

---

## The cost of living in the best livable cities in the world: a brief predictive quantitative analysis

---

Rafaela Costa Martins de Mello Dourado\* and  
Alessandra de Ávila Montini

Departamento de Administração,  
Faculdade de Economia Administração e Contabilidade,  
Universidade de São Paulo,  
Avenida Professor Luciano Gualberto,  
908, Cidade Universitária, São Paulo – SP, Brazil  
Email: rafaelamellodourado@gmail.com  
Email: amontini@usp.br  
\*Corresponding author

**Abstract:** There are a lot of studies, papers, summaries, reports, and articles regarding the cost of living in different cities around the world. Although these studies are rich and robust, only the final results are published, like cities rankings, or summaries of top cities and lowest ranking cities. This paper's goal is to study the most livable cities using living cost data. To accomplish this objective, a cluster analysis has been conducted using 2015 prices data and three clusters were obtained in result: high, medium and low cost. In addition, a multinomial logistic regression using 2014 prices data was adjusted to predict the cluster each city would fall into. This model could help companies or even people to decide which city to move to in order to decrease living costs. It can be important to avoid a wrong decision in case of an upcoming cluster change for a determined city.

**Keywords:** multinomial logistic regression; cluster; living costs; city; livable cities; predictive model.

**Reference** to this paper should be made as follows: de Mello Dourado, R.C.M and de Ávila Montini, Á. (2016) 'The cost of living in the best livable cities in the world: a brief predictive quantitative analysis', *Int. J. Multivariate Data Analysis*, Vol. 1, No. 1, pp.28–42.

**Biographical notes:** Rafaela Costa Martins de Mello Dourado currently works as a Statistical Modelling Analyst at Serasa Experian, Brazil. She received her Master's degree in Quantitative Methods and Computing Area of the Economy, Administration and Accounting College of Universidade de São Paulo. Her specialisations are data analysis and data mining at FIA in 2013. She received her Bachelor in Mathematics from the Mathematics and Statistics Institute IME-USP in 2011.

Alessandra de Ávila Montini currently occupies the position of Professor and researcher in Quantitative Methods and Computing Area at the Economy, Administration and Accounting College of Universidade de São Paulo. She received her Doctor's degree in Business Administration from the Economy, Administration and Accounting College FEA-USP in 2003, and Bachelor and Master's degrees in Statistics from the Mathematics and Statistics Institute IME-USP in 1995 and 2000, respectively. Her experiences are data mining, big data and applied quantitative methods areas. She is a Big Data Study Core Coordinator – CNPQ.

## **1 Introduction**

The curiosity and the analysis about costs of living are not new to human being. It is possible to find studies and even books from the last century comparing countries, cities, counties, like Wolman (1929) who discussed consumption and the standard of living in the USA; Kato et al. (1949) who investigated the food consumption and the cost of living in Tokyo; Lenzen (1998) who talked about different aspects of living using energy and greenhouse gas cost for Australia during 1993–1994; and finally the methods of calculating costs of living by Pollak (1981) and after by Porter-Hudak (1991).

On recent studies, *The Economist* (2015) has shown the urbanisation and rise of the megacities from 1950 to 2014 and it has also predicted the next year until 2030, when nearly 9% of the world's population will be living in just 41 megacities (more than 10 m inhabitants). Every day the world claims more and more for demographical changes and new dwelling ideas. Certainly livable cities and living costs are among the top subjects in the media and in different governments.

Alonso (1964), Mills (1967), and Muth (1969) developed a model that studies the urban economics using amenities, income, housing costs and transportation costs. With a brief comparison to *The Economist's* (2015) study cited before, in 1964 (when Alonso started the model) only 1.4% of the cities in the world were considered as megacities, and they grew to 6.4% in 2015. Certainly, those three American economists were visionaries. Besides, Glaeser (2008) said that Alonso-Muth-Mills model is useful not only in predicting housing prices, but also in predicting the location of different population groups.

