# 7 Cluster analysis

## Grouping workers by work location choice

*Alessandra Migliore and Cristina Rossi-Lamastra*

*Politecnico di Milano, Italy*

### 7.1 Background

Cluster analysis has been used in quantitative research since the 1970s to group observations into mutually exclusive groups. In so doing, it helps systematize information contained in data (Hair, 2009). Specifically, cluster analysis groups observations through *hierarchical* and *non-hierarchical* algorithms.[1]

Cluster analysis groups observations to maximize *within-cluster homogeneity* and *between-cluster heterogeneity* (Aldenderfer & Blashfield, 1984; Everitt et al., 2011). To assess homogeneity/heterogeneity, all cluster analyses consider measures of dissimilarity/similarity among observations (*e.g.*, the Euclidean distance or correlation matrices). Observations with low or high correlations are assigned to the same cluster, whereas those with high or low correlations are assigned to different clusters. Cluster analysis is well-suited for data exploration. Indeed, the number of clusters is not usually defined *a priori,* and how observations group together emerges *from the bottom up*. Accordingly, the methodology can offer rigorous support to preliminary ideas (*e.g.*, coming from data inspection) on how observations fit together and can be a valuable starting point for hypotheses formulation and testing (Johnson & Wichern, 2002).

Cluster analysis emerged as a meaningful methodology across several disciplines (Hair, 2009). Application domains include (but are not limited to) marketing (*e.g.,* partitioning of consumers into market segments, see Tsiptis & Chorianopoulos, 2009), biological sciences and genomics (*e.g.*, building groups of genes with similar expression patterns, see Eisen et al., 1998), and operation management (*e.g.,* grouping manufacturing strategies across industries, see Frohlich & Dixon, 2001). Cluster analysis has recently been applied to support coding (extensive) corpora of texts to unearth meaningful categories (Namey et al., 2007).

In recent years, workplace research has started to resort to cluster analysis, mainly for grouping (*i.e.,* profiling) workers based on multiple dimensions, including (self-reported) health and comfort (Kim & Bluyssen, 2020), control of indoor climate (Hong et al., 2020), indoor environmental quality (IEQ, Ortiz & Bluyssen, 2022), work motivation (Basińska, 2020), and resignation intention (Wang, 2021). Overall, there are only limited applications of cluster analysis to workplace research. For instance, workplace scholars have rarely used cluster analysis for

DOI: 10.1201/9781003289845-7

profiling workers based on *where* they decide to work and *when* they can adopt multi-local working. The diffusion of *flexible work* (Halford, 2005; Richardson & McKenna, 2013) and of *new ways of working* (Aroles et al., 2019) has put workers' location choices into the limelight, especially in the wake of the COVID-19 pandemic (Mallett et al., 2020). Accordingly, this chapter focuses on two-step cluster analysis and shows its adoption when analyzing workers' location choices.

Two-step cluster analysis uses *hierarchical* and *non-hierarchical* algorithms in tandem (Ketchen & Shook, 1996). Specifically, in the first step, the *hierarchical* clustering procedure developed by Ward (1963) determines the number of clusters and their centroids.[2] In the second step, partitioned or *non-hierarchical* clustering (*e.g.,* k-means clustering) allocates observations to clusters. Scholars (*e.g.,* Frohlich & Dixon, 2001; Ketchen & Shook, 1996) concur that combining the hierarchical and non-hierarchical approaches is more effective than resorting to only one. Accordingly, this chapter illustrates the potential of cluster analysis to advance workplace research on multi-local work. In this case, the unit of analysis is the individual worker and his/her own choices. We also discuss how this technique can help firms and other organizations to profile their workforce based on their work locations' preferences and needs; this profiling, in turn, can support organizations in designing their workplace policies.

## 7.2 Argument

Even if cluster analysis is a well-established methodology in social sciences, it is not commonly used in research on workplaces and research related to the built environment. Namely, few studies use (two-step) cluster analysis for grouping individuals and their relations with the workspaces. In workplace research, scholars mainly used cluster analysis to define workers' profiles in terms of roles, tasks, and workers' feelings. For instance, Soriano et al. (2020) grouped workers according to two main variables relating to their work type: degree of task complexity and degree of interaction with other people at work. After finding four groups of employees, the study associated each group with its recommended type of space, distinguishing the "fit" (i.e., workers in an adequate office space for their work type) and the "misfit" group (i.e., workers in an inadequate office space for their work type). This association is derived from the predefined assumption that a specific task needs a type of office. Thus, while the study identified *from the bottom-up* groups of employees based on the types of tasks that they perform at work (through cluster analysis), it assigned *a priori* the office environment "adequate" for each group, without considering employees' preferences for doing a certain activity in a certain space. Performing cluster analysis with data on individual choices and preferences could provide useful information on what employees want. This study shows how to apply cluster analysis when dealing with individual choices over work location. Studying workers' location choices is crucial in the contemporary working context. The increasing possibility of choosing work locations – also enabled by Information and Communication Technologies (ICT) – stimulates workers to reflect upon their preferences and needs. At the same time, organizations treasure the outcome of these reflections.

