Regression by principal components to identify the impact factors on life expectancy in the ranking of cities with the best rates of healthy habits

Lucas Dutra Mendes[1]*; Juliano Domingues da Silva2

1Optel Group. Bilingual Technical Support - Product Specialist. Rua James Clerk Maxwel, 280, Módulo 08 – Techno Park; 13069-380 Campinas, São Paulo, Brazil 2 State University of Maringá. Assistant teacher. Avenida Colombo, 5790 – Jardim Universitário; 87020-900 Maringá, Paraná, Brazil *corresponding author: lucas.dutra.mendes@gmail.com

Regression by principal components to identify the impact factors on life expectancy in the ranking of cities with the best rates of healthy habits

Summary

The present work sought to study the cities that have the best rates of healthy lifestyle habits, with the aim of understanding the factors that directly impact the life expectancy of a population, and the database studied was created by the Lenstore team and contains forty-four cities around the world, ranked with the best healthy lifestyle indexes. The analysis showed that the overload of daily life and healthy habits have a predominant factor in the population's life expectancy. Among the models used are; Linear Regression, Principal Component Analysis [PCA] and Principal Component Regression [PCR]. The main results found in the research are that everyday overload, generated by excessive work and pollution, negatively affects life expectancy while healthy habits positively affect life expectancy. This research helps to point out which factors public agents should prioritize to improve the population's quality of life.

Keywords: PCA; PCR; Box-Cox; Linear Regression; Life expectancy

Introduction

Nowadays, there is a lot of talk about the quality of life present in the world we live in, meaning that this topic can be assimilated differently by each person. us. Having a healthy quality of life or lifestyle can be referred to as a tree, containing several branches and each of these having its importance within this theme, which may be in the social, biological, economic, political, human, well-being, among others (Almeida et al., 2012). The perspective of life according to Minayo et al. (2000) can be understood in a way by elementary conditions, such as access to needs basic needs, such as drinking water, housing, health, leisure and food. According to Fonseca et al. (2019), chronic diseases linked to unhealthy habits, including cancer, diabetes and hypertension, which in Brazil alone accounts for 72% of the causes of death and observed the lower class as the most affected. We therefore realize the importance of knowing and study such behaviors, especially to understand the behaviors significant factors that define the quality of life and life expectancy of a population.

The database to be studied in this work (to see https://www.lenstore.co.uk/research/healthy-lifestyle-report/), it has to the following observations of forty-four cities, listed as the best in lifestyle healthy, being: hours of sunshine that each city receives per year, cost of a bottle of 300ml water, obesity index of the country each city is part of, life expectancy of the country, pollution that each city produces, annual average of hours worked in the country, index of happiness by country and outdoor activities in cities, number of restaurants that They provide meal delivery and monthly gym membership costs.

However, note the importance of studying and understanding the factors that explain the life expectancy of a population, especially in cities with the best rates of healthy habits, this way it will be possible to identify the preponderant factors, which in turn can be implemented through public policies in cities outside the ranking studied.

Highlighting the importance of technology, combined with data capture, mathematics and ability to analyze and process these, so that patterns can be studied, enabling analyzes to be carried out, assisting in decision making.
Guiding this work to the following questions: What are the important and common factors among the best ranked cities and which variables directly impact expectations life of a population? Therefore, the objective of this research is to analyze the impact of quality of life indicators of the best cities to live in, constant in the “healthy lifestyle cities report 2021” on the population’s quality of life.

Material and methods

The methodology of this work consists of the qualitative and quantitative analysis of the “healthy lifestyle cities report 2021” database created by the Lenstore team where forty-four cities around the world were analyzed and ranked with the best indices of healthy lifestyle. Data preparation through data wrangling, which is the art of import the data into R and make it ready to be visualized and modeled (Wickham and Grolemund, 2017).

