PATTERN CLASSIFICATION ON SPECIFICS OF PUBLIC SECTOR INVESTMENTS AND BUDGETING PRINCIPLES

LUKÁŠ BERNAT, RADKA MICHLOVÁ, HELENA MITWALLYOVÁ

Abstract:

The aim of the study is to find patterns in an exact complete data set containing the annual budget of all municipal subjects in the Czech Republic over the past 20 years. The focus of the analysis is on which resources could enable the development of and investment in municipal assets, especially estate property. The financial and real estate analysis (FAMA) method was chosen, which provides comparable indicators to calculate the debt service and other related features of subject performance on a municipal level. These indicators demonstrate whether municipal subjects follow responsible budgeting principles and/or how they utilize their own resources. Comparing similar studies using mentioned data and methodology there is a gap between context of data in time a relation chosen indicators. The reason of obstacle is to put data in time-series and properly analyze them. This appropriate items of indicators are aggregated so all the connections between them and other items are lost. In enormous amount of data study uses classification tools to unfold hidden patterns how does municipal budgeting develop in time without knowing details about each case in context of debt and assets. Study convert time dimension to static indicator of its dynamics a using pure K-Means classification conclude having 6 different clusters which differ each other in some of indicators. Within broader context of those clusters we propose an overview of municipal budgeting strategies. In big cities dominates financing of investment by debt and the rest of clusters differs usually significantly with small impact of their population size that is one of determinants budget income therefore essential budget part.

Keywords:

FAMA; classification, clustering municipalities; investment, budgeting, debt

JEL Classification: C55, C38, H72

Authors:

LUKÁŠ BERNAT, Vysoká škola Ekonomická v Praze, Národohospodářská fakulta, Katedra Filosofie , Czech Republic, Email: luk.bernat@gmail.com
RADKA MICHLOVÁ, Vysoká škola Ekonomická v Praze, Národohospodářská fakulta, Katedra regionálních studií , Czech Republic, Email: radka.michlova@seznam.cz
HELENA MITWALLYOVÁ, Vysoká škola Ekonomická v Praze, Národohospodářská fakulta, Katedra práva, Czech Republic, Email: mitwally@volny.cz

Citation:

1   Data of annual municipal budgets: The Case of Czech Republic

1.1   Budgeting and related indicators

The article deals with the financing of municipal and city assets in the Czech Republic, i.e., covering investment expenditures with external resources and the possible indebtedness of municipalities. The authors focus more closely on new possibilities of exploring the issue due to the large number of municipalities in the Czech Republic, their heterogeneous structure and differences in needs and financial resources.

Local government units have a long-term positive budget balance. The value of the property of towns and municipalities in the Czech Republic is still growing. The guaranteed revenues of local government units are also increasing, but these are mostly used for mandated expenditures. The funds used to cover investments are mainly obtained from surpluses of internal resources or from resources from their own economic activity. However, the latter option is not very extensive in the Czech Republic (cf. Toth, Michlová 2015). The gross savings of cities and municipalities, i.e., resources that can be used to cover investments, are growing. Interestingly, however, even as municipalities are increasing their internal resources available for capital expenditures, they are also using external financing to develop their assets. However, these external resources may have a negative impact on the cash flow of public entities and their overall indebtedness. The indebtedness of public budgets has been an accentuated topic in recent years. Although municipalities and cities contribute a marginal part of the negative balance of public budgets, it is still worth looking deeper into this trend.

The regular monitoring of municipalities and towns in the Czech Republic, published by the Ministry of Finance of the Czech Republic, shows that the bank balances of municipalities have been growing for a long time and their indebtedness has been decreasing, although the ratio of external resources to total assets exceeded the 25% threshold in the last year (2022) for 181 municipalities. However, it is questionable whether the rising level of municipal bank accounts is good news for public finances. Local governments either save for future capital expenditures or build up a reserve for unexpected expenses.

The debt of municipalities is still actual topic to focus on. In some size categories of municipalities the debt is continously growing, some other categories decreasing own debt over past years (Kameníčková 2019). As such reseaches are focusing on dominant subjects in categories resulting in arithmetic conclusion, we would like to gain deeper insight with context of municipal economy especially their development through investment and a value of real assets.

As the collection of data and its analysis developed extensively over past decades (Bernat, Löffler, and Štětinová 2021), it should be also reflected in contemporary research and related hypotheses. Our research focus on municipalities is an example of that and we would like to also reflect Data Science in that article. The goal is to extend an approach that uses just arithmetic values to demonstrate the pattern of focused subjects.

The case study uses a data set of annual financial statement reports provided by every city, town and village in the Czech Republic to the Ministry of Finance of the Czech Republic (for

detailed information see (Ministerstvo financí ČR 2022)). Each report contains hundreds of items and captures the years between 2001 and 2021 covering 6,238 municipal subjects which means 99.7% of all subjects in the Czech Republic.

All that data needs to be enriched by socio-economic statistics (Czech Statistical Office 2022) to provide a complete raw base for the calculation of the financial and estate analysis of cities and municipalities (hereinafter FAMA) which is essential for our analysis. FAMA consists of fewer items which are generalized indicators based on annual financial statements that municipalities have a duty to report.

The FAMA analysis provides a deeper insight into the municipal subject state given years comparable to each other and giving basic patterns if put into the context of time. In such a case we may, for example, analyze how the debt service of municipalities develops over time (compare Toth and Michlová 2014). The easiest way to proceed is to divide the debt service of each municipal subject per year by the number of population and then just calculate the average such per capita indicator for every year (*weighted average*), or just divide the total debt service by the total population per year (*average*), see Chart 1.

