Wikipedia Project Edition. 10 Languages. 2001-2020

Wilson Alexander Pinzón Rueda
Facultad Tecnológica
Universidad Distrital "Francisco José de Caldas"
Bogotá, Colombia
wapinzon@udistrital.edu.co, wapinzon@gmail.com

Carlos Andrés Arango Londoño
Departamento de Ingenieria Industrial
Universidad de LaSalle
Bogotá, Colombia
cararango@unisalle.edu.co

Jose Martin Díaz
Facultad de Ciencias Empresariales
Corporación Universitaria Minuto de Dios
Bogotá, Colombia
jdiazpulido@uniminuto.edu.co

Abstract

This text identifies the differences and similarities on temporal and recurrent cyclicity, by the editing activity for 10 Wikipedia Projects and the total for all 243 Wikipedia Projects, based on the monthly editorial activity from 2001 to 2020 in: Chinese, German, English, French, Italian, Japanese, Polish, Portuguese, Russian and Spanish; plus: total for this selected projects. This dataset was used to obtain linear regression model coefficients based on categorical lineal regression and Analysis Of Variance, ANOVA, to estimate coefficients by: language, year, moth of year. After this was used the Pearson Correlation Coefficients between linear regression model coefficients by language. We find differences in cyclicity, tendency for each language over this 20 years of edition

Keywords
Wikipedia, Digital Knowledge, Categorical Linear Regression, Cyclicity, Time Series

1. Introduction
The edition of public digital humankind knowledge have at least a public data source of detailed, historical, affordable and available information for various regions, languages and cultures. The Wiki Stats records the Wiki Media Projects activity, here we use the data about 10 Wikipedia Projects edition, one by language: Chinese, German, English, French, Italian, Japanese, Polish, Portuguese, Russian and Spanish; plus: total for this selected projects, and all 243 Wikipedia Projects Total. The edition activity requires the attention of dedicated and specialized authors as difference to the attention of a reader. Here was used monthly edition datasets since 2001-01-01 to 2021-02-28. This dataset about digital, online, public, and worldwide knowledge edition activity was used to obtain linear regression model coefficients based on categorical lineal regression and Analysis Of Variance, ANOVA, to estimate coefficients by: language, year, month of year. After this was used the Pearson Correlation Coefficients between linear regression model coefficients by language. With this procedure we identify the coevolution for 10 Wikipedia Projects edition; it could be used to identify the correlation ship between these digital online activities with other activities of the society

1.1 Objectives
Identify the differences and similarities in ciclicity between 10 Wikipedia Projects edition from 2001 to 2020.

2. Literature Review

The use and editing of Wikipedia pages indicates academic and technical activity (Toribio et al., 2020), it also implies political activism activity (Margolin et al., 2016); and also could be used to indicate the travel interest by a city, (Quian & Elías, 2017). Although the presence and use for editing and creating content could distort this measurements (Zheng et al., 2019). Kikkawa et al. (Kikkawa et al., 2020) identify that either through machines or human beings editorial activity and Wikipedia is linked to academic activity (Spasojevic et al., 2019) and the individual interests of those who make up society on issues and predictable amounts (Zhang et al., 2016). Lemmerich et al (Lemmerich et al., 2019) find the motivations for use Wikipedia Projects, they find different motivations between different speakers of different languages; they identify intrinsic knowledge as a common motivator as the main motivator. Ban et al. (Ban et al., 2017) identifies 283 Wikipedia Projects growth pattern since 2001 to 2015, characterizing several similar clusters based on characteristics such as the amount of activity and the delay in updating specific topics, the degree of studies, economic resource access to information technology resources.

In Wikipedia some texts collect special and intense attention in what has been identified as an "ideological battlefield" or "ideologized" in a similar way to the French Encyclopedia, (Spasojevic et al., 2019); on the other hand, the ideological bias is identified in the editing of texts (Martin, 2018). Therefore, from the beginning, the value and adequacy of the contents included here are discussed (Robinson, 2020); or on the other hand, a way to improve society could be use and edit knowledge with criteria, it is done on the Wiki projects services (Bruni et al., 2018), (Sydow et al., 2017), however there are positions against (Tsvetkova et al., 2016)

It should be noted that Wikipedia may be edited by anyone, however not everyone does it, (Ford & Wajcman, 2017). Lazer and his colleagues (Lazer et al., 2009) identify the fulfillment of social standards and power relations as an obstacle to the development and implementation of data techniques in the study of the social sciences, and coul be for edit Wikipedia Projects

In summary, social activity could be associated with access and editing activity in Wikipedia, since it reflects the use of social collaboration, including intercultural, (Gottschalk & Demidova, 2017), (Samoilenko et al., 2016) and in itself generating a positive effect both on social interaction and on increasing the quality of the contents of this encyclopedia (Tsikerdekis, 2016); Engels (Engel & Malone, 2019) proposes that this is an adequate measure of social interaction, within interest in the current text.

3. Methods

First, the monthly Wikipedia Projects editions datasets was taken from the Wiki Stats, as data source for all Wiki Media Projects, since 2001-01-01 to 2021-02-28, by selected languages or Wiki Projects: Chinese, German, English, French, Italian, Japanese, Polish, Portuguese, Russian and Spanish, plus: total for selected projects, and all 243 Wikipedia Projects. Descriptive statistics for each project was used, underpinning that different Wikipedia Projects Edition have different behavior because all projects have different behavior based on each culture, star Wikipedia Project, edition rate or language population. Next, linear regression procedure with categorical variables was used; also, statistics for verification of assumptions was obtained, a model is built for each language selected; standard and no standard regression coefficients was tabulated, monthly and annual. Finally, the statistical Pearson´s correlation coefficients are obtained between the behaviors between years for the monthly coefficients and then for the annual coefficients. From there the different types of results are interpreted.

