

# 카페 선정 요인 분석



데이터 마이닝 입문

14113344 이제석  
16102173 김원웅  
18101969 김상훈  
18102002 이현진

# 목차

## INDEX

1  
STEP

2  
STEP

3  
STEP

4  
STEP

5  
STEP

주제선정

데이터 수집

데이터 분석

분석 결과

분석의 한계

# STEP 1

## 주제선정



### Q 커피전문점 소비자 종합 만족도는?

단위 점 / 5점 만점 기준

평균  
만족도

2015년  
3.68

2017년  
3.74

2017년

2015년 대비

스타벅스

3.83

0.05 ▲

엔제리너스

3.73

0.15 ▲

탐앤탐스

3.73

0.10 ▲

투섬플레이스

3.73

자료 없음

할리스커피

3.72

0.02 ▲

커피빈

3.71

0.00

이디야커피

3.70

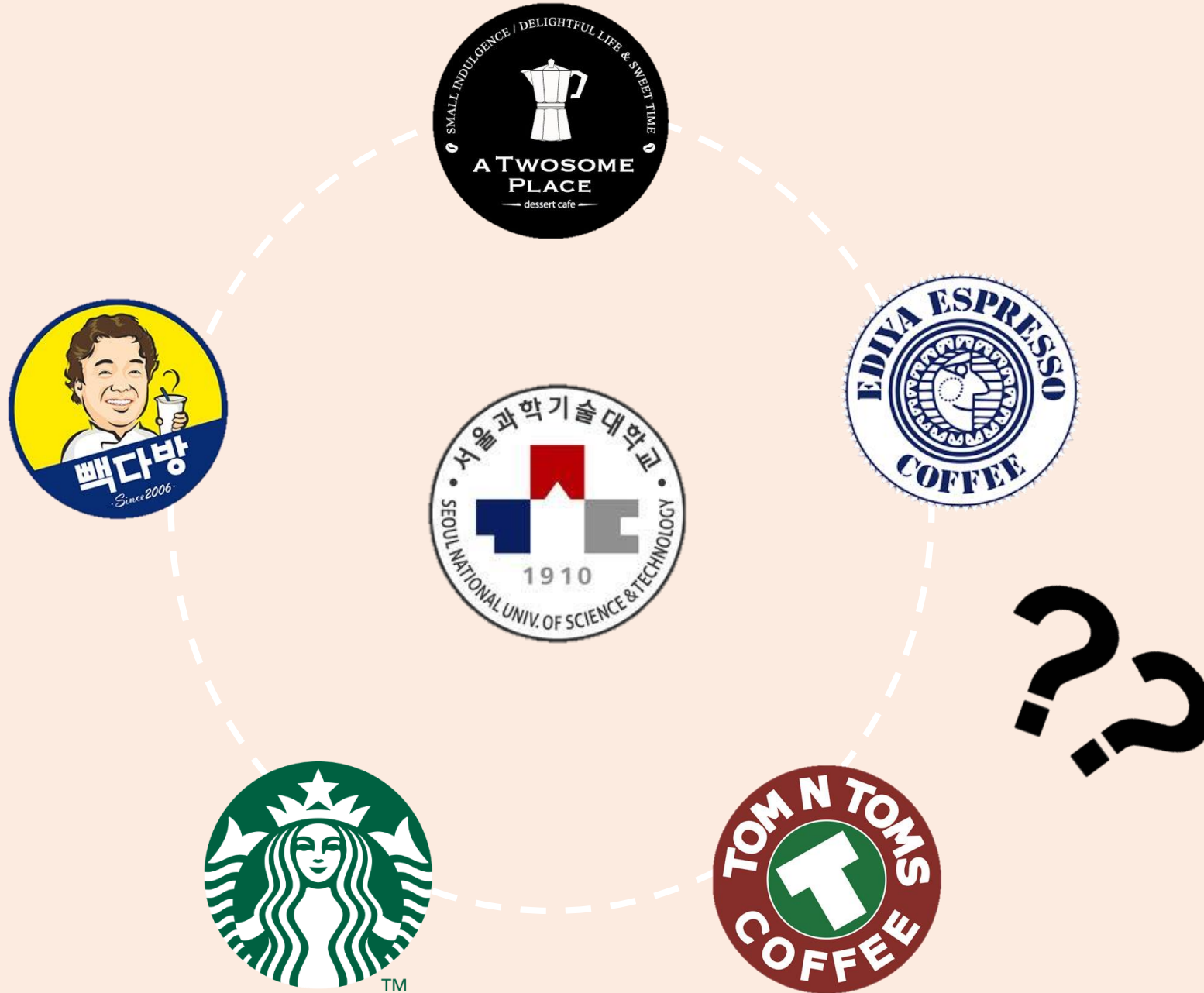
-0.02 ▼

그래픽 tong+ 자료 한국소비자원, 시장점유율 상위 7개 업체 선정 대상 커피전문점 이용자 1000명

지속적으로 성장하는 카페 시장

# STEP 1

주제선정



# STEP 1

주제선정



카페를 선정하는데 어떠한 요인이 중요할까?

## 데이터수집



엑셀 프로그램의 인터페이스를 보여줍니다. 상단에는 파일, 홈, 삽입, 레이아웃, 수식, 데이터, 검토, 보기 탭이 있습니다. 홈 탭이 선택되어 있으며, 글꼴, 배경, 정렬, 스타일 그룹이 표시되어 있습니다. 중앙에는 데이터 테이블이 표시되어 있으며, 행 번호와 열 번호가 표시되어 있습니다. 테이블의 내용은 다음과 같습니다:

	A	B	C	D	E	F	G	H	I
262	4	4	1	5	2	2	2	Starbucks	
263	5	5	1	1	3	1	5	Starbucks	
264	4	4	1	2	2	2	5	Starbucks	
265	3	3	3	3	4	1	3	Twosome	
266	3	2	2	3	5	2	4	Twosome	
267	4	3	3	2	4	1	3	Twosome	
268	3	2	2	1	5	2	4	Twosome	
269	4	3	3	2	5	1	5	Twosome	
270	4	2	4	1	3	1	1	Paikdabang	
271	3	3	5	1	2	2	2	Paikdabang	
272	2	2	4	2	3	3	1	Paikdabang	
273	2	3	5	3	2	2	2	Paikdabang	
274	2	4	4	1	3	1	3	Paikdabang	
275	2	5	5	1	2	2	4	Paikdabang	
276	2	3	4	2	1	1	1	Paikdabang	