Chart 1


Source: ARIS, UFIS, MON1TOR, calculation by authors
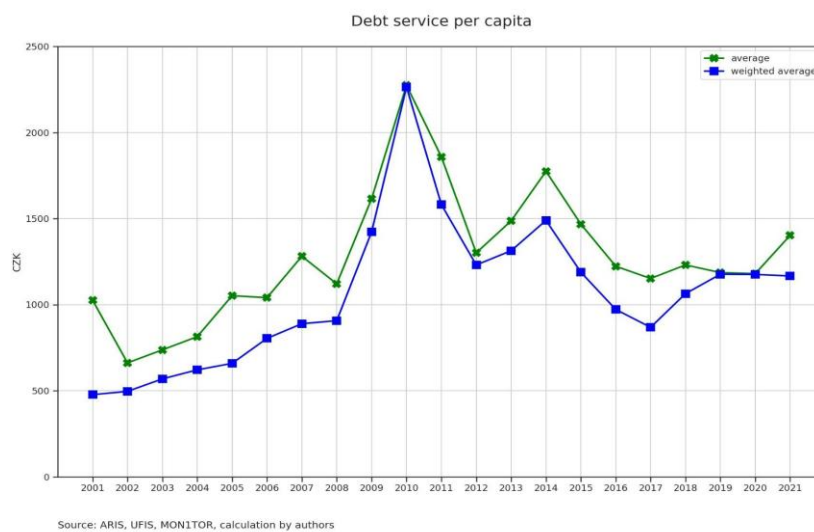
A simple analysis provides us a basic overview of the complete system behavior estimation, nevertheless it is obvious on Chart 1 that we are losing substantial detailed patterns. The arithmetic generalization in this case is limited by choice between the pure average where we can analyze only the pattern of the whole system, or to calculate the average by the municipal subject average which removes the weight of the population in the total average. Such differences can be explained by the distribution spread of debt service *per capita* (see Chart 2), where we see significant differences (spread over the whole scale) over time.

Chart 1 shows the generalized reality over time, showing that there was a large increase in debt among municipalities in 2009 and 2010. This reflects the development of the economy and refers to the economic recession of 2008. This will take time to be demonstrated on the

municipal level, as revenues are mostly made up of tax revenues. Thus, when tax revenues decline one year, there is a reduction in the amount of funds reallocated among municipal budgets in the following fiscal period.

Chart 2

Distribution of debt service per capita in municipalities

Source: ARIS, UFIS, MON1TOR, calculation by authors
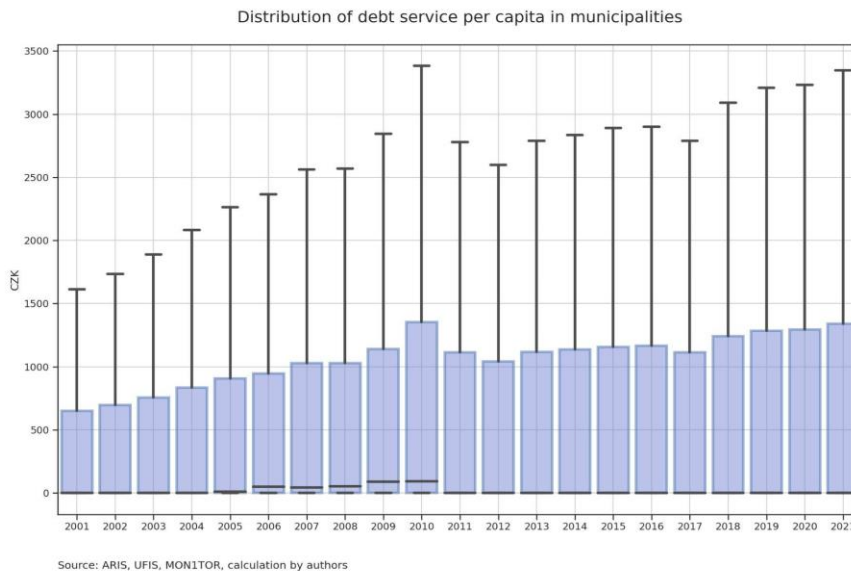
The Czech system of municipal subjects is characterized by high atomization and with vast differences that are obstacles for comparability. More than 90% of the subjects have a population less than 2,000 and the distribution visible on Chart 3 demonstrates that most of them have populations under 500. So, we may expect with certainty huge intervals in FAMA features over different sizes of subjects.

Chart 3

Distribution of municipalities by population (less than 2,000)

Source: ČSÜ, calculation by authors
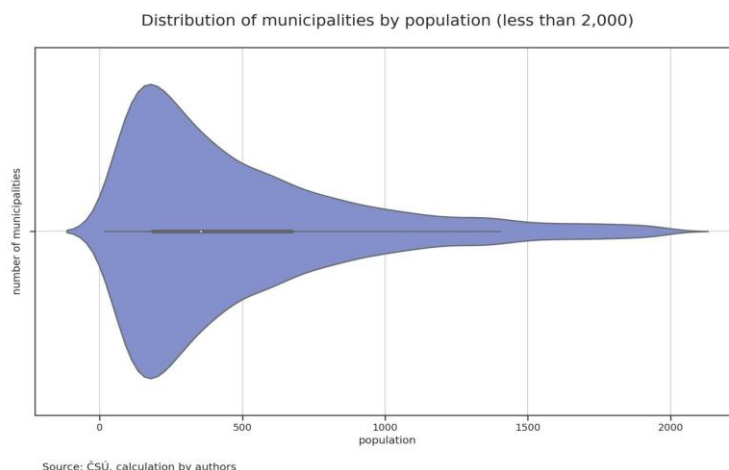
Based on the differences we saw in arithmetic cases, we can assume that also similar size subjects would differ in feature performances. Those differences, the relations and patterns of investment and debt related features are the subject analysis of this article.

1.2   Related research

Municipality-level budgeting has its specifics compared to the national level. A study (Ioannou 2023) shows that this is true across the EU states, pointing to the importance of examining the debt-income relationship on data from 2004-2017. Thus, the growing debts of public finances (Kose 2021) represent not only a matter of discourse at the state level, but a deeper interconnectedness in the participation of municipalities. The study (László 2019) shows the debt structure of individual local governments, which also plays a significant role in the analysis of the municipal budget. It then focuses in detail on the issue of municipal debts and their structure (Sedmihradská, Šimíková 2007).

What circumstances determine the creation of municipal debt in the context of US cities is shown in a paper (Cropf, Wendel 1998) a model explaining the influence of voters' decisions on budgeting. A time-series regression model is used here. They published a similar study (Benito, Bastida 2004) on Spain cities and later (Balaguer-Coll 2016) in the same country, but with a deeper focus on the specifics of individual entities and at the same time the level of debt, they concluded that it is not possible to unify homogeneous recommendations for such a complex structure as municipalities.