Duranton and Puga (2013) identified four key drivers of the population growth of cities in developed economies: transportation and housing supply; amenities; agglomeration effects; technology and shocks to specific cities or industries. With these engines of city growth comes the living costs and with high demand (because of population's growth) comes even higher costs, but there is no paper that provides a predictive cost model. Thinking of that, would it be possible to forecast future living costs of a city with the present pieces of information we have on costs?

The internet brought people closer to economics, education, culture and knowledge, and this device opened the doors to global studies with international data. The Economist Intelligence Unit (EIU), The Expatistan, Numbeo, The Best Places to Live, The Mercer and The WolframAlpha are some of the biggest websites that provide costs of living around the world. These websites, like The EIU, has rich and robust studies, but the companies only publish the final results, like cities ranking or summaries of top cities, or even bottom cities. The numbers and details about this kind of analysis are not published, although it is possible to buy the raw data by paying an expensive price for it.

This paper provides a brief study using the knowledge of The EIU by considering its cities rankings at the Numbeo raw data, which is collected by web users and it has an open access. At the end of the article you will be able to understand a little bit more about 61 of the 70 top cities to live and to predict which cluster (high, medium, or low costs) a city will be in the next year, using current year data.

## 2 Methodology

The Economist Intelligence Unit (2012) published the best cities ranking and report using, what was called as ‘Spatial Adjusted Livability Index’ to rank the cities around the world. This index is a simple average of seven indicators: green space, sprawl, natural assets, cultural assets, connectivity, isolation and pollution, and it was created by Filippo Lovato who won a competition provided by EIU and BuzzData. As a special report, EIU published more data than they are used to, so it is possible to find the top 70 livable cities around the world ranked by Lovato’s index.

From these 70 cities, 61 were found at The Numbeo website and 49 variables about living costs (USD) were collected in 2014 and in 2015 by each city, which can be seen at Table 1.

**Table 1** detailed living costs variables (USD)

CheapMeal	Meal in an inexpensive restaurant
MidMeal	Meal for 2 people in a mid-range restaurant
McCombo	McMeal at McDonald’s or equivalent combo meal
DomesticBeerRest	0.5 litre of domestic beer at restaurant
ImportedBeerRest	0.33 litre of imported beer at restaurant
Coke	0.33 litre of coke at restaurant
WaterRest	0.33 litre of water at restaurant
Milk	1 litre
Bread	500 g of fresh white bread
Eggs	1 dozen
Cheese	1 kg of local cheese
Water	1.5 litre bottle
Wine	1 bottle mid-range
DomesticBeer	0.5 litre
ImportedBeer	0.33 litre
Cigarette	1 pack of Marlboro
OneWayTicket	Local transportation
ChickenBreasts	Boneless, skinless, 1 kg
MonthlyPass	Regular price
Gasoline	1 litre
VolkswagenGolf	1.4, 90 Km Trendline or equivalent new car
Ap1Center	Rent an apartment with 1 bedroom in city centre
Ap1OutCenter	Rent an apartment 1 bedroom outside of centre
Ap3Center	Rent an apartment with 3 bedrooms in city centre
Ap3OutCenter	Rent an apartment with 3 bedrooms outside of centre
Basic	Electricity, heating, water, and garbage for 85 m <sup>2</sup> apartment
MobilePrepaid	1 minute of prepaid mobile tariff local without discounts or plans

*Source:* Numbeo website

**Table 1** detailed living costs variables (USD) (continued)

Internet	10 Mbps, unlimited data, cable/ADSL
FitnessClub	Monthly fee for 1 adult
TennisRent	Tennis court rent 1 hour on weekend
Cinema	International release, 1 seat
Jeans	1 pair of jeans Levis 501 or similar
Dress	1 summer dress in a chain store (Zara, H&M, ...)
Nike	1 pair of Nike running shoes mid-range
Shoes	1 pair of men leather business shoes
ApMeterCenter	Price per square meter to buy an apartment in city centre
ApMeterOutCenter	Price per square meter to buy an apartment in city centre
Monthly salary	Average monthly disposable salary after tax
Mortgage	Mortgage interest rate in percentages, yearly
TaxiStart	Taxi start normal tariff
Taxi1Km	Taxi 1 km normal tariff
Taxi1Hour	Taxi 1 hour waiting, normal tariff
Apple	1 kg
Orange	1 kg
Potato	1 kg
Lettuce	1 head
Cappuccino	Regular
Rice	White rice 1 kg
Tomato	1 kg

