Introduction

The importance of leasing and leasing loans[1] as easily available source of financing for corporate investments has increased significantly in the last decade. According to the data of the Polish Leasing Association[2], in 2010 leasing companies financed assets worth PLN 27.3 billion, while in 2021 this amount reached PLN 88.0 billion, which means more than a threefold increase. The share of financing provided by enterprises from the leasing industry in capital expenditure more than doubled in 2010-2021: from 9.3% to 19.7%. During this period, the share of financing by leasing and leasing loans in the total GDP of Poland also increased: from 1.9% to 3.4%. The status of leasing is also confirmed by a survey on access to financing for enterprises conducted by the European Central Bank in 2020[3]. Polish companies participating in the survey most often (in 63% of cases) indicated leasing as an important form of financing.

The Polish leasing market is also important compared to Europe. According to the Leaseurope[4] study, the value of financing by leasing and leasing loans in 2021 placed Poland in fifth place in Europe[5]. Polish enterprises accounted for 6.6% of the European leasing market. In turn, according to the PLA, the share of the leasing industry in total GDP in 2018 was higher in Poland (3.9%) than in countries such as the United Kingdom (3.4%), France (2.3%) or Germany (1.4%), whose leasing markets are in the top three of Leaseurope ranking mentioned above.

The Polish Leasing Association publishes incremental quarterly data based on members' reports. These statements include the value of assets financed by leasing and leasing loans broken down into vehicles, machines, IT equipment, other means of transports, other objects and real estate. For the purposes of this analysis, the data were converted to present

[1] A leasing is an agreement under which one party (lessor) transfers the right to use the leased asset to the other party (lessee) in exchange for fixed installment payments. A leasing loan is a service similar to a bank loan, with the difference that the financing entity is a leasing company (instead of a bank) and the loan is granted for purchase of a specific asset.

[2] The Polish Leasing Association (PLA) is an organization representing the leasing industry in Poland. The PLA affiliated 30 leasing companies and the Polish Vehicle Rental and Leasing Association in 2022.

[3] Survey on the access to finance of enterprises 2020, European Commission

[4] Leaseurope is an organization representing the leasing and car rental industry in Europe. In 2020, it had 46 members from 32 European countries, including the Polish Leasing Association.

[5] Annual Statistical Enquiry 2021, Leaseurope

the value of financing in a given quarter. Incorrect values and missing data were corrected using an expert method.

The aim of this analysis was to examine whether there were such close similarities between the companies affiliated in the Polish Leasing Association in the years 2003 – 2022 that it was possible to group these companies on the basis of the assets they financed.

This analysis was performed using various time series clustering methods. The purpose of cluster analysis (clustering) is to assign objects to groups (clusters) in such a way that the objects in a given group are as similar to each other as possible and, at the same time, the least similar to objects from other clusters. In other words, the goal is to divide the objects into homogeneous and well-separated groups. The analysis used the *R* environment, which is free software for statistical calculations. The analysis was performed using functions from `dtwclust` package, which provides a wide range of data clustering methods.

The analysis does not include enterprises that reported in less than 20% of the analyzed quarters (from Q1 2003 to Q1 2022). The final dataset consists of 46 companies (4). The input data has been standardized with the `dtwclust::zscore` function and is used as such throughout the analysis. Similarly, all analyzes use a slanted band window of size 5, which is 6.5% of the length of the series under study[6].

In the further part of the analysis, the names of measures and methods from the `dtwclust` package will be used. Their comparison with the full names is presented in tables 1, 2 and 3.

[6] As a rule of thumb, constraint window should have size of 5-10% of the series' length

The financed assets are divided into the following categories:

- Vehicles
  - Passenger vehicles
  - Commercial vehicles weighing up to 3.5 tonnes
  - Typical trucks weighing more than 3.5 tonnes
  - Truck tractors
  - Semitrailers and trailers
  - Buses
  - Other vehicles
- Machines
  - Construction equipment
  - Agricultural machines
  - Printing machines
  - Processing machines
  - Food machines
  - Medical equipment
  - Catering equipment
  - Forklifts
  - Other machines
- IT
  - Hardware
  - Software
  - Other IT equipment
- Other means of transport
  - Air transport
  - Water transport
  - Rail transport
- Other items
- Real estates
  - Industrial buildings
  - Commercial and service facilities
  - Office buildings
  - Hotels and recreational facilities
  - Other real estates

Table 1: Distance measures

Full measure name	**Name in `dtwclust` package**
Dynamic Time Warping distance	dtw
DTW distance with L2 norm	dtw2
Basic DTW distance	dtw_basic
Soft-DTW distance	sdtw
Keogh's DTW lower bound	lbk
Lemire's improved DTW lower bound	lbi
Shape-based distance	sbd
Fast global alignment kernels	gak

Table 2: Centroids calculation methods

Full method name	**Name in `dtwclust` package**
DTW Barycenter Averaging	dba
Soft-DTW centroids	sdtw_cent
Shape average of several time series	shape_extraction
Partition around medoids	pam
The average along each dimension	mean
The median along each dimension	median
Fuzzy c-means	fcm
Fuzzy c-medoids	fcmdd

Table 3: Agglomeration methods

Full method name	**Name in `dtwclust` package**
Ward's method	ward.D / ward.D2
Single linkage method (nearest neighbour method)	single
Complete linkage method	complete
Group average linkage method	average
Weighted average linkage method	mcquitty
Centroid linkage method	centroid
Weighted centroid linkage method	median

