Shchekotin, E., Myagkov, M., Goiko, V., Kashpur, V. (2021). Digital methods of analysis of subjective quality of life: case of Russian Regions. *Administratie si Management Public*, 36, 25-48. DOI: 10.24818/amp/2021.36-02

Digital methods of analysis of subjective quality of life: case of Russian Regions

Evgeniy SHCHEKOTIN[1], Mikhail MYAGKOV[2], Viacheslav GOIKO[3], Vitaliy KASHPUR[4]

***Abstract**: The paper presents a method for measuring subjective quality of life in the regions of the Russian Federation on the basis of digital data. Information about online activity of users in the largest social media in Russia - VKontakte was taken as the data source. Quality of Life Index was calculated based on the obtained data. The results show that overall users tend to negatively assess the quality of life in their regions, with the highest estimates of the quality of life observed in the "ethnic" areas of the Russian Federation – republics, autonomous territories, districts and regions, first of all, in North Caucuses, Siberia and the Russian Far East. The lowest quality-of-life assessments of are noted in some regions in West Siberia. The paper analyses how the results of measuring quality of life with digital methods correlate with objective social-and-economic and demographic indicators of regional development. Several regularities for the "ethnic" areas are revealed while in other areas (regions, territories, cities of federal status) no significant correlations were established. The paper also gives a comparative analysis of quality of life assessments obtained through traditional questionnaire methods and digital methods.*

Keywords: subjective quality of life, digital sociology, digital methods, social network, Russia

JEL: C83, I31, R59

DOI: 10.24818/amp/2021.36-02

[1] PhD, Center of Applied Big Data Analysis, Tomsk State University, Tomsk, Russia; Department of Sociology, Novosibirsk State University of Economics and Management, Novosibirsk, Russia, e-mail: evgvik1978@mail.ru. Corresponding author

[2] Professor PhD, Center of Applied Big Data Analysis, Tomsk State University, Tomsk, Russia; Institute of Education, National Research University Higher School of Economics, Moscow, Russia; Institute for Cognitive and Decision Sciences and Department of Political Science, Eugene, USA, e-mail: myagkov@skoltech.ru

[3] Center of Applied Big Data Analysis, Tomsk State University, Tomsk, Russia, e-mail: goiko@ftf.tsu.ru

[4] PhD, Center of Applied Big Data Analysis, Tomsk State University, Tomsk, Russia, e-mail: vitkashpur@mail.ru

Introduction

Digital sociology has been one of the most actively developing areas of sociology in recent years. Digital sociology studies a wide range of phenomena – whether analyzing the place of digital technologies in people's life or researching people's life on the basis of digital data created by users in the digital environment (Lupton, 2015; Androniceanu et al., 2021). Employing "digital methods" for studying various social phenomena, that occur both online and offline, also falls in the competence of digital sociology (Rogers, 2013). As a new research direction, digital sociology has become a subject of vivid discussions and acquired fervent advocates as well as fierce challengers, especially with regard to use of powerful computer technologies (such as Big Data and Data Mining) to study social behavior of humans.

In a polemic heat, some advocates of digital methods in social sciences tend to announce a new era of knowledge production based on Big Data, emerging new approaches to data analytics and machine learning which mark "the end of theory" and make the traditional scientific research method obsolete (Anderson, 2008). This point of view has become rather popular outside sociology and the academic science. Digital methods are most actively employed in the "corporate science" pursued by such companies as Microsoft, Facebook, Google, in particular branches of applied research, and so on (Burrows & Savage, 2014).

Critics, in their turn, bring forward a counter-accusation that digital method does not conform to the standards of the scientific principle; digital data are unrepresentative and nonobjective (Marres, 2017). The concerns of the disciples of academic sociology are quite understandable since use of digital sociology methods in humanities and social sciences, for example, big data analysis, requires reviewing fundamental ontological, epistemological and methodological principles of these sciences, and a series of basic concepts such as thing, agent, time, context, causality must be redefined (Wagner-Pacifici et al., 2015; Ciobanu et al., 2019). Use of digital methods equally leads to reconsidering many methodological principles, for instance, rejecting deduction in social sciences to the benefit of the induction principle, etc. In this paper, we do not set the goal to scrutinize vivid theoretical discussions regarding the value and reliability of digital methods. Let's just point out that we agree with the opinion of Burrows and Savage (2014) that if sociologists do not engage in the new research area based on digital technologies, sociology as a science may lose its predominant right to talk about social issues.

As an umbrella concept, digital sociology covers numerous separate areas, themes and research approaches united by use of digital methods or digital data in order to study social life. Digital sociology is actively applied, inter alia, in wellbeing studies and quality of life studies. The paper objective is to demonstrate with a particular empirical study how digital methods and digital data can be used to research the quality of life of the population in regions across Russia and make a comparative analysis of traditional approaches to measuring wellbeing and quality of life of the population on the basis of statistical and polling data, and digital methods.

Therefore, the research agenda of the paper can be formulated as follows:

1. To what extent data measuring the level of subjective quality of life, obtained through digital methods, and statistical indicators used to evaluate particular aspects of wellbeing and quality of life, conform?

To this purpose, we conducted a large-scale study of subjective quality of life of the population in Russian regions based on the data from the largest Russian social network - VKontakte. Then using correlation analysis, we identified presence of any correlations between the subjective level of quality of life, estimated on the basis of the social media data, and a set of statistical data characterizing qualify of life in a particular region.

2. To what extent data measuring the level of subjective quality of life, obtained though digital methods, conform to the findings of sociological surveys evaluating the wellbeing level of the population of Russia?

To get an answer to this question, we used correlation analysis to expose regularities between estimates of subjective quality of life made on the basis of data from VKontakte social media and data of sociological surveys carried out by the largest Russian sociological centers (VCIOM, FOM, Levada-Centre), concerning particular indicators of wellbeing and quality of life.

