

외래 방문객 로밍데이터 데이터 마이닝 분석보고서

대한민국 관광 3.0, 새로운 길을 찾다

부제: 외국인 이동경로 연구를 통한 질적 도약 및 균형 발전 솔루션제안

Report22



_ 현재 관광산업의 동향 및 이슈 조사/ 01_ 무엇이 문제이고 화제일까? 앞으로의 동향은?



2014.05.20. 비전 나눔식



오픈 데이터

관광의 질 향상

Increase the quality of tour

국내 관광 육성

Upgrade the domestic tour level

세부 타겟 관광객을 겨냥한 콘텐츠 개발

New contents aimed to target tourists

관광 3.0 을 통한 창조적 관광 가치 극대화

Maximize the creative tour-value by TOUR 3.0

외래 방문객 로밍데이터로 알아보는
여행하기 좋은 나라, 대한민국!



A_ 과제 파악/

01_ 전체 체계 개괄 및
주어진 데이터에 대한 이해

B_ 분석 방향 설정

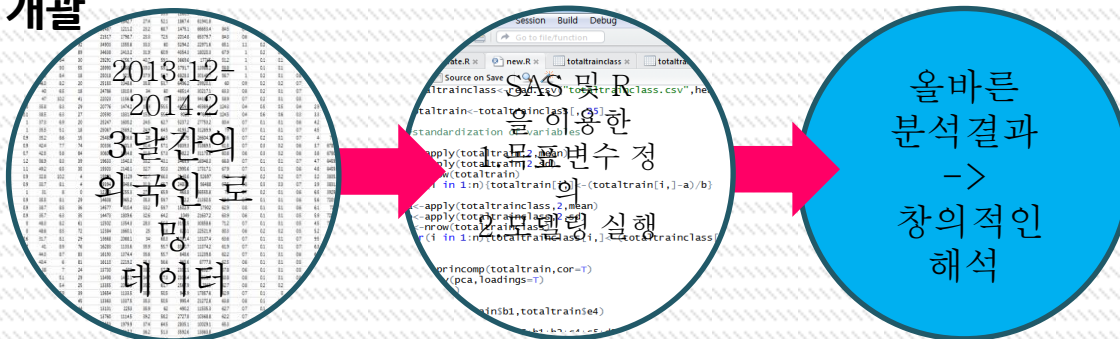
C_ 분석 시행

D_ 추가 분석

E_ 결과 해석과 관광마케팅 방안 기획 및 결론

A_ 과제 파악 > 01_ 전체 체계 개발 및 주어진 데이터에 대한 이해

- 전체 체계 개괄



Data
Information
Knowledge

데이터 마이닝 이용
정보 지식 창출, 활용
맞춤형 관광한국 소개

- 데이터 원본

roaming_sample_final.csv

id	date	time	country_cczip	latitude	longitude
12b6f1074	20140207	1248	N01	100453	37.55369
3d616fcb	20131218	1700	N02	157822	37.56119
e8ec075d	20131205	2157	N02	135741	37.51001
e501c08E	20140128	1617	N01	132030	37.66158
0d58f192	20131209	1408	N10	100391	37.55972
7be4a62d	20140207	1135	N01	690818	33.49213
521e98a8	20131218	1032	N02	110836	37.56974
937f64607	20140122	2009	N03	121854	37.55204
49751c6f	20140108	1334	N02	153801	37.47395
95746c613	20131203	1227	N02	400340	37.44706
1bf9736ea	20140211	1600	N10	435020	37.32816
185b6aa4f	20131207	1107	N01	400340	37.44705
8c260e3e2	20131230	1707	N02	618807	35.17664
8c48f1bac	20140204	1802	N02	403030	37.52097
1b31254ce	20140112	2132	N03	100809	37.56359
10bffff50d	20140211	1143	N02	137902	37.5162
52e133a6f	20131206	2303	N03	100450	37.56522
82ffc848c	20140212	1617	N02	701240	35.86721
1b7471e6f	20140213	1342	N02	477855	37.6867
0c2e1541f	20140228	1341	N01	135120	37.52529
99d6b7c04	20140110	1937	N01	100192	37.56507
131c677ec	20140205	1124	N01	415040	37.61921
875c928fa	20131224	1609	N01	750906	36.82292