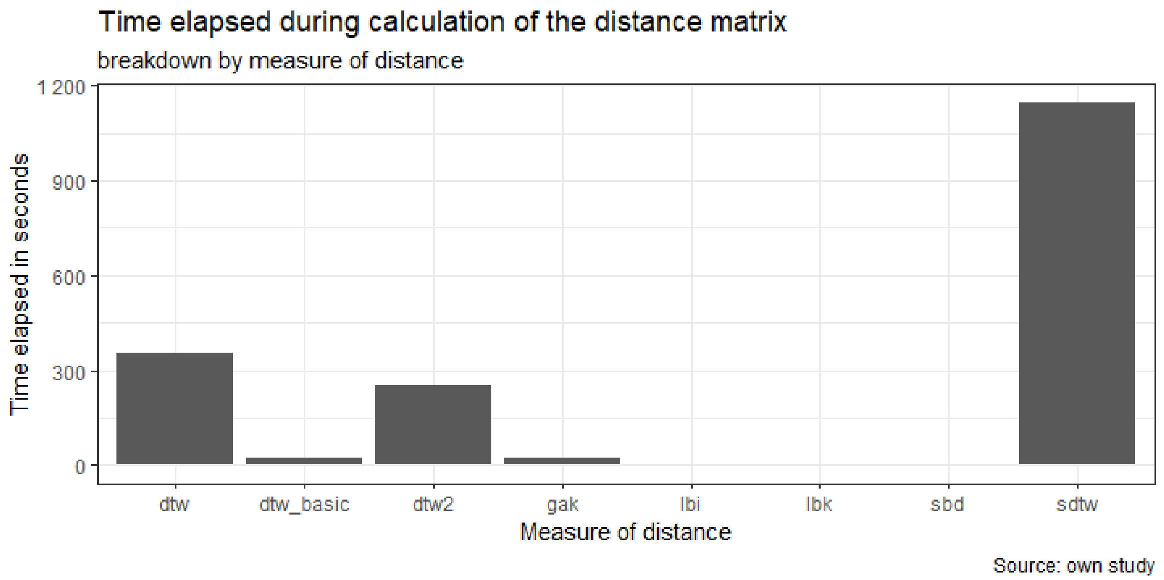
Table 4: Analysed leasing companies

Full company name	**Company code**
AKF LEASING POLSKA S.A.	AKF
ALIOR LEASING SP. Z O.O.	ALIOR
BGŻ LEASING SP. Z O.O.	BGZ
BMW FINANCIAL SERVICES POLSKA SP. Z O.O.	BMW
BNP PARIBAS LEASE GROUP SP. Z O.O.	BNP PARIBAS
BPH LEASING S.A.	BPH
CATERPILLAR FINANCIAL SERVICES POLAND SP. Z O.O.	CATERPILLAR
DE LAGE LANDEN LEASING POLSKA S.A.	DE LAGE LANDEN
DEUTSCHE LEASING POLSKA S.A.	DLP
DNB NORD LEASING SP. Z O.O.	DNB NORD
EUROPEJSKI FUNDUSZ LEASINGOWY S.A.	EFL
FORTIS LEASE POLSKA SP. Z O.O.	FORTIS
GETIN LEASING S.A.	GETIN
HANDLOWY-LEASING SP. Z O.O.	HL
IDEA LEASING S.A.	IDEA
IDEA GETIN LEASING S.A.	IDEA GETIN
IMPULS-LEASING POLSKA SP. Z O.O.	IMPULS
ING LEASE SP. Z O.O.	ING
KREDYT LEASE S.A.	KL
LEASYS POLSKA SP. Z O.O.	LEASYS
LEASING POLSKI SP. Z O.O.	LP
MAN FINANCIAL SERVICES POLAND SP. Z O.O.	MAN
Group MASTERLEASE	MASTERLEASE
MERCEDES-BENZ LEASING POLSKA SP. Z O.O.	MERCEDES-BENZ
MILLENNIUM LEASING SP. Z O.O.	MILLENNIUM
MLEASING SP. Z O.O.	MLEASING
NL-LEASING POLSKA SP. Z O.O.	NL
NOMA 2 SP. Z O.O.	NOMA
NORDEA FINANCE POLSKA S.A.	NORDEA
ORIX POLSKA S.A.	ORIX
PEAC POLAND SP. Z O.O.	PEAC
PEKAO SP. Z O.O.	PEKAO
PKO LEASING S.A.	PKO
RAIFFEISEN LEASING POLSKA S.A.	RAIFFEISEN
RCI LEASING POLSKA SP. Z O.O.	RENAULT
SANTANDER LEASING S.A.	SANTANDER
SANTANDER CONSUMER MULTIRENT SP. Z O.O.	SANTANDER CM
SCANIA FINANCE POLSKA SP. Z O.O.	SCANIA
SEB LEASING POLSKA SP. Z O.O.	SEB
SGB LEASING SP. Z O.O.	SGB
SG EQUIPMENT LEASING POLSKA SP. Z O.O.	SGE
SIEMENS FINANCE SP. Z O.O.	SIEMENS
VB LEASING POLSKA S.A.	VB
VOLKSWAGEN FINANCIAL SERVICES POLSKA SP. Z O.O.	VOLKSWAGEN
VOLVO FINANCIAL SERVICES USŁUGI FINANSOWE POLSKA SP. Z O.O.	VOLVO
WATIN LEASING & FINANCE S.A.	WATIN

Computational complexity

In order to select the distance measures for further analysis, the time of determining the distance matrix of the analyzed data using the `system.time` function was calculated. The results are presented below. `dtw`, `dtw2` and `sdtw` have quite high computational complexity compared to other measures, hence in further analysis they were replaced by `dtw_basic`.

Figure 1: Computational complexity of determining the distance matrix



A word on CVIs

`dtwclust` package provides a number of evaluation metrics called cluster validity indices. Internal CVIs measure the clustering purity, while external CVIs measure similarity of two clusterings. Both types of CVIs can be used for crisp and fuzzy partitions.

Table 5: Types of CVIs

	Indices	
Type	to be minimized	to be maximized
Internal crisp CVIs	Davies-Bouldin modified Davies-Bouldin COP	Silhouette Score Function Calinski-Harabasz Dunn
External crisp CVIs	Variation of Information	Rand Adjusted Rand Jaccard Fowlkes-Mallows
Internal fuzzy CVIs	Kwon Tang SC	Partition coefficient PBMF
External fuzzy CVIs	Soft Variation of Information	Soft Rand NMIM

Hierarchical clusterings

In hierarchical clustering five distance measures were used with their corresponding centroids calculation methods. The `gak` measure was an exception, as there is no method for assessing centroids derived from it. The `gak` was used two times, in combination with centroids obtained by the `dba` and `shape_extraction` methods. Each procedure has been calculated for eight different agglomeration methods available in `dtwclust` package (table 6). A total of 48 hierarchical clusterings were performed. In section below, these clusterings will be called in form distance measure – centroids calculation method – agglomeration method (e.g. `lbk_dba_complete`).

