

<https://doi.org/10.7236/JIIBC.2024.24.2.185>
JIIBC 2024-2-27

Elasticsearch를 이용한 부동산 시장 가격 분석 및 시각화

Analysis and Visualization of Real Estate Market Price using Elasticsearch

황승연*, 김정준**

Seung-Yeon Hwang*, Jeong-Joon Kim**

요약 2022년 대한민국의 부동산 시장이 하락되는 모습을 볼 수 있다. 이에 따른 원인에는 코로나19와 러시아의 우크라이나 침공이 가장 큰 원인으로 꼽히고 있다. 이 둘의 문제로 경기 침체에 불을 지폈으로써 물가가 떨어지면서 그이후로 환율과 금리 등이 높아지는 문제가 발생하였다. 기존에 활발했던 부동산 시장이 앞서 말한 문제들 때문에 실거래수가 줄어들어 높은 이자로 인해서 부동산 시장이 하락하는 모습을 볼 수 있다. 공공데이터 포털, KOSIS와 서울특별시에서 제공하는 데이터를 Logstash로 수집해서 Elasticsearch로 전달해 Kibana에서 제공하는 대시보드 기능을 이용해 인플레이션, 환율, 대출금리를 시각화로 나타내 원인들을 분석하고 결과를 도출했다. 그리고 서울특별시에서 가장 실거래수가 많은 노원구, 가장 적은 종로구의 특정 아파트 3개를 골라 매 월마다 변하는 실거래가를 Data Table로 나타냈다.

Abstract In 2022, we can see the real estate market in Korea going down. Corona 19 and the Russian invasion of Ukraine are cited as the biggest causes for this. These two problems ignited the economic recession, causing prices to fall and subsequently raising exchange rates and interest rates. Due to the aforementioned problems in the previously active real estate market, the number of actual transactions has decreased, resulting in a decline in the real estate market due to high interest rates. Data provided by the public data portal, KOSIS, and the Seoul Metropolitan Government were collected through Logstash, transferred to Elasticsearch, and visualized inflation, exchange rates, and loan interest rates using the dashboard function provided by Kibana, to analyze causes and derive results. In addition, three specific apartments in Nowon-gu and Jongno-gu, which have the highest number of actual transactions in Seoul, are selected and the actual transaction prices that change every month are displayed in the Data Table.

Key Words : Elasticsearch, Kibana, Logstash, Beats, Elastic Stack

*준회원, 안양대학교 컴퓨터공학과
**정회원, 안양대학교 ICT융합학부 소프트웨어전공
접수일자 2024년 2월 2일, 수정완료 2024년 3월 2일
게재확정일자 2024년 4월 5일

Received: 2 February, 2024 / Revised: 2 March, 2024 /
Accepted: 5 April, 2024

**Corresponding Author: jkim@anyang.ac.kr
Dept. ICT Convergence Engineering at Anyang University,
Korea

I. 서 론

대한민국의 부동산 시장을 살펴보면 하락하는 모습을 볼 수 있다. 이로 인해 다양한 방법으로 부동산 가격 변동등을 예측하거나 가상부동산에 관심을 보이기도 한다.^{[1][2]} 2020년도부터 지금까지 코로나19가 전세계의 경제를 타격했다. 그 결과로 출근하는 직원의 수가 줄어들며 공장의 생산성 또한 줄어들었다. 즉, 사회적인 공급이 줄어든 셈이다. 또한 2022년에는 러시아의 우크라이나 침공으로 인해 세계 경제에 큰 타격을 받았다. 앞에 말한 두 사례로 인해서 경기 침체에 불을 지폈으로써 물가 상승에 주요 원인이 되었다. 그래서 물가 안정을 위해서 대출금리가 인상된 것이다. 부동산 시장 급락의 주원인으로는 대출금리의 인상으로 뽑히고 있다. Elasticsearch는 검색 엔진도 가능하며 분석시스템과 통계분석에 특화되어있다. 사례로 엘라스틱 서치로 최적의 거리에 여행지 코스 추천을 해주는 시스템^[3]과 유의어 사전같은 검색 엔진도 만들 수 있다^[4]. 또한 정보 시스템의 위협에 벗어나고자 Logstash를 이용하여 보안로그 분석시스템을 만들기도 한다^[5]. 본 시스템은 Elasticsearch를 이용해 공공데이터 포털, KOSIS와 서울특별시에서 제공하는 데이터를 CSV 파일로 변환 후 Elasticsearch에 업로드를 통해 Kibana로 부동산 시장통계와 하락하는 원인들을 시각화로 통해 많은 데이터를 가공하고 분석하여 보기쉽게 보여줄 것이다.^[6]