*Source:* Numbeo website

The 61 cities being analysed are: Amsterdam, Athens, Atlanta, Bangkok, Beijing, Belgrade, Berlin, Bogota, Boston, Bucharest, Budapest, Buenos Aires, Cairo, Casablanca, Chicago, Dhaka, Guangzhou, Hanoi, Harare, Ho Chi Minh, Hong Kong, Istanbul, Jakarta, Johannesburg, Karachi, Kiev, Kuala Lumpur, Lagos, Lima, Lisbon, London, London (Canada), Los Angeles, Madrid, Manila, Mexico City, Miami, Moscow, Mumbai, Munich, Nairobi, New York, Paris, Phnom Penh, Rio de Janeiro, Rome, San Francisco, Santiago, Sao Paulo, Seoul, Shanghai, Shenzhen, Singapore, Stockholm, Sydney, Tashkent, Tehran, Tokyo, Toronto, Warsaw, Washington. Table 2 shows these 61 cities by continent and it is interesting to notice that this sample contains all the continents.

Numbeo is the world's largest database of user contributed data about cities and countries worldwide. Numbeo provides current and timely information on world living conditions including cost of living, housing indicators, healthcare, traffic, crime and pollution. At the time of the collection there were 2,237,262 prices in 5,605 cities entered by 267,503 contributors and the information was updated at 2015-12-28.

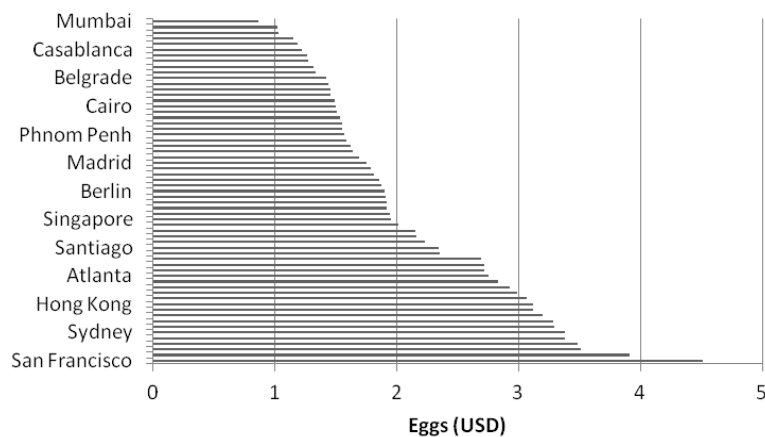
**Table 2** Sixty one cities by continent

<i>Continent</i>	<i>Cities</i>
Africa	6
Asia	19
Asia/Oceania	1
Europe	15
Europe/Asia	2
North America	11
Oceania	1
South America	6
<i>Total</i>	61

*Source:* Author

### 3 Data analysis

Before treating and working with the data of 2015, some dispersion measurements must be discussed. First of all, the correlation coefficient was calculated to each 49 variables. Numbeo website divides the cost of living data in 8 categories: restaurants, markets, transportation, utilities, sports and leisure, clothing and shoes, rent per month, apartment buying price, salaries and financing. Taking this split as effectual due to Numbeo's know-how, the variable which had the lowest correlation coefficient by each category was considered to analyse the data graphically. Taking the variable that has the lowest correlation coefficient can minimise the effects of collecting data by users, because there is subjectivity to people's choices, i.e., what they consider a cheap restaurant or a mid-range restaurant, for example.

