Telecom Italia(TIM) 공개 데이터 · 3억 2천만 건 · 2개월

CUBEAGENT · USE CASE BOOK · TELECOM

CubeAgent 활용 사례집 — 통신 데이터 편

도시의 62일을 문장으로 읽다.

밀라노 전역 1만 개 격자, 10분 단위로 쌓인
통화·SMS·인터넷 활동 데이터.
SQL 한 줄 없이 질문만으로 분석한 과정을
실제 조회 결과로 보여드립니다.

3.2억 건	10,000	62일	10분	229
통신 활동 레코드	격자 (100×100)	2013.11.1 ~ 2014.1.1	집계 단위	상대 국가



격자별 인터넷 트래픽 (전 기간 합계, 로그 척도)

이 사례집의 흐름

한 문장 조회에서 스스로 쌓는 분석 지식까지, 6단계

STEP 1 문장으로 추출·시각화
일별 추이·격자 지도
· 자유로운 질문 p.2

STEP 2 대화로 좁혀 가기
국가 → 시간 → 지역
· 자유로운 질문 p.4

STEP 3 스스로 계획하는 복합 분석
연말 감소 원인 분해
· 복잡한 분석 p.5

STEP 4 정의된 숫자, 검증되는 답
요일·공휴일의 함정
· 정확한 숫자 p.6

STEP 5 스스로 찾고 검증하는 탐색
SMS 급증일 추적
· 더 깊은 인사이트 p.7

STEP 6 쌓이는 분석 지식
다른 표현으로 재검색
· 더 깊은 인사이트 p.8

본 사례집의 모든 수치와 차트는 Telecom Italia 밀라노 통신 활동 공개 데이터를 적재한 FORESOFT 데모 환경에서 CubeAgent가 실제로 조회한 결과입니다. 트래픽 값은 원천 데이터의 활동량 지표로, 바이트·분 같은 물리 단위가 아닙니다. M은 백만 단위입니다.

FORESOFT

STEP 1 문장으로 추출·시각화 · 자유로운 질문

10분 단위 3억 2천만 행을 한 문장으로 요약합니다

날짜별 추이 하나를 보려 해도 10분 단위 활동 테이블을 시간 구간·달력과 조인하고, 세 가지 트래픽을 따로 합산한 뒤 화면을 만들어야 했습니다.

지금까지의 방식 BEFORE

1. **요건 정리** "통화"의 범위부터 합의
2. **테이블·키 확인** 통신활동·시간구간·달력 조인
3. **SQL 작성·검증** 3억여 행 합산, 수치 대조
4. **BI 화면 개발** Spotfire 추이 차트·배포
5. **수정 요청 반복** "구역별로도 볼 수 있을까요?"

작성해야 하는 쿼리 (예시)

```
SELECT d.날짜,
  SUM(a.인터넷트래픽)            AS 인터넷,
  SUM(a.발신통화 + a.수신통화)   AS 통화,
  SUM(a.SMS발신 + a.SMS수신)     AS SMS
FROM   통신활동 a
JOIN   시간구간 t ON a.시간구간키 = t.시간구간키
JOIN   달력     d ON t.날짜키     = d.날짜키
GROUP  BY d.날짜
ORDER  BY d.날짜;
```

'통화'가 발신만인지, 발신+수신인지부터 요청자와 다시 맞춰야 합니다.

CubeAgent AFTER

나 **날짜별 인터넷·통화 트래픽 추이 보여줘**

의도 TREND · .인터넷 트래픽 · .전체 통화 트래픽 · 차원 달력.날짜

평일·주말이 번갈아 나타나는 톱니 패턴이 뚜렷하고, **12월 23일 이후 급감**합니다. 1월 1일 인터넷(53.2M)은 기간 최고치인 11월 5일(111.1M)의 48%입니다. 통화는 요일 차가 더 커서 일요일이 평일의 절반 수준입니다.

