

마케팅 컨설팅 그룹을 지향하는 광고 회사 옴포스에는
꾸준히 티 안나게 별 짓 다 하는, 빅데이터실이 있는데

짧지만, 이게 저희 소개서입니다.

2023 ver.

“엠포스 빅데이터실은 무슨 일을 하나요?”

엠포스는 온라인광고로 통칭되는, **퍼포먼스 광고/마케팅**에 일가견 있는 회사입니다.

그 안에 속한 **빅데이터실**은 **데이터랩**이라는 이름으로 각종 무료 보고서를 발간하고 **빅데이터** 분석을 통해 **인사이트**를 찾는 다양한 **기업 및 기관**과 프로젝트를 진행하며

또문또문, **사내 타 부서**의 요청으로 광고주가 처한 상황을 면밀히 검토해주거나 또 가끔씩, 국내외 **언론사들**의 요청으로 시의성 있는 이슈를 파헤치기도 합니다.

그렇게 시간을 보내다 특정 이슈나 사안에 대한 데이터 분석을 진행하지 않을 때는

성장세 대비 활용도가 크지 않은 **광고 운영 데이터**에 대한 분석 기법을 연구하거나 우리가 발견한 통계적 분석 기법을 자동화 할 수 있는 **솔루션**을 만드는데 투자합니다.

아! 번외로, 하고 싶은 얘기가 생기면 카메라를 켜고 **유튜브** 콘텐츠를 찍기도 하고 소소하지만, 마케팅 관련 매체에 **칼럼**이라는 명목으로 빈번하게 기고도 합니다.

...

종합하면 우리는, **물어보면/ 찾아보고/ 대답하는** 일을 아주 잘 합니다. 그게 우리가 생각하는 **마케팅**이고 **광고**이며, 또 지향하는 **컨설팅**이고, 결국 우리가 추구하는 **데이터 분석**이니깐요.

“ 빅데이터가 왜 필요하세요? ”

새로운 요청을 접할 때면 우리는 늘 이런 질문을 드립니다

무엇을 알고자 하는지, 왜 빅데이터를 고려하게 되었는지 알아야

정말 빅데이터를 활용하기에 적합한지, 만약 아니라면
어떤 데이터를 활용할 수 있는지 대답할 수 있으니까요

빅데이터라고 해도, 여느 다른 데이터들과 마찬가지로
모든 이슈들에 정통한 대안이 되지 않습니다.

때로는 이미 가지고 있는 데이터만으로도 해결 가능할 수 있고
기존의 여론조사 방법이 더 적합한 경우도, ...사실 꽤 많습니다

...

그래서 우리는 특정 데이터를 무분별하게 주장하지 않고
다양한 데이터를 염두에 두고, **융합** 관점에서 고민합니다

빅데이터는 기존 데이터의 대체가 아니라 **별도 영역**이고
데이터 분석은 코딩에 머물지 않고 **해석을 담아야** 하며
해석은 번뜩이는 의견이 아닌 **관점에 기반**해야 하니까요

1. 데이터 분석 관련 업무 범위

- 시장의 경향성과 브랜드 위상, 광고 성과 측정까지 다양하게 진행하며 시장의 요구에 따라 특정 이슈나 사안에 대한 반응을 체크하여 제공
- 시장 환경과 동향, 경쟁 구조, 브랜드나 제품 인식, 거기다 정치/사회/문화에 이르는 다양한 분석 주제에 대응하기 위해서는, 역시 다양한 데이터 소스가 필요한데, 이에 대해서는 다음 페이지에 정리

Market Trend

시장의 트렌드를 분석하되 단기적 회자성과 붐, 트렌드를 구분할 줄 알고 시장의 환경과 구조 파악을 통해 현재의 상황을 면밀히 검토하여 제공

Brand Status

자사 브랜드나 제품에 대한 소비자 인식과 태도, 행동을 기반으로 브랜드가 도달하고자 하는 위치와 소비자 인식 상의 간극 측정

AD Performance

운영중에 축적된 온라인 광고 데이터를 활용, SA, DA 등에 대한 성과를 통계적으로 규명

Issue Tracking

정치, 사회, 문화 등
각 영역에 대한 이슈 점검



2. “꾸준히 연구” 중인 데이터 종류 소개

- 기술적으로 접근 가능한 여러 종류의 데이터에, 다양한 통계 기법을 적용해가며 활용 방안을 심도 있게 연구 중
- 2023년 현 시점에서의 주요 활용 데이터는 SNS, 검색, 쇼핑클릭 등 TEXT DATA이며
 - 앱 리뷰 데이터의 경우 단순 기능 평가 외 유의미한 리뷰 내용이 있을 경우 활용,
 - 광고 운영 데이터는 GA4 등의 플랫폼에 측정 용이한 변수들이 적절하게 세팅 되어 있을 경우, 아니면 고객에게 권한을 위임 받아 직접 세팅하여 활용,
 - 외부 데이터 수집은 해외 대표 커뮤니티 사이트인 레딧(Reddit) 등에 필요 정보가 있거나 고객사 내부에 보관중인 데이터 분석을 요청할 경우 대응