Table 6: Distance measures and corresponding centroids calculation methods used in hierarchical clustering

Distance measure	Centroid calculation method
dtw_basic	dba
lbk	dba
lbi	dba
sbd	shape_extraction
gak	dba
gak	shape_extraction

Internal CVIs and inertia

The internal CVIs (table 7) indicate four clusterings which can be considered as the best:

- `lbk_dba_average`
- `lbk_dba_median`
- `dtw_basic_dba_mcquitty`
- `dtw_basic_dba_average`

`lbk_dba` clusterings have the highest values of Silhouette index (0.4), as well as the lowest values of Davies-Bouldin indices (0.44) and respectably low COP index: 0.34 (minimum for all executed clusterings equals 0.26). However, their Score Function indices equal zero (maximum is 0.63), Calinski-Harabasz index is around 4 (maximum is 37.32) and Dunn index equals 0.49 (maximum is 0.68).

`dtw_basic_dba` clusterings have reasonable values of Silhouette index (0.32), low values of Davies-Bouldin indices (0.52) and the highest values of Dunn index (0.68). Similarly to lbk_dba clusterings, their Score Function indices equal zero. They have slightly higher values of Calinski-Harabasz index (12.84 for dtw_basic_dba_average, 10.83 for dtw_basic_dba_mcquitty) and COP index (0.52) than lbk_dba clusterings.

In order to determine the optimal number of clusters, the inertia measure was used:

$$inertia = 1 - \frac{\textit{whithin-cluster inertia}}{\textit{inter-cluster inertia}} \tag{1}$$

where values close to 1 indicate the best number of clusters.

From the clusterings mentioned above, lbk_dba_average has the highest value of inertia for number of clusters greater than four (table 7). It is also the first to exceed the value of inertia of 0.8 (for 11 clusters) and 0.9 (for 20 clusters). However, there are several lbk_dba clusterings that exceed the value of inertia of 0.8 for fewer clusters. lbk_dba_ward.D and lbk_dba_ward.D2 do that for 9 clusters, while lbk_dba_complete exceeds value of 0.8 for 7 clusters.

In all cases above, increasing the number of clusters over 10 does not increase the inertia significantly. Even lbk_dba_complete need 15 clusters to exceed value of inertia of 0.9. However, the greater number of clusters means that some companies may end up in single-element groups, which is an undesirable phenomenon (unless we are aiming to find outliers). For this reason, in the further part of the analysis, clusterings with up to 10 groups will be calculated.

The clusterings with increase of inertia greater or equal to lbk_dba_average have generally worse indices than four clusterings indicated by internal CVIs. They have relatively high values of Davies-Bouldin indices (2.39-2.75 versus 0.44 for lbk_dba_average) and low values of Dunn index (0.19-0.26 vs 0.5) and Silhouette index (0.2-0.3 vs 0.4). Nonetheless, they have higher values of Calinski-Harabasz index (8.18-8.94 vs 4.05) and lower values of COP index (0.26-0.3 vs 0.34).

External CVIs

In sections below, the similarity between lbk_dba_average and other clusterings mentioned above will be measured by external CVIs.

According to the Rand index, lbk_dba_mcquitty is the most similar clustering to lbk_dba_average (excluding clusterings with very low number of groups like 2 or 3 of course) – value of this index exceeds 0.9 for 6 clusters and reaches 1 for greater number of

Table 7: Internal CVIs for the best hierarchical clusterings

Best clustering by	Distance	Centroid	Linkage	Internal CVIs							Inertia > 0.8 for number of clusters
				Silhouette	Score Function	Calinski-Harabasz	Davies-Bouldin	Modified Davies-Bouldin	Dunn	COP	
Internal CVIs	dtw_basic	dba	mcquitty	0.32	0	10.83	0.52	0.52	0.68	0.52	16
	dtw_basic	dba	average	0.32	0	12.84	0.52	0.52	0.68	0.52	16
	lbk	dba	average	0.4	0	4.05	0.44	0.44	0.5	0.34	11
	lbk	dba	median	0.4	0	3.99	0.44	0.44	0.5	0.34	15
Inertia	lbk	dba	ward.D	0.22	0	8.66	2.42	2.42	0.24	0.27	9
	lbk	dba	ward.D2	0.22	0	8.91	2.39	2.39	0.24	0.27	9
	lbk	dba	mcquitty	0.2	0	8.94	2.75	2.75	0.19	0.3	11
	lbk	dba	complete	0.3	0	8.18	2.43	2.43	0.26	0.26	7

