AVAILABILITY OF SERVICES IN SOCIO-ECONOMIC CLUSTERS OF LATVIAN ADMINISTRATIVE UNITS

***Elita Benga[1], Juris Hazners[1], Anita Auzina[2]**
[1]Institute of Agricultural Resources and Economics, Latvia, Latvia
[2]Latvia University of Life Sciences and Technologies, Latvia
*Corresponding author's email: elita.benga@arei.lv

Abstract

The concepts of clusters and cluster analysis often are mixed up leading to confusion and certain misunderstandings. Cluster analysis encompasses unsupervised machine learning methods with datasets that are not predetermined. In turn, clusters are understood as regional groups of economic actors in a close proximity that have reached certain levels of specific resources, expertise and skills to stand out among the broader economy. Often the analysis on these clusters also is denominated as cluster analysis. The objective of the study is to assess the differences in availability of a number of services in clusters of administrative units at the parish level. K-means clustering method is selected for the forming of clusters based on their similarity. Five indicators are selected for the forming the clusters: Gross Value Added, Unemployment level, Number of Workplaces, Gross Salary and number of persons with the Higher Education. Eight clusters were identified. For the further analysis the following Service availability indicators were selected: Number of Pharmacies, General Practitioners, Automated teller machines, Gas stations, Postal offices, Food retail outlets, share of Kindergarten Enrolment and share of School Enrolment. Results revealed significant differences in cluster parameter indicators, though certain clusters exhibited similarities. A similar pattern was observed in service availability indicators. However, the relationship between cluster rankings for parameter and service availability indicators was inconclusive. The study is limited by data availability at the parish level, and a broader set of indicators would allow for greater variability in cluster parameters and variables.

Keywords: clusters, K-means clustering, services availability, parish.

Introduction

Regional development analysis involves examining the processes that affect the quality of life and economic well-being of populations. In the context of Latvia, parishes serve as the smallest administrative units but are not at all used as focal points for cluster-based analyses in regional development studies. Social scientists and policy makers over the last three decades have increasingly focused their attention on clusters often observing the concentration of production and innovation in specific geographic areas (Orsenigo, 2006). European Cluster Collaboration Platform definition states that clusters should be considered as regional ecosystems of related industries and competences featuring a broad array of inter-industry interdependencies. They are defined as groups of firms, related economic actors, and institutions that are located near each other and have reached a sufficient scale to develop specialised expertise, services, resources, suppliers and skills (European Cluster Collaboration Platform, 2025). Cluster analysis or clustering is a data analysis method that divides a set of objects into groups where objects in the same group are more similar to each other than to those in other groups. The number of clusters and their structure is not predefined. Time and again, these two concepts are misunderstood and improperly attributed. The analysis of the predefined clusters mistakenly is often called cluster analysis. Research in Latvia has so far mostly focused either on the identification of business clusters (Garanti & Zvirbule-Berzina, 2013; Liela et al., 2010) or on the evaluation of cluster policy (Kulakova & Volkova, 2013). The application of clustering the administrative units is less common. Analysing service access in socio-economic clustering at the parish level offers valuable insights into regional disparities. Understanding these spatial patterns is essential for applying cluster-based methods to assess territorial development, particularly in identifying areas of insufficient access to public services and economic opportunities.

Materials and Methods

Clustering algorithms are a type of unsupervised machine learning used to group data based on similarities. K-Means clustering is a method that partitions datasets into clusters. In K-Means Clustering clusters are distinct with their numbers assigned according to the user pre-selected total number of clusters. K-Means groups data by minimizing intra-cluster variation to create compact, distinct clusters. It requires specifying number of k clusters upfront. K-Means Clustering is based on the iterative relocation of data points between clusters. It is used to divide the units of a dataset into non-overlapping groups, or clusters, based on the selected unit characteristics, usually a number of variables. The expected result is a pre-determined number of distinct groups of panel data units with a high degree of similarity within each group and a low degree of similarity between groups. K-Means Clustering uses the Lloyd algorithm centroid model (Lloyd, 1982). A centroid is expressed as a mean of each pre-selected variable that characterizes the unit. First, the k initial centroids are chosen based on an arbitrarily selected single variable, either randomly or by dividing the range of specified variable into intervals. After that, iterations are done in two steps. In the first one, each case of the data set is assigned to a cluster based on its distance from the cluster's centroids, using one of the distance metrics. Euclidian distance is usually used:

$$d(P,Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

where $P = (p_1, p_2, \dots, p_n)$ - vector P;
$Q = (q_1, q_2, \dots, q)$ - vector Q ; n - number of data points;
$d(P,Q)$ - Euclidian distance between the vectors P and Q.
In the second step, the previous value of centroid (usually mean) is replaced by the mean of all cases assigned to the centroid in the first step. Then the sum of squared error (SSE) is calculated as a sum of squared minimum distances between the cases. These two step iterations are repeated either until the SSE does not become lower or assigned cluster numbers for each unit do not change.
The Caliński-Harabasz index (Caliński & Harabasz, 1974) is a metric for the evaluation of the remoteness between cluster centroids. The assessment is based solely on the dataset and the clustering results. CHI index is defined as the ratio of the between-cluster separation (BCSS) to the within-cluster dispersion (WCSS), normalized by their number of degrees of freedom:

$$CH = \frac{BCSS/(k-1)}{WCSS/(n-k)} \quad (2)$$

where $BCSS$ - weighted sum of squared Euclidean distances between each cluster centroid; $WCSS$ - sum of squared Euclidean distances between the data points and cluster centroid; k - number of clusters; n - number of data points; CH - Caliński-Harabasz index.
BCSS (Within-Cluster Sum of Squares) is the sum of squared Euclidean distances between the data points and their respective cluster centroids:

$$BCSS = \sum_{i=1}^{k} n_i \|c_i - c\|^2 \quad (3)$$

where n_i - number of points in cluster C_i, c_i - centroid of C_i, c_i - overall centroid of the data; $BCSS$ - weighted sum of squared Euclidean distances between each cluster centroid.
BCSS measures how well the clusters are separated from each other (the larger the better). WCSS (Within-Cluster Sum of Squares) is the sum of squared Euclidean distances between the data points and their

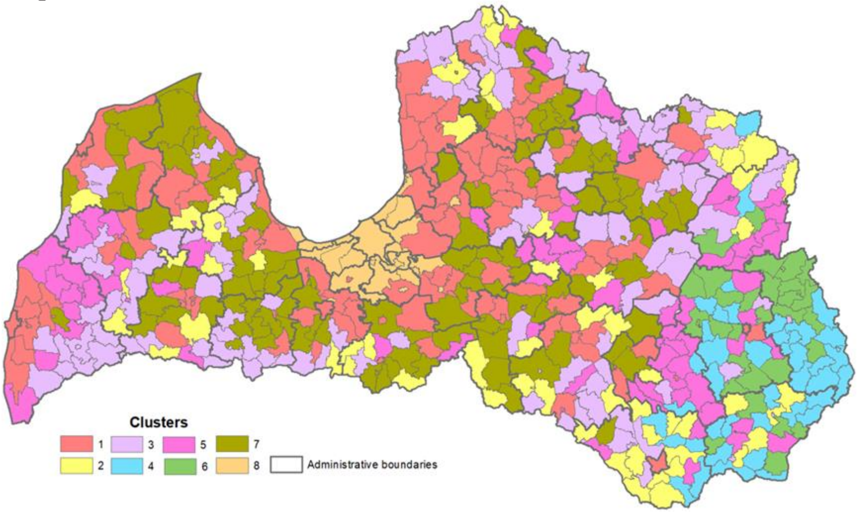
$$WCSS = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - c_i\|^2 \quad (4)$$