1. Literary review

Since we use social media as the data source in this study, the review is limited to similar studies. It should be pointed out that social media are the most popular source of data about wellbeing and quality of life, but not the only one. Other sources of digital traces can be users' search queries (Algan et al., 2019) or words in Google Books (Hills et al., 2019), but few such works have been completed. Overall, one can see three areas at the intersection of social media and wellbeing studies. The first is related to use of new information technologies for analyzing and processing data that can serve as a source of information in wellbeing studies. An example is use of the technology of automatic recognition of human emotions in group images in social media to determine "happiness" intensity (Dhall et al., 2015).

The second considers social media as a new social phenomenon that influences various sides of human wellbeing (Sabatini & Sarracino, 2017; Verduyn et al., 2017). In this case social media is taken as a significant factor that can have a negative or positive effect upon such parameters of human wellbeing as social capital (Burke et al., 2010; Ciobanu & Androniceanu, 2018), marriage satisfaction (Valenzuela et al., 2014), depression (Appel et al., 2016; McCloskey et al., 2015), loneliness (Song et al., 2014), social support (Lee et al., 2013), evaluating one's life as happy compared to other people (Chou & Edge, 2012) and so on.

Finally, the third area of studies sees social media as a self-sufficient data source for evaluating wellbeing. Taking into account the role that social media started playing in people's daily life in recent decades, the attempts seem to be quite productive (Haseeb et al., 2019). For instance, Hao et al. (2014) used machine learning technology to predict subjective quality of life of social media users. The authors took data in Sina Weibo on 1785 volunteers to learn an algorithm, asking them a priori to fill in questionnaires in order to evaluate positive and negative affects (PANAS) and assess psychological wellbeing (PWBS). Schwartz et al. (2016) undertook a similar study based on Facebook materials. In other work, Schwartz et al. (2013) evaluated subjective wellbeing in various US counties using Twitter. Chen at al. (2017) analyzed status updates on Facebook user pages to predict subjective wellbeing of users. Wu et al. (2015) used data from Sina Weibo to build the City Happiness Index. The authors analyzed statements about users' life in various cities.

Wang et al. (2014) studied profiles of Facebook users for a year and evaluated their wellbeing level under Diener's Satisfaction With Life Scale (SWLS) and compared the findings with the index of Facebook's Gross National Happiness, calculated on the basis of analyzing the number of positive and negative words in user status updates. The authors questioned possibility of applying linguistic analysis of online messages for analyzing users' psychological state. Yang and Srinivasan (2016) studied life satisfaction of the population on materials from Twitter. Studies on "happiness geography" also fall in this category, for example, relations between emotions and mobility (Mitchell et al., 2013). Interesting findings were obtained in several studies as to how positive and negative emotions correlate with life satisfaction measured on the basis of Facebook data. It appeared that positive emotions have low correlation with Facebook data on life satisfaction while negative emotions demonstrate such link. One can assume that social pressure can stimulate unhappy people to pretend that they are happier than they actually are which is less likely in case of negative emotions.

It can be pointed out that in the past ten years the number of studies where social media are the main source of information about subjective user wellbeing has been increasing rapidly. At the same time, an essential question is to what extent the results of digital and traditional studies of wellbeing and quality of life are consistent and whether digital studies enable to predict the "true" level of wellbeing. Bellet and Frijters (2019) identified two types of data for predicting the level of wellbeing on the basis of digital data: individual data (updating status in user accounts, likes put by a user, etc.) and aggregated data, i.e., not tied to a particular user, averaged for a particular group (search queries, average network values, tags, updates, words from Google Books, etc.). Bellet and Frijters concluded that aggregated indicators of wellbeing and quality of life derived from the text content of social media have a much higher prediction capacity compared to predictions based on individual data.

2. Study methodology on subjective quality of life based on social media data

The study methodology includes several consistent stages:

1) Developing a model of subjective quality of life.
2) Choosing communities in VK social media.
3) Classifying messages and posts in communities.

4) Automatically analyzing content in the chosen communities.
5) Building up an quality of life index for various regions of Russia.

2.1 The quality of life model

Research literature describes a vast number of approaches and models of wellbeing and quality of life, many of which have common components (Androniceanu A.-M et al., 2020). The most popular include the level of life, the level of social infrastructure development, the ecological state of the environment, the state of health, and so on.

We did not employ some popular quality of life indicators in our model due to the complexity of operationalizing and the specifics of the chosen communities in social media. For example, such things as income level, family life, demographic situation and social connections are not widely discussed in the mass regional communities. We have also excluded quality of work life, quality of leisure and recreation and inequality as separate parameters, because studies of these indicators require a deeper analysis. However, we have included several additional parameters. First of all, we have added the assessment of various administrative solutions and political situation in the region, because today the Russian authorities have direct and strong impact on the other parameters of quality of life. We have also included such parameters as general emotional state (psychological subjective well-being) and relationships between people to assess the emotional environment in the region. One more parameter of quality of life, very important but frequently forgotten in the Russian studies, is the situation with political rights and freedom (Veenhoven, 1996). Table 1 gives an overview of subjective quality of life parameters we have used in our study.

Table 1. Structure of quality of life parameters

Areas	Categories
Social	*Education* *Housing and utilities sector* *Healthcare* *Infrastructure* *Safety (situation in cities)* *Environment* *Relationships between people*
General emotional state	
Economic	*Work* *Products* *Taxes* *Lending and entrepreneurship* *Social support on behalf of the state*

Areas	Categories
Political	*Media freedom* *Remonstrative potential* *(resentment of population)* *Election freedom* *Attitude towards authorities* *Political resolutions* *Domestic politics*

(*Source:* Authors model)

2.2 Selection of VK communities

In our study we have included 83 out of 85 regions of the Russian Federation. We have not managed to collect data on two regions: Chechnya and Mordovia. In each of these 83 regions we have determined 3 largest cities and selected 10 VK communities discussing life there. We have defined them as 'urban communities'. We have filtered the communities according to several criteria:

1. They publish informative posts on social, economic and political life.
2. They publish posts of their subscribers with info on social, economic and political life.
3. The published posts contain sentiments on news and events.

