

이 보고서는 코스닥 기업에 대한 투자정보 확충을 위해 발간한 보고서입니다.

혁신성장품목보고서

YouTube 요약 영상 보러가기

데이터시각화

빅데이터 중요성 부각에 따른
데이터시각화 관심 증가

요약

배경기술분석

심층기술분석

산업동향분석

주요기업분석



작성기관

(주)NICE디앤비

작성자

정혜윤 전문위원

- 본 보고서는 「코스닥 시장 활성화를 통한 자본시장 혁신방안」의 일환으로 코스닥 기업에 대한 투자정보 확충을 위해, 한국거래소와 한국예탁결제원의 후원을 받아 한국IR협의회가 기술신용 평가기관에 발주하여 작성한 것입니다.
- 본 보고서는 투자 의사결정을 위한 참고용으로만 제공되는 것이므로, 투자자 자신의 판단과 책임하에 종목선택이나 투자시기에 대한 최종 결정을 하시기 바랍니다. 따라서 본 보고서를 활용한 어떠한 의사결정에 대해서도 본회와 작성기관은 일체의 책임을 지지 않습니다.
- 본 보고서의 요약영상은 유튜브로도 시청 가능하며, 영상편집 일정에 따라 현재 시점에서 미게재 상태일 수 있습니다.
- 카카오톡에서 “한국IR협의회” 채널을 추가하시면 매주 보고서 발간 소식을 안내 받으실 수 있습니다.
- 본 보고서에 대한 자세한 문의는 작성기관(TEL.02-2122-1300)로 연락하여 주시기 바랍니다.



한국IR협의회



데이터시각화

빅데이터 중요성 부각에 따른 데이터시각화 관심 증가

디지털 뉴딜 – D·N·A(Data, Network, AI) 생태계 강화

K-뉴딜의 10대 대표과제 중 「데이터 댐」 과제가 포함되어 있음.

- 데이터 수집·가공·거래·활용기반을 강화하여 데이터 경제를 가속화하고 5세대 이동통신(5G) 전국망을 통한 전 산업 5세대 이동통신(5G)·인공지능 융합 확산
- 분야별 빅데이터 플랫폼 확대, 공공데이터 14.2만 개 신속 개방, 인공지능학습용 데이터 1,300종 구축 등 데이터 확충
- 2025년까지 총사업비 18조 1천억 원 투자, 일자리 38만 9천 개 창출

정보통신(F) – 지능형데이터분석(F30) – 데이터시각화(F30002)

- 데이터시각화(Data Visualization)는 나열된 데이터를 지식화하기 위한 과정으로, 데이터가 의미하는 바를 직관적으로 이해할 수 있도록 표현하는 기술임
- 데이터시각화는 대량의 데이터를 동시에 전달할 수 있도록 지원하고, 새로운 관점의 지각적 추론을 가능하게 하며, 정보의 빠른 확산을 유도할 수 있음
- 최근 데이터시각화 기술은 빅데이터, 인공지능 가상현실 기술 등과 접목되어, 대량의 데이터를 효과적으로 전달하는 방향으로 발전하고 있음

■ 데이터시각화를 통한 빅데이터의 효과적인 정보 전달

빅데이터의 분석은 데이터 기반 의사결정에 도움을 주기 위한 것이다. 빅데이터는 단순한 선형적 구조로 표현하는 것이 어렵기 때문에 각 데이터의 특성에 맞는 시각화 방식을 이용하여 인사이트를 전달하는 것이 매우 중요하다. 다양한 시각화 방식을 통해 정보 전달 시간을 단축하고, 빅데이터에서 도출된 유의미한 정보를 효과적으로 전달할 수 있다. 빅데이터 시각화는 정보 시각화를 기반으로 하고 있으며, 전달하고자 하는 데이터의 특성에 따라 시간 시각화, 분포 시각화, 관계 시각화, 비교 시각화 등이 활용될 수 있다.

■ 디지털 뉴딜을 기반으로 한 데이터시각화 시장 상승 기대

한국형 디지털 뉴딜 정책을 통해 국내 데이터 산업이 활성화되고 있으며, 국내외 글로벌 기업들의 데이터 활용 관련 전략이 쏟아지고 있다. 데이터 산업 중에서도 빅데이터 분석 시장이 미국을 중심으로 활발하게 움직이고 있으며, 빅데이터 분석에 따른 데이터시각화의 중요성도 커지고 있다. 데이터시각화 시장에 동반하여 데이터시각화 툴 시장 성장세도 기대된다. 코로나19 바이러스로 인해 데이터시각화 툴의 사용이 증가하였으며, 마케팅 및 판매에 데이터시각화 툴의 활용도가 높아, 마케팅 및 판매에서의 데이터시각화 툴 사용이 시장 성장세에 가장 큰 기여를 할 것으로 전망된다.



I. 배경기술분석

4차 산업혁명을 주도하는 빅데이터 기술

빅데이터가 4차 산업혁명과 직결되면서 빅데이터 분석이 국가 경쟁력의 한 축으로 대두되고 있다. 또한, 빅데이터의 분석에서 더 나아가 분석 결과를 사용자들에게 전달하는 시각화의 중요성도 함께 부각되고 있다.

■ 4차 산업혁명을 주도하는 빅데이터 기술

빅데이터는 규모(Volume), 다양성(Variety), 변동성(Velocity, 데이터 변동 속도)이 모두 커서 기존의 방식으로는 분석하기 어려운 디지털 데이터를 의미한다. 여기서 말하는 기존 방식이란 숫자 데이터를 정형화된 형태로 저장하고 모집단의 특성과 추세를 통계적으로 분석하는 것을 의미한다. 빅데이터는 수치화가 어렵고 형태가 다양하고 빠르게 변화하는 비계량적인 비정형 변수들을 포함한다. 빅데이터 분석은 이렇게 비정형 데이터들에 대한 분석이 주를 이루기 때문에 분석 결과가 특정 값으로 나타나기보다는 시각화된 형태로 나타나는 경우가 많다.

빅데이터는 4차 산업혁명의 핵심 요소로 자리 잡고 있다. 빅데이터는 특히 인공지능, 사물인터넷, 증강현실, 가상현실, 자율주행차, 5세대 이동통신 등 향후 주목받는 신산업의 필수적인 요소로 평가되고 있다. 이에 따라 세계 각국은 데이터 수집 및 분석 기술 활성화에 수많은 노력을 기울이고 있다. 한국도 데이터 3법 개정과 디지털 뉴딜 추진을 시작으로 정책 및 제도 개선을 위한 노력을 기울이고 있다. 예를 들어, 한국 정부는 민간 수요가 높은 국세청 등 정부 기관의 핵심 데이터를 개방하거나 인공지능 모델 학습을 위한 학습용 데이터를 정부 주도로 구축하는 등의 정책을 펼치고 있다. 즉, 빅데이터 기술 수준이 4차 산업혁명의 신기술 수준, 그리고 신산업 발전에 직결되면서 데이터와 데이터 가공 기술은 국가 산업 발전의 경쟁력이 되고 있다.

[그림 1] 4차 산업혁명 대표 기술



*출처: 세계경제포럼(WEF) '직업의 미래'



데이터를 지배하는 기업이 4차 산업혁명을 주도할 것이라는 예측이 나오고 있으며, 데이터를 경쟁자보다 더 잘 활용하는 기업들이 2020년대의 산업을 주도하고 있다. 아마존의 경우 이용자의 패턴을 빅데이터로 분석, 예측하고 배달해주는 기술을 적용하였으며, 지멘스의 경우 스마트 공장의 센서 데이터를 이용하여 불량을 최소화하는 기술을 활용하고 있다. 데이터 활용뿐만 아니라 데이터 생태계를 주도하려는 기업들의 노력도 상당히 진행되고 있다. 글로벌 대기업들은 스스로 데이터를 확보할 수 있는 생태계를 구축하고, 참여 기업에 이익을 공유하는 방식으로 비즈니스를 확립 중이다. 구글, 아마존, 애플, 페이스북 등은 자사 플랫폼과 기업, 가입자 데이터를 연결하여 지속적으로 사업 영역을 확장시키고 있다.

