

**Data Lab**

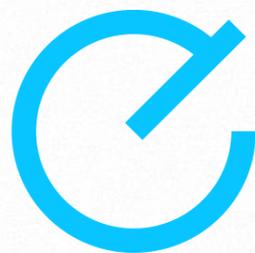
엠포스 데이터전략실

# 빅데이터 분석가들이 답답해서 시작한 구글 애널리틱스4 컨설팅 서비스

데이터 분석 관점에서의  
진단과 세팅, 그리고 분석까지

2024 eMFORCE DATA LAP PROPOSAL





**eMFORCE**  
Digital Marketing Group

## Data Strategy Office **Identity**

**"Pushing the boundaries of the impossible"**

“불가능의 경계를 조금씩 밀어내기”



## Google Analytics

Google Analytics(구글 애널리틱스, 이하 GA)는  
구글에서 제공하는 Marketing Data Platform으로

- ① 자사 사이트에 방문한 User 동향을 다양한 차트로 확인하려는 목적과
- ② Google만의 고유한 분석 시스템을 활용하기 위해 널리 이용되고 있음

# 01

## 우리가 구글 애널리틱스를 대하는 자세

# 구글 애널리틱스는 가장 진보한 추적 시스템

우리는 디지털 광고 업계 출신이 아니라  
사실 구글 애널리틱스를 잘 몰랐습니다.

그런데 알면 알수록 이런 결론에 도달하게 되더군요.

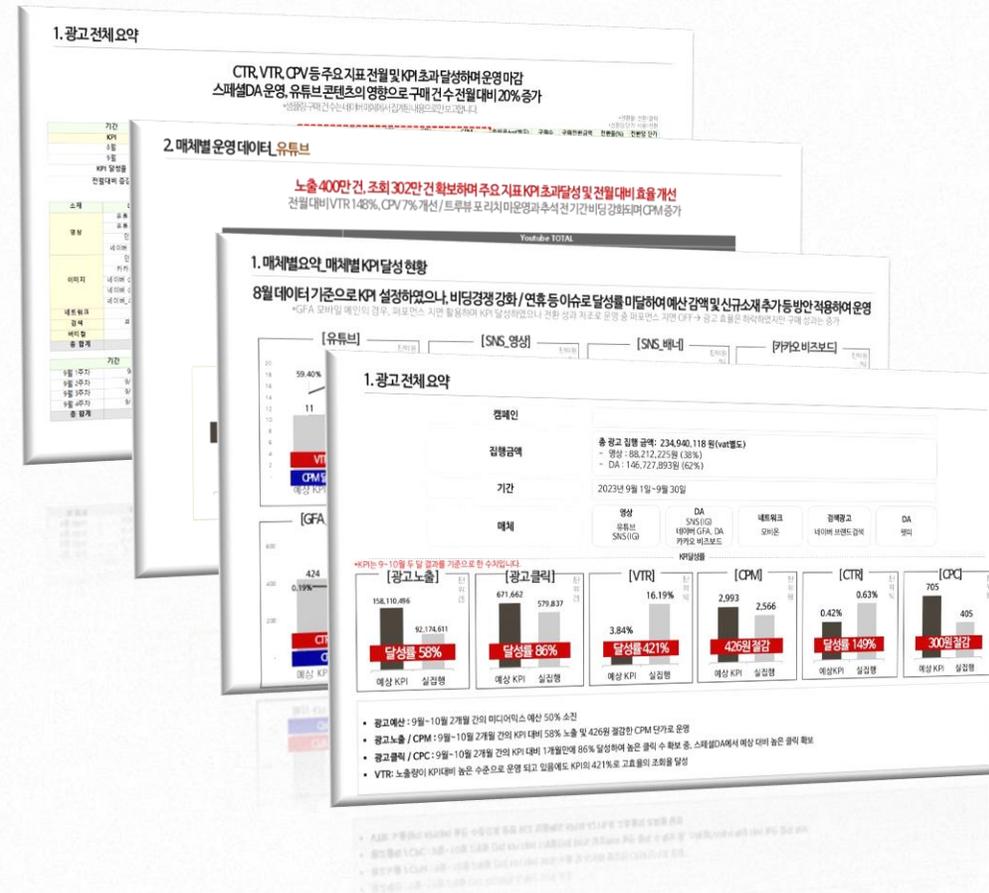
빅데이터의 70% 이상이 이미 온라인에서 생성되고 있고  
온라인 광고가 2019년 전체 광고 시장에서 50%를 돌파했으며  
광고 운영을 통해 생성되고 누적되는 자료만큼  
소비자의 온라인행동을 추적하기에 유용한 자료는 없다는 것.  
그리고 구글 애널리틱스는 가장 진보한 추적 시스템이라는 것.

그럼에도  
온라인 광고를 집행하고 데이터를 활용하는 많은 회사들이  
극히 일부의 단편적인 결과만을 활용하고 있고  
다양한 빅데이터 분석 기법이 범람하는 시대에도 불구하고  
간단한 수치 계산 정도로 성과를 예단하고 있는 것 같습니다.

데이터가 꼭 복잡한 산식을 거쳐야 유의미하다는 것은 아니지만  
데이터의 구조와 특징, 한계의 범위를 이해하지 않은채  
단순 계산되는 데이터는 논쟁거리가 되지 않을 정도로 무의미합니다.

이에 엠포스 데이터전략실은  
다양한 빅데이터 분석 통계 방법론을 적용하여  
소비자의 온라인 행동 전반을 연구하고 있습니다.

여러분의 구글 애널리틱스는 잘 활용되고 있으신가요?



## 구글에서 제시하는 GA4의 특징들을 보면,

[<https://support.google.com/analytics/answer/10089681?hl=ko>]

- 웹사이트와 앱 데이터를 모두 수집하여 고객 여정을 더욱 자세히 파악..
- 세션 기반 데이터 대신 이벤트 기반 데이터 사용..
- 쿠키 없는 측정, 행동 및 전환 모델링과 같은 개인 정보 보호 설정 포함..
- 예측 기능으로 복잡한 모델이 없는 가이드 제공..
- 미디어 플랫폼에 직접 통합하여 웹사이트 또는 앱에서 더 편리하게 작업...



## 구글에서 제시하는 GA4의 특징들을

(<https://support.google.com/analytics/>)

- 웹사이트와 앱 데이터를 모두 수집
- 세션 기반 데이터 대신 이벤트 기반
- 쿠키 없는 측정, 행동 및 전환 모델
- 예측 기능으로 복잡한 모델이 없는
- 미디어 플랫폼에 직접 통합하여 웹사



# So, WHAT?!

대체적으로 기능 구현에 대한 이야기, 심지어 용어도 직관적이지 않은.  
하지만 Analytics의 본질은, **결국 분석을 통해 유의미한 결과를 찾는 것**

타겟팅, 잠재고객, 고객 여정, 모델링, .. 다 좋은 말이지만  
*그래서 어떤 내용을 분석하고 어떻게 적용 가능한지는 별로 와 닿지 않죠.*

빅데이터 분석가로서 처음 마주한 GA4,  
우리는 정작 이런 내용들이 궁금했습니다.

“**왜 세션 수, 이벤트 같은 용어를 쓰는 건가요?**”

ID가 아닌 세션이 갖는 분석적 의미가 있을텐데요. 이벤트는 어떤 기준으로 세팅 되고 있나요?

“**GA4에서 분석할 수 있는 데이터 종류는 총 몇 개나 되나요?**”

세팅을 하려면 측정 가능한 범위나, 데이터로 분석 가능한 범위는 어디까지인지 알아야 하지 않나요?

“**업계마다 이 정도 데이터는 기본적으로 세팅되어 있어야 한다는 등의 기준이 있나요?**”

사이트의 목적이나 구조, 산업 특성에 따라 효과적인 분석으로 판단 가능한 변수의 종류나 기준치가 다를텐데요.