We have excluded the following communities:

1. Online shops and other commercial groups.
2. Groups with information on sports and cultural events and personalities.
3. Communities of public places (restaurants, clubs, cinemas, etc.).
4. Food delivery services.
5. Communities on health, nutrition, fitness, etc.
6. Communities on exchange of items and charity.
7. Communities with storytelling, stories and questions.
8. Dating communities.
9. Job offers communities.

In some regions (the Buryat Republic, Dagestan, Ingushetia, Tatarstan) we have included 10 regional groups not connected with any definite city. In these regions there are more than 3 large towns where the majority of the population is concentrated, so here we have included general regional groups to increase representativeness. In some regions with high population density in large towns we have selected only two or even one of them (like in Nenets Autonomous Area). We have searched for the communities manually and selected the largest ones corresponding to the above mentioned criteria. In doing so, we have built a cluster of 2410 communities.

2.3 Classification of messages and posts in communities

At the next stage of our study we have used the social media data collection and analysis platform of the University consortium of Big Data researchers (www.opendata.university), developed by the team of the Laboratory of Big Data in Social Sciences of Tomsk state University to download the materials from the selected communities. We have downloaded all messages, posts and comments for the period between January 1 and December 31, 2018. After that we have deleted all 'junk' like advertising, as well as the information beyond the scope of this study (job offers, sports and cultural events, free exchange, contests and campaigns, recipes, delivery and food, astrological predictions, sales of items, dating offers, discussions of private life of participants, etc.). We have deleted the 'junk' in two stages:

1) manual cleanup of approximately 60 thousand messages.

2) automatic cleanup based on the specially designed algorithm trained during manual cleanup. We have left only the messages of 20 words and longer and we have deleted all repeated messages.

After cleanup we have left only approximately 3 300 000 messages. At the same time we have categorized the messages according to topics (given in Table 1) and style (positive, negative, neutral).

2.4 Automatic content analysis in selected communities

The database set includes posts from the walls of the regional VK communities retrieved through its public API (Application Programming Interface). Each post had to be categorized into one of the 19 categories or as 'junk'. The method is based on machine learning technologies of retrieval of unknown patterns from these texts. To create an automatic algorithm of texts classification we have used the following conventional libraries of machine learning: Scikit Learn, Pandas, Numpy as well as a set of tools known as NLTK (Natural Language Toolkit) for the natural language analysis. The algorithm is based on Python 3 programming language.

At the data preprocessing stage we have deleted symbols belonging to neither English nor Russian alphabet. We have used stemming to bring all words to their basic forms. We have deleted all rare terms, which could have been typos. To be able to use different classification methods we had to present the texts in the vector form. To transform the texts into the terms significance vector we have applied TF-IDF (TF – term frequency, IDF – inverse document frequency), where the word weight is used in proportion to the frequency these words occur in a document and in inverse proportion to the frequency of the words use in all documents from the sample. TF-IDF is frequently applied to present documents as numeric vectors reflecting the significance of each term from a certain term set (a number of terms determines the vector dimension) in each document. For each vector we have taken into account not only separate terms but also diagrams, that is, pairs of consecutive terms.

We have conducted a number of experiments to determine the best data for learning. To do that we have checked several data sets including: number of comments, likes, reposts, views, number of words in a post. All these values have been scanned in accordance to their average value in a community the corresponding text was taken from. We have also considered the vectors of significance for the comments to the posts retrieved the same way, as well as for the posts texts. After that we have determined the best data set for this task: significance vectors of the words in the posts, the scanned number of words and the number of comments, likes, reposts and views.

Based on the obtained data we have built the machine learning models. We have conducted a number of experiments, where we have categorized the samples as the one for learning and the one of a text, with their subsequent validation to select the most accurate model to classify categories and attitudes. After that we have checked the models of Gradient Boosting, where we have used a prediction model as an ensemble of weak predictive models and the Random Forest with variations of hyper parameters. In the end we have selected the models demonstrating the best result for the corresponding task. We have implemented the gradient boosting from LightGBM library. We have validated the data to determine the accuracy of the obtained models. The accuracy of the categories classifier is 73%, the accuracy of the attitudes classifier is 77,5%.

2.5 Subjective quality of life index for regions of the Russian Federation

After that we have built the subjective quality of life index (QLI) for each selected region. The method of aggregates such as the index is widely used in various well-being and life quality studies. In this article we have also used the indices of online activities to calculate the SWBI. We have calculated the QLI according to the following formula:

$$I_{kjt} = A_{kjt}/B_k \quad (1)$$

where I_{kjt} is the Online Activity Index (OAI) for the region (k) for this subjective quality of life value (j) for the certain attitude of message. This index determines the intensity of discussion of this topic in the certain attitude in the certain region. The online activity index demonstrates how topical the subject of the message is (quality of life value) for the region and subjective assessments of this value by the social network users.

A_{kjt} is the value on online activity in the certain region for the certain value of subjective well-being; it is calculated according to the formula:

$$A_{kjt}= L+2\times C+5\times R \quad (2)$$

where L is the amount of likes collected by the messages on or around a certain value of subjective quality of life in the certain region in the certain attitude.

C is the number of comments collected by the messages related to the certain value of subjective quality of life in the certain region. We have equated each comment with 2 likes, because, in our view, this action of a user is an evidence of importance of these messages for the commentator. Here a like is a passive form of demonstrating support of this message.

R is a number of reposts of the messages on the certain value of subjective quality of life in the certain region. We have equated each repost with 5 likes, because, in our view, a repost is an evidence of complete and active support of this message by the user. This action means that the user not only expresses his or her consent with this message but also openly demonstrates his or her solidarity with the message. As compared to various forms on online activities, a repost is an evidence of the greatest topicality of this subject for the user.