277	5	3	5	1	2	1	1	Paikdabang	
278	2	3	4	2	1	1	2	Paikdabang	
279	2	3	3	1	2	1	3	Paikdabang	
280	5	2	3	2	1	2	1	Paikdabang	
281	2	2	5	1	1	3	2	Paikdabang	
282	4	4	3	2	1	5	3	Tomntoms	
283	3	3	2	3	2	4	2	Tomntoms	
284	4	1	3	5	3	5	3	Tomntoms	
285	3	1	2	3	1	5	2	Tomntoms	
286	3	1	3	4	2	5	4	Tomntoms	
287	3	2	2	3	3	5	3	Tomntoms	
288	3	2	1	4	1	4	4	Tomntoms	
289	3	3	2	3	2	5	3	Tomntoms	

## STEP 2

데이터수집

“

나이  
카페 사용 빈도

”

“

카페 선택 시  
중요도

”

브랜드 이미지  
커피 맛  
가격  
공간여부  
위치(거리)  
디저트  
운영시간  
인테리어

“

가장 선호하는  
카페

”

# STEP 3

## 데이터분석

```
> install.packages("caret")
'C:/Users/user/Documents/R/win-library/3.6'의 위치에 패키지(들)을 설치합니다.
(왜냐하면 'lib'가 지정되지 않았기 때문입니다)
--- 현재 세션에서 사용할 CRAN 미러를 선택해 주세요 ---
URL 'https://cran.seoul.go.kr/bin/windows/contrib/3.6/caret_6.0-84.zip'을 시도합니다
Content type 'application/zip' length 6236317 bytes (5.9 MB)
downloaded 5.9 MB

패키지 'caret'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다

다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
C:\Users\user\AppData\Local\Temp\RtmpYtzGs8\downloaded_packages
>
>
> library(caret)
필요한 패키지들 로딩중입니다: lattice
필요한 패키지들 로딩중입니다: ggplot2
Registered S3 methods overwritten by 'ggplot2':
  method      from
[.quosures    rlang
c.quosures    rlang
print.quosures rlang
>
> cafe=read.csv("c:/data_analysis/cafe.csv")
> cafe
```