When appointed with autonomy in deciding where to work, workers may choose different locations for their work. Some workers choose to work exclusively at the office; others prefer to work solely from home, while others mix both locations (Halford, 2005; Hislop & Axtell, 2007). The motivation to choose one workspace or the other, or to alternate between them, may depend on different factors, including work tasks and the cost of traveling (*i.e.,* in monetary terms but also in terms of time and effort) (Brown & O'Hara, 2003). Firms and other organizations are interested in workers' preferences and needs regarding work locations because they should decide about their *location flexibility* (*e.g.*, where workers can work, how far from the firm's premises, and when they have to be at the office) to balance their objectives (including those dealing with real estate assets) and those of their workers. For instance, some of the recent workplace flexibility (or inflexibility) policies adopted by firms due to the pandemic have created tensions between workers' quest for flexibility and the organizations' implementation costs.

Organizations address these issues in different ways. Some disregard their workforce's preferences and needs and define their work arrangement policies with a *top-down* approach; a case in point is the recent Tesla CEO's *ultimatum* to his employees to return to the office (Nicholas & Hull, 2021). Other organizations rely on descriptive results obtained from samples of workers and generalize them to the whole workforce, in line with a *one-size-fits-all approach* according to which, supposedly, all workers' preferences and needs coincide with those of the sampled workers. Such systems are dangerous: recent studies (Morning Consult, 2022) report that 55% of remote workers would consider resigning if their firms tried to force them to return to the office. More and more people are quitting their job, a phenomenon popularized as *great resignation*[3] or *great discontent* (Hirsch, 2021), partially motivated by inflexible work arrangements. Although we lack data to understand whether this phenomenon is a short-lived trend – which media amplify – or a long-lasting effect of the COVID-19 pandemic (Ksinan Jiskrova, 2022), it showcases the importance for organizations to make decisions that their workers embrace (Bailey & Rehman, 2022). In such a context, cluster analysis may help decision-makers understand *how many* groups with different preferences and needs emerge from data covering several dimensions (*e.g.*, workers' age, gender, commuting time, and family burden).

Even though other methods exist that can address these issues (*e.g.,* mostly qualitative methods, including focus groups, one-to-one interviews), cluster analysis offers some unique advantages. First, it allows managing large datasets such as the workforce of large firms with many employees and, thus, high heterogeneity in workers' preferences and needs. Second, it enables the replicability of the analysis. Indeed, thanks to widely diffused statistical software packages (*e.g.,* STATA, R, or SPSS), cluster analysis has become an easy-to-run procedure that requires just an adequate knowledge of statistics. Third, partially related to the second point, cluster analysis allows for quick simulations of different scenarios (e.g., what happens when an additional variable is loaded in the analysis). It is worth noting that cluster analysis may return misleading results because of shortcomings in data collection, variable selection, execution of the procedure steps, and tests run on the outcome

(Ketchen & Shook, 1996). In the next section, an example is provided of an application to showcase how two-step cluster analysis works and how to improve its application for future inquiry.

## 7.3 Example of application/use

This example shows how the authors applied cluster analysis on a large dataset containing information about how Italian academics chose their work locations during the COVID-19 pandemic. The investigation aimed to understand *how many* and *which types* of location choices would emerge for these workers once the strict lockdown in Italy was over (*i.e.,* at the beginning of May 2020, Bontempi, 2021). We expected to find two main groups: those who persisted in working only from home and those who moved between home and the university office. Nonetheless, some surprising evidence is unearthed.

Although some seminal contributions rely on a non-hierarchical algorithm (also known as the k-means or iterative method; see, *e.g.*, Miller & Roth, 1994) where the number of clusters is defined in advance, a two-step clustering procedure is nowadays strongly advised because of its higher validity and reliability (Frohlich & Dixon, 2001; Ketchen & Shook, 1996), especially in workplace research (Soriano et al., 2020). A hierarchical method first determines the number of clusters and cluster centroids; then, it uses them as inputs of the subsequent non-hierarchical algorithm. The following sections explain how to collect and prepare data for the analysis, how two-step cluster analysis groups the observations in the two steps of the procedure, and how to interpret the results.

