



Contents

분석 아이디어 배경

빅데이터 처리 아키텍처

활용데이터 탐색 및 분석

분석서비스 활용방안

분석서비스 기대 및 파급효과

분석 아이디어 배경

커피시장
분석

서울 음료제조
전력지도

비/브랜드 카페
폐업률 추이

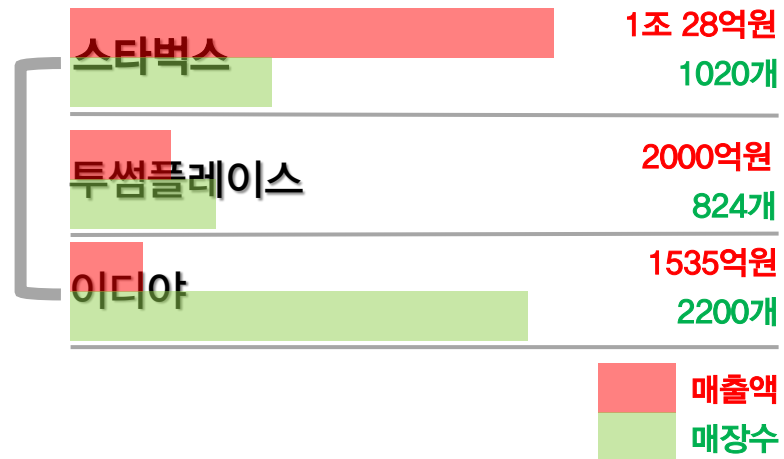


국민 1인당 커피 섭취량



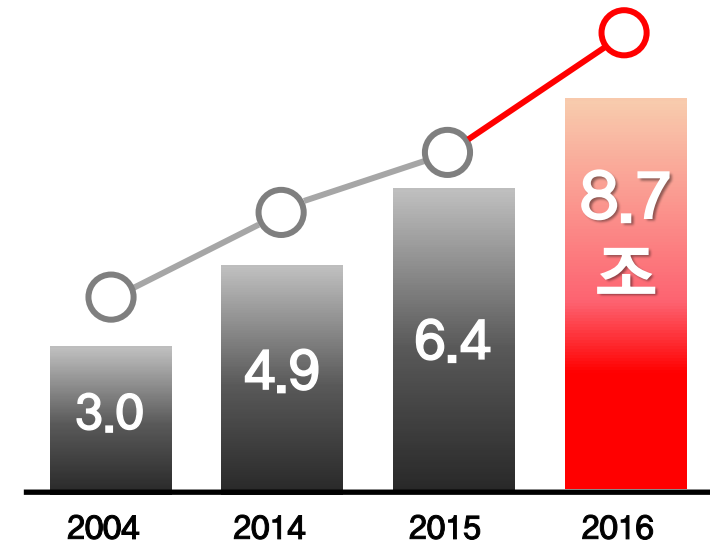
1인 커피 구매량
[큰 증가]

주요커피 전문점 현황

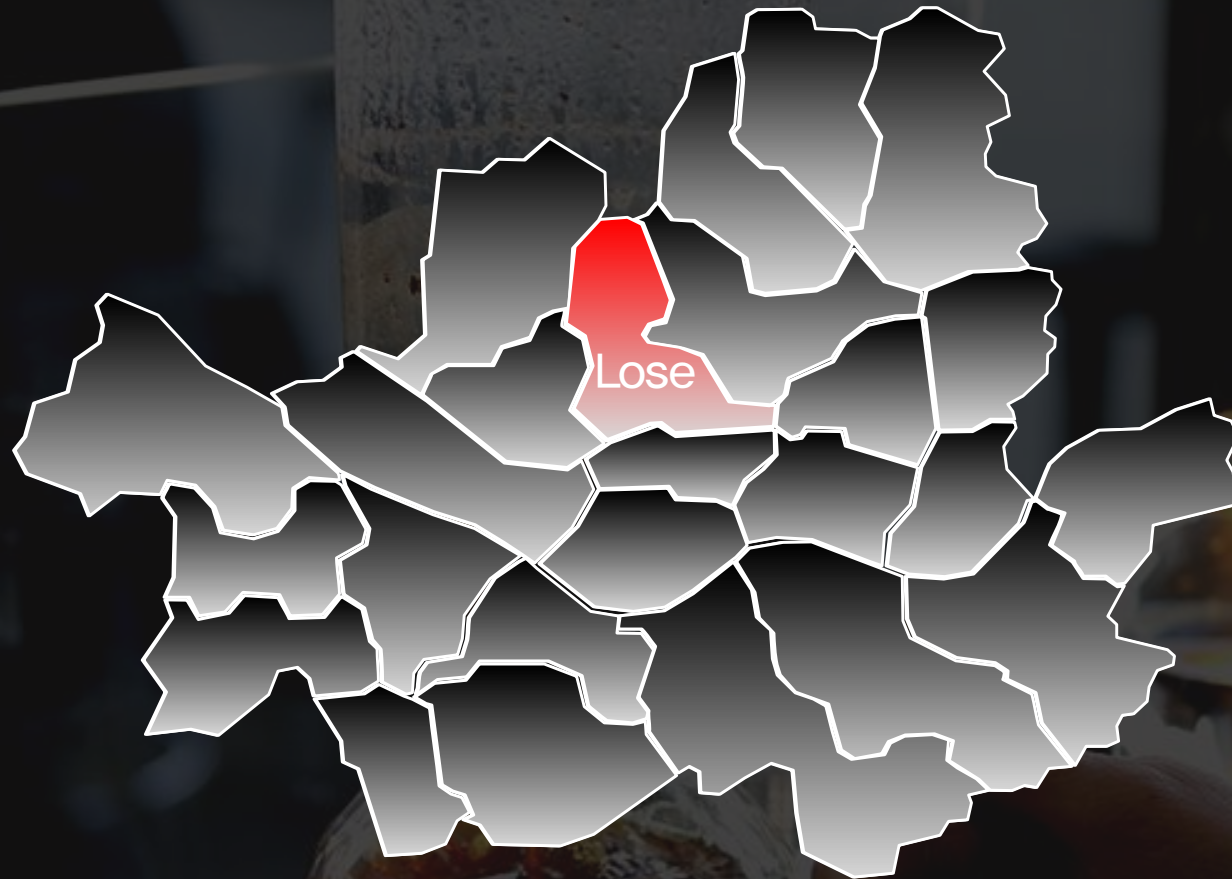


시장 점유비중
[확대]

커피 판매액 추이

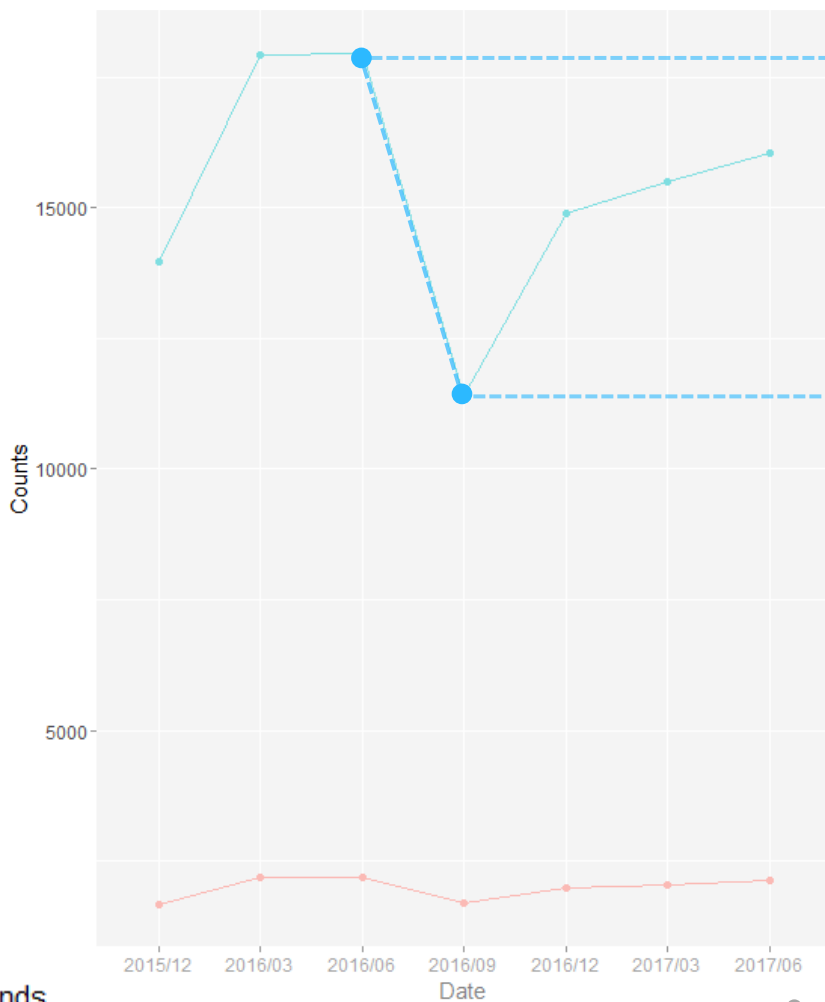


커피 판매액
[압도적 증가 추세]



