

Yelp's Hotel Advertisement Business Project





목차

1. Analysis Outline

- Business objective
- Process of analysis

2. Data Outline

- Business Data Set
- User Data Set

3. Exploratory Data Analysis

- Hotel's Location 시각화 분석
- 평점, 리뷰, 체크인, Tip 시각화 분석
- 연도별 리뷰 및 월별 리뷰 횟수

4. Unsupervised Learning

- Hotel 사업자 세분화
- Hotel 사업자 세분화 기준 변수
- Hotel 사업자 세분화 결과

5. Profiling

- Hotel 사업자 Sub Category
- Hotel 사업자 Profiling
- Target Group Location in NV & AZ
- Target Group 리뷰 분석

6. Business Proposal

- Target Group 광고 Concept

1

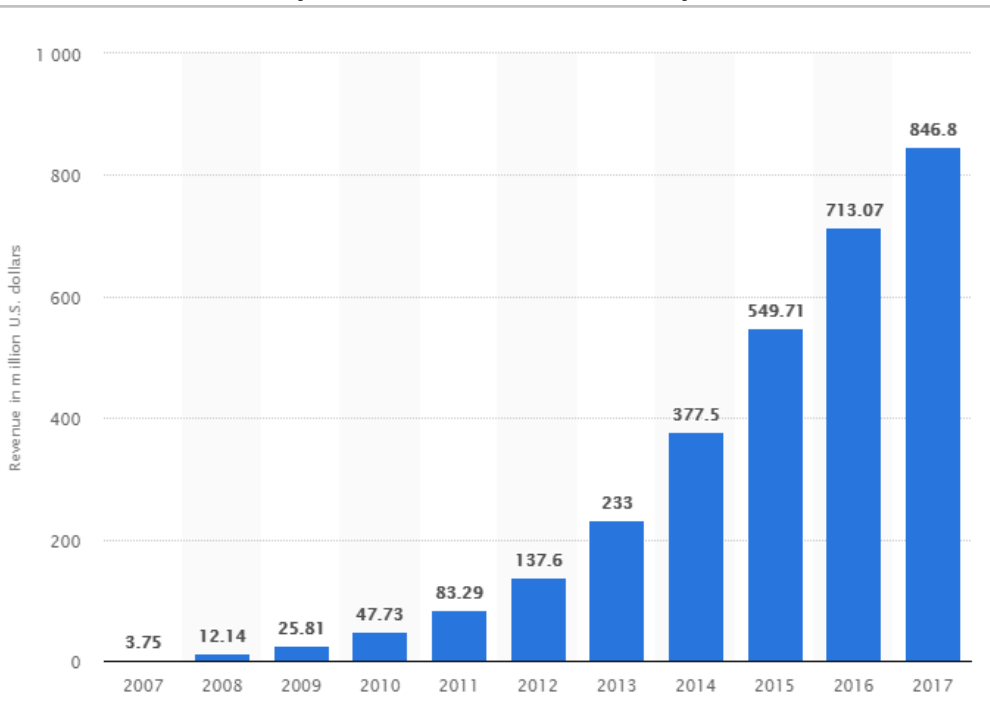
Analysis Outline

분석 목적과 분석 절차에 대해 간단하게 설명한다.



Business objective

Yelp's annual net revenue from 2007 to 2017
(in million U.S. dollars)