### *7.3.1 Data collection*

After the first wave of the COVID-19 pandemic (that in Italy ended in May 2020, Bontempi, 2021), a survey was administered via email to the entire population of tenured Italian academics, whose contacts are publicly listed by the Italian Ministry of University and Research (MUR).[4] The target population consisted of 52,630 Italian scholars. Participation in the survey was voluntary and confidential; the survey stayed open from July 24 to September 24, 2020. Among the others, the survey collected data on how often academics worked from multiple work locations (*i.e.,* weekly frequency of access from "never" to "more than five times per week" to the home, the university office, and other third locations of work). The variables included in the cluster analysis were based on these data. *Covid_University* captured the weekly frequency of access to the university for working during the COVID-19 pandemic; *Covid_Home* captured the weekly frequency of access to the home for working during the COVID-19 pandemic; *Covid_Otherspace* captured the weekly frequency of access to other spaces[5] for working during the COVID-19 pandemic.

Noteworthy, there are (in principle) no limitations in the number of variables that one can include in a two-step cluster; however, it is preferable to have a limited number of variables, wisely chosen according to the literature.[6] In the presence of high correlations, one can reduce the number of cluster variables through a principal component analysis (PCA, *e.g.*, Ortiz & Bluyssen, 2022). PCA reduces the correlated variables into fewer independent components, thus, ultimately, solving the problem of multicollinearity[7] (Pacáková & Poláčková, 2013). Ketchen and Shook (1996) criticized PCA because it drops the components with low eigenvalues (a measure of the amount of variance explained by a member). The excluded components may provide unique, important information, and this exclusion may result in a sub-optimal set of clusters. Following Ketchen and Shook (1996), we advise repeating cluster analysis by trying multiple methods for addressing multicollinearity (*e.g.,* PCA or variable standardization[8]) to see how each method may differently affect the results.

### *7.3.2 Data analysis*

We received 11,634 responses, which required cleaning to get rid of incomplete responses. Finally, we obtained 7,865 usable and consistent answers (response rate: 14.94%). The sample included 3,853 women (48.99%) and 4,012 men (51.01%); respondents were on average 51 years old, work in universities located in the North (48.29%), Centre (25.86%), South (25.85%) of Italy, and belong to many scientific fields. Once we selected the variables, we adopted the two-step cluster analysis.

As a first step, we used the hierarchical cluster procedure developed by Ward (1963) to determine the number of clusters and their centroids. We adopted Ward's partitioning and squared Euclidean distance because of its robustness and solidity in maximizing within-cluster homogeneity and between-cluster heterogeneity (Aldenderfer & Blashfield, 1984; Basińska, 2020; Everitt et al., 2011; Frohlich & Dixon, 2001). We referred to the Duda–Hart stopping rule and the Calinski–Harabasz pseudo-F yields equivalent clustering to determine clusters' number and centroids.[9] We also visually inspected the dendrogram[10] to confirm the number of resulting clusters during this step. Namely, starting from the top-down diagram, we detected the number of branches, while starting from the bottom-up diagram, we looked at the points of joining of the branches. Both stopping rules and dendrogram inspection suggested the existence of four clusters in our data.

As a second step, we assigned sampled observations to the four clusters through the k-medians non-hierarchical clustering method. The k-median method allows using the vectors of medians of the variables as centroids.[11] This gave us more consistent and reasonable results than using the k-means clustering method, which instead uses vectors of means.

Finally, to check whether original variables significantly differ across clusters, we ran the one-way analysis of variance (ANOVA) for pairwise comparison of means with a Scheffe post-hoc test. This test was crucial to understanding if the clusters were reasonably defined and profile distinct groups of workers. In addition, to check whether academics in each cluster changed their habits because of the COVID-19 pandemic, we performed matched pairs t-tests within each cluster for variables capturing time spent at different work locations before (*Before_University; Before_Home; Before_Otherspace*[12]) and during COVID-19 (*Covid_University; Covid_Home; Covid_Otherspace*).