빅데이터는 현재 고도화 단계로 데이터 활용 및 데이터를 기반으로 한 사업화 추진에 초점을 두고 있다. 2000년대 후반에는 빅데이터 발전 초기 단계로 데이터 수집과 인프라 구축에 중점을 두고 있었으며, 2010년대 초반에는 빅데이터 발전 단계로 데이터 분석 기술 강화에 중점을 두었다. 한편, 2010년대 초반까지 IT, 통신 및 금융 산업에서 빅데이터를 주로 활용했던 것과 달리, 2010년대 중반 이후부터는 전 산업에서 빅데이터 활용성을 논의 중이다.

■ 빅데이터 분석과 시각화

빅데이터 기술은 데이터를 수집, 저장, 처리, 분석하고, 그리고 분석된 데이터를 시각화하는 기술을 의미한다. 빅데이터는 기존의 단순한 선형적 구조로 표현하는 것이 어렵기 때문에 데이터의 분석뿐만 아니라, 분석된 데이터의 시각화가 매우 중요하다.

[그림 2] 빅데이터 분석과 시각화



*출처: 빅데이터 처리과정, 박두순외 5명, "울산 빅데이터, 통합센터 신설 등 고도화해야" focus daily 재인용



• 빅데이터 수집

데이터베이스는 특정 그룹이나 조직의 내부와 외부에 존재하는 데이터들의 수집을 기반으로 생성된다. 빅데이터 수집은 로그 수집기, 센싱, 크롤링, 오픈 API 기술을 통해 수행된다. 로그 수집기를 통해 조직 내부에 존재하는 웹 서버의 로그를 수집하고, 웹 로그, 트랜잭션 로그, 클릭 로그, 데이터베이스 로그 데이터 등을 수집할 수 있다. 센싱 기술은 스마트팩토리 등에 있는 센서를 통해 데이터를 수집하는 기술을 의미한다. 크롤링 기술은 주로 웹 로봇을 이용하여 조직 외부에 존재하는 소셜 데이터 및 인터넷 데이터에 공개되어 있는 자료를 수집해오는 기술을 의미한다. 오픈 API 기술은 데이터의 생산, 공유, 참여 환경인 웹을 구현하는 프로그래밍 기술을 통해 필요한 데이터를 수집하는 기술을 의미한다.

• 빅데이터 저장 및 처리

작은 데이터라도 모두 저장하고 실시간으로 데이터를 처리할 수 있도록 대부분 분산 저장 시스템을 활용한다. 저장된 데이터들을 실시간으로 처리하기 위해 분산 병렬 처리 방식을 주로 활용하며, 일반적으로 하둡(Hadoop)¹⁾ 분산 파일 시스템을 통해 이기종간의 하드웨어로 구성된 컴퓨터 클러스터에서 대용량 데이터 처리를 수행한다.

빅데이터 처리는 빅데이터에서 유용한 정보와 숨어있는 지식을 찾아내기 위한 데이터 가공 및 분석 과정을 지원하는 기술을 의미한다. 대규모 데이터 처리를 위해서는 확장성이 필요하며, 데이터 생성 및 처리 속도를 해결하기 위한 처리 시간 단축, 실시간 처리 지원 및 비정형 데이터 처리 지원이 필요하다. 이를 위해 하둡, NoSQL 및 구글 맵리듀스(MapReduce) 등을 통해 데이터 처리를 지원하고 있다.

• 빅데이터 분석

빅데이터 분석 기술은 데이터를 효율적으로 정확하게 분석하여 비즈니스 등의 영역에 적용하기 위한 기술로, 빅데이터에 숨어있는 패턴과 지식을 찾아내기 위한 기술을 의미한다. 빅데이터 분석을 통해 찾아진 패턴과 지식을 토대로 비즈니스 영역에서는 의사결정을 수행한다. 빅데이터 분석 기술에는 통계 분석, 데이터 마이닝, 텍스트 마이닝, 평판 분석, 소셜 네트워크 분석 기술이 포함된다.

통계 분석은 R, SAS(Statistical Analysis System) 등을 통한 다양한 통계 기법을 통해 수행된다. R은 빅데이터 분석 기술 도구로, 기본적인 통계 기법부터 데이터 마이닝 기법까지 구현이 가능한 기술 도구이며, SAS는 통계 해석을 중심으로 한 소프트웨어 패키지이다. 최근에는 기계 학습, 신경망 등을 이용한 데이터 마이닝 기술이 가장 각광받고 있다. 평판 분석은 소셜 미디어 등의 정형 또는 비정형 텍스트의 긍정 선호도를 판별하여 분석하는 기술로, 주로 특정 서비스 및 상품에 대한 시장 규모 예측 및 입소문 분석 등에 활용된다. 소셜 네트워크 분석은 소셜 네트워크의 연결 구조 및 연결 강도 등을 바탕으로 사용자의 명성 및 영향력을 분석하는 기술이다. 이러한 빅데이터 분석을 통해 도출된 빅데이터의 패턴과 지식은 비즈니스 의사결정에 활용되고 있어, 비선형적인 결과 값을 알아보기 쉽게 표시하는 데이터시각화 기술의 중요도가 증가하고 있다.

1) 하둡(Hadoop): 분산 파일 시스템과 맵리듀스를 구현한 빅데이터 처리 기술의 대표적인 프레임워크



II. 심층기술분석

데이터시각화를 통한 빅데이터의 인사이트 전달

빅데이터 분석 결과는 합리적인 의사결정의 근거 자료가 된다. 합리적인 의사결정 지원을 위해 빅데이터의 분석 결과를 적절한 시각화 형태로 전달하는 것이 중요하며, 이러한 니즈가 바로 데이터시각화의 핵심이다.

■ 정보 시각화를 기반으로 한 빅데이터 시각화

데이터시각화는 데이터 분석 결과를 쉽게 이해할 수 있도록 시각적인 수단으로 정보를 전달하는 과정을 의미한다. 빅데이터 내에 수많은 패턴들을 시각화하여 핵심 개념과 아이디어를 직관적이고 명확하게 이해할 수 있도록 하는 것이 데이터시각화의 목적이다. 데이터시각화는 차트, 그래프, 맵과 같은 시각적 요소를 사용하여 빅데이터에서 추세, 이상 값 및 패턴을 쉽게 이해할 수 있도록 하는 기법이다. 빅데이터의 세계에서 데이터시각화 도구는 막대한 양의 정보를 분석하고 데이터 기반 의사결정을 내리는 데 필수적이다.

빅데이터 시각화는 정보 시각화를 기반으로 하고 있다. 정보 시각화는 대규모 수량, 비수량 데이터를 색채, 통계, 이미지 등을 활용해서 시각적으로 표현하는 것을 의미한다. 정보 시각화 방법에는 시간 시각화, 분포 시각화, 관계 시각화, 비교 시각화 및 공간 시각화가 있다. 다양한 시각화 방식을 통해 정보 전달 시간을 단축하고, 빅데이터에서 도출된 유의미한 인사이트를 비즈니스에 빠르게 적용하도록 할 수 있다.