Following the methodology of this work, Multiple Linear Regression was also used with the objective of understanding the variables that explain life expectancy, the Shapiro-Francia test to analyze the adherence of residuals to normality, the variance inflation value [VIF] test, to check the presence of multicollinearity, Breusch-Pagan test to identify the presence of heteroscedasticity, normalization of the variable Y by Box-Cox to adapt the residuals to normality and treat the phenomenon of heteroscedasticity in the regression model, Principal Component Analysis [PCA], with the objective of treating the phenomenon of multicollinearity, grouping the variables and capturing non-latent observations in the studied base, and, finally , Principal Component Regression [PCR] with the premise of obtaining the preponderant factors obtained in the PCA and that explain life expectancy.

Table 1 illustrates the variables to be studied in the database, with a contextualization of each variable and the way in which each of them is measured.

Table 1. Concept of variables used in the research.

Variable	Contextualization	Measure	Source
Hours of sunshine	Amount of sun exposure is linked to physical and mental health and disease prevention.	Average hours of sun exposure in the city in a year.	https://www.pucrs.br/blog/efeitos-e-beneficios-da-exposicao-luz-solar-para-imunidade/
Water bottle cost	Access to water, being a basic need, directly impacts the health of a population. The less access, the greater the health problems.	£/300ml	There is no source described. https://www.numbeo.com/cost-of-living/
Obesity index	The higher the obesity levels, the lower the healthy habits and the decrease in life expectancy Life expectancy of the	Percentage by country.	https://ourworldindata.org/obesity
Life expectancy	population of each country.	Years of life by country.	https://shorturl.ae/BbMVp
Pollution index	High pollution levels negatively impact the life of a community.	in Pollution index by city, the higher, the worse.	https://www.numbeo.com/pollution/rankings.jsp
Annual working hours	The more hours worked, time for leisure and any less physical activities a person has, the more people have an on their impact quality of life.	Hours year by city.	https://data.oecd.org/emp/hours-worked.htm
happiness index	The higher the index, the happier the population The is in the country they live in.	index the higher the better, by country	https://worldhappiness.report/ed/2021/
Outdoor activities by city.	Outdoor activities provided by the cities observed.	Number of activities provided per city.	https://www.lawnstarter.com/blog/studies/best-cities-for-spring/
Food for delivery	Restaurants available for food delivery, showing the ease of access to food.	Number of restaurants per city	There is no source described.
Academy Price	Accessibility of The physical activities in closed spaces.	Monthly £ per city	There is no source described.

Source: Original data obtained in the research.

Results and discussion

Data Wrangling

As the observed variables are in English and in a not very user-friendly format of reading and understanding, organizing the data is the first step to be followed. With

aid of the R language, the variable names were renamed into Portuguese and were Shorter names were used to facilitate understanding and analysis. A point to note are the variables “Water” and “Gym”, where both have the symbol “£” before each observation. This symbol has been removed and these variables have been converted to dollars American. The obesity variable has the symbol % before each observation and this has also been removed. Moving forward with data manipulation, the next step is filling in missing values, also known as missing values. First value to be filled in is in Fukuoka city for the pollution variable. This data cannot be found in the values are concentrated in the variable hours worked per year, for the cities: Johannesburg database references and, therefore, data from the city of Osaka, which is approximately 600 kilometers away from Fukuoka and It has a pollution index of 53.24. The sunshine hours variable has missing data for the city of Geneva, in which the value of 1887 hours was obtained from the database references per year of sunshine. The remaining missing, 2189 hours - São Paulo, 1706 hours - Shanghai and Beijing, 2168 hours - Hong Kong, 2148 hours - Mumbai, 2122 hours - Taipei, 2085 hours - Jakarta, 2018 hours - Buenos Aires, 1606 hours – Bangkok, 2090 hours. All data were found in the database references with the exception of annual hours worked in the city of Cairo, where data from the city of Johannesburg was used, being 2189 hours per year. A Figure 1 below illustrates the database after applying Data Wrangling.