인터넷 트래픽 (일별)



통화 트래픽 (일별)



답변과 함께 남는 조회 정의 (실제 생성값)

```
{
  "intent": "TREND",
  "measures": [".인터넷 트래픽",
               ".전체 통화 트래픽", ".전체 SMS 트래픽"],
  "dimensions": ["달력.날짜"],
  "intentParameters": { "timeGranularity": "DAY",
                        "aggregationMode": "SUM" },
  "ccq": { "chartType": "line", "xAxis": "달력.날짜",
           "yAxis": ".인터넷 트래픽" }
}
```

'통화'의 정의는 모델이 들고 있습니다

.전체 통화 트래픽은 발신과 수신을 합한 값으로 모델에 한 번 정의돼 있고, 발신·수신만 보고 싶으면 **.발신 통화 트래픽**처럼 이미 정의된 측정값을 고르면 됩니다.

10분 단위 원천 행을 날짜로 올려 합산하는 경로(통신활동 → 시간구간 → 달력)도 모델의 관계로 정해져 있습니다.

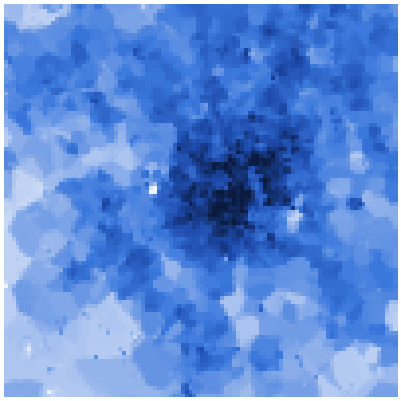
STEP 1 문장으로 추출·시각화 · 공간·시간·국가

어디서, 언제, 누구와 질문의 축만 바꾸면 됩니다

격자 지도, 시간대 패턴, 상대 국가, 행정 구역. 분석 축이 달라도 묻는 방식은 같습니다.

공간 DISTRIBUTION

나 격자별 인터넷 트래픽을 지도로 보여줘



11% 상위 1% 격자(100개)의 트래픽 비중

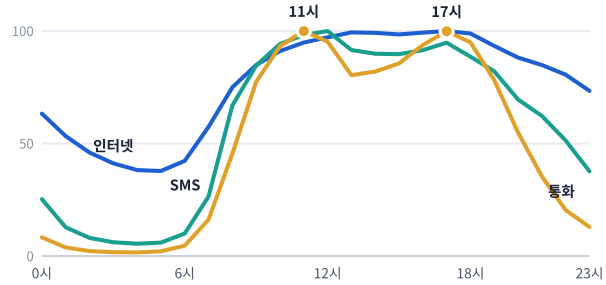
48% 상위 10% 격자 (1,000개)의 비중

적음 많음 (로그)

트래픽은 도심 한가운데로 몰립니다. 가장 많은 격자(61, 52)는 중앙값 격자의 약 46배입니다. 격자 1만 개는 쿼리 한 번의 행 제한(1,000행)을 넘어 에이전트가 격자Y 구간으로 나눠 10번 조회했습니다.

시간 DISTRIBUTION

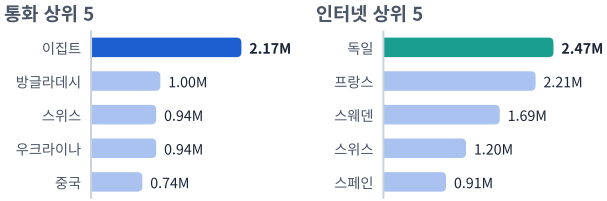
나 시간대별 통화·SMS·인터넷 패턴 비교해줘



통화는 11시·17시에 두 번 정점, SMS는 12시, 인터넷은 13~18시 내내 고원입니다. 밤 0시 인터넷은 정점의 63%가 남지만 통화는 8%로 떨어집니다.

국가 RANKING

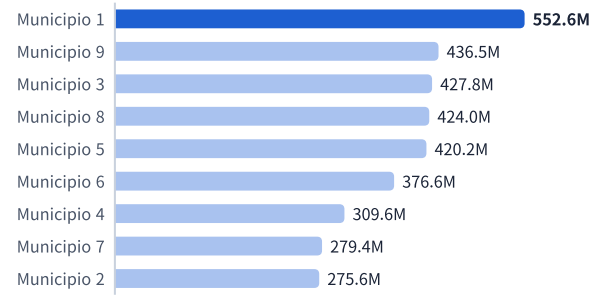
나 이탈리아 말고 어느 나라와 많이 통신해?