[그림 3] 주요 시각화 방법



*출처: 데이터베이스 진흥원, 데이터 분석 전문가 과정 3장 시각화 방법

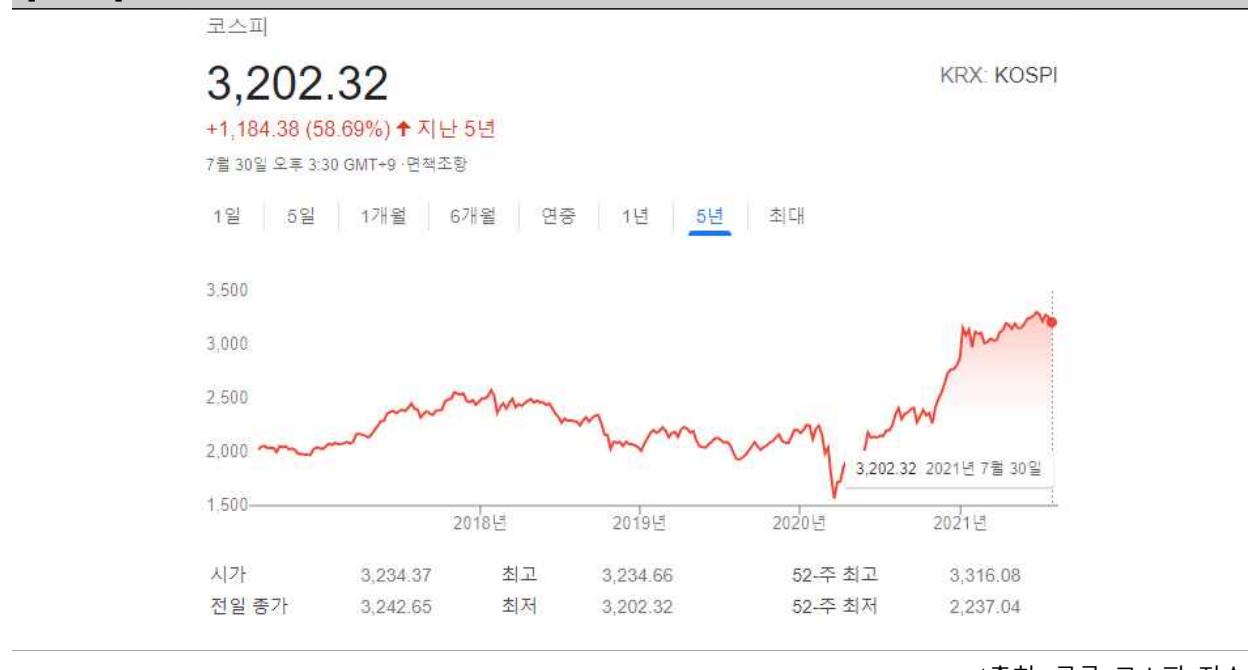


■ 시간 시각화 기법

빅데이터에서 시간과 관련된 데이터를 전달하고자 할 경우 시간 시각화 기법이 사용된다. 시계열 데이터에서 가장 큰 특징은 트렌드이다. 시간의 흐름에 따라 인구 분포가 변화하고, 사업 영역이 변화하고, 사람들의 취향이 변화한다. 이러한 변화가 얼마나 진행되었는지 측정해서 기록한 것이 시계열 데이터이다. 변화의 패턴을 찾아서 트렌드를 알기 위해서는 개별 측정 데이터보다는 전체적인 변화의 추이를 알아야만 한다. 즉, 한가지 구간에 대한 특정 값보다는 다수 간에 걸친 데이터 변화량의 전달이 더 중요하다. 시간 시각화를 통해 전체적인 데이터의 변경 추이, 급격하게 데이터가 증가하거나 감소된 구간이 있는지, 눈에 띠는 특정 구간이 있는지, 일정한 패턴이 있는지 여부를 확인할 수 있다.

시간 시각화는 막대 그래프, 누적 막대 그래프, 점 그래프 등으로 표시될 수 있다. [그림 4]는 코스피 지수의 추이를 시간 시각화 형태로 나타낸 것이다. 코스피 지수의 5개년 트렌드는 우상향하는 방향으로 이동하고 있는 것을 확인할 수 있다. 2020년 초에 기존 트렌드와 다르게 급격하게 지수가 감소한 부분을 확인할 수 있다. 이러한 그래프를 통해 해당 시점에 코로나19로 인해 주식 시장에 부정적인 영향을 상당히 주었다는 것을 확인할 수 있다.

[그림 4] 시간 시각화 예시: 코스피 지수 추이



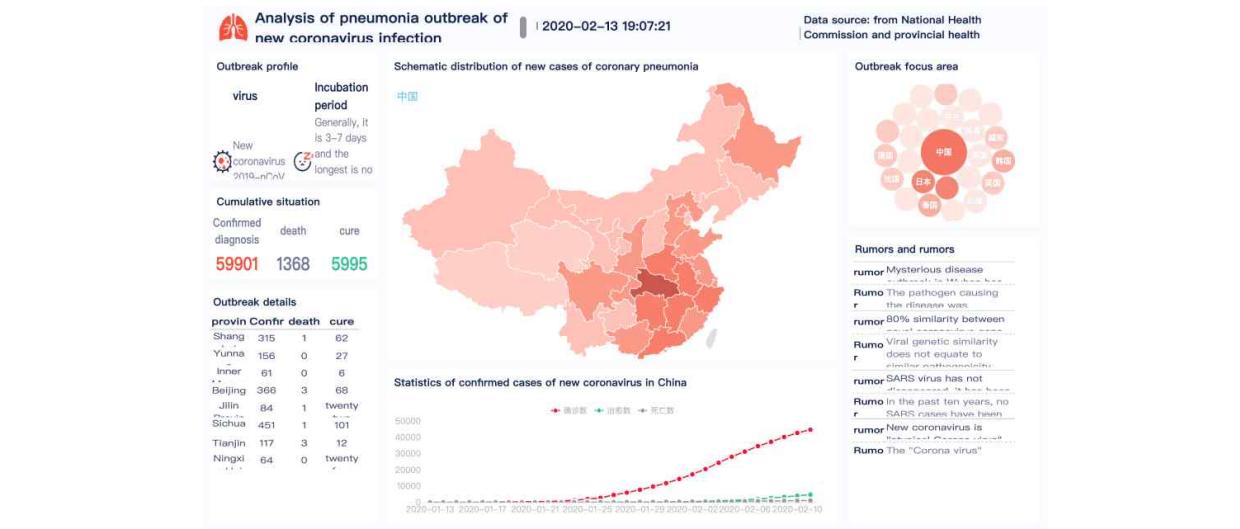
*출처: 구글 코스피 지수

■ 분포 시각화 기법

분포 시각화는 데이터의 전체적인 분포를 확인하기 위한 시각화 방식이다. 시간 시각화의 단위가 시간인 것과 달리, 분포 시각화의 단위는 분류, 세부 분류 및 가짓수이다. 분포 시각화 자료를 통해 빅데이터의 최대, 최소, 전체 분포를 확인할 수 있다. 분포 시각화에는 트리맵 차트, 파이 차트, 도넛 차트, 누적 연속 그래프 등이 사용된다.

[그림 5]는 코로나19 확산을 분포 시각화 형태로 분석한 대시보드이다. 확진자 수가 많은 지역은 더 짙은 색으로 표시되어 있어 어느 지역에 확진자가 가장 많은지 한눈에 확인이 가능하다.

[그림 5] 분포 시각화 예시: 코로나19 확산 분석 대시보드



*출처: FineReport "코로나19 사태에 데이터 시각화를 어떻게 활용할까?"

[그림 6]은 분포 시각화에서 가장 많이 쓰이는 형태 중 하나인 트리맵 형태로 나타난 S&P500의 동향을 나타낸 차트이다. S&P500 중 규모가 큰 종목들일수록 사각형의 크기가 크게 나타나고, 종목의 변동에 따라 색상으로 차트가 표시가 된다. 하기 차트를 통해 사용자들은 전체적인 주가 동향 및 섹터(기술, 금융, 커뮤니케이션 섹터 등) 별 주가 동향을 한눈에 확인할 수 있다.