	Posicao	Cidades	Sol	Agua	Obesidade	Vida	Poluicao	Trabalhadas	Felicidade	Atividades	Restaurantes	Academia
1	1	Amsterdam	1858	2.50	20.4	81.2	30.93	1434	7.44	422	1048	45.37
2	2	Sydney	2636	1.92	29.0	82.1	26.86	1712	7.22	406	1103	54.16
3	3	Vienna	1884	2.52	20.1	81.0	17.33	1501	7.29	132	1008	33.46
4	4	Stockholm	1821	2.24	20.6	81.8	19.63	1452	7.35	129	598	48.50
5	5	Copenhagen	1630	2.85	19.7	79.8	21.24	1380	7.64	154	523	42.29
6	6	Helsinki	1662	2.08	22.2	80.4	13.08	1540	7.80	113	309	45.80
7	7	Fukuoka	2769	1.01	4.3	83.2	53.24	1644	5.87	35	539	72.63
8	8	Berlin	1626	2.02	22.3	80.6	39.41	1386	7.07	254	1729	33.94
9	9	Barcelona	2591	1.55	23.8	82.2	65.19	1686	6.40	585	2344	49.14
10	10	Vancouver	1938	1.40	29.4	81.7	24.26	1670	7.23	218	788	40.35
11	11	Melbourne	2363	2.04	29.0	82.1	25.90	1712	7.22	243	813	47.96
12	12	Beijing	2671	0.34	6.2	75.4	85.43	2168	5.12	223	261	50.21

Figure 1. Database after changes
Source: Original data obtained in the research

Multiple Linear Regression

The analysis begins with the Pearson correlation map between the predictor variables, where the variable life expectancy, to be studied and, therefore, removed from the map of correlations. Figure 2 illustrates the correlation map between the predictor variables where

we can observe strong correlations in modulus above 0.5 between multiple variables, being is a strong characteristic of the presence of multicollinearity.

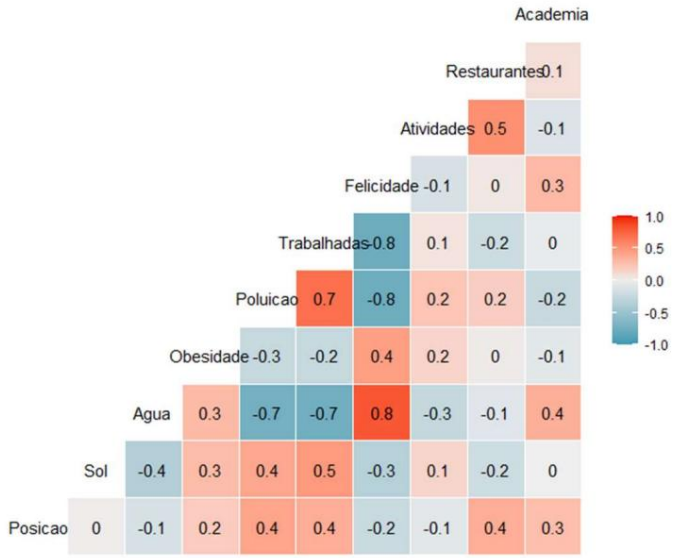

Figure 2. Correlation matrix
Source: Original results obtained in the research

Generating the multiple linear regression model with the life expectancy variable as Y (dependent or response) and the others as X variables (independent or predictors), the following output was obtained, illustrated in Figure 3.

```
Residuals:
    Min      1Q  Median      3Q     Max
-10.223  -1.919   1.146   1.905   5.271

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  77.6525192 12.7285857   6.101 6.38e-07 ***
Sol          -0.0005499  0.0012888  -0.427  0.67230
Agua         -1.4237850  1.1542117  -1.234  0.22583
Obesidade    -0.0889106  0.0720917  -1.233  0.22592
Poluicao      0.0510994  0.0435259   1.174  0.24855
Trabalhadas  -0.0129272  0.0047077  -2.746  0.00957 **
Felicidade    3.1876377  1.1782964   2.705  0.01059 *
Atividades    0.0057991  0.0050368   1.151  0.25763
Restaurantes -0.0007419  0.0005729  -1.295  0.20409
Academia      0.1028749  0.0365539   2.814  0.00807 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.267 on 34 degrees of freedom
Multiple R-squared:  0.7001,    Adjusted R-squared:  0.6207
F-statistic:  8.82 on 9 and 34 DF,  p-value: 1.073e-06
```

Figure 3. Summary Multiple Linear Regression Model.
Source: Original results obtained from the research

The model summary shows that the p-value of the F test distribution is statistically different from 0 to 95% confidence, therefore being statistically significant to predict the phenomenon studied. Among the variables studied, only the intercept, hours worked, Happiness Index and Gym were statistically significant.