where k - number of clusters; c_i - centroid of C_i; χ - data point; $WCSS$ - within cluster sum of squares.
WCSS measures the compactness of the clusters (the smaller the better). The objective of the centroid-based clustering is the simultaneous maximizing the BCSS and minimizing the WCSS.
Finding the optimal number of clusters is usually a two-step procedure. The Elbow method, proposed by Thorndike (Thorndike, 1953) is a heuristic graphical method for finding the optimal number of clusters in a k-means clustering algorithm. The WCSS values are plotted on the y-axis with the corresponding k values plotted on the x-axis. The optimal number of clusters is determined by a forming of an elbow in the graph. The optimal number of clusters suggested by an Elbow method could be retained only if the corresponding CHI index has the highest value.
Data for creating the clusters from 591 administrative units for 2022 are obtained from Central Statistical Bureau (CSB, 2025). Five indicators are selected as cluster variables: Per capita Gross Value Added, Gross Salary, share of persons with Higher Education in total population, per capita Workplaces and Unemployment Rate. Data on variables for service availability for 2022 are obtained from Central Statistical Bureau (CSB, 2025), Central Bank (Latvijas Banka, 2025), National System of Educational Information (Valsts izglītības informācijas sistēma, 2025), National Health Office (Nacionālais Veselības dienests, 2025), Latvian Post (VAS Latvijas Pasts, 2025), SIA Kalifeks (2025), SIA Venipak Latvija (2025), SIA DPD Latvija (2025), SIA ELVI Latvija (2025), SIA Iepirkumu grupa (2025), SIA Lats veikali (2025), SIA Latvijas Neatkarīgo Tirgotāju Kooperācija (2025), SIA Lidl Latvija (2025), SIA Maxima Latvija (2025), SIA Mego (2025), SIA RIMI Latvia (2025), SIA SPAR Latvia (2025), uzziņu dienests 1188 (2025), SIA Vesko (2025). Nine indicators are selected as service availability variables: Parcel machines, Pharmacies, General Practitioners, Automated teller machines, Gas stations, Postal offices, Food Retail outlets on 10,000 inhabitants; Share of kindergarten enrolment and Share of school enrolment within the respective population group.

Results and Discussion

K-means clustering with Euclidean distance is performed with pre-selected user defined cluster numbers ranging from two to ten. As the cases with the cluster numbers above eight yielded no results (only eight clusters had at least one unit) and the plot at the eight-cluster case shows the slight elbow, the optimal number of the clusters is eight. The corresponding CH index value also is the highest. Hence, the proceeding with the eight clusters is justified. After the seven iterations, eight clusters were established. The clusters are plotted on the map 'Figure 1'.
As seen from the map, units belonging to the same cluster do not necessarily be geographically closely located. Even within one district parishes can be different clusters. Mapping of the clusters points towards a pronounced monocentricity with a well performing area around the Capital City (Cluster 8).
There are districts divided into two clusters indicating on relatively similar situation. At the same time, there are districts with parishes attributed to five or more clusters indicating on uneven development of the administrative units.

Figure 1
Latvian map with eight clusters



It is important to note that while some clusters (1, 2, and 5) are distributed across the entire country, two clusters (4 and 6) are exclusively observed in the eastern part of country which is Latgale region. This suggests that certain territories exhibit regional variations and, based on the selected indicators, reflect differing levels of development. Arbitrary cluster titles are assigned to each cluster. The average values of the cluster variables for eight clusters are shown in 'Figure 2'.

Figure 2
Eight clusters and parameter averages

Clusters	Unemployment rate	Gross Value Added per capita	Share of workplaces in total population	Average Gross Salary	Share of persons with higher education (age 19 + years)
8 - Metropolitan area	5.2	14,495.4	60.0	1,491.7	41.0
1 - State cities and well performing counties	7.6	9,095.9	53.4	1,150.7	27.4
7 - Affluent counties	8.1	9,423.7	50.0	1,080.3	18.9
5 - Average counties	10.9	6,185.0	49.3	1,002.7	20.2
3 - Average performing counties	10.1	6,223.0	47.4	985.7	18.8
2 - Below the average perfoming counties	11.6	5,718.7	44.1	955.2	16.7
6 - Poor counties	21.8	4,601.6	44.8	922.8	19.5
4 - Extremely poor counties	24.5	3,933.1	38.7	828.3	13.1