	Desk Research [거시적 통계 자료]	SNS Data [자발적 언급 내용]	Search Data [포털사이트 검색 행동]	Shopping Click [쇼핑 상품 클릭 지수]	App Review [어플 이용 후기]	Performance Data [광고 운영 데이터 전반]	Data Crawling [국내외 직접 수집]
종류	언론 기사, 논문, 서적, 국내 외 통계 자료 등	트위터, 블로그, 커뮤니티, 인스타그램 등 4개 SNS 채널	네이버 사이트를 메인으로 구글, 유튜브, 카카오톡 일부 포함	네이버 사이트에서 판매되는 온라인 쇼핑 클릭 지수	구글플레이 및 앱스토어 리뷰 내용 추출, 분석	<ul style="list-style-type: none"> • SA, DA 등 광고 지표 분석 • GA4 등 플랫폼 점검/세팅 	해외 레딧(Reddit) 등 접근 가능 채널
추세 분석	○	○	○	○	○	○	○
키워드 분석	-	○	○	○	○	○	○
성/연령 분석	-	-	○	○	-	○	-
통계 기법 활용	-	○	○	○	○	○	○
통계적 분석 기법 종류	-	WordCloud, SNA, TF-IDF, LDA, Multi-Correspondence, Clustering, R Shiny, ..	WordCloud, Clustering, ..	WordCloud, Clustering, ..	WordCloud, SNA, TF-IDF, LDA, Multi-Correspondence, Clustering, R Shiny, ..	WordCloud, Clustering, R Shiny, ..	WordCloud, SNA, TF-IDF, LDA, Multi-Correspondence, Clustering, R Shiny, ..
분석 접근 방향	개요 점검/ 가설 탐색	인식 및 태도/ 시장 트렌드	관심 행동/ 시장 트렌드	잠재 구매 행동/ 시장 트렌드	브랜드 평판 점검 [단순 기능 평가 제외]	광고 운영 성과 - 유입 키워드 Portfolio, - 사이트 유입 흐름	추가 발굴

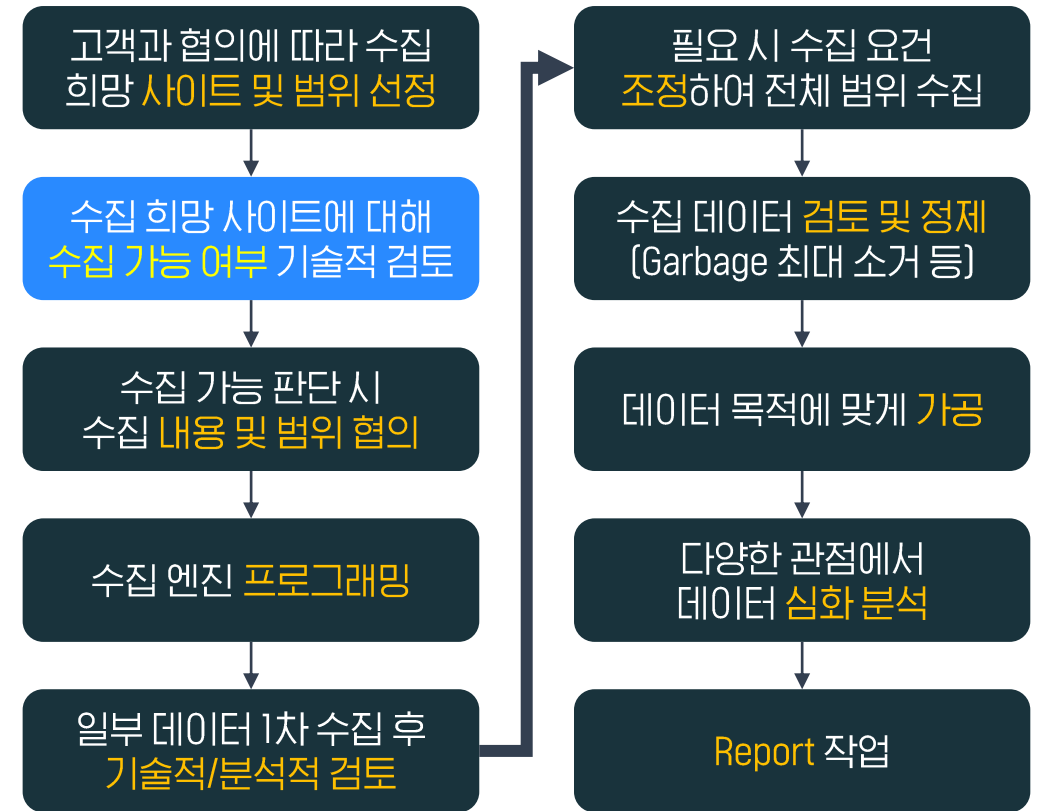
3. 주로 활용하는 데이터 분석 기법 소개

- 엠포스 빅데이터실에서는 텍스트 분석을 위해 하기와 같이 6가지의 통계적 분석 기법을 활용하며 분석된 결과를 공유하기 위한 목적으로 R Shiny 활용
- 하기 6가지의 통계적 분석 기법은 요구된 이슈 정의, 시장 특성이나 상황에 따라 변별하여 사용