Figure 4 presents the test to verify the adherence of residues to normality of Shapiro-Francia, where to confirm the null hypothesis H0, and, therefore, the adherence of given normality, the p-value must be greater than 0.05, however, the p-value is less than this value, indicating that the distribution of the data does not follow a normal distribution.

```
        Shapiro-Francia normality test

data:  OLS_Vida$residuals
W = 0.9182, p-value = 0.005673
```

Figure 4. Shapiro-Francia Normality Test.
Source: Original results obtained from the research

Multiple Nonlinear Regression

To treat residuals that did not adhere to normality, the variable Y was normalized through the BOX-COX transformation and the lambda value generated for the BOX-COX model is 11.72205. The summary of this model, presented in Figure 5, shows an insignificant improvement in adjusted R^2, only two statistically significant variables and the verification of residue adherence to Shapiro-Francia normality, Figure 6, indicates that the data adheres to normality, after normalization of the Y variable.

```
Residuals:
       Min          1Q      Median          3Q         Max
-7.773e+20 -3.287e+20 -3.284e+19  3.227e+20  8.519e+20

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.221e+21  1.816e+21   1.223   0.2299
Sol           2.373e+17  1.839e+17   1.290   0.2056
Agua          4.276e+19  1.647e+20   0.260   0.7967
Obesidade    -2.172e+19  1.029e+19  -2.111   0.0422 *
Poluicao     -3.769e+18  6.211e+18  -0.607   0.5479
Trabalhadas  -1.555e+18  6.717e+17  -2.315   0.0268 *
Felicidade    2.362e+20  1.681e+20   1.405   0.1691
Atividades    4.871e+17  7.187e+17   0.678   0.5025
Restaurantes  1.295e+16  8.175e+16   0.158   0.8750
Academia      1.055e+19  5.216e+18   2.022   0.0511 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.661e+20 on 34 degrees of freedom
Multiple R-squared:  0.7073,    Adjusted R-squared:  0.6298
F-statistic: 9.129 on 9 and 34 DF,  p-value: 7.348e-07
```

Figure 5. Summary BOX-COX model.
Source: Original results obtained from the research

```
        Shapiro-Francia normality test

data:  OLS_BC_vida$residuals
W = 0.98388, p-value = 0.6962
```

Figure 6. Shapiro-Francia normality test.
Source: Original results obtained from the research

Multicollinearity

The VIF (Variance Inflation Factor) test, which can be defined by VIF = ———— , where Tolerance = 1 ÿ R², used to diagnose the presence of multicollinearity in the model, and the output of this test is the same for the linear model and the BOX-COX model, as there was no transformation in the variables X. According to (Fávero and Belfiore , 2017), many authors state that VIF above ten indicates multicollinerity problems, however, a VIF equal to or greater than four already represents a high value for this test. When checking Figure 7, high VIF can be seen, close to or greater than four for four predictor variables, therefore indicating the presence of a strong correlation between them.

	Variables	VIF
1	Sol	2.124419
2	Agua	4.685874
3	Obesidade	2.177049
4	Poluicao	3.562757
5	Trabalhadas	5.340954
6	Felicidade	5.496669
7	Atividades	1.653832
8	Restaurantes	2.551331
9	Academia	2.049322

Figure 7. VIF Linear Model & BOX-COX.
Source: Original results obtained from the research