The cluster parameter average values for each cluster are paired and coloured in four hues from light green to deep red according to the position of the value within the respective range. Clusters 1 and 8 have only green (very light to dark green) cells indicating on good socio-economic performance. The highest Gross Value Added per capita (GVA per capita), share of persons with Higher Education, Average Gross Salary and lowest Unemployment rate are recorded in Metropolitan area (Cluster 8). On the other hand, Cluster 4 has the lowest average values for every parameter variable. Cluster 1 comprises six State Cities, 29 townships and 89 rural parishes, wherein all five parameters exhibit considerable similarity to the highest - performing Cluster 8, except for GVA per capita, which is 1.5 times lower, and the proportion of persons with higher education, which also is 1.5 times lower. In Cluster 7, indicator values are relatively similar to those in Cluster 8 and Cluster 1, differing slightly in two parameters: the share of persons with Higher Education and the share of Workplaces in the total population. However, it exhibits a higher Average Gross Salary compared to Cluster 1. Clusters 4 and 8 show significant disparities in unemployment rates, with differences over fourfold, and GVA per capita differs more than three times. However, the Average Gross Salary is about half in Cluster 8 compared to Cluster 4. Clusters characterized by low Average Gross Salaries also exhibit significantly high unemployment rates. Clusters 5 and 3 are relatively similar, displaying moderate values across all indicators. Notably, Cluster 5 outperforms Cluster 7 in the share of persons with Higher Education. Cluster 2 exhibits relatively poor performance in the share of persons with Higher Education indicator, both in absolute terms and relative to the average values recorded in Clusters 3 and 5. If compared to Cluster 8, the indicators for Unemployment rate, GVA per capita, and share of persons with Higher Education are 2.2 to 2.5 times worse. Clusters 4 and 6 demonstrate the weakest performance across all indicators, particularly in the Unemployment Rate and GVA per capita, where their values are 4.4 and 3.6 times worse, respectively, if compared to Cluster 8. Moreover, Cluster 4 exhibits the most unfavourable situation as it has the highest Unemployment rate among all clusters, and for all other indicators, it stands only at 37% to 90% of Cluster 6 levels.

The calculated values for the service availability variables for eight clusters are shown in 'Figure 3'.

Figure 3

Service availability indicator average values in eight clusters

	Clusters							
	1	2	3	4	5	6	7	8
Number of Units	124	62	115	41	76	32	114	27
	Per 10,000 inhabitants							
Parcel machines	8.0	1.4	6.3	0.8	5.8	4.3	5.6	7.8
Pharmacies	4.6	2.0	4.4	1.2	4.3	3.0	3.6	4.1
General Practitioners	7.5	4.3	6.2	2.9	6.7	3.4	5.4	7.8
Automated teller machines	4.5	0.5	3.2	0.4	3.1	2.2	2.2	3.7
Gas stations	4.2	1.8	2.9	1.2	4.4	1.7	3.5	2.9
Post offices	5.9	15.1	14.5	18.0	15.9	17.2	13.4	3.9
Food retail outlets	9.4	9.5	9.7	3.3	9.6	6.5	11.6	5.8
	Per population group							
Share of kindergarten enrollment	39.5	7.5	25.0	8.4	25.4	33.4	21.4	38.7
Share of school enrollment	29.2	13.5	25.3	10.4	22.3	22.0	31.0	63.5