No.	통계적 분석 기법 명칭	요약	기법 설명
01	워드 클라우드 (WordCloud)	키워드 종류/빈도 확인 시장/브랜드 개요 확인	문서내 키워드 간 게시 빈도를 키워드의 상대적 크기로 시각화. “어떤 키워드가 가장 많이 나오지?”
02	소셜 연결망 분석 / SNA (Social Network Analysis)	키워드 간 관계 확인 시장/브랜드 구조 점검	키워드 간 연관성을 수많은 연결선으로 시각화 “어떤 키워드들끼리 뭉쳐 있지?”
03	TF-IDF (Term frequency-Inverse Document Frequency)	빈도 기반의 키워드 조합 확인 Tone & Manner 구조화	가중치에 따라 핵심 키워드 및 연관 키워드 조합을 차트화 “상위 키워드 옆에는 어떤 키워드들이 붙어있지?”
04	잠재 디리클레 할당 / LDA (Latent Dirichlet Allocation)	키워드 그룹화 확인 언급 주제 유형 정의	키워드 간 연관성을 기반으로 그룹화, 숨겨진 주제(Topic) 찾기 “몇 개의 그룹으로 묶이지? 그룹별 어떤 키워드들이 묶였지?”
05	시계열 군집 분석 / (Time Series Clustering Analysis)	키워드 그룹화 확인 유사 반응 정의	키워드의 기간별 정보량을 기반으로 그룹화 “몇 개의 추세로 나뉘지?” “추세가 비슷한 그룹의 키워드들은?”
06	다중 대응일치분석 / MCA (Multiple Correspondence Analysis)	그룹 간 관계 확인 소비자 인식 구조화	키워드 그룹 간의 동반 언급 횟수를 기반으로 상대적 거리 파악 “스마트워치가 무게에 더 가깝나, 가격에 더 가깝나?”
07	R Shiny	임시 대시보드 구현/ URL 기반 결과 공유	위 01~06까지의 통계 분석 결과들을 대시보드 형태로 구현해주는 R 프로그래밍

4-1. 추가 업무 범위 상세 : 특정 사이트 데이터 별도 수집

- 기존에 자사 수집 범위에 포함되지 않는 특정 사이트나, 혹은 자사 SNS에서 수집되는 범위라도 일부가 아닌 **전문[全文]** 수집이 필요한 경우, 기술적 수집 가능 여부를 판단하여 활용