254	5	5	2	1	3	2	2	Starbucks
255	4	4	3	3	2	3	4	Starbucks
256	5	5	2	4	3	3	4	Starbucks
257	4	4	1	3	2	3	4	Starbucks
258	5	5	2	2	3	3	2	Starbucks
259	4	4	1	1	2	2	4	Starbucks
260	5	5	2	2	3	1	4	Starbucks
261	4	4	1	5	2	2	2	Starbucks
262	5	5	1	1	3	1	5	Starbucks
263	4	4	1	2	2	2	5	Starbucks
264	3	3	3	3	4	1	3	Twosome
265	3	2	2	3	5	2	4	Twosome
266	4	3	3	2	4	1	3	Twosome
267	3	2	2	1	5	2	4	Twosome
268	4	3	3	2	5	1	5	Twosome
269	4	2	4	1	3	1	1	Paikdabang
270	3	3	5	1	2	2	2	Paikdabang
271	2	2	4	2	3	3	1	Paikdabang
272	2	3	5	3	2	2	2	Paikdabang
273	2	4	4	1	3	1	3	Paikdabang
274	2	5	5	1	2	2	4	Paikdabang
275	2	3	4	2	1	1	1	Paikdabang
276	5	3	5	1	2	1	1	Paikdabang
277	2	3	4	2	1	1	2	Paikdabang
278	2	3	3	1	2	1	3	Paikdabang
279	5	2	3	2	1	2	1	Paikdabang
280	2	2	5	1	1	3	2	Paikdabang
281	4	4	3	2	1	5	3	Tomntoms
282	3	3	2	3	2	4	2	Tomntoms
283	4	1	3	5	3	5	3	Tomntoms
284	3	1	2	3	1	5	2	Tomntoms
285	3	1	3	4	2	5	4	Tomntoms
286	3	2	2	3	3	5	3	Tomntoms
287	3	2	1	4	1	4	4	Tomntoms
288	3	3	2	3	2	5	3	Tomntoms