The disparities are marked in service availability among clusters. Clusters 1 and 8 have a good availability in a majority of the services. The availability of postal services seemingly is negatively correlated with the availability of Teller Machines (ATM) as we can see in Cluster 4. The situation in Clusters 1 and 8 is similar. However, Cluster 1 records more favourable situation across all services except Postal services. This stands in steep contrast to Cluster 4, where postal service accessibility is the highest, yet all other indicators are among the weakest. This suggests that Post offices could serve as key locations for providing essential services otherwise unavailable in these areas, such as ATMs, parcel lockers, and goods collection points. The highest workload on General Practitioners is observed in Clusters 1 and 8, with workload indices of 7.5 and 7.8, respectively, whereas the lowest workload is recorded in Cluster 4 (0.4). The availability of Pharmacies and General Practitioners is most limited in Clusters 4, 2 and 6, which also rank lowest in terms of overall situation. The overall assessment based on the selected essential services highlights Clusters 1 and 8 as having the most favourable conditions, aligning with their strong performance across other development criteria. Clusters 3 and 5 can be classified as having a moderately favourable situation, as service availability is relatively balanced across all service types. In contrast, Clusters 6 and 7, despite being in different stages of development, exhibit lower service availability indicators, particularly in access to Pharmacies and ATMs, compared to the mid-level clusters (Clusters 3 and 5). Similarly, in evaluating kindergarten and school attendance, Clusters 2 and 4 exhibit the most unfavourable conditions, with attendance rates up to five times lower than those observed in Clusters 1 and 8.

The data suggests that clusters provide a nuanced framework for understanding the distribution of key services across regions, identifying areas of both strength and vulnerability. Post offices, for instance, when considered as part of a cluster analysis, reveal their potential to function not just as logistical nodes but as multifaceted service points, offering insights into how infrastructure can be optimised for broader regional needs.

The disparities in service availability across clusters emphasise how the use of clusters can effectively pinpoint specific regional deficiencies, particularly in sectors like healthcare and education. For example, Clusters 2, 4, and 6, which display significant service gaps, can be seen as regions requiring more targeted policy intervention and investment. Conversely, Clusters 3 and 5 offer valuable data on regions with relatively balanced service distribution, which can serve as models for incremental improvements in less developed areas.

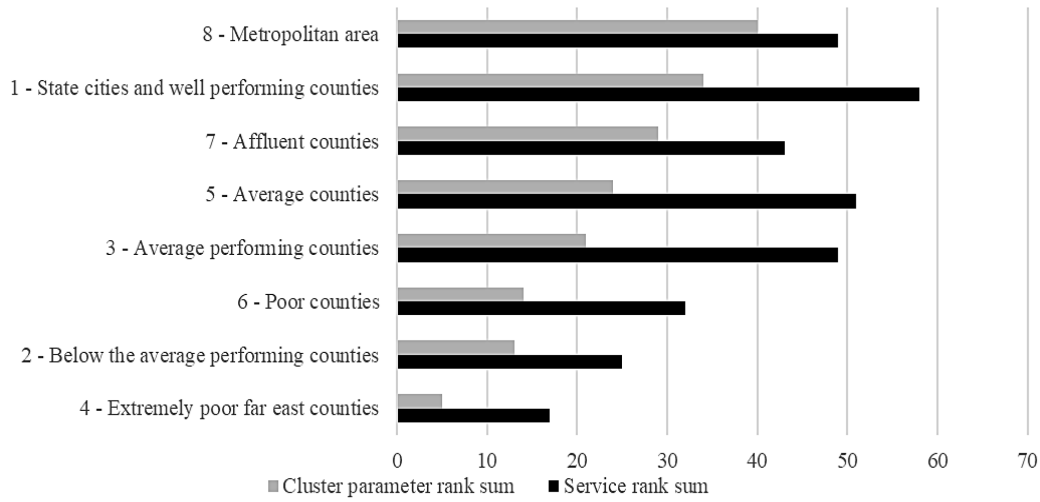
Moreover, the consistently low service indicators in Clusters 6 and 7, irrespective of their broader development contexts, illustrate how cluster analysis can uncover underlying systemic issues that might otherwise remain undetected in broader regional assessments. These findings underscore the utility of clusters in dissecting regional heterogeneity and directing targeted interventions.

Ultimately, clusters, as a methodological tool, offer a sophisticated approach to analysing regional development, enabling policymakers to identify spatial inequalities and allocate resources more effectively based on empirical, data-driven insights.

