

# 대표 프로젝트 수행 결과 작성 양식

## 국내 맥주 수출을 위한 타겟 시장 및 최적합 맥주 선정

- 팀명: *Break*
- 팀원: 김선림, 김한준, 류현정, 윤선영
- 훈련과정명: 서비스 산업 데이터를 활용한 빅데이터 분석 실무
- 운영기관명: 멀티캠퍼스

# 목차

구분
I. 프로젝트 배경
II. 프로젝트 팀 구성 및 역할
III. 프로젝트 수행절차 및 방법
IV. 프로젝트 수행 결과
V. 느낀 점

# I. 프로젝트 배경

## ■ 프로젝트 주제

- 국내 맥주 수출을 위한 타겟 시장 선정 및 최적합 맥주 선정

## ■ 프로젝트 배경

- 주세법 개정 : 소규모 양조장의 외부 유통 허용, 중소맥주업체의 과세 인하, 수제 맥주 업체 수 2배 이상 증가
- 수입맥주의 저관세 : 수입 맥주 대거 등장
- 국내 맥주 시장의 경쟁 심화 → 소비자는 다양한 맥주를 선택 가능

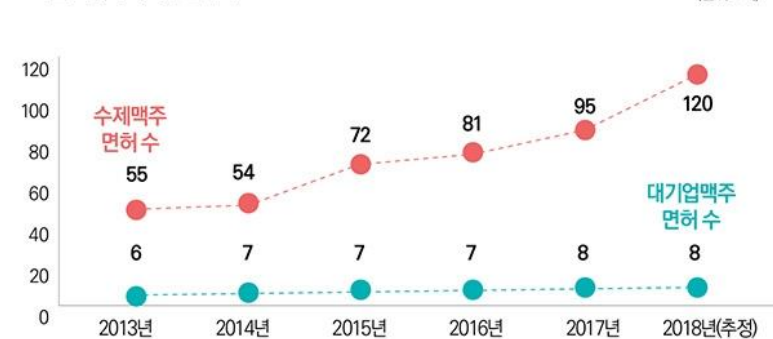
하우스 맥주(소규모 맥주제조자) 관련 주세법령 주요 변화

	2014년 3월31일 이전	현행 (2014년 4월1일 개정 시행령 발효)	홍종학 의원 개정안 발의 예정 (2014년 중)
외부유통	외부유통 금지	다른 면허자의 영업장 판매 허용 (맥주 전문점, 레스토랑 등 일반 음식점)	슈퍼, 편의점 등에 가정용 판매 허용 (종합주류도매상 통하지 않고 직접 납품)
세율인하	주세 72% (단 과세표준 계산시 제조원가의 1.1배X80%)	주세 72% (단 과세표준 계산시 제조원가의 1.1배X60%)	주세 5%

자료: 새정치민주연합 홍종학 의원실



■ 국내 수제맥주 면허 수



(자료: 한국수제맥주협회)

※ 업체수는 면허발급일 기준. 대기업 3사를 제외한 맥주 면허(국세청 국세통계 및 식약처 통계 기준 협회 제작성)

※ 대기업 3사 면허 수: 오비 3개, 하이트 3개, 롯데 2개

# I. 프로젝트 배경

- 프로젝트 목적

- 경쟁력 갖춘 국내 맥주 상품 해외 수출

- 타겟 국가 선정

- 타겟 국가에서 선호하는 맥주의 특징과 국내 맥주 제품별 특징 비교

- > 타겟 국가에 수출하기에 가장 적합한 국내 맥주 도출을 위한 분석

# I. 프로젝트 배경

## ■ 프로젝트 구조

1. 프로젝트 주제 및 방향 설정 : 업무분배 / 프로젝트방향 설정/ 일정수립

2. 데이터수집 : 프로젝트 중 추가적인 데이터 탐색상황을 고려하여

초기 데이터 수집 일정과 추가적 데이터 수집일정을 별도로 구성

3.전처리 : 얻고자 하는 변수와 초기 로우 데이터 간 양식의 차이가 존재하여 데이터 정제에 가장 큰 비중 두며

다수의 인원이 작업한 코드와 파일을 통일성 있게 정리 과정 포함

4.분석 : 통계량, 인덱스의 활용, 3D 산점도, 덴드로그램을 통한 최적의 국내 맥주 분석 및 시각화

5. 마무리 : 결과보고서 작성

# I. 프로젝트 배경

- 기대효과
  - 국내 맥주 수출 시 적절한 수출국 선정
  - 수출 타겟 국가의 선호 맥주들의 특징 확인
  - 타겟 국가의 선호 맥주 특징을 통한 수출할 국내 맥주 선정

## II. 프로젝트 팀 구성 및 역할

훈련생 명	역할
김선림 (팀 리더)	<ul style="list-style-type: none"><li>• 시장조사 및 데이터수집</li><li>• 데이터 분석 등 기타 작업 백업</li></ul>
김한준 (팀원)	<ul style="list-style-type: none"><li>• 데이터 분석 및 시각화</li><li>• 일정관리</li></ul>
류현정 (팀원)	<ul style="list-style-type: none"><li>• 데이터 수집 및 전처리</li><li>• 결과보고서 작성</li></ul>
윤선영 (팀원)	<ul style="list-style-type: none"><li>• 데이터 수집 및 전처리</li><li>• 데이터 분석</li></ul>

## Ⅲ. 프로젝트 수행절차 및 방법

구분	기간	활동	비고
사전 기획	• 8/20(목) ~ 8/21(금)	• 프로젝트 기획 및 주제 선정	• 아이디어 선정
	• 8/24(월)	• 기획안 작성/ 일정 수립/ 업무 분배	
개발	• 8/21(금) ~ 9/4(금)	• 데이터 수집 및 전처리	• 음주가능인구 추산 • 국가 선정 시, 활용
	• 9/2(수)~9/6(일)	• 데이터 분석 및 시각화	• Gplot, 3d 산점도, hclust 활용
	• 9/4(금)	• 팀 별 중간보고	• 피드백 및 일정 검토
수정/보완	• 9/7(월) ~ 9/9(수)	• 피드백 반영 및 프로젝트 마무리 • 결과 보고서 작성	• 최적화, 오류 수정
총 개발기간	• 8/20(목) ~ 9/10(목)(2주)		



# IV. 프로젝트 수행 결과

## 결과 제시 1. 데이터 수집

### ■ 데이터 수집

#### • 맥주 스타일별 데이터

- 맥주 스타일별 ABV, IBU, SRM 등 각 변수 별 범위 수치

#### • 맥주 수출 타겟 국가 선정 관련 데이터

- 2018년 국가별 맥주 소비량 데이터
- 국가별 음주 가능 연령
- 각 국가의 연령별 인구 데이터
- 국가별 주류 관세율
- 국가별 맥주 수입액