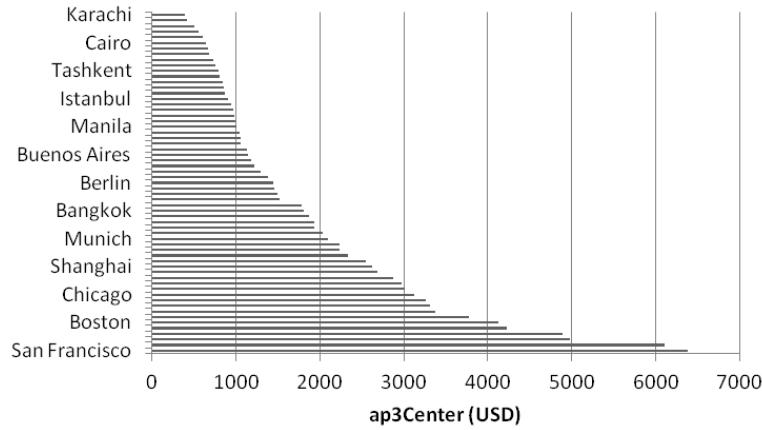
**Figure 1** 1 dozen of eggs in USD

*Source:* Numbeo website data

Figure 1 shows egg’s (lowest correlation coefficient in markets category) prices by city. San Francisco has the highest price (\$4.51) and Kiev has the second lowest price (\$1.02) of the 61 cities. Kiev is also one of the cheapest cities like Mumbai with lowest egg’s price (\$0.86) and 4th lowest McCombo’s price (\$3.03). It is interesting to notice that Sydney was the 5th most expensive city in the world according to EIUs summary (2015), having the 6th most expensive eggs (\$3.38). The highest price variation is between San Francisco and Los Angeles (1st most expensive and 2nd most expensive eggs) which is \$0.60.

Figure 2 shows apartment with three bedrooms (lowest correlation coefficient in rent per month category) prices by city. San Francisco has the highest price (\$6,383.73) and Karachi has the lowest price (\$394.42) of the 61 cities. It is interesting to notice that Hong Kong, Singapore, and Sydney are the three most expensive cities out of ten in the world according to Economist Intelligence Unit’s (2015) summary and these cities are also the three most expensive cities out of ten in this category.

**Figure 2** Rent an apartment with 3 bedrooms in city centre in USD



Source: Numbeo website data

Except squareMeterAp variable, which has a huge square deviation, in all variables analysed before, average value range does not contain the minimum value, it means that data is more scattered at high prices than at low prices.

To start a cluster analysis with 2015 data, some treatments must be done. First, missing values were estimated.<sup>1</sup> The second and last treatment before cluster analysis is

to standardise the data. All data was standardised by z-scores  $\left( \frac{x_i - \bar{x}}{\sigma} \right)$ .

By setting a hierarchical cluster using in these treated data, the agglomeration schedule shows how cities are grouped from 61 to 1 cluster as it can be seen at Table 4. It was used the between-groups linkage cluster method and Euclidean distance.

The coefficient’s value variation from 60 to 59 stages is higher than 59 to 58 stages, however the numbers seem to stabilise before 58 stages. With the goal to study more groups, it was considered the break of 59 to 58 stages as it was indicated at Table 3, which it means that three clusters will be created.

**Table 3** Agglomeration schedule, 2015 data

<i>Agglomeration schedule</i>						
<i>Stage</i>	<i>Cluster combined</i>		<i>Coefficients</i>	<i>Stage cluster first appears</i>		<i>Next stage</i>
	<i>CI1</i>	<i>CI2</i>		<i>CI1</i>	<i>CI2</i>	
1	32	35	2.017	0	0	3
...	...	...	...	...	...	...
46	21	51	7.229	0	0	51
47	5	15	7.697	41	0	48
48	2	5	7.892	45	47	54
49	22	41	8.016	42	26	51
50	3	14	8.079	43	0	53
51	21	22	8.138	46	49	55
52	26	28	8.425	44	0	56
53	3	12	8.662	50	0	54
54	2	3	8.887	48	53	56
55	21	38	9.020	51	0	58
56	2	26	9.610	54	52	57
57	1	2	9.924	0	56	58
58	1	21	10.986	57	55	59
59	1	25	12.585	58	0	60
60	1	8	15.936	59	0	0

*Source:* Numbeo website data, SPSS Software

Dendrogram supports the decision of three clusters considering the cutting line that could be done at Figure 3.