[그림 6] 분포 시각화 예시: S&P500 동향 차트



*출처: finviz.com

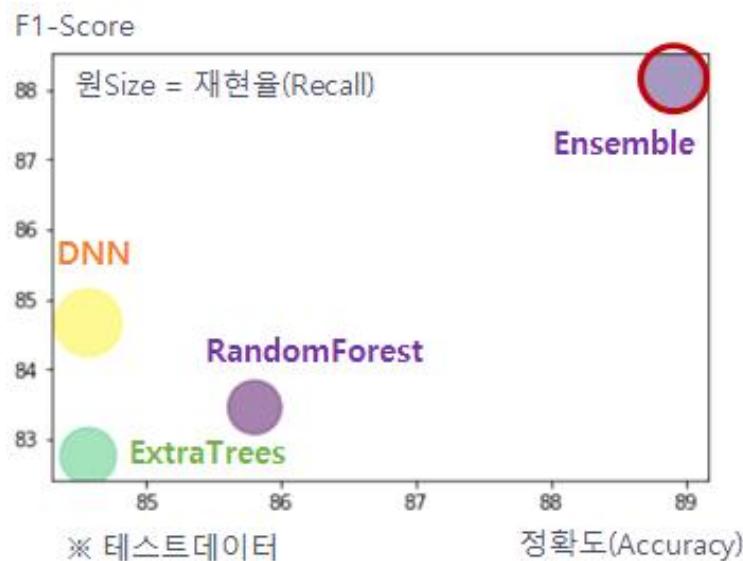


■ 관계 시각화 기법

빅데이터를 통해 확인할 수 있는 중요한 지표는 데이터 간의 관계이다. 집단 간에 유사점이 있는지, 특정 집단의 소집단이 존재하는지, 집단 간의 상관관계가 있는지 여부를 관계 시각화를 통해 확인할 수 있다. 관계 시각화에는 버블 차트, 히스토그램, 스캐터 플랫 등이 사용된다. 관계 시각화에 가장 많이 사용되는 차트는 이 중에서도 버블 차트이다.

[그림 7]은 인공 신경망 모델들의 성능을 버블 차트로 나타낸 것이다. 가로축은 정확도, 세로축은 F1-Score로 구현하여 모델 각각에 대한 성능을 나타내며, 버블의 크기로 재현율을 표시하고 있다. 해당 버블 차트를 통해 신경망 모델들의 전체적인 정확도 및 F1-Score 추이를 알 수 있으며, 특정 신경망 모델의 성능이 다른 신경망 모델들에 비해 압도적으로 높은 것을 한 눈에 확인할 수 있다. 즉, 사용자들은 버블 차트를 통해 신경망 모델들의 성능 추이 및 모델들 각각의 개별 성능을 한눈에 비교할 수 있다.

[그림 7] 관계 시각화 예시: 신경망 모델별 성능 버블 차트



*출처: Github, "[Python] 버블차트로 모델 성능 표현하기", 2019.01

■ 비교 시각화 기법

비교 시각화 기법은 하나 이상의 변수를 기준으로 대상을 비교할 수 있도록 하는 기법이다. 한 개 또는 두 개의 변수로 대상을 비교하는 것은 비교적 간단하나, 비교 대상이 많아지거나 비교 변수의 개수가 늘어날 경우 고도화된 시각화 기법이 필요하다. 시각화 기법을 활용하지 않을 경우, 비교해야 할 대상의 숫자에 비례하여 변수의 숫자를 곱한 만큼의 단위 비교 목록을 체크해야 한다는 문제가 존재한다. 비교 시각화 기법으로는 데이터를 전체적으로 시각화한 후, 흥미로운 부분을 선택적으로 자세히 시각화하는 것이 일반적인 접근 방법이다. 데이터를 한눈에 볼 수 있도록 값을 색으로 나타내어 표로 도식화하는 히트맵(Heatmap) 또는 히트 행렬(Heat matrix)이 비교 시각화 기법에 일반적으로 사용되고 있다. 그 외에도 비교 시각화 기법에는 스타 차트, 평행 좌표계, 다차원 적도법 등이 사용된다.



[그림 8] 비교 시각화 예시: 미국의 주별 범죄 동향



*출처: Tableau public "Crime Waves in the States"

[그림 8]은 미국의 주별 범죄 종류 및 연도에 따른 범죄 횟수를 히트맵으로 나타낸 것이다. 히트맵을 이용하여 주별 범죄 발생 동향, 범죄 종류별 범죄 발생 동향, 그리고 연도별 범죄 발생 동향을 확인할 수 있다. 해당 히트맵은 주, 범죄 종류 및 연도 3가지 변수를 이용하여 시각화 자료를 표현한 것이다. 십만 건 이상의 미국 주별 범죄 현황을 데이터 각각으로 보면 동향을 파악하는 것이 어렵지만, 위와 같은 히트맵을 활용할 경우, 단 한 장의 차트 만으로도 다양한 관점의 분석이 가능하다. 즉, 히트맵은 전체 데이터를 한눈에 볼 수 있다는 장점이 있으며, 가지고 있는 데이터가 여러 개의 변수로 구성되더라도 일정 기준에 따라 분류하거나 단위에 따라 나누어 예외적인 값을 찾아낼 수 있다.

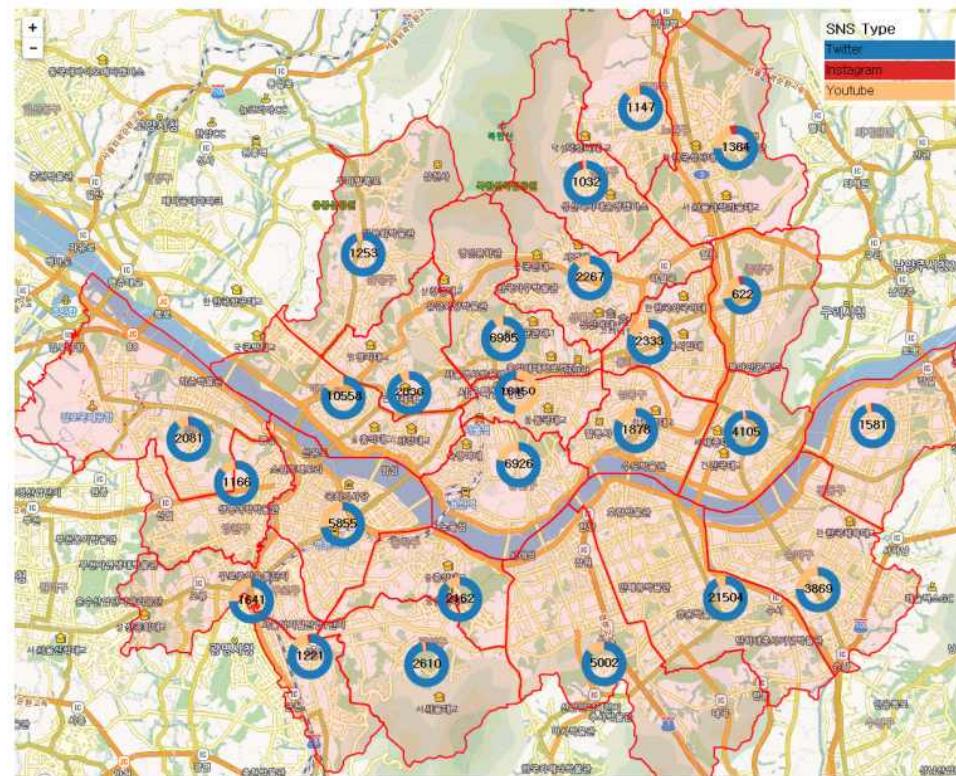
■ 공간 시각화 기법

공간 시각화는 지도 맵핑 방식을 통해 구현된다. 구체적으로, 공간 시각화는 지도에 특정 시간 단위의 값을 맵핑하거나, 지도에 다양한 분류 값을 맵핑하는 시각화 방식이다. 지도는 직관성을 폭넓게 활용한 시각화의 한 분야로, 지도를 읽는 것과 통계 그래픽을 읽는 방법은 매우 유사하다. 지도는 위치와 거리에 대한 직관적인 정보를 내포하고 있다. 위치와 색상 구분을 명확히 하고, 라벨이 위치를 가리지 않도록 정확한 투사 방식을 통해 지도를 기반으로 한 시각화 결과를 생성할 수 있다. 예를 들어, 시간 변수를 반영한 공간 시각화 방식에는 애니메이션으로 특정 지역의 경제적인 성장 또는 몰락을 시각화할 수 있다.

