed and Emerging Markets. *Systems,* 12(12), 541. https://doi.org/10.3390/systems12120541

Ashraf, M. A., Rahim, M. Z. B. A., Qureshi, I. M., & Hanif, M. (2024). Impact of Entrepreneurial Mindset and Alertness on Converting Education into Entrepreneurial Intentions: A Study of Pakistani University Students. *Sustainability*, 16(21), 9345. https://doi.org/10.3390/su16219345

Bacalan, R., Cupin, M., Go, L. A., Manuel, M., Ocampo, L., Kharat, M. G., & Promentilla, M. A. (2019). The incubatees' perspective on identifying priority enabling factors for technology business incubators. *Engineering Management Journal*, *31*(3), 177-192. https://doi.org/10.1080/10429247.2018.1540225

Bertoni, F., Martí, J., & Reverte, C. (2019). The impact of government-supported participative loans on the growth of entrepreneurial ventures. *Research Policy*, *48*(1), 371-384. https://doi.org/10.1016/j.respol.2018.09.006

Bolukbas, U., & Guneri, A. F. (2018). Knowledge-based decision making for the technology competency analysis of manufacturing enterprises. *Applied Soft Computing*, *67*, 781-799. https://doi.org/10.1016/j.asoc.2017.11.023

Busulwa, R., Birdthistle, N., & Dunn, S. (2020). *Startup accelerators: A field guide*. John Wiley & Sons.

Cheah, S. L. Y., Ho, Y. P., & Li, S. (2021). Search strategy, innovation and financial performance of firms in process industries. *Technovation*, *105*, 102257. https://doi.org/10.1016/j.technovation.2021.102257

Collins, M., Dasgupta, S., & Schapire, R. E. (2001). A generalization of principal components analysis to the exponential family. *Advances in neural information processing systems*, *14*. https://doi.org/10.7551/mitpress/1120.003.0084

Crammond, R. J. (2024). Evaluating intellectual capital within developmental entrepreneurial environments: conceptualising the sequential role of education and ecosystems. *Journal of Intellectual Capital*, 25(5/6), 914-941. https://doi.org/10.1108/JIC-05-2024-0148

Cubero, J. N., Gbadegeshin, S. A., & Consolación, C. (2021). Commercialization of disruptive innovations: Literature review and proposal for a process framework. *International Journal of Innovation Studies*, *5*(3), 127-144. https://doi.org/10.1016/j.ijis.2021.07.001

Dai, S., Xue, H., Jiang, Y., Zhang, W., & Zhang, X. (2018). Sustainable entrepreneurship team scale development: A complex systems perspective. *Sustainability*, *10*(11), 4199. https://doi.org/10.3390/su10114199

Deyanova, K., Brehmer, N., Lapidus, A., Tiberius, V., & Walsh, S. (2022). Hatching start-ups for sustainable growth: a bibliometric review on business incubators. *Review of Managerial Science*, 16(7), 2083-2109. https://doi.org/10.1016/j.tej.2020.106904

Estep, J., Daim, T., & Shaygan, A. (2021). R&D project evaluation: Technology transfer focus. *The Electricity Journal*, *34*(2), 106904. https://doi.org/10.1016/j.tej.2020.106904

Font-Cot, F., Lara-Navarra, P., Sánchez-Arnau, C., & Sánchez-Pérez, E. A. (2025). Startup Survival Forecasting: A Multivariate AI Approach Based on Empirical Knowledge. *Information,* 16(1), 61. https://doi.org/10.3390/info16010061

Geissdoerfer, M., Vladimirova, D., Van Fossen, K., & Evans, S. (2018). Product, service, and business model innovation: A discussion. *Procedia Manufacturing*, *21*, 165-172. https://doi.org/10.1016/j.promfg.2018.02.107

Gersho, A., & Gray, R. M. (2012). *Vector quantization and signal compression* (Vol. 159). Springer Science & Business Media.

Hagen, B., Zucchella, A., Cerchiello, P., & De Giovanni, N. (2012). International strategy and performance—Clustering strategic types of SMEs. *International Business Review*, *21*(3), 369-382. https://doi.org/10.1016/j.ibusrev.2011.04.002

Heckler, C. E. (1996). A step-by-step approach to using the SAS™ system for factor analysis and structural equation modeling. https://doi.org/10.2307/1270628

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178-210. https://doi.org/10.1016/j.ins.2022.11.139

Innocenti, N., & Zampi, V. (2019). What does a start-up need to grow? An empirical approach for Italian innovative start-ups. *International Journal of Entrepreneurial Behavior & Research*, *25*(2), 376-393. https://doi.org/10.1108/IJEBR-04-2018-0194

Jang, J. S. R., Sun, C. T., & Mizutani, E. (1997). Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. *IEEE Transactions on automatic control*, *42*(10), 1482-1484. https://doi.org/10.1109/TAC.1997.633847

Jha, V. K., & Pande, A. S. (2024). Making sustainable development happen: Does sustainable entrepreneurship make nations more sustainable?. *Journal of Cleaner Production*, 440, 140849. https://doi.org/10.1016/j.jclepro.2024.140849

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical* *and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1016/j.jclepro.2024.140849

Karambakuwa, J. K., & Bayat, M. S. (2022). Understanding entrepreneurship training in incubation hubs. *Indonesian Journal of Innovation and Applied Sciences (IJIAS)*, *2*(3), 168-179. https://doi.org/10.47540/ijias.v2i3.568

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. International Journal, 1(6), 90-95.