<https://www.statista.com/statistics/278022/yelps-annual-net-revenue/>

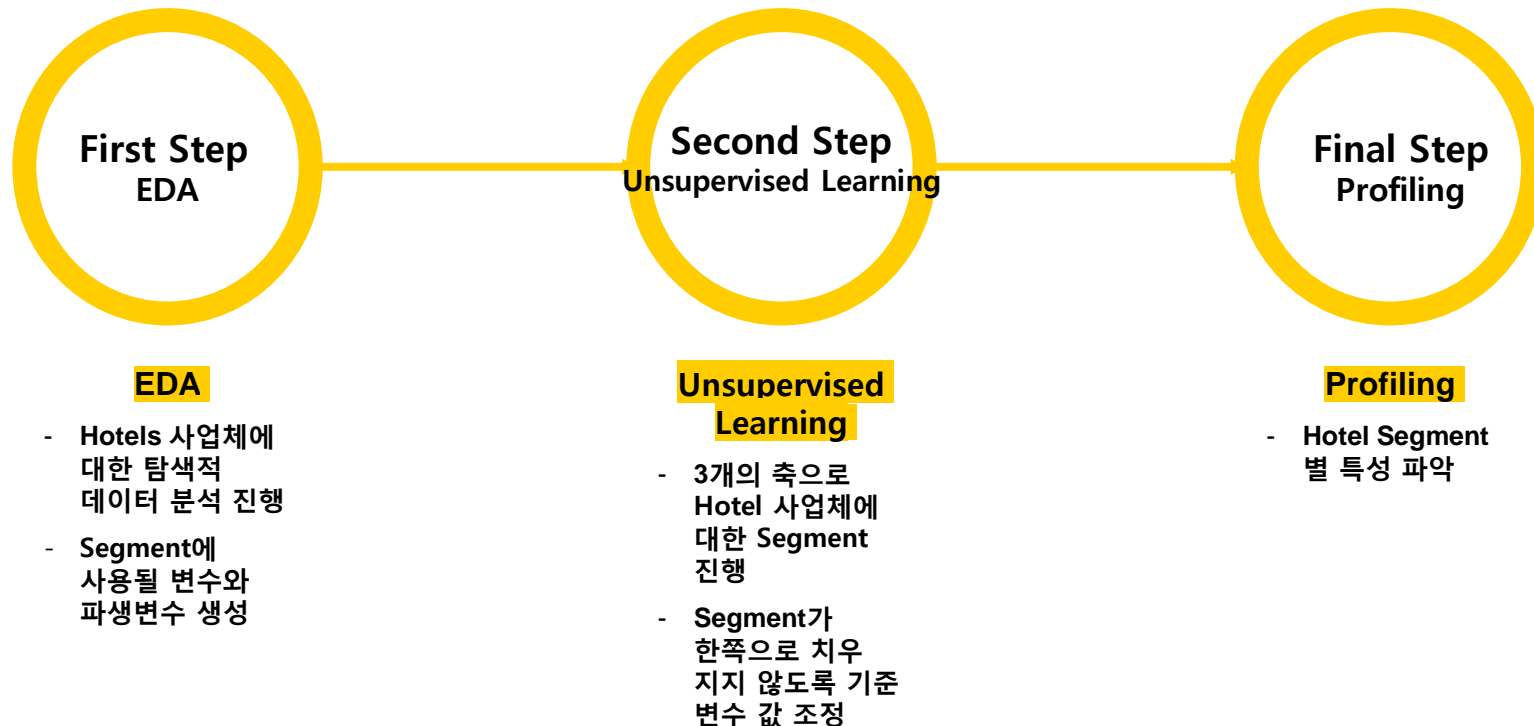
- Yelp의 2017년 순수익은 846.8백만 달러로 매년 급속한 성장을 보이고 있음
- 이러한 급성장 속에서 신규 가맹점 확보를 통해 현재 입지를 더욱 강화시키는 것이 필요함
- 현재 174,567개 사업체 중 Hotel 업체는 5,736개로 약 3.3% 차지하고 있음
- 타 업종에 비해, Hotel 가맹점의 수가 상당히 적은 상황으로 Hotel 업종의 신규 가맹점 확보가 필요
- 기존에 등록된 Hotel 업체들의 데이터 분석을 통해 비즈니스 세분화를 진행
- 세분화된 Hotel 업체들의 특성을 파악하고 Target 집단을 선정
- 선정된 집단에 남긴 고객들의 리뷰를 분석한 뒤 가장 효율적인 광고 concept 작성
- 선정된 집단과 유사한 비 가맹업체들을 선별하여 분석 결과물을 가지고 영업 진행



Process of Analysis



This process of Analysis has three steps



2

Data Outline

분석에 사용된 총7개 Data table을

Business Data User Data로 각각 구분하여 전체 데이터에 대한 간단한 요약을 진행한다.



Business Data set

Observations: 174,567

Variables: 13

```
$business_id <chr> "FYWN1wneV18bWNgQjJ2GNg", ...
$name <chr> "Dental by Design",...
$neighborhood <chr> "", "Summerlin", ...
$address <chr> "4855 E Warner Rd, Ste B0"
$city <chr> "Ahwatukee", "M
$state <chr> "AZ", "PA", "AZ"
$postal_code <chr> "85044", "15317
$latitude <dbl> 33, 40, 34, ...
$longitude <dbl> -112.0, -80.1, -1
$stars <dbl> 4.0, 3.0, 1.5, ...
$review_count <int> 22, 11, 18, ...
$isOpen <int> 1, 1, 1, 0, ..
$categories <chr> "Dentists;C
```

Observations: 174,567

Variables: 8

```
$business_id <chr> "FYWN1w
$monday <chr> "7:30-17:0
$tuesday <chr> "7:30-17:0
$wednesday <chr> "7:30-17:0
$thursday <chr> "7:30-17:0
$friday <chr> "7:30-17:0
$saturday <chr> "None", "8
$sunday <chr> "None", "N
```

Observations: 3,911,218

Variables: 4

```
$business_id <chr> "3Mc-LxcqeguOXOVT_2ZtCg", ...
$weekday <chr> "Tue", "Wed", ...
$hour <chr> "0:00", "0:00", ...
$checkins <int> 12, 4, 1, 1, 1, ...
```

Observations: 152,041

Variables: 82

```
$business_id <fct> FYWN1wneV18bjJ2GNg, ...
$AcceptsInsurance <fct> Na, Na, Na, Na, ...
$ByAppointmentOnly <fct> Na, Na, Na, Na, ...
$BusinessAcceptsCreditCards <fct> Na, Na, Na, False, ...
$BusinessParking_garage <fct> True, Na, Na, False, ...
$BusinessParking_street <fct> Na, Na, False, Na, ...
$BusinessParking_validated <fct> Na, False, Na, True, ..
$BusinessParking_lot <fct> Na, False, Na, Na, ..
$BusinessParking_valet <fct> Na, Na, Na, Na, ...
$HairSpecializesIn_coloring <fct> Na, Na, False, Na, ...
$HairSpecializesIn_africanamerican <fct> Na, Na, Na, Na, ...
```

<Yelp business Data Set>

174,567개 사업체에 대한 데이터로 기본적인 정보를 가지고 있음

<Yelp checkin Data Set>

사업체에 대한 요일별 checkin을 나타내는 데이터로 3,911,218개의 레코드가 있음

<Yelp business hours Data Set>

174,567개 사업체에 대한 요일별 운영 시간 정보를 가지고 있음

<Yelp business attributes Data Set>

152,041개 사업체에 대한 각각의 Attributes를 가지고 있음

Ex) 발렛파킹 여부, 신용카드 사용 여부 등.