With the question of municipalities' debt comes the discussion of how to bail out these debts so that municipal budgets are sustainable. On a sample of US data from the late 1990s, the study (Weber 2010) mentions the possibilities of using municipal property sales and their effect on the real estate market.

The issue of applying classification using big public sector data is addressed for example (Jasova et al, 2017 or Kaderabkova et al, 2011), while the used method of fuzzy c means clustering related to data on the current epidemiological situation in European countries classifies the development of the epidemiological situation in individual actors (compare Cinaroglu 2021). The method brings an innovative approach and is also useful for the case of municipal budgeting, if we want to know the individual details in the time series and opens the potential for further research on public budgets.

A view of wealth from the point of view of public finances offers (Benbrahim et al. 2021) using k-means classification. They cluster individual states according to wealth, but not only in the traditional way according to total GDP and GDP per capita, but also by extension to the country area. In a similar vein, the study (Atalay 2022) dealing with the post-pandemic state of Java villages from an economical perspective identifies what effect COVID-19 had on public finances of municipalities and where the government should concentrate financial assistance. A similar perspective for comparing the context of several indicators at a lower level is addressed (Sugianto 2021, or Kaderabkova, 2021), which with its study compares poverty at the level of complex communities of municipalities Bangka Belitung Island Province.

The direct use of the K-means clustering dealing with the relation of income and expenditure in municipal budgets can be found in the study (Novaliendry 2015) Padang, which attempts to predict future development of budgets using multi-linear regression. With a greater emphasis on time series analysis, ARIMA solves the predictions of municipalities' budgets (Rahmat 2020) on the data in the Province of North Sumatra.

2   Methodology

2.1   Obstacles to model

The approach of this analysis to model the data is from a time series perspective for each subject to capture dynamic relations between concerned features. There are many obstacles to model those FAMA data, such as pure linear regression. Let us name a few of them that we will also face when designing our model. All those are of a kind that we are not able to isolate from generalized item particular acts related to focused features.

Chart 4

Source: ARIS, UFIS, MON1TOR, calculation by authors
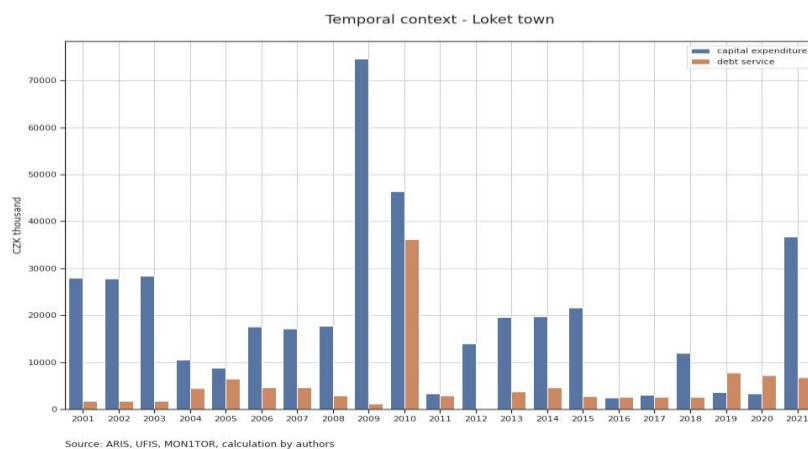
The first is related to the payoff that must be paid after the realized investment supported by debt. In most cases we can estimate the investment by the increase in capital expenditures (e.g., year 2005 on Chart 4) followed by a constant increase of the debt service for some period. As the information about the period is not given, it is very difficult to separate the temporal framework of debt.

This brings another issue of overlapping payoffs and investment in a given period, which makes it even harder to identify the temporal context and its boundaries of debts and payoffs. Municipalities can accumulate many debts and/or payoffs regardless of those two. Even more, they do not need to be dependent on each other at all and can also be for a single period only.

Also, seasonality is problematic in the case of municipalities. That sounds counterintuitive because in the Czech Republic there is a regular 4-year political cycle, nevertheless that does not automatically mean a cyclical reflection on economic features. The example on Chart 4 seems to show that, at least, the first four periods fit into the political cycle, but election years (2002, 2006, 2010 and 2014) or pre-election ones (as populist gestures) do not fit to the cycle. There is also room for other factors like establishing and breaking up coalitions outside of a cycle, or the long-term vs. short-term continuity of applied politics.

2.2   Data and dynamics

To overcome all obstacles and capture the dynamics, the data must be pre-calculated before

fitting to model. As described above it is hard to identify and mark off vectors, so it is necessary to capture dynamics by scalar values that need to be calculated. We wanted to avoid using absolute values, otherwise we would lose the perspective of the given municipality.

For chosen features from FAMA (see the Appendix), it is first necessary to find the potential of any feature in a given municipal subject. The basis for the calculation is the *scope* of potential determined by the maximum value that the subject reached, subtracted by the minimum recorded value (1). Those values can be any time during the concerned period because their purpose is to depict the worst/best case.

(1) $\max_i x_i - \min_i x_i$

The *scope* is insignificant without putting it in proper context, so the calculation also needs to be extended by the weight (2) that the scope has in the maximal potential. The *weight* is an important indicator of how much the feature can change related to its potential.

(2) $(\max_i x_i - \min_i x_i) / \max_i x_i$

Those indicators are just related to a static point of view that is telling us the landscape peaks, valleys and their differences. We also need to capture smaller hills and dents between them along the path from the beginning till the end. For that purpose, we will calculate the sum of all differences (3) that are showing an *increase* step-by-step in time.

As the *increase* could possibly take on a value of 0 in marginal cases or equal to the scope, we would not know whether something happened in between, so the *net increase* also needs to be calculated as a sum of only positive differences of the subsequent values.

(3) $\Sigma (x_i - x_{i-1})$

To place the *increase* indicator or *net increase* respectively in the subject context is to quantify a portion of how much the feature change increased compared to the subject scope. That can be related to the *effort* (*net effort* respectively) that the government put to change related feature (e.g., investment).