“**사이트에 접속한 소비자들의 이동 경로는 어떻게 파악이 되나요?**”

첫 방문, 재 방문, 장바구니 담고 구매.. 이런 건 우리도 알죠? 그런데 순서를 특정할 수 있나요?

“**디지털 광고 말고 마케팅 전반에 적용할 수도 있지 않나요?**”

작금의 온라인 소비 행동은 그 자체가 소비 전반의 행동이잖아요. 그럼 트렌드 분석이 되지 않나요?

구글애널리틱스, GA4를 제대로 활용하려면 적어도 다음과 같은 요건들을 충족해야 합니다.

### 해당 산업과 자사 **사이트의 구조를 반영한 설계**가 필요합니다.

- ✓ 커머스과 금융 사이트에서의 소비자 행동은 다를 수 밖에 없으며, 같은 커머스 계열이라 할지라도 종합몰과 단일 브랜드몰에서의 행동 또한 차이가 있을 수 밖에 없음을 고려해야 합니다.
- ✓ 예를들어, 우리가 잘 쓰지 않는 '스크롤(Scroll)' 이벤트는 보통 방문자가 페이지의 90% 이상을 내렸을 때 발생하는데 이상태 그대로 설정하는 건 사이트의 특성이 전혀 고려되지 않는 결정일 수 있습니다.

### 구글 애널리틱스와 **빅쿼리**(BigQuery)가 **연동되어** 있어야 합니다.

- ✓ 구글 애널리틱스에서 보여지는 데이터 자체를 보는 것 이상으로 데이터의 활용 가치를 높이려면 빅쿼리와 연동을 통해 가장 날 것의 데이터를 확보한 상태에서 분석이 이뤄져야 합니다.
- ✓ 빅쿼리에 애널리틱스 정보가 연동되면 파이썬 등 코드 작업을 통해 좀 더 많은 분석을 진행할 수 있습니다.

### 제공되는 지표 그대로가 아닌 **별도의 정제와 가공, 방법론**이 필요합니다.

- ✓ 애널리틱스를 통해 제공되는 정보는 이미 잘 정제되고 가공된 자료이기는 하지만 다양한 방법론을 적용하기에는 무리가 있습니다. "데이터 분석 관점에서의 세팅"이라는 것은 결국 여러가지 분석이 가능하도록 데이터의 구조를 만들어 놓는 것을 의미하는데 이러한 과정에서 비로서 필요없는 정보와 의외로 유의미한 정보가 선별되는 경우가 많습니다.

## 우리가 했던 실제 분석 사례 3가지

01. 관심도 분석 - 군집 분석 활용
02. 다채널 유입 경로 분석 (MCF)
03. 쇼핑몰 장바구니 분석 - 연관규칙 분석 활용

그래서 우리는 실제로 이런 분석을 진행 했습니다.

# 01. GA 관심도 분석을 통한 소비자 유형 세분화

‘관심도 카테고리’ 데이터는 하기와 같이 GA 상에서 제공되는 자체 지표로 웹사이트 방문자의 주요 관심 종류를 알 수 있어, 타겟 정의에 효과적

관심도 카테고리(도달범위)	획득			동작			전환 <span>전자상거래</span>			
	사용자	신규 방문자	세션	이탈률	세션당 페이지수	평균 세션 시간	거래수	수익	전자상거래 전환율	
	76 전체 대비 비율(%): 44.19% (172)	68 전체 대비 비율(%): 45.33% (150)	90 전체 대비 비율(%): 45.23% (199)	0.00% 평균 조회: 3.52% (-100.00%)	4.15 평균 조회: 3.55 (16.89%)	00:01:11 평균 조회: 00:01:31 (-22.48%)	0 전체 대비 비율(%): 0.00% (0)	US\$0.00 전체 대비 비율(%): 0.00% (US\$0.00)	0.00% 평균 조회: 0.00% (0.00%)	
1. Food & Dining/Cooking Enthusiasts/30 Minute Chefs	46 (4.55%)	40 (4.40%)	52 (4.55%)	0.00%	3.73	00:00:57	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
2. Lifestyles & Hobbies/Outdoor Enthusiasts	41 (4.06%)	36 (3.96%)	48 (4.20%)	0.00%	3.88	00:01:21	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
3. Home & Garden/Do-it-Yourselfers	39 (3.86%)	34 (3.74%)	46 (4.02%)	0.00%	3.70	00:01:26	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
4. Lifestyles & Hobbies/Green Living Enthusiasts	38 (3.76%)	35 (3.85%)	43 (3.76%)	0.00%	3.86	00:02:03	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
5. News & Politics/Avid News Readers	36 (3.56%)	31 (3.41%)	41 (3.58%)	0.00%	3.66	00:01:36	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
6. Beauty & Wellness/Frequently Visits Salons	33 (3.26%)	27 (2.97%)	37 (3.23%)	0.00%	4.11	00:01:23	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
7. Media & Entertainment/Book Lovers	32 (3.17%)	27 (2.97%)	39 (3.41%)	0.00%	5.18	00:01:34	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
8. Shoppers/Value Shoppers	32 (3.17%)	28 (3.08%)	36 (3.15%)	0.00%	5.00	00:01:14	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
9. Beauty & Wellness/Beauty Mavens	30 (2.97%)	27 (2.97%)	33 (2.88%)	0.00%	4.00	00:00:33	0 (0.00%)	US\$0.00 (0.00%)	0.00%	
10. Lifestyles & Hobbies/Fashionistas	30 (2.97%)	26 (2.86%)	32 (2.80%)	0.00%	4.19	00:01:10	0 (0.00%)	US\$0.00 (0.00%)	0.00%	

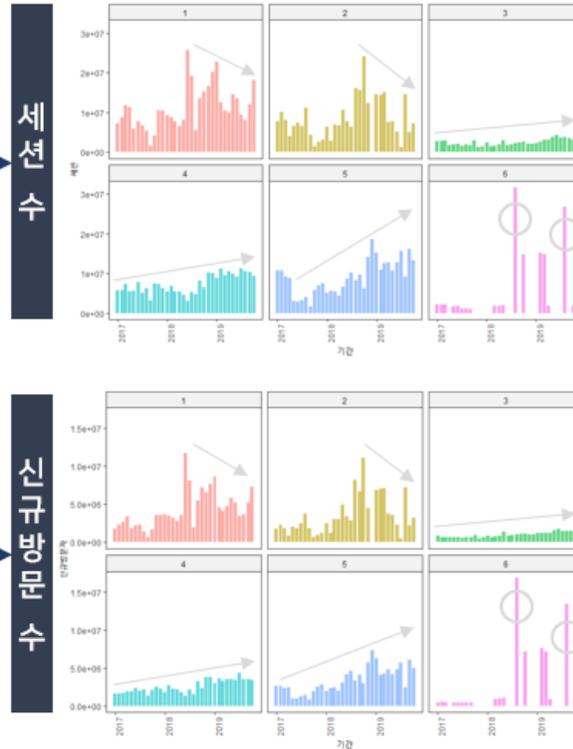
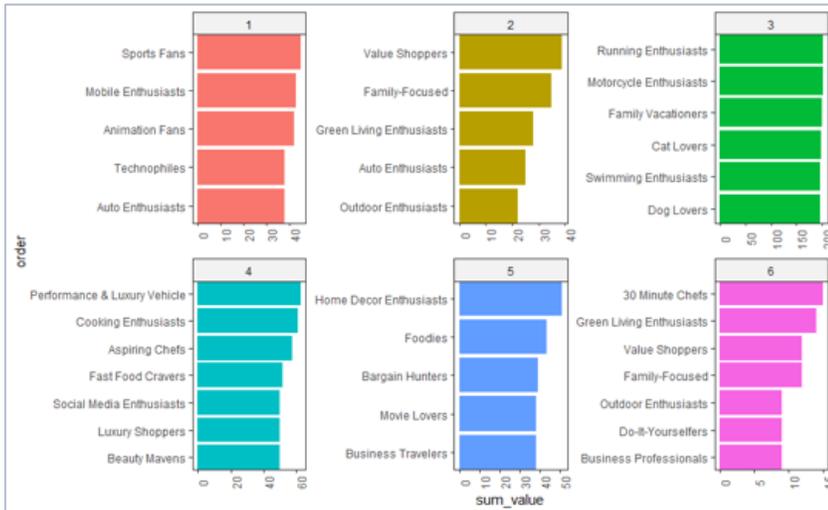
# 01. GA 관심도 분석을 통한 소비자 유형 세분화

