

South Korea's GNI: A Big Data Analysis

Namkil Kang

Professor, Far East University, South Korea

Abstract:

The main goal of this paper is to provide an in-depth analysis of 23 articles of Google written in 2024 concerning South Korea's GNI. A major point of this paper is that the keyword *year* was the most occurred one, followed by the keyword *GNI*, the keyword *capita*, the keyword *period*, the keyword *Korea*, the keyword *income*, and the keyword *bank*, in that order. A point to note is that in the word cloud, the keyword *year* serves as the most central and pivotal one. A further point to note is that the keywords *countries*, *dollars*, *advance*, *Korea*, *expenditure*, and *estimate* are the core ones that consist of 23 articles of Google. The sentiment analysis shows that while the total figure of the positive ones is +382, that of the negative ones is -40. From all of this, it seems quite clear that South Korea's economy is in better shape despite all the negative factors.

Keywords: big data, word cloud, keyword, cluster, sentiment analysis, network

1. Introduction

The main goal of this paper is to provide an in-depth analysis of 23 articles of Google written in 2024 concerning South Korea's GNI. A specific goal of this paper is to provide a sentiment analysis, keyword analysis, network analysis, word cloud analysis, and cluster analysis. First, we aim to provide a detailed analysis of keywords that occurred in 23 articles of Google. We attempt to classify 50 keywords into descending order and describe 50 major keywords that showed up in 23 articles of Google. These 50 keywords showed up frequently in 23 articles of Google and they are the representative ones that consist of 23 articles of Google. Second, we aim at going over a word cloud in which the representative keywords occur. We try to describe the core ones that are made up of 23 articles of Google. Python is supposed to produce the word cloud, depending on the so-called importance of the relevant keywords. Third, we

aim at describing the relevant network that is supposed to mean the links among major keywords. While the core ones are bound to turn up in the center, the less core ones are bound to show up outside the center. From this, it seems quite evident that the so-called central and pivotal words are bound to show up in the center of the map. Fourth, we aim at contemplating one cluster in which the so-called keywords with similar aspects form a cohesive group. Finally, we attempt to provide a detailed sentiment analysis through which we evaluate the status of South Korea's economy.

2. Results

2.1. Samples

In what follows, we briefly illustrate 23 samples that are the first sentences of 23 articles of Google. Take a look at the following samples. Python produced the first sentences of 23 articles of Google:

Table 1 Samples

Number	The First Sentence
1	For the first time, South Korea's Gross Nation...
2	GNI per capita (formerly GNP per capita) is th...
3	South Korea posted a gross national income per...
4	South Korea's gross national income per capita...
5	Per capita gross national income (GNI) last ye...
6	GNI (formerly GNP) is the sum of value added b...
7	Real gross domestic product (real GDP, chained...
8	South Korea outstripped Japan in terms of per ...
9	South Korea's real Gross National Income (GNI)...
10	The gross national income (GNI) per capita in S...
11	Last year, South Korea's gross national income...

12	South Korea's gross national income (GNI) per c...
13	Korea's per capita gross national income (GNI)...
14	Korea's gross national income (GNI) per capita...
15	South Korea's gross national income per capita...
16	Growth in South Korea's real gross national in...
17	The latest value for GNI per capita, Atlas met...
18	Gross national income (GNI) in the first quart...
19	South Korea exceeded Japan's gross national in...
20	Korea's gross national income (GNI) per capita...
21	South Korea's gross national income (GNI) per capita...
22	Korea's per capita gross national income (GNI)...
23	For the first time, South Korea's Gross Nation...

Note that each sentence is made up of 23 articles of Google. After obtaining 23 articles of Google written in 2024 concerning South Korea's GNI, we eliminated many unnecessary words. To be more specific, we got rid of adjectives, relative pronouns, adverbs, definite articles, indefinite articles, to-infinitives, gerunds, etc. Put differently, we obtained nouns only after eliminating many unnecessary words. Only nouns suffice for five analyses.

2.2. Keywords

This section is devoted to contemplating the so-called pivotal keywords that turned up in 23 articles of Google. Have a look at the following fifty keywords, which we classified into descending order by using python:

Table 2 Keywords

Number	Keyword	Frequency	Number	Keyword	Frequency
1	year	165	26	national	26
2	GNI	157	27	point	26
3	capita	115	28	service	26
4	percent	110	29	investment	24
5	Korea	108	30	statistics	23
6	income	68	31	product	22
7	bank	60	32	accounts	21
8	gdp	53	33	goods	21
9	south	52	34	rate	21
10	quarter	47	35	country	20
11	rate	46	36	Taiwan	20
12	growth	45	37	construction	19
13	Koreas	44	38	figure	19
14	US	37	39	population	19
15	economy	33	40	consumption	18
16	gross	32	41	won	18
17	years	32	42	percentage	17
18	currency	31	43	data	16
19	exchange	31	44	Choi	14
20	dollar	30	45	ratio	14
21	increase	30	46	series	14
22	time	27	47	base	13
23	value	27	48	government	13
24	exports	26	49	Korean	13