- 주어진 변수들

id	각각의 고객의 암호화된 ID = 중복정보분석으로 추적가능
date	통화가 발생한 날짜기록, 년월일 구별X ex)20140221
time	로밍이 발생한 시간기록, 시분 구별X ex)1240
country_code	고객의 국적정보(10개국) ex)N01~N10
zip	우리나라 우편번호
latitude	기지국의 위도 정보
longitude	기지국의 경도 정보

기지국 정보를 종합하여
로밍이 이루어진
장소 파악 가능

Associate Rule Mining
& Sequence Pattern Mining

이용한 외국인 이동경로 연구



A_ 과제 파악

B_ 분석 방향 설정/ 01_ 세부 목표와 현대 동향과의 연관성
02_ 분석의 목표와 모델링 방법
03_ 기본 개념 및 분석 방향 구체화

C_ 분석 및 실행

D_ 추가 분석

E_ 결과 해석과 관광마케팅 방안 기획 및 결론



국적
"NATIONALITY"



외국인 관광객은 다양한 국적으로 이루어져있다.

= 국적별 여행패턴이 상이할 것임

외국인 관광객이 “주로 많이 통화하고 여행하는 장소”를 이용
“주 여행지 콘텐츠 개발”, “입국 경로, 여행 패키지 순서” 추론

> 국가별 이동경로 연구를 통해 안내문, 안내원 최적화 배치,
1차, 2차적 산업연계 집중 강화 가능, 효과적 효율적 유치전략 가능

B_ 분석 방향 설정 > 01_ 세부목표와 현재 동향과의 연관성



지역 “REGION”



수도권, 비수도권 별 “**관광객 이동경로 탐색**”을 통해
“동선 최적화”, “지역균형 발전 방법 모색”, “관광객 니즈충족”
또한 “숨은 관광객 방문 장소” 탐색, 콘텐츠 개발!

> 관광 3.0에서 제안한 빅데이터 플랫폼을 구현해 각 부처 산업 연계로 균형적인 질적 성장이 가능!
추가로 외국인 방문자들이 ‘특수한’ 목적, ‘특수’장소 방문 여부 조사





체류시간 "EXTEND OF STAY"



통화 시간 분석을 통해 “활동 시간 중 **컨텐츠 부족 시간대**” 탐색
주요 관광지 이외 “모든 관광 동선”이 “복합적 관광산업”이 되도록!

> 국가별, 시간대별 통화량 분석을 통해 비 이동 시간, 체류 장소의 콘텐츠를 강화하는 방안을 찾는다.



- 분석의 목표

외국인 관광객이 “주로 많이 통화하고 여행하는 장소”는 어디이고,
“어떤 순서로 여행”을 하는가.

> 결론적으로 어느 지역을 활성화 시키고 개발해야 하는가를 알아보자!

- 모델링 방법

연관규칙을 찾기위한

1. Association Rule Mining(장바구니 분석)
2. Sequential Pattern Mining

> 모든 나라를 포함한 모델링 VS 국적별 모델링
→ 직접 모델링을 통하여 분석한 후 결정

Association Rule Mining

> 데이터베이스에서 발견되는 연관된 규칙을 발견해 나가는 데이터마이닝의 기법

> 목적 : “관광객이 방문하는 도시에 대한 연관규칙분석”

> 시간과 방문순서를 무시하고, 순수하게 방문한 지역에만 중점을 둠

Ex) “서울, 인천을 방문하면
→ 수원을 방문한다.”

Sequential Pattern Mining

> Association Rule Mining에서 발견한 형태로,
순차적으로 방문하는 고객에 대한 분석

> 목적 : “관광객이 방문하는 도시들의 패턴”

> 하루에 방문하는 도시들 = 하나의 Item set

Ex) “서울, 인천을 방문한 관광객은
→ 다음날 수원을 방문하고
→ 그 다음날 부산을 방문한다.”