## 133개의 세부 관심 항목을 군집화 분석을 통해 6개의 집단으로 유형화

총 133개의 세부 관심 항목 중 상대적으로 변별력 낮은 다수의 항목을 소거하고 72개의 최종 관심 항목을 추출한 뒤  
**군집화 분석을 통해 6개의 집단으로 유형화** → 이후 6개로 구성된 집단에서 최종적으로 최근 트렌드 정의에 필요한 키워드를 추출하여 재정리

최종 추출된 6개의 Cluster에 포함된 하위 관심 항목 종류

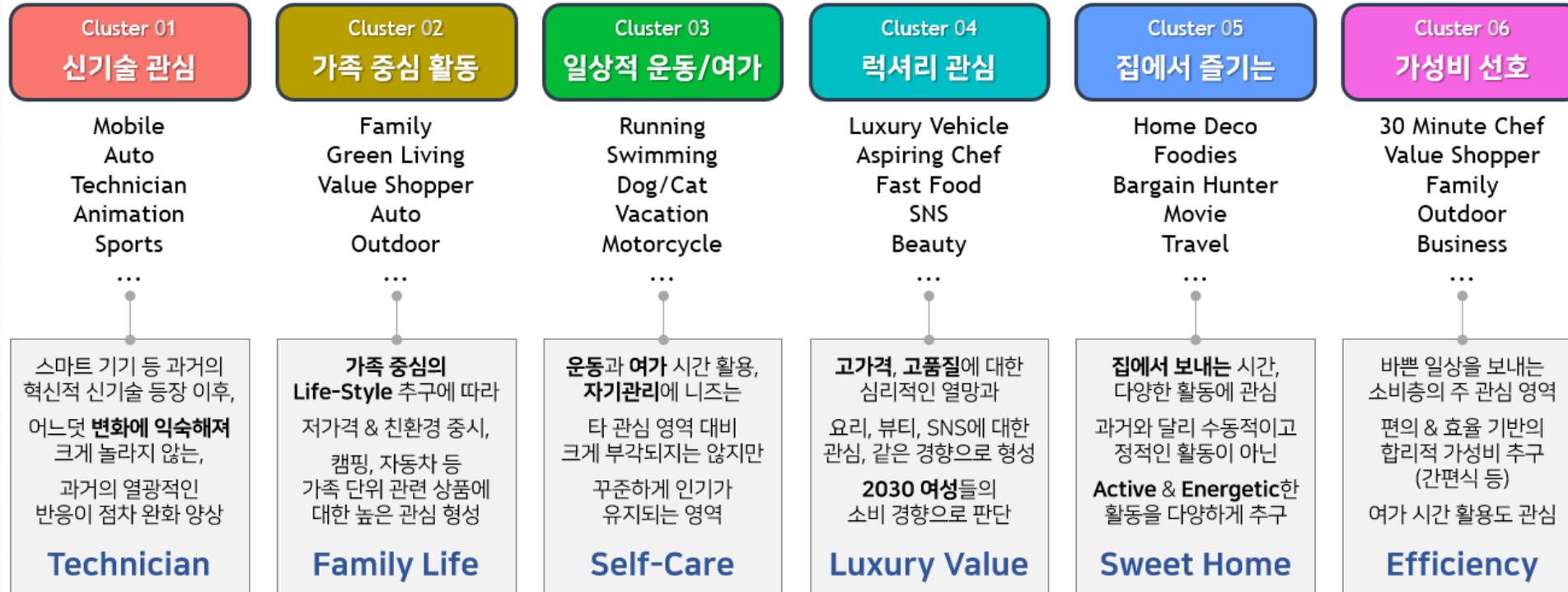
\*하기 그래프는 상위 5위까지만 표기된 결과로  
 실제 분석 과정에서는 전체 포함 항목을 고려하였음



- Cluster 01**
  - 신기술에 대한 관심 두드러짐
  - 세션 및 신규 방문 최근 감소세
- Cluster 02**
  - 가족 중심 활동 관심 두드러짐
  - 세션 및 신규 방문 최근 감소세
- Cluster 03**
  - 러닝, 수영 등 일상적 운동 관심 두드러짐
  - 세션 및 신규 방문 미미한 증가세
- Cluster 04**
  - 럭셔리 쇼핑 & 요리 관심 두드러짐
  - 세션 및 신규 방문 눈에 띄게 증가세
- Cluster 05**
  - 집 안에서의 활동 관심 두드러짐
  - 세션 및 신규 방문 큰 폭 증가
- Cluster 06**
  - 간편식 선호 두드러짐
  - 최근 등장, 비정기적이나 급격히 증가 형성

## 01. GA 관심도 분석을 통한 소비자 유형 세분화

하위에 포함된 다양한 **관심 항목 간의 관계를 비교 분석**하여 해당 유형에 대한 **경향성 정의**



## 01. GA 관심도 분석을 통한 소비자 유형 세분화

# 관심도 분석을 통한 기대 방향

STP 전략 수립의 일환으로 활용하는 것이 의미

### Market Segmentation

미시적 차원



거시적 차원



### Strategic Targeting

미시적 차원



거시적 차원



### Communication Positioning



GA '관심도 데이터'의 주된 활용 목적은

소비자 유형화를 기반으로

- ① 쉽게 자각하기 어려운 현상을 진단하고
- ② 타겟을 재정비하여
- ③ 향후 전략 수립을 위한 기반 자료로서 활용하고자 하는 것

그리고 이런 분석도 진행했습니다.

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

MCF보고서에는 거의 모든 디지털 채널에서 발생한 상호작용이 포함되어 전환 및 거래로 이어진 전반적인 과정을 확인 가능하며 여러 마케팅 채널이 매출 전환에 기여하는 과정을 보여주어 유입 단계별 성과 파악에 용이



최종 전환이 발생하기까지의 이동 경로는 물론 개별 방문 채널에 대한 기여 정도 분석을 통해 각 마케팅 채널의 성격과 효과까지 분석 가능

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

# 소비자의 일반적인 온라인 구매 여정을 예로 들어보면,

다음과 같이 다양한 마케팅 채널을 통해 인지되고 관심이 형성되는 과정을 거치는데  
 기존에 온라인 마케팅에서 주로 활용되는 성과 측정은 최종 전환에만 집중하고 있어 소비자가 어떤 초기 단계를 통해 구매에 이르렀는지까지는 확인 불가

< Online Consumption Process Example >



< 소비자 >



“최초 인지”

“자발적  
정보탐색”

“관심자극”

“호감형성”

“구매자극”

“구매결정”



< 최종구매 >

### 새롭게 고려해야 할 성과 측정 기준

소비자의 구매 여정에서 접하게 되는 채널들에 대한 중요성이 부각  
 또한 각 채널은 시기마다, 또는 이슈마다 소비자에게 자극을 주는  
 성격이 변하게 되므로, **시시각각 변하는 성격 정보** 획득이 필요

### 기존 성과 측정 기준

구매발생 채널 정보만 획득 가능  
 구매 여정 측정 불가

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

## 다채널 유입경로 보고서(MCF)의 장점

MCF 보고서는 [아래 예시로 든 과정에서] 최종 클릭 채널인 「배너 광고」에 전환 성과를 부여하는 것 외에도 각 유입 단계 채널이 상호 어떤 영향을 미치는지 전반적인 성과 분석 정보를 제공