4. Data Collection

Table 1 shows Wikipedia Projects edition data in millions, for each year and selected language Wikipedia projects and the total of editions in all available languages, which have been gradually incorporated each year. The used data is monthly basis and here, due to their length, they are represented on an annual basis. For Total, 243 Wikipedia Projects in 2001 editions begins in 0,2M (millions) growing in 2005 to 47,4 M; in 2010 to 184,9M; down to 171,3 M to 2015 and growing again to 196,1M in 2020. For Total selected, 10 selected Wikipedia Projects in 2001 editions begins in 0,1M growing in 2005 to 39,14 M; in 2010 to 133,3M; down to 113,7 M to 2015 and growing again to 126,1M in 2020.

Table 2 exhibit descriptive statistics by Table 1, for 20 year. For 243 Wikipedia Projects: 196,07 is the maxim (Max); with 135,61 mean; and 74,60 for standard deviation (Std. D.); and 1,82 mean over standard deviation ratio; with 0,81 in Pearson Correlation coefficient with date of year. For 10 selected Wikipedia Projects: 135,11 is the maxim (Max); with 93,23 mean; and 50,07 for standard deviation (Std. D.); and 1,86 mean over standard deviation ratio; with 0,71 in Pearson Correlation coefficient with date of year. The Major means by language are: English, 46,13M; German, 9,51M; French, 8,44M and Spanish 6,09M.; all projects have Pearson Correlation coefficient with the date over 0,53.

Table 1. Selected Wikipedia Projects Annual Editions, in millions. 2001- 2020

Year	Chinese	German	English	French	Italian	Japanese	Polish	Portuguese	Russian	Spanish	Total Sel.	Total Wkpds.
2.001	-	0,0	0,1	0,0	0,0	-	0,0	0,0	-	0,0	0,1	0,2
2.002	0,0	0,1	0,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,7	0,8
2.003	0,0	0,5	1,6	0,2	0,0	0,2	0,1	0,0	0,0	0,1	2,6	3,1
2.004	0,2	3,4	6,9	1,0	0,3	1,1	0,4	0,2	0,1	0,3	14,0	16,4
2.005	0,9	7,2	20,4	3,1	1,2	2,5	1,3	0,8	0,5	1,3	39,1	47,4
2.006	1,7	12,8	56,0	7,4	3,7	5,4	3,3	2,8	1,7	3,8	98,6	120,3
2.007	2,5	14,0	73,3	10,7	6,1	6,9	4,3	3,8	3,5	6,8	131,7	167,5
2.008	2,8	13,4	72,6	11,2	7,1	6,3	4,1	4,2	5,3	8,1	135,1	176,3
2.009	2,8	13,0	68,0	10,9	7,2	5,9	4,1	4,0	7,2	8,6	131,7	177,1
2.010	3,1	13,4	64,8	11,7	7,9	5,7	4,6	4,4	8,6	9,2	133,3	184,9
2.011	3,2	13,2	58,3	12,2	8,1	4,7	4,2	4,4	8,8	8,9	125,9	182,1
2.012	5,2	13,3	57,3	12,7	8,5	4,8	4,5	4,9	9,7	9,3	130,3	193,5
2.013	5,1	12,4	53,8	12,1	7,8	4,5	4,0	4,0	8,4	8,4	120,5	187,4
2.014	3,9	10,2	48,3	10,1	6,2	3,6	2,9	2,9	6,6	6,9	101,7	154,1
2.015	4,6	11,4	53,9	10,8	7,2	4,0	2,9	3,1	7,3	8,5	113,7	171,3
2.016	3,8	10,6	56,8	11,0	7,2	4,3	3,3	3,0	6,9	7,3	114,3	187,6
2.017	4,8	10,2	57,4	10,5	8,3	4,2	3,1	3,1	6,8	8,2	116,6	181,2
2.018	4,7	10,4	55,2	10,9	7,9	4,2	3,8	2,9	6,8	7,9	114,7	182,2
2.019	4,7	10,1	55,1	10,2	7,7	4,3	2,8	3,0	6,8	9,0	113,8	182,4
2.020	5,9	10,5	62,3	12,1	7,7	5,4	3,3	2,9	6,8	9,3	126,1	196,1
Total	60,1	190,2	922,6	168,9	110,2	77,9	57,0	54,4	101,6	121,7	1.864,6	2.712,2

Table 2. Year Descriptive Selected Wikipedia Projects Annual Editions, in millions. 2001- 2020