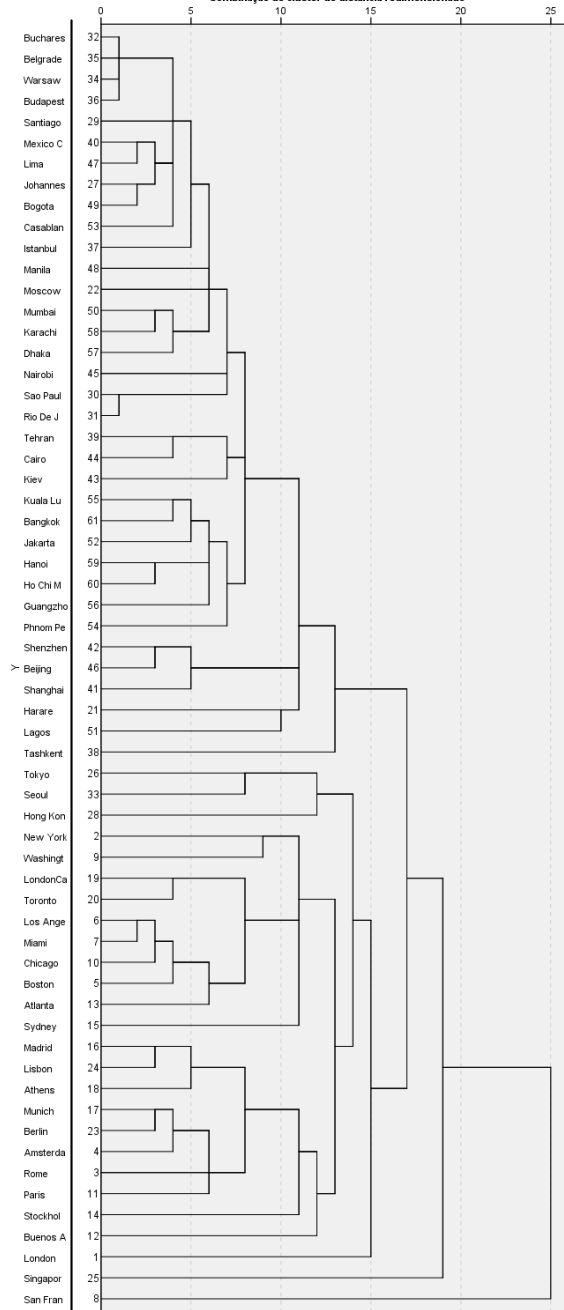
Next step was setting a k-means cluster analysis to create three clusters and the output is show at Table 4.

**Table 4** Number of cities in each cluster

<i>Cluster</i>	<i>Cities</i>
1	16
2	8
3	37
<i>Total</i>	61

*Source:* Numbeo website data, SPSS Software

Figure 3 Dendrogram using between-groups linkage



Source: Numbeo website data, SPSS Software



Analysing and understanding each cluster, cluster 1 contains Singapore, Sydney, Hong Kong, and Seoul, which are four of the top ten most expensive cities in the world according to Economist Intelligence Unit's (2015) summary. Cluster 3 contains Karachi, Mumbai, and Tehran, which are three of the top ten cheapest cities in the world according to Economist Intelligence Unit's (2015) summary. So, it brings us to this conclusion: cluster 1 could be classified as high living costs cities, cluster 2 as medium living costs cities, and cluster 3 as low living costs cities.

Each city was classified inside one cluster, an ordinal qualitative variable that will be called as cluster2015. In order to predict the future cluster of each city using current living costs, a multinomial logistic regression could be adjusted using 2014 data.

First, missing values were estimated<sup>2</sup> in 2014 data. After that, the outliers were treated using capping method according to Sinclair and Blackwell (2002). Capping method replaces all values that are below percentile 1 to percentile 1 value and replaces all values that are above percentile 99 to percentile 99 value. It helps to minimise the data dispersion to fit a model without losing any record.

As it is expected because of the nature of the costs of living variables, there are many variables with high correlation, what may cause a multicollinearity problem inside a predictive model. That is why all variables were analysed using a correlation matrix with Pearson's correlation. All correlations equal or above 0.7 were considered as a strong correlation and using previously knowledge about the variables, the one who could be more important to explain cluster2015 was preserved and the other ones were not considered to the study. It remained 23 variables.