< Online Consumption Process Example >



< 소비자 >



“최초 인지”

“자발적  
정보탐색”

“관심자극”

“호감형성”

“구매자극”

“구매결정”

< 최종구매 >

기존 전환 데이터

최종 클릭 채널인  
“배너 광고”에만 전환 1건 부여,  
그 외 채널은 전환 값 추적 불가

### 다채널 유입경로 보고서(MCF)

최종 클릭 채널인 “배너광고”에는 **직접 전환 값**을 부여하고, 이전 방문 채널에는 **첫 클릭 전환, 지원 전환** 등 유입단계에 맞는 개별 전환 값을 부여함으로써 단계별 마케팅 성과 측정 가능

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

# 4가지 차트 유형으로 해석 가능한 기본 성격

첫 클릭 & 중간 클릭이 최종 클릭보다 많는지 적은지 등에 따라 각 채널이 소비자 구매 단계에서 어떤 역할로 기여하는지 확인 가능

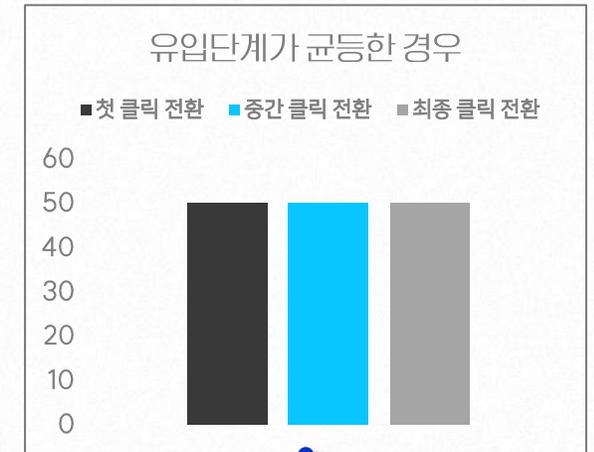
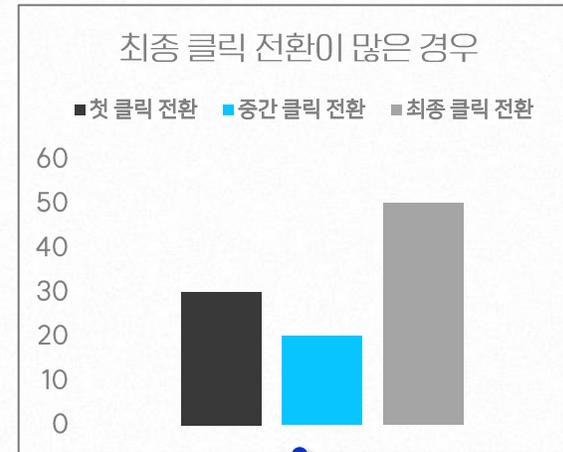
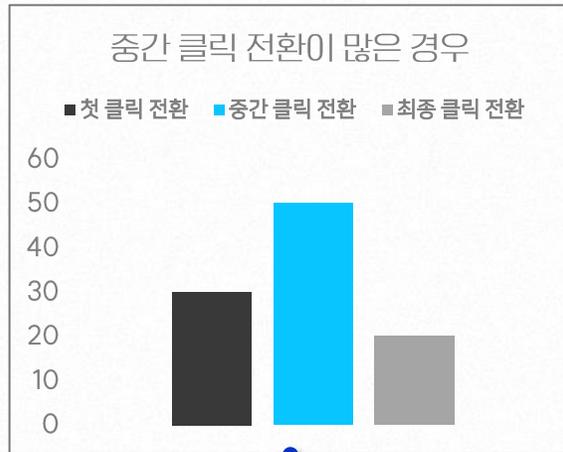
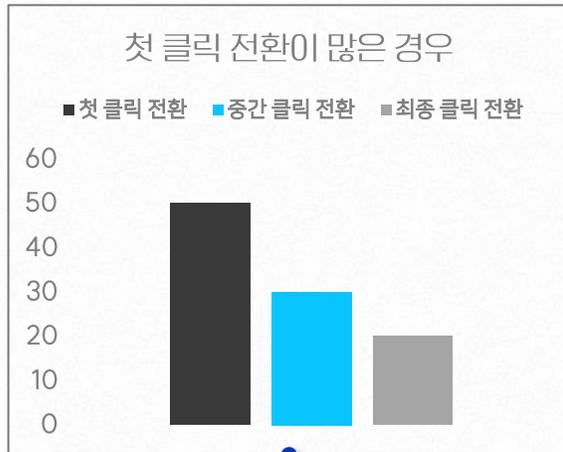
< 채널 성격을 이해하기 위한 기본 차트 유형 >