[그림 9]는 서울시의 각 구를 기준으로 각 구에 포함된 SNS 데이터의 총 건수를 숫자로 나타내고 있으며, 파이 차트는 해당 구의 SNS 데이터의 총 건수에서 트위터, 인스타그램, 유튜브가 차지하는 비율을 보여주고 있다.



[그림 9] 공간 시각화 예시: 서울시 구별 SNS 데이터 분포



*출처: 서양모, 김원균 “공간빅데이터를 위한 정보 시각화 방법” 한국공간정보학회지 v.23, no. 6, pp. 2015. 12

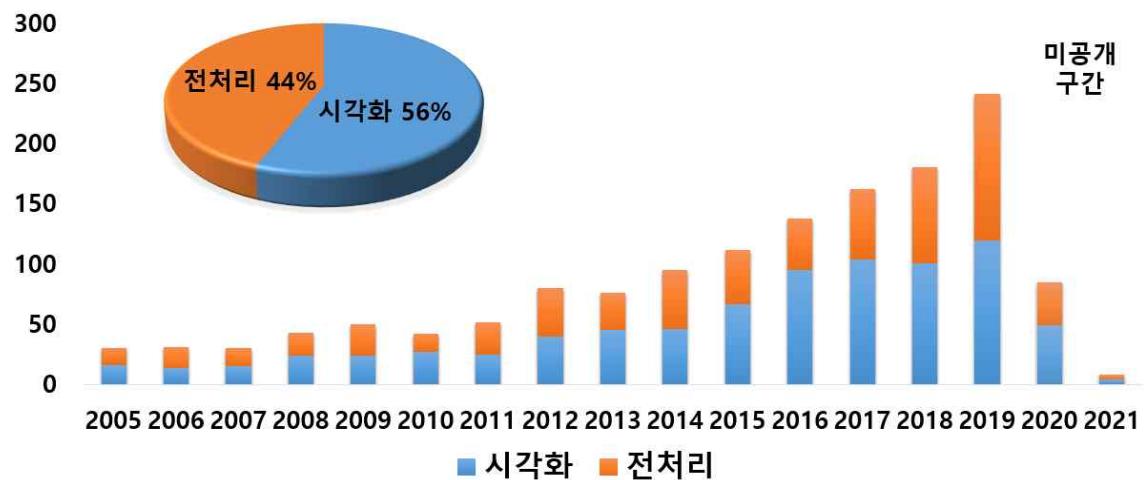
■ 데이터시각화 기술 관련 특허동향

[그림 10]은 데이터시각화 기술과 관련된 특허 출원 동향을 연도별, 기술별로 나타내었다. 데이터시각화 관련 기술은 미국의 유수 대기업들이 글로벌 시장의 대부분을 선점하고 있어, 해외 출원 특허를 함께 검색 시 해외 출원인들 위주로 검색 결과가 도출되어 국내 기업들의 출원 동향을 확인하기 어려우므로 특허 검색은 국내에 한정하여 진행하였다. 특허 검색 시기와 관련하여, 빅데이터가 활용되기 시작한 시점이 2000년대 중반이라는 점을 고려하여, 2005년 이후 출원 건으로 한정하였다. 전체 조사 특허 건수는 총 1,458건이며, 데이터시각화 기술은 크게 전처리 기술과 시각화 기술로 분류된다. 전처리 기술은 빅데이터를 시각화하기 위해 빅데이터를 전처리하는 작업들을 의미한다. 위에 기술한 빅데이터의 수집부터 분석까지가 빅데이터 전처리 기술에 해당한다. 시각화 기술은 전처리된 빅데이터를 사용자들이 한눈에 알아볼 수 있도록 시각화하는 기술을 의미한다. 기술 분류 별 출원 비중을 살펴보면, 전처리 기술이 44%, 시각화 기술이 56%를 차지하였다. 일반적인 빅데이터 전처리 기술들은 빅데이터 시각화 기술보다 훨씬 더 많은 비중을 차지하나, 해당 특허 분석은 빅데이터 시각화를 위한 전처리 기술로 한정하였으므로, 시각화 기술보다 더 적은 비중을 차지하는 것으로 나타내고 있다. 2000년대 중반 이후 데이터시각화 관련 출원이 꾸준히 증가하였으며, 특히 2019년에 관련 출원이 이전 년도에 비해 가파르게 증가하였다. 이와 관련하여, 세일즈포스의 태블로는 한국에 2014년 런칭된 이후, 2019년에 상당히 많은 고객사를 확보하였는데, 태블로의 국내 시장에서의 성장 및 출원 동향을 확인하였을 때, 국내 시장에서 데이터시각화에 대한 니즈가 2019년도를 기점으로 상당히 증가한 것을 확인할 수 있다.



[그림 10] 연도별 특허출원 동향

(단위: 년, 건)

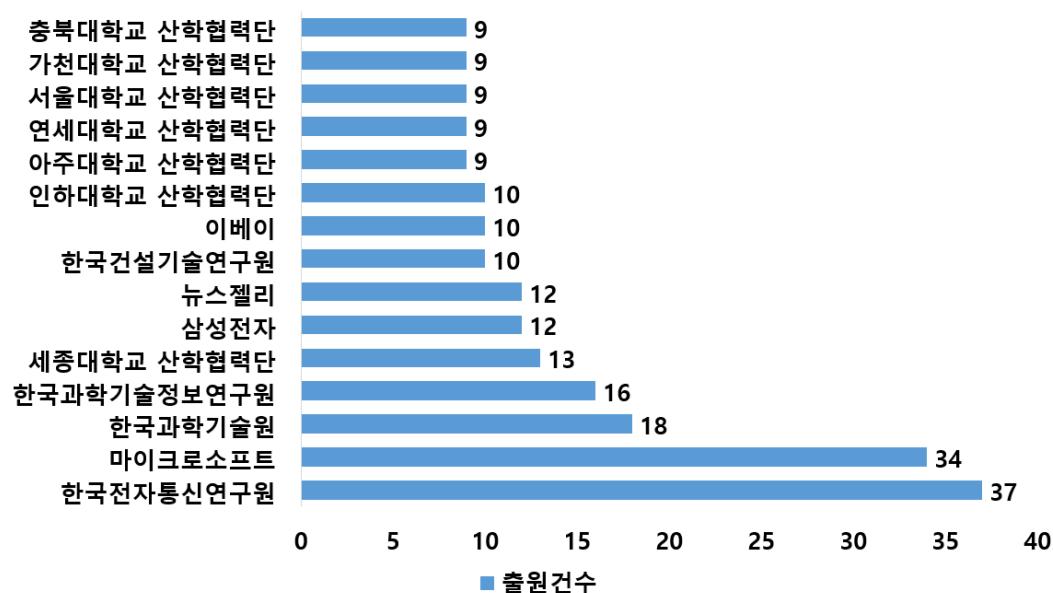


*출처: 원텔립스 DB, NICE디앤비 재구성

[그림 11]은 데이터시각화 기술과 관련된 출원 특허를 검색하여 확인된 주요출원인을 나타내었다. 주요출원인은 한국전자통신연구원, 마이크로소프트, 한국과학기술원 순이었다. 세계 시장을 선점하고 있는 구글과 세일즈포스의 독주로 국내 대기업들은 데이터시각화와 관련된 기술 개발에 큰 투자를 진행하고 있지 않고 있다. 이에 따라, 국내 대기업들은 주요 출원인에 거의 나타나고 있지 않으며, 국내에서는 연구원(산학협력단 포함)들을 위주로 특허 출원이 진행된 것을 알 수 있다. 세일즈포스의 경우 미국 위주로 출원을 진행하고 있으며, 한국에서의 사업 영위는 별론으로 한국 특허 출원은 진행하고 있지 않다. 주요출원인 중 뉴스젤리가 데이터시각화에 특화된 솔루션을 제공하는 기업으로, 뉴스젤리의 특허 출원 전부가 데이터시각화 관련 특허이다.