#### • 타겟 국가의 맥주 관련 데이터

- 미국 맥주 판매 순위 및 점유율
- 미국 맥주 상세 정보(ABV, IBU, SRM, 맥주 스타일 등)

#### • 국내 맥주 데이터 관련 데이터

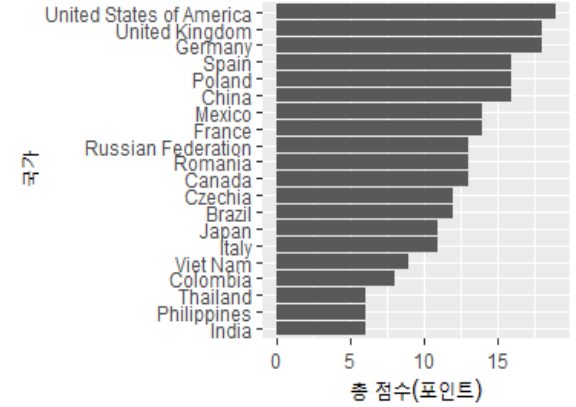
- 국내 맥주 상세 정보(ABV, IBU, SRM, 맥주 스타일 등)


#### AMERICAN-STYLE LIGHT LAGER

**Color:** Very light to pale  
**Clarity:** Chill haze should not be present  
**Perceived Malt Aroma & Flavor:** Very low  
**Perceived Hop Aroma & Flavor:** Absent to very low  
**Perceived Bitterness:** Absent to very low  
**Fermentation Characteristics:** Low fruity-estery aromas and flavors are acceptable. Diacetyl should be absent. Corn, rice, or other grain or sugar adjuncts are often used. These beers are characterized by an extremely high degree of attenuation. Final gravity is often less than 1.000(0 °Plato).  
**Body:** Low with dry mouthfeel  
**Additional Notes:** These beers are high in carbonation. Flavor attributes typical of beer are usually very low when present. Calories should not exceed 125 per 12-ounce serving. Low carb beers should have a maximum carbohydrate level of 3.0 gm per 12 oz. (356ml).  
 • Original Gravity(°Plato) 1.024-1.040(6.1-10 °Plato)  
 • Apparent Extract/Final Gravity(°Plato) 0.982-1.008 (minus 2.1-2.1 °Plato)  
 • Alcohol by Weight(Volume) 2.8%-3.5% (3.5% - 4.4%)  
 • Hop Bitterness (IBU) 4-10  
 • Color SRM (EBC) 1.5-4(3-8 EBC)

(Table 1) Global Beer Consumption by Country in 2018

2018 Rankin g	2017 Rankin g	Country	2018			2017		
			Total Consumption (thousand kl)	Global Market Share	Growth Rate 2017-2018	Total Consumption (thousand kl)	Global Market Share	Growth Rate 2016-2017
1	1	China	39,362	20.9%	-2.0%	40,166	21.4%	-2.0%
2	2	United States	24,029	12.7%	-0.4%	24,137	12.9%	-0.4%
3	3	Brazil	12,662	6.7%	0.8%	12,565	6.7%	0.8%
4	4	Mexico	8,982	4.8%	0.3%	8,532	4.6%	0.3%
5	5	Germany	8,321	4.4%	1.3%	8,218	4.4%	1.3%
6	6	Russia	7,819	4.1%	-2.4%	8,008	4.3%	-2.4%
7	7	Japan	5,108	2.7%	-0.1%	5,116	2.7%	-0.1%
8	8	United Kingdom	4,866	2.6%	5.7%	4,594	2.5%	5.7%
9	9	Vietnam	4,667	2.5%	7.1%	4,356	2.3%	7.1%
10	10	Spain	4,051	2.1%	0.0%	4,050	2.2%	0.0%
11	11	Poland	3,741	2.0%	-1.5%	3,798	2.0%	-1.5%
12	12	South America	3,447	1.8%	2.4%	3,365	1.8%	2.4%
13	13	India	2,831	1.5%	5.7%	2,678	1.4%	5.7%
14	14	Colombia	2,362	1.3%	3.6%	2,281	1.2%	3.6%
15	15	Thailand	2,248	1.2%	1.6%	2,212	1.2%	1.6%
16	16	France	2,183	1.1%	2.7%	2,106	1.1%	2.7%
17	17	Philippines	2,158	1.1%	8.8%	1,984	1.1%	8.8%
18	17	Canada	2,061	1.1%	-0.4%	2,070	1.1%	-0.4%
19	21	Czech Republic	2,033	1.1%	4.7%	1,942	1.0%	4.7%
20	20	Italy	2,032	1.1%	3.2%	1,968	1.1%	3.2%
21	18	South Korea	2,014	1.1%	0.4%	2,006	1.1%	0.4%
22	23	Romania	1,935	1.0%	3.4%	1,875	1.0%	3.4%
23	24	Australia	1,892	1.0%	1.3%	1,869	1.0%	1.3%
24	22	Argentina	1,825	1.0%	-2.7%	1,875	1.0%	-2.7%
25	25	Ukraine	1,814	1.0%	-0.4%	1,821	1.0%	-0.4%





United Nations

Population Division

Department of Economic and Social Affairs

World Population Prospects 2019

Total population (both sexes combined) by single age, region, subregion and country, annually for 1950-2100 (t)

Estimates, 1950 - 2020

POP/DB/WPP/Rev.2019/INT/F03-1

© 2019 by United Nations. Made available under a Creative Commons license CC BY 3.0 IGO. <http://creativecommons.org/licenses/by/3.0/>

of citation: United Nations, Department of Economic and Social Affairs, Population Division (2019). World Population Prospects 2019. Online Edition

Index	Variant	subregion, country	Notes	Country code	Type	Parent code	Reference date (as of 1 July)	Total population by age, both sexes combined (thousands)				
								0	1	2	3	4
1	Estimates	WORLD		900	World	0	1950	78 436	71 960	66 650	62 389	59 062
2	Estimates	WORLD		900	World	0	1951	82 782	76 585	70 734	66 874	61 914
3	Estimates	WORLD		900	World	0	1952	86 210	79 964	74 088	69 400	65 049
4	Estimates	WORLD		900	World	0	1953	89 607	82 614	77 034	72 330	68 224
5	Estimates	WORLD		900	World	0	1954	89 975	84 394	79 270	74 593	71 055
6	Estimates	WORLD		900	World	0	1955	90 381	85 052	81 174	76 958	73 012