B_k is a total number of subscribers in all selected communities in the region. This value demonstrates a relative value of online activity for this region.

k is a number of each region (1 to 83). The study involves 83 out of 85 regions of the Russian Federation. We have not managed to collect reliable data for two regions: Mordovia and Chechnya.

j is a topic of messages, that is, an indicator of well-being, which we have included into the model of life quality (1 to 19) (according to Table 1)

t is the attitude of messages (0, 1 or 2).

Thus, I_{kjt} shows intensity of discussion of a topic in the selected communities in the selected region. It is an evidence of urgency and topicality of this subject for the population of the region. I_{kjt} has been calculated for each attitude, that is, for each region OAI has three values: one for positive attitude, one for negative attitude and one for neutral attitude. Attitude is defined as an emotional evaluation of a message, thus, the positive attitude means that the message contains some positive evaluation or expression of approval of some news or situation mentioned in the message; the negative attitude means it contains disapproval, resentment towards the contents of the message, while the neutral attitude means the message is purely informative and contains no evaluation.

We have calculated the monthly value of the index for each subject and each attitude; after that we have calculated the mean monthly value for each subject and each attitude. Mean monthly values have been calculated by addition of the online activity index for definite attitude and dividing the sum by 12. Later, by simple subtraction of mean values of the online activity index for positive and negative attitudes, we have calculated the mean value of the subjective quality of life index QLI_{kj} for the corresponding region and for the corresponding topic/value of subjective quality of life:

$$QLI_{kj\Delta m} = I_{kj1m} - I_{kj2m} \tag{3}$$

where I_{kj1m} is the mean monthly online activity index with the positive attitude I_{kj2m} is the mean monthly online activity index with the negative attitude.

We have not included neutral messages into the QLI. We have calculated the total Russian value of subjective quality of life QLI_{tot} as the sum of the index for all quality of life subjects/values for each region:

$$QLI_{tot} = \Sigma\, QLI_{kj\Delta m} \quad (4)$$

The suggested method is perfect to study the life quality of the ‘digital’ population in the region, that is, the part of the population, which can be referred to as active users of social media. This study does not include the population groups, which, for whatever reasons, do not use social media. The second limitation is caused by technical aspects of this method: we have not used botnets in the social networks to manipulate the public opinion. We have not managed to level up this issue. Another limitation is related to psychological peculiarities of the users’ behavior. It is a well-known fact that users tend to react to negative messages.

At it was pointed out, users tend to more actively react to negative events and phenomena (Liebrecht et al., 2019; Stafford, 2014; Trussler & Soroka, 2014).

3. Statistical indicators for comparative analysis

For a comparative analysis of the study findings and statistical data characterizing the quality of life of the population in regions, we took data published on the website of the Federal State Statistical Service (https://www.gks.ru/). Statistical indicators used in the analysis are listed in Tables 4-7. Data on statistical indicators are published on a yearly basis and with a breakdown by particular regions. Therefore, to perform Pierson correlations we used statistical indicators for particular regions in 2018 as the predictor variable and the Quality of Life Index for particular regions in a year – as the dependent variable.

4. Sociological surveys for comparative analysis

Choosing sociological surveys for the comparative analysis with the findings of the online study, we were guided by the following several rules:

1) Surveys are conducted on a regular basis by a reputable research body. We used open-access data produced by VCIOM, FOM and Levada-Centre.

2) Surveys should cover some quality of life indicators that we use in our study (summarized in Table 1). The sociological surveys taken for the comparative analysis are presented in Table 2.

Since all surveys that we used for the comparative analysis are national and there is no regional data breakdown and the surveys are carried out on a monthly or quarterly basis, we recalculated our Quality of Life Index (OLI). We estimated it for all regions by each quality of life indicator. To this purpose, we calculated QLI in the positive and negative modes separately in each month for all regions of Russia on each quality-of-life indicator:

$$QLI_{njt} = \Sigma\, QLI_{knit} \quad (5)$$

where k – number of regions (from 1 to 83)
n – number of months (from 1 to 12)
j – quality-of-life indicator (from 1 to 19)
t – message mode (1 – positive or 2 – negative).

It means that we got two national indices per each month of 2018 on two modes. Then we calculated the overall Quality of Life Index QLI_{total} for each month as the difference between QLI_{nj1} in the positive mode and QLI_{nj2} in the negative mode. We also normalized QLI pro rata to the share of the population of a particular region in the total population of Russia.

We used Spearman correlations for monthly sociological surveys; and Kendall correlations for quarterly (Happiness Index, Economic Protest Potential, Political Protest Potential) or bi-quarterly (Consumer Sentiment Index, Social Sentiment Index, Family Index, Index of Russia, Expectations Index, Power Index) surveys. Data from sociological surveys were used as the predictor variable and the national Quality of Life Index values per each months of a year – as the dependent variable.

Table 2. Correlation analysis of QLI indicators and data of sociological surveys

Data of sociological survey / **QLI Indicators**	**Housing-and-Utilities**	**Medicine**	**Job**	**Goods**	**Market relations**	**Protest potential**	**Political decisions**	**Home policy**
Social Assessment Indices	0.46		0.52		0.55			0.71
Social Sentiments Indices – “To what extent does your current life suit you?”								0.46
Social Sentiments Indices – “Do you think that in a year you (your family) will live better than now?”						0.50	0.53	0.50
Social Sentiments Indices – “How would you evaluate the current financial		-0.49				0.58	0.71	

Data of sociological survey / QLI Indicators	Housing-and-Utilities	Medicine	Job	Goods	Market relations	Protest potential	Political decisions	Home policy
situation of your family?”								
Indices on the situation in Russia – “How would you assess the situation in Russia in general?”					0.58			0.61
Indices on the situation in Russia - “How would you assess the current political situation in Russia in general?”			0.49		0.52	0.49		
Indices on the situation in Russia – “To what extent do you agree that things in Russia are developing in the right direction?”			0.52		0.55			0.55
Inflation perception indices - How would you evaluate the price growth?	-0.52	0.52		-0.73				-0.45
Inflation perception indices - In your opinion, how will prices for the main consumer goods and services change in the next 1-2 months?	-0.64			-0.55	-0.55			-0.58
Frustration with the authorities - In the past month, did you hear critical						-0.71	-0.46	