Figure 2: Inertia of `dtw_basic_dba` and `lbk_dba` clusterings

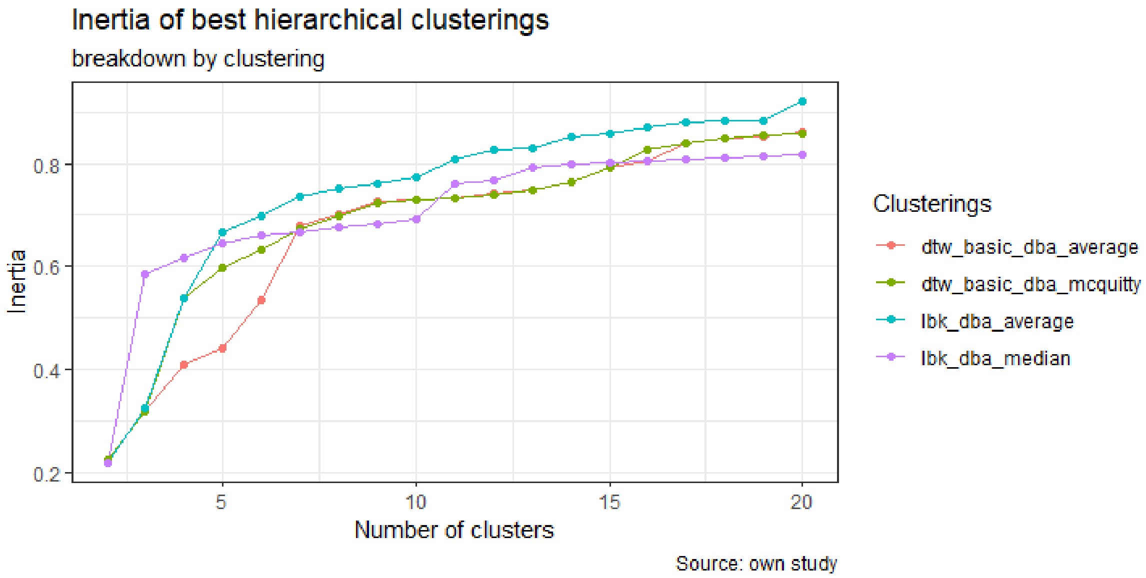
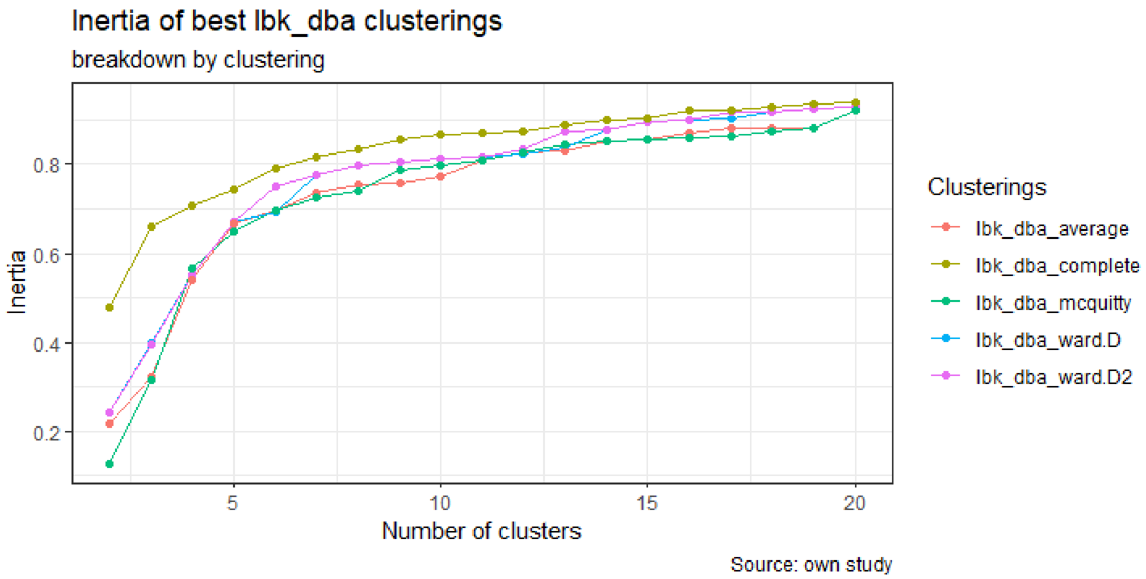


Figure 3: Inertia of `lbk_dba` clusterings



groups. lbk_dba_ward.D and `lbk_dba_ward.D2` have the value of Rand index higher than 0.9 for 10 clusters. Other clusterings are less similar to `lbk_dba_average`.

The adjusted Rand index, the Jaccard index and the Fowlkes-Mallows index indicate that only `lbk_dba_mcquitty` should be considered as similar to `lbk_dba_average`, as it is the only clustering which indices values exceed 0.9 (excluding clusterings with low number of groups again).

The Variation of Information index also shows that only `lbk_dba_mcquitty` could be considered as similar to lbk_dba_average. Given that internal CVIs of this clustering are generally worse than those of lbk_dba_average and that lbk_dba_mcquitty does not have greater inertia for lower number of clusters than its comparison, there is no reason two consider this clustering as substitute for lbk_dba_average.

Summary of **lbk_dba_average** clustering

The lbk_dba_average clustering with 11 groups is presented below. The first cluster contains only one company (SGE), which reported in all analyzed quarters and focused mostly on machines (especially construction equipment) and vehicles. The companies from the second cluster reported in almost all quarters, but leased almost exclusively the machines (medical equipment, agricultural, processing and other machines). The companies from the third group reported in less than half of quarters (IDEA from 2014 and LP from 2016) and specialized in passenger vehicles and truck tractors. Fourth cluster focused mostly on passenger vehicles and to some extent on other machines, reporting in almost all quarters as well. The fifth cluster specialized in agricultural machines. Companies from this group reported in most quarters.

The sixth group is not very homogeneus: although it is dominated by companies, which leased passenger vehicles, there are also entities specializing in trucks and truck tractors. Moreover, companies from this group reported in various quarters. In seventh cluster there are two companies: CATERPILLAR leased almost only construction equipment (and reported in whole analyzed period), while AKF specialized in various machines (especially agricultural and processing machines) and reported only in 2013-18. VB is in another one-piece group and focused on various vehicles and machines (passenger cars, trucks, semitrailers, trailers, construction equipment and agricultural machines can be distinguished), which reported in 2003-14. The ninth cluster contains companies which leased passenger cars and other machines (both of them reported in 2003-17), while the tenth group focused mostly on truck tractors (reported in almost all quarters). The last cluster is another very little homogeneous group with companies of various specialization (though focused mostly on vehicles and machines).