II. 관련연구

1. Elastic Stack^[7]

일반적으로 데이터 작업 수행은 수집, 가공, 분석, 관리, 시각화 등의 단계를 가지고 있다. Elastic Stack은 앞서 말한 작업 수행을 할 수 있는 오픈소스 데이터 분석 플랫폼이다. Elastic Stack 구조를 살펴보면 Kibana, Elasticsearch, Beats, Logstash로 데이터를 다루는 기술들이 속해있다. Elasticsearch는 저장과 처리, Kibana는 시각화, Beat와 Logstash는 데이터를 수집하는 목적을 가지지만, 본 연구에서는 데이터를 가공해 원하는 데이터만을 수집하기 위해서 Logstash를 사용했다.

2. Elasticsearch^[8,9]

ElasticSearch는 Apache Lucene 기반의 Java 오픈

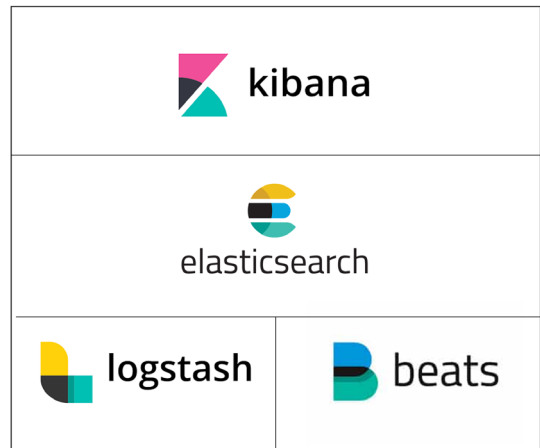


그림 1. Elastic Stack 구조
Fig. 1. Elastic Stack structure

소스 분산형 RESTful 검색 엔진으로 NoSQL 기반의 문서 지향 데이터베이스이다. Kibana, Logstash없이도 단독으로 사용이 가능하며 주로 빅데이터 같이 많은 양의 데이터를 실시간으로 검색, 처리, 저장을 수행할 수 있다. Elasticsearch는 여러 노드를 가질 수 있으며 REST API를 이용해 다른 서버와 HTTP를 통해 통신이 가능하다. 또한 DBMS와 연동이 가능해서 Beats와 Logstash와 같은 수집기로 데이터를 DBMS에 저장할 수 있다.

3. Kibana^[10]

Kibana는 Elasticsearch와 함께 동작 되도록 설계된 오픈소스 분석 및 시각화 플랫폼이다. 데이터의 형태를 만들고, Elastic Stack을 탐색할 수 있게 하는 시각화 및 관리를 할 수 있다. 가지고 있는 데이터로 바 그래프, 파이 그래프 등의 다양한 Chart 종류를 가지고 있으며, 지도를 통한 시각 자료도 제공한다.

4. Logstash

Logstash는 csv, DB 등 다양한 소스에서 로그 등의 데이터를 수집, 집계, 파싱하여 Elasticsearch로 전달한다. 즉, 서버측 데이터 처리 파이프 라인으로 데이터 수집 엔진이다. 데이터를 필요한 모양으로 가공이 가능하며 파이프라인을 컨트롤 하기 위한 도구이다. 동작과정은 3단계로 구성된다. Input에는 Logstash가 수집할 경로를 입력한다. Filter는 데이터를 가공하는 곳이다. 마지막으로 Output은 Logstash가 보낼 곳을 명시한다.

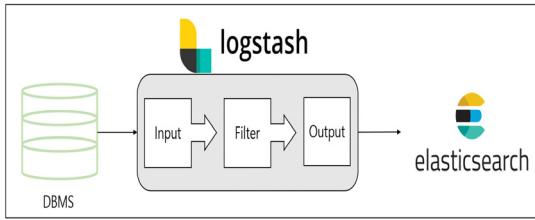


그림 2. Logstash 동작 구조
 Fig. 2. Logstash action structure

III. 실험 방법 및 결과

1. 실험환경

본 시스템을 구현하기 위해 Elasticsearch를 통해서 통계분석 시스템을 만들었다. 운영체제는 Window11과 CPU는 AMD 라이젠5-4세대 5600G를 사용했다. Elasticsearch, Kibana와 Logstash는 8.2.3버전을 이용했다.

2. 실험 방법

공공데이터 포털, KOSIS에서 제공하는 인플레이션, 환율, 대출금리 데이터와 서울특별시에서 제공하는 실거래수와 매매가 데이터를 이용하여 Logstash를 통해 데이터를 수집 및 파싱하여 Elasticsearch로 전달해 Kibana를 이용해 시각화를 제공한다.