통화와 인터넷의 상대국 구성이 전혀 다릅니다. 통화는 이집트·방글라데시, 인터넷은 독일·프랑스·스웨덴이 상위입니다. 국가 코드가 매핑되지 않은 트래픽은 순위에서 제외했습니다.

행정 구역 RANKING

나 밀라노 구역별 인터넷 트래픽 순위는?



도심 Municipio 1이 552.6M으로 1위, 가장 적은 Municipio 2의 2배입니다. 시 경계 밖(Outside Milano)은 2,050.5M으로 전체의 37%입니다.

시간대 차트는 세 측정값의 규모가 달라 각 측정값의 최대 시간대를 100으로 환산했습니다. 구역 차트는 Outside Milano(시 경계 밖 전체)를 제외했습니다.

STEP 2 대화로 좁혀 가기 · 자유로운 질문

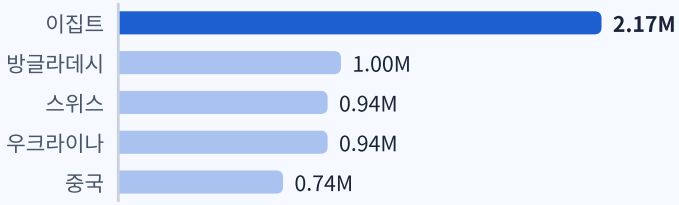
국가 → 시간 → 지역, 질문을 이어 가며 좁혀 갑니다

앞선 대화의 조건(국가 = 이집트)을 이어받아, 같은 대상을 시간과 공간의 축으로 차례로 비춰 봅니다.

나 **이탈리아 말고 통화가 가장 많은 나라는?**

1

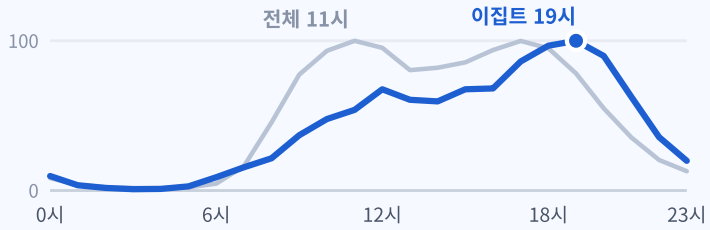
이집트가 2.17M으로 1위, 2위 방글라데시(1.00M)의 2.2배입니다.
국가 코드가 매핑되지 않은 '미상' 트래픽(5.7M)은 특정 국가가 아니므로 순위에서 제외했습니다.



나 **이집트 통화는 주로 언제 해?**

2

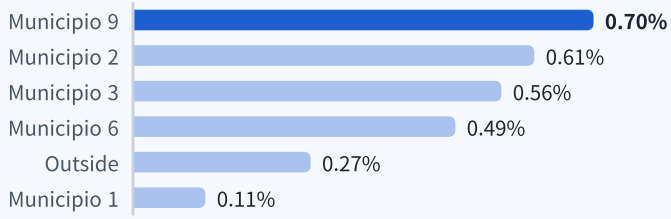
저녁 19시가 정점입니다. 전체 통화의 정점(11시)과 8시간 차이가 납니다.
17~20시에 집중되고, **97.5%가 발신**입니다. 업무 통화보다는 퇴근 후 개인 통화에 가까운 패턴입니다.



나 **그럼 어느 지역에서 많이 걸어?**

3

북부 주거 지역 Municipio 9입니다. 구역 전체 통화 중 이집트 비중이 0.70%로, 도심 Municipio 1(0.11%)의 6.3배입니다.
건수는 면적이 넓은 외곽이 가장 크므로 **구역별 비중**으로 비교했습니다. 평균은 0.38%입니다.



POINT 질문 3개로 **상대 국가 → 통화 시간 → 거주 지역**의 윤곽이 나왔습니다. 다음 질문이 이전 조건을 그대로 이어받고, 건수·비중 중 어떤 기준으로 비교할지도 에이전트가 판단해 이유를 밝힙니다. 특정 커뮤니티 대상 요금제나 마케팅을 기획할 때의 출발점이 됩니다.