Figure 4: Dendrogram of `lbk_dba_average` clustering with 11 groups

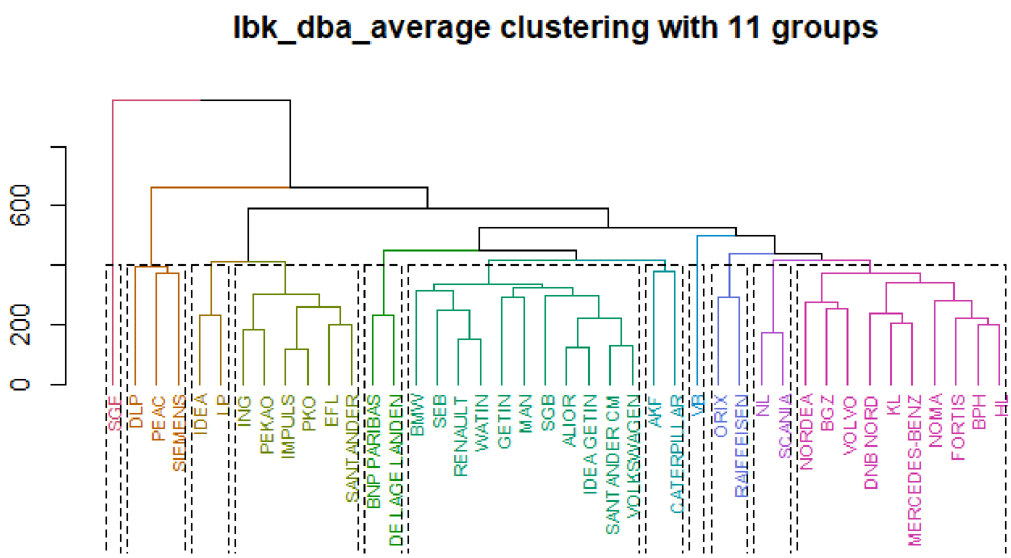


Table 8: `lbk_dba_average` clustering

Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**	**Group 9**	**Group 10**	**Group 11**
SGE	DLP	IDEA	ING	BNP PARIBAS	BMW	AKF	VB	ORIX	NL	NORDEA
	PEAC	LP	PEKAO	DE LAGE LANDEN	SEB	CATERPILLAR		RAIFFEISEN	SCANIA	BGZ
	SIEMENS		IMPULS		RENAULT					VOLVO
			PKO		WATIN					DNB NORD
			EFL		GETIN					KL
			SANTANDER		MAN					MERCEDES-BENZ
					SGB					NOMA
					ALIOR					FORTIS
					IDEA GETIN					BPH
					SANTANDER CM					HL
					VOLKSWAGEN					

Partitional clusterings

In partitional clustering five distance measures were used with their corresponding centroids calculation methods. The `gak` measure was an exception, as there is no method for assessing centroids derived from it. The `gak` was used four times, in combination with centroids obtained by the `dba`, `pam`, `mean` and `median` methods. Each procedure has been calculated for 2 to 10 clusters. A total of 72 partitional clusterings were performed. In section below, these clusterings will be called in form distance measure – centroids calculation method – number of clusters (e.g. `lbk_dba_5`).

Table 9: Distance measures and corresponding centroids calculation methods used in partitional clustering

Distance measure	Centroid calculation method
dtw_basic	dba
lbk	dba
lbi	dba
sbd	shape_extraction
gak	dba
gak	pam
gak	mean
gak	median

Internal and external CVIs

The internal CVIs (table 10) indicate five clusterings which can be considered as the best:

- `gak_mean_8`
- `gak_median_2`
- `gak_mean_2`
- `gak_mean_6`
- `gak_median_4`

The `gak_mean_8` clustering has one of the highest value of Silhouette index in analyzed partitional clusterings (0.25, while maximum for all partitional clusterings is 0.28), Score Function index (0.63) and Dunn index (0.32, maximum is 0.33). It has mediocre values of Davies-Bouldin index (0.95, minimum is 0.7), modified Davies-Bouldin index (0.97, minimum is 0.7) and COP index (0.23, minimum is 0.14), as well as very low value of Calinski-Harabasz index (6.45, while maximum is 45.56).

The gak_median_4 clustering has very high value of Silhouette index (0.27), while all of clusterings mentioned above have Score Function index of 0.63. On the other hand, the gak_median_2 has the highest value of Calinski-Harabasz index (45.56) and the lowest value of Davies-Bouldin indices (0.7 in both cases). These five clusterings have similar value of Dunn index (around 0.3).

Hierarchical clustering has shown that the analyzed data set is so diverse that limiting to clusterings divided into two groups is an oversimplification. In turn, the gak_mean_6 and gak_median_4 clusterings have worse CVIs than gak_mean_8, hence the latter clustering will be analyzed below.

Comparing the lbk_dba_average_11 and gak_mean_8 clusterings, one can see that the former has better values of Silhouette, Davies-Bouldin and Dunn indices, while the latter prevails in Score Function, Calinski-Harabasz and COP indices. At the same time external CVIs do not show much similarities between the two clusterings – although the Rand index is relatively high (0.79), but the other indices are much worse (adjusted Rand: 0.19, Jaccard: 0.19, Fowlkes-Mallows: 0.31, Variation of Information: 0.76).

Summary of gak_mean_8 clustering

The gak_mean_8 clustering is presented in table 12. The first group is dominated by companies which leased vehicles (especially passenger cars) and reported before 2018. The second cluster is one-piece: SEB financed mostly buses and other machines in 2007 – 2010. The third group also consists mostly of car leasing companies, however there are a bunch of entities specialized in agricultural machines. Companies from this cluster reported throughout the analyzed period.