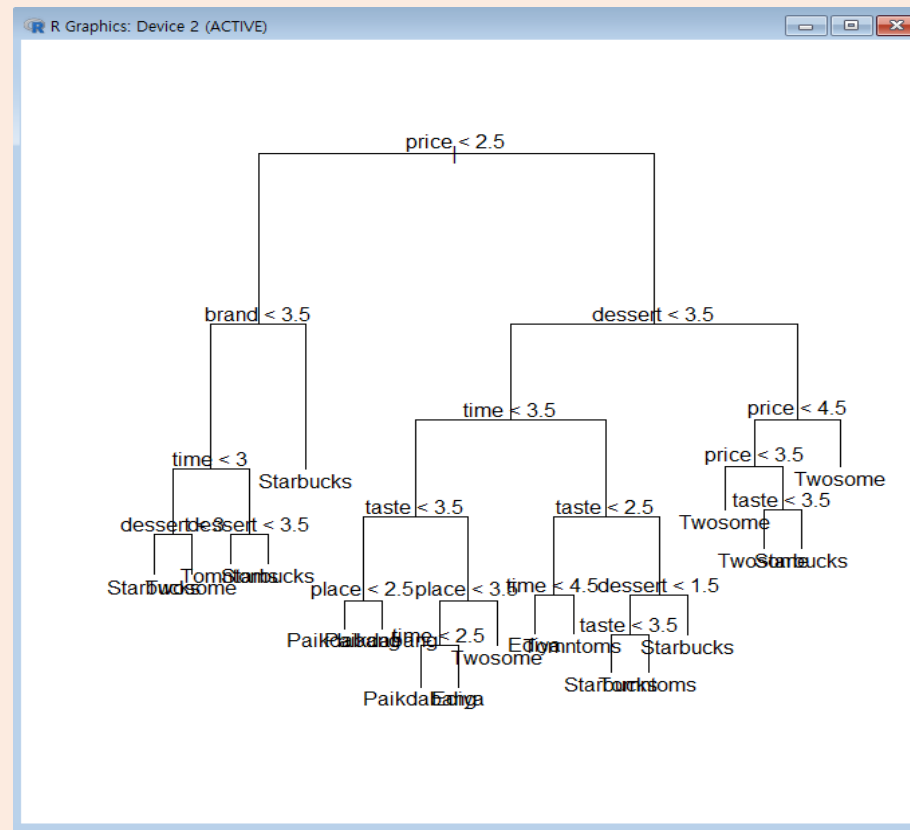
# STEP 3

## 데이터분석

```
> set.seed(1234)
> intrain=createDataPartition(y=cafe$preference,p=0.7,list=FALSE)
> train=cafe[intrain,]
> test=cafe[!intrain,]
>
> install.packages("tree")
'C:/Users/user/Documents/R/win-library/3.6'의 위치에 패키지(들)을 설치합니다.
(왜냐하면 'lib'가 지정되지 않았기 때문입니다)
URL 'https://cran.seoul.go.kr/bin/windows/contrib/3.6/tree_1.0-40.zip'
Content type 'application/zip' length 178909 bytes (174 KB)
downloaded 174 KB

패키지 'tree'를 성공적으로 압축해제하였고 MD5 sums 이 확인되었습니다

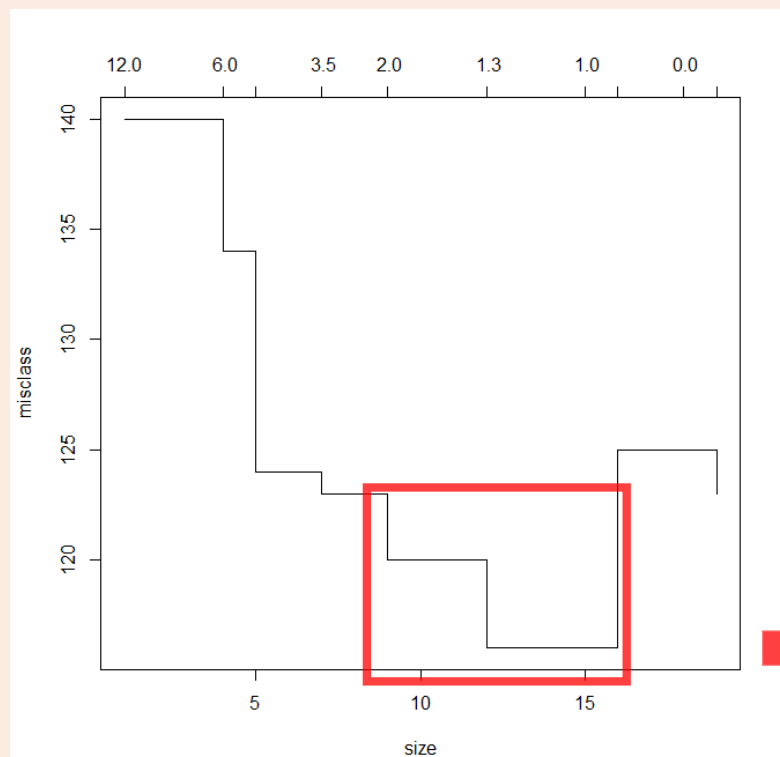
다운로드된 바이너리 패키지들은 다음의 위치에 있습니다
C:\Users\user\AppData\Local\Temp\RtmpYtzGs8\downloaded_packages
>
> library(tree)
>
> tree_model=(preference~.,data=train)
예리: 여기치 않은 ', '입니다 in "tree_model=(preference~.,"
> tree_model=tree(preference~.,data=train)
>
>
>
> plot(tree_model)
> text(tree_model)
```