STEP 3 스스로 계획하는 복합 분석 · 복잡한 분석

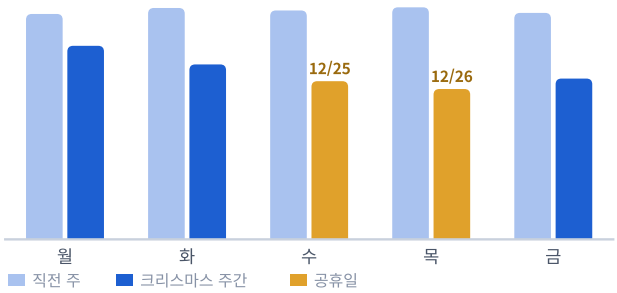
“왜 줄었지?”라고 물으면 분석 단계를 스스로 설계합니다

나 **크리스마스 주간에 인터넷 트래픽이 왜 이렇게 줄었어?**

분석 계획 — ① 직전 주 같은 요일과 비교해 규모 확인 → ② 공휴일 효과 분리 → ③ 구역별 분해 → ④ 시간대별 분해. 비교 기간: 직전 주 12/16~20(월~금) vs 크리스마스 주간 12/23~27(월~금).

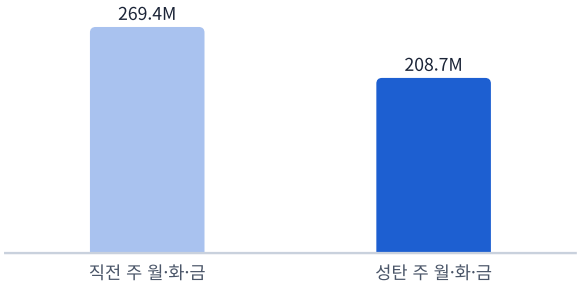
① 전체 규모

−26.8% 인터넷 451.2M → 330.4M
통화는 −38.6% (59.7M → 36.7M)

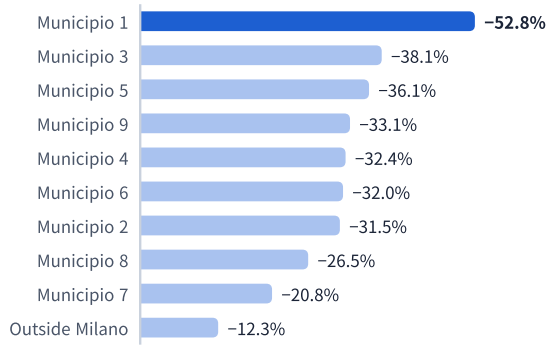


② 공휴일 효과 분리

−22.5% 12/25·26 공휴일을 빼고
월·화·금끼리만 비교해도 감소

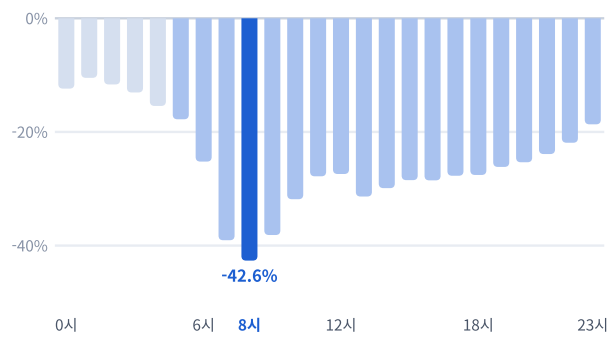


③ 구역별 감소율



도심 Municipio 1은 **절반 이하**(−52.8%)로, 외곽은 −12.3%에 그쳤습니다.

④ 시간대별 감소율



출근 시간 08시가 −42.6%로 가장 크고, 새벽 0~4시는 −10~−15%입니다.

결론: 감소의 주요인은 ‘통근·업무 인구의 이탈’입니다.

공휴일을 통제해도 감소가 이어지고, 업무 지구일수록·출근 시간대일수록 감소가 큽니다. 연휴 기간 기지국 용량을 도심에서 주거 지역으로 재배분하는 계획의 근거로 쓸 수 있으며, 다음으로는 **같은 기간 외국 번호 트래픽 변화**(관광 유입)를 확인해 볼 것을 제안합니다.