Year Stat./ Language	Chinese	German	English	French	Italian	Japanese	Polish	Portuguese	Russian	Spanish	Total Sel.	Total Wks.
Max	5,86	13,97	73,25	12,74	8,51	6,87	4,57	4,95	9,65	9,28	135,11	196,07
Min	0,00	0,00	0,07	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,08	0,17
Mean	3,00	9,51	46,13	8,44	5,51	3,90	2,85	2,72	5,08	6,09	93,23	135,61
Std. D.	1,94	4,71	24,92	4,65	3,26	2,08	1,58	1,62	3,41	3,63	50,07	74,60
Mean/ Std. D.	1,55	2,02	1,85	1,81	1,69	1,87	1,80	1,68	1,49	1,68	1,86	1,82
Variation C.	0,65	0,50	0,54	0,55	0,59	0,53	0,56	0,60	0,67	0,60	0,54	0,55
Pear. C.C.(t)	0,94	0,57	0,65	0,77	0,83	0,53	0,59	0,57	0,78	0,81	0,71	0,81
Kurtosis C.	-1,17	0,24	-0,34	-0,49	-0,85	-0,23	-0,55	-0,71	-1,31	-0,91	-0,32	-0,49
Assimetry C.	-0,38	-1,26	-1,09	-1,13	-0,98	-0,86	-0,94	-0,73	-0,56	-0,95	-1,19	-1,16

The kurtosis coefficient, as an indicator of the centering towards the mean of the data, indicates that three languages are close to a normal concentration: German, French and Portuguese; the editions for the English, Spanish, and Russian languages, present more homogeneous or flattened concentrations with respect to the standard normal form; Chinese and Polish languages tend to focus on their average, Regarding the coefficient of asymmetry, only the Portuguese edition in its monthly average of editions resembles a normal probability distribution function. Groupings to the left

of the middle are identified for the languages: English, German, French, Spanish, Russian. The editions of the Polish and Chinese language projects are centered on the right.
Table 3 exhibit descriptive statistics by Table 1, 240 months. For 243 Wikipedia Projects: 31,46M is the maxim; with 11,35 mean; and 6,27 for standard deviation; 1,81 mean over standard deviation ratio; with 0,81 in Pearson Correlation coefficient with date of year. For 10 selected Wikipedia Projects: 135,11 is the maxim ; with 93,23 mean; and 50,07 for standard deviation; and 1,86 mean over standard deviation ratio; with 0,79 in Pearson Correlation Coefficient (Pear. C. C. (t)) with date The Major means by language are: English, 3,86M; German, 0,79M; French, 0,71M, and Spanish 0,51M.; all projects have Pearson Correlation coefficient with the date over 0,51.

Table 3. Monthly Descriptive Selected for Wikipedia Projects, Editions, in millions. 2001- 2020

Monthly Stat. / Languaje	Chinese	Deustch	English	French	Italian	Japanese	Polish	Portuguese	Russian	Spanish	Total Sel.	Total Wkps.
Max	1,15	1,55	6,74	2,11	1,25	0,62	0,95	0,89	1,40	1,53	17,19	31,46
Mean	0,25	0,79	3,86	0,71	0,46	0,33	0,24	0,23	0,42	0,51	7,80	11,35
Std. Dev.	0,18	0,39	2,06	0,40	0,28	0,17	0,15	0,14	0,28	0,31	4,15	6,27
M/ Std. Dev.	1,43	2,02	1,88	1,79	1,64	1,87	1,61	1,59	1,49	1,65	1,88	1,81
CV	0,70	0,50	0,53	0,56	0,61	0,53	0,62	0,63	0,67	0,61	0,53	0,55
Pear. C. C. (t)	0,85	0,55	0,64	0,74	0,78	0,52	0,51	0,51	0,76	0,78	0,70	0,79
Kurtosis C.	2,38	-0,13	-0,55	-0,18	-0,61	-0,51	3,01	0,97	-0,80	-0,55	-0,50	-0,37
Assimetry C.	0,64	-1,02	-0,96	-0,69	-0,49	-0,68	0,43	0,03	-0,31	-0,52	-1,00	-0,82

5. Results and Discussion

5.1 Numerical Results

Table 4 contains the standardized coefficients of the effect in the editing of the Wikipedia projects. The coefficients in terms of units are in Table 5, they are identified with the underlined and bold text at p-value less than 0,05; all the values was compared in reference to the month of December, month 12. For German, Chinese, Spanish language, an increased effect is identified in month three, March, most likely associated with increased student activity.

Table 4. Standardized monthly linear regression coefficients of Wikipedia projects 2001-2020

Month	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*	Total
1	0,04	-0,03	-0,02	0,01	0,00	0,07	0,01	0,02	0,01
2	-0,02	-0,04	-0,03	-0,02	-0,03	-0,06	-0,01	-0,01	-0,03
3	0,04	0,02	0,05	0,05	0,02	0,05	0,07	0,05	0,04
4	0,00	0,06	0,00	-0,01	-0,01	-0,04	0,02	-0,01	0,00
5	-0,01	-0,02	0,02	0,00	-0,01	-0,05	0,05	0,01	0,00
6	-0,03	-0,03	-0,01	-0,03	-0,03	-0,06	0,01	0,00	-0,02
7	-0,02	0,07	-0,01	-0,04	-0,03	-0,07	0,04	0,00	-0,01
8	0,00	0,04	0,02	-0,01	-0,02	-0,06	0,00	0,00	0,01
9	-0,03	-0,03	0,01	-0,01	-0,01	-0,06	-0,02	-0,02	-0,01
10	0,00	-0,02	0,04	0,01	0,00	-0,05	-0,01	0,01	0,00
11	0,00	-0,03	0,02	0,00	-0,01	-0,06	-0,01	-0,01	-0,01