4-2. 추가 업무 범위 상세 : 라이브 커머스 쇼핑 댓글 실시간 수집

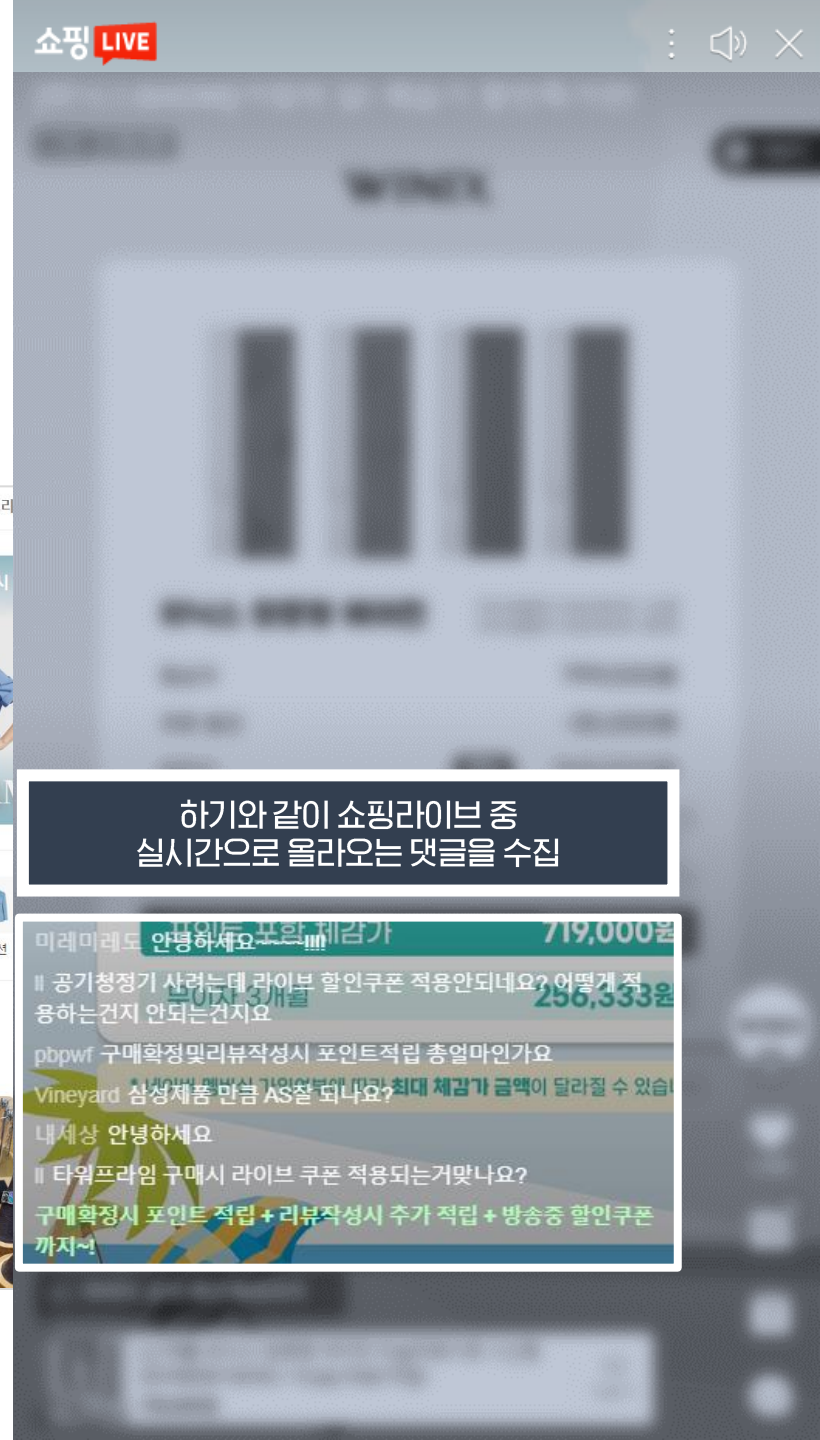
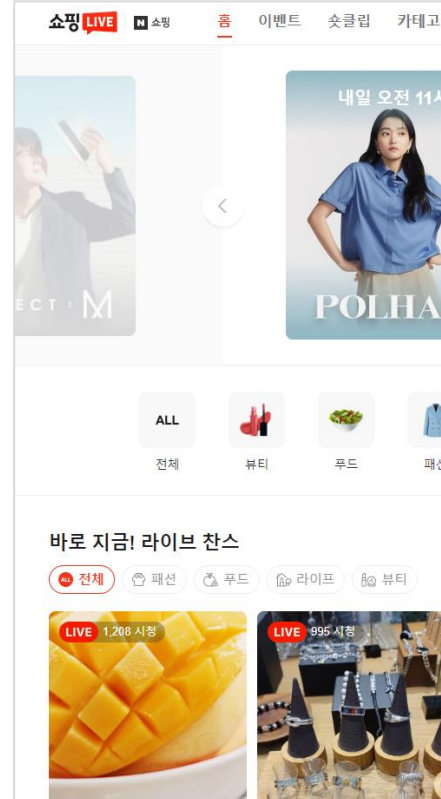
- 새로운 매체로 각광받기 시작한 포털사이트의 '라이브 커머스 쇼핑'에 대한 실시간 댓글 수집 가능
- 해당 댓글 수집은, **실시간 방송 중에 함께 수집이 가능**하다는 장점이 있어 활용도가 높을 것으로 기대, 다만, 아직까지는 모든 방송에서 유의미한 댓글이 생성되는 것이 아니기에 좀 더 연구가 필요한 영역
- 관련한 상세 분석은, 엠포스 데이터랩 사이트에서 보고서로 상세 확인 가능 (<http://bigdata.emforce.co.kr>)

```

1 # config 파일 절대 경로 지정
2 with open(os.path.abspath(os.getcwd()) + '/config/config.json', 'r') as f:
3     CONFIG = json.load(f)
4 CHROME_DRIVER = CONFIG['CHROME_DRIVER'] # 크롬 드라이버 경로
5 URL_FIR = CONFIG["URL_FIR"] # 크롤링할 url 전반부
6 URL_LAT = CONFIG["URL_LAT"] # 크롤링할 url 후반부
7 USER_AGENT = CONFIG["USER_AGENT"] # USER_AGENT
8
9
10 class ShoppingliveCrawler():
11     # 함수 호출
12     def __init__(self, CHROME_DRIVER, URL_FIR, URL_LAT, USER_AGENT):
13         self.chrome_driver = CHROME_DRIVER
14         self.URL_FIR = URL_FIR
15         self.URL_LAT = URL_LAT
16         self.USER_AGENT = USER_AGENT
17         self.driver = None
18         self.initialize_driver()
19