SEMMA, Data Mining,

For Tourist Pattern



A_ 과제 파악

B_ 분석 방향 설정

C_ 분석 및 실행/ 01_ RAW 데이터 처리 후 완성된 데이터
02_ SEMMA 방법론(cycle1,cycle2,cycle3,cycle4)

D_ 추가 분석

E_ 결과 해석과 관광마케팅 방안 기획 및 결론

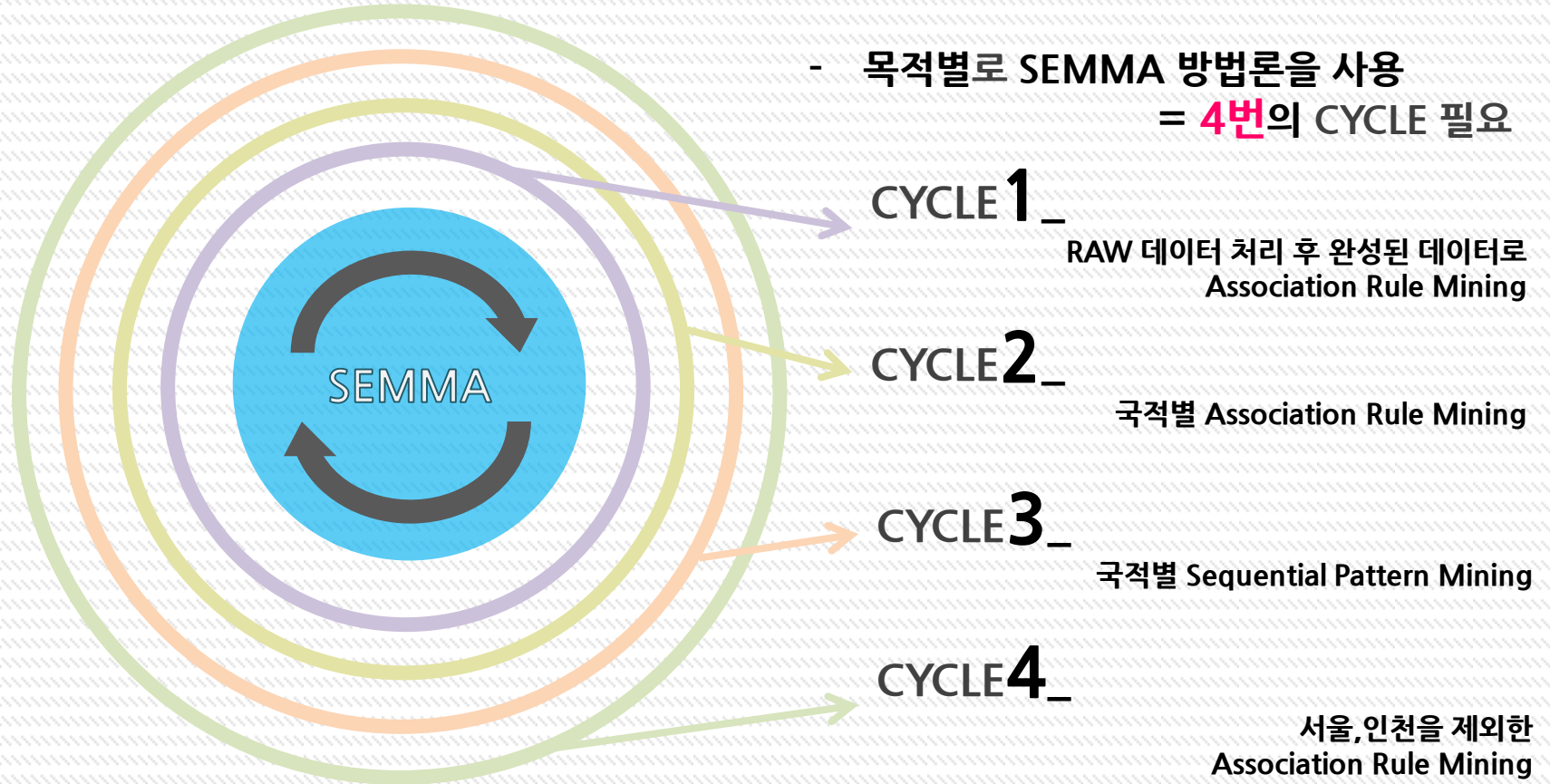
C_ 분석 및 실행 > 01_RAW 데이터 처리 후 완성된 데이터

- RAW 데이터 처리 후 완성된 데이터의 변수

id	date	time	Country_code	zip	latitude	longitude	SIDO	GUGUN	DONG	Location
각각의 고객의 암호화된 ID	통화가 발생한 날짜기록 ex)2014-02-21	로밍이 발생한 시간기록 ex)12:40	고객의 국적정보 (10개국) ex)N01~N10	우리나라 우편번호	기지국의 위도 정보	기지국의 경도 정보	기지국의 시,도 정보	기지국의 구,군 정보	기지국의 동 정보	SIDO+ GUGUN+ DONG