#### • 데이터 출처

##### • 맥주 스타일별 데이터

- IBA 2019 - 2019 STYLE GUIDELINE : <https://www.iba.org.au/common/Uploaded%20files/INDIES-STYLE-GUIDELINES-2019.pdf>

##### • 2018년 국가별 맥주 소비량

- 기린 맥주 일본 웹사이트 : [https://www.kirinholdings.co.jp/english/news/2019/1224\\_01.html](https://www.kirinholdings.co.jp/english/news/2019/1224_01.html)

##### • 각 국가의 연령별 인구 데이터

- UN 웹사이트 : <https://population.un.org/wpp/>

##### • 국가별 맥주 수입액

- World's Top Exports : <http://www.worldstopexports.com/beer-imports-by-country/>

##### • 국가별 주류 관세율 데이터

- 통합무역정보 서비스 : <http://m.tradenavi.or.kr/CmsWeb/viewPage.req?idx=PG0000000596>

##### • 미국 맥주 판매 순위 및 점유율

- <https://www.usatoday.com/story/money/2020/03/18/the-most-popular-beer-brands-in-america/111416118/>

##### • 미국 맥주 상세 정보 / 국내 맥주 상세 정보

- 제조사 홈페이지 및 검색, 맥주 스타일별 데이터 참고

# IV. 프로젝트 수행 결과

## 결과 제시 2. 데이터 전처리

### ■ 데이터 전처리

#### • 맥주 스타일별 데이터 데이터셋

- 맥주 스타일별 ABV, IBU, SRM 값은 각 범위의 중간값

#### • 각 국가별 음주 가능 인구 수

- 년도 : 2018
- 국가 : 2018년 국가별 맥주 소비량 데이터 기준으로 1~20위의 국가
- 나이 : 음주 가능 나이 ~ 100세

#### • 각 국가별 1인당 맥주 소비량

- 각 국가별 1인당 맥주 소비량 = 2018년 국가별 맥주 소비량 / 각 국가별 음주 가능 인구수

#### • 타겟 국가 선정

- 변수 : 국가별 맥주 총 소비량, 국가별 1인당 맥주 소비량, 주류 관세율, 맥주 수입액

#### • 미국 맥주 상세 정보

- N/A: 맥주 스타일별 데이터 데이터셋의 값 해당 맥주 스타일의 ABV, IBU, SRM 값은 각 범위의 중간값

#### • 국내 맥주 상세 정보

- 제조사 통해 변수 별 범위 확인 : 변수별 범위의 중간값
- N/A : 맥주 스타일별 데이터 데이터셋의 값 해당 맥주 스타일의 ABV, IBU, SRM 값은 각 범위의 중간값

나라	실질 음주 가능 나이	음주가 능나이 인구	맥주소 비량	1인당 맥주 소비 량
1 United States of America	21	240272301	2.4029e+10	100.00737
2 Mexico	18	85939544	8.9820e+09	104.51533
3 Germany	18	69350757	8.3210e+09	119.98427
4 Spain	18	38531315	4.0510e+09	105.13526
5 Poland	18	31176800	3.7410e+09	119.99307
6 Czechia	18	8722354	2.0330e+09	233.07928
7 Romania	18	15880539	1.9380e+09	122.03616
8 China	18	1122854336	3.9362e+10	35.05530
9 Brazil	18	154877611	1.2662e+10	81.75488
10 Russian Federation	18	115511092	7.8190e+09	67.69047
11 Japan	20	105240440	5.1080e+09	48.53648
12 United Kingdom	16	54559727	4.8560e+09	89.00338
13 Viet Nam	18	69529374	4.6670e+09	67.12271
14 India	21	836786651	2.8310e+09	3.38318
15 Colombia	18	35628888	3.3620e+09	66.29452
16 Thailand	20	53001474	2.2480e+09	42.41391
17 France	18	50985807	2.1630e+09	42.42357
18 Philippines	18	67375735	2.1580e+09	32.02934
19 Italy	16	51974609	2.0320e+09	39.09601
20 Canada	19	29593508	2.0610e+09	69.64365

Style	OG_Median	FG_Median	ABV_Median	IBU_Median	SRM_Median
1 American-Style Amber Light Lager	1.0320	1.005	3.95	11.5	8.00
2 American-Style Ice Lager	1.0500	1.010	5.55	13.5	5.00
3 American-Style India Pale Lager	1.0575	1.012	6.30	50.0	4.50
4 American-Style Lager	1.0440	1.010	4.60	10.0	3.00
5 American-Style Light Lager	1.0320	1.000	3.95	7.0	2.75
6 Australasian, Latin American or Tropical-Style Light Lager	1.0380	1.008	4.60	13.5	3.50
7 Classic Irish-Style Dry Stout	1.0430	1.010	4.70	35.0	N/A
8 Golden or Blonde Ale	1.0495	1.012	4.60	20.0	5.00
9 International-Style Pilsener	1.0470	1.009	4.95	23.5	3.50
10 Belgian-Style Witbier	1.0470	1.008	5.20	13.5	3.00
11 American-Style Malt Liquor	1.0550	1.007	5.50	17.5	4.00
12 Vienna-Style Lager	1.0510	1.015	4.05	25.0	14.00
13 American-Style Amber/Red Ale	1.0530	1.014	4.15	35.0	14.50
14 German-Style Pilsner	1.0495	1.009	3.90	32.5	3.50
15 Session India Pale Ale	1.0450	1.011	3.50	47.5	8.00
16 American-Style India Pale Ale	1.0650	1.013	5.50	60.0	9.00
17 American-Style Stout	1.0625	1.016	5.45	47.5	N/A
18 American-Style Imperial Stout	1.0900	1.025	7.50	65.0	N/A
19 American-Style Dark Lager	1.0510	1.009	3.80	19.0	19.50
20 American Wheat Beer with Yeast	1.0460	1.011	3.60	22.5	7.00