[그림 11] 주요출원인 및 출원건수

(단위: 건)



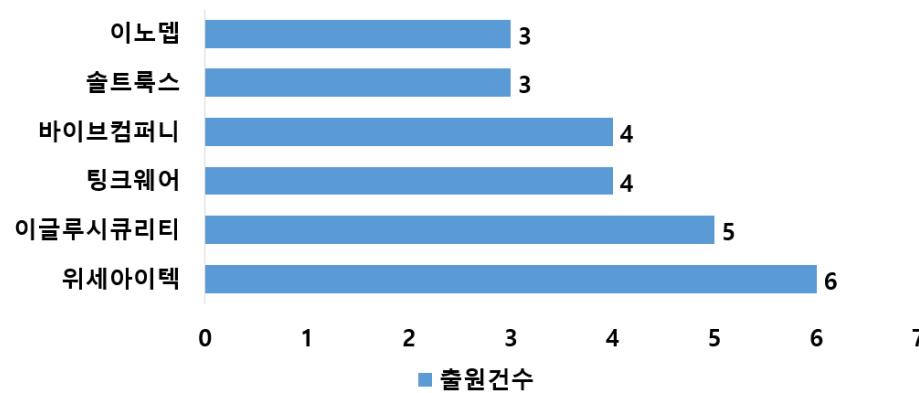
*출처: 원텔립스 DB, NICE디앤비 재구성



[그림 12]는 데이터시각화 기술과 관련된 출원 특허를 검색하여 확인된 주요 코스닥 출원인을 나타내었다. 관련 특허를 출원한 코스닥 기업으로는 위세아이텍, 이글루시큐리티, 텅크웨어, 바이브컴퍼니, 솔트룩스 및 이노뎁이 확인되었다. 해당 기업들은 모두 데이터시각화를 주 사업으로 영위하는 기업들은 아니며, 인공지능 및 빅데이터 기술을 주로 연구개발하고 있는 회사이다. 해당 기업들의 특허 출원 건수의 대부분은 데이터시각화가 아닌 인공지능 또는 빅데이터 관련 내용이나, 빅데이터와 관련된 기술 개발을 수행하면서 데이터시각화 관련 내용도 일부 출원을 진행한 것으로 확인된다. 또한, 세일즈포스와 같이 데이터시각화만을 주력하여 연구하고 있는 코스닥 기업은 국내에서는 없는 것으로 확인된다.

[그림 12] 주요 코스닥 출원인 및 출원건수

(단위: 건)



*출처: 원텔립스 DB, NICE디앤비 재구성

[그림 13]은 주요출원인별 주요기술 동향, [그림 14]는 주요 코스닥 출원인별 주요기술 동향을 나타내었다. 주요 기업들은 데이터시각화를 위한 빅데이터 전처리 기술과 데이터시각화 기술에 대한 연구개발을 모두 수행 중인 것으로 나타났다. 특히, 한국전자통신연구원의 경우 데이터시각화 자체 기술 위주로 연구를 진행하였으며, 마이크로소프트의 경우 데이터시각화를 위한 빅데이터 전처리를 위주로 연구를 진행한 것을 확인할 수 있다. 코스닥 기업들의 경우 데이터시각화가 주 사업이 아니기 때문에 전체적으로 데이터시각화 자체 기술보다는 데이터시각화를 위한 빅데이터 전처리 기술에 더 집중하고 있는 것으로 확인되었다.

[그림 13] 주요출원인별 주요기술 동향 (단위: 건)

[그림 14] 주요 코스닥 출원인별 주요기술 동향 (단위: 건)



*출처: 원텔립스 DB, NICE디앤비 재구성



III. 산업동향분석

디지털 뉴딜을 기반으로 한 데이터시각화 시장 상승세 기대

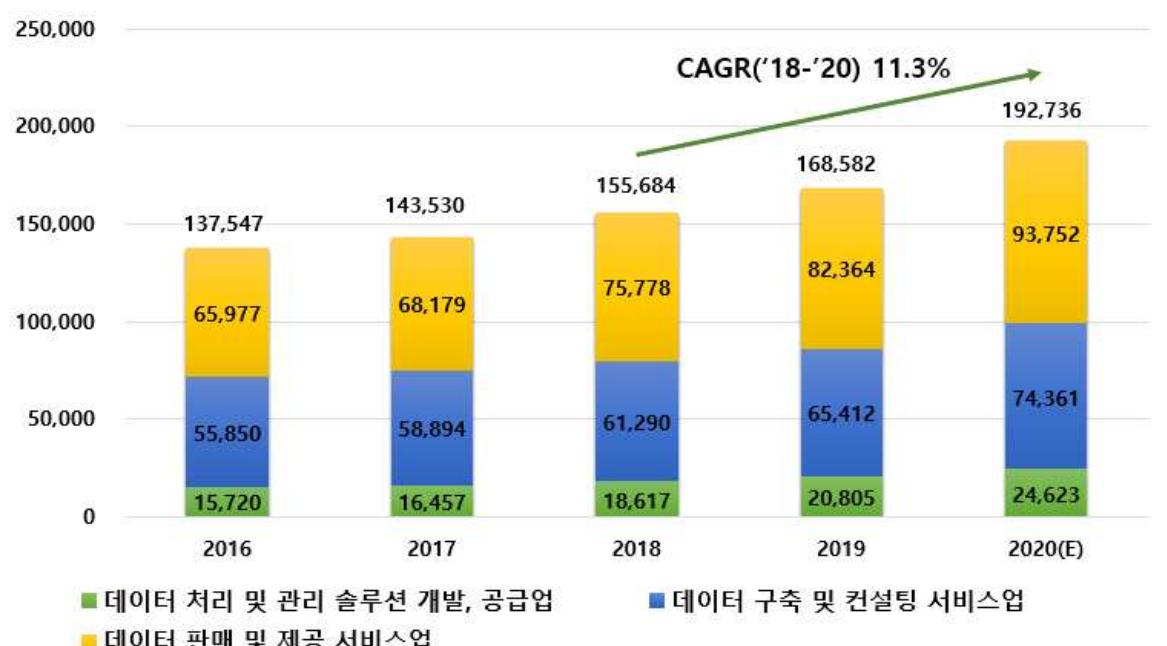
한국형 디지털 뉴딜 정책을 통해 국내 데이터 산업이 활성화되고 있으며, 데이터 산업 활성화에 따른 데이터시각화에 대한 수요 증가가 기대된다.

■ 데이터 중요성에 대한 국내외 인식 증가

전 세계 국가 정부와 대규모 데이터를 활용하는 글로벌 기업들은 2020년 한 해 동안 데이터 활용 및 보호, 수익 창출이 가능한 주요 전략들을 쏟아냈다. 데이터 산업을 구성하는 생태계 구성원들인 기업들은 이러한 흐름에 적극적으로 동참하는 추세이다. 한국 정부도 데이터 3법을 시행하여 데이터 활용 권한을 분명히 하고, 데이터 가공, 결합 및 분석을 위한 기본 환경을 마련하였다. 한국 정부는 국가 경제 위기에 대응하고자 국비 114조 원을 투입한 한국형 디지털 뉴딜 정책을 발표하여 데이터 댐 건설, 지능형 정부 구현에 적극적인 지원을 하고 있다. 과학기술정보통신부는 데이터 댐 프로젝트의 7대 핵심사업으로 AI학습용 데이터 구축, AI바우처와 AI데이터 가공 바우처 사업, AI융합 프로젝트(AI+X), 클라우드 플래그십 프로젝트, 클라우드 이용 바우처 사업, 빅데이터 플랫폼 및 센터 구축을 선정하였다. AI 바우처 사업의 경우, 수요기업 중 비(非) ICT 기업 비율이 2019년 41.2%에서 2020년에 84.8%로 크게 증가하였다. 이는 AI 및 데이터가 ICT 기업뿐만 아니라 전 산업분야에 확산되고 있다는 것을 반증하는 결과이다. 이러한 국내외 정책과 기업들의 움직임에 따라 데이터 산업의 시장이 최근 2년간 급격하게 성장한 바 있다.