The investment ratio and coefficient of ordinary expenditure cover are special cases, which are already relative in the time series, so in this case we record the mean and standard deviation. For selected features, paired Pearson correlations were calculated in order to group municipal subjects that have similar relations between those feature progressions.

We will keep those indicators separately for the model. It is not necessary to include them for

further calculation, because even subjects with similar indicators could differ in other indicators closely related to the similar one. As the model is looking for classified similarities in a few clusters, it is desirable to have a reference to such single indicators, which is possible to interpret.

3   Model

As mentioned before, we only know ahead which size category municipalities are and that is the only known feature in the data in question that apparently affects economic behavior. Specifically, the size category is one of the factors for tax redistribution to the municipalities, so it affects the amount of money in the budget that the subjects receive (for more information see MFČR 2022). We want to identify which subject belongs to a category according to their debt and investment patterns, and we do not have such categories prior to the analysis.

In that particular case, we would need unsupervised machine learning method classification (see Bernat, Löffler, and Štětinová 2021), so that we do not have any labels providing information about the category they belong to. The relevant discussion here is which of the many available methods to use on a scale starting from simple ones like K-means clustering to much more sophisticated ones like fuzzy clustering (compare Łuczak and Kalinowski 2022). As the purpose of this paper is to identify basic patterns that determine the behavior of municipalities, we would be satisfied with the very basic machine learning model.

This study provides extension to basic K-means clustering of static features that characterize subject of research (compare Rovan, Sambt 2003) by putting it in to context of time. That is added value to research of municipalities development, because indicators of budget are cumulative, so it is impossible to isolate particular items relation between indicators. So we decided to calculate features as scalar indicator of dynamics.

Similarly were analyzed indicators of financial performance in local government in study "Evaluating financial performance in local government: maximizing the benchmarking value." (Zafra-Gómez, López-Hernández, Hernández-Bastida 2009) This study is using commonly accepted financial indicators related to financial fitness.

3.1   Pre-processing analysis

The data needs to be pre-processed before being input to the model to increase the computational performance by the standardization of the scale to unit variance (Pedregosa et al. 2011). That is also needed to perform a Principal Component Analysis (PCA) orthogonal linear transformation. The goal is to express a complete dataset with the least number of features. Similar attitude is used for various municipal analysis e.g. on land-use data (Kušová, Tesitel, Matějka and Bartoš 2023). Alternative approach to use PCA is apply on regression function (see Yamaha 2015) or directly to feature affecting country development (Olawale and Garwe 2010).

As RUD is a very complex set of indicators, we need only its fragment, which is directly or indirectly related to municipality investments. Indirect features are those which somehow

determine features which are directly related to investments (e.g., sources that are necessary to realize investments). That reduces the number of features from 82 to 23. We could certainly build a model for all features, but we would lose the purpose of this study - to classify municipal subjects by their relation to investments.

As we can see on Chart 5, there is almost no linear dependency between the features which is a good sign for the model that used features would not needlessly overload the computational capacity, so that all of them are important to describe the model. Based on the PCA calculation, we can exclude 17 of the 23 features (see Appendix) so that the remaining 6 features will describe 95% of the data variability. Similarly, if we calculate the same for all features, only 29 of the 82 features describe 95% percent of the data variability.

Chart 5
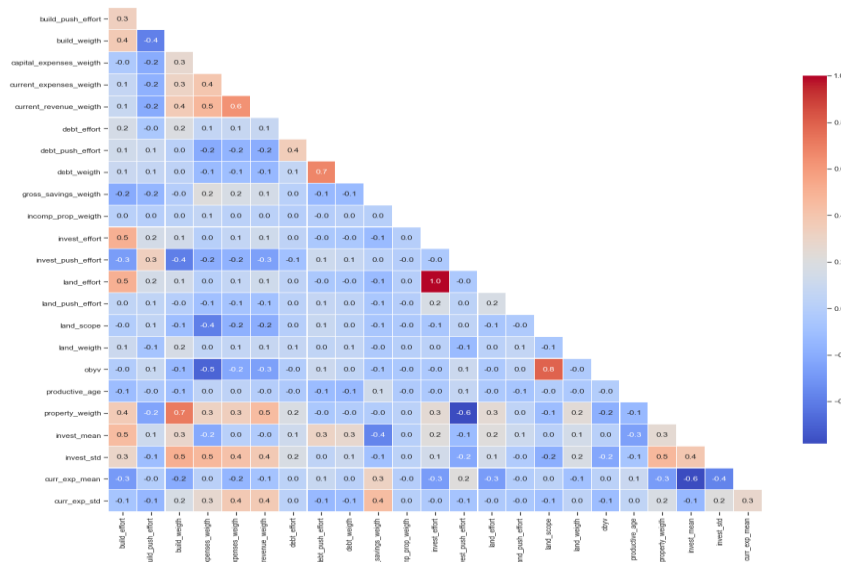
Some very interesting features were eliminated from the list. Let's focus on them closely. For example, the *population* of the subject, which is one of determinants of municipality income from the state budget, is on the list of surprisingly missing features the model described. Looking at the correlation matrix, we can also conclude that the only feature which has linear dependency is to *land scope* (0.8), to *capital expenditure weight* (-0.5) and the others are about 0, so even not affecting each other due to the per capita calculation in some cases.

The other eliminated feature that is also related to *population* is the *productive age ratio*, which describes the ratio of people in productive age to the total population. We could expect that the more productive people inhabiting the municipality would also affect the income of the municipality, whose determinant is also tax paid to the state in a given district. So, the *productive age ratio* is doubtless influencing the state income. Its correlation to other features is negligible and close to 0.

As we are concerned with investment relations, we are surprised by the excluded feature

property weight. Property is the output of realized investment, so we expect it to be a part of the model. Luckily, this eliminated feature is just a *weight*, i.e., the ratio of the scope to the max value. For explanation of the data variance, other property measures are still important, and therefore remain.

By the number of reduced features, we still keep the most important attribute and capture the maximum information about the dataset. Those features are combined with the PCA to just 2 dimensions for the further classification calculation. The number of components for the dimensionality reduction was chosen simply for the purpose of a better visual interpretation of newly-arising clusters. Afterwards, the important features are identified by their belonging to clusters.