1 4 , 0 8 3 , 9 2 2 kw/h

[아이디어 배경] 폐업 현황 : 비브랜드 일반 커피숍



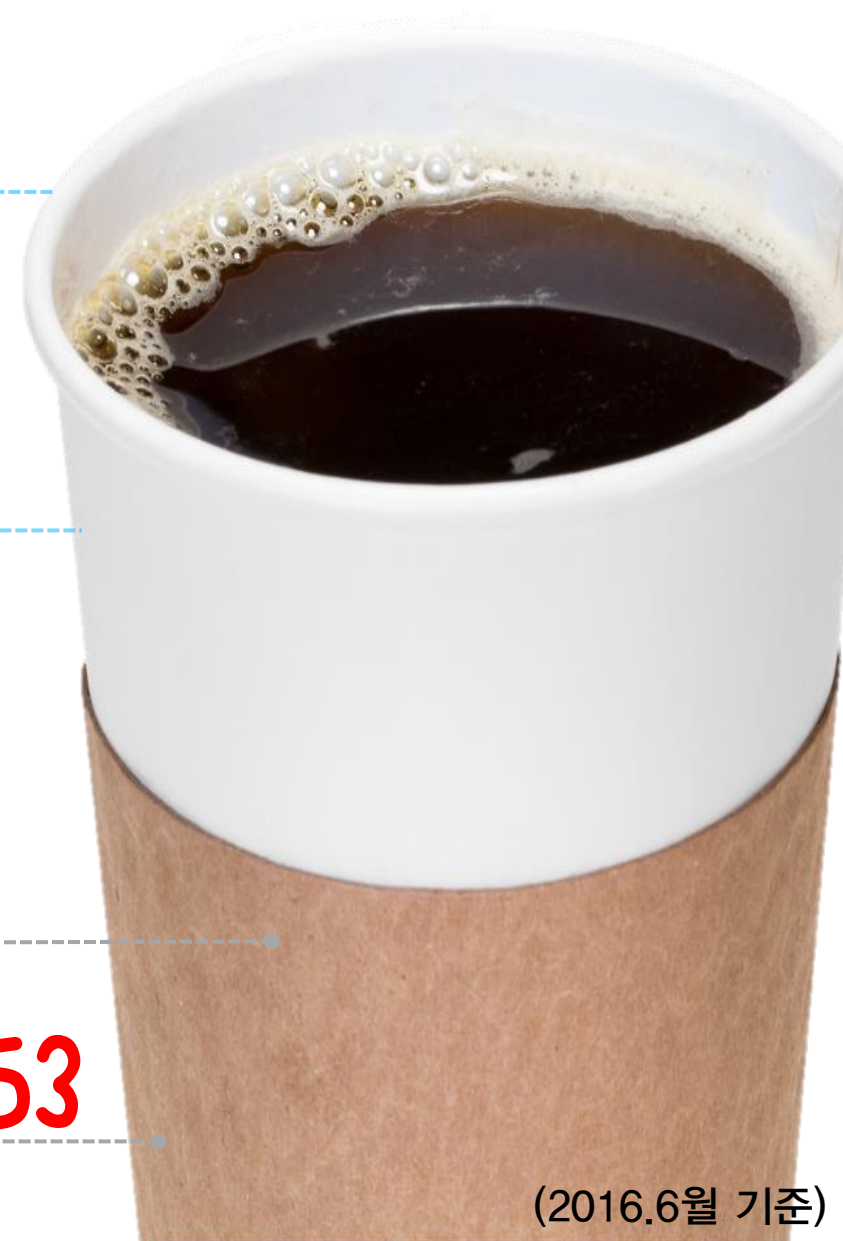
비 브랜드 3곳 중 1곳 이상 **폐업**

1분기 내 폐업을

36%

비 브랜드 1분기 내 폐업

6,553



(2016.6월 기준)



 **Angel-in-us Coffee**



 **CAFFÉ PASCUCC**



HOLLYS  **COFFEE**



DROPTOP



STARBUCKS®



A TWOSOME PLACE



A person is pouring coffee from a siphon brewer into a glass. The coffee is dark and frothy. The background is blurred, showing a kitchen or cafe setting.

Big spot Motto

[**위치** 가 곧 경쟁력]

2017.9.22

‘데이터 기반 커피숍 창업’

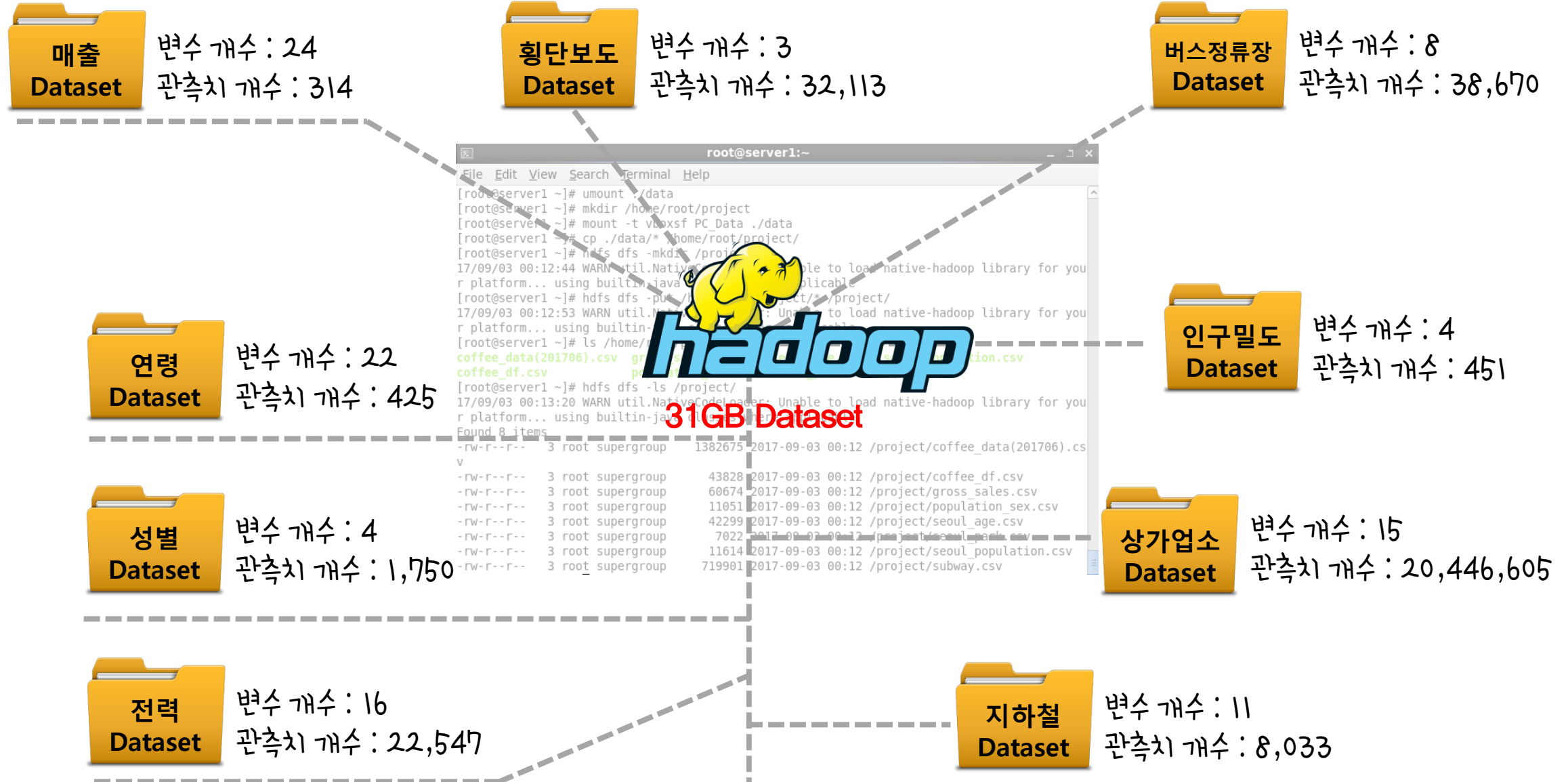
빅데이터 처리 아키텍처

빅데이터 처리
아키텍처

활용데이터
정의



[데이터처리 아키텍처] : Hadoop



[데이터처리 아키텍처] : PySpark



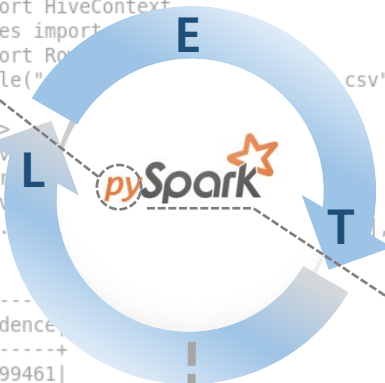

데이터 처리
Pandas Library

```
root@localhost:/home/root/hadoop-2.7.0
File Edit View Search Terminal Help

>>> from pyspark.sql import HiveContext
>>> from pyspark.sql.types import Row
>>> from pyspark.sql import Row
>>> growth_csv=sc.textFile("hdfs://.../growth.csv",use_unicode=True)
>>> type(growth_csv)
<class 'pyspark.rdd.RDD'>
>>> growth_csv=growth_csv.repartition(1)
>>> header=growth_csv.first()
>>> growth_csv=growth_csv.filter(lambda row: row[0] != header)
>>> df_growth=growth_csv.toDF()
>>> df_growth.show(5)
+-----+-----+
|Competitiveness|Residence|
+-----+-----+
|0.305198939|-0.347999461|
|0.700694597|-0.600810327|
|-0.30025905|-0.21899425|
|0.703462266|-0.786753024|
|-0.537376748|-0.568863503|
+-----+-----+
only showing top 5 rows

>>> df_growth.printSchema()
root
 |-- Competitiveness: string (nullable = true)
 |-- Residence: string (nullable = true)

>>> from pyspark.sql import HiveContext
>>> hc=HiveContext(sc)
>>> growth_csv.write.format("orc").saveAsTable("Final_df")
```