[그림 15] 국내 데이터 산업 시장 규모

(단위: 억 원)



*출처: 과학기술정보통신부 "2020년 데이터산업현황조사 주요 결과 요약", 2020.12., NICE디엔비 재구성



과학기술정보통신부(2020)에 따르면 국내 데이터 산업 시장 규모는 2019년 16조 8,582억 원으로 2018년 대비 8.3% 성장했으며, 2020년은 19조 2,736억 원으로 추정된다. 특히, 2018년부터 2020년까지 2년간 연평균 성장률은 11.3%로 최근에 높은 성장률을 보이는 것을 확인할 수 있다. 즉, 2018년도를 기점으로 국내 데이터 산업이 급격하게 증가하였음을 알 수 있다.

■ 빅데이터 분석 시장 성장에 따른 데이터시각화 시장 활성화

데이터 산업에서 빅데이터에 대한 중요성이 대두되면서 빅데이터 분석에 대한 관심도 함께 증가하고 있다. 빅데이터를 활용하기 위해서는 빅데이터 분석이 필수적이며, 빅데이터 분석을 통해 빅데이터 내에서 유의미한 결과를 도출해 낼 수 있고, 이를 비즈니스에 적용할 수 있다.

Statista(2019)에 따르면 빅데이터 및 데이터 분석 시장은 2018년 1,688억 달러에서 성장하여 2022년에는 2,743억 달러를 기록할 것으로 예상된다. 미국은 전 세계에서 가장 큰 빅데이터 및 데이터 분석 시장으로, 전 세계 빅데이터 및 데이터 분석 부문 매출의 대부분이 미국에서 창출되고 있다. 미국의 가장 대표적인 빅데이터 분석 전문 기업은 팰런티어테크놀로지로, 2003년 설립되어 미국 정부를 주 고객으로 서비스를 제공하고 있다. 팰런티어테크놀로지는 빅데이터 분석 결과를 시각화하여 제공하고 있으며, 특히 데이터 간의 연결 관계를 시각화하여 잘 보여주고 있다.

빅데이터 분석을 통해 도출된 결과는 데이터시각화 방식에 따라 정보 전달 효율에 상당한 영향을 줄 수 있기 때문에 빅데이터 분석 시장과 함께 빅데이터의 분석 내용을 전달하기 위한 데이터시각화 시장도 활성화될 것으로 보인다.

■ 데이터시각화 시장에 동반한 데이터시각화 툴 시장의 성장 기대

MarketsandMarkets(2021)에 따르면 글로벌 데이터시각화 툴 시장은 2021년 59억 달러에서 연평균 11.6% 성장하여 2026년 102억 달러로 성장할 것으로 예측된다. 이는 빠른 비즈니스 결정, 증강현실 및 가상현실의 발전에 따른 데이터시각화에 대한 수요 증가와 기업들의 3D 형태 데이터를 기반으로 한 상호 작용 증가에 따른 것으로, 데이터시각화 툴 시장은 급격하게 발전할 것으로 예측된다. 또한, 코로나19 바이러스로 인해 뉴스에서 선형 차트, 막대 차트 등 데이터시각화 도구들의 사용이 증가하였고, 대중들에게 정보 전달을 위한 도구로 사용 빈도가 크게 증가하는 등 코로나19 바이러스의 확산이 데이터시각화 툴 시장에 영향을 미치고 있다.

한편, 데이터시각화 툴은 비즈니스를 기반으로 마케팅, 판매, 인적 자원 관리, 운영 및 재무에 사용되는데, 마케팅 및 판매를 기반으로 한 데이터시각화 툴의 사용이 가장 크게 시장 성장에 기여할 것으로 전망된다. 이는 데이터시각화 도구를 통해 관리자가 매출에 대한 분기별 목표, 마감된 거래의 비율 등 영업 성과를 모니터링할 수 있기 때문이다.



IV. 주요기업분석

글로벌 대기업들이 리딩하는 데이터시각화 기술

데이터시각화 기술은 세일즈포스를 포함한 글로벌 대기업들이 리딩하고 있다. 국내에서는 아직까지는 활발한 연구개발이 이루어지고 있지는 않으나, 빅데이터 및 인공지능 기술을 개발하는 코스닥 기업들의 관련 시장 진입이 시작되고 있는 상태이다.

■ 국외 데이터시각화 기술 선도 기업: 세일즈포스, 구글, 마이크로소프트

1. 세일즈포스(Salesforce), 전 세계 독보적인 데이터시각화 솔루션 제공

세일즈포스의 태블로는 간단한 조작만으로 데이터를 분석하고 시각화된 자료를 공유할 수 있도록 한 솔루션이다. 태블로 소프트웨어는 2003년 후발주자로 비즈니스 인텔리전스(Business Intelligence, 이하 BI) 시장에 뛰어들어, 기존 BI 산업의 강자였던 SAP를 제치고 세계 BI 솔루션 시장에서 1위를 차지하였다. 태블로 소프트웨어는 이후 2019년도에 세일즈포스에 인수되어, 현재 세일즈포스 사에서 태블로 솔루션 판매를 담당하고 있다. 태블로는 간편한 시각화 과정을 내세워 전 세계 사용자들의 선택을 받고 있다. R에서 시각화를 하려면 코드를 입력하고, 데이터 임포트(import), 라이브러리 가동, 축의 변수명 지정 등 번거로운 절차를 거쳐야 했던 것과 달리, 태블로에서는 간단히 드래그 앤 드롭을 통해 데이터시각화를 수행할 수 있다. 태블로 솔루션은 국내에 2014년도에 런칭되어, 2019년 상반기 기준 1,000여 개 이상의 고객을 확보하였다.

[그림 16] 태블로를 이용한 대시보드 생성 예시



*출처: medium.com/@saerombang11 “태블로로 쉽게 데이터 시각화하기”



2. 구글, 다양한 구글 서비스와의 호환성을 내세운 데이터시각화 솔루션 제공

구글은 2016년도에 BI 데이터시각화 시장 발전 트렌드에 따라 구글 데이터 스튜디오를 출시하였다. 구글의 데이터 스튜디오는 다른 구글 서비스들과의 호환성을 기반으로 빠르게 성장하고 있으며, 웹 분석 솔루션의 대표 주자로 자리매김한 구글 애널리틱스, 구글 스프레드시트 등 구글에서 제공하는 다른 서비스들과 곧바로 데이터 호환이 가능하다. 특히 구글 데이터 스튜디오는 구글 애널리틱스에서 사용되는 대부분의 지표를 지원하고, 무료 솔루션이라는 점에서 다른 플랫폼보다 우위를 가진다.

[그림 17] 구글 데이터 스튜디오 예시



*출처: 구글 데이터 스튜디오 도움말

3. 마이크로소프트, 마이크로소프트 기본 서비스와 호환성을 강조한 BI 솔루션

마이크로소프트의 Power BI는 비즈니스 분석 플랫폼으로, 다양한 데이터의 원본(엑셀, CSV, XML 데이터 파일 등)에 연결하여 데이터 통합 분석을 수행한 후, 대시보드를 통해 시각적 분석을 생성한다. Power BI는 기존에 일반적으로 사용되던 마이크로소프트 서비스를 기반으로 손쉽게 동작이 가능하다. 또한, 사용자들의 진입이 쉽다는 점에서 Power BI의 점유율이 점차 증가하고 있는 추세이다.



■ 국내 데이터시각화 기술 관련 기업: 뉴스젤리, 와이즈넛, 비아이매트릭스

1. 뉴스젤리, 데이터시각화 특화 솔루션 제공

뉴스젤리는 데이터시각화에 특화된 기업으로, 2014년에 설립된 후 2016년에 데이터시각화 솔루션인 DAISY를 출시하였다. DAISY는 은평구, 성남시, 기상청, 농협, KB에서 사용되고 있는 솔루션으로, 대용량 데이터의 활용이 자유롭지 않은 엑셀과 차별화하여 대시보드 단위로 차트를 관리할 수 있도록 한다. 뉴스젤리는 15건의 등록 특허를 보유하고 있으며 연간 예산의 39% 이상을 R&D에 투자하고 있다. 뉴스젤리는 데이터진흥원에서 진행하는 2021년 데이터 바우처 지원 사업의 공급기업으로 데이터시각화 및 분석 솔루션을 제공하고 있다.