분석에 용이한 단위로 변환

새로 추가된 변수들



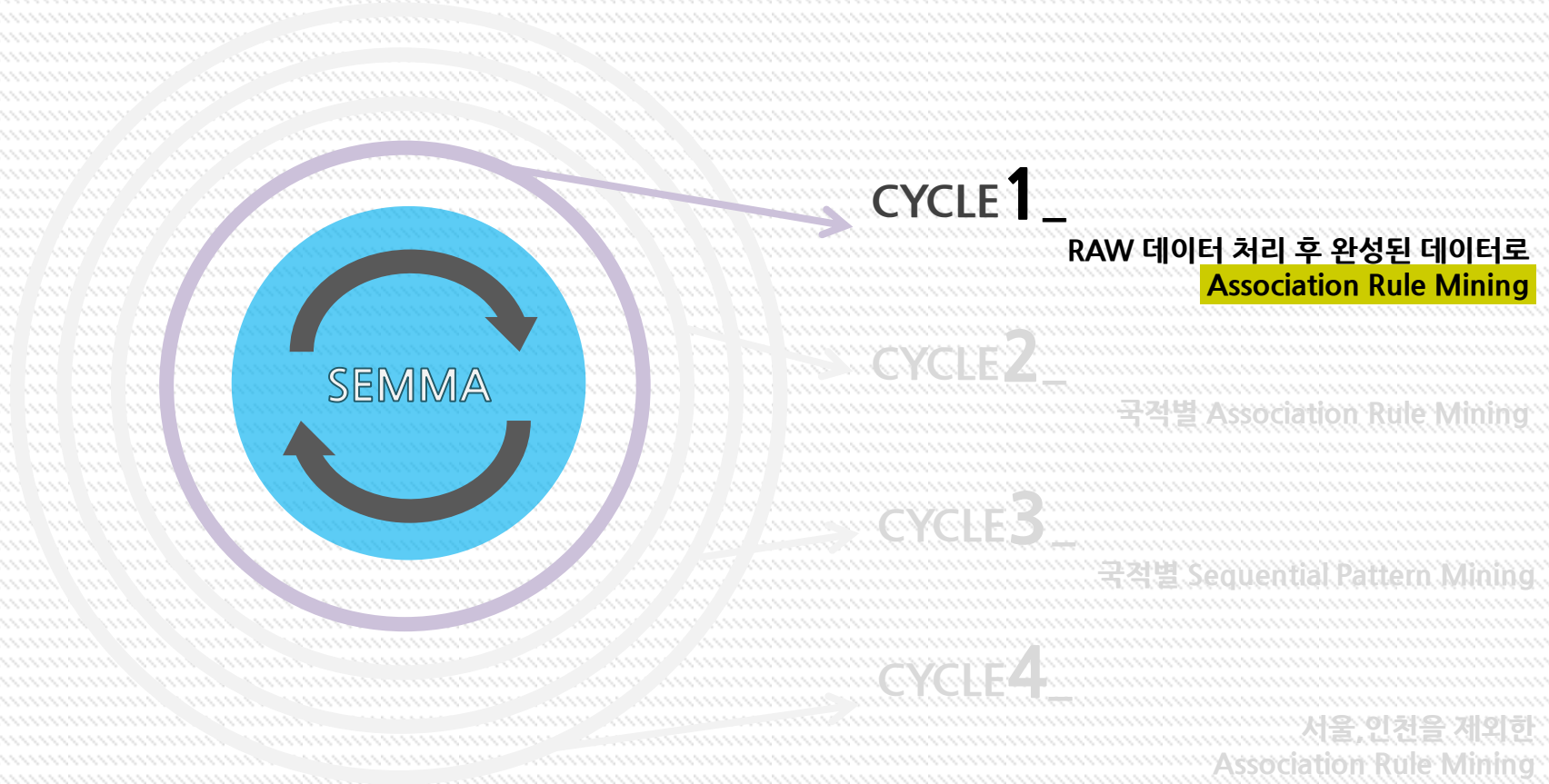
- 목적별로 SEMMA 방법론을 사용
= 4번의 CYCLE 필요

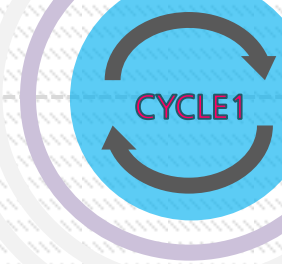
CYCLE 1_
RAW 데이터 처리 후 완성된 데이터로
Association Rule Mining

CYCLE 2_
국적별 Association Rule Mining

CYCLE 3_
국적별 Sequential Pattern Mining

CYCLE 4_
서울, 인천을 제외한
Association Rule Mining





lhs	rhs	support	confidence	lift
1 {제주서귀포시증문동}	=> {제주제주시연동}	0.01020204	0.796875	12.371361
2 {제주서귀포시증문동}	=> {N01}	0.01180236	0.921875	2.179968
3 {광주서구풍암동}	=> {N02}	0.0110022	0.9821429	3.149283
4 {대전중구옥계동}	=> {경기가평균설악면}	0.010002	0.8333333	69.430556
5 {경기가평균설악면}	=> {대전중구옥계동}	0.010002	0.8333333	69.430556
6 {대전중구옥계동}	=> {부산강서구송정동}	0.010002	0.8333333	59.511905
7 {부산강서구송정동}	=> {대전중구옥계동}	0.010002	0.7142857	59.511905
8 {대전중구옥계동}	=> {N02}	0.01160232	0.9666667	3.099658
9 {전북전주시덕진구산정동}	=> {N02}	0.01020204	0.9272727	2.97334
10 {울산동구전하동}	=> {부산강서구송정동}	0.010002	0.8196721	58.5363
11 {부산강서구송정동}	=> {울산동구전하동}	0.010002	0.7142857	58.5363
12 {울산동구전하동}	=> {N02}	0.01120224	0.9180328	2.943711
13 {경기가평균설악면}	=> {N02}	0.01160232	0.9666667	3.099658
14 {부산부산진구부전2동}	=> {N02}	0.0120024	0.7407407	2.375217
15 {서울구로구구로2동}	=> {N01}	0.010002	0.8064516	1.907025
16 {인천중구항동7가}	=> {N01}	0.01060212	0.9298246	2.198767
17 {부산강서구송정동}	=> {N02}	0.01320264	0.9428571	3.023312