User Data set

Observations: 1,326,100

Variables: 22

```
$ user_id      <chr> "JJ-aSuM4pCFPdkfoZ34q0Q", ...
$ name        <chr> "Chris", "Tiffany",...
$ review_count <int> 10, 1, 6, 3,...
$ yelping_since <chr> "2013-09-24", "2017-03-02",...
$ friends     <chr> "0njfJmB-7n84DIlgUByCNw, ...
$ useful      <int> 0, 0, 0, 0, 4,...
$ funny       <int> 0, 0, 0, 0, ...
$ cool        <int> 0, 0, 0, 0,...
$ fans        <int> 0, 0, 0, 0,...
$ elite       <chr> "None", "No
$ average_stars <dbl> 3.7, 2.0, 4.7
$ compliment_hot<int> 0, 0, 0, 0, ...
```

Observations: 5,261,668

Variables: 9

```
$ review_id    <chr> "vkVSCC7xljirAl4UGfnKEQ",...
$ user_id      <chr> "bv2nCt5Qv5vroFtqKGopiW",...
$ business_id  <chr> "AEx2SYEUJmTxVVB18LICwA",...
$ stars        <int> 5, 5, 5,...
$ date         <date> 2016-05-28,...
$ text         <chr> "Super simple place...",...
$ useful       <int> 0, 0, 0, 0,...
$ funny        <int> 0, 0, 0, 0,...
$ cool         <int> 0, 0, 0, 0,...
```

Observations: 1,098,324

Variables: 5

```
$ text         <chr> "Great breakfast large...",...
$ date         <chr> "2015-08-12",...
$ likes        <int> 0, 0, 0, 0,...
$ business_id  <chr> "jH19V2l9fIsInNhDzPmdkA",...
$ user_id      <chr> "ZcLKXikTHYOnYt5VYRO5sg",...
```

<Yelp User Data Set>

1,326,100명에 대한 정보를 가지고 있음

Yelp를 처음 시작한 일자 및 리뷰 작성한 횟수 등

<Yelp Review Data Set>

5,261,668개의 리뷰 정보를 가지고 있으며

Business ID, User ID 및 리뷰가 작성된 일자 등의 정보를 가지고 있음

<Yelp Tip Data Set>

1,098,324개의 Tip 정보를 가지고 있으며

Tip이란? 해당 Business의 Main 리뷰를 뜻함

3

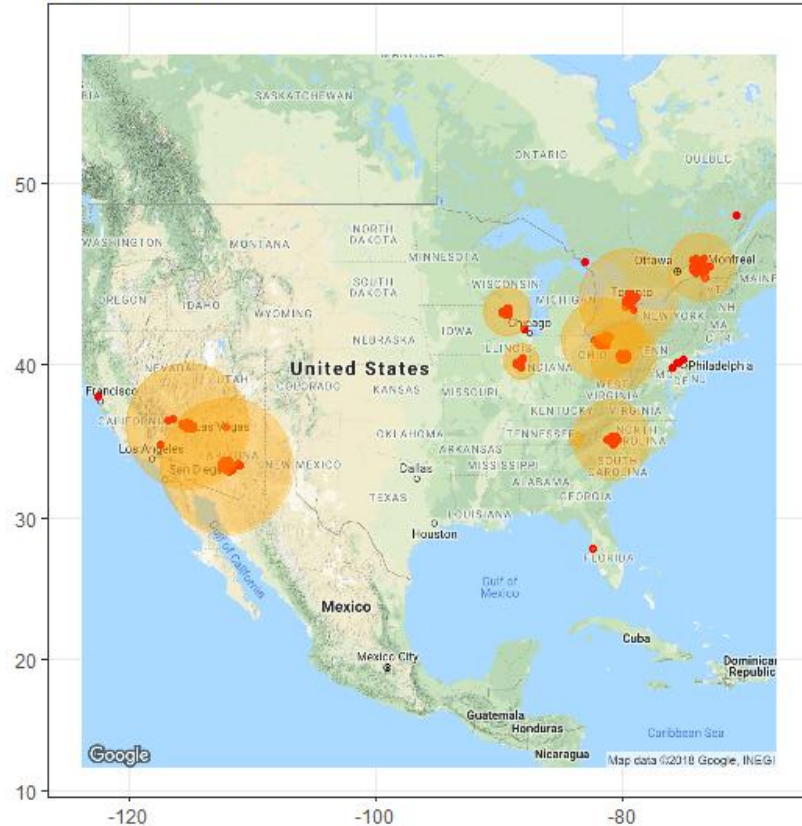
Exploratory Data Analysis

Hotel 사업체에 대한 EDA를 진행하고, Segment에 사용될 기준변수를 정의한다.

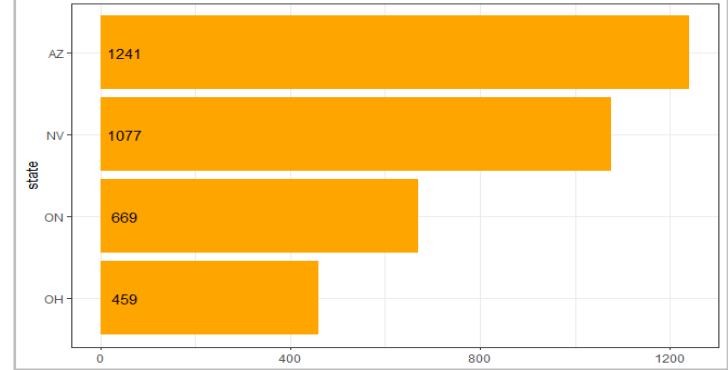


Hotel's Location 시각화 분석

Hotel Location



Hotel 사업체가 가장 많은 지역

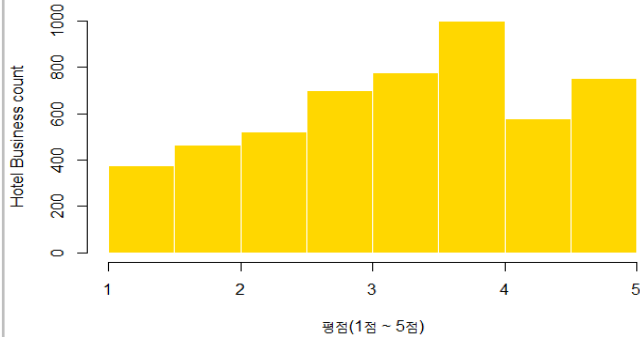


- 미국의 경우 Arizona, Nevada, Ohio주에 Hotel이 가장 많이 있음
- 캐나다의 경우 Ontario 지역에 가장 많은 Hotel이 있음