STEP 4 정의된 숫자, 검증되는 답 · 정확한 숫자

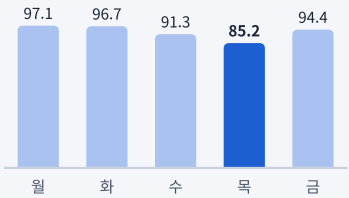
그럴듯한 숫자가 아니라 비교해도 되는 숫자로 답합니다

요일 비교처럼 단순해 보이는 질문에도 함정이 있습니다. 기간 안에 요일마다 며칠이 들어 있는지, 공휴일이 어느 요일에 끼었는지에 따라 결론이 뒤집힙니다.

나 **평일 중에 통화가 가장 적은 요일은?**

흔한 답 · 요일별 합계

목요일



합계 85.2M, 월요일보다 12% 적어 보입니다. 하지만 62일 중 목요일은 **8일**, 나머지 요일은 9일입니다.

① 일평균으로 환산

수요일?



이번엔 수요일(10.14M)이 최저입니다. 그런데 수요일에는 **크리스마스와 새해 첫날**이 들어 있습니다.

② 공휴일·연말 통제

차이 없음



공휴일 6일과 12/23 이후를 빼면 평일 일평균은 11.28~11.64M, 차이는 약 3%입니다.

평일 간에는 뚜렷한 차이가 없고, 월요일이 약간 낮은 정도입니다. 요일별 합계는 기간에 포함된 요일 수(목요일만 8일)와 공휴일 위치에 따라 왜곡되므로, 일평균으로 바꾸고 달력의 공휴일 정보로 통제한 뒤 비교했습니다. 뚜렷한 차이는 평일과 주말 사이에 있습니다(토 7.33M, 일 5.45M).

구분	원천 테이블에 직접 SQL을 생성하는 방식	CubeAgent (시맨틱 모델 기반)
지표 정의	'통화'가 발신인지 발신+수신인지 매번 해석	.전체 통화 트래픽 / .발신 통화 트래픽 등 **정의된 측정값** 사용
달력 속성	공휴일 목록을 따로 구해 쿼리에 하드코딩	달력 차원의 **공휴일 여부·명칭**으로 바로 통제
코드 품질	매핑 안 된 국가 코드가 '미상'으로 섞여 순위에 등장	국가 **매핑상태** 속성으로 미상을 분리해 제외·명시
필터 전파	조인 경로가 맞는지 결과를 보고 짐작	필터가 원천 테이블까지 전달되는지 **조회 전 검증**

21.2%
전체 SMS 중 국가 코드 'UNKNOWN' 비중

국가명으로 SMS 순위를 내면 '미상'이 이탈리아 다음 2위가 됩니다(132.4M). 통화에서는 1.0%, 인터넷에서는 0.03%에 불과해 **SMS 분석에서만 크게 왜곡**됩니다. CubeAgent는 이런 데이터 품질 정보를 답변과 함께 밝힙니다.

트래픽의 '단위'
측정값 설명도 모델의 일부

이 데이터의 트래픽 값은 바이트나 통화 분이 아니라 원천 데이터가 제공하는 **활동량 지표**입니다. 그래서 인터넷과 통화의 절대 크기를 직접 비교하지 않고, 필요할 때는 각자의 최댓값을 100으로 환산해 패턴을 비교했습니다.

STEP 5 스스로 찾고 검증하는 탐색 · 더 깊은 인사이트

“평소와 다른 날”을 찾고 가설을 세워 스스로 검증합니다

무엇을 봐야 할지 모르는 질문에도, 에이전트가 기준선을 세우고 벗어난 날을 찾은 뒤 원인을 좁혀 갑니다.

나 **두 달 동안 평소와 다른 날이 있었어?**

① 관찰: SMS/통화 비율의 이탈 발견

SMS를 통화로 나눈 일별 비율은 평일 1.02, 주말 1.24 안팎인데 **12/25는 2.22, 1/1은 1.87**로 튑니다.

② 가설: 공휴일이라서? 기각

다른 공휴일(11/1 1.28, 12/8 1.30, 12/26 1.41)은 주말 수준에 그칩니다. 공휴일만으로는 설명되지 않습니다.