## IV. 프로젝트 수행 결과

### 결과 제시 3. 탐색적 데이터 분석

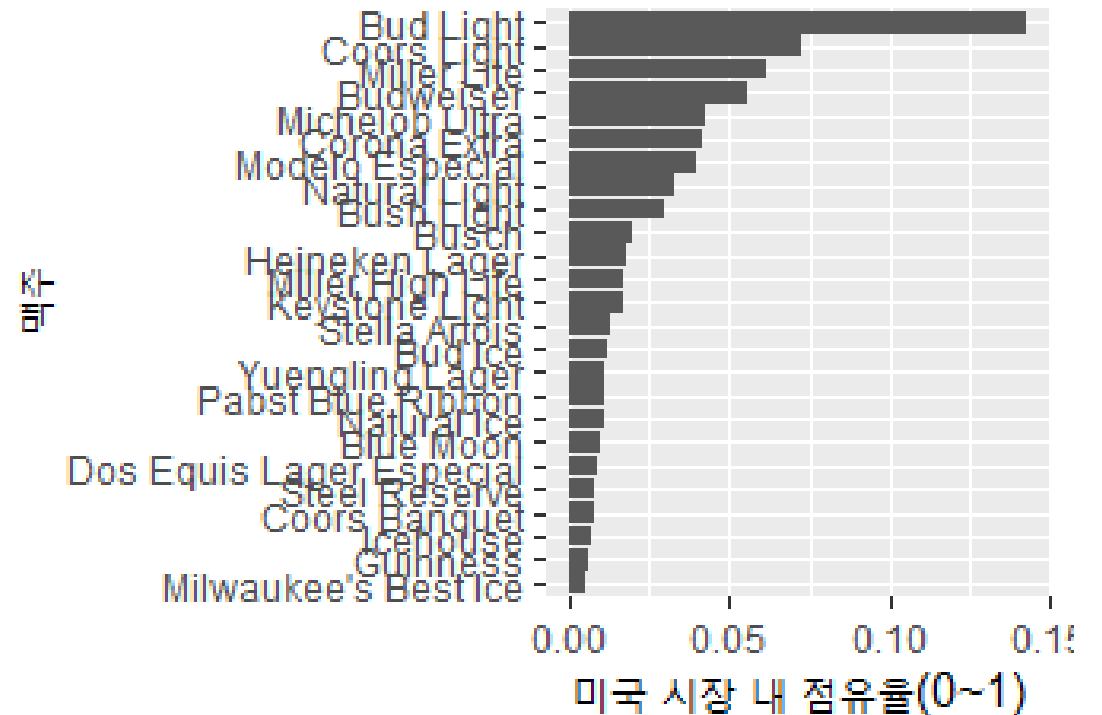
#### ■ 타겟시장인 미국 시장의 Top25 데이터 파악 및 점유율 시각화

	Name	ABV	IBU	SRM	Market_Share	Style
1	Bud Light	4.20	10.0	2.75	0.1430	American-Style Light Lager
2	Coors Light	4.20	10.0	2.75	0.0720	American-Style Light Lager
3	Miller Lite	4.20	10.0	2.75	0.0610	American-Style Light Lager
4	Budweiser	5.00	12.0	5.00	0.0550	American-Style Ice Lager
5	Michelob Ultra	4.20	10.0	2.00	0.0430	American-Style Light Lager
6	Corona Extra	4.60	10.0	2.00	0.0420	American-Style Lager
7	Modelo Especial	4.50	7.0	3.00	0.0400	American-Style Light Lager
8	Natural Light	4.20	10.0	2.00	0.0330	American-Style Light Lager
9	Bush Light	4.10	5.0	2.75	0.0300	American-Style Light Lager
10	Busch	5.00	12.0	2.00	0.0200	American-Style Lager
					⋮	

- TOP 25 맥주가 전체 미국 시장의 약 70%를 점유

```
> sum(us_ranking$Market_Share)
[1] 0.7001
```

- TOP 25 맥주별 미국 시장 내 점유율 시각화



# IV. 프로젝트 수행 결과

## 결과 제시 3. 탐색적 데이터 분석

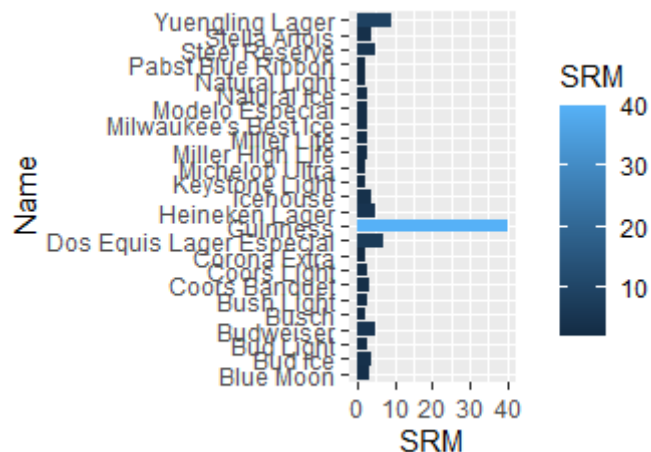
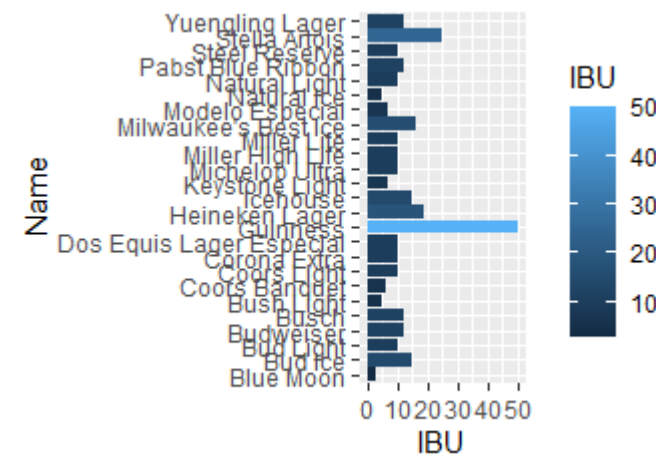
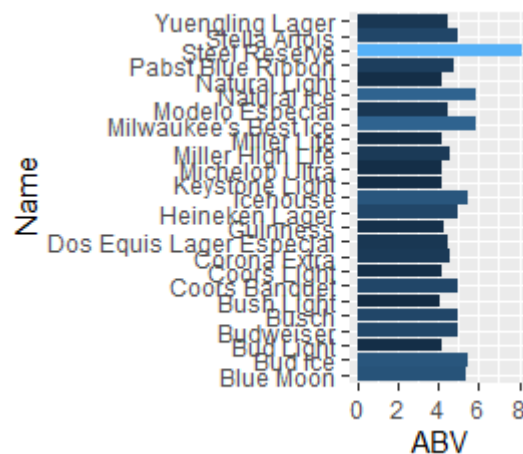
### 미국 내 맥주 판매 순위 Top 25의 ABU, IBU, SRM

```
> summary(us_ranking)
```

Name	ABV	IBU
Length:25	Min. :4.100	Min. : 3.00
Class :character	1st Qu.:4.200	1st Qu.:10.00
Mode :character	Median :4.600	Median :10.00
	Mean :4.894	Mean :12.42
	3rd Qu.:5.000	3rd Qu.:12.00
	Max. :8.100	Max. :50.00