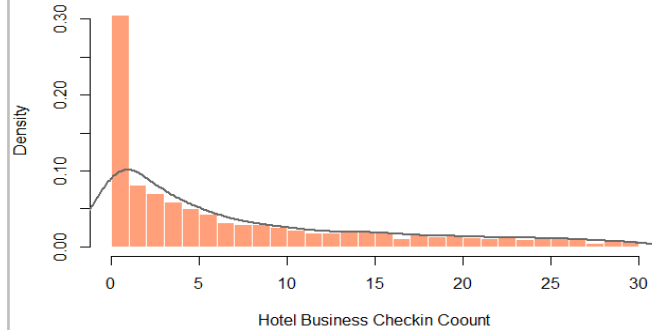


평점, 리뷰, 체크인, Tip 시각화 분석

Hotels 사업체들의 평점 Histogram



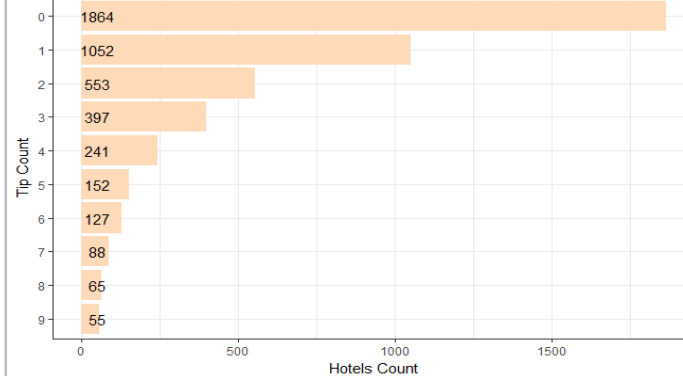
Hotels 사업체들의 Checkin Histogram



Hotels 사업체들의 리뷰 개수 Histogram



Tip 개수별 Hotels 사업체 분포

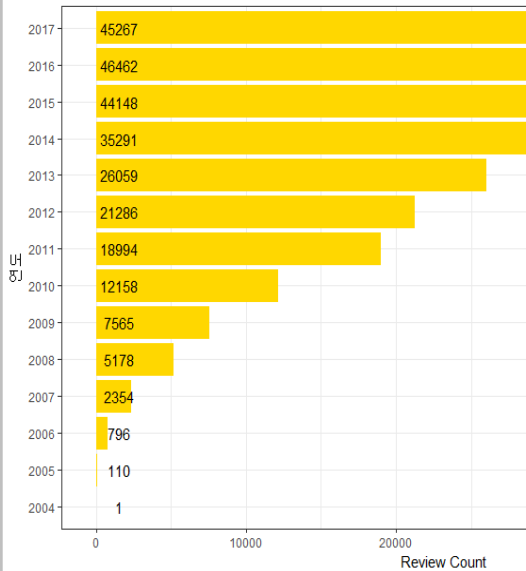


- 평점, 3.5점을 받은 업체들이 대부분이며 5점 받은 업체들도 많음 상대적으로 낮은 평점을 받은 업체들은 적음
- 리뷰 및 체크인 횟수의 경우 5건 이하인 업체들이 대부분이며, 편차가 심함
- Tip 횟수의 경우 0건 또는 1건 있는 업체들이 대부분임



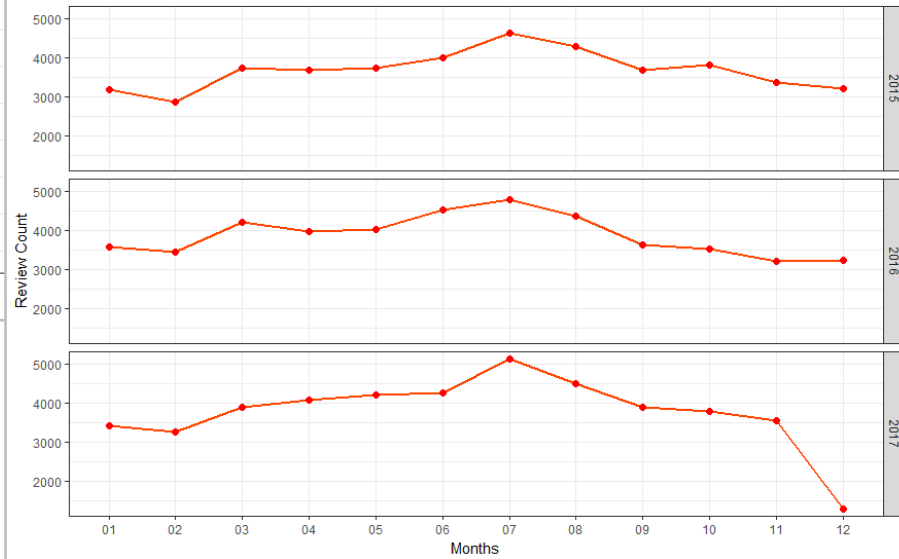
연도별 리뷰 및 월별 리뷰 횟수

Hotel 사업체들의 연도별 리뷰 개수



- 매년, 업체들에 대한 리뷰 횟수가 증가하고 있음
- Yelp 이용 고객들의 증가와 함께 상승한 것으로 판단
- 월별 리뷰 횟수의 경우, 7월에 가장 많은 리뷰가 발생
- 여름 휴가철에 대한 계절적 요인으로, 이용 고객들이 증가함에 따라 리뷰 또한 상승한 것으로 판단

Hotel 사업체들의 월별 리뷰 개수



4

Unsupervised Learning

Hotel Business Segment 를 위한 기준 변수를 정의하고 Model을 구축한다.