Data of sociological survey / QLI Indicators	Housing-and-Utilities	Medicine	Job	Goods	Market relations	Protest potential	Political decisions	Home policy
statements about Russian authorities from people around you? (Yes)								
Frustration with the authorities - In the past month, did you hear critical statements about Russian authorities from people around you? (No)						0.70	0.57	
Frustration with the authorities - In the past month, were you frustrated, outraged with actions of Russian authorities? (Yes)						-0.49	-0.62	
Frustration with the authorities - In the past month, were you frustrated, outraged with actions of Russian authorities? (No)	0.46				0.52		0.46	
Sentiments of the environment – “Which sentiments, in your opinion, dominate today among your relatives, friends, colleagues and acquaintances? (Calm)	0.53	-0.47					0.47	0.72

Data of sociological survey / QLI Indicators	Housing-and-Utilities	Medicine	Job	Goods	Market relations	Protest potential	Political decisions	Home policy
Sentiments of the environment - “Which sentiments, in your opinion, dominate today among your relatives, friends, and acquaintances? (Worries)	-0.56	0.50					-0.47	-0.75
Home policy of the government							0.48	0.55
Government economic policy						0.46	0.46	0.55
Government social policy						0.46	0.68	0.46
Assessing the current state of things in Russia – In the right direction					0.49			
Assessing the current state of things in Russia – On a wrong path					-0.57			
Consumer Sentiments Index								0.83
Index of the authorities						0.83		
Social protest potential						-0.53	-0.69	

(*Source:* Authors calculation)

5. Results

Table 3 represents the results of calculating the Quality of Life Index in 83 regions of the Russian Federation in 2018.

Table 3. The overall QLI by regions in 2018 (the total of all quality-of-life indicators for a particular region)

Region	$QLI_{\Delta av}$	Region	$QLI_{\Delta av}$
Adygeya	-0.04637	Nenetsky Autonomous District	-0.07537
Altai Territory	-0.18246	Nizhny Novgorod Region	-0.06259
Amur Region	-0.02586	Novgorod Region	-0.09415
Arkhangelsk Region	-0.07722	Novosibirsk Region	-0.10585
Astrakhan Region	-0.04684	Omsk Region	-0.07532
Bashkortostan	-0.03384	Orenburg Region	-0.19532
Belgorod Region	-0.05497	Orel Region	-0.04635
Bryansk Region	-0.08234	Penza Region	-0.08399
Buryatia	-0.04452	Perm Territory	-0.00481
Vladimir Region	-0.06636	Primorye Territory	-0.0101
Volgograd Region	-0.02053	Pskov Region	-0.00093
Vologda Region	-0.15559	Altai Republic	-0.00463
Voronezh Region	-0.04727	Rostov Region	-0.09866
Dagestan	-0.03179	Ryazan Region	-0.00225
Jewish Autonomous Region	-0.00869	Samara Region	-0.00711
Zabaikalie Territory	-0.03046	St Petersburg	-0.00326
Ivanovo Region	-0.0245	Saratov Region	-0.15107
Ingushetia	-0.00102	Sakha (Yakutia)	-0.00924
Irkutsk Region	-0.0589	Sakhalin Region	-0.11088
Kabardino-Balkaria	-0.00212	Sverdlovsk Region	-0.05464
Kaliningrad Region	-0.07651	Sevastopol	-0.05698
Kalmykia	-0.05434	North Ossetia Alania	-0.00453
Kaluga Region	-0.04576	Smolensk Region	-0.01138
Kamchatka Territory	-0.00313	Stavropol Territory	-0.03365
Karachaevo-Cherkessia	-0.00374	Tambov Region	-0.11993
Karelia	-0.06089	Tatarstan	-0.05849
Kemerovo Region	-0.10255	Tver Region	-0.05037
Kirov Region	-0.02238	Tomsk Region	-0.11716
Komi	-0.03621	Tula Region	-0.03143
Kostroma Region	-0.09263	Tyva	-0.00853
Krasnodar Territory	-0.04428	Tyumen Region	-0.07588

Region	$QLI_{\Delta av}$	Region	$QLI_{\Delta av}$
Krasnoyarsk Territory	-0.05809	Udmurtia	-0.09363
Crimea	-0.0212	Ulyanovsk Region	-0.069
Kurgan Region	-0.07369	Khabarovsk Territory	-0.02739
Kursk Region	-0.05522	Khakassia	-0.0664
Leningrad Region	-0.08755	Khanty-Mansiisky Autonomous District	-0.0202
Lipetsk Region	-0.04889	Chelyabinsk Region	-0.7561
Magadan Region	-0.00687	Chuvashia	-0.1696
Marii El	-00412	Chukotka Autonomous District	0.009571
Moscow	-0.01925	Yamalo-Nenetsky Autonomous District	-0.4034
Moscow Region	-0.03322	Yaroslavl Region	-0.8844
Murmansk Region	-0.08631		

(*Source:* Authors calculation)

The obtained data show that negative QLI values dominate in the overwhelming majority of the regions, with the ex caption of the Chukotka Autonomous District, which is the only area where positive online activity index is higher than the negative index value. The overall positive QLI value in Chukotka is due to such indicators as the state of infrastructure, work, and social support from the state, assessment of political decisions and home policy of the authorities. Other indicators are either negative or equal zero. Areas with tentatively "high" QLI, i.e., negative but close to 0 comprise a group of North Caucuses Republics (Ingushetia, Kabardino-Balkaria, Karachaevo-Cherkessia, and North Ossetia Alania) and the Altai Republic. The Pskov and Ryazan regions, the Kamchatka and Perm Territories, and St Petersburg also have a rather low negative QLI. The Orenburg region and several regions in Western Siberia (the Altai Territory, the Tomsk, Novosibirsk and Kemerovo regions) are on the opposite pole of the list with the highest negative QLI. The Volgograd, Saratov, Tambov and Sakhalin regions are also in the same group.