Only three other coefficients were significant, so it would be identified that for the period in question there is no systematic effect of the months. It should be remembered that the edition data for these 20 years changes due to the increase and access. It should be remembered that the editing data for these 20 years changes due to the increase and access and therefore it is expected to observe this effect in the year coefficients, this can be observed in Table 6 of standardized coefficients and in Table 7 of the annual coefficients without standardize. Therefore, in Table 5 are the No Standardized Monthly Linear Regression Coefficient. Bold and underlined shows significance with a p-value less

than 0,05. Table 6 exposes the Annual Standardized Linear Regression Coefficients. Therefore, in Table 7 and considering as a basis the effect of the last year studied, 2020, with a coefficient of zero, it can be seen that the coefficients increase little by little in each of the languages. This will be discussed in the graphical results

Table 5. No Standardized Monthly linear regression coefficients of Wikipedia projects 2001-2020

Month	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*	Total
1	**62.241**	-16.992	-18.520	16.494	-26.941	37.103	5.068	24.242	134.584
2	-25.937	-27.270	-34.055	**-34.514**	-243.654	-31.389	-4.432	-15.600	-573.588
3	**61.027**	**13.730**	**58.400**	68.561	145.031	26.370	36.945	**52.425**	1.007.201
4	6.883	36.315	3.291	-14.129	-57.947	-23.182	12.384	-11.540	42.718
5	-10.789	-10.084	25.473	-2.017	-41.959	-26.004	27.392	6.815	-43.997
6	-40.251	-15.591	-16.230	**-48.729**	-241.712	-30.411	5.077	-660	-462.011
7	-23.831	40.718	-8.172	-64.635	-221.530	-40.172	20.369	1.221	-210.840
8	2.809	22.372	18.283	-14.032	-140.760	-31.051	983	3.130	151.508
9	-38.413	-17.143	9.203	-14.115	-104.483	-35.162	-12.952	-18.288	-158.684
10	-5.967	-13.500	44.668	18.277	23.278	-28.551	-2.806	12.232	-941
11	-1.979	-20.539	23.858	1.679	-95.096	-30.519	-4.073	-9.089	-293.840

Table 6. Annual Standardized Linear Regression Coefficients of Wikipedia projects 2001-2020

Year	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*	Total
2001	-0,47	-0,52	-0,54	-0,48	-0,50	-0,34	-0,38	-0,44	-0,54
2002	-0,46	-0,52	-0,54	-0,48	-0,49	-0,34	-0,38	-0,44	-0,54
2003	-0,45	-0,52	-0,54	-0,47	-0,48	-0,33	-0,38	-0,44	-0,53
2004	-0,31	-0,49	-0,52	-0,43	-0,44	-0,30	-0,36	-0,43	-0,49
2005	-0,13	-0,43	-0,47	-0,34	-0,32	-0,18	-0,27	-0,41	-0,40
2006	0,12	-0,33	-0,32	-0,14	0,00	0,06	-0,02	-0,33	-0,19
2007	0,18	-0,25	-0,14	0,02	0,16	0,19	0,11	-0,21	-0,05
2008	0,15	-0,21	-0,06	0,04	0,15	0,16	0,16	-0,10	-0,02
2009	0,13	-0,21	-0,03	0,03	0,11	0,15	0,13	0,03	-0,02
2010	0,15	-0,18	0,01	0,07	0,08	0,22	0,19	0,11	0,00
2011	0,14	-0,17	-0,01	0,09	0,02	0,17	0,18	0,12	0,00
2012	0,15	0,05	0,01	0,11	0,02	0,20	0,25	0,18	0,03
2013	0,11	0,04	-0,04	0,08	-0,02	0,15	0,12	0,10	0,01
2014	0,00	-0,10	-0,13	-0,01	-0,06	0,01	-0,02	-0,01	-0,09
2015	0,06	-0,01	-0,04	0,03	-0,01	0,01	0,02	0,03	-0,03
2016	0,02	-0,10	-0,10	0,03	0,01	0,07	0,00	0,00	0,02
2017	0,00	0,00	-0,05	0,01	0,02	0,04	0,01	0,00	0,00
2018	0,01	0,00	-0,07	0,03	0,00	0,13	-0,02	0,00	0,00
2019	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
2020	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Table 8 presents the spearman correlation coefficients for the annual variables. All the coefficients in this table are significant with a reliability of 95%. In any case, the variables are correlated at least between 74%, between the Chinese and German projects; and 98% between the Portuguese and Polish project. In no case is a negative or inverse behavior identified. Table 9 presents the spearman correlation coefficients for the monthly variables. Some of the coefficients of this table are significant, 34 of the 45, with a reliability of 95%. Several cases identify a negative or inverse behavior 5 of the 34, with which the behavior differs between months for those cases. The languages with the least editing behavior towards the end of the year are the projects in German, Polish, Portuguese, Russian and the total number of world projects.

Table 7. Non-standardized annual linear regression coefficients of Wikipedia projects 2001-2020