Hotel 사업자 세분화

세분화 기준 변수



평점(Star)

호텔 이용 고객들이 부여한 평점

Low

Middle

High



최근성

처음 댓글 달린 연도와 마지막 댓글 달린 연도에 따른 분류

period_low

period_middle

period_high

period_very_high



인지도

사업장 Review 횟수

사업장 Tip 횟수

1

2

3

4

5

Unsupervised Learning

비지도학습이란, 입력값에 대한 목표치가 주어지지 않은 상황에서 데이터 자체에 대한 유용한 패턴을 찾아내는 학습 방법이다.

군집화(Clustering)

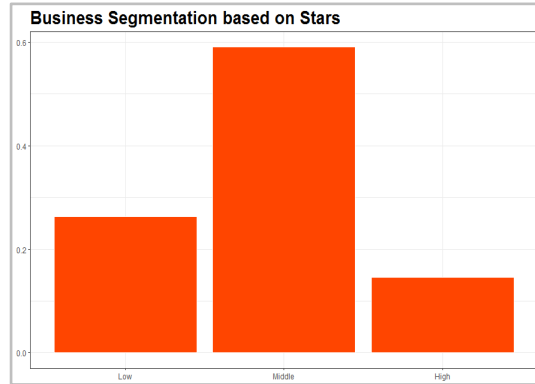
데이터가 쪼그라져 있을 때 레이블이 없다고 해도 데이터간 거리에 따라 대충 두 세 개의 군집으로 나눌 수 있다. 이렇게 x만 가지고 군집을 학습하는 것

분포 추정(Underlying Probability Density Estimation)

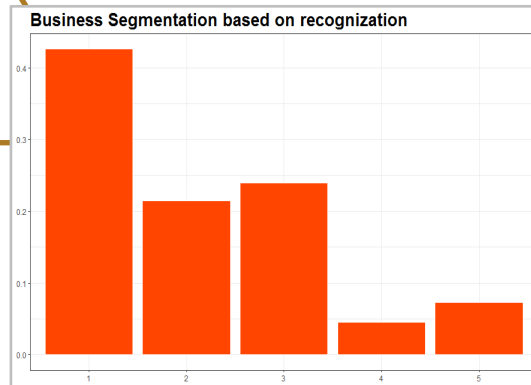
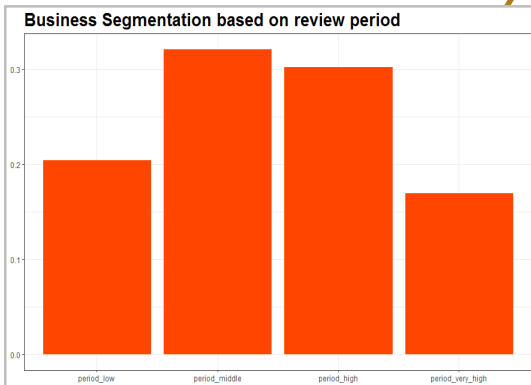
군집화에서 더 나아가서, 데이터들이 쪼그라져 있을 때 애네들이 어떤 확률 분포에서 나온 샘플들인지 추정하는 문제



Hotel 사업자 세분화 기준 변수



- **Stars Segment**
- Low : 2.5점 이하 / Middle : 4.5점 이하 / High : 4.5점 초과
- **Review period Segment**
- Period_low : 처음 리뷰 달린 연도가 2015년 이후
- Period_middle : 처음 리뷰 달린 연도가 2012년 이후
- Period_high : 처음 리뷰 달린 연도가 2009년 이후
- Period_very_high : 처음 리뷰 달린 연도가 2004년 이후



- **Recognition**
- 1 / 2 / 3 / 4 / 5
- 리뷰 및 Tip 횟수를 나타내는 변수 사용
- 변수들에 대한 점수 부여, Recognition 변수 생성
- 5점 척도로 나타냈으며, 점수가 높을 수록 인지도가 높음



Hotel 사업자 세분화 결과

60개 집단으로 세분화

Review Period	Stars	Recognition	Business
period_low	Low	1	234
period_low	Low	2	48
period_low	Low	3	20
period_low	Low	4	4
period_low	Low	5	1
period_low	Middle	1	288
period_low	Middle	2	85
period_low	Middle	3	64
period_low	Middle	4	8
period_low	Middle	5	6
period_low	High	1	222
period_low	High	2	60
period_low	High	3	17
period_low	High	4	3
period_low	High	5	1
period_middle	Low	1	317
period_middle	Low	2	117
period_middle	Low	3	69
period_middle	Low	4	12
period_middle	Low	5	8
period_middle	Middle	1	397

사업체 수가 가장 많은 TOP5 집단

period_high	Middle	3	400
period_middle	Middle	1	397
period_middle	Low	1	317
period_high	Middle	1	299
period_low	Middle	1	288