dollars domestic due earnings economic economies economists economy ecosbokorkr education effect effects eg electricity employees end engines enterprises environment equipment estate estimate estimates estimation euro event exchange expansion expenditure expenditures export exports face facilities facility fact factor factors fall february federal figure figures final finance fishing flight fluctuations food forecast forecasts forestry formation france funds gambling gas gauge gdp germany gnp goals goods government governor greenback gross gu hand head headquarters health height home household households illegal importance imports improvements includes income income1 incomes increase increases indicator indicators industries industry inflation inflations information infrastructure input institutions insurance intellectual intelligence interest international investment investments island it italy january japan japans jeong jump june jung kim kingdom korea korean koreans koreas krw kyung lead lee level line livelihood living lockdown look losses manufacturing march margin mark market measure measurement meetings memory method methods mexico mid milestone minister ministry mok money month months nation national nations new news no nominal note notes october oecd official officials oil optimism outbound output overtake pandemic parentheses payments pension people peoples per percent percentage performance period person petroleum phase photo place point points policies policy poor population populations possibility power pre preconditions preliminary president press price prices prime private producers product product1 production products projections projects promising property prostitution q1 q2 q3 q4 quarter quarterly quarters range rankings rate rates rating ratio ratios reading real rebound receipts recession recovery reduction reference refers reforms reorganization report reports reserves resident residents respects rest result results revamped revision rise sa sale sang savings scope search sector security self semiconductor semiconductors seoul september series services services1 sets shift shipments shrank side sixth size slowdown source south spending spike stabilizes stagnation standard standards statement statements states statistics stats status stays stood storage strategy study subsidies suk sum sun super supply support surpassing sustainability system table tables taiwan taiwanese taiwans target targets taxes team technology tensions term terms thousand thursday tightening time timing tore trade trading transactions transportation trend tuesday types uk uncertainties united us use use valuation value values variety various volume volumes water weakness wealth website wednesday whereas woes won work world years yen yens yeol yong yonhap yoon

As indicated in Table 3, these keywords have one thing in common. That is to say, they are closely related to one another. The so-called cluster is somewhat larger than topics, but somewhat smaller than a document. Keywords consist of topics, which constitute a document. The so-called cluster that means a group has similar properties in common, hence differing to topics. Python produced about seven clusters, but we provided

one of them.

2.6. The Sentiment Analysis

This section is focused on providing the sentiment analysis of 23 articles of Google. We classified each word into a positive one and a negative one. After classifying each keyword into two types, we multiplied the frequency of the relevant keyword by +1 or -1. Have a look at Table 4:

Table 4 The Sentiment Analysis

Number	Keyword	Frequency	Positive or Negative	Score
1	GNI	157	positive	157
2	income	68	positive	68
3	growth	45	positive	45
4	export	26	positive	26
5	Japan	26	positive	26
6	Taiwan	20	positive	20
7	population	19	negative	-19

8	depreciation	12	negative	-12
9	inflation	9	negative	-9
10	product	9	positive	+9
11	rise	8	positive	+8
12	trade	8	positive	+8
13	recovery	7	positive	+7
14	expansion	4	positive	+4
15	semiconductors	4	positive	+4

It must be emphasized that while the number of the positive keywords is 12, that of the negative keywords is 3. More specifically, while the total figure of the positive ones is +382, that of the negative ones is -40. Consequentially, this leads to the figure of +342. From all of this, it seems quite evident that South Korea's economy is in better shape despite all the negative factors. For the analyses of AI and big data, see Kang (2024a, 2024b, 2024c, 2024d, 2024e, 2024f, 2024g).

3. Conclusion

To sum up, we have provided an in-depth analysis of 23 articles of Google written in 2024 concerning South Korea's GNI. In section 2.2, we have maintained that the keyword *year* was the most occurred one, followed by the keyword *GNI*, the keyword *capita*, the keyword *period*, the keyword *Korea*, the keyword *income*, and the keyword *bank*, in that order. In section 2.3, we have shown that in the word cloud, the keyword *year* serves as the most central and pivotal one. In section 2.4, we have contended that the keywords *countries*, *dollars*, *advance*, *Korea*, *expenditure*, and *estimate* are the core ones that consist of 23 articles of Google. In section 2.5, we have argued that 23 articles of Google let us know how good South Korea's economy was despite all the negative factors.

References

- [1] Kang, N (2024a). A Big Data Analysis of a Hot Political Issue. *Studies in Linguistics* 70, 149-165
- [2] Kang, N (2024b). The Ukraine War in BBC News: A Big Data Analysis. *International Journal of Applied Engineering and Technology* 6(1), 1548-1555
- [3] Kang, N (2024c). A Study of the Prediction of the Model Logistic Regression in Machine Learning: Focusing on a Survey. *Semiconductor*

Optoelectronics 43(1), 86-94

- [4] Kang, N (2024d). The Spring of Seoul: A Big Data Analysis. *Semiconductor Optoelectronics* 43(1), 264-272
- [5] Kang, N (2024e). A Study of the Prediction of the Three Models Logistic Regression, Gaussian Naive Bayes and Multi-Layer Perceptron Classifier in Machine Learning. *Semiconductor Optoelectronics* 43(1), 253-263
- [6] Kang, N (2024f). An Analysis of the Prediction Power of Four Models in Machine Learning and Deep Learning: Focusing on the Reviews of the Movie Noryang. *A Jungang Journal of English Language and Literature* 66(1), 101-122
- [7] Kang, N(2024g). An Association Analysis of Shopping Items in Machine Learning. *Stochastic Modelling and Computational Sciences* 4(1), 97-109