첫 단계에서의 유입이 가장 높은 마케팅 채널

중간 단계(비교 등)의 유입이 가장 높은 마케팅 채널

최종 전환 단계의 유입이 가장 높은 마케팅 채널

처음 / 중간 / 최종 단계의 유입 및 전환 기여가 균등한 채널



구매 전환 가능성은 낮은 반면 소비자 유입에 긍정적인 영향을 미치는 채널

소비자의 구매 전환을 촉진시키는 채널

유입과 구매 전환 모두 균등한 채널

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

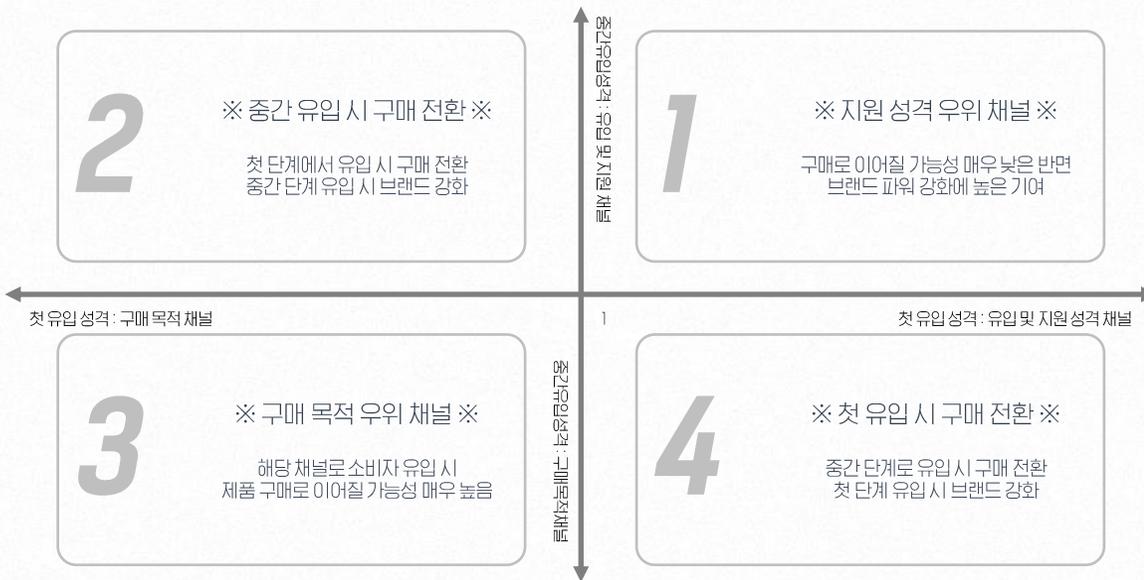
## 마케팅 채널 성격을 정의하기 위한 기본 구조 도식화

## Marketing Channel Performance Evaluation Model



## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

## MKT Channel Portfolio, Strategy Diagnosis



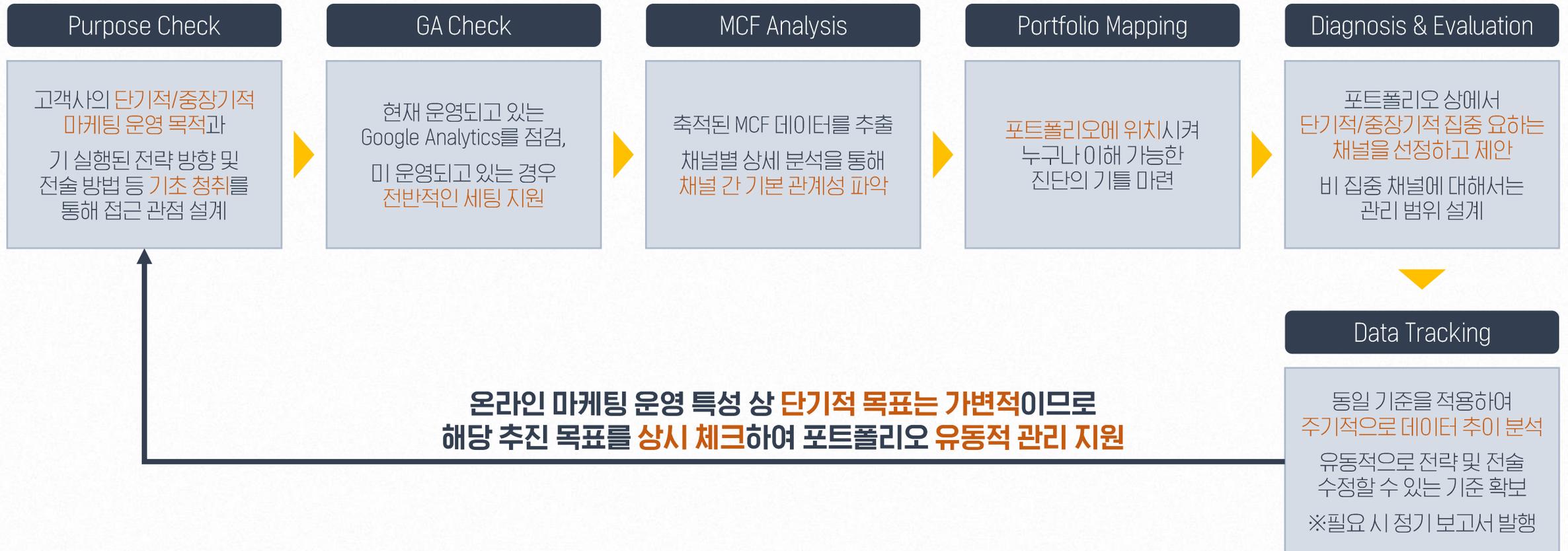
현재의 채널별 운영 취지에 맞게  
성과가 이뤄지고 있는지 진단하려면,

- ① 4분면에서 개별 채널들의 **현 위치를 우선 확인**하고
- ② 전략적 목적에 **맞게 위치해 있는가를 판단**해야 하며
- ③ 이전 시점의 위치에서 **변화가 있는지**  
어떠한 이벤트 영향으로 **변화되었는지를 탐색**,
- ④ 해당 위치를 더욱 공고히 할지, 초기 목적에 맞게  
타 영역으로 이동시킬지 **전략 방향 설계**
- ⑤ 마지막으로 해당 전략을 수행하기 위한 **전술 개발**

## 02. 다채널 유입경로(MCF) 분석을 통한 구매 여정 파악

# MKT Channel Performance Evaluation Model

고객사가 처한, 혹은 의도하는 마케팅 운영 현황 청취를 시작으로 소비자가 유입되는 전체 채널의 역할을 재정의하며 지속적인 체계적인 관리를 통해 각각의 채널에 대한 기대 역할이 안정적으로 자리매김할 수 있도록 상시 지원



세번째 사례로는 이런 것도 있습니다.

## 03. 함께 구매할 가능성이 높은 상품 추출

한 달간의 빅쿼리 GA4 데이터를 통계적 분석 기법에 적용하여, 연관 상품 추천을 위한 장바구니 분석 진행

### 목적

한 달간의 쇼핑몰 실제 데이터를 활용하여 주문 ID별 고유 구매 상품 목록으로부터 상품 간의 규칙을 분석하여 **연관 상품 추천, 페이지 체류 시간 확보 및 매출 증대 목표**

### 방법

**연관 규칙 분석(Association Rule Analysis)**은 하나의 거래나 사건에 포함되어 있는 **항목(item)들 간의 관련성을 찾는 분석방법으로, 일종의 규칙 기반 학습 방법으로 마케팅에서는 일명 장바구니 분석(Market Basket Analysis)**이라고 불리는 기법  
연관 규칙 분석의 대표적인 평가 지표에는 **하기처럼 3가지가 존재**

- **지지도** : 전체 중 A와 B가 동시에 포함된 비율
- **신뢰도** : A 포함 중 B가 포함된 비율
- **향상도** : A 비포함일 때 B 대비 A 포함일 때 B의 비율

### 03. 함께 구매할 가능성이 높은 상품 추출

## 빅쿼리에 파이썬을 연동하여 전처리 및 데이터 분석 진행

GA4에서는 한도 초과 시 데이터 샘플링 발생,  
 모든 Raw Data 확인을 위해 BigQuery 연동

- ✓ 최소 GA4 속성 관리자 권한 필요
- ✓ 속성 1개 당 빅쿼리 1개 연동 가능

BigQuery 다운로드 가능 데이터 용량 제한,  
 모든 Raw Data 확보를 위해 Python 연동

- ✓ 빅쿼리 프로젝트의 서비스 계정 발급
- ✓ "BigQuery 관리자" 권한으로 추가

BigQuery-Python 연동

GA4-BigQuery 연동

BigQuery에서 데이터 적재

Python에서 장바구니 분석 진행

결과 도출 및 활용 제안

Raw Data 형태 확인

행	event_date	event_timestamp	event_name	event_params key	event_params value string_value	eve_int_value
14	20230731	169072920...	purchase	transaction_id	20230730F2CA46	null
				page_title	주문완료, 링크 사는 즐거움 SS..	null
				ga_session_id	null	1690726592
				coupon	null	1101764462
				session_engaged	1	null
				ga_session_number	null	1
				value	null	188200
				page_location	https://pay.ssg.com/order/ord...erComplete.ssg?ordNo=20230730F2CA46	null
				shipping	null	0
				ignore_referrer	true	null
				page_referrer	https://pay.ssg.com/order/ord...	null
				affiliation	null	6005
				currency	KRW	null

데이터 전처리

Item	Adjustment	Afternoon with the baker	Alfajores	Argentina Night	Art Tray	Bacon	Baguette	Bakewell
Transaction								
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0