To compress the data into a dimension, we use the principal component analysis (Pedregosa et al. 2011). The dimensionality reduction is performed not only for visual presentation purposes, but also to improve the computational performance algorithm and the eradication of correlated features to avoid any bias toward any set of features.

## 3.2   Clustered municipalities

The RUD methodology consists of 13 size categories to classify the municipal subject. As this is also one of the determinants of income which is closely related to the focused model subject features, we can therefore expect that the size category could be a label for classification in our case, so we will also focus on how clusters fit size categories.

Unsupervised learning classification is specific in that we do not know the number of final clusters, but we need to define it previously as a parameter for the model. To determine the number of clusters, the so-called Elbow method was used, which is computing Within Clusters Sum of Squares (WCSS) that identify a threshold when putting an additional cluster that does not improve the model significantly (4), so it could cause overfitting (compare Bernat, Löffler, and Štětinová 2021).

$$(4) WCSS = \sum_{C_k}^{C_n} \left(\sum_{d_i \in C_i}^{d_m} distance(d_i, C_k)^2 \right)$$

To determine the maximum number of clusters as the threshold, where WCSS stops computing (*n*) is the number of size categories to make a proper comparison. Our model therefore has a target value of 6 as an optimal number of clusters we are looking for (see Chart 5). This first information depicts the difference we can see between the predetermined number of categories by size and the natural number of clusters according to pattern municipalities formed by investment and related indicators.

Chart 6
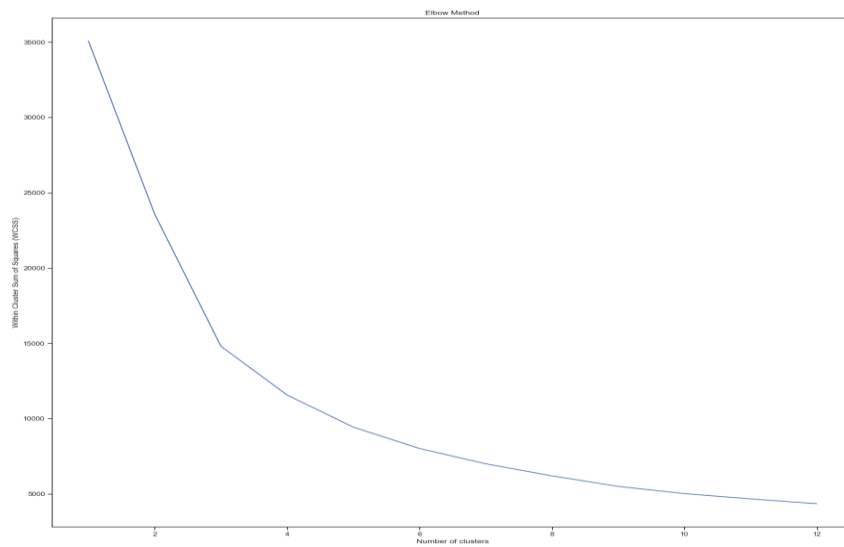
As proposed during the methodology determination, the method for classification is the simplest clustering K-means as a continuation of the Principal Component Analysis. Chart 7 shows the spread of normalized and dimensional reduced values. At first glance, there is no visible pattern that could arise just from the distinct concentration of elements, so the variance is unequal. We need to calculate it to recognize the desired 6 clusters.

Chart 7
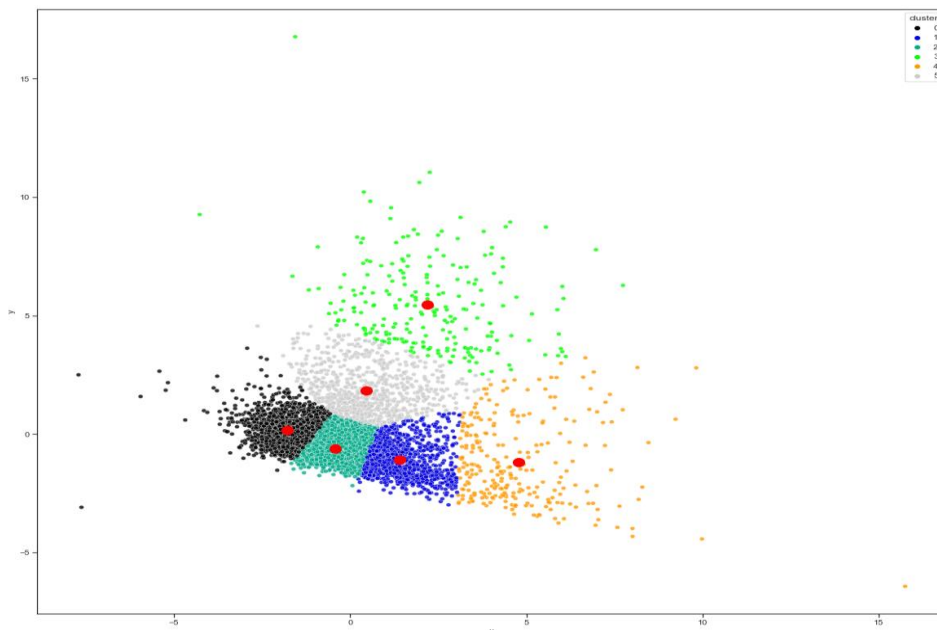
In the first step, we need to separate *n* groups of items (in our case 6) with equal variance as defined with within-cluster sum-of-squares to find the cluster nearest the "centroids" to minimize inertia. In the second step, we re-define centroids by calculating the mean of all the samples assigned to the initial centroids (scikit-learn developers, 2022). These two steps are

repeated until the centroid does not move significantly, so such clusters should then be internally coherent. In other words, within-cluster variances are minimized (compare Sculley 2010).

In our case, the model computation took 26 repetitions to reach the optimal inertia. The centroids of each cluster are shown on Chart 7 by 6 bigger red dots. All items falling in a particular cluster are differentiated by colors. So far, we reach the point that we have clusters and all items are labeled by their belonging to clusters. This does not tell much about municipality patterns appropriate to the clusters; any characteristics that differentiate them from other clusters. All we have is a 2-dimensional array with virtual values that is not possible to interpret (not even the population size).