To fit a multinomial logistic model to predict the cluster2015 variable using costs of living data of 2014, the absolute Pearson's correlation with the 23 variables to the cluster2015 was calculated and sorted from the highest to the lowest.<sup>3</sup> So, the model started to adjust by forward method respecting the variable's order by absolute correlation showed by Table 5.

**Table 5** Variables sorted by absolute Pearson's correlation with cluster2015 variable

<i>Variable</i>	<i>Absolute Pearson's correlation with cluster2015</i>
midMeal	0.689
domesticBeerRest	0.669
basic	0.645
apMeterCente	0.602
potato	0.597
bread	0.559
taxiStart	0.536
<i>mortgage</i>	<i>0.500</i>
<i>rice</i>	<i>0.454</i>
<i>gasoline</i>	<i>0.442</i>
jeans	0.409
<i>mobilePrepaid</i>	<i>0.334</i>
<i>cheese</i>	<i>0.330</i>
tennisRent	0.302

*Source:* Numbeo website data, SPSS Software

**Table 5** Variables sorted by absolute Pearson's correlation with cluster2015 variable (continued)

<i>Variable</i>	<i>Absolute Pearson's correlation with cluster2015</i>
<i>Nike</i>	<i>0.219</i>
<i>fitnessClub</i>	<i>0.183</i>
<i>domesticBeer</i>	<i>0.168</i>
<i>milk</i>	<i>0.103</i>
<i>dress</i>	<i>0.045</i>
<i>volkswagenGolf</i>	<i>0.040</i>
<i>Wine</i>	<i>0.027</i>
<i>internet</i>	<i>0.010</i>

Source: Numbeo website data, SPSS Software

A multinomial logistic model was fitted using stepwise method, Table 6 contains the output of the final model with Log likelihood = -21.04448. Except from the basic variable at cluster of medium costs of living, all the variables have p-values < 0.1, which means that these variables are significant to explain the cluster2015 variable considering 90% of confidence. Although the basic variable at cluster of medium costs of living had a p-value above 0.1, this variable was considered into the model because it had a p-value below 0.1 at cluster of high costs of living.

**Table 6** Final output multinomial logistic model where \*indicates p-value < 0.1, \*\*indicates p-value < 0.05, and \*\*\*indicates p-value < 0.01

<i>cluster2015</i>	<i>Coefficient</i>	<i>p-value</i>
<i>High (1)</i>		
domesticbeerrest	1.72528	0.007***
basic	0.04185	0.029**
potato	3.36033	0.063*
Constant	-16.59685	0.004***
<i>Medium (2)</i>		
domesticbeerrest	1.89751	0.006***
basic	0.01987	0.354
potato	4.46472	0.023**
Constant	-16.72709	0.007***
<i>Low (3)</i>		
(Base outcome)		

Source: Numbeo website data, Stata Software

Low category was considered as base outcome, so some relations can be discussed using odd ratios from Table 7. The chance of a city being at a high cluster in relation to being at a low cluster is multiplied by 5.6 when domestic beer price (0.5 litre at restaurant) is increased in \$1.00, ceteris paribus. In other words, the chance is 460% bigger. The chance of a city being at a high cluster in relation to being at a low cluster is multiplied

by 1.04 when basic costs are increased in \$1.00, *ceteris paribus*. Therefore, the chance is 4% bigger. The chance of a city being at a high cluster in relation to being at a low cluster is multiplied by 28.79 when potato price (1 kg) is increased in \$1.00, *ceteris paribus*. In other words, the chance is 27.79% bigger. The chance of a city to be at medium cluster in relation of be at low cluster is multiplied by 86.89 when potato price (1 kg) is increased in \$1.00, *ceteris paribus*. In other words, the chance is 85.89% bigger. So, it seems that potato price can say a lot about living costs when a cluster comparison is done.