Heteroscedasticity

Figure 8 shows the output of the Breusch-Pagan test, used to verify the presence of heteroscedasticity in the models studied so far. This test evaluates the correlation between the predictor variables in relation to the models' error terms. The null hypothesis, H0, for p-values below 0.05 indicates the presence of heteroscedasticity, since the alternative hypothesis H1, for values above 0.05 indicates that the model is not heteroscedastic. By comparing both outputs, we can see that the linear model

has a strong presence of heteroscedasticity, as the probability value > Chi2 is lower to 0.05. In turn, the non-linear model, normalizing Y by BOX-COX indicated improvement in this test, therefore nullifying the presence of heteroscedasticity.

```
        Test Summary                        Test Summary
-------------------------------     -------------------------------
DF            =     1               DF            =     1
Chi2          =     16.6139         Chi2          =     0.05204538
Prob > Chi2   =     4.581403e-05    Prob > Chi2   =     0.8195416
```

Figure 8. Breusch-Pagan Linear and Nonlinear Model.
Source: Original results obtained from the research

Principal Component Factor Analysis [PCA]

This is an exploratory multivariate technique, with the aim of treating the multicollinearity present in the base and obtain new observations that capture the behavior set of original variables (Fávero and Belfiore, 2017). After creating the correlation matrix and applying Bartlett's sphericity test, which returned a p-value of 1.283726e-88, it is possible to conclude that the matrix is factorable, validating the application of this technique.

Figure 9 displays the summary of the PCA model generated, which adopted the latent roots method, and kept the main components with standard deviation values greater than one, to obtain the factors. It is important to highlight that 86.23% of the total base variance was captured with these four components.

```
Importance of components:
                          PC1    PC2    PC3    PC4     PC5     PC6     PC7     PC8     PC9
Standard deviation     1.9307 1.2747 1.1674 1.0249 0.66653 0.56561 0.47993 0.37915 0.30990
Proportion of Variance 0.4142 0.1805 0.1514 0.1167 0.04936 0.03555 0.02559 0.01597 0.01067
Cumulative Proportion  0.4142 0.5947 0.7461 0.8629 0.91222 0.94776 0.97336 0.98933 1.00000
```

Figure 9. Summary PCA Model.
Source: Original results obtained from the research

Below, Figures 10 and 11 show the weights that each main component capture in each variable.

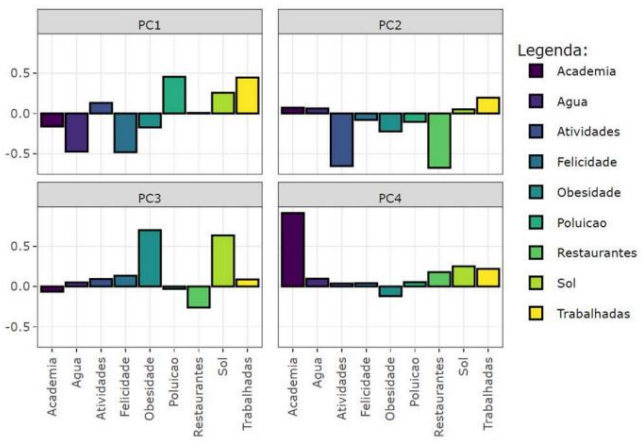

Figure 10. Weights captured by the main components.
Source: Original results obtained from the research

	PC1	PC2	PC3	PC4
Sol	0.25632378	0.05173649	0.63367667	0.24974038
Agua	-0.47373063	0.06177191	0.04994505	0.09687390
Obesidade	-0.17383443	-0.22367118	0.69886180	-0.11965068
Poluicao	0.45472310	-0.10514722	-0.02923054	0.05284818
Trabalhadas	0.44639406	0.19619218	0.08742518	0.21799024
Felicidade	-0.48013600	-0.08016724	0.13317232	0.04025332
Atividades	0.13039188	-0.65382201	0.09315924	0.03697810
Restaurantes	0.00657942	-0.67453001	-0.26154649	0.17797806
Academia	-0.16137640	0.07092758	-0.06506842	0.91046819

Figure 11. Weights captured by the main components.
Source: Original results obtained from the research

When analyzing the first main component, it is possible to observe that it negatively captures the variables Water, referring to the cost of a 300 ml bottle of water and Happiness, referring to the happiness index of the studied population, positively capturing large portion of the variables Worked, this being the annual average of hours worked and Pollution, referring to the pollution index that each city presents, and Sun referring to the average of annual hours of sunlight in each city. Taking these observations into account, concluded that the greater the number of hours worked, the pollution increases, the likely exposure to the sun, the happiness index decreases and access to water becomes easier due to the cost, however this observation may indicate a difficulty in drinking water during the day, due to the long journeys. Therefore, when obtaining the main factors, this will be named "Overload_Cotidiano".