분산처리

Extraction
Transformation
Loading

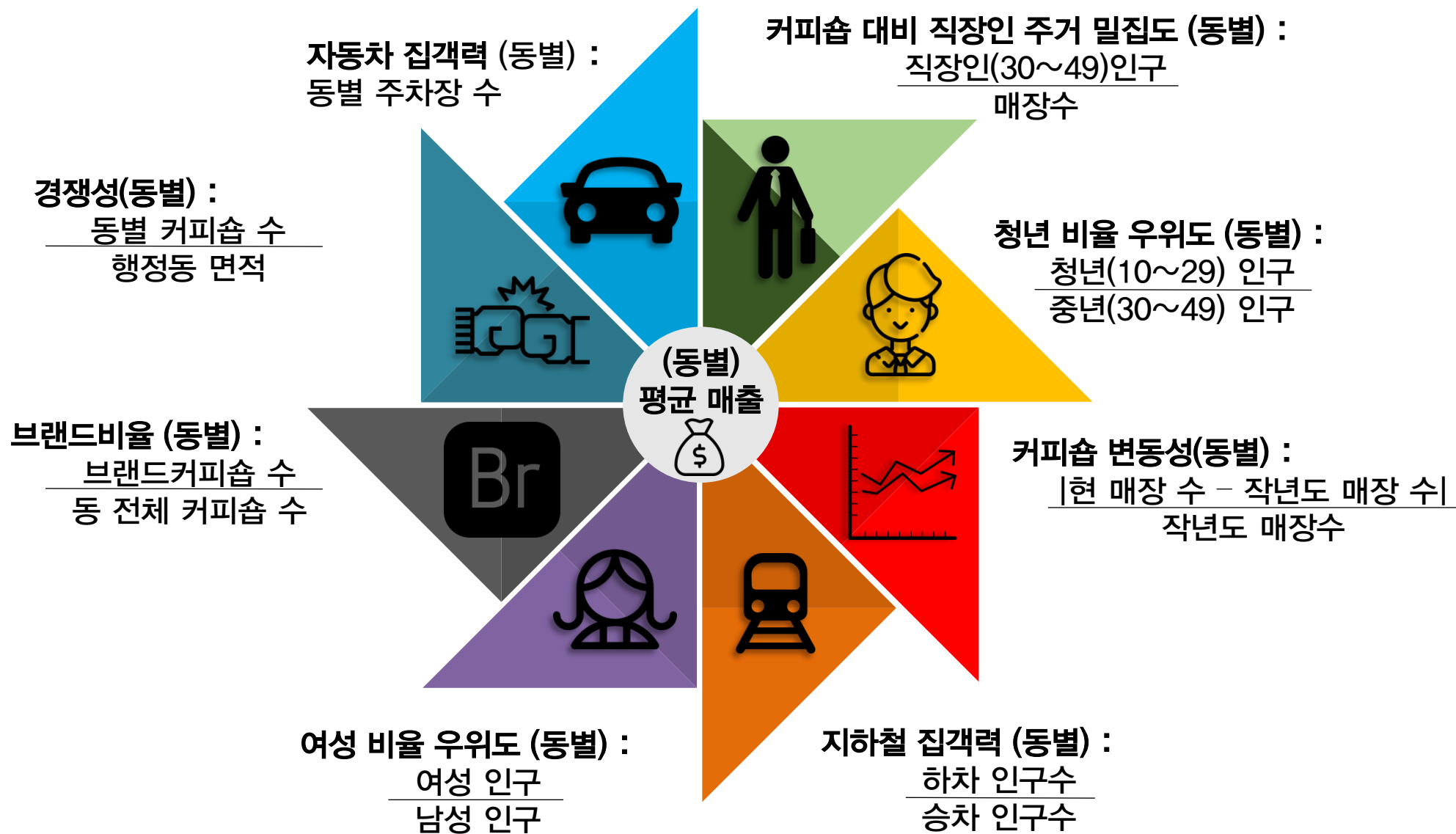
분석모형 :