Thus, we see a rather prominent division of Russian regions by types of administrative arrangements. To improve the correlation analysis accuracy, let's divide all regions of the Russian Federation into two groups: A) ethnic republics, autonomous districts and regions, and B) all other – territories, regions and cities of federal status. The results of the correlation analysis of statistical indicators and QLI for the A Group are given in Tables 4 and 5, and for the B Group – in Tables 6 and 7. For perception convenience, we keep the indices that register at least weak relations (the correlation rate $r \geq |0.25|$).

Table 4. Correlation analysis of the statistical indicators characterizing regional economic development, and QLI (the A Group)

	Average wage in the region	**GRP**	**Gini index**	**Per capita monetary income by regions**	**Available resources in urban area**	**Murder and attempted murder**	**Base index of consumer prices**	**Number of unemployed**	**Number of active enterprises**	**Number of " dead" enterprises**	**Consumer price indices**	**Cost-of-living index for particular cities**
Education		-0.46									0.37	
Housing-and-Utilities	-0.36	-0.74	-0.30	-0.47				0.26			0,.9	-0.44
Healthcare		-0.51		-0.26							0.5	
Infrastructure	0.44	0.57	0.53	0.54		-0.29		-0.27	-0.40	-0.43	-0.28	0.52
Safety (urban situation)			-0.26			-0.27				-0.30		
Ecology		-0.50		-0.27							0.38	
Relations between people		-0.27							-0.34	-0.35		
General emotional state						-0.42			-0.73	-0.69		
Job	-0.27	-0.69	-0.36	-0.42	-0.25						0.54	-0.32
Goods	-0.32	-0.73	-0.44	-0.47	-0.32						0.55	-0.36
Taxes		0.36		0.28					-0.28	-0.28	-0.37	0.31
Lending and entrepreneurship						-0.54			-0.51	-0.47		
Social support from the state									-0.34	-0.35	0.39	
Freedom of mass media			-0.27			-0.60		-0.36	-0.54	-0.51		
Protest potential (frustration of the population)	-0.35	-0.78	-0.44	-0.51							0.60	-0.44
Freedom of elections		-0.58	-0.25	-0.32				0.29			0.40	-0.30
Attitude to the authorities	-0.54	-0.77	-0.43	-0.67	-0.33			0.27			0.37	-0.66

	Average wage in the region	GRP	Gini index	Per capita monetary income by regions	Available resources in urban area	Murder and attempted murder	Base index of consumer prices	Number of unemployed	Number of active enterprises	Number of " dead" enterprises	Consumer price indices	Cost-of-living index for particular cities
Political decisions		-0.48					0,34				0.66	
Home policy		-0.59		-0.26							0.60	

(*Source:* Authors calculation)

Table 5. Correlation analysis of the statistical indicators characterizing regional demography, and QLI (the A Group)

	Net migration rate	Population change	Population natural increase rate	Life expectancy at birth	Number of students	Population younger than the working age	Working-age population	Population older than the working age
Education	0.33	0.27						
Housing-and-Utilities	0.25				0.38			
Healthcare	0.28	0.27						
Infrastructure					-0.53			-0.25
Safety (urban situation)			0.36			0.31		-0.38
General emotional state	-0.28				-0.43	0.25		-0.32
Taxes					-0.49			
Lending and entrepreneurship					-0.26			
Social support from the state					-0.25	0.25	0.29	-0.36
Protest potential (frustration of the population)					0.41			
Attitude to the authorities	0.28			0.26	0.56			
Political decisions							0.25	
Home policy					0.27			

(*Source:* Authors calculation)

Table 6. Correlation analysis of the statistical indicators characterizing regional economic development, and QLI (the B Group)

	Average wage in the region	**Gini index**	**Per capita monetary income by regions**	**Available resources in urban area**	**Murder and attempted murder rate**	**Number of unemployed**	**Consumer price indices**	**Cost-of-living index for particular cities**
Education		0.25						
Healthcare					0.33			
Infrastructure	0.29		0.32	0.25				0.33
Safety (urban situation)			0.25				0.28	0.35
Ecology							0.26	
Relations between people			0.26					
General emotional state							-0.26	
Taxes	0.26		0.29	0.31				0.33
Social support from the state							0.27	0.29
Freedom of mass media						-0.41		
Political decisions			0.28				0.25	0.30
Home policy								0.26

(*Source:* Authors calculation)

Table 7. Correlation analysis of the statistical indicators characterizing regional demography, and QLI (the B Group)

	Number of students	**Working-age population**
Ecology	0.25	
Taxes		0.33
Social support from the state		0.33

(*Source:* Authors calculation)

We see two very different pictures for different groups of regions: the A Group regions have some rather strong correlations that cannot be considered a coincidence, between data on social-and-economic and demographic development in a region and online estimates of the quality of life in the same region. Such relations are not registered for the regions from the B Group, the highest correlation rate in the B Group is -0.41. These findings demonstrate practically complete absence of interrelations for the B Group between social-and-economic and demographic development of the regions and online quality-of-life estimates by the residents of regions. It can partly be explained by the considerable heterogeneity of the regions within the B Group on various parameters.

Let's now move to a comparative analysis of QLI and results of sociological surveys. To simplify data perceptions, we did not include the quality-of-life parameters and indices in Table 2, if no significant correlations were found. For example, we did not expose any considerable correlations on such quality-of-life indicators as: education, medicine, infrastructure, safety, relations between people, general emotional state, taxes, social support from the state and freedom of mass media. For such indicators as ecology, we revealed significant correlation with the results of sociological surveys on Social Protest Potential (-0.47); for the "attitude to the authorities" indicator, significant correlation is found with the results of sociological surveys on Social Sentiments Index (0.87) and Expectations Index (0.87); for the Freedom of Elections indicator, correlation is observed with the Social Sentiments Index – "Do you believe that in a year you (your family) will live better or worse than now?" (-0.47). We also did not identify any significant relations between quality-of-life indicators and some sociological surveys – Happiness index, Economic Protest Potential – Quite possible, Economic Protest Potential – Would take part, Political Protest Potential - Quite possible, Political Protest Potential - Would take part, Family Index and Index of Russia. Other results are presented in Table 8 (only values for $p<0.05$ are given).