Year	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*	Total
2001	**-841.507**	**-396.921**	**-758.228**	**-859.500**	**-4.636.652**	**-233.943**	**-250.817**	**-570.143**	**-15.296.761**
2002	**-836.369**	**-396.866**	**-757.488**	**-857.639**	**-4.591.273**	**-231.575**	**-250.806**	**-570.133**	**-15.242.888**
2003	**-802.876**	**-393.988**	**-752.101**	**-842.470**	**-4.510.145**	**-226.822**	**-250.658**	**-569.776**	**-15.051.716**
2004	**-560.563**	**-376.641**	**-730.807**	**-772.627**	**-4.066.841**	**-200.687**	**-235.579**	**-562.924**	**-13.942.330**
2005	**-239.678**	**-324.043**	**-652.755**	**-604.178**	**-2.944.940**	**-124.968**	**-179.776**	**-532.509**	**-11.354.734**
2006	**222.152**	**-251.604**	**-443.709**	**-242.477**	32.473	39.785	-10.319	**-424.338**	**-5.264.969**
2007	**319.539**	**-190.485**	**-194.769**	37.363	**1.472.393**	**126.190**	**69.485**	**-276.145**	**-1.322.374**
2008	**271.558**	**-163.216**	**-83.111**	78.799	**1.417.925**	**109.257**	**106.271**	**-129.299**	-588.259
2009	**238.441**	**-162.481**	-44.425	53.290	**1.032.610**	**105.024**	**87.497**	34.292	-522.056
2010	**274.212**	**-137.642**	9.093	**124.034**	**761.436**	**147.246**	**122.692**	**148.657**	125.733
2011	**260.370**	**-131.527**	-18.452	**159.976**	229.147	**112.733**	**117.686**	**162.687**	-100.549
2012	**266.847**	39.335	14.479	**204.110**	148.138	**138.452**	**163.493**	**236.267**	849.144
2013	**191.357**	27.742	-55.917	**147.810**	-144.635	**101.800**	**80.955**	**129.911**	343.083
2014	7.725	**-74.332**	**-182.015**	-13.174	**-605.284**	8.922	-9.912	-19.097	**-2.437.258**
2015	**102.258**	-10.158	-51.035	45.342	-137.049	6.528	12.001	36.279	-985.163
2016	35.799	**-75.950**	**-145.443**	58.560	105.435	45.494	2.152	5.441	440.607
2017	5.756	2.940	-70.722	16.241	159.506	28.954	7.579	-4.163	-120.716
2018	24.328	-299	**-97.618**	49.635	-18.113	**85.762**	-10.590	-2.079	-64.762
2019	1.756	2.140	-70.722	11.241	151.506	21.954	7.179	-41163	-111.716
2020	4.328	-289	-47.618	49.635	-18.123	8.762	-10.590	-2.999	-4.762
Intersect.	842.858	397.586	749.422	866.767	4.726.471	251.757	243.839	566.402	15.344.483

Table 8. Annual Spearman Correlation Coefficients by Wikipedia Projects 2001-2020

	Year	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*
German	0,59								
Chinese	0,93	0,74							
Spanish	0,81	0,88	0,91						
French	0,77	0,94	0,90	0,98					
English	0,65	0,97	0,76	0,92	0,95				
Polish	0,61	0,97	0,78	0,92	0,96	0,97			
Portuguese	0,60	0,95	0,79	0,94	0,97	0,95	0,98		
Russian	0,80	0,79	0,91	0,96	0,93	0,79	0,85	0,89	
Total	0,82	0,92	0,91	0,98	0,99	0,95	0,95	0,94	0,92

The Table 9 indicates a variation between the month of year and the editions in a different way for each language. With no one significance to Chinese, French and English. Other languages show a tendency to increase from final part in each year. In a less sensible way that could be done in the annual model.

Table 9. Monthly Spearman correlation coefficients by Wikipedia projects 2001-2020

	Month	*German*	*Chinese*	*Spanish*	*French*	*English*	*Polish*	*Portuguese*	*Russian*
German	**-0,42**								
Chinese	-0,03	0,16							
Spanish	**0,30**	**0,35**	0,15						
French	-0,10	**0,78**	-0,13	**0,72**					
English	0,03	**0,73**	0,06	**0,75**	**0,94**				
Polish	**-0,44**	**0,90**	-0,06	0,13	**0,71**	**0,67**			
Portuguese	**-0,45**	**0,45**	**0,50**	**0,38**	**0,31**	**0,39**	**0,36**		
Russian	**-0,34**	**0,80**	0,15	**0,54**	**0,73**	**0,70**	**0,76**	**0,68**	
Total	**-0,20**	**0,80**	**0,39**	0,70	**0,82**	**0,84**	**0,68**	**0,61**	**0,84**

5.2 Graphical Results

The graphical elements in figure 1 compare the monthly behavior of the Wikipedia project between months. There, the relevance of the month of March for all languages and the relevance of the month of June, month 7 for Chinese language. There a sustained growth is observed for all languages with a maximum growth rate between the years 2004 to 2007. A stable growth zone between 2008 and 2012 and a sustained decrease in the publishing process from 2012 to 2021. With this, the systematic decrease in page publishing each year is evident, directly associated with the data in the table 6 and table 7, and a similar behavior between years and languages in table 8 and table 9.