The companies from the fourth cluster focused on various machines, while entities from the fifth group specialized in trucks and truck tractors. The company BGZ was an exception here, as it leased mostly agricultural machines, so perhaps it may be better suited to the fourth group. The sixth cluster consists of two companies involved in the leasing of passenger and commercial vehicles, which reported before 2010. The only company in the seventh cluster, SGB, specialized in passenger vehicles and truck tractors and reported in whole analyzed period. The last group is the least homogeneous – it consists of companies which leased pas-

Table 10: Internal CVIs for the best partitional clusterings

Distance	Centroid	Number of clusters	Internal CVIs						
			Silhouette	Score Function	Calinski-Harabasz	Davies-Bouldin	Modified Davies-Bouldin	Dunn	COP
gak	mean	8	0.25	0.63	6.45	0.95	0.97	0.32	0.23
gak	median	2	0.22	0.63	45.56	0.70	0.70	0.31	0.43
gak	mean	2	0.22	0.63	24.37	0.92	0.92	0.29	0.37
gak	mean	6	0.19	0.63	7.75	1.07	1.12	0.30	0.28
gak	median	4	0.27	0.63	24.22	1.29	1.29	0.29	0.35

Table 11: Internal CVIs for the best hierarchical and partitional clusterings

Clustering	Internal CVIs						
	Silhouette	Score Function	Calinski-Harabasz	Davies-Bouldin	Modified Davies-Bouldin	Dunn	COP
lbk_dba_average_11	0.40	0	4.05	0.44	0.44	0.50	0.34
gak_mean_8	0.25	0.63	6.45	0.95	0.97	0.32	0.23

senger and commercial vehicles, trucks, but also various machines. Moreover, they reported in different parts of the analyzed period.

Table 12: `lbk_dba_average` clustering

Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**
BMW	SEB	ALIOR	AKF	BGZ	RENAULT	SGB	BPH
DNB NORD		BNP PARIBAS	CATERPILLAR	MAN	WATIN		FORTIS
GETIN		EFL	DE LAGE LANDEN	NL			HL
KL		IDEA	DLP	SCANIA			NOMA
MERCEDES-BENZ		IDEA GETIN	SGE	VB			NORDEA
ORIX		IMPULS	SIEMENS	VOLVO			PEAC
RAIFFEISEN		ING					
		LP					
		PEKAO					
		PKO					
		SANTANDER					
		SANTANDER CM					
		VOLKSWAGEN					

TADPole clusterings

The TADPole algorithm uses dtw_basic distance and dba centroids by default. This clustering method demands using one of lower bounds (lbk or lbi) and providing cutoff distance. Since the lower values of this parameter proved that they are not able to provide satisfactory results, the analysis was carried out on three values of cutoff distance: 100, 250 and 500. Each procedure has been calculated for 2 to 10 clusters. A total of 54 TADPole clusterings were performed. In section below, these clusterings will be called in form distance measure – centroids calculation method – lower bound - number of clusters (e.g. dtw_basic_dba_lbi_3).

Internal and external CVIs

The internal CVIs indicate that dtw_basic_dba_lbk_10 and dtw_basic_dba_lbi_8 provide the best results, simultaneously indices prove that increasing the cutoff distance over 100 does not improve CVIs significantly. dtw_basic_dba_lbk_10 clusterings have slightly lower Davies-Bouldin and COP indices, as well as marginally higher Dunn index, while dtw_basic_dba_lbi_8 clusterings are a little better considering Silhouette and Calinski-Harabasz indices.

The external CVIs prove that dtw_basic_dba_lbk_10 and dtw_basic_dba_lbi_8 are fairly similar with Rand index at 0.93, adjusted Rand index at 0.77, Jaccard index equal 0.69, Fowlkes-Mallows index at 0.82 and Variation of Information index equal 0.18.

Comparing these TADPole clusterings with lbk_dba_average_11 and gak_mean_8, it can be observed that dtw_basic_dba_lbk_10 and dtw_basic_dba_lbi_8 have generally lower values of Silhouette index and higher values of Davies-Bouldin indices. However, their Calinski-Harabasz, Dunn and COP indices are at the same – or better – level when comparing to the best hierarchical or partitional clusterings. Simultaneously, the external CVIs does not show many similarities between best clusterings: although the Rand index is relatively high for each pair of clusterings compared, other indexes do not confirm the similarities - with exception of the TADPole clusterings mentioned above.

Summary of **dtw_basic_lbi_8** clustering

The dtw_basic_lbi_8 clustering is presented in table 16. The companies from the first group focused on various machines (especially on agricultural and processing machines). They generally reported in all analyzed periods, with exception of AKF, which was active in

Table 13: Internal CVIs for the best TADPole clusterings

Distance	Centroid	Lower bound	Number of clusters	Cutoff distance	Internal CVIs						
					Silhouette	Score Function	Calinski-Harabasz	Davies-Bouldin	Modified Davies-Bouldin	Dunn	COP
dtw_basic	dba	lbk	10	100	0.16	0	4.01	1.09	1.13	0.45	0.33
dtw_basic	dba	lbk	10	250	0.16	0	4.04	1.11	1.14	0.45	0.33
dtw_basic	dba	lbk	10	500	0.16	0	4.04	1.07	1.10	0.45	0.33
dtw_basic	dba	lbi	8	100	0.17	0	4.74	1.22	1.25	0.44	0.37
dtw_basic	dba	lbi	8	250	0.17	0	4.88	1.14	1.18	0.44	0.36
dtw_basic	dba	lbi	8	500	0.17	0	4.86	1.16	1.18	0.44	0.36

Table 14: Internal CVIs for the best crisp clusterings

Clustering	Internal CVIs						
	Silhouette	Score Function	Calinski-Harabasz	Davies-Bouldin	Modified Davies-Bouldin	Dunn	COP
lbk_dba_average_11	0.40	0	4.05	0.44	0.44	0.50	0.34
gak_mean_8	0.25	0.63	6.45	0.95	0.97	0.32	0.23
dtw_basic_dba_lbk_10	0.16	0	4.01	1.09	1.13	0.45	0.33
dtw_basic_dba_lbi_8	0.17	0	4.74	1.22	1.25	0.44	0.37

Table 15: External CVIs for the best crisp clusterings

Clustering	Comparison to	External CVIs				
		Rand	Adjusted Rand	Jaccard	Fowlkes-Mallows	Variation of Information
lbk_dba_average_11	gak_mean_8	0.79	0.19	0.19	0.31	0.76
lbk_dba_average_11	dtw_basic_dba_lbk_10	0.79	0.21	0.20	0.34	0.71
lbk_dba_average_11	dtw_basic_dba_lbi_8	0.77	0.24	0.23	0.39	0.68
gak_mean_8	dtw_basic_dba_lbk_10	0.90	0.65	0.55	0.71	0.41
gak_mean_8	dtw_basic_dba_lbi_8	0.85	0.53	0.45	0.63	0.50
dtw_basic_dba_lbk_10	dtw_basic_dba_lbi_8	0.93	0.77	0.69	0.82	0.18

2013-2018. The second cluster contains mostly companies specializing in passenger cars, however there were also entities which leased truck tractors or agricultural machines. This also applies to the only company from the third group – BGZ. The fourth cluster contains two companies specializing in passenger cars, while the fifth group is another very little homogeneous group with companies of various specialization: all types of vehicles, as well as other machines. The companies from the sixth group focused on construction equipment, while entities from the seventh group leased mostly truck tractors – exactly like the only company from the last cluster.