4   Interpretation of result

To distinguish the characteristics and patterns, we need to take some steps back. As we labeled a set of reduced data, we can also join those labels back to the dataset with calculated indicators. There are still too many, so we need to focus on statistics and compare them across clusters to see which relate to each other.

Chart 8
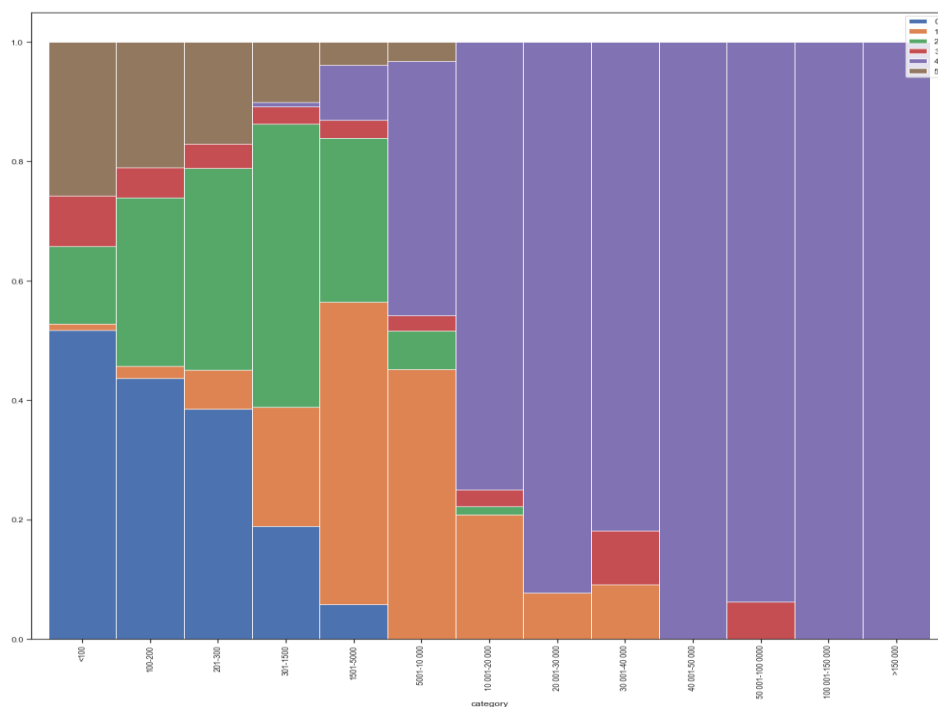
Before we do this, let's focus on the frequently mentioned relation of municipality sizes and clusters related to debt services, investments and other related FAMA indicators. On Chart 8, we see the relative distribution of municipality size categories rating differentiated by colors in the columns by the clusters they belong to.

Chart **9**
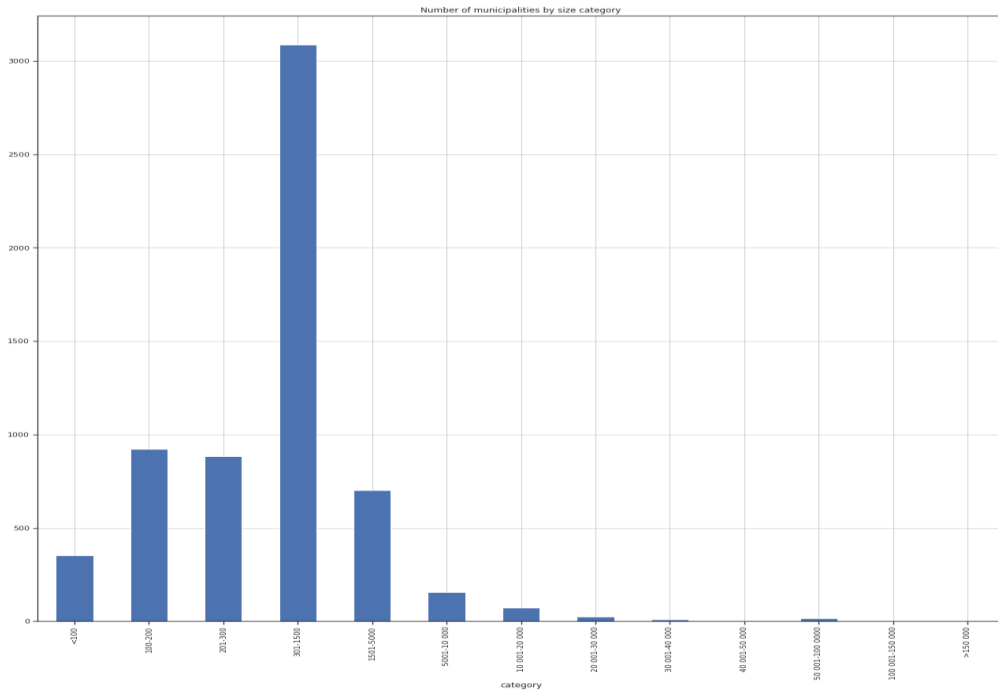
This chart describes the spread of clusters over size categories, but do not forget its relativity. The conclusion that the purple cluster #4 dominates the rest is completely wrong. In Table 1, we can see that even cluster #4 overlaps 10 size categories; this is represented in only about 3% of all municipalities. On the contrary, cluster #3 (red color) on Chart 9 looks like it is minor, even though it is overlapping 7 size categories, but it dominates the proportion in municipalities with more than 35%. We detect a high contrast between clusters #1 and #3, which cover about 65% of all municipalities, and clusters #4 and #5, which only cover about 6%, even though they overlap more size categories.

Table **1**

Cluster number	Municipality count	Count ratio	Size category occurrences
0	911	0.146	9
1	1893	0.303	6
2	784	0.125	6
3	2227	0.358	7
4	189	0.030	10
5	234	0.038	9

An important finding that Chart 9 depicts is the likelihood that a particular size category belongs to some cluster. As we can see, the purple #4 cluster dominates all the bigger cities with more than 5,000 inhabitants. That means that there is very little variability in patterns concerning our topic. The red #3 cluster is also interesting, as it overlaps the widest spread of sizes, which leads us to conclude that in this case there is weak linkage between the size of a municipality and its pattern regarding debts and investments. The remaining clusters determine more consistent groups from a size point of view.