## STEP 3

### 데이터분석

```
>  
> plot(tree_model)  
> text(tree_model)  
>  
>  
> cv.tree=cv.tree(tree_model,FUN=prune.misclass)  
>  
> plot(cv.tree)
```



```
> tree_predict=predict(prune.trees,test,type='class')  
>  
> confusionMatrix(tree_predict,test$preference)  
Confusion Matrix and Statistics
```

	Reference				
Prediction	Ediya	Paikdabang	Starbucks	Tomntoms	Twosome
Ediya	0	0	1	0	0
Paikdabang	1	6	3	1	2
Starbucks	8	3	13	3	4
Tomntoms	2	0	0	4	1
Twosome	1	5	11	1	14

```
Overall Statistics  
  
Accuracy : 0.4405  
95% CI : (0.3322, 0.553)  
No Information Rate : 0.3333  
P-Value [Acc > NIR] : 0.02632  
  
Kappa : 0.2493  
  
McNemar's Test P-Value : 0.05502
```

분산이 가장 낮은 값을 설정

## STEP 3

### 데이터분석

```
> prune.trees=prune.misclass(tree_model,best=9)
> plot(prune.trees)
> text(prune.trees)
> tree_predict=predict(prune.trees,test,type='class')
> confusionMatrix(tree_predict,test$preference)
Confusion Matrix and Statistics
```

		Reference				
Prediction		Ediya	Paikdabang	Starbucks	Tomntoms	Twosome
Ediya		0	0	1	0	2
Paikdabang		1	6	3	1	0
Starbucks		8	3	13	2	4
Tomntoms		2	0	0	5	1
Twosome		1	5	11	1	14

#### Overall Statistics

Accuracy : 0.4524

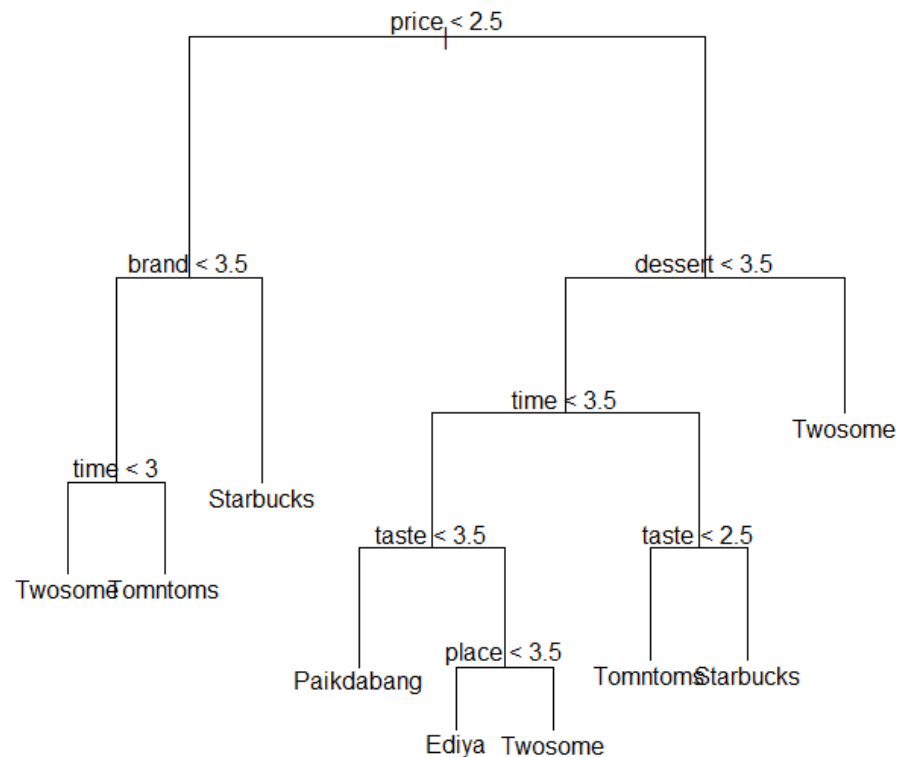
95% CI : (0.3434, 0.5648)