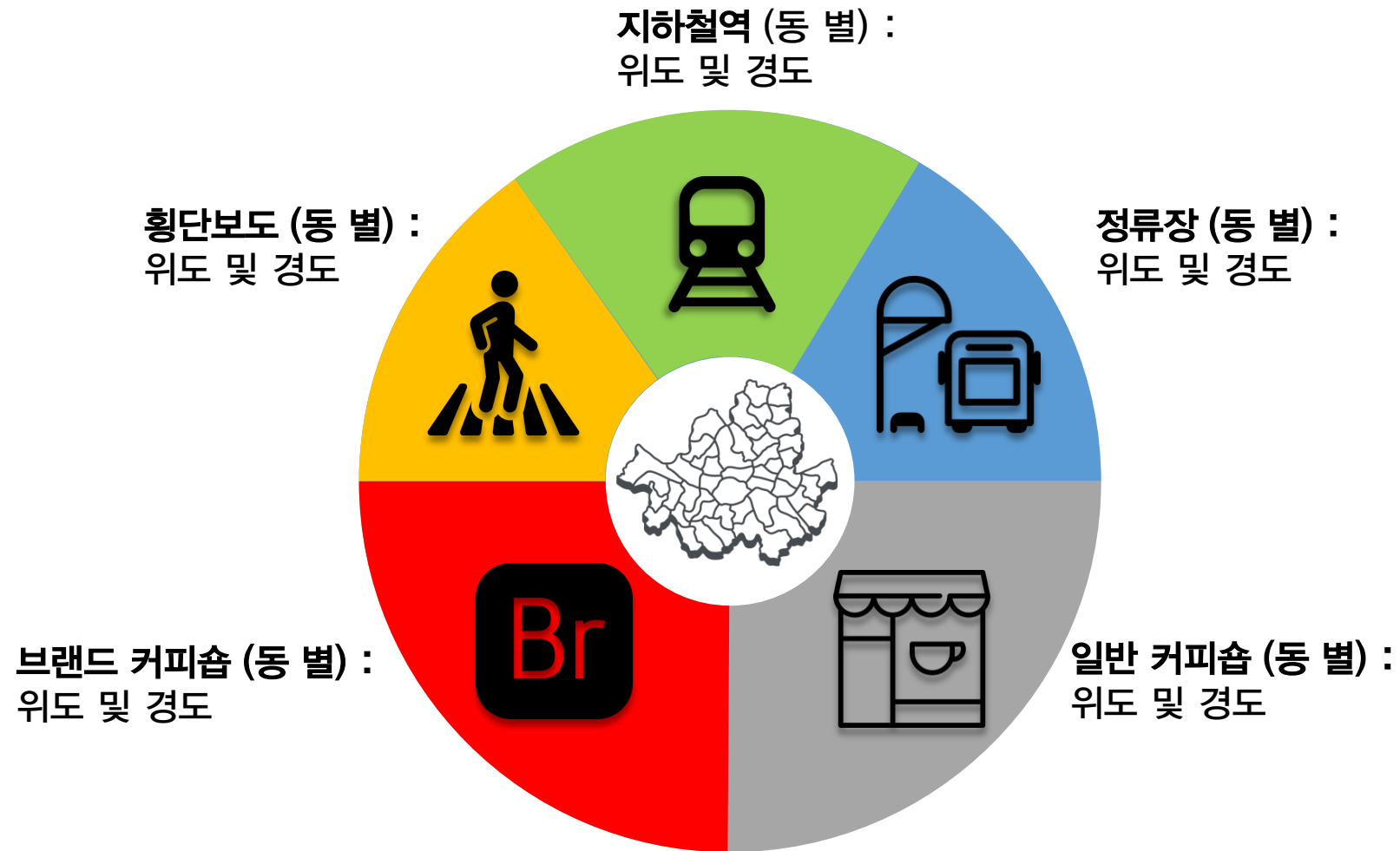
Bigspot Indicator
Deep Learning
Dataset

3GB Dataset

Geospatial
Visualisation
Dataset

: 공간정보 시각화





활용데이터 탐색 및 분석

활용데이터
탐색

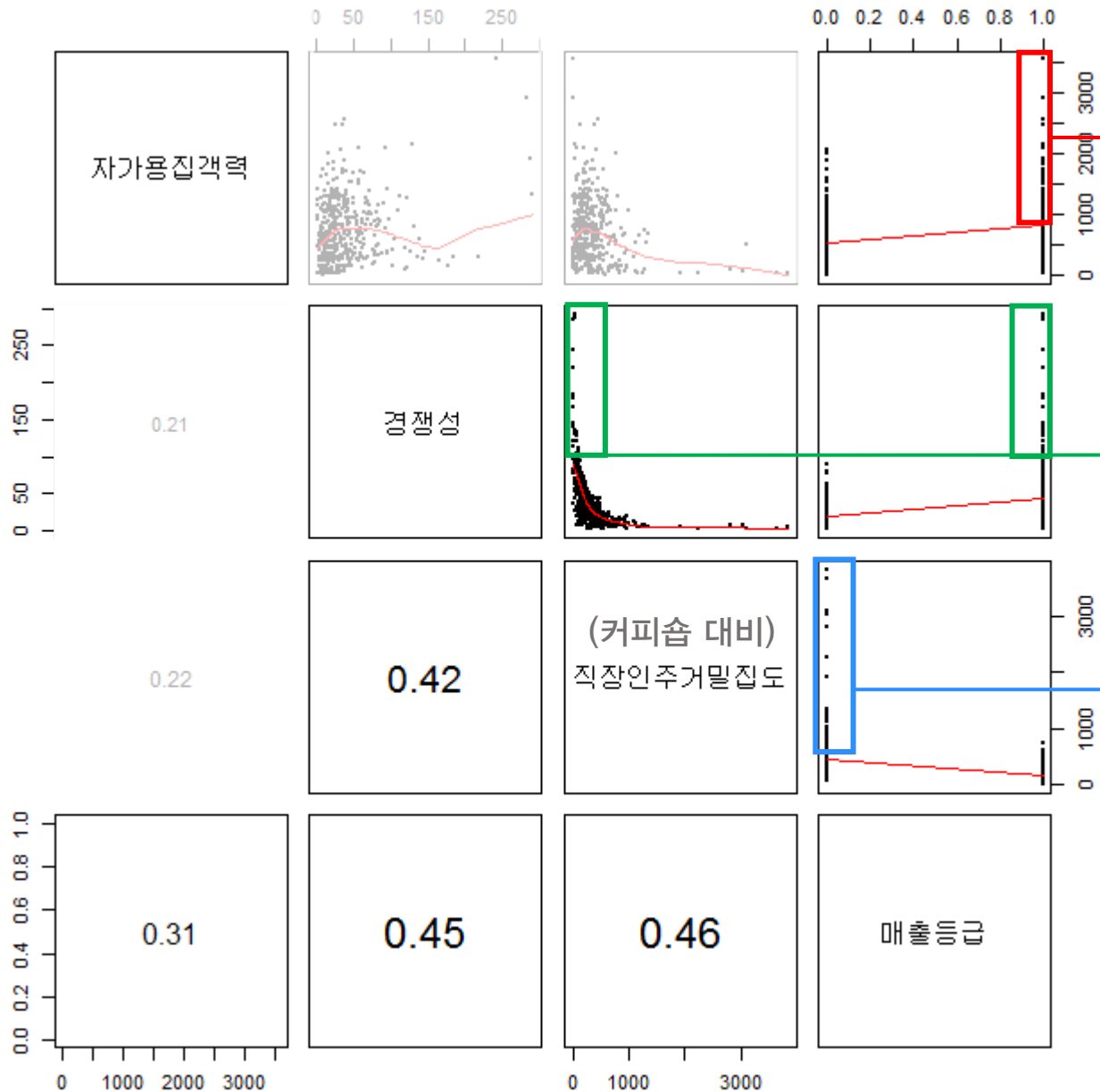
활용데이터
분석 및 결과

k-means 군집분석

로지스틱 회귀분석
: Big Spot 지표개발

딥러닝
: 상위매출 행정동 예측분류

[활용데이터 탐색] : 상권요인 및 매출등급



자가용 집객력이 높을때
주로, **평균이상**의 매출확보 행정동 분포

경쟁성 높을때, 주로 **주거밀집도 低**
평균이상의 매출확보 행정동 분포 : **상권배후지**

커피숍 매장 당 직장인 주거 밀집도가 높을 때
주로, **평균이하**의 매출확보 행정동 분포

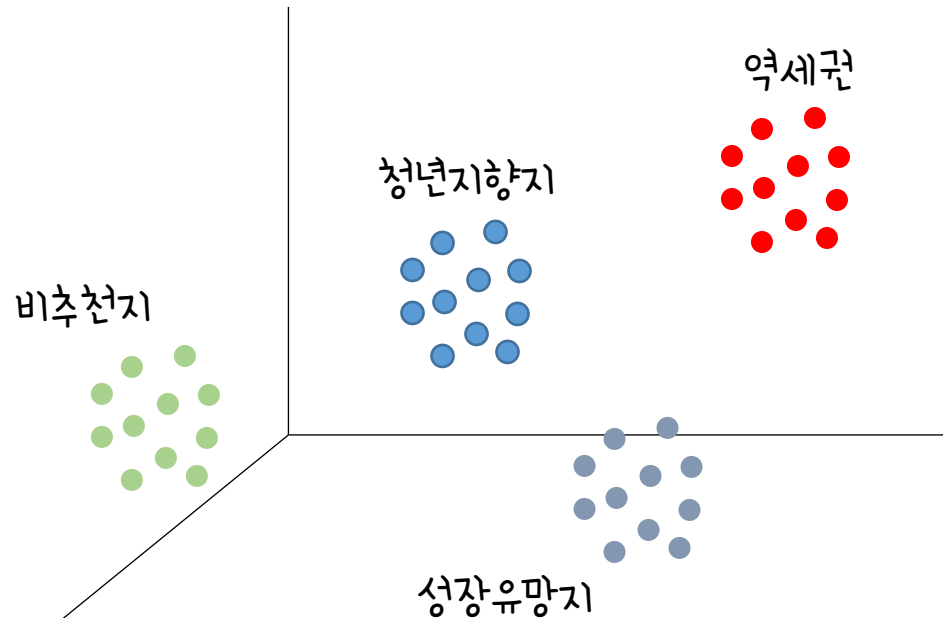
경쟁성 高, 주거밀집도 低, 자가용집객력 高
=
외부인구 흡인력이 높은 **상권배후지**