Therefore, we see that the results of online studies and results of sociological surveys on some aspects demonstrate quite strong and logically explainable correlations. It suggests presence of solid interrelations between these data.

6. Conclusions

In the paper, we have attempted to study relations between the traditional approaches to measuring quality of life on the basis of statistical indicators (objective quality of life) and data of sociological surveys (subjective quality of life) and a new approach to assessing quality of life of the population that relies on social media data. The obtained data lead to several conclusions essential for developing research in this area. As can be noted, we were not able to establish general correlations between the statistical parameters of social-and-economic and demographic development and quality-of-life assessment in a particular region on the basis of social media data. The exceptions are some rather interesting and noticeable interrelations between the level of gross regional product and some quality-of-life indicators for the regions from the A Group (republics, autonomous districts and regions). The result is not surprising: a lot of international studies show no linear correlation between the level of social-and-economic development of a particular society and wellbeing and happiness sentiments of the members of this society (for example, Global End of Year Survey, Happy Planet Index , etc.).

At the same time, we observe good correlations on some quality-of-life parameters measured using digital methods and questionnaires. It encourages particular optimism and reasonable expectations. One can assume that further improvement of quality-of-life measurement on the basis of social media data, comparative studies, when quality of life is researched with digital and questionnaire surveying methods will enable us to create a rather reliable tool to measure quality of life in the near future. In our opinion, it is essential that digital and questionnaire methods capture different sides of human wellbeing: the former registers emotional wellbeing, and the latter - satisfaction with life. The difference between these two edges of wellbeing is that emotional wellbeing, the emotional quality of human daily experience – the frequency and intensity of experiencing joy, stress, sadness, anger and attachment, that make life pleasant or unpleasant, while satisfaction with life means assessing life, general statements, considerations about life made by people when they think about life (Kahneman & Deaton, 2010). Evidently, quality of life measured on the basis of social media data is much closer to emotional wellbeing since it registers human immediate experiences, feelings that overwhelm a person at a particular moment of time. Sociological surveys (Tamulevičienė & Androniceanu, 2020) tend to measure satisfaction with life as questions asked by interviewer's force respondents to contemplate and assess their lives.

Authors Contributions

Conceptualization, Methodology, Investigation, Writing - Original Draft, Evgeniy Shchekotin; Software, Data Curation, Validation, Formal analysis, Viacheslav Goiko; Supervision, Mikhail Myagkov; Writing - Review & Editing, Vitaliy Kashpur.

Conflict of Interest Statement

Authors declare that they do not have any competing financial, professional, or personal interests from other parties.

References

Algan, Y., Murtin, F., Beasley, E., Higa, K., & Senik, C. (2019). Well-being through the lens of the Internet. *PLoS ONE*, *14*(1), e0209562. doi: 10.1371/journal.pone.0211586

Algan, Y., Murtin, F., Beasley, E., Higa, K., & Senik, C. (2019). Well-being through the lens of the Internet. *PLoS ONE*, *14*(1), e0209562. doi: 10.1371/journal.pone.0211586

Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. Retrieved from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Androniceanu, A., Kinnunen, J., & Georgescu, I. (2021). Circular economy as a strategic option to promote sustainable economic growth and effective human development. *Journal of International Studies, 14*(1), 60-73. doi:10.14254/2071- 8330.2021/14-1/4

Androniceanu, A.-M., Kinnunen, J., Georgescu, I., & Androniceanu, A. (2020). A multidimensional approach to competitiveness, innovation and well-being in the EU using Canonical Correlation Analysis. *Journal of Competitiveness*, *12*(4), 5-21. https://doi.org/10.7441/joc.2020.04.01

Appel, H., Gerlach, A. L., & Crusius, J. (2016). The interplay between Facebook use, social comparison, envy, and depression. *Current Opinion in Psychology, 9*, 44-49. doi: 10.1016/j.copsyc.2015.10.006

Bellet, C. and Frijters, P. (2019). Big Data and Well-being. In Helliwell, J., Layard, R., Sachs, J. (Eds.), *World Happiness Report 2019*. New York: Sustainable Development Solutions Network, 97-122.

Burke, M., Marlow, C., & Lento, T. (2010). Social Network Activity and Social Well-Being. *Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, Atlanta, Georgia, USA, April 10-15*, 1909-1912. doi: 10.1145/1753326.1753613

Burrows, R. and Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society*, *1*(1), 1-6. doi: 10.1177/2053951714540280

Chen, L., Gong, T., Kosinski, M., Stillwell, D., & Davidson, R. L. (2017). Building a profile of subjective well-being for social media users. *PloS one*, *12*(11), e0187278. doi: 10.1371/journal.pone.0187278

Chou, H.-T. G. and Edge, N. (2012).'They Are Happier and Having Better Lives than I Am': The Impact of Using Facebook on Perceptions of Others' Lives. *Cyberpsychology, behavior, and social networking*, *15*(2), 117-121. doi: 10.1089/cyber.2011.0324

Ciobanu, A., Androniceanu, A. & Lazaroiu, G. (2019). An integrated psycho-sociological perspective on public employees' motivation and performance. *Frontiers in Psychology*, 10, Article number: 36. Doi: 10.3389/fpsyg.2019.00036

Ciobanu A. and Androniceanu A. (2018). Integrated human resources activities - the solution for performance improvement in Romanian public sector institutions, *Management Research and Practice*, *10*(3), September, 60-79.