**Table 7** Relative risk ratio (RRR) by cluster2015 categories

<i>cluster2015</i>	<i>RRR</i>
<i>High (1)</i>	
domesticBeerRest	5.61407
basic	1.04274
potato	28.79857
_cons	0.00000006
<i>Medium (2)</i>	
domesticBeerRest	6.66928
basic	1.02007
potato	86.89512
_cons	0.00000005
<i>Low (3)</i>	
(Base outcome)	

*Source:* Numbeo website data, Stata Software

In order to verify the accuracy of multinomial logistic model, Table 8 shows % of correct classified cases by cluster and calculates the model's overall efficiency. With 84% of correct classified cases, there is evidence that this could be a reasonable predictive model. It is relevant to explain that to choose which cluster the model would classify a city, the biggest probability was considered.

**Table 8** Model's overall efficiency

<i>Estimated/real</i>	<i>High</i>	<i>Medium</i>	<i>Low</i>	<i>% correct classified</i>
High	14	0	2	88%
Medium	5	2	1	25%
Low	1	1	35	95%
<i>Model's overall efficiency</i>				84%

*Source:* Numbeo website data, Stata Software

Ultimately, the final equations of probability's models of each cluster classification to predict the future living costs cluster of a city using current data are shown inside appendix.

#### 4 Results

This paper showed a brief quantitative analysis about costs of living using cluster analysis and multinomial logistic regression. The results show an association between livable index and living costs.

This study corroborates with Economist Intelligence Unit’s (2015) summary by identifying five of the ten most expensive cities described at the summary (Singapore, Paris, Sydney, Hong Kong, Seoul) and three of the ten cheapest cities (Karachi, Mumbai, Tehran). Besides, 61 cities were classified into three clusters which are presented at Table 9.

**Table 9** City by cluster (high, medium, and low living costs)

<i>City</i>	<i>Cluster (living costs)</i>
Amsterdam	High
Berlin	High
Buenos Aires	High
Hong Kong	High
London	High
London – Canada	High
Madrid	High
Munich	High
Paris	High
Rome	High
Seoul	High
Singapore	High
Stockholm	High
Sydney	High
Tokyo	High
Toronto	High
Atlanta	Medium
Boston	Medium
Chicago	Medium
Los Angeles	Medium
Miami	Medium
New York	Medium
San Francisco	Medium
Washington	Medium
Athens	Low
Bangkok	Low
Beijing	Low
Belgrade	Low

*Source:* Numbeo website data, SPSS Software

**Table 9** City by cluster (high, medium, and low living costs) (continued)

<i>City</i>	<i>Cluster (living costs)</i>
Bogota	Low
Bucharest	Low
Budapest	Low
Cairo	Low
Casablanca	Low
Dhaka	Low
Guangzhou	Low
Hanoi	Low
Harare	Low
Ho Chi Minh City	Low
Istanbul	Low
Jakarta	Low
Johannesburg	Low
Karachi	Low
Kiev	Low
Kuala Lumpur	Low
Lagos	Low
Lima	Low
Lisbon	Low
Manila	Low
Mexico City	Low
Moscow	Low
Mumbai	Low
Nairobi	Low
Phnom Penh	Low
Rio de Janeiro	Low
Santiago	Low
Sao Paulo	Low
Shanghai	Low
Shenzhen	Low
Tashkent	Low
Tehran	Low
Warsaw	Low

*Source:* Numbeo website data, SPSS Software

Analysing these clusters, the chance of a city be considered as higher living costs has a big correlation with the potato price.

Is possible by using living costs of current year to predict which cluster a city will be in the next year (low, medium, high). This model could help companies or even people to

decide which city to move to be able to decrease living costs. It can be important to avoid a wrong decision if a cluster change is coming in a city.

It could be useful to help describe the economy situation of each city by combining another piece of information, like politics. It could also help investors to identify opportunities, and it could be used by government to see a crisis coming and take steps to avoid serious consequences.