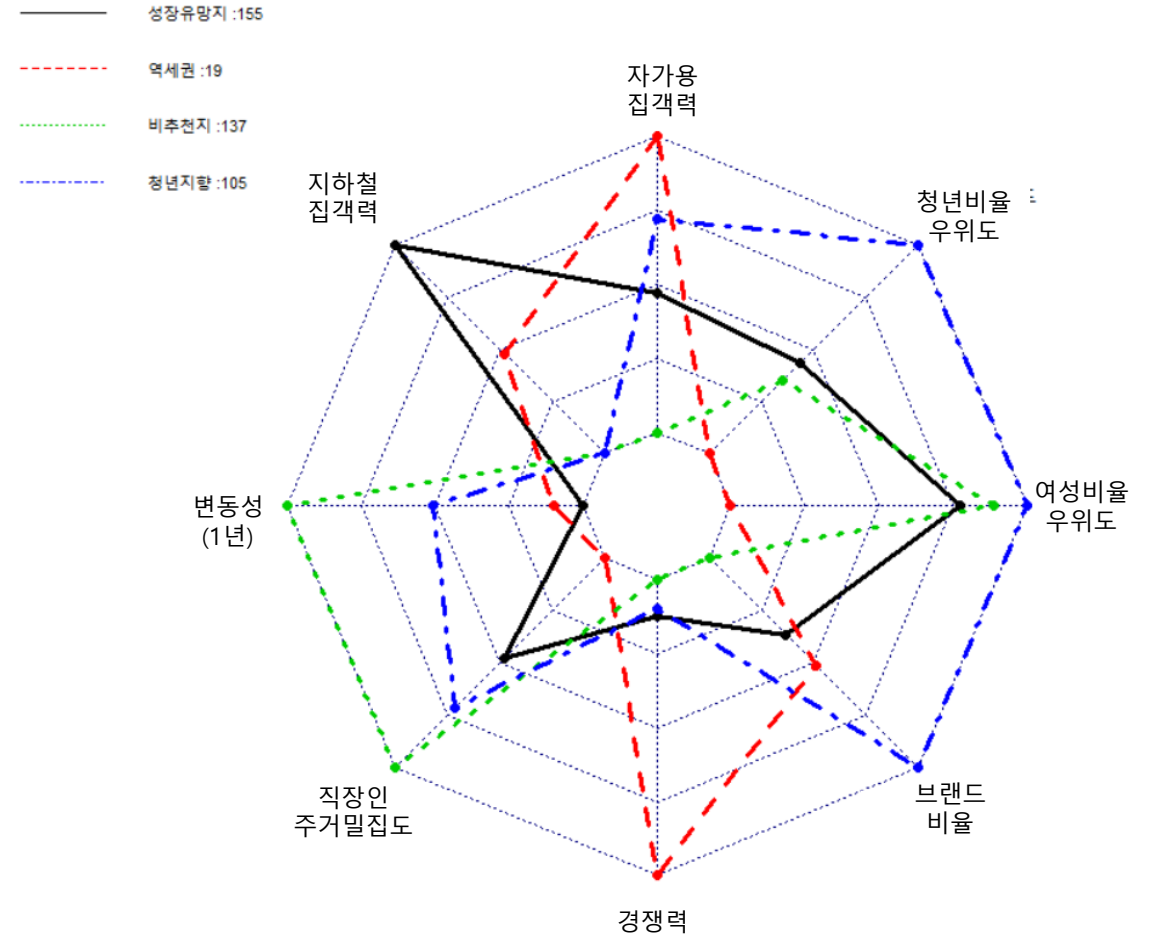
상관분류 위한 군집분석

K-means 군집분석

- : eigenvector의 고유값을 통해 3~4개의 분류기준
- : 3개보다 4개일때 군집의 특성이 더 잘 나뉘어짐



8가지 상권변수를 통한 서울특별시 416개 행정동 4상권 분류





행정구 별 Big Spot 지표 개발

Bigspot 지표(indicator)란?

: 상위매출 범주에 속할 각 지역(행정동)별 확률을 0~100의 지표로서 나타낸 것

$$\log\left(\frac{Y}{1-Y}\right) = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 = Z$$