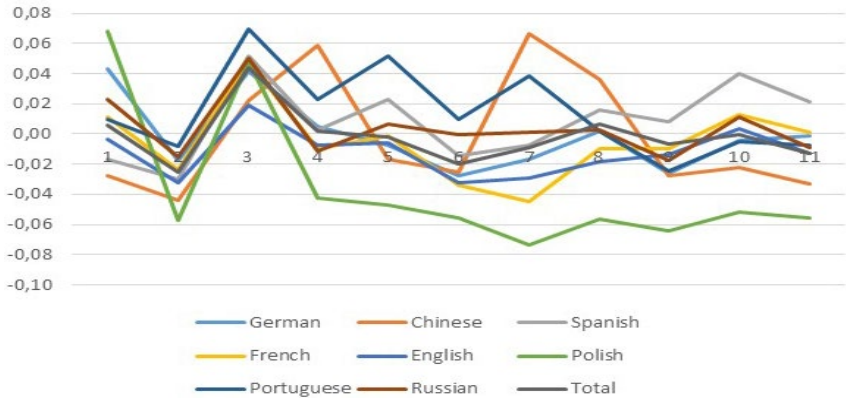

Figure 1. Monthly standardized coefficients by Wikipedia projects 2001-2020

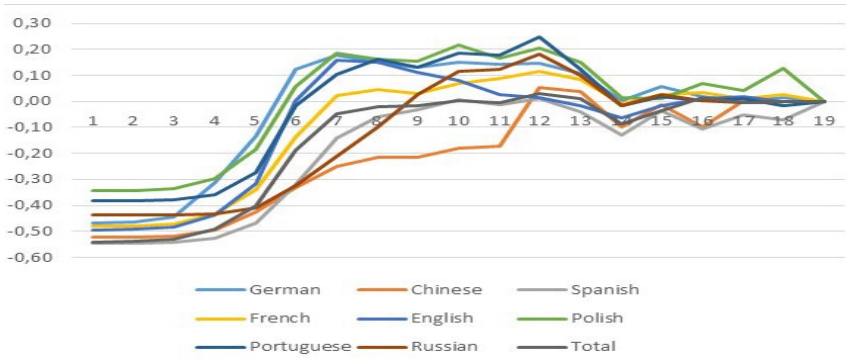

Figure 2. Annual standardized coefficients by Wikipedia projects 2001-2020

5.3 Proposed Improvements

It is proposed to compare this work not only with the editing and reading behaviors, a combined effect with the readings or identified access or new visitors to the sites. On the other hand, the contingent public health effect of the year 2020 that led to administrative closure.
Wavelet coefficients and spectral decomposition could also be used for the study of this access as signals and editing with the detail of the daily information that is available.
These variables could be linked with others of knowledge generation or economic activity

5.4 Validation

The identification of the adequacy of the underpinnings of the statistical model is in Table 10. Based on categorical regression Model, the quantity of variables for year and month and the independence between both therefore the multi collinearity free is granted. In any case de Durbin Watson Coefficient is over 1, except from Japanese, indicating a relevant first order autocorrelation between itself. All determinations coefficients are over 0,78 showing the proportion of variation in the data included in the models. The identification of the adequacy of the model is in Table 11. This have statistics: F Fisher; P value of F, % Total of information "% Total" and proportion of information of the model " % Modeled" by each source or variable type: Year "DYY", Month" DMoY", Model. P values are minor to 0,05 for all variables and languages except to Month of year for: French, Italian and Portuguese. Blacked and underlined for statistically significant variables and models. All models are significant, for all models Year variable is significant

Table 10. Regression statistics of monthly editions of Wikipedia projects 2001-2020

Month Stat/ Lang.	Chinese	German	English	French	Italian	Japanese	Polish	Portuguese	Russian	Spanish	Total Sel.	TotalWkps.
R^2	0,83	0,95	0,97	0,92	0,90	0,95	0,78	0,87	0,95	0,92	0,97	0,94
R^2Adj.	0,81	0,95	0,97	0,91	0,88	0,94	0,75	0,85	0,95	0,91	0,96	0,93
DW	2,08	1,02	1,12	1,51	1,55	0,96	1,69	1,57	1,59	1,38	1,22	1,28
AIC	5.475	5.555	6.210	5.676	5.584	5.177	5.456	5.320	5.405	5.578	6.606	6.949
SBC	5.587	5.667	6.322	5.788	5.695	5.288	5.568	5.431	5.517	5.689	6.718	7.060
PC	0,22	0,06	0,03	0,10	0,13	0,06	0,28	0,17	0,06	0,11	0,04	0,08

Table 11. Statistical Inference for Regression of monthly editions of Wikipedia projects 2001-2020

Stat.	Source	Chinese	Deustch	English	French	Italian	Japanese	Polish	Portuguese	Russian	Spanish	Total
F	DYY	51,69	213,77	401,90	126,96	90,88	200,39	36,33	67,02	209,01	115,	315,
	DMoY	1,85	2,80	2,40	1,79	1,43	2,24	2,26	1,41	1,77	1,61	2,35
	Modelo	34,00	138,97	260,09	82,56	59,17	129,96	24,27	43,75	135,49	75,2	204,
P	DYY	**9,6E-70**	**1,0E-127**	**2,1E-155**	**1,7E-105**	**9,7E-92**	**6,5E-125**	**4,5E-57**	**1,2E-79**	**9,9E-127**	**1,1E-101**	**1,2E-144**
	DMoY	**4,8E-02**	**2,0E-03**	**8,1E-03**	5,6E-02	1,6E-01	**1,4E-02**	**1,3E-02**	1,7E-01	6,1E-02	9,8E-02	**9,6E-03**
	Model	**1,7E-65**	**9,8E-123**	**3,5E-150**	**1,3E-100**	**6,0E-87**	**7,6E-120**	**1,9E-53**	**5,1E-75**	**1,2E-121**	**9,7E-97**	**1,7E-139**
% Total	DYY	81,78	94,63	97,16	91,68	88,91	94,56	75,50	85,58	94,78	91,1	96,4
	DMoY	1,61	0,68	0,32	0,71	0,77	0,58	2,58	0,99	0,44	0,70	0,39
	Model	83,39	95,35	97,46	92,42	89,73	95,05	78,18	86,59	95,24	91,7	96,7
% Mod	DYY	98,07	99,24	99,69	99,21	99,09	99,48	96,58	98,83	99,52	99,3	99,6
	DMoY	1,93	0,71	0,33	0,77	0,86	0,61	3,31	1,15	0,46	0,76	0,41