```

네이버 라이브 쇼핑 - 실시간/다시보기 댓글 크롤링 코드 일부



하기와 같이 쇼핑라이브 중 실시간으로 올라오는 댓글을 수집

미레미레도 안녕하세요
 # 공기청정기 사려는데 라이브 할인쿠폰 적용안되네요? 어떻게 적용하는건지 안되는건지요
 pbpwf 구매확정및리뷰작성시 포인트적립 총얼마인가요
 Vineyard 삼성제품 만큼 AS잘 되나요? 최대 체감가 금액이 달라질 수 있습
 내세상 안녕하세요
 # 타워프라임 구매시 라이브 쿠폰 적용되는거맞나요?
 구매확정시 포인트 적립 + 리뷰작성시 추가 적립 + 방송중 할인쿠폰 까지~!

4-3. 추가 업무 범위 상세 : 데이터 분석 관점에서, GA4 등 시각화 플랫폼 설계와 세팅

- 다양한 관점에서 적극적으로 활용되지 못하고 있는 **광고 운영 데이터의 활용성을 극대화**시키고자
데이터 분석 관점에서 기존 운영 체계를 점검하고, 추가 분석 요소를 발굴하여 새롭게 세팅,
중국에는 **루커스튜디오** 등을 통해 시각화시켜 이후에도 **동일한 관점에서 데이터를 추적할 수 있는 체계 마련**

Step 01

기존 플랫폼 사전 점검

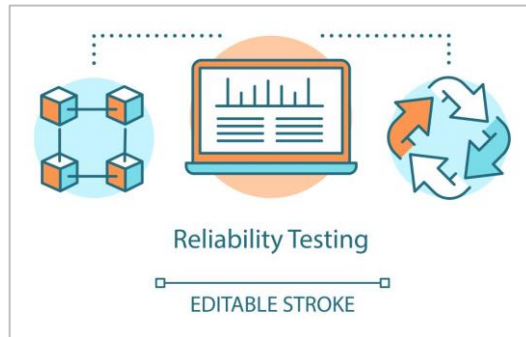
Google Analytics 등 기존의 플랫폼을
데이터 분석 및 기술적 관점에서 점검,
유지 가능 영역과 개선 필요 영역 등
체크리스트를 만들어 고객과 소통
***접근 권한 공유 필요**



Step 02

기본 요소 설계 및 세팅

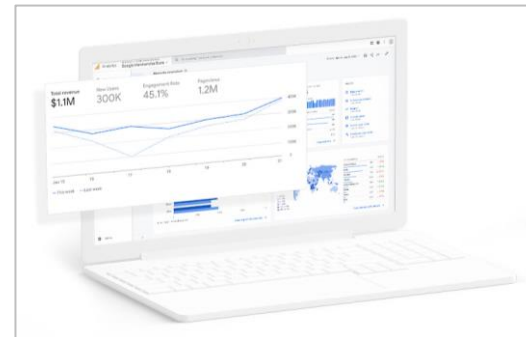
고객과 최종 합의된 내용을 기반으로
기본 데이터를 면밀히 세팅,
일정 기간 동작 상태를 확인하고
안정화시키는데 주력
***고객사 개발실과 직접 소통 필요**



Step 03

추가 세팅 가능 요소 발굴

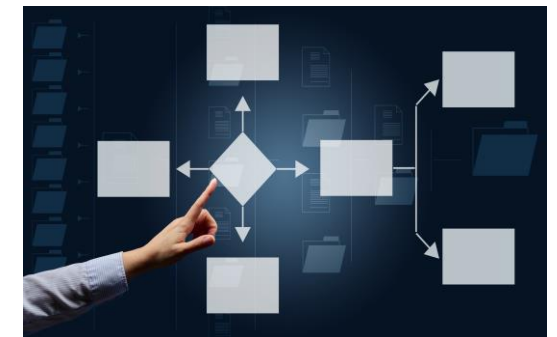
빅데이터 분석 관점에서의
유입 키워드 포트폴리오 구성 및
사이트 유입 흐름 파악 등 분석 위해
추가로 세팅 가능/필요한 요소를
모두 발굴하여 분석 솔루션에 반영



Step 04

유익한 분석 체계 정립

최종 세팅 된 데이터 체계를 기반으로
데이터를 심도 있게 분석, 체계 마련
✓ 데이터 분석 보고서로 고객 전달
✓ 분석 기반 유의미한 데이터 발굴
✓ Tracking 가능한 분석 체계 마련