- 오래됐고 평점과 인지도가 중간인 업체들이 400개로 가장 많음
- Hotel 운영 기간이 길수록 인지도 또한 증가하는 것이 일반적이거나 세분화 결과, 운영 기간에 관계없이 인지도가 낮은 업체들이 많음
- 따라서, 인지도에 영향을 미치는 특정 요인이 있을 것으로 판단

5

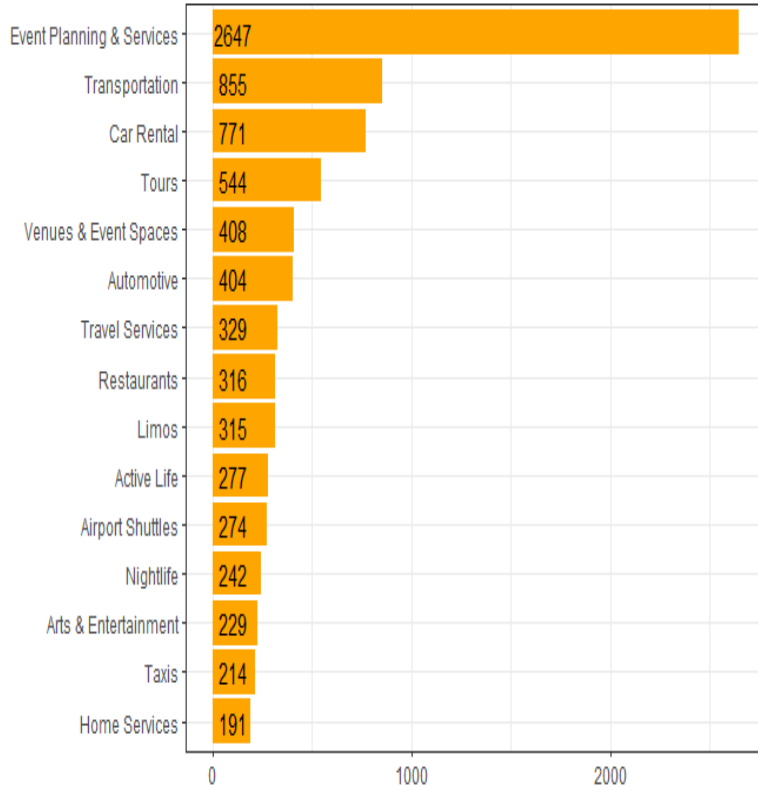
Profiling

Sub Category를 활용한 프로파일링 진행
Target Group을 선정하여 고객들이 남긴 리뷰 분석



Hotel 사업자 Sub Category

Hotel 사업체들의 Sub Category



Casinos

Casinos

Event

Event Planning & Services

Venues & Event Spaces

Party & Event Planning

Meal

Restaurants

Bed & Breakfast

Bars

Food

useful

Breakfast & Brunch

Limos

Airport Shuttles

- 506개 Sub Category를 [Casinos / Event / useful / Meal] 4개로 분류
- Sub Category 빈도수를 고려하여 비슷한 항목끼리 분류될 수 있도록 정의



Hotel 사업자 Profiling

period_review	stars_seg	recog_score	Hotel	likes	checkins	Casino	Event	AZ	NV
period_low	Low	1	234	9	3	0%	35%	21%	22%
period_low	Middle	1	288	9	7	0%	43%	20%	18%
period_low	High	1	222	7	3	1%	30%	31%	26%
period_high	Middle	1	299	13	19	0%	52%	12%	5%
period_high	Middle	2	242	37	47	1%	57%	23%	10%
period_high	Middle	3	400	71	84	0%	72%	30%	14%
period_very_high	Middle	3	212	87	98	0%	78%	24%	15%
period_very_high	Middle	5	227	1,601	4,441	28%	82%	29%	52%
<div> <div></div> <div></div> <div></div> <div></div> </div>									

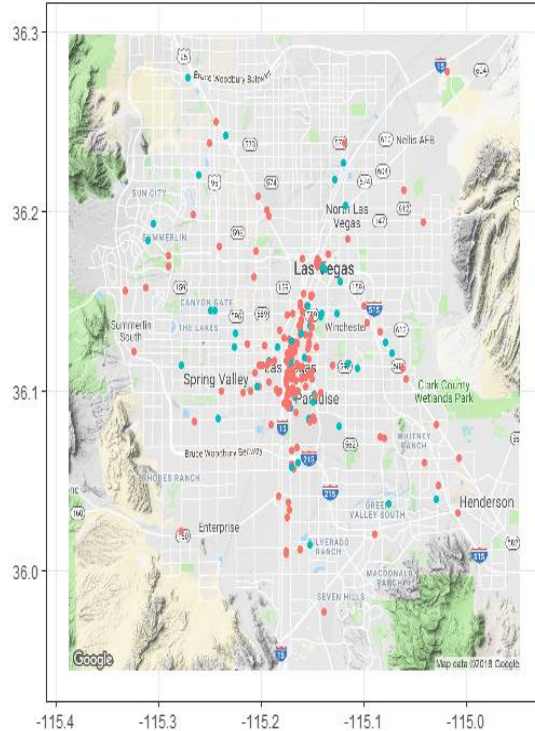
- 운영기간이 길고 인지도가 높은 Hotel
- Casino를 운영하는 업체비중이 높음
- Event 서비스를 대부분 제공
- Nevada주에 많음
- likes 및 checkins 수가 높음
- Casino 및 다양한 Event 서비스를 제공하는 것으로 보아 규모가 큰 Hotel들이 많을 것으로 판단

- 운영기간이 짧고 평점 및 인지도가 낮은 Hotel
- Likes 및 checkins 수가 낮음
- Event 서비스를 제공하는 업체비중이 낮음