6. Conclusion

Wikipedia page editing has a similar behavior regardless of the project language on an annual basis.
The behavior in months is similar but with lower degrees of correlation and statistical significance between languages.
In the period 214 to 2015, the edition of pages of Wikipedia projects independently of the same and the scale of edition of each project decreases
The linear regression model with categorical variables is used for each year and each month within each language, so the statistics are presented in both Table 10 and Table 11. From there it is identified that the models capture between 75% and 95% of the variability total information this in the adjusted coefficient of determination. In all languages of the variables since they are growing time series or with pronounced cyclicality, so time series models could be incorporated for other purposes. On the other hand, always higher values of the Bayesian Information Criterion, SBC, are obtained with respect to the Aikake Information Criterion, which indicates a good fit of the proposed models. As they are categorical unsaturated models, a quasi-orthogonal design matrix is obtained, with which there are no multicollinearity effects, given the selection of the data, implying an adaptation of the estimators by ordinary least squares, which is the model used. Similar behaviors could occur for other electronic services or for other stages of this service

References

Ban, K., Perc, M., & Levnajić, Z. (2017). Robust clustering of languages across wikipedia growth. *Royal Society Open Science*, *4*(10). https://doi.org/10.1098/rsos.171217

Bruni, S., Calonaci, B., Giaccai, S., Margheri, M., & Viti, E. (2018). Wikipedia (and his brothers) to network and improve services. *JLIS.It*, *9*(3). https://doi.org/10.4403/jlis.it-12501

Engel, D., & Malone, T. W. (2019). Integrated information as a metric for group interaction. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0205335

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, *47*(4). https://doi.org/10.1177/0306312717692172

Gottschalk, S., & Demidova, E. (2017). MultiWiki: Interlingual text passage alignment in wikipedia. *ACM Transactions on the Web*, *11*(1). https://doi.org/10.1145/3004296

Kikkawa, j., Takaku, m., & Yoshikane, f. (2020). Analyses of Wikipedia Editors Adding Bibliographic References based on DOI Links. *Joho Chishiki Gakkaishi*, *30*(1). https://doi.org/10.2964/jsik_2020_004

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Social science: Computational social science. In *Science* (Vol. 323, Issue 5915, pp. 721–723). https://doi.org/10.1126/science.1167742

Lemmerich, F., Sáez-Trumper, D., West, R., & Zia, L. (2019). *Why the World Reads Wikipedia*. 618–626. https://doi.org/10.1145/3289600.3291021

Margolin, D. B., Goodman, S., Keegan, B., Lin, Y. R., & Lazer, D. (2016). Wiki-worthy: collective judgment of candidate notability. *Information Communication and Society*, *19*(8). https://doi.org/10.1080/1369118X.2015.1069871

Martin, B. (2018). Persistent Bias on Wikipedia: Methods and Responses. *Social Science Computer Review*, *36*(3), 379–388. https://doi.org/10.1177/0894439317715434

Quian, A., & Elías, C. (2017). Wikipedia y sus relatos colaborativos como indicador de interés ciudadano: aplicación a la cultura hacker-wikileaks. *Prisma Social*, *18*, 85–123.

Robinson, A. (2020). The ascent of Wikipedia. *Science*, *370*(6517). https://doi.org/10.1126/science.abf2433

Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., & Strohmaier, M. (2016). Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity. *EPJ Data Science*, *5*(1). https://doi.org/10.1140/epjds/s13688-016-0070-8

Spasojevic, N., Bhargava, P., & Hu, G. (2019). DAWT: Densely annotated Wikipedia texts across multiple languages. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, 1655–1662. https://doi.org/10.1145/3041021.3053367

Sydow, M., Baraniak, K., & Teisseyre, P. (2017). Diversity of editors and teams versus quality of cooperative work: experiments on wikipedia. *Journal of Intelligent Information Systems*, *48*(3), 601–632. https://doi.org/10.1007/s10844-016-0428-1

Toribio, M. I. H., Claes, F., & Deltell, L. (2020). Spanish/castilian on wikipedia: Voices and discussion forum. *Doxa Comunicacion*, *2020*(31), 283–302. https://doi.org/10.31921/doxacom.n31a14

Tsikerdekis, M. (2016). Personal communication networks and their positive effects on online collaboration and

outcome quality on Wikipedia. *Journal of the Association for Information Science and Technology*, *67*(4). https://doi.org/10.1002/asi.23429

Tsvetkova, M., Garciá-Gavilanes, R., & Yasseri, T. (2016). Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration. *Scientific Reports*, *6*. https://doi.org/10.1038/srep36333

Zhang, H., Zhang, S., Wu, Z., Huang, L., & Ma, Y. (2016). A Method for Predicting Wikipedia Editors' Editing Interest. *International Journal of Web Services Research*, *13*(3). https://doi.org/10.4018/ijwsr.2016070101

Zheng, L., Albano, C. M., Vora, N. M., Feng, M. A. I., & Nickerson, J. V. (2019). The roles bots play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). https://doi.org/10.1145/3359317

Ban, K., Perc, M., & Levnajić, Z. (2017). Robust clustering of languages across wikipedia growth. *Royal Society Open Science*, *4*(10). https://doi.org/10.1098/rsos.171217