Two ranking parameters can be established. First, the clusters are assigned ranks for each of the parameters according to the respective average value. Second, the clusters are assigned ranks for each of the service availability variable. Higher rank assumes better performance. Third, the ranks are summed to get the total rank. Total ranks for parameters and service availability variables are shown in 'Figure 4'.

Figure 4

Service availability indicator average values in eight clusters



As seen from the chart, the relationship between the parameter ranks and service availability ranks is not unambiguous. Affluent parishes in Cluster 7 perform lower than might be expected.

Differences can be observed between the ranking of clusters based on their parameters and their ranking in terms of service availability. While the cluster ranking positions Clusters 8, 1, 7, and 5 at the top, the service availability ranking places Cluster 1 first, followed by Clusters 5, 3, and only then Cluster 8. However, the bottom three clusters remain unchanged in both rankings, with Clusters 6, 2, and 4 consistently occupying the lowest positions. This indicates that these clusters (6, 2 and 4) are in the most challenging situations, both from an economic perspective and in terms of service accessibility.

The clustering of administrative units is not common in the EU countries. In the UK, Office for National Statistics in the latest Census in 2021 include experimental statistics with clustering local authorities against subnational indicators in England (Office for National Statistics, 2025). Clusters are defined as similar local authorities. Using a k-means clustering method, seven different models on different themes using data available are created allowing users to understand the similarities between local authorities, while providing local authorities with control groups for investigating the impact of policy interventions. The cluster forming variables were based on themes such as urban vs. rural areas, health, well-being, connectivity, educational performance, population density. After establishing the clusters, headline metrics for comparisons per theme were selected, such as gross value added per hour worked; average travel time to nearest employment centre by public transport or walking; gigabit capable broadband; pupils at expected standards by the end of primary school, apprenticeships achievements per capita; healthy life expectancy; life satisfaction. Analysis of selected headline metrics across seven different themes produced four clusters of local authorities in England: Higher health and well-being, lower connectivity (52 Local Authorities); Higher connectivity, lower health and well-being (78 Local Authorities); Higher health and well-being, moderate educational performance (114 Local Authorities); Higher health and productivity, lower well-being (56 Local Authorities). Similarly to Latvia, the number of Local Authorities differs significantly between the clusters. To present the findings of the analysis, an interactive map has been created. Results for each cluster are presented in this tool across all seven themes. Local authorities of interest can be selected from the drop-down menu, and text summarising the clusters is shown.

Conclusions

Clusters and clustering are distinct yet often misinterpreted concepts that need to be clearly delineated in both research and policy contexts.

1. This study has demonstrated that clustering administrative units-particularly rural parishes-can uncover significant patterns in socio-economic situation and service provision. The analysis revealed notable disparities in parish development outcomes, where some clusters exhibit considerable deficits in both economic and service-related indicators.

2. Identifying these spatial disparities is essential for applying cluster-based analyses to enable justified territories-specific interventions.
3. Future research can be focused on clusters created based upon the themes other than socio-economic performance, such as education, health, connectivity, service availability with the analysis of the performance in other themes.
4. The development of interactive tools for visualising regional disparities at the parish level can facilitate data-driven approaches to policy formulation, enhancing the effectiveness of regional development strategies.

References

Caliński, T. & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics, 3*(1), 1-27. https://doi.org/10.1080/03610927408827101

Central Statistical Bureau. (CSB). (2025, January 8). *Oficiāls statistikas portals datubāze* [Official Statistics Portal Database]. https://data.stat.gov.lv/pxweb/lv/OSP_PUB/START__POP__IR__IRS

European Cluster Collaboration Platform. (2025). *The European online hub for industry clusters*. https://www.clustercollaboration.eu/

Garanti, Z. & Zvirbule-Berzina, A. (2013). In search for regional clusters in Latvia. *Journal of Business Management*, 7, 93-104. https://www.ba.lv/wp-content/uploads/2015/08/no7jbm_2013_ready.pdf