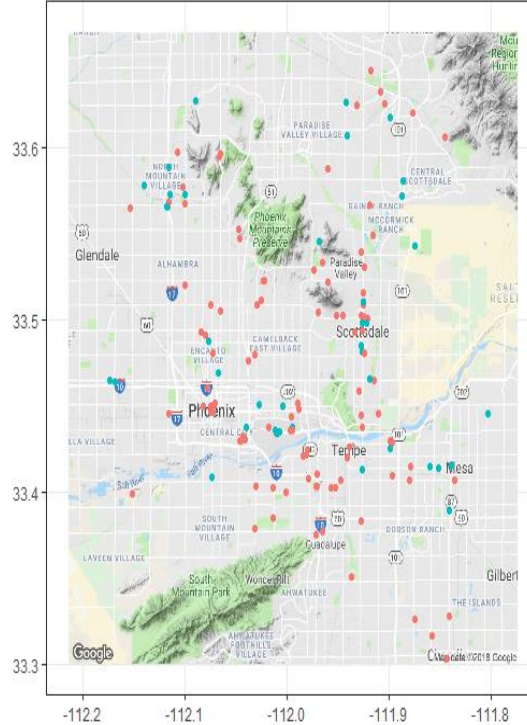
Target Group Location in NV & AZ

Hotel's location in Nevada



• 운영시간이 길고 평점 및 인지도가 높은 Hotel • 운영시간이 짧고 평점 및 인지도가 낮은 Hotel

Hotel's location in Arizona



• 운영시간이 길고 평점 및 인지도가 높은 Hotel • 운영시간이 짧고 평점 및 인지도가 낮은 Hotel

- Nevada 주와 Arizona 주에 위치한 Hotel location 시각화 자료
- Nevada주의 경우 Las Vegas에 Hotel 대부분이 모여 있음
- Arizona주의 경우 Phoenix와 Scottsdale에 Hotel 대부분 모여 있음



운영기간이 짧고 평점 및 인지도가 낮은 Hotel 리뷰 분석

해당 Group Hotel들의 경우 리뷰수가 적고 부정적 리뷰가 많음

N E G A T I V E	<p>Eat before you come here. There are hardly any restaurant in this terminal.</p> <p>I have been trying to get a phone connection the room is cool</p> <p>Don't stay here if you have other choices. My pillow case had stains on it, the walls are filthy, the paint on the door is coming off. Booger on my picture in the room, stickers on the fridge. I can go on and on but I think you get the idea.</p> <p>Listen to Matt H...because I didn't and I have bed bug bites.</p> <p>Rip off artist.</p> <p>Staff presence non existent</p> <p>DO NOT USE THIS Company!! Worst most disgusting experience of my life.</p> <p>The trunk was bungee corded down, filthy dirty, dents everywhere in car, filthy dirty cups, ripped seats, the list goes on and on.</p> <p>Will never fly jet blue again... Costumer service was very poor...</p> <p>The female in front of the carryon baggage x-ray machine had a personality of a troll. No "please" in it's vocabulary.</p> <p>Bring your own bedding, curtains, and preferably your own hotspot for wifi.</p> <p>Go somewhere else! Not safe!</p> <p>Discussing! Needs a lot of work.</p> <p>**If planning an extended stay** Bring your own toilet paper, shower curtain, kitchen utensils Etc. Website says its included buy talking with a resident it hasn't been included for about a year. Also LAUNDRY..is a prepaid buy from them (2.00 for card) the cost ls a joke to reload it only takes \$5,10,20's and a wash load is 1.75ea ..</p> <p>This place is a big scam don't fall for their tricks!</p>
P O S I T I V E	<p>Great cars</p> <p>Branch manager Joel and team are excellent!</p> <p>Great</p> <p>The provide helmets upon request & sell goggles that are highly recommended! Have fun!</p> <p>Awesome place to work! Managers, and co-workers are all cool and easy to work with.</p> <p>nicole was awesome</p>



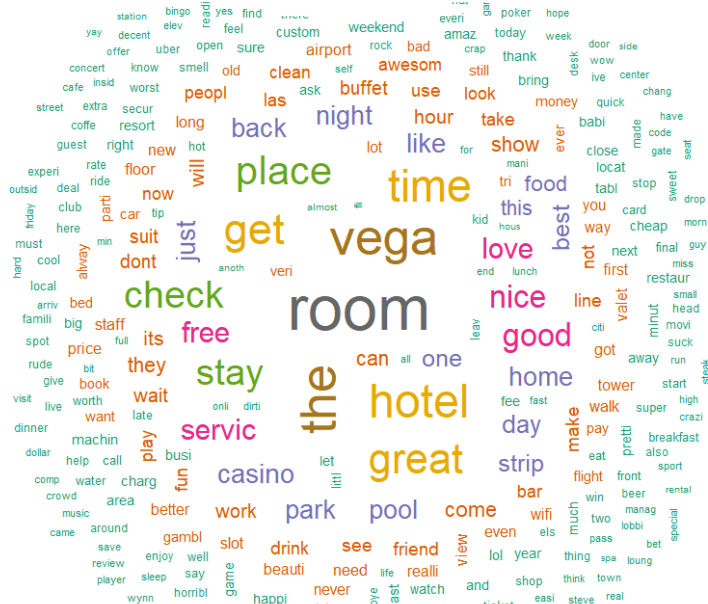
운영기간이 길고 평점 및 인지도가 높은 Hotel 리뷰 분석(1)