③ 가설: 통화가 줄어든 착시? 기각

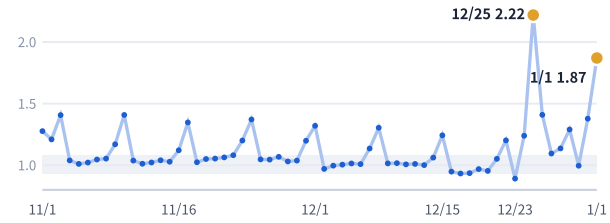
12/25 SMS는 12.07M으로 주말(약 7~9M)이 아니라 **평일 수준**입니다. 분모가 아니라 SMS 자체가 늘었습니다.

④ 시간대 추적: 자정과 성탄 오전 핵심

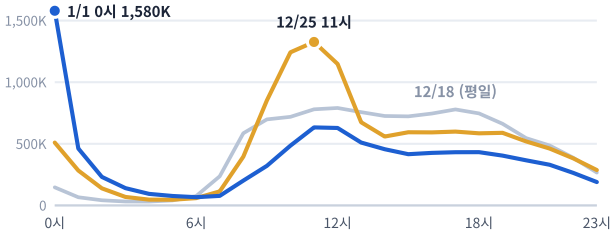
1월 1일 0시 SMS 1.58M은 평소(12/18 0시)의 10.8배, 평일 정점 시간의 2배입니다. 12/25는 10~12시 SMS가 평일의 1.7배인데 같은 시간 통화는 오히려 적습니다.

일별 SMS ÷ 통화 비율

회색 띠 = 평상 평일 범위(0.93~1.08)



시간대별 SMS 트래픽



에이전트가 적용한 검증 방법

비율로 정규화
요일·계절에 따른 전체 증감을 지우고 구성 변화만 보이게

기준선 분리
평일과 주말의 정상 범위를 따로 계산

대조군 비교
다른 공휴일과 비교해 ‘공휴일 효과’ 가설 검증

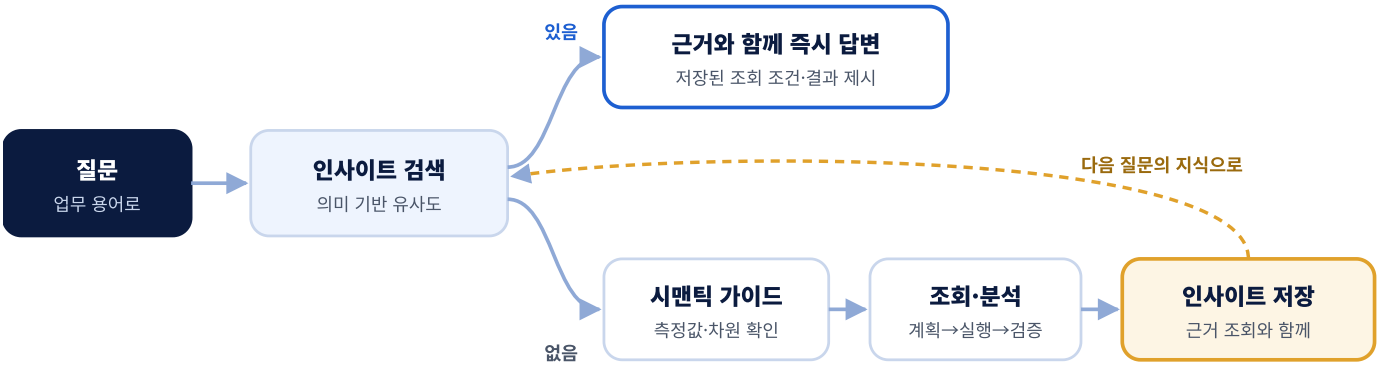
분자·분모 분해
비율 이탈이 SMS 증가인지 통화 감소인지 확인

결론 | 새해·성탄 인사 메시지로 인한 **이벤트성 급증**입니다. SMS 설비 용량은 평일 정점이 아니라 **1월 1일 0시**를 기준으로 계획해야 한다는 판단까지 에이전트가 근거와 함께 제시했습니다.

STEP 6 쌓이는 분석 지식 · 더 깊은 인사이트

분석은 끝나도 지식은 남습니다

검증을 마친 분석 결과는 근거 조회와 함께 인사이트로 저장됩니다. 다음 질문은 이 지식에서 먼저 답을 찾고, 없을 때만 새로 분석합니다.