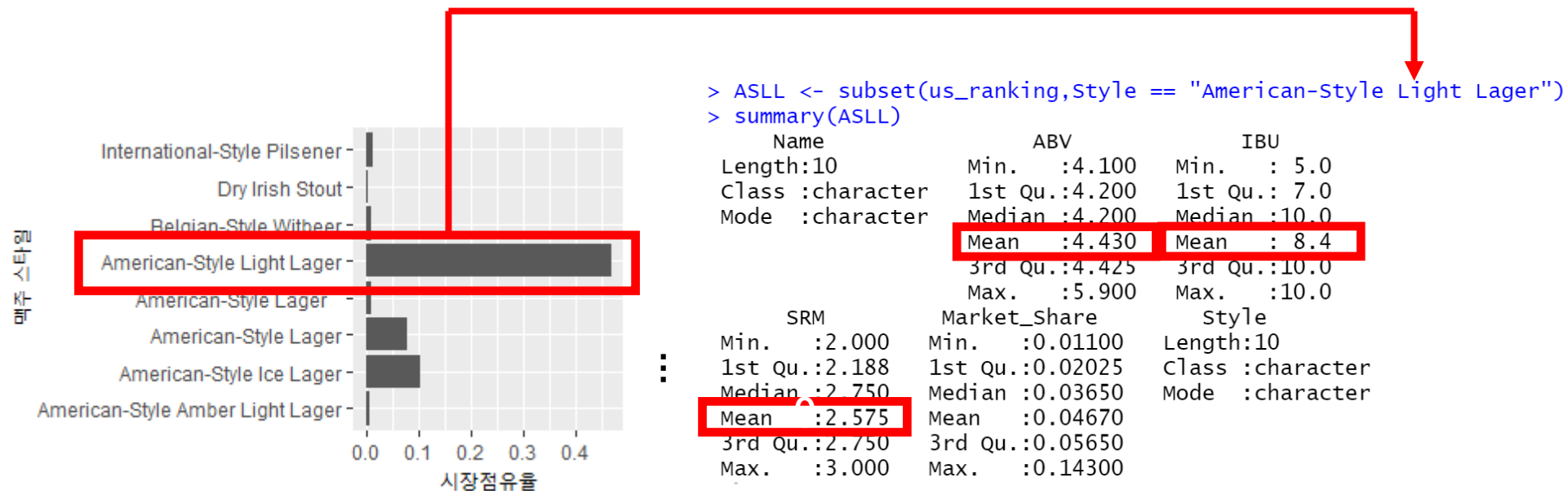
SRM	Market_Share	Style
Min. : 2.000	Min. :0.0051	Length:25
1st Qu.: 2.750	1st Qu.:0.0100	Class :character
Median : 3.000	Median :0.0170	Mode :character
Mean : 4.958	Mean :0.0280	
3rd Qu.: 4.000	3rd Qu.:0.0400	
Max. :40.000	Max. :0.1430	