3. 실험 결과

부동산 시장에 영향을 주었던 인플레이션, 달러 환율, 금리 등의 원인들과 서울특별시에서 가장 실거래수가 많은 노원구 가장 적은 종로구 2곳의 선정해 아파트 매매 실거래가를 Kibana를 이용해 대시보드로 나타냈다.

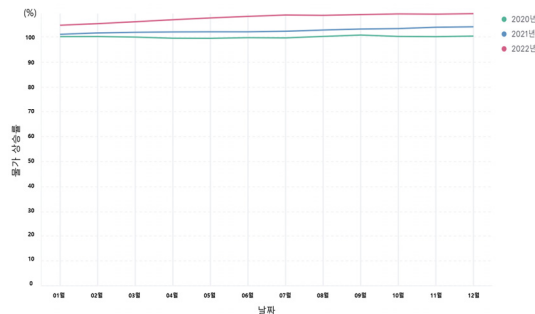


그림 3. 인플레이션 대시보드
 Fig. 3. Inflation Dashboard

그림3은 Kibana에서 제공하는 Chart의 종류 중 Line Chart를 이용하였다. 그림3을 보면 해를 거듭할수록 물가가 지속적으로 올라가는 모습을 볼 수 있다. 특히, 2022년에 러시아의 우크라이나 침공으로 인해서 2020년에서 2021년 증가률보다 2021년에서 2022년 증가률이 더 높은 것으로 육안으로 쉽게 확인해볼 수 있다. 또한, 2022년을 보면 날이 갈수록 물가가 증가하는 모습도 볼 수 있다.

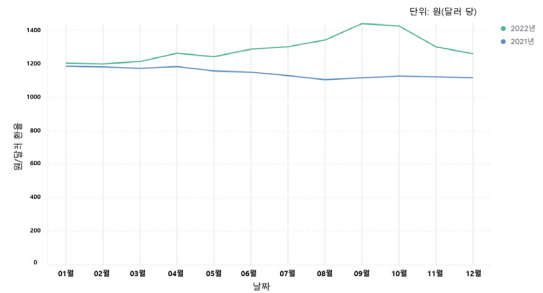


그림 4. 달러 환율 변동 대시보드
 Fig. 4. Dollar Exchange Rate Changes Dashboard

그림 4는 달러 환율 변동 그래프이다. 2021년과 2022년을 비교해보면 2022년이 2021년보다 높은 것을 볼 수 있다. 특히 2022년 5월부터 지속적으로 증가하면서 9월과 10월에는 1달러에 1400원을 넘는 모습을 볼 수 있다.

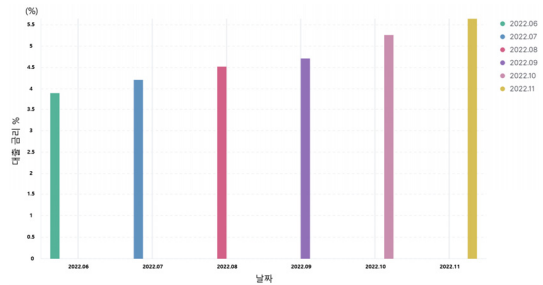


그림 5. 대출 금리 변동 대시보드
 Fig. 5. Dashboard of changes in loan interest rates

그림 5는 Kibana에서 제공하는 Chart의 종류 중 Bar Chart를 이용하였다. 그림 5를 보면 6월부터 11월 까지의 대출 금리의 그래프이다. 월마다 약 0.5%씩 증가하는 것을 볼 수 있다. 이는 부동산 시장에 큰 영향을 미친다. 왜냐하면 보통 매매를 하기 위해서는 자신의 자본만으로 사는 것이 아닌 자본과 대출을 통해서 매매를 하기 때문이다.

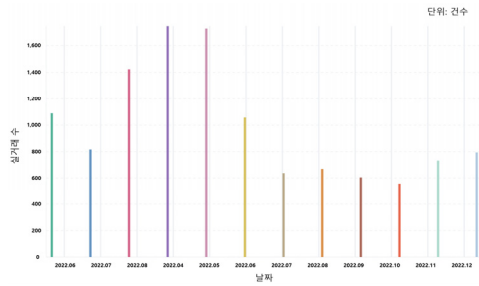


그림 6. 서울특별시 아파트 매매 실거래 수 대시보드
Fig. 6. Dashboard of actual number of sales of apartments in Seoul

그림 6은 2022년 서울특별시 아파트 매매 실거래 수를 나타낸 것이다. 3월부터 5월까지의 활발한 거래량을 나타냈다. 그러나 그 이후로는 약 800 이하의 거래량을 나타냈다. 이는 그림 3에서 6월부터 11월까지 대출금리가 꾸준히 증가함에 따라 아파트 매매 실거래 수가 떨어진 것이다. 왜냐하면 높은 금리에 대출을 받으면 갚아야 할 이자가 감당이 안되기 때문이다. 그래서 대출이자는 아파트 매매 실거래 수에 영향을 미친다는 것을 알 수 있다.