Table 16: `dtw_basic_lbi_8` clustering

Group 1	**Group 2**	**Group 3**	**Group 4**	**Group 5**	**Group 6**	**Group 7**	**Group 8**
AKF	ALIOR	BGZ	BMW	BPH	CATERPILLAR	MAN	SGB
DLP	BNP PARIBAS		GETIN	DNB NORD	NORDEA	NL	
PEAC	DE LAGE LANDEN			FORTIS		SCANIA	
SGE	EFL			HL			
SIEMENES	IDEA			KL			
	IDEA GETIN			MERCEDES-BENZ			
	IMPULS			NOMA			
	ING			ORIX			
	LP			RAIFFEISEN			
	PEKAO			RENAULT			
	PKO			SEB			
	SANTANDER			VB			
	SANTANDER CM			VOLVO			
	VOLKSWAGEN			WATIN			

Fuzzy clusterings

The analyzes above show that in the case of the leasing companies, crisp clustering is not a good enough method. Although the majority of companies have dominant categories of financed items, they often also lease other assets on a secondary basis. An additional difficulty is that companies generally reported in different periods. The solution may be fuzzy clustering, which allows for the possibility that each object belongs to a given group only to a certain extent.

In fuzzy clustering five distance measures were used, each with two calculation methods (fuzzy c-means and fuzzy c-medoids). Each procedure has been calculated for 2 to 10 clusters and for three convergence criterion values: 0.001, 0.01 and 1. A total of 270 fuzzy clusterings were performed. In section below, these clusterings will be called in form distance measure – centroids calculation method – number of clusters (e.g. `gak_fcm_5`).

Internal and external CVIs

The internal CVIs (table 17) indicate four clusterings which can be considered as the best:

- `lbi_fcmdd_4`
- `lbi_fcmdd_3`
- `sbd_fcmdd_3`
- `gak_fcmdd_3`

The `sbd_fcmdd_3` and `gak_fcmdd_3` clusterings have relatively high MPC index (maximum for all fuzzy clusterings is 0.36) and low Kwon and Tang indices. On the other hand, `lbi_fcmdd_3` and `lbi_fcmdd_4` clusterings have very high values of PBMF index.

The external CVIs generally prove that these four clusterings are similar to each other, which is confirmed by Soft Rand index (1 for all pairs), Soft Variation of Information index (0.04 – 0.07) and NMIM index (0.69-0.81).

Summary of `lbi_fcmdd_4` clustering

The membership matrix of `lbi_fcmdd_4` clustering is presented in table 19. The representative of the first group is SGE, a company specializing in construction equipment and agricultural machines. However, other companies do not belong to this group to any significant degree. The second cluster is represented by EFL (focused on passenger and commercial vehicles, as well as other machines) and there is four other companies, which

Table 17: Internal CVIs for the best fuzzy clusterings

Distance	Centroid	Number of clusters	Convergence criterion	Internal CVIs				
				Modified Partition Coefficient	Kwon	Tang	SC	PBMF
lbi	fcmdd	4	0.01	0.26	12.6	12.8	0.25	1 400 000
lbi	fcmdd	3	1.00	0.24	13.1	13.6	-0.11	934 000
sbd	fcmdd	3	0.01	0.33	12.1	5.07	0.05	0.27
gak	fcmdd	3	0.01	0.34	7.90	0	-0.03	0

Table 18: External CVIs for the best fuzzy clusterings

Clustering	Comparison to	External CVIs		
		Soft Rand	Soft Variation of Information	NMIM
lbi_fcmdd_4	lbi_fcmdd_3	1.00	0.07	0.70
lbi_fcmdd_4	sbd_fcmdd_3	1.00	0.07	0.70
lbi_fcmdd_4	gak_fcmdd_3	1.00	0.07	0.69
lbi_fcmdd_3	sbd_fcmdd_3	1.00	0.04	0.80
lbi_fcmdd_3	gak_fcmdd_3	1.00	0.04	0.81
sbd_fcmdd_3	gak_fcmdd_3	1.00	0.05	0.76

membership in this cluster is higher than 50% and their profile is similar to EFL. The third group representative is BPH, which specialized in passenger vehicles, trucks and other machines, as well as reported in 2003 – 2008. Other companies, which membership in this cluster is higher than 50%, have similar characteristics to BPH. The last cluster is represented by IDEA GETIN, which focused on passenger and commercial vehicles, truck tractors and agricultural machines, as well as reported in 2016-2020. ALIOR is very similar to this company (membership value of 88% in this group) – the main difference is it did not financed agricultural equipment to a significant extent. In addition, 17 companies showed a similar level of belonging to two or more groups.

Summary of `gak_fcmdd_3` clustering

The `gak_fcmdd_3` clustering provides more consistent membership matrix (table 20). The first cluster is represented by AKF (various machines, especially agricultural and processing, reported in 2013-2018) with SIEMENS as the most similar company (membership value of 58%, focused on medical equipment, agricultural, printing and other machines, reported since 2005). The representative of the second group, PKO, focused on passenger vehicles, truck tractors and other machines, as well as reported in all quarters. There are several companies in this cluster, which membership values were higher than 90% - they generally financed similar assets, however some of them did not report in whole analyzed period. The last cluster is represented by VB (passenger vehicles, trucks, truck tractors, semitrailers and trailers) and the most similar companies are VOLVO (trucks and truck tractors) and MERCEDES-BENZ (passenger and commercial vehicles, trucks and truck tractors) – all three companies have stopped reporting in 2013-2014. There are also 15 companies, which have a similar membership value with two or three groups.