## IV. 프로젝트 수행 결과

### 결과 제시 3. 탐색적 데이터 분석

- 미국 내 맥주 판매 순위 Top 25의 맥주 스타일별 시장 점유율

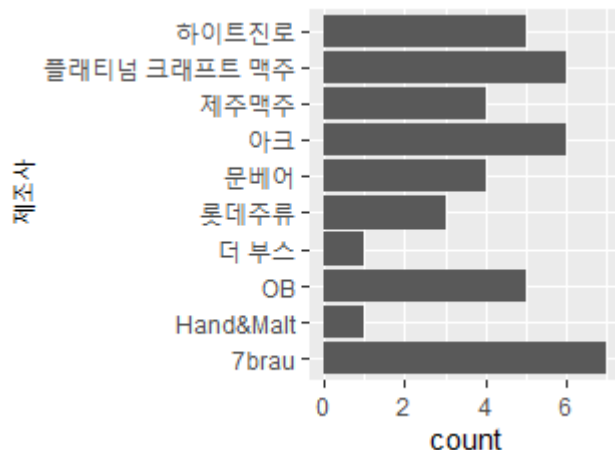


## IV. 프로젝트 수행 결과

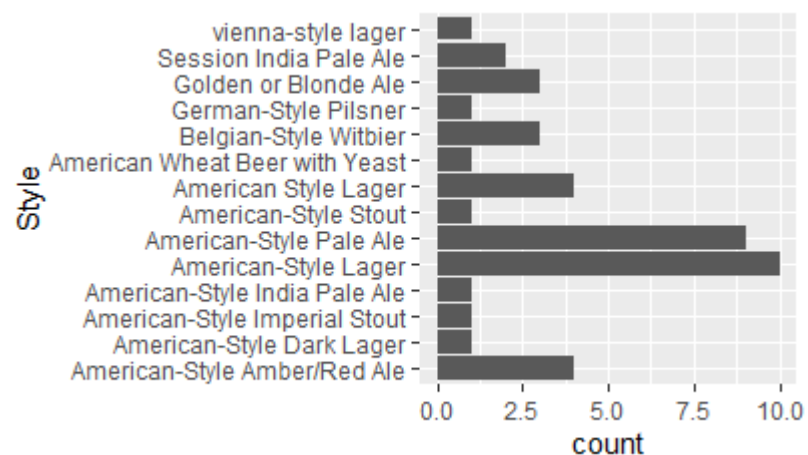
### 결과 제시 3. 탐색적 데이터 분석

- 수집한 국내 맥주 데이터 파악
  - 수출에 적합한 맥주 도출을 위해 브랜드 및 맥주 스타일에 대한 제한 두지 않음

#### • 제조사별 상품 개수



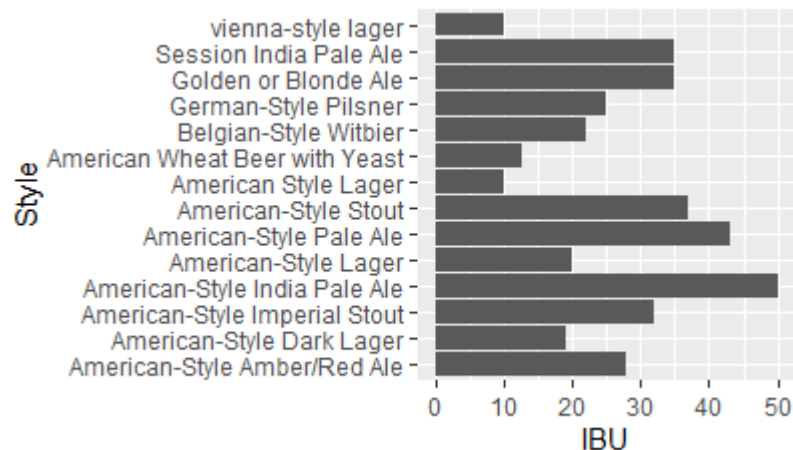
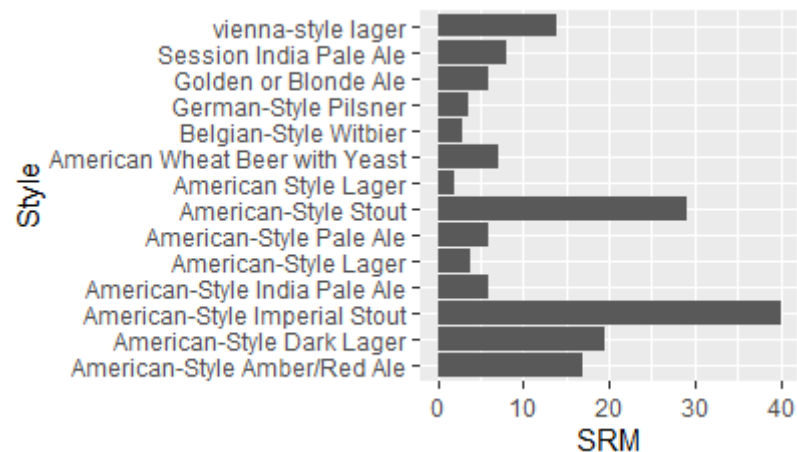
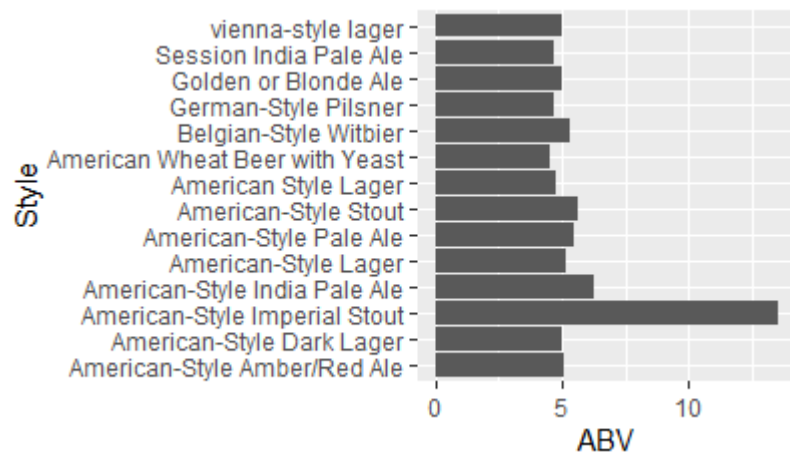
#### • 스타일별 상품 개수



## IV. 프로젝트 수행 결과

### 결과 제시 3. 탐색적 데이터 분석

- 국내 맥주 스타일별 ABV, IBU, SRM 값 시각화
  - ABV, IBU, SRM 값은 각 범위의 중간값



# IV. 프로젝트 수행 결과

## 결과 제시 3. 탐색적 데이터 분석

### ■ 미국 내 선호도가 가장 높은 맥주 스타일과 가장 유사한 국내 맥주 도출

- 미국 내 선호도가 가장 높은 맥주 스타일 : American-Light Lager
- American-Light Lager의 변수별 평균

#### • American Light Lager 변수의 평균

```
> us_mean <- c(mean(ASLL$ABV), mean(ASLL$IBU), mean(ASLL$SRM))
> us_mean
[1] 4.430 8.400 2.575
```

#### • 맥주 데이터셋의 각 변수 평균값 비교

- American-Light Lager 각 변수 평균값과 국내 맥주의 평균값의 차가 가장 낮은 국내 맥주 도출

```
> ko_beer[which.min(abs(ko_beer$ABV - us_mean[1])),]
# A tibble: 1 x 6
  제조사      name      ABV    IBU    SRM Style
  <chr>      <chr>    <dbl> <dbl> <dbl> <chr>
1 하이트진로 Hite extra ~ 4.5  12.5  3.81 American-Style ~

> ko_beer[which.min(abs(ko_beer$IBU - us_mean[2])),]
# A tibble: 1 x 6
  제조사      name      ABV    IBU    SRM Style
  <chr>      <chr>    <dbl> <dbl> <dbl> <chr>
1 하이트진로 Dry d    4.8    10     2 American Style Lager

> ko_beer[which.min(abs(ko_beer$SRM - us_mean[3])),]
# A tibble: 1 x 6
  제조사      name      ABV    IBU    SRM Style
  <chr>      <chr>    <dbl> <dbl> <dbl> <chr>
1 아크      해운대    4.2    15     2.5 Golden or Blonde Ale
```

- 각 변수별로 평균값의 차가 최소인 국내 맥주가 다르게 도출됨.

- 프로젝트의 의도는 모든 변수와의 차가 최소인 맥주 상품을 도출  
→ 변수(ABV, IBU, SRM)의 합계를 새로운 변수로 활용하게 됨.



## IV. 프로젝트 수행 결과

### 결과 제시 3. 탐색적 데이터 분석

- **탐색적 데이터 분석을 통한 결과 도출**
  - 미국 내 선호도가 가장 높은 맥주 스타일 : American-Light Lager
  - American-Light Lager의 대표적인 특징과 가장 유사한 국내 맥주 도출 : OB사의 Cass Light