## 7.4 Results

After the abovementioned tests, we confirmed the emergence of four clusters grouping Italian academics according to their location choices. After discussion between the authors, we labeled the four clusters as (1) *home-centric*, (2) *between home and university*, (3) *multi-located*, and (4) *university-centric* (Table 7.1). The labels point to each cluster's main features (*i.e.*, the main location that academics accessed). Therefore, the cluster *home-centric* (Cluster 1) includes a large group of academics (4,564 observations, 58.03% of the sample) that worked solely from home in the observed period (on average 5.395 times per week). The cluster *between home and university* (Cluster 2) collected those who balanced their research activity between home and university but rarely use other spaces and covered one-fourth of the academics in the sample (1,187; 25.26%). The cluster *multi-located* (Cluster 3) isolated those who did research from other spaces (on average 4.508 times per week) more often than from home (on average 3.947 times per week) and sometimes accessed also the campus (on average 1.614 times per week); this group consists of a few academics (368, 4,68%). Finally, the cluster *university-centric* (Cluster 4) summed up those working mainly on-campus (on average, 4.786 times per week) – a relatively small percentage of academics (946, 12.03%).

Based on the Scheffe post hoc test results, we found significant differences across clusters for each variable. The same superscript label in Table 7.1 indicates that the variable's mean is not significantly different in various clusters; this happens only in one case (*Covid_Home* in Clusters 2 and 3).

For relevance, we compare the mean frequencies of access to university, home, and other spaces before and during the COVID-19 period through matched pair t-test. This test's results show that the means of all the variables measuring the frequency of access to the different work locations are different before and during-COVID, suggesting that all academics changed their work location choices because of the pandemic (see Figure 7.1).

Important information that emerged from the clusters helped interpret the phenomenon of multi-location work, including clusters' dimensions, and workers'

*Table 7.1* Cluster analysis results

| | | *Cluster 1 – Home-centric (n=4,564)* | *Cluster 2 – Between Home and University (n=1,987)* | *Cluster 3 – Multi-located (n=368)* | *Cluster 4 – University-centric (n=946)* |
|---|---|---|---|---|---|
| Variables | Sample mean | Mean | Mean | Mean | Mean |
| *Covid_University* | 1.424 | 0.255 [d] | 2.474 [b] | 1.614 [c] | 4.786 [a] |
| *Covid_Home* | 4.430 | 5.395 [a] | 3.846 [b] | 3.947 [b] | 1.188 [c] |
| *Covid_Otherspaces* | 0.491 | 0.184 [d] | 0.355 [c] | 4.508 [a] | 0.699 [b] |

*Note*: Based on ANOVA tests, the means of all the variables are significantly different among clusters at 99%. Note that the highest mean of each variable is labelled with "a," the next highest mean with "b" and "c," and the lowest mean with "d." The same superscript label indicates that the variable's mean is not significantly different in the various clusters.

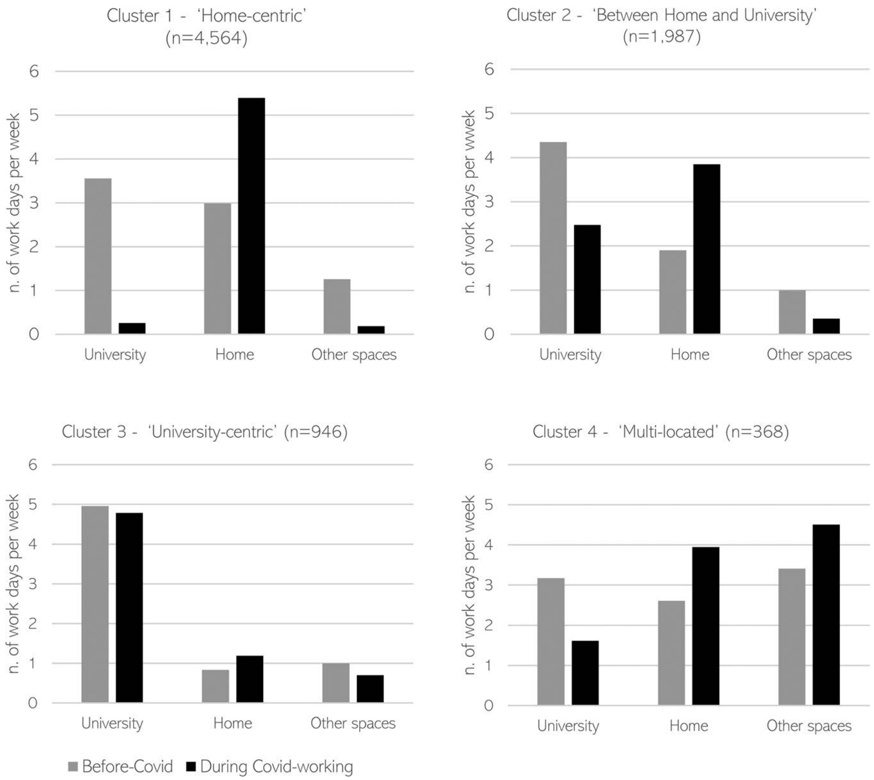


*Figure 7.1* Means of frequency of access to university, home, and other spaces before and during COVID-19 in each cluster