Y : 상위 매출범주

X_1 : 직장인 주거밀집도 $\beta_1 < 0$

X_2 : 자가용 집객력 $\beta_2 > 0$

X_3 : 변동성(1년) $\beta_3 < 0$

$$Y = \frac{e^Z}{1+e^Z} : \text{Big Spot Indicator}, (0 < Y \leq 1)$$

유의한 매출요인 및 방향성 탐색

```
> summary(result)
```

Call:

```
glm(formula = grade ~ 직장인주거밀집도 + 자가용 집객력 + 변동성(1년), family = binomial(link = "logit"), data = xy_data)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.52040	-0.96306	-0.00002	0.65055	2.20766

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
직장인주거밀집도	-2.5871	0.2842	-9.102	< 2e-16 ***
자가용 집객력	0.5881	0.1421	4.138	3.51e-05 ***
변동성(1년)	-0.3591	0.1722	-2.086	0.037 *

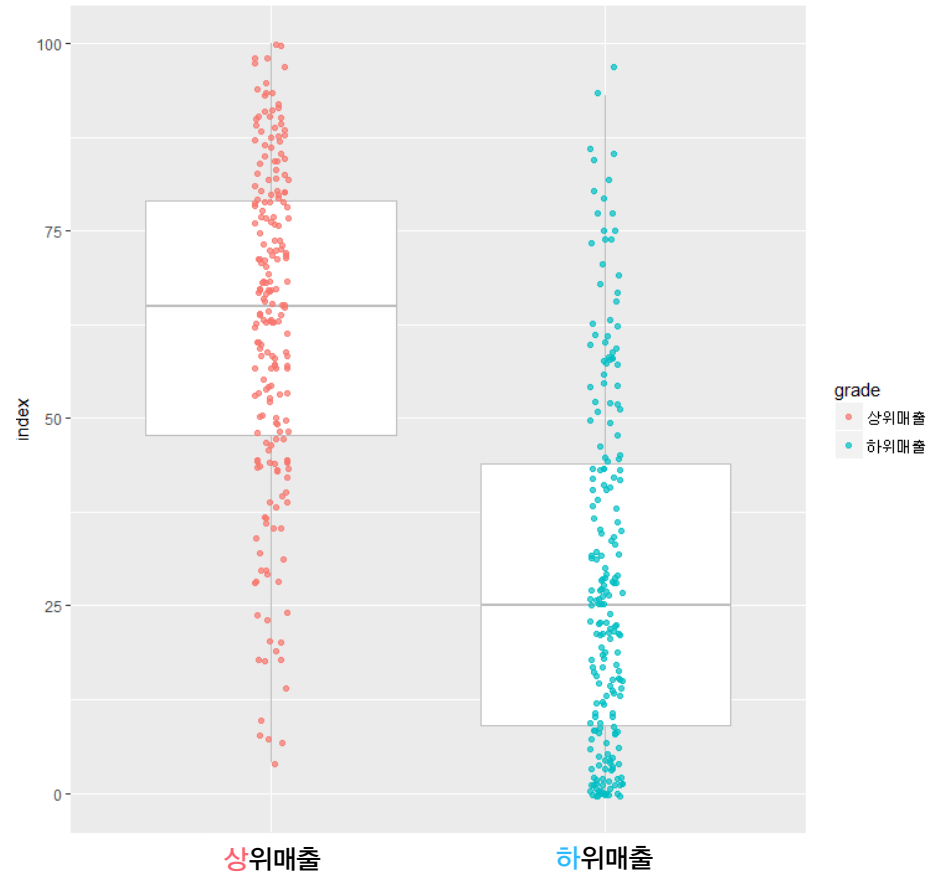
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

커피숍 대비 직장인 주거 밀집도(-), 자가용집객력(+), 커피숍 변동성(-)

매출 상위 범주에 속할 확률, 즉 BigSpot 지표 **高**

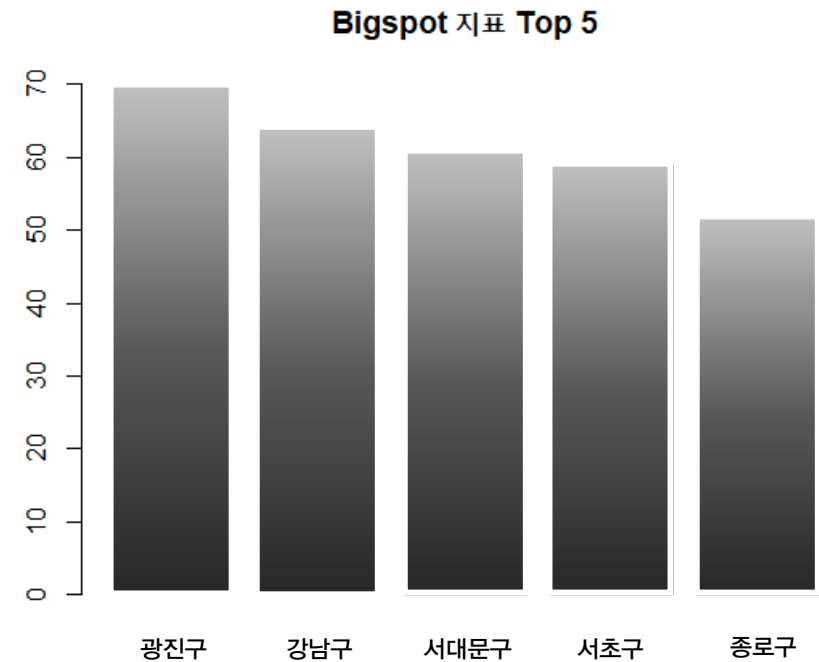


매출 범주별 Big Spot 지표 분포



Big Spot 지표의 **합리적 분포**

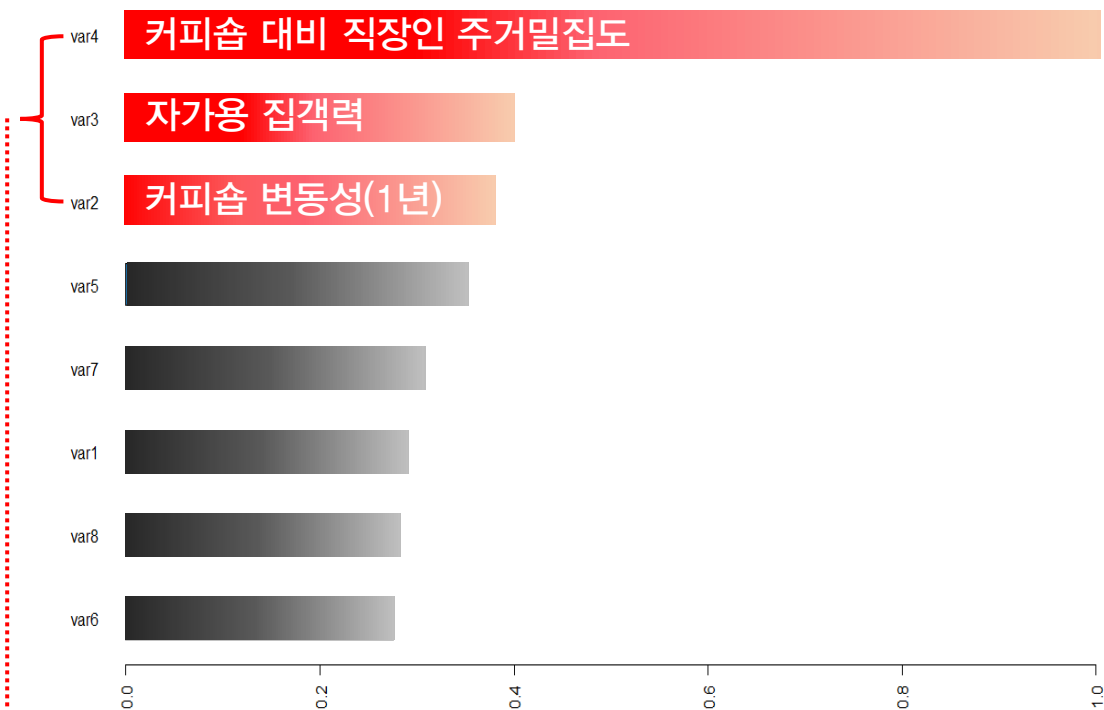
Big Spot 지표 상위 5개 행정구



서울 25개구 한눈에 **직관적 판단**



Variable Importance



커피숍 대비 직장인 주거밀집도, 자가용집객력, 커피숍 변동성(1년)
매출범주에 대해 가장 중요

(방향은 알 수 없지만, 로지스틱 회귀모형과 동일 결과로서 추정)

Confusion Matrix

	Predicted No	Predicted Yes
Actual No	67%	33%
Actual Yes	13%	87%