4.1   Cluster specific characteristics

To compare the differences and conversely the common similarities, we use distribution statistics of model features shown on box-plots (see the Appendix). There we can see specifics that differentiate clusters from each other. Table 1 is a summary overview of all those characteristics based on where they fit the most. That brings us back to the original feature indicator values before the dimensionality reduction just labeled by clusters.

From the results, we can see that clusters really differ and also have common patterns within their own group. Let's highlight some of the most important and interesting findings regarding clusters. All those findings also have some relations to the size categories.

The first finding is related to owned buildings and their dynamics over time. Cluster #3 is mostly independent of size categories and rather decreasing their building assets value over time. That distinguishes the cluster from the rest, even though the weight varies a lot in the set of those municipalities. The same situation as in the case of buildings is regarding the land effort and push effort. Cluster #3 has negative progress over time regarding those indicators and also the invest effort should be highlighted, because this cluster is the only one whose values are slightly above 0, so almost 50% is negative.

Members of cluster #4 - the biggest municipalities, are specific with the lowest capital expenditures and on the other hand a high debt push effort. Compared to the other subjects in clusters, they have the smallest portion of savings, so in that case we can conclude that bigger towns and cities invest using mostly sourcing not from their own sources. They are using debts and loans for their investments according to the assumption that it is easier for them to find allocation in budgets.

The rest of the clusters cover the smallest subjects, which is more than 60% of the whole. A low build effort is the first factor that differentiates cluster #0 from the others as they have it significantly higher. This cluster is specific by 0 debt changes, so we can say that those municipalities do not use debt as a solution for their activities and also do not have much debt in the past. Of course, this has consequences in the form that the cluster has lower investments than the others.

Clusters #1 and #2 differs the most in the property they have, because #1 has significantly more. On the other hand, #2 has a higher invest effort financed by debt, which seems logical, because cluster #1 pronouncedly fills its budgets with a higher amount of money in the form of current revenue.

The last cluster #5 is specific in increasing the building value (is higher than market value, because real estate prices increase is stable over time and gradual mild (see Hromada 2021)) with lower capital expenses and lower current income. They use debt funding but not in all cases, because the spread of debt is half negative, so some of them also decrease debt. Their invest effort is high similarly to cluster #1 and #2.

Figure 1

CLUSTER	0	1	2	3	4	5
BUILD EFFORT	Mostly low	High	High	Mostly negative	High positive spread	High
BUILD PUSH EFFORT	Slightly increase	Slightly increase	Slightly increase	Decrease	High	High
BUILD WEIGHT	High	Moderate	Very high	Wide spread	Low	Wide spread
CAPITAL EXPENSES WEIGHT	High	High	High	High	Low	Low
CURRENT EXPENSES WEIGHT	Moderate	Moderate	Higher many outliers	Moderate	Low	Moderate
CURRENT REVENUE WEIGHT	Moderate	Moderate	High	Wide spread	Low	Low
DEBT EFFORT	Zero A lot of outliers	Low	Slightly above zero	Slightly under zero	Wide spread around zero	Wide spread around zero
DEBT PUSH EFFORT	Moderate	High	Moderate	Moderate	High	High
DEBT WEIGHT	High spread	No effect	High spread	No effect	No effect	No effect
INVEST EFFORT	High with spread	High	High	Low	High with spread	High
INVEST PUSH EFFORT	High	High	Moderate	Wide spread	High	High
LAND EFFORT	High with .25lower	High	High	Zero with both side	High with .25lower	High

	spread			spread	spread	
LAND PUSH EFFORT	Moderate	Moderate	Moderate	Low	High	Moderate

5   Conclusion

The method proved that it is a valuable and scalable tool to analyze hypotheses on complex data even though the simplest variant was used. The classification overcomes difficulties essential for given data and brings us to clear results. The clustering uncovered unseen ties and patterns inherent in the subjects. Compared to pure arithmetic aggregation, that is a huge advantage, which enables us to see details which are hidden, but those details make the difference.

As we saw, in the case of municipalities there is a pattern that subjects that have a population of more than 10,000 use loans for their investment compared to savings in the case of smaller and the smallest subjects. We also found a cluster that overlaps ordinary categories but contains more than 35% of all the subjects. There were also interesting results related to the rest of the clusters, because they were not homogeneous as may have been expected, but distinguished significantly in many ways that would remain hidden for pure mathematics.

Wide options on how to use the classification for the purpose (but not only) of focusing on bigger details, how municipalities use their budget for investment and which sources are used, remain for further research. That was not even the purpose of our research but huge opportunities to focus on fine details can follow.

Comparing to arithmetical generalization we can conclude that smaller municipalities also decreasing their debt and if so they split into 2 categories differing by increasing/decreasing value of their building assets (compare Kameníčková 2019). The more significant case are municipalities about 2000 population which differ the most and cover almost all clusters which differ in trends of debts, building value and investment expenditure.

6   References

Atalay, M. *Investigating the Impact of the COVID-19 Pandemic on Economic Activities Using Multidimensional Scaling and K-Means Clustering Analysis.* EKOIST Journal of Econometrics and Statistics, 2022, 37: 171-198.

Balaguer-Coll, Maria Teresa; Prior, Diego; Tortosa-Ausina, Emili. *On the determinants of local government debt: Does one size fit all?.* International Public Management Journal, 2016, 19.4: 513-542.

Benbrahim, S., Chauche, S. N., Toumache, R., & Lequad, E. A. *Countries' economic segmentation using k-means clustering for the year 2021.*

Benito, Bernardino; Bastida, Francisco. *The determinants of the municipal debt policy in Spain.* Journal of Public Budgeting, Accounting & Financial Management, 2004, 16.4: 492-525.

Bernat, Lukáš, Vladimír Löffler, and Barbora Štětinová. 2021. *BIG DATA a umělá inteligence pro*

manažery: Praktický návod, jak držet krok s dobou v 21. století. 1st ed. Karlovy Vary: Vladimír Löffler.