BigQuery에서 적재된 데이터를  
 빈발 품목 집합 형태로 전처리

\* 빈발 품목 집합(Frequent Item Set)이란, 각 품목별로 발생 횟수(비율)가 특정 값 이상인 품목을 모아놓은 집합



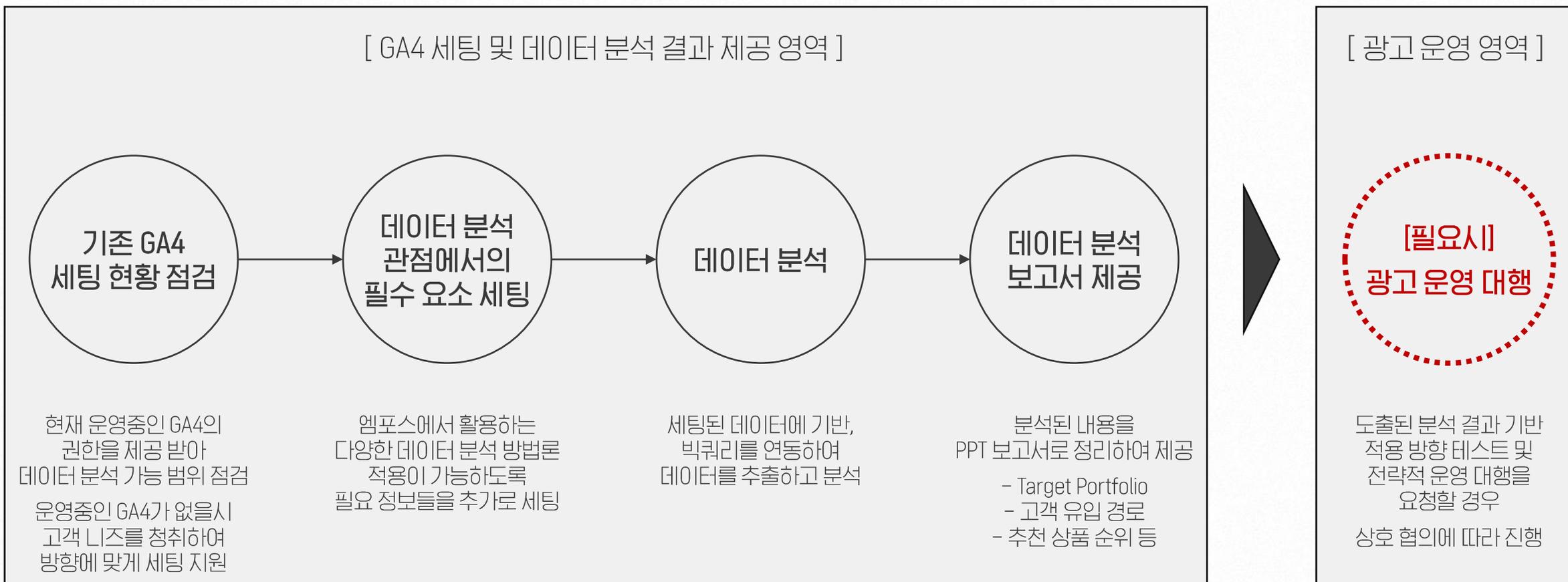
# 우리가 제공할 수 있는 서비스 소개

03

제공 서비스 소개

구글 애널리틱스 컨설팅 서비스 업무 범위

고객으로부터 GA4 계정을 공유받아 **세팅**과 **분석** 지원



## 구글 애널리틱스 컨설팅 서비스 진행 과정

- 01

GA4 운영 현황 및 고객 니즈 청취

사전미팅을 통해 광고 운영목표(KPI) 및 방향성, 최근 광고 운영 성과, 현재 관련 시장의 주요 이슈, GA4 활용 정도 등을 청취하여 GA4 운영 방향 설계
- 02

GA4 데이터 세팅 현황 진단

데이터 활용에 대한 고객 니즈에 기반해 현재 GA4가 어떻게 세팅되어 있는지를 “데이터 분석 관점에서” 면밀히 검토하여 진단
- 03

고객 니즈 기반 GA4 세팅

고객 요구 내용과 기본 데이터 분석 시 필요한 추가내용 등을 점검하여 기본적인 분석이 가능하도록 GA4를 데이터 분석 관점에서 세팅
- 04

빅쿼리 연동 데이터 분석

세팅 된 데이터의 심층 분석을 위해 **빅쿼리를 연동**하여 다양한 데이터를 추출하고 여러 각도로 분석하여 PPT 보고서 형태로 고객에게 전달
- 05

심층 분석 위한 추가 데이터 세팅

기본 데이터 분석에서 발견하기 어려운 소비자의 Needs, segmentation, 사이트 유입 Flow등을 심층 탐색하기 위해 추가 세팅 가능한 데이터를 정의하여 고객에게 제안

# 99

## 엠포스 데이터전략실 소개

# “ 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 -

## “엠포스 데이터전략실은 무슨 일을 하나요?”

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

그 안에 속한 데이터전략실은 **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 아직 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, **몰어보면/ 찾아보고/ 대답하는** 일을 아주 잘 합니다. 그게 우리가 생각하는 **마케팅**이고 **광고**이며, 또 지향하는 **컨설팅**이고, 결국 우리가 추구하는 **데이터 분석**이니까요.

## “ 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.**

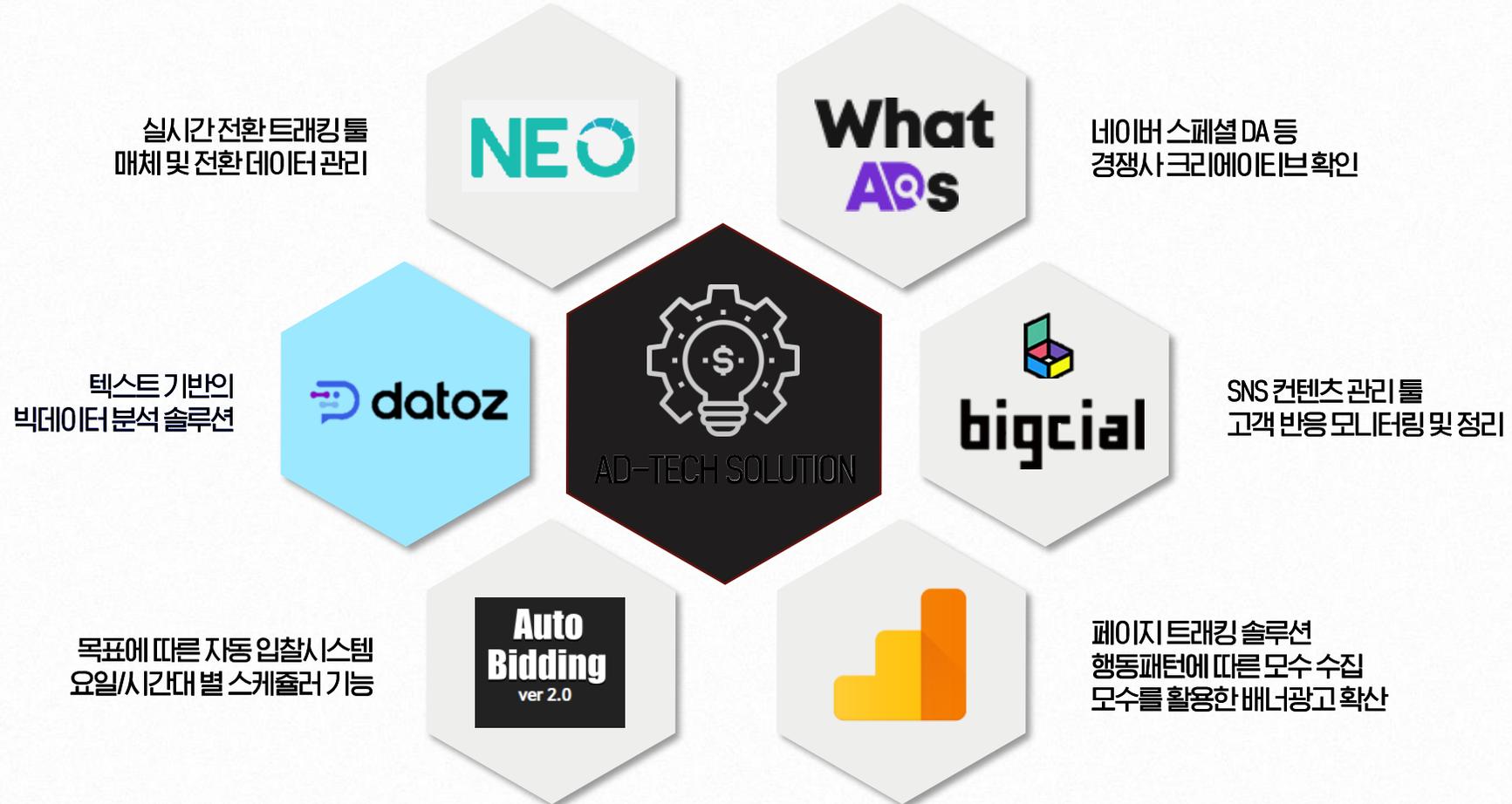
때로는 이미 가지고 있는 데이터만으로도 해결 가능할 수 있고 기존의 여론조사 방법이 더 적합한 경우도, ...사실 꽤 많습니다

...