Table 19: `lbi_fcmdd_4` clustering

Company	Group 1	Group 2	Group 3	Group 4
AKF	0.15	0.15	0.24	0.46
ALIOR	0.02	0.06	0.04	0.88
BGZ	0.14	0.17	0.33	0.37
BMW	0.08	0.14	0.24	0.54
BNP PARIBAS	0.12	0.21	0.13	0.54
BPH	0.00	0.00	1.00	0.00
CATERPILLAR	0.15	0.15	0.29	0.41
DE LAGE LANDEN	0.16	0.16	0.19	0.48
DLP	0.38	0.26	0.17	0.19
DNB NORD	0.12	0.18	0.55	0.16
EFL	0.00	1.00	0.00	0.00
FORTIS	0.06	0.08	0.76	0.11
GETIN	0.08	0.21	0.19	0.53
HL	0.06	0.08	0.76	0.10
IDEA	0.06	0.18	0.05	0.71
IDEA GETIN	0.00	0.00	0.00	1.00
IMPULS	0.06	0.53	0.07	0.34
ING	0.18	0.43	0.11	0.29
KL	0.15	0.26	0.43	0.16
LP	0.08	0.19	0.11	0.62
MAN	0.09	0.13	0.25	0.53
MERCEDES-BENZ	0.07	0.18	0.60	0.16
NL	0.13	0.26	0.27	0.34
NOMA	0.06	0.10	0.75	0.09
NORDEA	0.13	0.17	0.40	0.30
ORIX	0.11	0.30	0.36	0.24
PEAC	0.23	0.21	0.29	0.27
PEKAO	0.12	0.66	0.07	0.15
PKO	0.07	0.32	0.08	0.53
RAIFFEISEN	0.10	0.55	0.19	0.17
RENAULT	0.07	0.16	0.50	0.27
SANTANDER	0.12	0.61	0.06	0.20
SANTANDER CM	0.05	0.14	0.16	0.64
SCANIA	0.08	0.13	0.25	0.53
SEB	0.10	0.12	0.46	0.32
SGB	0.10	0.18	0.24	0.48
SGE	1.00	0.00	0.00	0.00
SIEMENS	0.21	0.22	0.24	0.33
VB	0.18	0.32	0.37	0.13
VOLKSWAGEN	0.06	0.17	0.20	0.57
VOLVO	0.09	0.12	0.53	0.25
WATIN	0.08	0.15	0.40	0.37

Table 20: `gak_fcmdd_3` clustering

Company	Group 1	Group 2	Group 3
AKF	1.00	0.00	0.00
ALIOR	0.01	0.99	0.01
BGZ	0.21	0.21	0.58
BMW	0.29	0.41	0.30
BNP PARIBAS	0.21	0.68	0.12
BPH	0.15	0.16	0.69
CATERPILLAR	0.33	0.30	0.37
DE LAGE LANDEN	0.44	0.24	0.33
DLP	0.43	0.25	0.32
DNB NORD	0.18	0.20	0.62
EFL	0.05	0.86	0.09
FORTIS	0.25	0.26	0.50
GETIN	0.26	0.36	0.39
HL	0.17	0.18	0.65
IDEA	0.06	0.90	0.04
IDEA GETIN	0.11	0.81	0.08
IMPULS	0.01	0.98	0.01
ING	0.04	0.93	0.03
KL	0.12	0.13	0.75
LP	0.15	0.73	0.12
MAN	0.26	0.34	0.39
MERCEDES-BENZ	0.09	0.10	0.81
NL	0.12	0.20	0.68
NOMA	0.19	0.20	0.61
NORDEA	0.26	0.26	0.49
ORIX	0.22	0.29	0.49
PEAC	0.48	0.28	0.25
PEKAO	0.03	0.92	0.04
PKO	0.00	1.00	0.00
RAIFFEISEN	0.17	0.27	0.56
RENAULT	0.23	0.23	0.54
SANTANDER	0.04	0.93	0.03
SANTANDER CM	0.01	0.97	0.01
SCANIA	0.20	0.35	0.45
SEB	0.31	0.31	0.39
SGB	0.28	0.44	0.28
SGE	0.33	0.31	0.36
SIEMENS	0.58	0.20	0.23
VB	0.00	0.00	1.00
VOLKSWAGEN	0.15	0.67	0.18
VOLVO	0.08	0.08	0.84
WATIN	0.29	0.29	0.42

Conclusion

The analyzed dataset proved to be a challenge for time series clustering methods when it comes to division of leasing companies into consistent and well-separated clusters. While among the analyzed enterprises one can find those that are so similar to each other that clustering them is not problematic, there are also entities that share characteristics with two or more groups. The fuzzy clustering confirmed this phenomenon - regardless of the chosen method and the number of groups, there were always companies that could be assigned to several clusters at the same time (about one third of the analyzed entities).

These problems resulted mainly from two reasons. Only one third of companies reported in the entire analyzed period - as a result, in some clusters there were entities that financed items from different categories, and their only similarity was reporting at the same time. Moreover, there is a visible advantage of companies for which the most important financed items were vehicles (in particular passenger vehicles).

The analyzes performed highlighted the most frequently leased items. Vehicles, which were a significant specialization of most companies, played a key role in the characteristics of the surveyed enterprises. The aforementioned passenger cars stood out here, but also commercial vehicles, trucks, truck tractors, trailers and semitrailers. On the other hand, among machines, construction equipment and agricultural machinery were the most important categories of financed items.

The other categories (IT, other means of transport, other items and real estates) did not have a major impact on the obtained clusterings - they are usually a side item in the portfolio of leasing companies. The exceptions are: IT for ORIX and SGE, trains for NORDEA, other items for NOMA and real estates for FORTIS (especially commercial and service facilities) and ING (industrial, commercial and office buildings).