## References

- Alonso, W. (1964) *Location and Land Use: Toward a General Theory of Land Rent*, Cambridge, Harvard University Press, Cambridge, MA.
- Duranton, G. and Puga, D. (2013) *The Growth of Cities*, CEPR Discussion paper, No. DP9590.
- Economist Intelligence Unit (2012) *Best Cities Ranking and Report*, A special report from the Economist Intelligence Unit [online] [http://pages.eiu.com/rs/eiu2/images/EIU\\_BestCities.pdf](http://pages.eiu.com/rs/eiu2/images/EIU_BestCities.pdf) (accessed 6 January 2016).
- Economist Intelligence Unit (2015) *Worldwide Cost of Living Report 2015*, A summary from the Economist Intelligence Unit [online] [http://www.eiu.com/public/topical\\_report.aspx?campaignid=WCOL2015](http://www.eiu.com/public/topical_report.aspx?campaignid=WCOL2015) (accessed 6 January 2016).
- Glaeser, E.L. (2008) *The Economic Approach to Cities*, Harvard Institute of Economic Research Discussion paper, No. 2149; KSG Working paper, No. RWP08-003.
- Kato, S., Nagamine, S., Nishihara, A., Yokota, F. and Matsuno, N. (1949) 'Investigations on the food consumption and the cost of living of certain Salarymen's families in Tokyo', *The Japanese Society of Nutrition and Dietetics*, Vol. 7, No. 2, pp.75–80.
- Lenzen, M. (1998) 'Energy and greenhouse gas cost of living for Australia during 1993/94', *Energy*, Vol. 23, No. 6, pp.497–516, Elsevier.
- Mills, E.S. (1967) 'An aggregative model of resource allocation in a metropolitan area', *American Economic Review Papers and Proceedings*, Vol. 57, No. 2, pp.197–210.
- Muth, R.F. (1969) *Cities and Housing*, University of Chicago Press, Chicago, IL.
- Pollak, R. (1981) 'The social cost of living index', *Journal of Public Economics*, Vol. 15, No. 3, pp.311–336.
- Porter-Hudak, S. (1991) 'A numerical methods approach to calculating cost-of-living indices', *Journal of Econometrics*, Vol. 50, Nos. 1–2, pp.91–105.
- Sinclair, A. and Blackwell, G.H. (2002) *Applied Mineral Inventory Estimation*, pp.167–179, Cambridge University Press, UK.
- The Economist* (2015) 'Bright lights, big cities: urbanisation and the rise of the megacity' [online] <http://www.economist.com/node/21642053> (accessed 16 February 2016).
- Wolman, L. (1929) *Recent Economic Changes in the United States*, pp.13–78, National Bureau of Economic Research, USA.

**Notes**

- 1 It was considered the average of series of each variable to estimate missing values. There were 15 missing values in 2015 data.
- 2 It was considered the average of series of each variable to estimate missing values. There were 23 missing values in 2014 data.
- 3 Some variables cause a convergence problem to the model because the matrix becomes not invertible, these variables were not considered and it is indicated in italic font at Table 5.

**Appendix**

Final equations of probability's models of each cluster classification to predict the future living costs cluster of a city using current data:

- Probability of a city with *low living costs*:

$$p_{i_{low}} = \frac{1}{1 + e^{(-16.6+1.7*domesticBeerRest_i+0.04*basic_i+3.4*potato_i)} + e^{(-16.6+1.9*domesticBeerRest_i+0.02*basic_i+4.5*potato_i)}}$$

- Probability of a city with *medium living costs*:

$$p_{i_{medium}} = \frac{e^{(-16.7+1.9*domesticBeerRest_i+0.02*basic_i+4.5*potato_i)}}{1 + e^{(-16.6+1.7*domesticBeerRest_i+0.04*basic_i+3.4*potato_i)} + e^{(-16.7+1.9*domesticBeerRest_i+0.02*basic_i+4.5*potato_i)}}$$

- Probability of a city with *high living costs*:

$$p_{i_{high}} = \frac{e^{(-16.6+1.7*domesticBeerRest_i+0.04*basic_i+3.4*potato_i)}}{1 + e^{(-16.6+1.7*domesticBeerRest_i+0.04*basic_i+3.4*potato_i)} + e^{(-16.7+1.9*domesticBeerRest_i+0.02*basic_i+4.5*potato_i)}}$$