4-4. 추가 업무 범위 상세 : R Shiny를 활용한 임시 대시보드 구현

- R shiny는 R 코딩 언어를 사용하는 일종의 패키지로, 통계적으로 분석된 결과를 시각화 할 수 있으며 특히 **웹 브라우저(URL)** 기반의 **대시보드** 형태로 구현되기 때문에, URL만 있으면 언제 어디서든 접속이 가능
- 광고 운영 데이터 중, 유입 키워드를 활용해 분석한 결과를 **예로 들면**,
 - 수 년 동안 발생한 수 십 만 개의 유입 키워드들을, 일별 전환 추이에 따라 통계적으로 **군집화(Clustering)**하여 **6개의 그룹 도출**
 - 그룹별로 포함된 수 만 개의 유입 키워드들을 일일이 확인 할 수 없으니 **R Shiny를 통해 대시보드로 구현하여 URL 제공**
 - 대시보드 상에서 **KPI 기준이나 기타 변수 설정 등을 통해 그룹별 유의미성, 포함된 유입 키워드 종류 등을 모두 확인 가능**

Advertising Data Time Series Clustering Analysis

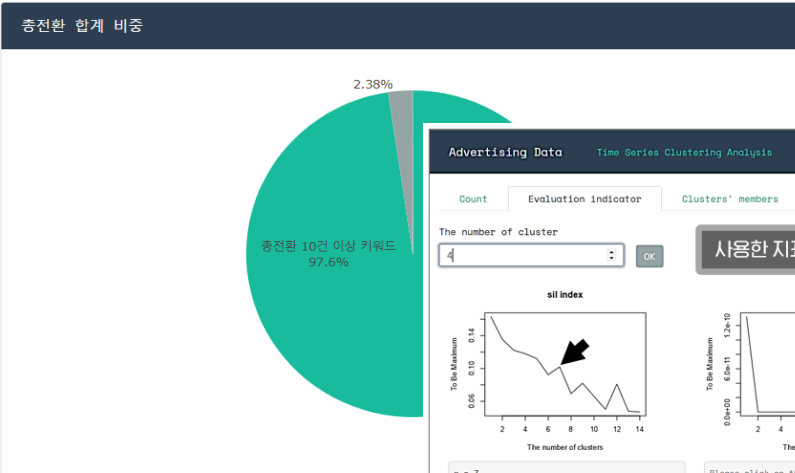
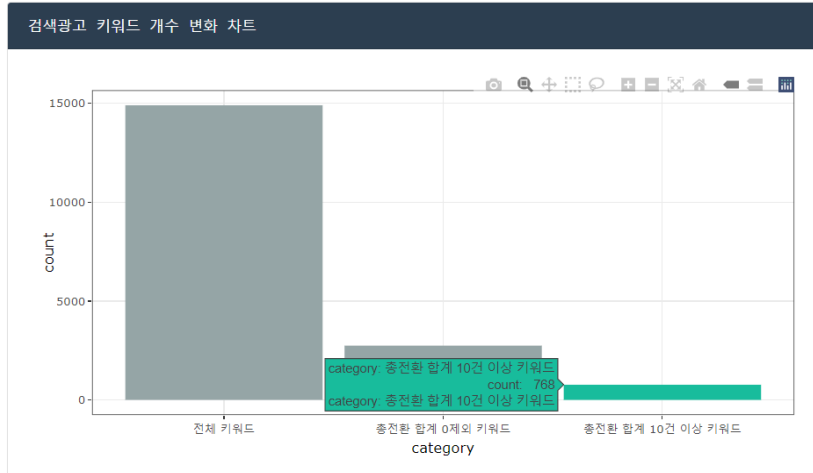
Count Evaluation indicator Clusters' members Cluster detail