Bruni, S., Calonaci, B., Giaccai, S., Margheri, M., & Viti, E. (2018). Wikipedia (and his brothers) to network and improve services. *JLIS.It*, *9*(3). https://doi.org/10.4403/jlis.it-12501

Engel, D., & Malone, T. W. (2019). Integrated information as a metric for group interaction. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0205335

Ford, H., & Wajcman, J. (2017). 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, *47*(4). https://doi.org/10.1177/0306312717692172

Gottschalk, S., & Demidova, E. (2017). MultiWiki: Interlingual text passage alignment in wikipedia. *ACM Transactions on the Web*, *11*(1). https://doi.org/10.1145/3004296

Kikkawa, j., Takaku, m., & Yoshikane, F. (2020). Analyses of Wikipedia Editors Adding Bibliographic References based on DOI Links. *Joho Chishiki Gakkaishi*, *30*(1). https://doi.org/10.2964/jsik_2020_004

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Social science: Computational social science. In *Science* (Vol. 323, Issue 5915, pp. 721–723). https://doi.org/10.1126/science.1167742

Lemmerich, F., Sáez-Trumper, D., West, R., & Zia, L. (2019). *Why the World Reads Wikipedia*. 618–626. https://doi.org/10.1145/3289600.3291021

Margolin, D. B., Goodman, S., Keegan, B., Lin, Y. R., & Lazer, D. (2016). Wiki-worthy: collective judgment of candidate notability. *Information Communication and Society*, *19*(8). https://doi.org/10.1080/1369118X.2015.1069871

Martin, B. (2018). Persistent Bias on Wikipedia: Methods and Responses. *Social Science Computer Review*, *36*(3), 379–388. https://doi.org/10.1177/0894439317715434

Quian, A., & Elías, C. (2017). Wikipedia y sus relatos colaborativos como indicador de interés ciudadano: aplicación a la cultura hacker-wikileaks. *Prisma Social*, *18*, 85–123.

Robinson, A. (2020). The ascent of Wikipedia. *Science*, *370*(6517). https://doi.org/10.1126/science.abf2433

Samoilenko, A., Karimi, F., Edler, D., Kunegis, J., & Strohmaier, M. (2016). Linguistic neighbourhoods: explaining cultural borders on Wikipedia through multilingual co-editing activity. *EPJ Data Science*, *5*(1). https://doi.org/10.1140/epjds/s13688-016-0070-8

Spasojevic, N., Bhargava, P., & Hu, G. (2019). DAWT: Densely annotated Wikipedia texts across multiple languages. *26th International World Wide Web Conference 2017, WWW 2017 Companion*, 1655–1662. https://doi.org/10.1145/3041021.3053367

Sydow, M., Baraniak, K., & Teisseyre, P. (2017). Diversity of editors and teams versus quality of cooperative work: experiments on wikipedia. *Journal of Intelligent Information Systems*, *48*(3), 601–632. https://doi.org/10.1007/s10844-016-0428-1

Toribio, M. I. H., Claes, F., & Deltell, L. (2020). Spanish/castilian on wikipedia: Voices and discussion forum. *Doxa Comunicacion*, *2020*(31), 283–302. https://doi.org/10.31921/doxacom.n31a14

Tsikerdekis, M. (2016). Personal communication networks and their positive effects on online collaboration and outcome quality on Wikipedia. *Journal of the Association for Information Science and Technology*, *67*(4). https://doi.org/10.1002/asi.23429

Tsvetkova, M., Garciá-Gavilanes, R., & Yasseri, T. (2016). Dynamics of Disagreement: Large-Scale Temporal Network Analysis Reveals Negative Interactions in Online Collaboration. *Scientific Reports*, *6*. https://doi.org/10.1038/srep36333

Zhang, H., Zhang, S., Wu, Z., Huang, L., & Ma, Y. (2016). A Method for Predicting Wikipedia Editors' Editing Interest. *International Journal of Web Services Research*, *13*(3). https://doi.org/10.4018/ijwsr.2016070101

Zheng, L., Albano, C. M., Vora, N. M., Feng, M. A. I., & Nickerson, J. V. (2019). The roles bots play in Wikipedia. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW). https://doi.org/10.1145/3359317

Biography

Wilson Alexander Pinzón Rueda is Research Professor of Engineering, director of research groups of the Facultad Tecnologica of the Universidad Distrital Francisco José de Caldas, he is Industrial Engineer of the Universidad Distrital Francisco José de Caldas, master in Industrial Engineering of the Universidad de los Andes. Experience as a teacher in engineering programs at Universidad de la Salle, researcher teacher in master and doctorate programs at Universidad Distrital Francisco José de Caldas. He has published articles in journals and conferences.

Carlos Andrés Arango Londoño is a member of industrial engineering investigation group and serves as an adjunct professor at Universidad de la Salle. Mr Arango is a mechanical Engineering from Universidad Tecnologica de Pereira and a master in industrial engineering from Universidad de los Andes. He has taught courses in operations management, produccion planning and scheduling and operations research for engineers. His interest are production planning, operations management and logistics processes.

Jose Martin Diaz is an Industrial Engineer and a member of the Facultad de Ciencias Empresariales of Corporacion Universitaria Minuto de Dios. He has taught courses in operations management, produccion planning and scheduling and operations research for engineers. His interest are production planning, operations management and logistics processes. Experience in leading research groups and monitoring different projects, formulating research projects, coordinating / creating research teams.