distributions among clusters. For instance, academics belonging to the *home-centric cluster* tended to do research from home already about three times per week before the pandemic. On the contrary, those in the university-centric cluster spent limited time at home both during and before the pandemic. This suggests some inertia in work location choices, which may entail home-centric academics being reluctant to return to the university after the pandemic.

Additional analysis – that we cannot describe in this chapter – was necessary to explain the motivations of cluster adherence and advance knowledge of academics' location choices. Therefore, we employed a multivariate analysis to learn more about the demographic characteristics of academics in the four clusters.

## 7.5 Implications

### *7.5.1 Method relevance to research*

The reported application shows that cluster analysis is highly relevant for workplace research. In the first place, cluster analysis can inform research that aims at profiling categories of users of a workplace to understand their needs, their relations with other users, and their relations with the built environment. Cluster analysis will likely advance workplace research in the future.

First, researchers can easily explore the composition of the clusters (*e.g.*, by statistical tests such as t-test and ANOVA test) to know who belongs to each cluster. For instance, does the cluster include more women than men? More young workers than older ones? More workers engaged in collaborative work than workers carrying on individual tasks? Do employees performing similar work tasks belong to the same cluster?

Second, scholars can understand the temporal evolution of workers' groups by comparing cluster analyses conducted on the same organization in different periods. For instance, they can investigate whether and how clusters of workers tweak after a focal event (*e.g.*, the change in the office location). Clustering workers according to their commuting features (*e.g.*, travel costs, travel time, perceived effort to reach the office, Brown & O'Hara, 2003) sheds light on how many workers likely prefer to work near their main office. Therefore, we believe that the use of cluster analysis contributes to conversations on remote and flexible work.

Third, scholars can compare different organizations by observing how their workers group together according to different measures (*e.g.,* work tasks, work motivations, workplace attachment, etc.). In turn, this might offer interesting insights into heterogeneity in organizational cultures and employees' attachment.

Fourth, by combining cluster analysis with other methodologies (*e.g.,* econometric models), researchers may unearth what influences adherence to the clusters. For instance, one can understand why some workers decide to work from a specific venue and how this choice affects their performances.

Fifth, cluster analysis opens avenues for future methodological advancements in workplace research by overcoming some shortcomings of qualitative methods (*e.g.*, managing large datasets, identifying groups from the bottom up, dealing with subjectivity, and enabling replicability). Along this line of reasoning, scholars can combine cluster analysis with qualitative approaches. This mixed-method approach (Bryman & Bell, 2007) is gaining momentum in social sciences as it allows the depiction of a highly nuanced picture of phenomena.

Finally, cluster analysis may pave new ways of analyzing qualitative data. A telling example consists in coding qualitative text-based data (*e.g.,* data from interviews) into quantitative variables and then applying cluster analysis to these variables (Namey et al., 2007). For instance, one can think of grouping workers according to their experiences in an environment.

### *7.5.2 Method relevance to practice*

Cluster analysis results can inform workplace management and corporate real estate strategies. Indeed, amid organizational and environmental constraints, it holds the potential to support organizations in grouping their workforce according to their features, preferences, and needs in a systematic way, rather than relying on anecdotal information.

Profiling workers offers interesting insights to human resources managers (HRMs) who want to promote job autonomy of their workers (Basińska, 2020). It is reasonable to expect that profiling workers according to their motivations improves HRMs' decisions about redesigning the job of some groups (e.g., granting autonomy in task execution).

Likewise, workplace managers' decisions can benefit from cluster analysis because it helps to understand and quantify how different workers' groups may appreciate different work environments, thus ultimately nurturing workers' well-being (Manganelli et al., 2018). For example, it might be wise to locate workers with interdependent tasks on the same floor of an office building or allow those performing independent tasks to work from home. Overall, cluster analysis helps in the design of flexible working policies. This is a timely issue as more and more workplace managers are looking for answers to the following questions: how many days per week can workers work out of the office? Which work locations should they have access to? Should the organization establish satellite offices in some locations (e.g., near where workers reside)?