KPI select Count select

총전환 10

▶ Analysis

공유 받은 URL을 통해 대시보드에 접속한 후 왼쪽과 같은 **select 영역을 조절**해 원하는 조건의 값 확인 가능



Advertising Data Time Series Clustering Analysis

Count Evaluation indicator Clusters' members Cluster detail

The number of cluster: 4

사용한 지표들을 종합하여, 최종 클러스터링 개수 선정

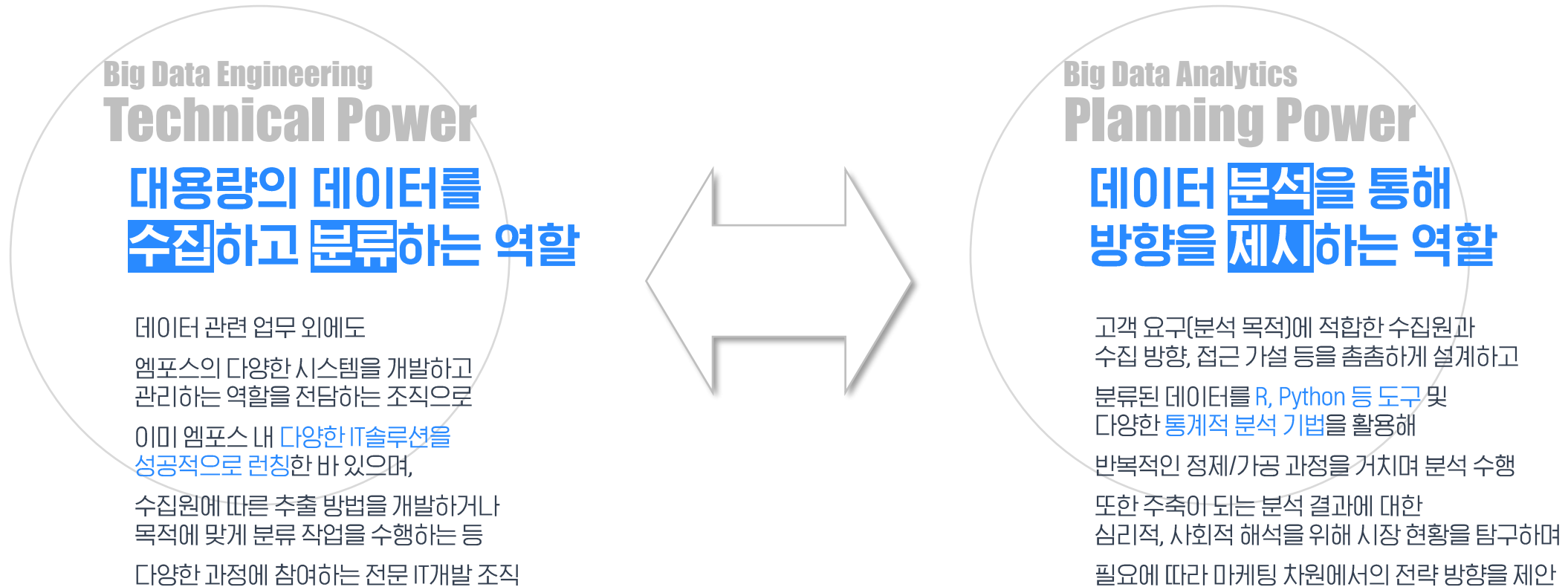
sil index SF index CH index DB index

Modified DB index Dunn index COP index

마우스로 클릭 시, 해당 지표에서의 적합한 클러스터링 개수 (x 값) 을 반환함.

5. 기술적 & 분석적 역량 소개

- 단편적인 결과도출에 그치지 않고 **데이터 연구**라는 본래 취지를 중심에 두고 데이터를 수집하고 분류하는 **Engineering** 영역과 과정을 설계하고 데이터를 분석/해석하는 **Analytics** 영역을 이원화하여 대응
- 또한 엠포스의 Planning 구성원들은 다양한 각도에서의 데이터 해석을 위해 브랜드 전략 컨설팅, 광고/마케팅, 통계, IT솔루션 등 **각기 다른 배경을 가진 전문가들로 구성**



“ 왜 빅데이터인가? ”

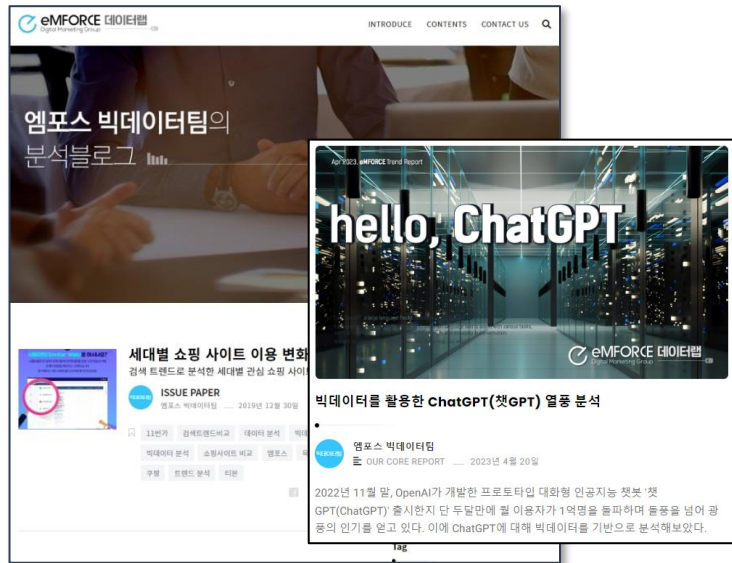
“소비자는 자신과 비슷한 입장의 소비자에게만
허심탄회하게 털어놓는 경향이 있으며,
더욱이 **자신이 꾸며 놓은 환경 속에서** 만큼은
마음속 깊숙이 담고 있는 걱정과 바람을
자연스럽고 분명하게 드러내게 된다.”