No Information Rate : 0.3333

P-Value [Acc > NIR] : 0.01537

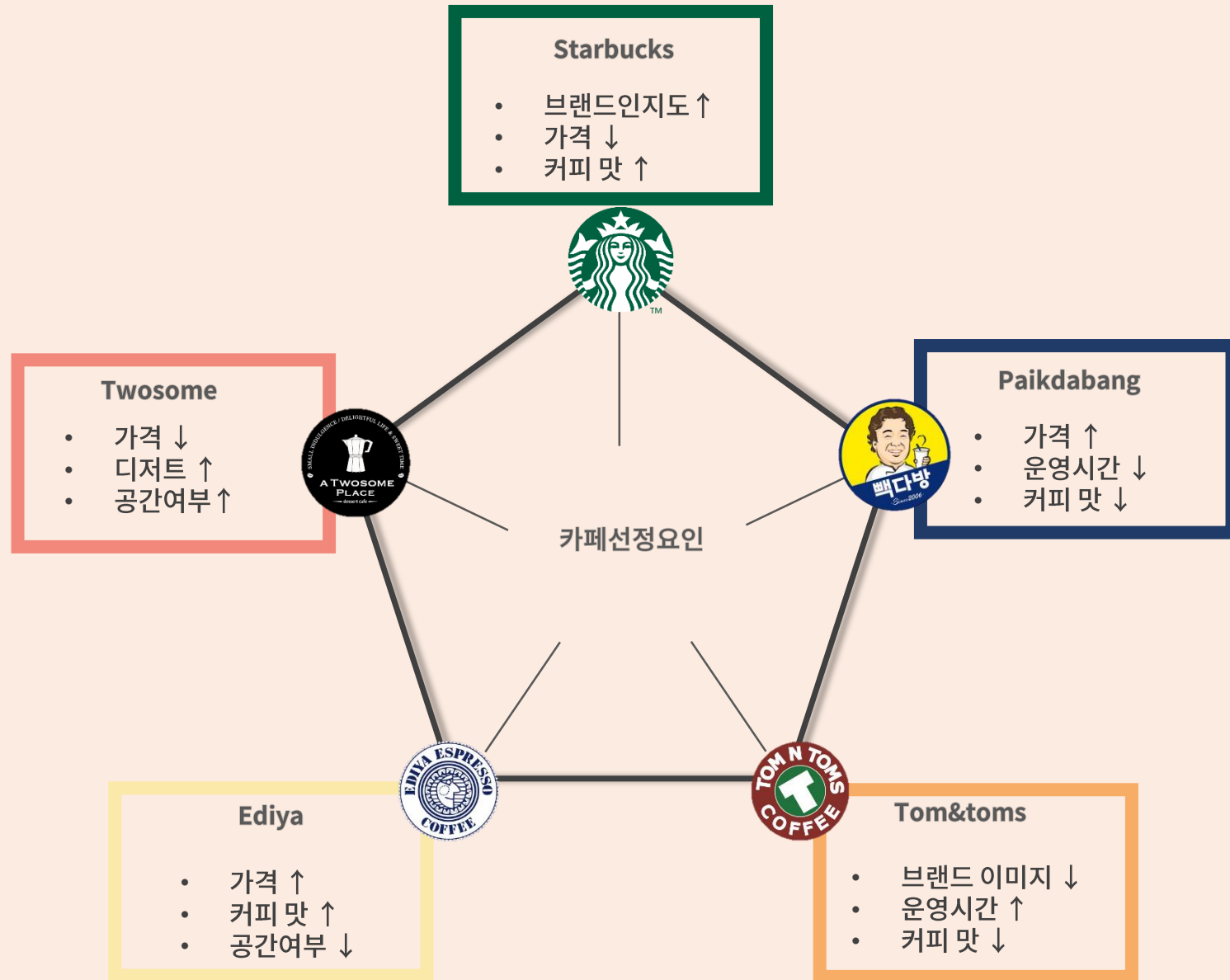
Kappa : 0.2685

Mcnemar's Test P-Value : 0.02884



## STEP 4

분석결과



## STEP 5

분석의 한계



데이터의 신뢰도



결과 해석의 부족



가지치기와 정확도



경계의 모호함

# Q&A

감사합니다