# IV. 프로젝트 수행 결과

## 결과 제시 4. 모델링

- 미국 선호 맥주와 가장 유사한 국내 맥주 도출 : 3D 산점도 시각화
  1. 미국과 한국 맥주 데이터에서 변수 NAME, ABV, IBU, SRM 추출
  2. 추출한 데이터들을 병합하여 새로운 데이터 프레임에 담음
  3. ABV, IBU, SRM 변수를 기준으로 3D 산점도로 시각화

```
#로딩 및 필요행, 열만 추출
us_beer <- read_xlsx(file.choose())
us_beer <- select(us_beer, c(2:5))
is.na(us_beer)
us_beer <- us_beer[c(1:25),]
us_beer$gubun <- c("US") #국가 변수 추가

kobeer <- read_xlsx(file.choose())
ko_beer <- select(kobeer, c(2:5))
names(ko_beer) <- c("Name", "ABV", "IBU", "SRM")
is.na(ko_beer)
tail(ko_beer)
ko_beer$gubun <- c("KOR") # 국가 변수 추가

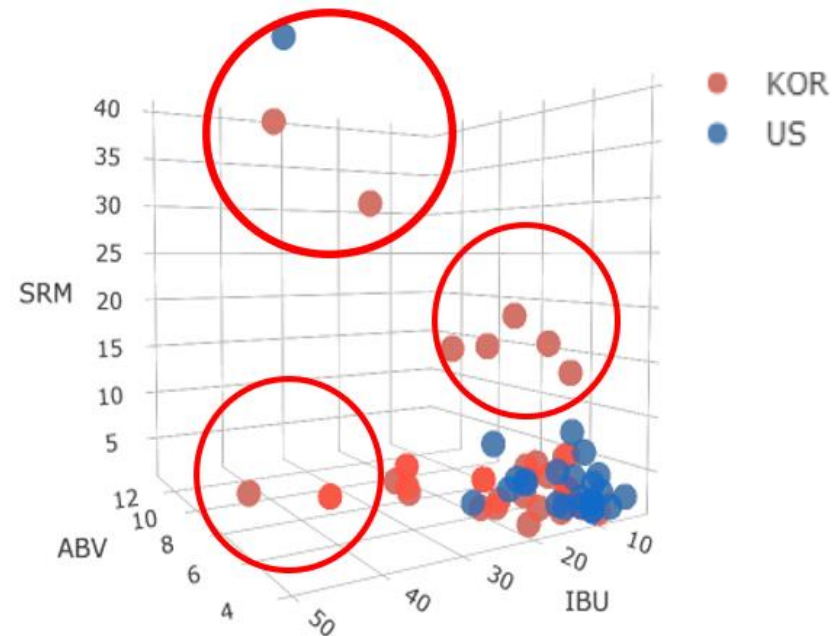
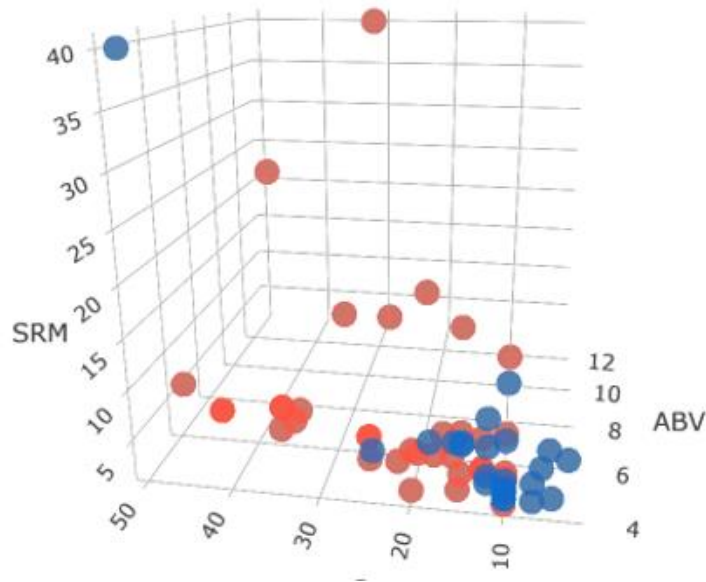
#데이터 병합, 형변환
beers <- rbind(us_beer, ko_beer)
beers$ABV <- as.numeric(beers$ABV)

#산점도 확인 (2차원으로는 3개의 변수를 동시에 볼 수 없기때문에 3차원 함수 사용)
p <- plot_ly(beers, x = beers$ABV, y = beers$IBU, z = beers$SRM,
             color = beers$gubun, colors = c('#BF382A', '#0C4B8E'), alpha = .7) %>%
  add_markers() %>%
  layout(scene = list(xaxis = list(title = 'ABV'),
                       yaxis = list(title = 'IBU'),
                       zaxis = list(title = 'SRM')
                     ))
```

# IV. 프로젝트 수행 결과

## 결과 제시 4. 모델링

- 미국 선호 맥주와 가장 유사한 국내 맥주 도출 : 3D 산점도 시각화
  - 제외 : 미국 선호 맥주와 밀집되어 있지 않은 국내 맥주, 집단과 떨어져 있는 미국 맥주
  - 밀집된 집단만 유의미한 데이터로 판단
    - > 서로 가까운 거리에 군집된값만 재샘플링
      - $x = 6(\text{ABV})$ ,  $y = 25(\text{IBU})$ ,  $z = 9(\text{SRM})$  이내 값들만 샘플링



# IV. 프로젝트 수행 결과

## 결과 제시 5. 계층적 클러스터링

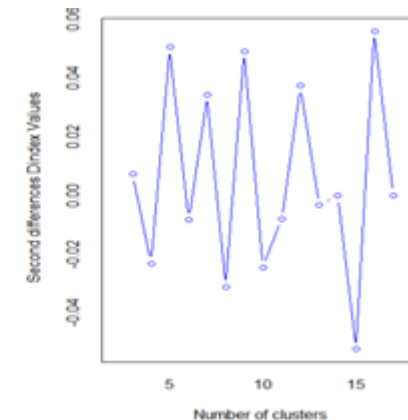
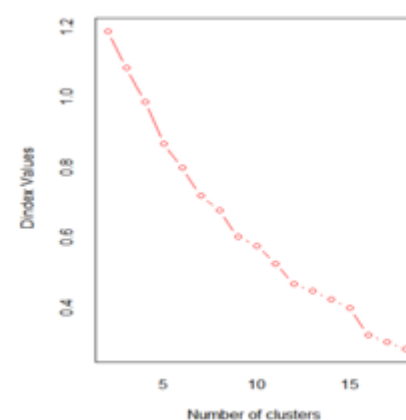
- 미국 선호 맥주와 가장 유사한 한국 맥주 도출 : 계층적 클러스터링
  1. ABV : 6이하, IBU : 25이하, SRM : 9이하인 데이터 재추출 -> 모델링에 사용할 연속형 변수 생성
  2. NbClust 통해 군집 개수 추천 받음 : 16개 그룹
  3. 16개의 그룹으로 시각화 : 유사한 맥주끼리 그룹핑

```
# 데이터 재추출
re_beers <- beers
re_beers$ABV <- ifelse(beers$ABV <= 6, beers$ABV, 0)
re_beers$IBU <- ifelse(beers$IBU <= 25, beers$IBU, 0)
re_beers$SRM <- ifelse(beers$SRM <= 9, beers$SRM, 0)
re_beers <- re_beers[!(re_beers$ABV == 0 | re_beers$IBU == 0 | re_beers$SRM == 0),]

# 모델링에 사용할 연속형 변수 추출
x <- re_beers[,c(2:4)]
y <- re_beers$Name

## 계층적 클러스터링
scaled_x <- scale(x) # 정규화
dist_x <- dist(scaled_x) # 거리 알아보기
cluster_x_avg <- hclust(dist_x, method = "average") # 평균거리법으로 군집화
NbClust(scaled_x, distance="euclidean", min.nc = 2, max.nc = 18, method = "average")
# 군집개수 추천받기
cutree_x_avg <- cutree(cluster_x_avg, k=16) # k개의 군집화에 따른 각각의 값 확인
table(cutree_x_avg)

plot(cluster_x_avg, hang=-1, cex=.6, labels = re_beers$Name, main = "맥주 클러스터링", xlab = "맥주")
# 그룹 시각화
rect.hclust(cluster_x_avg, k=16) # 그룹표시
```



```
* Among all indices:
* 5 proposed 2 as the best number of clusters
* 1 proposed 3 as the best number of clusters
* 2 proposed 4 as the best number of clusters
* 2 proposed 5 as the best number of clusters
* 2 proposed 6 as the best number of clusters
* 1 proposed 9 as the best number of clusters
* 6 proposed 16 as the best number of clusters
* 4 proposed 18 as the best number of clusters
```