Finally, firms have been reflecting upon their real estate assets over the last decades. The current adverse macro-economic contingencies (e.g., the increase in electricity prices, the need to reduce CO2 emissions, the endurance of the COVID-19 pandemic) have urged such reflections and are now calling to action. In such a scenario, grouping employees through cluster analysis is a tool easing the design and implementation of firms' real estate strategies. For example, as the pandemic has made people accustomed to remote working, how much space do firms need for their core activities? How much space could eventually be leased or sold? Where should firms (re)locate their premises to promote workers' well-being and reduce their environmental impacts?

## 7.6 Conclusions

Nowadays, firms are strongly oriented toward a user-centered approach when designing their workplace and real estate strategies. In such a contest, cluster analysis can be a powerful tool for achieving the goal of understanding in more depth people behaviors. This chapter showcases how cluster analysis – especially the two-step procedure[13] we discussed above – can advance theoretical and empirical research on the workplace in multiple directions. Moreover, it brings an array of exciting practical implications. The robustness and reliability of the cluster analysis results speak in favor of adopting it beyond its mainstream fields of applications (e.g., strategic management, marketing). As any other method, cluster analysis has limitations that open avenues for future research and call relevant stakeholders to take actions.

First, the technique requires a large amount of data. This is undoubtedly a limitation for many workplace researchers and practitioners, given the well-known difficulties in data gathering (e.g., high data collection costs and workers' reluctance to disclose information about their habits). One can overcome data shortage in the short run by resorting to secondary sources (e.g., the European Working Condition

Survey – EWCS[14]). In the long run, we believe that the quest for data will encourage researchers and organizations to develop data collection systems based on cutting-edge technologies (e.g., artificial intelligence, machine learning, and IoT-based solutions). These systems have pros and cons. On the one side, they collect data quickly and automatically, thus reducing the time and effort required to gather data directly from people. On the other side, they pose entangled privacy issues.

Second, adopting cluster analysis requires good statistical skills and tailored software. Accordingly, researchers and practitioners must climb the learning curve by developing new skills and mastering new abilities in computer science. Indeed, implementation mistakes may lead to false evidence and, consequently, to wrong decisions.

## 7.7 Further Reading

A list of four recommended papers can be found below for those who intend to deepen cluster analysis technique and application. The first paper is one of the seminal works on how to apply cluster analysis and on its main limitations. The other three papers show three applications of cluster analysis on workplace research. While the paper from Basińska groups employees based on work motivation and performance contributing to psychology, the last two papers apply cluster analysis contributing to research in the indoor built environment.

- Ketchen, D., & Shook, C. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal, 17*(6), 441–459.
- Basińska B. A. (2020). Work motivation profiles and work performance in a group of corporate employees: A two-step cluster analysis, *Annals of Psychology*, *23*(3), 227–245.
- Kim, D. H., & Bluyssen, P. M. (2020). Clustering of office workers from the OFFICAIR study in The Netherlands based on their self-reported health and comfort. *Building and Environment, 176*, 106860. https://doi.org/10.1016/j.buildenv.2020.106860
- Soriano, A., Kozusznik, M. W., Peiró, J. M., & Mateo, C. (2020). The role of employees' work patterns and office type fit (and misfit) in the relationships between employee well-being and performance. *Environment and Behavior*, *52*(2), 111–138. https://doi.org/10.1177/0013916518794260

## Notes

1 *Hierarchical algorithms* progress through several steps that build a tree-like structure by either adding observations to (*i.e.*, agglomerative) or deleting them from (*i.e.*, divisive) clusters. The most popular hierarchical algorithms used are agglomerative; they include single linkage, complete linkage, average linkage, centroid method, and Ward's method (Hair, 2009). Selection among the different algorithms depends on the underlying structure of the data (*i.e.*, sample size, distribution of observations, and types of variables included) (Ketchen & Shook, 1996).

*Non-hierarchical algorithms* (also referred to as k-means or iterative methods) partition a data set into a pre-specified number of clusters (Hair, 2009).

2 The centroid of a cluster is the *center point of the cluster along input variables* (Ketchen & Shook, 1996). For each cluster, it corresponds to the vector of the means of the variables used for cluster analysis.