그림6에서 3월부터 5월까지 활발한 거래량 있다는 것을 알 수 있다. 이를 기준으로 서울특별시에서 3월부터 5월까지 가장 거래량이 많은 노원구의 실거래가와 적은 종로구의 실거래가의 데이터를 가져와서 Kibana에서 제공하는 Chart의 종류 중 Data Table을 이용해 인플렉션, 달러 환율, 대출 금리와 서울 특별시 아파트 매매 실거래 수와 실거래 가의 변화와 관련이 있는지 알아보기 위해서 비교해 보았다.

계약일	지역	거래금액(단위: 만원)
03.07	노원A	80,000
03.26	노원A	82,200
04.11	노원A	83,000
04.16	노원A	82,000
04.20	노원A	85,000
05.21	노원A	82,500
07.25	노원A	72,000
07.26	노원A	70,500
08.12	노원A	75,000
08.30	노원A	67,000
12.05	노원A	58,400

그림 7. 노원구 A 아파트 실거래가 대시보드
Fig. 7. Nowon-gu A Apartment Actual Transaction Price Dashboard

계약일	지역	거래금액(단위: 만원)
03.18	노원B	82,000
04.01	노원B	80,000
04.08	노원B	79,800
04.11	노원B	82,000
04.14	노원B	80,000
08.20	노원B	69,500
09.07	노원B	64,500
10.07	노원B	61,000
12.05	노원B	58,500

그림 8. 노원구 B 아파트 실거래가 대시보드
Fig. 8. Nowon-gu B Apartment Actual Transaction Price Dashboard

계약일	지역	거래금액(단위: 만원)
01.13	노원C	51,800
03.19	노원C	49,900
04.16	노원C	52,000
04.28	노원C	45,000
05.09	노원C	48,500
05.28	노원C	55,300
06.30	노원C	55,000
08.04	노원C	47,000

그림 9. 노원구 C 아파트 실거래가 대시보드
Fig. 9. Nowon-gu C Apartment Actual Transaction Price Dashboard

그림 7, 8, 9는 2022년 서울특별시 노원구 특정 아파트 3개의 실거래가를 가져왔다. 1년 중에 보통 1분기에 가장 높은 가격으로 매매되는 것을 볼 수 있다. 그러나 C 아파트를 제외하고 3, 4분기에는 1분기에 비해서 가격이 떨어진 모습을 볼 수 있다.

그림 10, 11, 12는 2022년 서울특별시 종로구 특정 아파트 3개의 실거래가를 가져왔다. 종로구 A아파트의 경우 1월에는 높은 가격에 매매되었으나 그 이후로는 3천에서 4천이상 떨어진 모습을 볼 수 있다. 이외에 종로구 B와 C 아파트는 매매가가 큰 변화가 없다.

계약일	지역	거래금액(단위: 만원)
01.06	종로A	15,700
02.19	종로A	11,800
05.21	종로A	12,450
09.26	종로A	12,500
09.28	종로A	12,510
10.10	종로A	12,350
10.13	종로A	12,500

그림 10. 종로구 A 아파트 실거래가 대시보드
Fig. 10. Jongno-gu A Apartment Actual Transaction Price Dashboard

계약일	지역	거래금액(단위: 만원)
01.03	종로B	19,500
01.08	종로B	19,000
02.23	종로B	19,500
04.26	종로B	18,000
05.13	종로B	16,000
09.07	종로B	16,000
11.15	종로B	18,600
12.30	종로B	20,200

그림 11. 종로구 B 아파트 실거래가 대시보드
 Fig. 11. Jongno-gu B Apartment Actual Transaction Price Dashboard

계약일	지역	거래금액(단위: 만원)
03.18	종로C	13,000
06.17	종로C	13,000
07.29	종로C	13,200
08.22	종로C	13,300
10.05	종로C	13,400
11.05	종로C	12,900

그림 12. 종로구 C 아파트 실거래가 대시보드
 Fig. 12. Jongno-gu C Apartment Actual Transaction Price Dashboard