[그림 18] 뉴스젤리 DAISY를 이용한 데이터시각화 예시



*출처: medium.com/@saerombang11 “태블로로 쉽게 데이터 시각화하기”

2. 와이즈넛, 빅데이터 솔루션 제공

와이즈넛은 2000년도에 설립된 인공지능 기반 챗봇 및 빅데이터 수집, 분석 솔루션 전문 기업으로, 빅데이터 의미분석 솔루션을 제공하고 있다. 와이즈넛의 WISE BIC Analyzer는 비정형 빅데이터의 의미를 분석하고 빠른 의사결정을 지원하는 솔루션이다. 해당 솔루션은 텍스트 마이닝 및 분류 기술을 기반으로 구현되었다. WISE BIC Analyzer는 형태소 분석, 자연어 처리, 표현어 추출 기술을 포함하는 감성 분석용 언어처리를 통해 감성 분석 결과를 지수화하고, 긍정 또는 부정 트렌드를 분석하며, 이슈나 급상승 키워드를 분석하여 제공하고 있다.



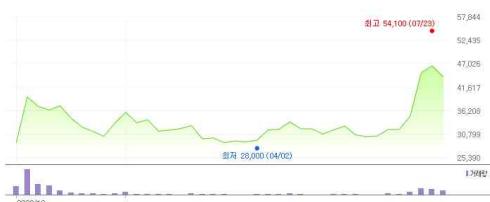
3. 비아이매트릭스, BI 솔루션 제공

비아이매트릭스는 기업의 경영 환경에 대응할 수 있도록 의사결정을 지원하는 빅데이터 및 BI 솔루션을 제공하고 있다. 비아이매트릭스의 i-STREAM은 데이터 전처리부터 데이터 분석까지 지원하는 데이터 분석 솔루션이며, i-AUD는 대시보드 화면 제작에 최적화된 BI 솔루션이다. i-AUD는 실시간 데이터를 반영하는 동적 차트, 현황판, 플래시 타입 차트 등 대시보드 시스템 구축을 지원하여 데이터시각화를 통한 의사결정을 지원하고 있다.

■ 국내 데이터시각화 기술 관련 코스닥 기업: 바이브컴퍼니, 솔트룩스, 엑셈

[바이브컴퍼니] 바이브컴퍼니는 2000년 7월 다음커뮤니케이션 내부에서 사내 인큐베이팅을 통해 다음소프트로 설립되었으며, 2020년 8월 14일 바이브컴퍼니로 상호를 변경한 후 기술특례상장을 통해 2020년 10월 28일에 코스닥 시장에 상장되었다. 바이브컴퍼니의 Sometrend Biz는 빅데이터 분석 솔루션으로 SNS와 온라인 플랫폼에서 발생하는 고객, 브랜드, 상품 등에 대한 데이터를 수집, 분석 및 시각화하여 인사이트를 제공하는 솔루션이다. Sometrend Biz는 트위터, 블로그, 인스타그램, 뉴스 등 채널 별 언급 추이를 확인하고, 카테고리에 따른 연관어를 분류하여 제공함으로써 기업들에게 고객의 관심 변화를 실시간으로 확인할 수 있도록 한다. 또한, 해당 솔루션은 연관어 시각화 데이터를 통해 기업의 평판을 확인할 수 있다.

[표 1] 바이브컴퍼니 주가추이 및 기본 재무현황(2018, 2019년도 K-IFRS 별도기준, 2020년도 K-IFRS 연결기준)

Performance		Fiscal Year	2018년	2019년	2020년
(단위: 원)		매출액(억 원)	141.0	163.0	255.7
		증감률 YoY(%)	0.5	15.6	56.9
영업이익(억 원)		영업이익률(%)	13.4	-3.4	-33.5
영업이익률(%)		순이익(억 원)	9.5	-2.1	-13.1
순이익(억 원)		EPS(원)	7.7	-7.8	-25.7
EPS 증감률(%)		EPS 증감률(%)	174	-164	-53
P/E (x)		P/E (x)	-35.3	적전	적전
EV/EBITDA(x)		EV/EBITDA(x)	-	-	-67.0
ROE(%)		ROE(%)	7.5	-5.1	-
P/B(x)		P/B(x)	-	-	5.6

*출처: 네이버금융, NICE디앤비 재가공

[솔트룩스] 솔트룩스는 2000년에 설립된 인공지능 및 빅데이터 기술 기업인 시스메타를 시작으로, 2006년에 사명을 솔트룩스로 변경한 기업이다. 솔트룩스는 2020년 07월 기술특례상장 제도를 통해 코스닥 시장에 상장되었다. 솔트룩스의 Bigdata Suite는 정형 빅데이터와 비정형 빅데이터의 융합 분석 솔루션이다. Bigdata Suite는 기업 및 공공 빅데이터에 대한 시맨틱 검색/분석, 지능화, IoT 센서와 생산 및 운영시스템 로그와 같은 스트리밍 빅데이터에 대한 실시간 분석, 예측 기능을 제공하고 있다. Bigdata Suite의 시각 분석 엔진은 다양한 형식의 데이터 소스를 연결하여 인터랙티브한 시각화 라이브러리들을 이용하여 시각적 분석 기능을 제공한다.



[표 2] 솔트룩스 주가추이 및 기본 재무현황(K-IFRS 연결기준)

Performance		Fiscal Year	2018년	2019년	2020년
	(단위: 원)	매출액(억 원)	154.8	183.7	216.6
	증감률 YoY(%)	26.9	18.7	17.9	
		영업이익(억 원)	1.9	0.3	3.6
		영업이익률(%)	1.3	0.1	1.7
		순이익(억 원)	0.5	8.3	5.1
		EPS(원)	21	203	116
		EPS 증감률(%)	-68.7	866.7	-42.9
		P/E (x)	-	-	295.3
		EV/EBITDA(x)	-	-	102.1
		ROE(%)	0.6	2.6	1.12
		P/B(x)	-	-	3.0

*출처: 네이버금융, NICE디앤비 재가공

[엑셈] 엑셈은 2001년에 설립된 후 2014년에 상장된 빅데이터 솔루션 기업으로, Flamingo를 통해 빅데이터 수집, 저장, 정제, 분석 및 모니터링이 가능한 솔루션을 제공하고 있다. 하둡 기반으로 빅데이터 플랫폼을 사용하는 기업들이 빅데이터 운영 중 발생한 이슈를 해결하는데 해당 솔루션을 활용할 수 있다. Flamingo를 통해 빅데이터를 관리하는 기업들은 플랫폼 모니터링, 데이터 정제/처리를 위한 워크플로우 기능을 이용하여 빅데이터 통합 분석 및 성능 관리를 손쉽게 수행할 수 있다.

[표 3] 엑셈 주가추이 및 기본 재무현황(K-IFRS 연결기준)

Performance		Fiscal Year	2018년	2019년	2020년
	(단위: 원)	매출액(억 원)	323.6	344.5	391.8
	증감률 YoY(%)	-4.0	6.5	13.7	
		영업이익(억 원)	44.9	63.2	96.5
		영업이익률(%)	13.9	18.3	24.6
		순이익(억 원)	44.9	63.2	96.5
		EPS(원)	180	140	242
		EPS 증감률(%)	1,385.7	-22.2	72.9
		P/E (x)	15.5	20.2	18.4
		EV/EBITDA(x)	9.5	6.9	10.0
		ROE(%)	11.6	7.5	11.6
		P/B(x)	1.5	1.4	2.0

*출처: 네이버금융, NICE디앤비 재가공