Kulakova, N. & Volkova, T. (2013). Evaluation of cluster policy in Latvia. *Journal of Business Management,* 7, 83-92. https://www.ba.lv/wp-content/uploads/2015/08/no7jbm_2013_ready.pdf

Liela, E., Zeibote, Z., & Stale, L. (2010). Business clusters for improving competitiveness and innovation of enterprises-experience of Latvia. *Journal of Business Management,* 3.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28, 129-137. https://www.stat.cmu.edu/~rnugent/PCMI2016/papers/LloydKMeans.pdf

Nacionālais Veselības dienests. (2025, January 8). *Ģimenes ārsti atbilstoši teritorijām* [Family doctors by territory]. https://www.vmnvd.gov.lv/lv/gimenes-arsti-atbilstosi-teritorijam?utm_source=https%3A%2F%2Fwww. Google.com%2F

Office for National Statistics. (2025, March 9). *Data and analysis from Census 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/clusteringlocalauthoritiesagainstsubnationalindicatorsengland/2023-02-24

Orsenigo, L. (2006). Clusters and clustering: Stylized facts, issues, and theories. In M. Feldman, P. Braunerhjelm (Eds.) *Cluster genesis.* University of Brescia and Cespri, Bocconi University, Italy. https://doi.org/10.1093/acprof:oso/9780199207183.003.0010

SIA DPD Latvija. (2025, January 8). *Saņemšana* [Receiving]. https://www.dpd.com/lv/lv/sanemsana/pickup-tikls/pickup-punktu-saraksts/

SIA ELVI Latvija. (2025). *Veikali* [Shops]. https://elvi.lv/elvi-veikali/

SIA Iepirkumu grupa. (2025). *Veikali* [Shops]. https://etop.lv/lv/veikali

SIA Kalifeks. (2025, January 8). *Piegādes noteikumi* [Delivery terms]. https://kalifeks24.lv/omniva-pakomatu-adreses

SIA Lats veikali. (2025). *Veikali* [Shops]. https://latts.lv/lats-veikali

SIA Latvijas Neatkarīgo Tirgotāju Kooperācija. (2025). *Veikali* [Shops]. https://aibe.lv/veikali/

SIA Lidl Latvija. (2025). *Klientu atbalsts* [Customer Support]. https://klientu-serviss.lidl.lv/SelfServiceLV/s/article/Kur%C4%81s-pils%C4%93t%C4%81s-7-oktobr%C4%AB-tiks-atv%C4%93rti-Lidl-veikali

SIA Maxima Latvija. (2025). *Veikalu ķēde* [Chain of stores]. https://www.maxima.lv/veikalu-kedes

SIA Mego. (2025). *Kontakti* [Contacts]. https://mego.lv/kontakti

SIA RIMI Latvia. (2025). *RIMI veikali* [RIMI Shops]. https://www.rimi.lv/veikali

SIA SPAR Latvija. (2025). *Veikali* [Shops]. https://spar.lv/veikali/

SIA Venipak Latvija. (2025, January 8). *Pakomāti* [Parcel machines]. https://venipak.com/lv/pakomati/

SIA Vesko. (2025). *Kontakti* [Contacts]. https://veskoveikals.lv/

Thorndike, R. L. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267-276.

Uzziņu dienests 1188. (2025, January 8). *Degvielas uzpildes stacijas* [Filling stations]. https://www.1188.lv/katalogs/degvielas-uzpildes-stacijas-1540

Valsts izglītības informācijas sistēma. (2025, January 8). *Vispārējās izglītības iestāžu un izglītojamo skaita aktuālie dati* [National System of Educational Information. Current data on the number of general education institutions and students]. https://www.viis.gov.lv/dati/visparejas-izglitibas-iestazu-un-izglitojamo-skaita-aktualie-dati

VAS Latvijas Pasts. (2025, January 8). *Privātpersonām* [For individuals]. https://pasts.lv/lv/privatpersonam/nosutit/pakomati/#pakomatu-adreses-laiki (in Latvian).