Le Trinh, T. (2019). Factors affecting startup performance of small and medium-sized enterprises in Danang city. *Entrepreneurial Business and Economics Review*, *7*(3), 187-203. https://doi.org/10.15678/EBER.2019.070310

Linde, Y., Buzo, A., & Gray, R. (1980). An algorithm for vector quantizer design. *IEEE Transactions on communications*, *28*(1), 84-95. https://doi.org/10.1109/TCOM.1980.1094577

Luo, X., Huang, F., Tang, X., & Li, J. (2020). Government subsidies and firm performance: Evidence from high-tech start-ups in China. *Emerging Markets Review*, *49*, 100756. https://doi.org/10.1016/j.ememar.2020.100756

Matricano, D. (2020). The effect of R&D investments, highly skilled employees, and patents on the performance of Italian innovative startups. *Technology Analysis & Strategic Management*, *32*(10), 1195-1208. https://doi.org/10.1080/09537325.2020.1757057

Olufunke, B., Babatunde, M., & Gbenga, A. (2020). University Business Incubation and Entrepreneurial Society: An Empirical Review. *ADEDIPE, Oluwaseyi Ayodele OLADEJI*, Ige Olubunmi, 223.

Pearce, D. D., & Pearce II, J. A. (2020). Distinguishing attributes of high-growth ventures. *Business Horizons*, *63*(1), 23-36. https://doi.org/10.1016/j.bushor.2019.10.003

Petrovic, S. (2006, October). A comparison between the silhouette index and the davies-bouldin index in labelling ids clusters. In *Proceedings of the 11th Nordic workshop of secure IT systems* (Vol. 2006, pp. 53-64). Citeseer.

Prokop, D., & Thompson, P. (2023). Defining networks in entrepreneurial ecosystems: The openness of ecosystems. *Small Business Economics*, 61(2), 517-538. https://doi.org/10.1007/s11187-022-00710-w

Pugliese, R., Bortoluzzi, G., & Balzano, M. (2021). What drives the growth of start-up firms? A tool for mapping the state-of-the-art of the empirical literature. *European Journal of Innovation Management*, *25*(6), 242-272. https://doi.org/10.1108/EJIM-03-2021-0163

Quero-Gervilla, M. J., Díaz-Méndez, M., Ventura-Fernández, R., & Gummesson, E. (2022). Co-patenting, co-ownership, and co-ideation as drivers for university business innovation: the case of public universities in Spain. https://doi.org/10.1108/TQM-03-2022-0087

Rai, R. S., Prasad, A., & Murthy, B. K. (2025). Incubation support for academia-based entrepreneurship: an empirical study. *Journal of Asia Business Studies*. https://doi.org/10.1108/JABS-02-2024-0102

Rannikko, H., Tornikoski, E. T., Isaksson, A., & Löfsten, H. (2019). Survival and growth patterns among new technology-based firms: Empirical study of cohort 2006 in Sweden. *Journal of Small Business Management*, *57*(2), 640-657. https://doi.org/10.1111/jsbm.12428

Ratten, V. (2020). Coronavirus disease (COVID-19) and sport entrepreneurship. *International Journal of Entrepreneurial Behavior & Research*, *26*(6), 1379-1388. https://doi.org/10.1108/IJEBR-06-2020-0387

Rompho, N. (2018). Operational performance measures for startups. *Measuring Business Excellence*, *22*(1), 31-41. https://doi.org/10.1108/MBE-06-2017-0028

Seo, S. Y., Kim, S. D., & Lee, M. S. (2018). The effects of knowledge assets on the performances of startup firms: Moderating effects of promotion focus. *The Journal of Asian Finance, Economics and Business*, *5*(4), 187-199. https://doi.org/10.13106/jafeb.2018.vol5.no4.187

Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)* (pp. 747-748). IEEE. https://doi.org/10.1109/DSAA49011.2020.00096

Silva Júnior, C. R., Siluk, J. C. M., Neuenfeldt Júnior, A., Francescatto, M., & Michelin, C. (2022). A competitiveness measurement system of Brazilian start-ups. *International Journal of Productivity and Performance Management*, *72*(10), 2919-2948. https://doi.org/10.1108/IJPPM-02-2022-0098

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, 8, 80716-80727. https://doi.org/10.1109/ACCESS.2020.2988796

Solano, G., Larrañeta, B., & Aguilar, R. (2020). Absorptive capacity balance and new venture performance: cultivating knowledge from regional clusters. *Technology Analysis & Strategic Management*, *32*(11), 1264-1276. https://doi.org/10.1080/09537325.2020.1760236

Staniewski, M. W., & Awruk, K. (2019). Entrepreneurial success and achievement motivation–A preliminary report on a validation study of the questionnaire of entrepreneurial success. *Journal of Business Research*, *101*, 433-440. https://doi.org/10.1016/j.jbusres.2019.01.073

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering* (Vol. 336, p. 012017). IOP Publishing. https://doi.org/10.1088/1757-899X/336/1/012017

Thukral, E. (2021). COVID-19: Small and medium enterprises challenges and responses with creativity, innovation, and entrepreneurship. *Strategic Change*, *30*(2), 153-158. https://doi.org/10.1002/jsc.2399

Udell, G. G. (1990). Are business incubators really creating new jobs by creating new business and new products. *Journal of Product Innovation Management*, *7*(2), 108-122. https://doi.org/10.1111/1540-5885.720108

Van Rijnsoever, F. J., & Eveleens, C. P. (2021). Money Don't matter? How incubation experience affects start-up entrepreneurs' resource valuation. *Technovation*, *106*, 102294. https://doi.org/10.1016/j.technovation.2021.102294

Verbano, C., Crema, M., & Venturini, K. (2015). The identification and characterization of open innovation profiles in Italian small and medium-sized enterprises. *Journal of Small Business Management*, *53*(4), 1052-1075. https://doi.org/10.1111/jsbm.12091