Totals : 77%

행정동 매출등급 예측분류



데이터 탐색 결과

상위매출 행정동



1. 커피숍 대비 직장인주거 밀집도가 낮은 **상권배후지**
2. 자가용 집객력이 높아 **외부인구 흡인력 高**

커피숍 상위매출 중요요인

by 로지스틱 & 딥러닝



데이터 분석 결과

BigSpot

자가용집객력(+), 직장인 주거 밀집도(-), 변동성(-)
행정동 BigSpot 지표 高

상권	구군시	행정동	자가용집객력	변동성_1년	직장인주거 밀집도	BigSpot
역세권	마포구	서교동	4.126723	0.5732340	-0.8344156	100
역세권	강남구	신사동	2.236774	0.5357302	-0.7831695	97
역세권	강남구	역삼1동	5.317109	-0.4393688	-0.8182314	100

'상위매출 행정동 분류'

상권	구군시	행정동	Predict	Actual
역세권	마포구	서교동	1	1
역세권	강남구	신사동	1	1
역세권	강남구	역삼1동	1	1

분석서비스 활용방안

서울특별시
상권군집 분류

행정구
BigSpot 지표

행정동
딥러닝 분석

환경분석

4P

Price
Product
Place
Promotion

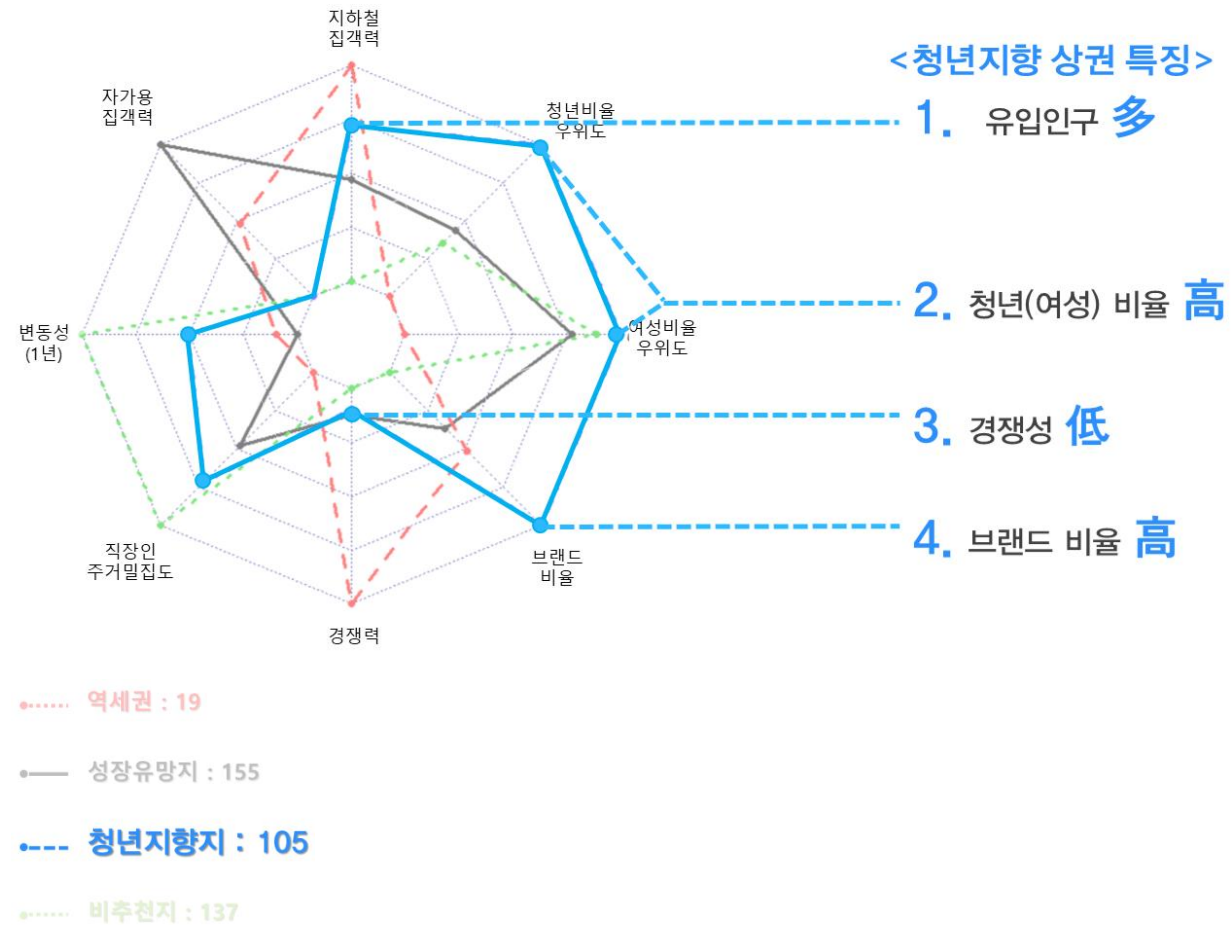
청년을 목표 고객으로
관찰은 커피숍 상권
없을까!?



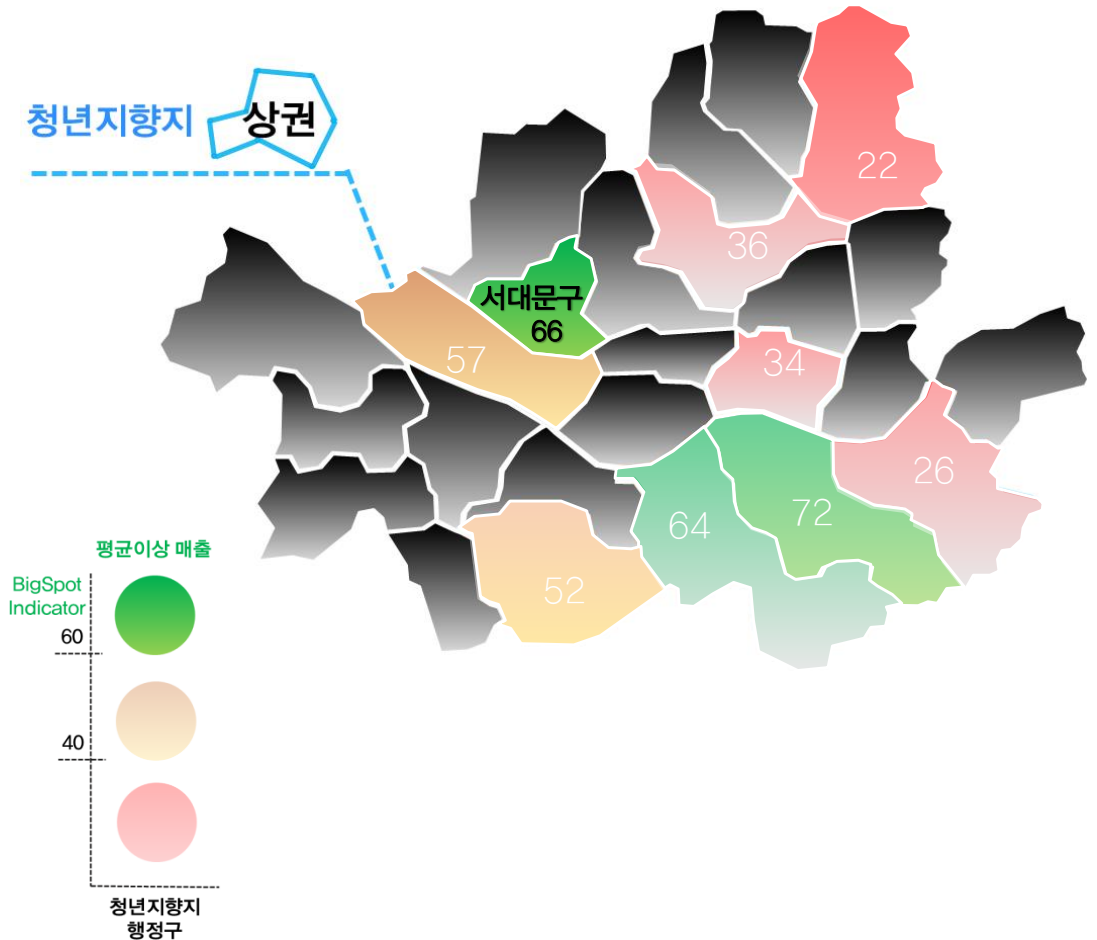
고객



청년 지향지



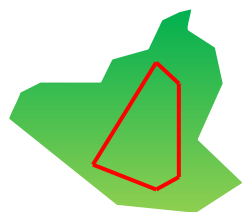
Big Spot 분석





서대문구 신촌동

Big Spot 高 행정구



서대문구

BigSpot : 66

행정동 : 신촌동

매출 : 상위

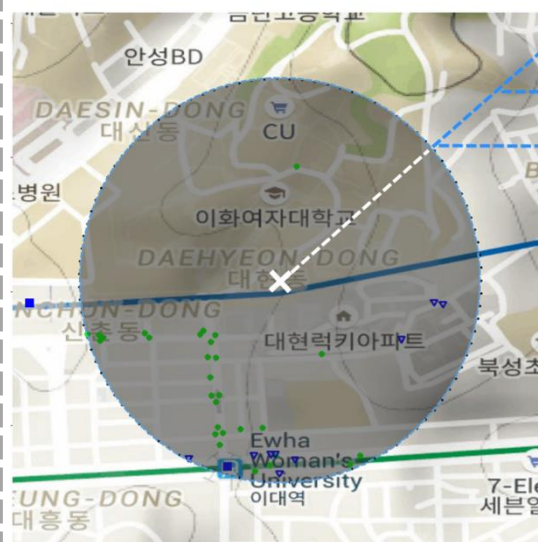
GU	Dong	Actual	Predicted
서대문구	북아현동	0	0
서대문구	신촌동	1	1
서대문구	연희동	1	1
서대문구	충현동	1	1

평균이하
매출

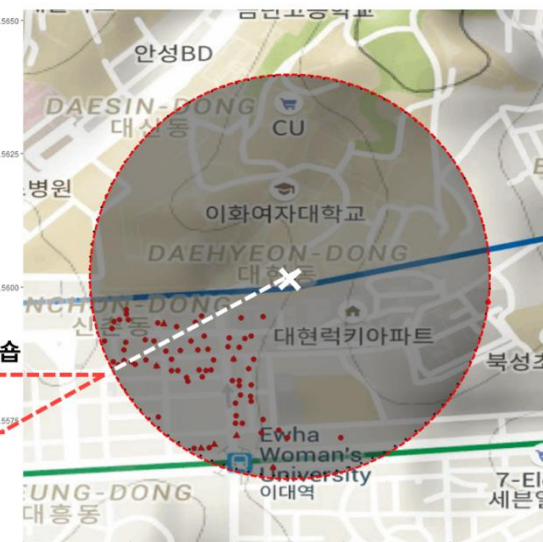
평균이상
매출 행정동

커피숍 상권 긍정 및 위협요인

긍정요인



부정요인



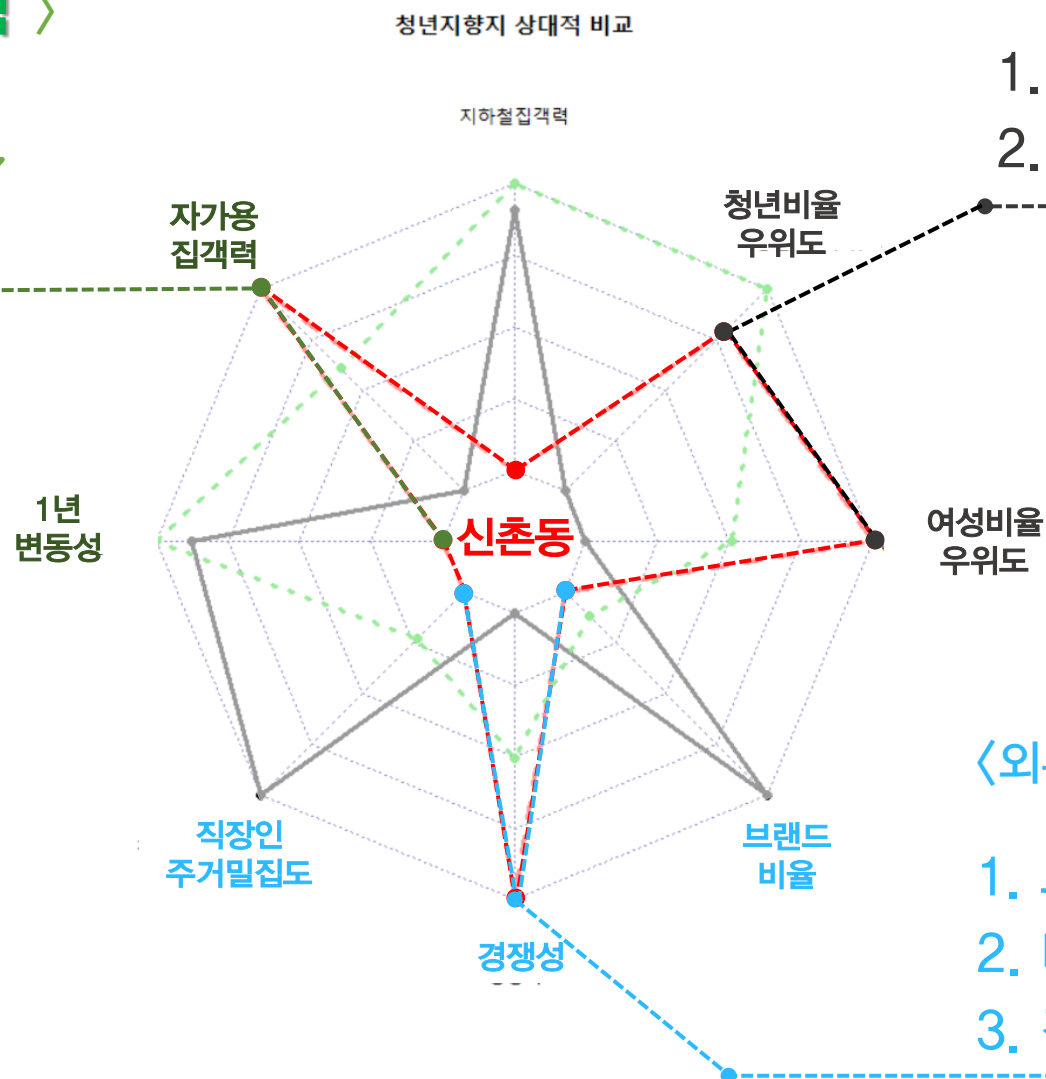


〈외부 환경분석 : 높은 집객력〉

1. 자가용 집객력(유입인구) 多
2. 1년 대비 변동성 低