The second main component, in turn, negatively and very strongly captures the variables, Activities, referring to the outdoor activities available in cities and Restaurants, which refers to the number of delivery options. As the main weights of this component are negative, multiplying it by -1 is appropriate so that the four captured components are vectorially in the same direction, illustrated in figure 12.
Therefore, this main component can be named “Leisure”, representing the leisure offered by each city to its inhabitants.

The third main component, in turn, vigorously and positively captures the Obesity and Sun variable. For this reason, this main component was named of “Sedentarism”.

The analysis of the fourth, and last main component, shows that only the Variability of the Academy variable is captured in a significant way, with the Restaurants, Sun and Work variables being positive and discreetly captured.
Therefore, this main component will be called “Habitos_Saudaveis”.

	PC2
Sol	-0.05173649
Agua	-0.06177191
Obesidade	0.22367118
Poluicao	0.10514722
Trabalhadas	-0.19619218
Felicidade	0.08016724
Atividades	0.65382201
Restaurantes	0.67453001
Academia	-0.07092758

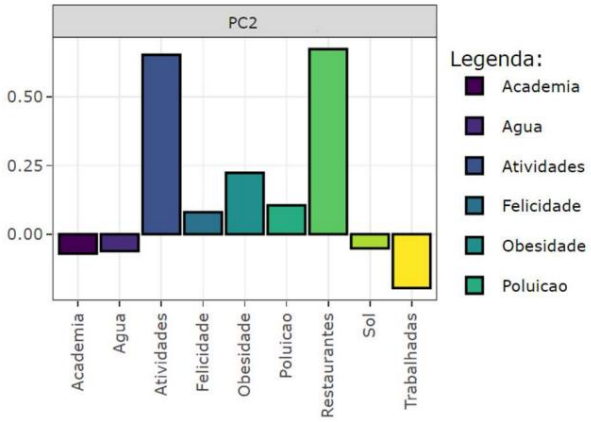

Figure 12. Result of multiplying Principal Component 2 by -1.
Source: Original results obtained from the research

Principal Component Regression [PCR] and Y Box-Cox Normalization

The PCR model was created with the new variables obtained through the PCA technique, including at this point the Life variable to be predicted, however, normalized by Box-Cox. The summary of this model shows that the Overload_daily variable is the most significant in explaining life expectancy, followed by Healthy_Habits. The other variables were not statistically significant at 95% confidence to explain the Life expectancy. Below is Figure 13 illustrating the model summary.

```
Residuals:
       Min         1Q     Median         3Q        Max
-9.031e+20 -3.811e+20 -5.710e+19  3.056e+20  1.169e+21

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)           1.600e+21  7.487e+19  21.365  < 2e-16 ***
Sobrecarga_Cotidiano -5.606e+20  7.573e+19  -7.402 6.06e-09 ***
Lazer                 8.154e+19  7.573e+19   1.077   0.2882
Sedentarismo         -9.345e+19  7.573e+19  -1.234   0.2246
Habitos_Saudaveis     1.833e+20  7.573e+19   2.420   0.0203 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.966e+20 on 39 degrees of freedom
Multiple R-squared:  0.6189,    Adjusted R-squared:  0.5798
F-statistic: 15.83 on 4 and 39 DF,  p-value: 8.857e-08
```

Figure 13. Summary PCR + Box-Cox Y model.
Source: Original results obtained from the research