Arizona



- Arizona주에 위치한 Hotel의 경우 “great”, “home”, “airport”단어들이 주로 나타난 것으로 확인 됨

Nevada



- Nevada주에 위치한 Hotel의 경우 “room”, “vega”, “hotel”단어들이 주로 나타난 것으로 확인 됨

6

Business Proposal

신규 가맹점 영업을 위한 Target Group을 선정하고
기존에 등록된 Hotel 업체들의 리뷰분석을 통해 영업을 필요한 광고(홍보)방안 제시

● 운영기간이 짧고 평점 및 인지도가 낮은 Hotel 광고 Concept

Hotel Via Claimed

★★★★★ 80 reviews

Write a Review

Add Photo Share Save

\$\$\$ · Hotels

138 King St
San Francisco, CA 94107
bt 3rd St & 2nd St
South Beach

Get Directions

(415) 200-4977

hotelvia.com

Send to your Phone

Photo of Hotel Via - San Francisco, CA, United States

See all 133

"Make sure to checkout the **rooftop bar** where theres an amazing panoramic view of the city." in 24 reviews

"I love small boutique hotels and love this place is right across the street from the greatest ballpark ever- **AT&T Park**!" in 23 reviews

Today Open 24 hours Open now

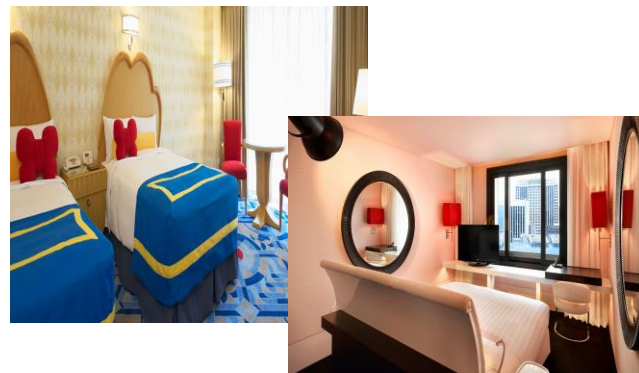
\$\$\$ Price range Pricey

Hours

Mon Open 24 hours

Tue Open 24 hours

- 해당 업체들의 리뷰 분석 결과
- 긍정적 리뷰보다 부정적 리뷰가 더 많았음
- Hotel에 대한 기본적인 시설들에 대한 부정적 리뷰가 많음
Ex) 얼룩진 침대 시트, 더러운 벽지 및 식기구, 부족한 화장지, Wifi 부재 등..
- 따라서, 기본적인 시설(침대,벽지,화장지...)에 대한 서비스를 강화하고 관련 서비스를 홍보할 수 있도록 광고 제작
- Yelp 홈페이지의 업체 정보 페이지에서 홍보 사진을 깔끔한 이미지의 객실 안 사진 위주로 홍보





운영기간이 길고 평점 및 인지도가 높은 Hotel 광고 Concept (AZ)



- Arizona에 위치한 Hotel 리뷰 분석 결과
- 부정적 리뷰보다 긍정적 리뷰가 더 많았음
- 긍정적 리뷰 분석 결과
- Pool에 대한 긍정적 언급이 많았음
- 또한, 가족단위로 놀러 오는 것에 대한 Tip이 많았고
- Hotel 위치에 대한 언급이 많았음
- 따라서, 해당 집단과 비슷한 신규 Hotel 가맹점 영업 시
- 리뷰 분석 결과를 토대로,
- 수영장 시설과 가족 단위의 이미지로 홍보하는 방안을 제시
- 추가로, Hotel 위치에 대한 언급도 강조



운영기간이 길고 평점 및 인지도가 높은 Hotel 광고 Concept (NV)

The Cosmopolitan of Las Vegas
Claimed

★★★★★ 4183 reviews Details

\$\$\$ · Hotels, Casinos Edit

3708 Las Vegas Blvd S
Las Vegas, NV 89109
The Strip
Get Directions
(702) 698-7000
cosmopolitanlasvegas.com
Message the business
Send to your Phone

Can smoke cigars on the terrace?

Yes they are a smoke friendly hotel
6 months ago · 5 people found this helpful
View 11 more answers

See all 12 questions for The Cosmopolitan of Las Vegas

Recommended Reviews for The Cosmopolitan of Las Vegas

Your trust is our top concern, so businesses can't pay to alter or remove their reviews. [Learn more.](#)

Search within the reviews 🔍 Sort by Yelp Sort Language English (4174)

Photo of The Cosmopolitan of Las Vegas - Las Vegas, NV, United States

See all 8594

More business info

Tue	Open 24 hours
Wed	Open 24 hours
Thu	Open 24 hours
Fri	Open 24 hours
Sat	Open 24 hours Open now
Sun	Open 24 hours

[Edit business info](#)

Accepts Credit Cards **Yes**
Accepts Apple Pay **No**
Accepts Google Pay **No**
Good for Kids **No**
Wi-Fi **Free**
Dogs Allowed **Yes**

- Nevada에 위치한 Hotel 리뷰 분석 결과
- 부정적 리뷰보다 긍정적 리뷰가 더 많았음
- 긍정적 리뷰 분석 결과
- Casino, Parking에 대한 정보가 담긴 글이 많았음
- 또한, 위치에 대한 언급이 많았으며
- 공항에서부터의 거리에 따라 긍정/부정 리뷰에 영향을 많이 끼침
- 따라서, 해당 집단과 유사한 신규 Hotel 영업 시
- Parking / Casino / Location에 대한 광고 Concept을 위주로 영업 진행
- Ex) Parking 서비스를 강화하고, Yelp 사이트 신규 등록 시 Parking에 대한 정보(홍보글)를 메인으로 등록