Cinaroglu, Songul. *Comparison of Fuzzy C-Means and K-Means Clustering Performance: An Application on Household Budget Survey Data*. In: Intelligent and Fuzzy Techniques: Smart and Innovative Solutions: Proceedings of the INFUS 2020 Conference, Istanbul, Turkey, July 21-23, 2020. Springer International Publishing, 2021. p. 54-62.

Cropf, R. A.; Wendel, G. D. *The determinants of municipal debt policy: a pooled time-series analysis*. Environment and Planning C: Government and Policy, 1998, 16.2: 211-224.

Czech Statistical Office. 2022. "Public database VDB." https://vdb.czso.cz/vdbvo2/.

Hromada, Eduard, et al. Development of the real estate market in the Czech Republic in connection with the Covid-19 pandemic. In: *Proceedings of Economics and Finance Conferences*. International Institute of Social and Economic Sciences, 2021.

Ioannou, Stefanos. *Regional and municipal debt in the Eurozone: a cross-country analysis*. Regional Studies, 2023, 57.1: 97-111.

Jasova, E., Kaderabkova, B., Cermakova, K. (2017) Use of the method of the stochastic trend for NAIRU estimation in the Czech Republic and Slovakia at the macro- and meso-levels, Economic Research-Ekonomska Istraživanja, 30:1, 256-272, DOI: 10.1080/1331677X.2017.1305782

Kaderabkova, B., Jasova, E. (2021). How the Czech Government got the Pandemic Wrong. Proceedings of the 15th Economics and Finance Conference, Prague. DOI: 10.20472/EFC.2021.015.005

Kadeřábková, B., & Jašová, E. (2011). Analysis of the Indicator NAIRU on the Sector Level. *Politická ekonomie*, *59*(4), 508-525. doi: 10.18267/j.polek.802

Kameníčková, Věra. Je dluh obcízávažný problém? *Deník veřejné správy,* 2019.

Kose, M. Ayhan, et al. Global waves of debt: Causes and consequences. World Bank Publications, 2021.

Kušová, Drahomíra, et al. Socio-economic conditions in selected biosphere reserves. *Silva Gabreta*, 2006, 12.3: 157-169.

László, Viktor. Local Government Debts in the EU countries. In: 27th NISPAcee Annual Conference, May. 2019. p. 24-26.

MFČR. 2022. "Metodická podpora Územní rozpočty." Ministerstvo financí. https://www.mfcr.cz/cs/verejny-sektor/uzemni-rozpocty/metodicka-podpora.

Ministerstvo financí ČR. 2022. "Monitor - kompletní přehled veřejných financí." Monitor - an information portal of the Ministry of Finance of the Czech Republic. https://monitor.statnipokladna.cz/.

Novaliendry, D., Hendriyani, Y., Yang, C. H., & Hamimi, H. (2015). *The optimized K-means clustering algorithms to analyzed the budget revenue expenditure in Padang*. Proceeding of the Electrical Engineering Computer Science and Informatics, 2(1), 61-66.

Olawale, Fatoki; Garwe, David. Obstacles to the growth of new SMEs in South Africa: A principal component analysis approach. *African journal of Business management*, 2010, 4.5: 729.

Pedregosa et al. 2011. "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (85): 2825-2830.

Rahmat, R. F., et al. *Prediction of Regional Revenue and Expenditure Budget using Autoregressive Integrated Moving Average (ARIMA)*. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing, 2020. p. 012064.

Rovan, Jože; Sambt, Jože. Socio-economic differences among Slovenian municipalities: A cluster analysis approach. *Development in Applied Statistics*, 2003, 19: 265-278.

scikit-learn developers. n.d. "Clustering." 2022.

Sculley, David. 2010. *Web-scale k-means clustering*. N.p.: Proceedings of the 19th international conference on World wide web. https://www.eecs.tufts.edu/~dsculley/papers/fastkmeans.pdf.

Sdmihradská, Lucie; Šimíková, Ivana. *Municipal debt in the Czech Republic*. Debt Management in Transition Countries: Experience and Perspectives. Bratislava: NISPAcee, 2007, 43-59.

Sugianto, Castaka Agus; Bokings, Tri Pratiwi Olivia Riska. *K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province*. Journal of Computer Networks, Architecture and High Performance Computing, 2021, 3.1: 58-67.

Toth, Petr, and Radka Michlová. 2014. *Ekonomické aktivity obcí a měst*. 1st ed. Plzeň: Vydavatelství a nakladatelství Aleš Čeněk.

Weber, Rachel. Selling city futures: *The financialization of urban redevelopment policy*. Economic geography, 2010, 86.3: 251-274.

Yahaya, Onipe Adabenege, et al. Principal Components Analysis of Local Government Revenue in Nigeria: 1993–2014. *Public Policy and Administration Research*, 2015, 5.10: 38-47.

Zafra-Gómez, José Luis; López-Hernández, Antonio Manuel; Hernández-Bastida, Agustín. Evaluating financial performance in local government: maximizing the benchmarking value. *International Review of Administrative Sciences*, 2009, 75.1: 151-167.

Appendix

feature	explained variance	meaning
build_effort	0.189818	Building value change
build_push_effort	0.143029	Building value change (only positive increase)
build_weigh	0.090156	Building value scope related to its potential
capital_expenses_weigh	0.067568	Capital expenditure scope related to its potential
current_expenses_weigh	0.056306	Current expenditure scope related to its potential
current_revenue_weigh	0.045601	Current revenue scope related to its potential
debt_effort	0.044431	Debt change
debt_push_effort	0.043032	Debt change (only positive increase)
debt_weigh	0.040874	Debt scope related to its potential
gross_savings_weigh	0.039295	Gross savings scope related to its potential
incomp_prop_weigh	0.037699	Incomplete property scope related to its potential
invest_effort	0.033930	Realized investment change

invest_push_effort	0.030014	Realized investment change (only positive increase)
land_effort	0.026807	Owned land value change
land_push_effort	0.023670	Owned land value change (only positive increase)
land_weigth	0.021862	Owned land scope related to its potential
population	0.015886	Total population of municipality

Boxplot grouped by cluster