실제 사례 · 이 사례집을 만들며 저장하고 다시 찾은 지식

INSIGHT 재사용

① 처음 질문 → 분석 → 인사이트 #758로 저장

나 크리스마스 주간에 밀라노 구역별 인터넷 트래픽이 직전 주보다 얼마나 줄었나

② 며칠 뒤, 다른 사람이 다른 표현으로

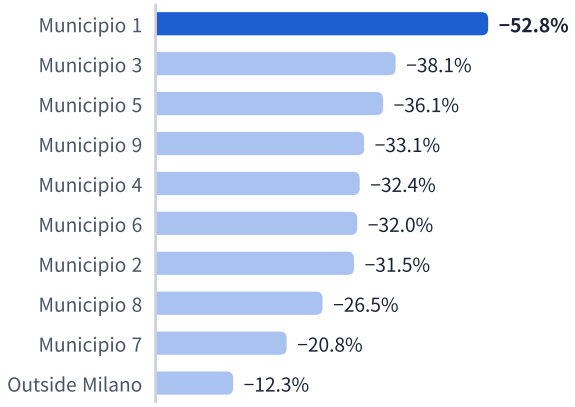
나 연말 연휴에 밀라노 도심 데이터 사용량이 얼마나 줄었어?

저장된 인사이트 #758 · 의미 유사도 **0.84** (다른 후보 0.67)

크리스마스 주간 도심 Municipio 1은 직전 주 대비 −52.8%, 외곽은 −12.3%

'크리스마스 주간'↔'연말 연휴', '인터넷 트래픽'↔'데이터 사용량'처럼 단어가 달라도 의미로 찾아, 새 조회 없이 근거와 함께 답합니다.

인사이트에 저장된 근거 · 구역별 인터넷 트래픽 증감률



인사이트 하나에 저장되는 것

대표 질문과 제목
무엇을 물었고 무엇이 결론인지

요약과 해석
수치, 비교 기준, 공휴일 통제 같은 주의점

근거 조회 정의
측정값·차원·필터·의도 (다시 실행 가능)

결과 데이터
당시 조회된 행 그대로

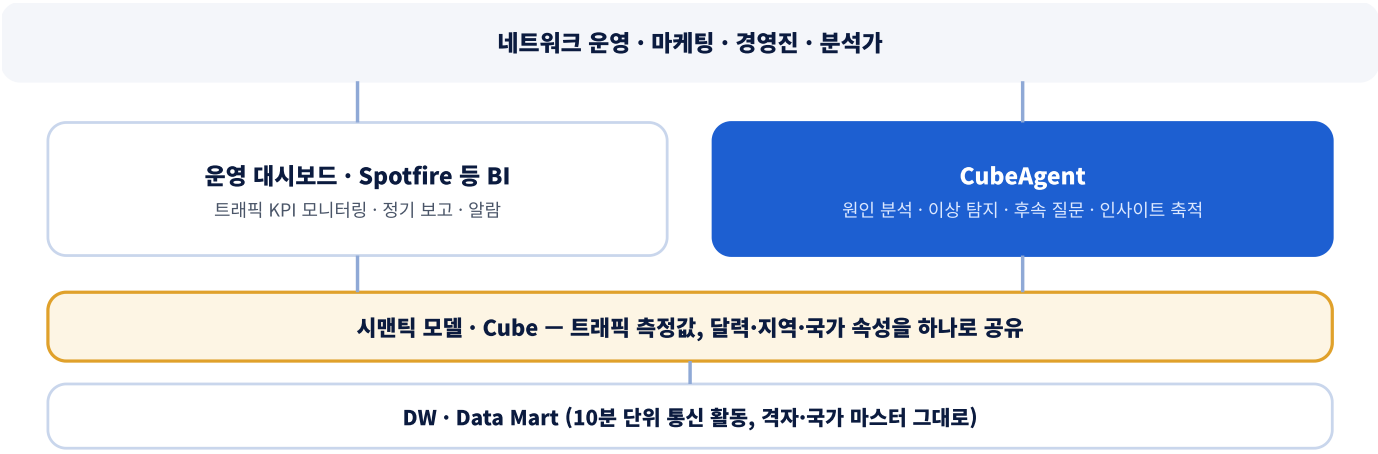
차트 정의
차트 형태, 축, 색 구분 기준

의미 벡터
표현이 달라도 찾아지는 의미 검색용

WITH BI 기존 BI와 함께 · 기존 BI 활용

운영 대시보드는 그대로, “왜?”는 CubeAgent에게

트래픽 KPI를 매일 지켜보는 대시보드는 계속 BI가 맡습니다. 대시보드에서 이상한 숫자를 본 순간 생기는 질문을 CubeAgent가 이어받습니다.