〈잠재고객 : 여성청년 多〉

1. 청년비율 高
2. 여성비율 高



〈외부 환경분석 : 상권배후지〉

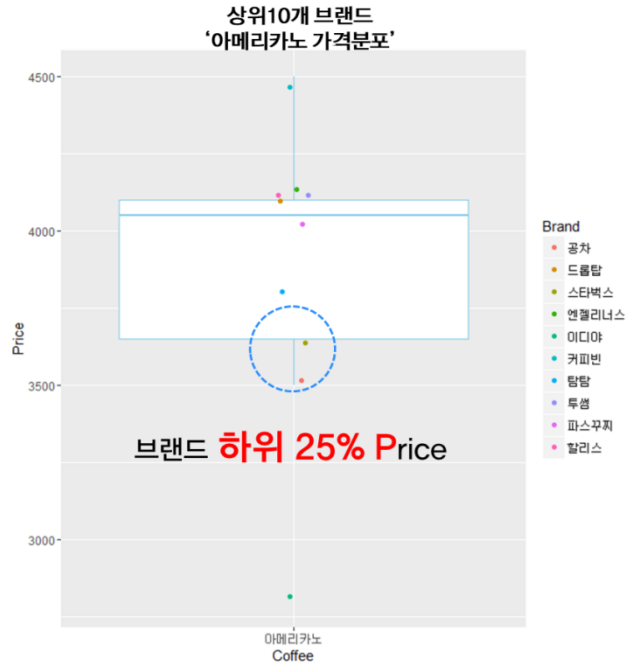
1. 브랜드 비율이 低
2. 비브랜드숍 간 경쟁성 高
3. 직장인 주거밀집도 低

[분석서비스 활용] : 4P Marketing Mix



Price

Q. “얼마의 가격을?”



[행정구]

By ‘Big Spot 지표 高’

[행정동]

By ‘딤러닝 매출상위 분류’

[동 内 세부위치]

By ‘공간정보’

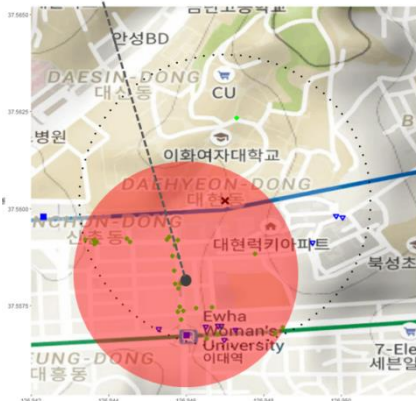
Place

Q. “어느 상권에?”

Promotion

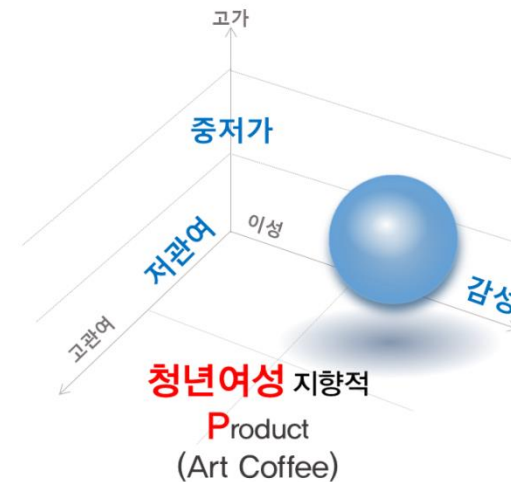
Q. “어떻게 광고?”

공정요인 다수 분포 = 유동인구 多
옥외 Promotion



Product

Q. “어떤 커피?”

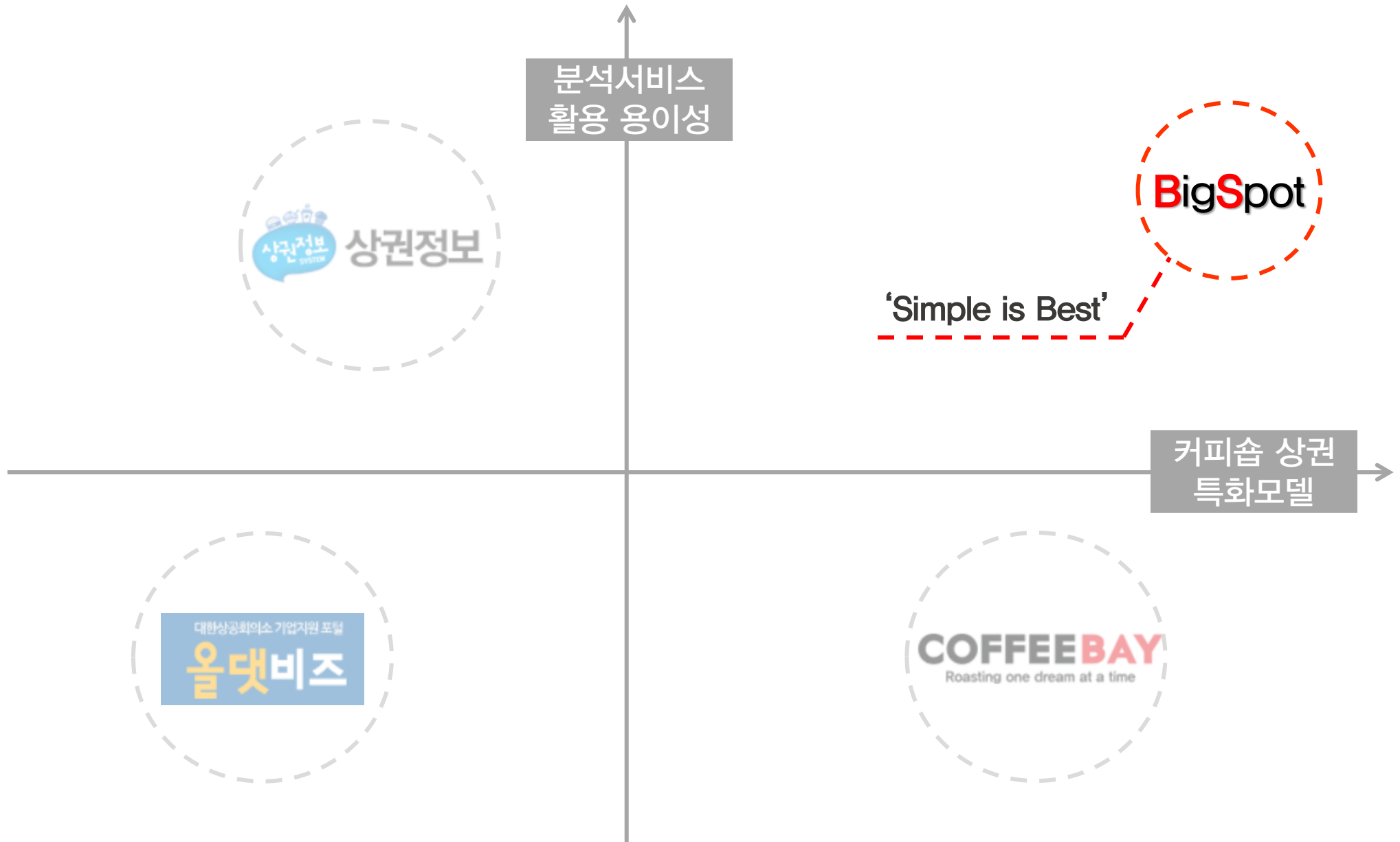


분석모델 파급효과

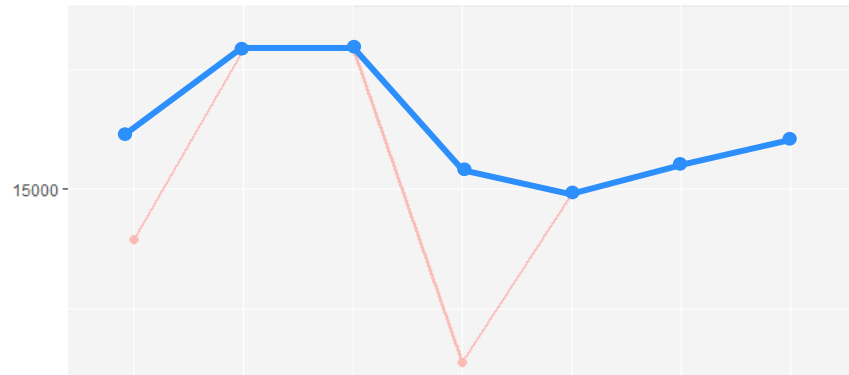
분석서비스
경쟁성

분석 서비스
파급효과

분석 서비스
web



[분석서비스 파급효과] : 사회적 비용 절감



1Q 비브랜드 폐업

16%

3,591

180,000,000,000,000

(15평,1분기)



STARBUCKS®

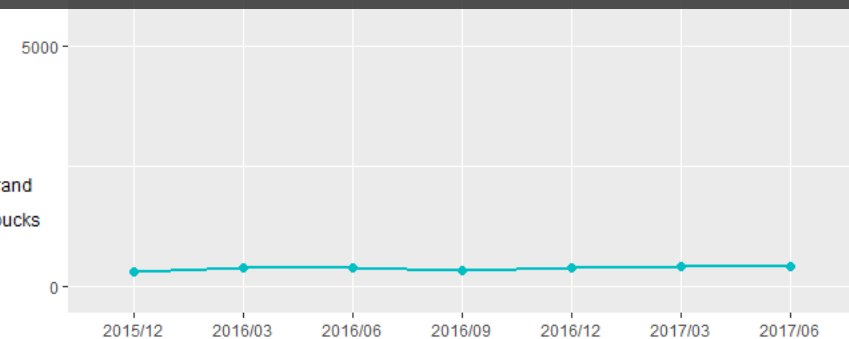
스타벅스 폐업율

14%



Name

NoBrand
Starbucks



Cafe! Here is Blue-Ocean

당신의 창업 BIGSPOT과 함께

$$\text{Business} + \text{Analysis} = \text{Big Spot}$$

【 분석서비스 흐름도 】



서울 특별시
(4 상권)

상권 프로파일링을 통한
고객 맞춤형 상권 제시

서울 행정구
(25개)

지표를 활용한 직관적 해석

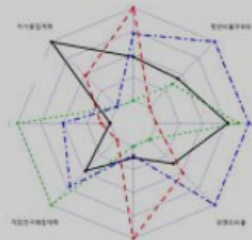
서울 행정동
(424개)

신뢰성 있는 세부분석

행정동 세부위치
(500m 이내)

필수적인 공간요인 상기

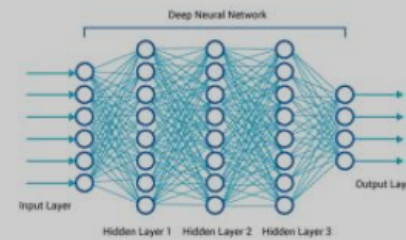
군집분석



Bigspot 지표



딥러닝



공간 시각화