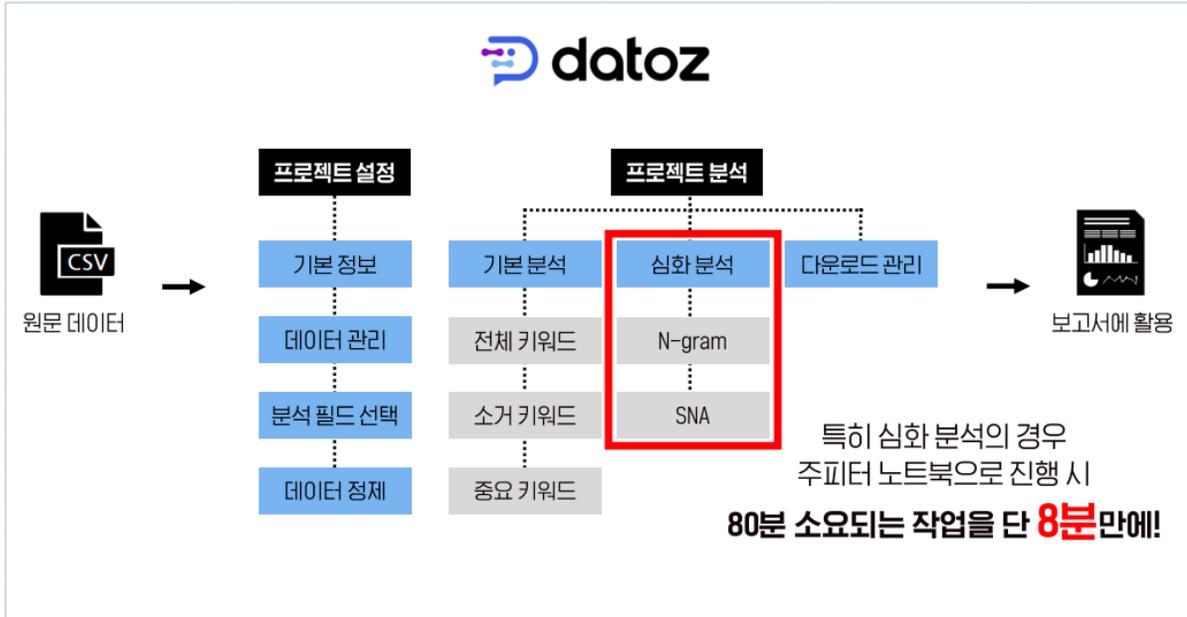
그래서 우리는 특정 데이터를 무분별하게 주장하지 않고 다양한 데이터를 염두에 두고, **융합** 관점에서 고민합니다

**빅데이터**는 기존 데이터의 대체가 아니라 **별도 영역**이고 **데이터 분석**은 코딩에 머물지 않고 **해석을 담아야** 하며 **해석**은 번뜩이는 의견이 아닌 **관점에 기반**해야 하니까요

엠포스는 이미 다양한 솔루션을 보유하고 있는 AD-Tech 기업입니다.



그 중 데이토즈(Datoz)라는, 코딩을 못해도 다양한 텍스트 분석을 할 수 있게 지원하는 무료 솔루션도 개발해 제공중입니다.



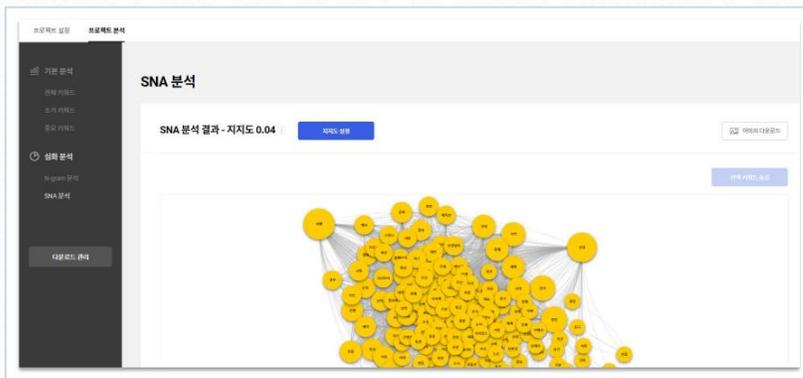
데이토즈는, 본인이 가지고 있는 엑셀 포맷의 텍스트 데이터를 업로드하면 파이썬, R 등 데이터 분석을 위한 코딩을 하지 않고도 문장 속 키워드를 분리, 분해해주거나 키워드 간 연결 관계를 규명해주는 등 다양한 데이터 분석이 가능한 솔루션이다.

게다가 엠포스 내부 서버를 이용하기 때문에

개인 컴퓨터 사양에 따라 속도 차이가 나지 않고 코딩을 짜서 하는 속도보다 빠를 때가 많다.

물론 아직은 베타 버전이라 불편할 수 있지만 분석의 맛이라도 보고 싶은 분들은 아래 URL 클릭! 마음껏 이용해 보시길!

URL : <https://www.datoz.net>



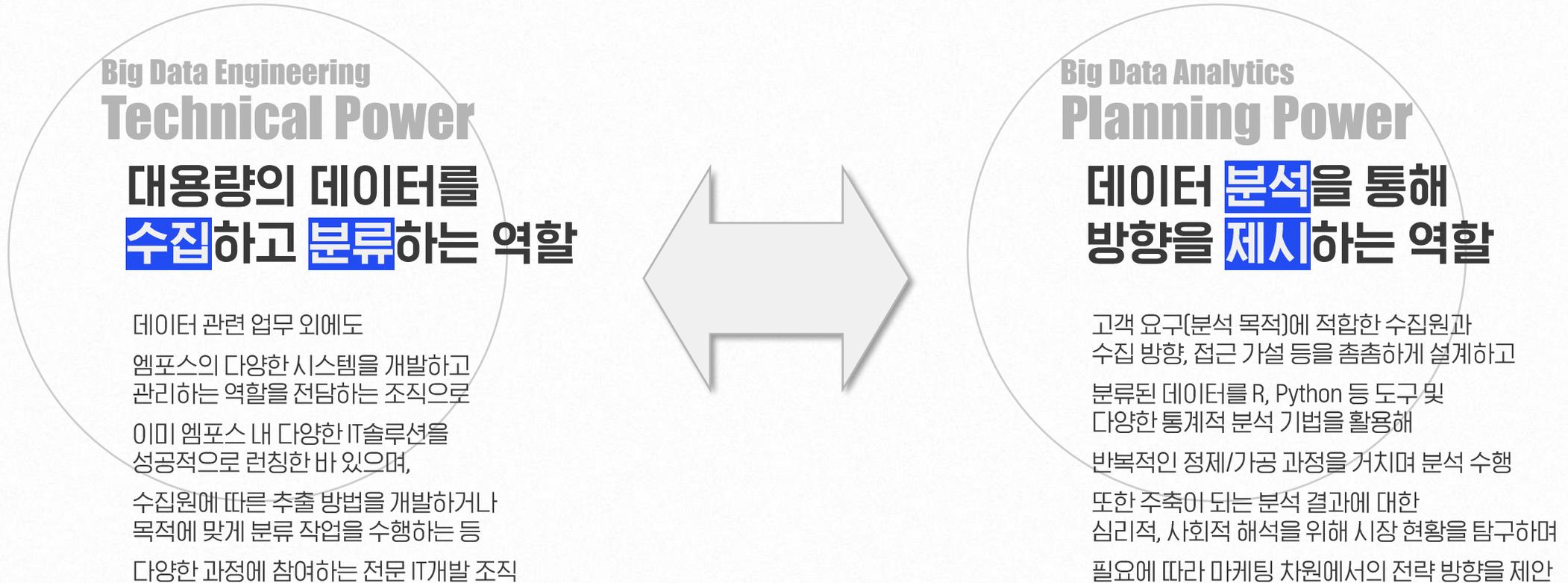


우리가 다루는 데이터는  
정량과 정성 / 온라인과 오프라인을 넘나듭니다.