이 사례집의 질문은 어디서 시작될까요

대시보드에서 본 것 (BI)	이어지는 질문 (CubeAgent)	사례
12월 말 일별 트래픽 그래프가 꺾였다	“크리스마스 주간에 인터넷 트래픽이 왜 이렇게 줄었어?”	STEP 3
12/25 SMS 막대가 유독 높다	“두 달 동안 평소와 다른 날이 있었어?”	STEP 5
국제 통화 상위 국가 목록	“이집트 통화는 주로 언제, 어디서 해?”	STEP 2
요일별 통화량 막대 차트	“평일 중에 정말 목요일이 가장 적은 거야?”	STEP 4

지금: 질문마다 개발 요청 BEFORE

- 이상 원인: 요청 → 쿼리 작성 → 검증 → 리포트
- 후속 질문: 재요청 → 쿼리 수정 → 화면 수정
- 정기 KPI: 요청 → 쿼리 작성 → 대시보드 → 운영

도입 후: 질문의 성격대로 AFTER

- 이상 원인: 문장으로 질문 → 답변
- 후속 질문: 이어서 질문 → 답변
- 정기 KPI: 질문으로 검증 → 대시보드 → 운영

통신 분석 조직의 역할이 달라집니다

네트워크·마케팅 현업
용량 계획, 요금제 기획에 필요한 질문을 직접 던지고 후속 질문까지 이어 갑니다.

데이터 분석가
추출 작업자에서 지표 설계자·검증자로. 측정값 정의와 인사이트 품질을 관리합니다.

BI 개발
매일 볼 운영 대시보드에 집중합니다. 지표 정의는 같은 모델을 공유합니다.

도입 안내

우리 데이터로, 우리 질문으로 확인해 보십시오

3억 행이 넘는 10분 단위 데이터도, 잘 정의된 시맨틱 모델 위에서는 문장 하나로 시작합니다. 계획하고 검증하는 에이전트, 그리고 쌓이는 지식이 그 뒤를 잇습니다.

STEP 1
문장으로 추출·시각화
3억여 행 → 일별 추이, 1만 격자 지도

STEP 2
대화로 좁혀 가기
이집트 통화: 19시 정점, 북부 주거 지역

STEP 3
복합 분석
연말 −26.8%를 공휴일·구역·시간대로 분해

STEP 4
정확한 숫자
'목요일 최저'를 일평균·공휴일 통제로 교정

STEP 5
스스로 탐색
1월 1일 0시 SMS 10.8배 급증 추적

STEP 6
지식 축적
다른 표현의 질문을 유사도 0.84로 재검색

도입은 이렇게 진행됩니다

01 진단
데이터·질문 진단
트래픽·가입자·요금 데이터 구조와 부서별 대표 질문을 함께 정리합니다.

02 모델링
시맨틱 모델 구성
기존 DW·Cube를 활용해 측정값, 달력·지역·코드 속성을 AI Ready 모델로 정리합니다.

03 PoC
실제 질문으로 검증
대표 질문의 답을 기존 운영 리포트 수치와 대조해 정확도를 확인합니다.

04 확산
운영·확대
업무별 스킬과 인사이트를 축적하고 부서와 데이터 영역을 넓힙니다.

POC · DEMO 문의

쿼리 한 줄 없이, 다음 회의의 질문에 답해 보세요.

Tel 02-2088-1501, 010-9733-2366 | Fax 02-6918-0898
E-mail sspark@foresoft.co.kr | Web www.foresoft.co.kr
08501 서울시 금천구 가산디지털2로 184, 벽산디지털밸리2차 1109호

FORESOFT
데이터, 그리고 더 나은 내일