V. 결 론

본 논문은 대한민국의 부동산 시장이 하락하는 원인들의 데이터와 서울특별시에서 아파트 매매가 많은 노원구 아파트와 매매가 적은 종로구 아파트 매매가 데이터를 Logstash로 수집하고 Elasticsearch로 전달해 Kibana에서 제공하는 대시보드 기능을 이용해 시각화를 나타냈다. 코로나와 러시아의 우크라이나 침공으로 인해서 2022년에 물가가 상승 되면서 높은 금리를 기록한 것을 볼 수 있었다. 이는 부동산 시장에도 큰 영향을 미쳤다. 2022년 1분기에 비해서 3, 4분기에는 낮은 가격으로 매매되는 것을 볼 수 있었다. 이는 금리가 큰 영향을 미쳤다. 2021년에 대출을 받아 아파트를 매매한 사람들이 현재의 금리 때문에 대출이자를 감당하지 못하고 싸게 내놓는 경우가 많아졌다. 이로 인해서 3, 4분기에 아파트 매매 실거래 수가 줄어든 것을 알 수 있었다. 이처럼 Elasticsearch를 통해 Kibana에서 제공하는 대시보드로 많은 양의 데이터를 손쉽게 보면서 정확한 데이터 분석 결과를 도출한 것을 볼 수 있었다. 이후 2023년에 발생하는 이슈들에 따라 부동산 시장에 영향을 준다면 추후적으로 시각화로 나타낼 예정이다.

References

- [1] Jongyoung Yoo, "A Study on Transaction Service of Virtual Real Estate based on Metaverse", The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 22, No. 2, pp.83-88, Apr. 30, 2022.
- [2] Sanghyun Nam, Taeho Han, Leeju Kim, Eunji Lee, "Real-Estate Price Prediction in South Korea via Machine Learning Modeling", The Journal of The Institute of Internet, Broadcasting and Communication (IIBC), Vol. 20, No. 6, pp.15-20, Dec. 31, 2020.
- [3] Jun-Yeong Kim, Kyeong-Ho Jo, Jun Park, Se-Hoon Jung, Chun-Bo Sim, "A Development of Optimal Travel Course Recommendation System based on Altered TSP and Elasticsearch Algorithm", Journal of Korea Multimedia Society, Vol. 22, No. 9, pp. 1108-1121, 2019.
DOI: <https://doi.org/10.9717/kmms.2019.22.9.1108>
- [4] PIAO XIANGHUA, YIN HELIN, Yeong Hyeon Gu, Seong Joon Yoo, "Design of environmental technology search system using synonym dictionary", Proceedings of the Korean Society of Broadcast Engineers Conference, pp. 582-586, 2020.
- [5] Bong-Hwan Lee, Dong-Min Yang "A Security Log Analysis System using Logstash based on Apache ElasticSearch", Journal of the Korea Institute of Information and Communication Engineering, Vol. 22, No. 2, pp. 382-389, 2018.
DOI: <https://doi.org/10.6109/jkiice.2018.22.2.382>
- [6] DaeKyeong Park, Dongkyoo Shin, Dongil Shin, "Host-based intrusion detection research using CNN and Kibana", Proceedings of the Korea Information Processing Society Conference, pp. 920-923, 2020.
DOI: <https://doi.org/10.3745/PKIPS.y2020m11a.920>
- [7] Elasticsearch's official web page, Available at <https://www.elastic.co/kr/>
- [8] Elastic Search: Introduction, Basics, Architecture and Usage of Elastic Search [Internet], Available: <https://hassantariqblog.wordpress.com/>.
- [9] "Big Data Processing for Full-Text Search and Visualization with Elasticsearch", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 12, 2017.
DOI: <https://doi.org/10.14569/IJACSA.2017.081211>
- [10] Seong Woo Hwang Bo, Kang Sun Lee, Yong Jun Kwon, "A Simulation Output Analysis Environment by utilizing Elastic Stack", Journal of the Korea Society for Simulation, Vol. 27, No. 3, pp. 65-73, 2018.
DOI: <https://doi.org/10.9709/JKSS.2018.27.3.065>

저 자 소 개

황 승 연(준회원)



- Seung-Yeon Hwang received his BS in Department of Computer Science at Korea Polytechnic University in 2019. He is currently studying MS in Department of Computer Science at Anyang University. His research interests include Database System, Big Data, Data Analysis, Machine Learning, etc.

김 정 준(정회원)



- Jeong Joon Kim received his BS and MS in Computer Science at Konkuk University in 2003 and 2005, respectively. In 2010, he received his PhD in at Konkuk University. He is currently a professor at the department of Computer Science at Anyang University. His research interests include Database Systems, Big Data, Semantic Web, Geographic Information Systems (GIS) and Ubiquitous Sensor Network (USN), etc.

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2022R1F1A1062953).