\*\*\*\*\* Conclusion \*\*\*\*\*

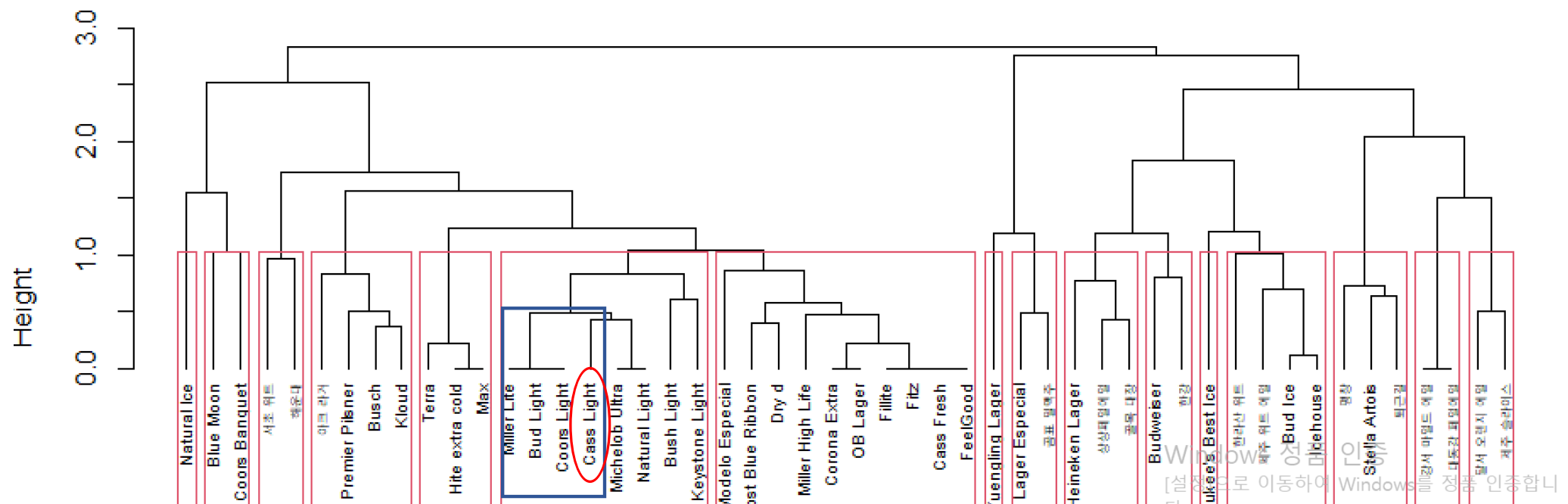
```
* According to the majority rule, the best number of clusters is 16
```

# IV. 프로젝트 수행 결과

## 결과 제시 5. 계층적 클러스터링

- 계층적 클러스터링을 통한 결과 도출
  - 미국시장 점유율 1~3위 맥주와 가장 근거리인 국내 맥주 : OB사의 Cass Light

맥주 클러스터링



## V.느낀 점

### 작성요령

- ❖ [느낀 점]은 프로젝트를 수행하면서 느끼거나 경험한 성찰이나 반성, 성과, 자신의 경력계획 등과 연관시켜 팀별 공통 의견 또는 개인 의견을 작성할 수 있다. 프로젝트를 마치고 수행상 어려움, 갈등요소 등을 작성하고 이를 해결한 방법을 작성한다. 또한 프로젝트 수행에서 개인, 우리 팀이 잘한 부분과 아쉬운 점을 작성한다. 그리고 프로젝트 수행을 통해 자신의 진로설계와 취업분야 탐색 및 의사결정 등 도움된 사항이 있었다면 구체적으로 작성한다.

## V. 느낀 점

구분	내용
프로젝트 수행상 어려움 극복 사례	<ul style="list-style-type: none"><li>정형 데이터를 가져오지 못하여 데이터 마이닝과 데이터 분석 과정에서 결과 도출이 힘들었다. 회귀 분석 과정에서 상관계수가 너무 낮고, 다중공산성이 발생하는 등 어려움을 겪었지만 원래 의도한 분석 과정에는 다행히 영향을 끼치지 않아 무사히 마무리할 수 있었다.</li></ul>
프로젝트에서 잘한 부분	<ul style="list-style-type: none"><li>온라인으로 진행되는 팀 프로젝트였지만, 비교적 소통이 원만하였다. 예상대로 되지 않은 부분도 많았음에도, 계획한 시간 내에 마무리할 수 있었다.</li></ul>
프로젝트에서 아쉬운 부분	<ul style="list-style-type: none"><li>주제에 대한 배경지식이 다소 부족했던 점과 데이터 접근이 힘들어서 전처리에서 시간이 많이 소요된 점이다.</li></ul>

## V.느낀 점

구분	내용
프로젝트를 통한 진로설계, 취업분야 탐색 및 결정 등 도움	<ul style="list-style-type: none"><li>이 프로젝트를 통해 빅데이터 관련 활용을 처음 해보면서 데이터 접근성에 대한 중요함을 깨닫고, 분석과 마이닝 과정을 다시 익히면서 성장할 수 있었다. 수업 시간에는 어렵다고만 생각했지만 시행착오를 겪으면서 활용해보니 마냥 불가능한 일은 아니라는 것을 깨달았다. 관련 취업 분야에 취업하고 싶은 마음이 커졌다.</li></ul>
기타	<ul style="list-style-type: none"><li>대학 졸업 후 많은 회사에 지원을 하고 결과가 좋지 않아 우울한 나날들을 보냈었는데, 새로운 기술을 배우는 것에 도전하면서 재미를 느끼고 프로젝트 진행에서도 데이터를 다루는 작업을 하면서 나도 할 수 있다는 자신감을 얻었다.</li></ul>