Dhall, A., Goecke, R. & Gedeon, T. (2015). Automatic group happiness intensity analysis. *IEEE Transactions on Affective Computing*, *6*(1), 13-26. doi:10.1109/TAFFC. 2015.2397456

Hao, B., Li, L., Gao, R., Li, A. & Zhu, T. (2014). Sensing Subjective Well-Being from Social Media. In Ślęzak, D., Schaefer, G., Vuong, S.T., Kim, Y.S. (Eds.), *Active Media Technology. AMT 2014. Lecture Notes in Computer Science, vol 8610*. Springer, Cham, 324-335. doi: 10.1007/978-3-319-09912-5_27

Haseeb, M., Hussai, H. I., Kot, S., Androniceanu, A. & Jermsittiparsert, K. (2019). Role of social and technological challenges in achieving a sustainable competitive advantage and sustainable business performance. *Sustainability*, *11*(14), Article Number: 3811

Hills, T., Proto, E. & Sgroi, D. (2019). Historical analysis of national subjective well-being using millions of digitized books. *Nature: Human Behaviour*, *3* (12), 1271-1275. doi: 10.1038/s41562-019-0781-5

Kahneman, D. and Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *PNAS, 107 (38),* 16489-16493. doi: 10.1073/pnas.1011492107

Lee, K.-T., Noh, M.-J. & Koo, D.-M. (2013). Lonely people are no longer lonely on social networking sites: The mediating role of self-disclosure and social support. *Cyberpsychology, Behavior, and Social Networking, 16* (6*),* 413-418. doi: 10.1089/cyber.2012.0553

Liebrecht, C., Hustinx, L. & van Mulken, M. (2019). The Relative Power of Negativity: The Influence of Language Intensity on Perceived Strength. *Journal of Language and Social Psychology*, *38*(2), 170-193. doi:10.1177/0261927X18808562

Lupton, D. (2015). *Digital Sociology*. London, New York: Routledge.

Marres, N. (2017). *Digital Sociology: The Reinvention of Social Research.* Cambridge: Polity Press.

McCloskey, W., Iwanicki, S., Lauterbach, D., Giammittorio, D.M. & Maxwell, K. (2015). Are Facebook 'Friends' Helpful? Development of a Facebook-Based Measure of Social Support and Examination of Relationships Among Depression, Quality of Life, and Social Support. *Cyberpsychology, behavior and social networking, 18*(9*),* 499-505. doi: 10.1089/cyber.2014.0538

Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S. & Danforth, C.M. (2013). The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS ONE, 8(5),* e64417. doi: 10.1371/journal.pone.0064417

Rogers R. (2013). *Digital Methods*. Cambridge, MA: MIT Press.

Sabatini, F. and Sarracino, F. (2017). Online Networks and Subjective Well-Being. *Kyklos, 70*(3*),* 456-480. doi: 10.1111/kykl.12145

Schwartz, H.A., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Lucas, R.E., Agrawal, M., Park, G.J., Lakshmikanth, S.K., Jha, S., Seligman, M.E. & Ungar, L. (2013). Characterizing geographic variation in well-being using tweets. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.* Cambridge, MA, 583-591.

Schwartz, H.A., Sap, M., Kern, M.L., Eichstaedt, J.C., Kapelner, A., Agrawal, M., Blanco, E., Dziurzynski, L., Park, G., Stillwell, D., Kosinski, M., Seligman, M.E. & Ungar, L.H. (2016). Predicting individual well-being through the language of social media. *Pacific Symposium on Biocomputing, 21,* 516-527. doi: 10.1142/97898147 49411_0047

Stafford, T. (2014, July 29). Psychology: Why Bad News Dominates the Headlines. *BBC*. Retrieved from https://www.bbc.com/future/article/20140728-why-is-all-the-news-bad

Song, H., Zmyslinski-Seelig, A., Kim, J., Drent, A., Victor, A., Omori, K. & Allen, M. (2014). Does Facebook make you lonely?: A meta analysis. *Computers in Human Behavior, 36,* 446-452. doi: 10.1016/j.chb.2014.04.011

Tamulevičienė, D. and Androniceanu, A. (2020.) Selection of the indicators to measure an enterprise's value and its changes in the controlling system for medium-sized enterprises. *Entrepreneurship and Sustainability Issues, 7*(3), 1440-1458. Doi: https://doi.org/10.9770/jesi.2020.7.3(1)

Trussler, M. and Soroka, S. (2014). Consumer Demand for Cynical and Negative News Frames. *The International Journal of Press/Politics, 19*(3), 360-379. doi: 10.1177/1940161214524832

Valenzuela, S., Halpern, D. & Katz, J.E. (2014). Social network sites, marriage well-being and divorce: Survey and state-level evidence from the United States. *Computers in Human Behavior, 36,* 94-101. doi: 10.1016/j.chb.2014.03.034

Veenhoven, R. (1996). Happy life-expectancy. A comprehensive measure of quality-of-life in nations. *Social Indicators Research, 39,* 1-58. doi: 10.1007/BF00300831

Verduyn, P., Ybarra, O., Resibois, M., Jonides, J. & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review, 11*(1), 274-302.

Wagner-Pacifici, R., Mohr, J. W. & Breiger, R. L. (2015). Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences. *Big Data & Society 2(2),* 1-11. doi: 10.1177/2053951715613810

Wang, N., Kosinski, M., Stillwell, D.J. & Rust, J. (2014). Can Well-Being be Measured Using Facebook Status Updates? Validation of Facebook's Gross National Happiness Index. *Social Indicators Research*, *115*, 483-491. doi:10.1007/s11205-012-9996-9

Wu, K., Ma, J., Chen, Z. & Ren, P. (2015). Analysis of Subjective City Happiness Index Based on Large Scale Microblog Data. In Cheng, R., Cui, B., Zhang, Z., Cai, R., Xu, J. (Eds.), *Web Technologies and Applications. APWeb 2015*. Lecture Notes in Computer Science, vol 9313. Springer, Cham, 365-377. doi: 10.1007/978-3-319-25255-1_30

Yang, C. and Srinivasan, P. (2016). Life Satisfaction and the Pursuit of Happiness on Twitter. *PLoS ONE*, *11*(3), e0150881. doi: 10.1371/journal.pone.0150881