	Others		Meaningful Big Data					Experimental Performance Data		Extra
	CRM (기업 보유 데이터)	Desk Research (거시적 통계 자료)	보도자료 (기업 이슈)	소셜 데이터 (온라인 텍스트)	검색어 트렌드 (소비자 검색)	쇼핑인사이드 (소비자 검색)	App Review (어플 이용 후기)	Performance (광고 운영 데이터 전반)	Shopping Live (쇼핑라이브 댓글 등)	Data Crawling (국내외 직접 수집)
세부 내용	리뷰, 댓글, 매출 등 고객 내부 데이터 제공 시 통합 분석	재무제표, 각종 통계 자료	보도 언론사 및 보도 내용, 시기, 혹은 댓글 등	트위터, 블로그, 커뮤니티, 인스타 등 자발적 언급 내용	네이버, 구글 등에서의 검색 키워드 종류 등	온라인 쇼핑 상품 클릭 행동 데이터	구글플레이 및 앱스토어 리뷰 내용	유입 키워드 및 사이트 로그 데이터, GA4 등 점검/세팅	실시간 댓글 반응 모니터링 & 원문 내용 상세	해외 레딧(Reddit) 등 접근 가능 채널
추세 분석	○	○	○	○	○	○	○	○	○	○
키워드 분석	-	-	○	○	○	○	○	○	○	○
성/연령 분석	-	-	-	-	○	○	-	○	-	-
분석 접근 방향	제공 목적 기반 심층 분석	시장 개요 및 가설 탐색	언론 평판 및 홍보 내용	소비자 인식 및 태도	소비자 관심 행동	잠재 구매 행동	솔루션 평판	광고 운영 현황 및 성과 측정	구매 점점 호감 정도	광고 운영 현황 및 성과 측정

# 기술적 & 분석적 역량 소개

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



# 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



# 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



# 추가 업무 범위 상세 : 라이브커머스 쇼핑 댓글 실시간수집

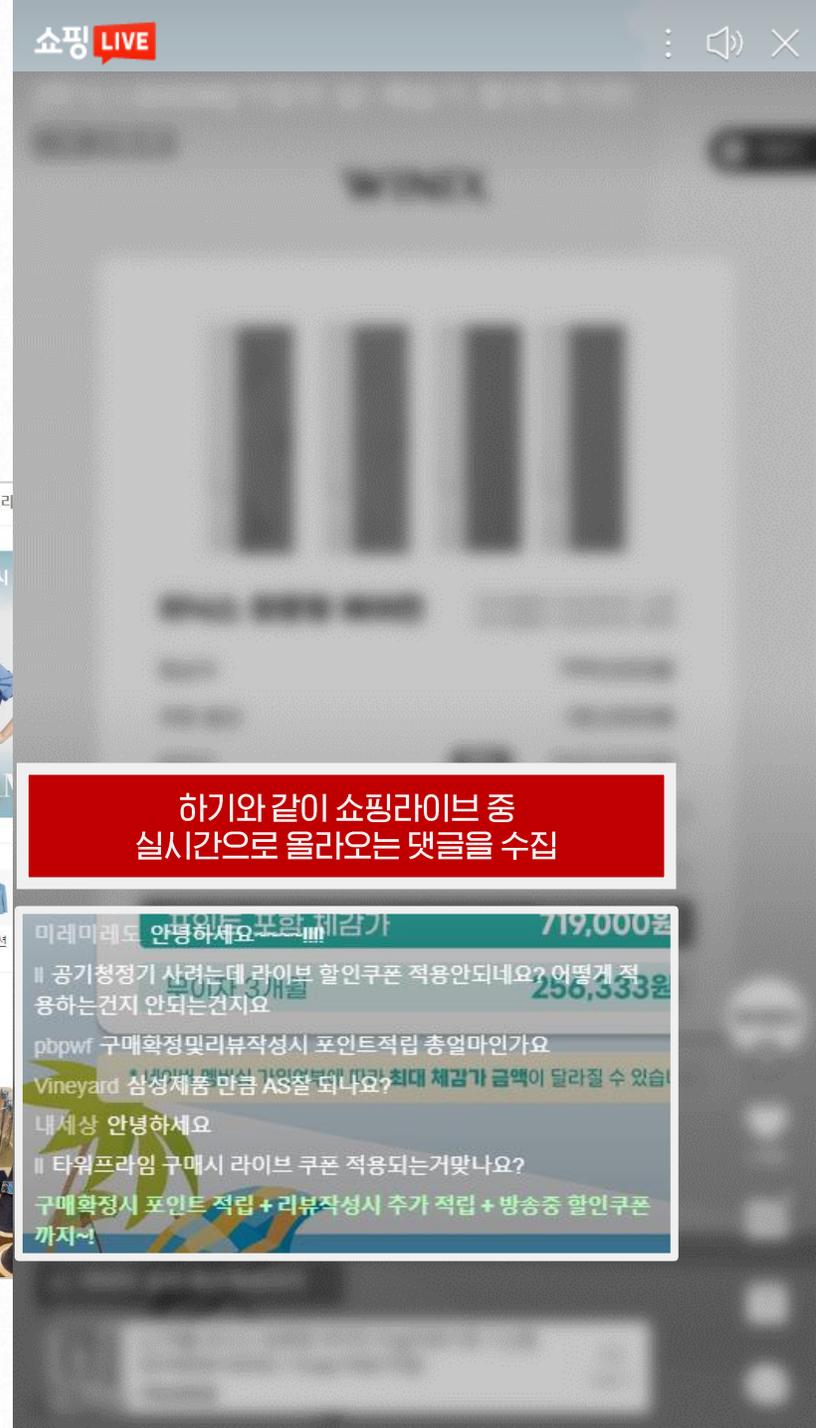
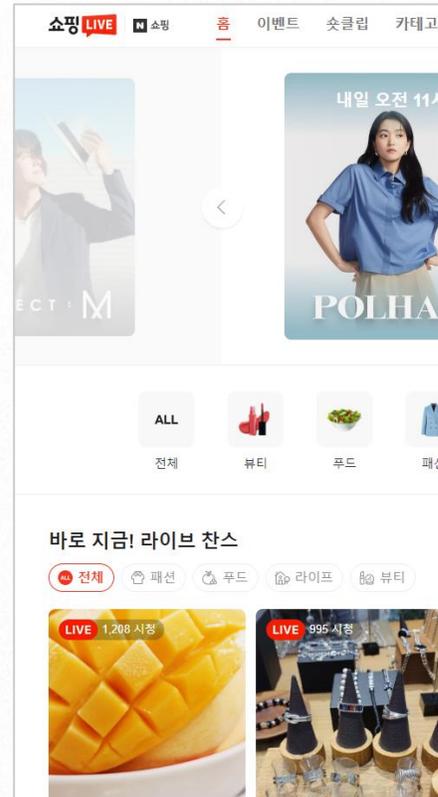
- 새로운 매체로 각광받기 시작한 포털사이트의 '라이브 커머스 쇼핑'에 대한 실시간 댓글 수집 가능
- 해당 댓글 수집은, 실시간 방송 중에 함께 수집이 가능하다는 장점이 있어 활용도가 높을 것으로 기대, 다만, 아직까지는 모든 방송에서 유의미한 댓글이 생성되는 것이 아니기에 좀 더 연구가 필요한 영역
- 관련한 상세 분석은, 엠포스 데이터랩 사이트에서 보고서로 상세 확인 가능 (<http://bigdata.emforce.co.kr>)

```

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



# Data Lab

엠포스 데이터전략실

<http://bigdata.emforce.co.kr> ..... T : 02. 6177. 1871 ..... Email : khbak@emforce.co.kr