- 필립 코틀러 (Philip Kotler)의 『마켓 4.0』 중에서 -

e M F O R C E C o n t a c t P o i n t

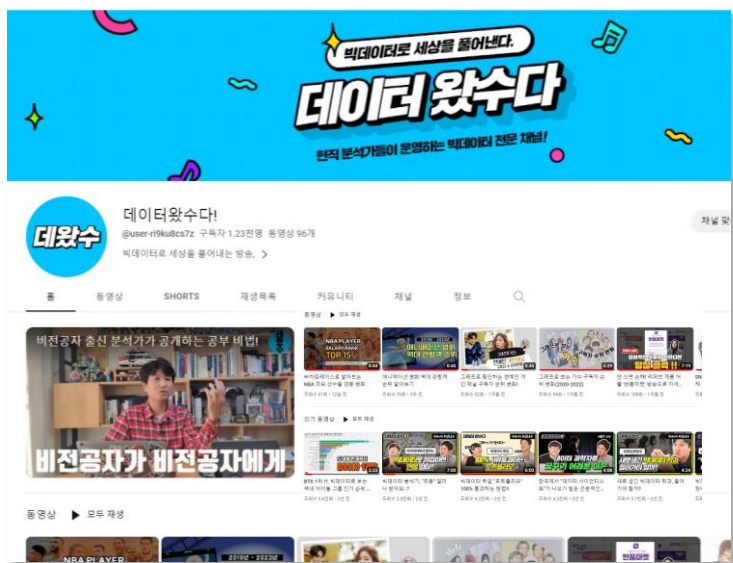
누구나 추정 가능한 내용을 상투적 기교로 분석하는 것이 아닌
누구도 주목하지 않았던 영역에서의 우리만의 발견을 목표로

자체 운영 사이트를 통해
매번 Insight는 아니지만, 어쨌든 심층 분석 보고서 제공



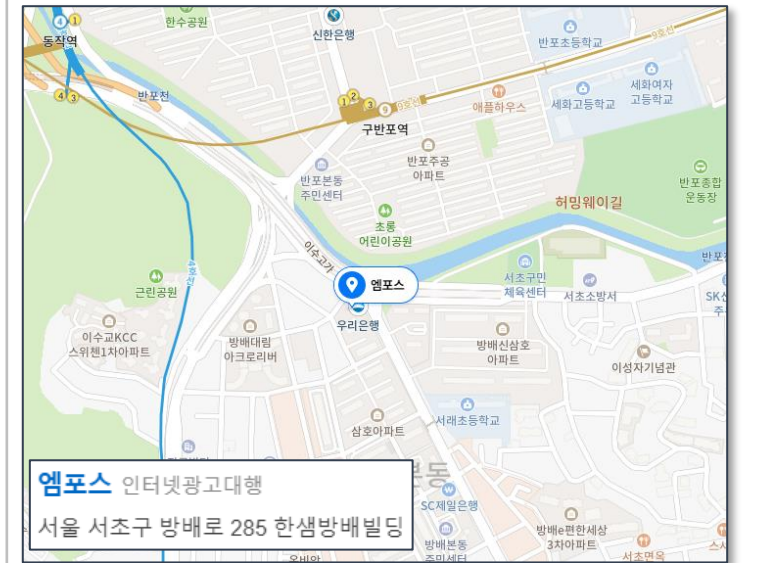
비정기적으로 발간되지만, 언제나 시의 적절한 이슈를
콕 집어 상세히 분석한 결과를 내놓는, 데이터 연구소.
더욱이, 어떤 회사도 흉내 낼 수 없는 **무료**라는 것!
<http://bigdata.emforce.co.kr/>

YouTube 채널을 통해
빅데이터에 대한 의미있는 잡담 나누기

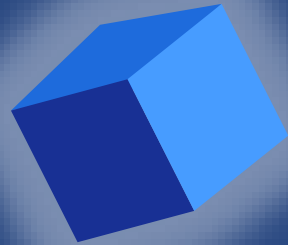


빅데이터와 관련된 소소한 주제들로 수다를 떨지만
듣다 보면 귀가 트이고 눈이 휘둥그레지며,
생각이 많아지는 공식 유튜브 채널 **[데이터왔수다]** 운영
<https://www.youtube.com/channel/UC9vx06TNSIb7IBiURhqYmcQ/>

업무 협업이 필요한 경우
언제든 엠포스 빅데이터실로~



역세권, 숲세권은 아니지만, 럭셔리 한 인테리어를 품은
접견실이 마련되어 있고, 본의 아니게 교통 요충지로
안 막히는 시간대가 없지만, 지상 주차가 가능한 엠포스



End Of Document

Contact Us

Website URL <http://bigdata.emforce.co.kr>

N 박경하 실장 **T** 02. 6177. 1871

eMAIL khbak@emforce.co.kr