The Shapiro-Francia test, Figure 14, indicates the adherence of the error terms to normality, while the Breusch-Pagan test, Figure 15, shows that there is no heteroscedasticity present in the model and the VIF test, Figure 16, indicates that there is no multicollinearity, this phenomenon being overcome by applying the PCA model, making the new observations orthogonal to each other.

```
        Shapiro-Francia normality test

data:  OLS_BC_PCR$residuals
W = 0.96847, p-value = 0.2284
```

Figure 14. Shapiro-Francia normality test.
Source: Original results obtained from the research

```
        Test Summary
-----------------------------
DF            =     1
Chi2          =     0.6339087
Prob > Chi2   =     0.4259252
```

Figure 15. Breusch-Pagan test.
Source: Original results obtained from the research

	variables	VIF
1	Sobrecarga_Cotidiano	1
2	Lazer	1
3	Sedentarismo	1
4	Habitos_Saudaveis	1

Figure 16. VIF test.
Source: Original results obtained from the research

The results of this research show that the main factors that explain the quality of life index of the cities listed in the "healthy lifestyle cities report 2021" are daily overload and healthy habits, while leisure and sedentary lifestyle do not have significant explanatory power. According to Almeida et al. (2012), quality of life can have factors explained by aspects of the social, biological, economic, political, human, well-being, among others. As the scope of this research seeks to evaluate only cities that already have a higher quality of life status, it was noted that biological and social aspects such as sedentary lifestyle and leisure are not problematic factors in explaining the quality of life in the population. However, the results of this research do not provide evidence that these factors are not important for other cities, considered to have a more varied quality of life status. Therefore, future studies can explore an analysis to evaluate whether the explanatory factors of life expectancy may differ between cities with different quality of life status.

Furthermore, Minayo et al. (2000) and Fonseca et al. (2019) argue that basic conditions, such as access to basic needs, such as drinking water and healthy lifestyle habits, are essential for increasing the population's life expectancy. You results of this research corroborate this perspective, by showing the effects important aspects of everyday overload (composed of high levels of pollution and excess work and low levels of happiness) to harm the life expectancy index, while also highlighting the positive effect of healthy habits, such as dedicating hours of the day to go to gyms.

Conclusion

This research sought to study and understand the factors that explain and determine life expectancy, especially guiding public policies in unlisted cities, allowing the application of short and long-term measures with the aim of increasing quality and life expectancy of its population.

The analysis showed, through the PCR model, that the observation Sobrecarga_quotidiano, proved to be negative and statistically significant in explaining life expectancy, or In other words, the longer the hours worked, the more pollution is generated, the access to water and the of happiness decrease drastically, and consequently reducing life expectancy.

In turn, the Habitos_Saudaveis observation, taken from the PCA model and which is showed positive and statistically significant to explain the longevity of a population through the PCR model, explained the importance of physical activity in life expectancy, especially in closed environments such as gyms and gyms.

Therefore, investing in and encouraging more flexible working hours or even even reduced, reduction in travel time between home and work, reduction in pollution levels generated, facilitating access to water which can also be linked to basic sanitation, combining this with the practice of physical activities in gyms and gyms, can be public policies to be implemented in regions with the aim of increasing quality of life and consequently the life expectancy of the population.

References

Almeida, MAB, Gutierrez, GL, Marques, R. Quality of Life. Each USP. São Paulo.

Fávero, LP, & Belfiore, P. 2017. Data analysis manual: statistics and multivariate modeling with Excel®, SPSS® and Stata®. Elsevier Brazil.

Fonseca, CD, Ting, MLB, & Sarti, FM. Evolution of lifestyle indicators and potential influence of public policies on the health of the Brazilian population: 2006-2017. São Paulo, 2019.

Lenstore. Healthy lifestyle cities report 2021. Available at: https://www.lenstore.co.uk/-research/healthy-lifestyle-report/

Minayo, MCDS, Hartz, ZMDA, Buss, PM Quality of life and health: a necessary debate. Ciência & Saúde Coletiva, 5, 7-18, 2000.

Wickham, H., & Grolemund, G. (2017). R for data science: Import. Tidy, transform, visualize, and model data, 1.