3 The term Great Resignation was coined by Anthony Koltz, Professor of Management at Mays Business School at Texas, A&M University for describing his prediction of a huge wave of people quitting their jobs in the near future. See the following editorial: Impact of 'The Great Resignation' on organizational knowledge and skills (2022). Business Information Review. https://doi.org/10.1177/02663821221101814

4 These lists include all the Italian scholars tenured in Italian public universities but exclude Ph.D. students, post-doc researchers, and research grant holders — Source of the lists: https://cercauniversita.cineca.it/php5/docenti/cerca.php

5 In line with the extant literature, "other spaces" in our survey include (i) in transit, (ii) at other universities, research centres, labs, or companies, (iii) third spaces as coworking spaces, archives, public libraries, bars, open-air, and parks, and (iv) other environments related to fieldworks or private offices.

6 Of note, we might acknowledge that even if cluster analysis is an explorative methodology, it draws on a post-positivist paradigm that recognizes the relevance of the researcher in shaping scientific knowledge.

7 Usually, the term multicollinearity refers to correlation among the explanatory variables in multivariate regressions (Goldberger, 1991, pp. 245–253). In the case of cluster analysis, it refers to high correlation among cluster variables.

8 Standardization is a procedure of re-scaling variables. Standardization transforms variables so that each has a mean of zero and a standard deviation of one. In the case of cluster analysis, it allows variables to contribute equally to the definition of clusters. However, it may reduce heterogeneity among variables (Ketchen and Shook, 1996).

9 *Duda–Hart* and *Calinski–Harabasz pseudo-F* stopping rules imply the calculation of indices whose values help understand the best number of clusters resulting from the dataset. These stopping rules follow the hierarchical algorithm of the cluster analysis.

Calinski–Harabasz pseudo-F stopping rules requires computing the Calinski–Harabasz pseudo-F measure, which looks at the sum of squared distances within the partitions (*i.e.*, data in the same cluster), and compares it to that in the unpartitioned data (*i.e.*, data in different clusters), taking account of the number of clusters and number of cases (Calinski and Harabasz, 1974).

The Duda–Hart index is similar to the Calinski–Harabasz pseudo-F. It involves the computation of the Duda–Hart index, which is simply the sum of squares in the two clusters divided by the sum of squares in the combined cluster (Duda et al., 2000).

10 Branching diagram with a tree-like structure showing the relations between observations used in cluster analysis. The dendrogram first groups individual observations and then merges groups until only a unique group is obtained.

11 We used the Stata commands cluster *wardslinkage* and cluster *kmedians* to run the first and second steps of the cluster analysis.

12 *Before_University* captures the weekly frequency of access to the university for working before the COVID-19 pandemic; *Before_Home* captures the weekly frequency of access to the home for working before the COVID19 pandemic; *Before_Otherspace* captures the weekly frequency of access to other spaces for working before the COVID-19 pandemic.

13 As explained in the background section, two-step cluster analysis uses hierarchical and non-hierarchical algorithms in tandem.

14 For further information on the survey see https://www.eurofound.europa.eu/surveys/european-working-conditions-surveys-ewcs

## References


Aldenderfer, M. S., & Blashfield, R. K. (1984). *Cluster analysis*. Sage.

Aroles, J., Mitev, N., & de Vaujany, F. (2019). Mapping themes in the study of new work practices. *New Technology, Work and Employment*, *34*(3), 285–299.

Bailey, J. R., & Rehman, S. (2022, February 14). How to overcome return to office resistance. *Harvard Business Review*. https://hbr.org/2022/02/how-to-overcome-return-to-office-resistance

Basińska, B. A. (2020). Work motivation profiles and work performance in a group of corporate employees: A two-step cluster analysis. *Annals of Psychology*, *23*(3), 227–245.

Bontempi, E. (2021). The Europe second wave of COVID-19 infection and the Italy "strange" situation. *Environmental Research*, *193*, 110476. https://doi.org/10.1016/j.envres.2020.110476

Brown, B., & O'Hara, K. (2003). Place as a practical concern of mobile workers. *Environment and Planning A*, *35*(9), 1565–1587. https://doi.org/10.1068/a34231

Bryman, A., & Bell, E. (2007). *Business research methods* (2nd ed.). Oxford University Press.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, *3*(1), 1–27.

Duda, R. O., Hart, P. E., & Stork, D. G. (2000). *Pattern classification* (2nd ed.). Wiley.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Science USA*, *95*(25), 14863–14868.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Frohlich, M., & Dixon, R. (2001). A taxonomy of manufacturing strategies revisited. *Journal of Operations Management*, *19*(5), 541–558.https://doi.org/10.1016/S0272-6963(01)00063-8

Goldberger, A. S. (1991). *A course in econometrics*. Harvard University Press.

Hair, J. F. (2009). *Multivariate data analysis* (7th ed.). Prentice Hall.

Halford, S. (2005). Hybrid workspace: Re-spatialisations of work, organisation and management. *New Technology, Work and Employment*, *20*(1), 19–33. https://doi.org/10.1111/j.1468-005X.2005.00141.x

Hirsch, P. B. (2021). The great discontent. *Journal of Business Strategy*, *42*(6), 439–442. https://doi.org/10.1108/JBS-08-2021-0141

Hislop, D., & Axtell, C. (2007). The neglect of spatial mobility in contemporary studies of work: The case of telework. *New Technology, Work and Employment*, *22*(1), 34–51. https://doi.org/10.1111/j.1468-005X.2007.00182.x

Hong, T., Chen, C., Wang, Z., & Xu, X. (2020). Linking human-building interactions in shared offices with personality traits. *Building Environment*, *170*, 106602. https://doi.org/10.1016/j.buildenv.2019.106602

Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis*. Prentice-Hall.

Ketchen, D., & Shook, C. (1996). The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, *17*(6), 441–459. https://www.jstor.org/stable/2486927

Kim, D. H., & Bluyssen, P. M. (2020). Clustering of office workers from the OFFICAIR study in he Netherlands based on their self-reported health and comfort. *Building and Environment*, *176*, 106860. https://doi.org/10.1016/j.buildenv.2020.106860

Ksinan Jiskrova, G. (2022). Impact of COVID-19 pandemic on the workforce: From psychological distress to the Great Resignation. *Journal of Epidemiol Community Health*, *76*(6), 525–526. https://doi.org/10.1136/jech-2021-218561

Mallett, O., Marks, A., & Skountridaki, L. (2020). Where does work belong anymore? The implications of intensive homebased working. *Gender in Management*, *35*(7/8), 657–665. https://doi.org/10.1108/GM-06-2020-0173

Manganelli, L., Thibault-Landry, A., Forest, J., & Carpentier, J. (2018). Self-determination theory can help you generate performance and well-being in the workplace: A review of the literature. *Advances in Developing Human Resources*, *20*(2), 227–240. https://doi.org/10.1177/1523422318757210

Miller, J. G., & Roth, A. (1994). A taxonomy of manufacturing strategies. *Management Science*, *40*(3), 285–304. http://www.jstor.org/stable/2632800

Morning Consult (2022). New Workers, New Normal. Analyzing the Workforce Transformation Amid COVID-19. *Morning Consult*. https://morningconsult.com/return-to-work/

Namey, E., Guest, G., Thairu, L. N., & Johnson, L. (2007). Data reduction techniques for large qualitative data sets. In G. Guest & K. M. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 137–162). Rowman & Littlefield.

Nicholas, K., & Hull, D. (2021). Elon Musk's ultimatum to Tesla execs: Return to the office or get out. *Bloomberg*. https://www.bloomberg.com/news/articles/2022-06-01/musk-s-tesla-ultimatum-return-to-office-or-work-somewhere-else

Ortiz, M. A., & Bluyssen, P. M. (2022). Profiling office workers based on their self-reported preferences of indoor environmental quality and psychosocial comfort at their workplace during COVID-19. *Building and Environment*, *211*, 108742. https://doi.org/10.1016/j.buildenv.2021.108742

Pacáková, Z., & Poláčková, J. (2013). Hierarchical cluster analysis – Various approaches to data preparation. *AGRIS on-Line Papers in Economics and Informatics*, *5*(3), 1–11.

Richardson, J., & McKenna, S. (2013). Reordering spatial and social relations: A case study of professional and managerial flexworkers. *British Journal of Management*, *25*(4), 724–736. https://doi.org/10.1111/1467-8551.12017

Soriano, A., Kozusznik, M. W., Peiró, J. M., & Mateo, C. (2020). The role of employees' work patterns and office type fit (and misfit) in the relationships between employee well-being and performance. *Environment and Behavior*, *52*(2), 111–138. https://doi.org/10.1177/0013916518794260

Tsiptis, K., & Chorianopoulos, A. (2009). *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

Wang, B. (2021). Evaluation method of the excellent employee based on clustering algorithm. In H. N. Dai, X. Liu, D. X. Luo, J. Xiao, & X. Chen (Eds.), *Blockchain and trustworthy systems* (pp 593–600). Springer.

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845