<완성된 Association Rule>

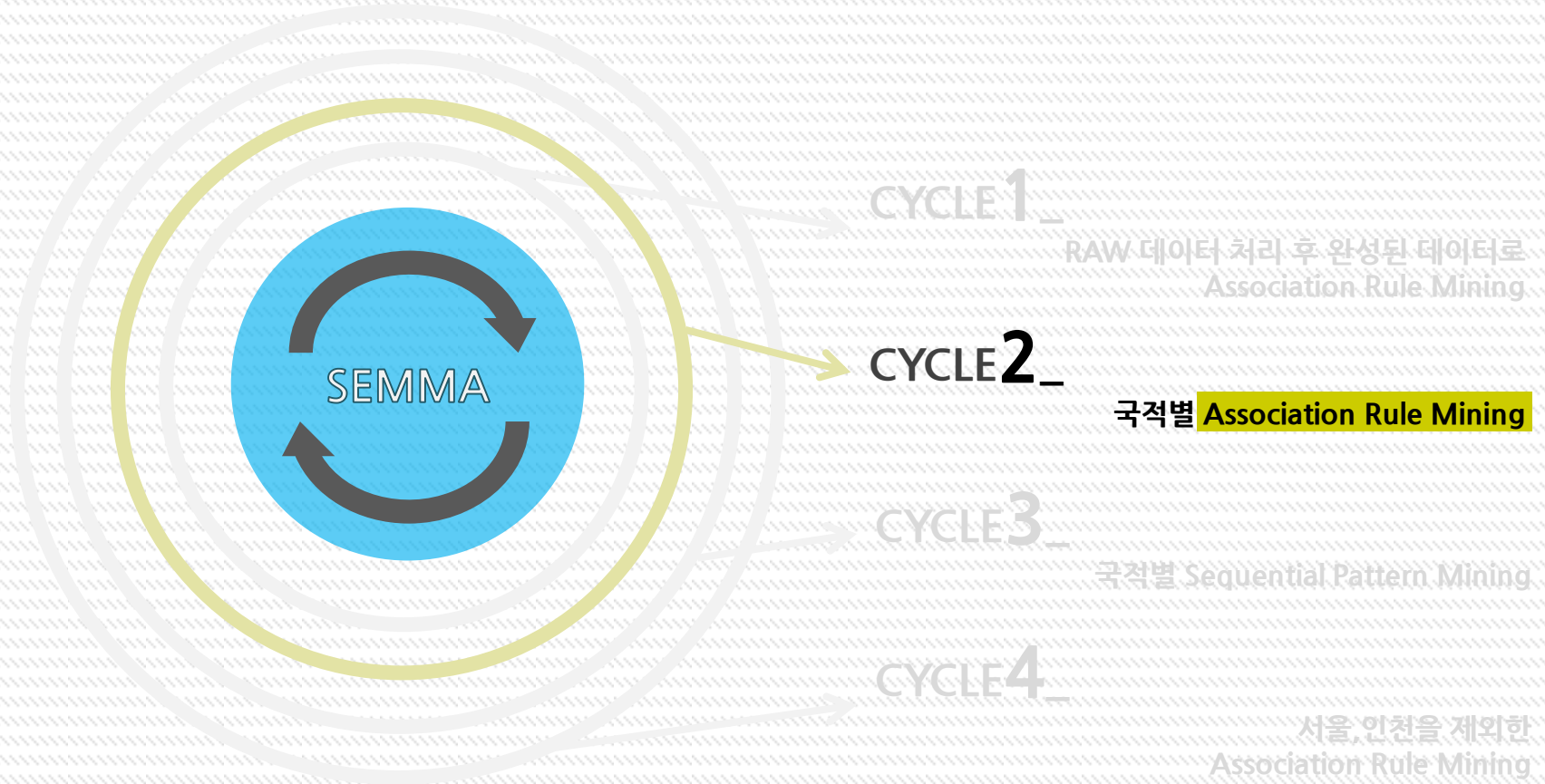
> 모두 140개의 rule이 생성, but,

Rule을 파악할 수 있는 최소 3번 이상 통화한 사람 중
N01~N10상관없이
랜덤 5000명을 표본 추출

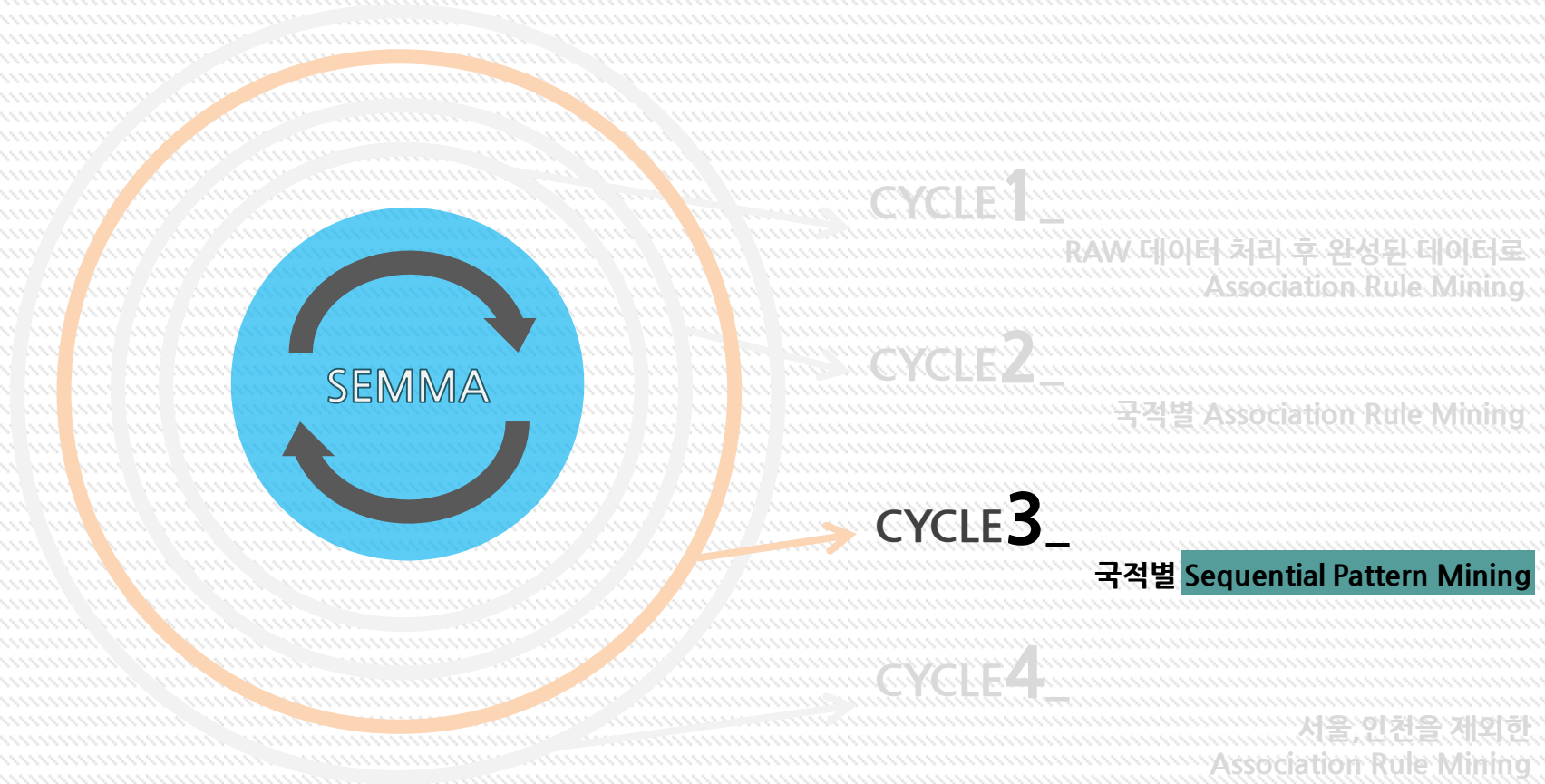
이 조건 때문에,
로밍 이용 고객수가 많은 나라에
Rule이 집중될 수 밖에 없음

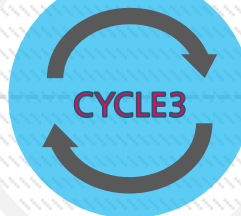
국가별 분석이 필요
(다음 SEMMA cycle2)

C_ 분석 및 실행 > 02_ SEMMA방법론



C_ 분석 및 실행 > 02_ SEMMA 방법론





sequence_name	sequence_value	support	ratio	5000
<(명동)(을지로2가)>	<(2390, 2398)>	324	0.0648	
<(명동)(충무로1가)>	<(2390)(2406)>	247	0.0494	
<(명동)(신당3동)>	<(2390)(2395)>	187	0.0374	
<(명동)(장충동1가)>	<(2390)(2402)>	182	0.0364	
<(명동)(신사동)>	<(2390)(1963)>	173	0.0346	
<(충무로1가)(인천국제공항)>	<(2406)(2638)>	169	0.0338	
<(명동)(태평로2가)>	<(2390)(2411)>	168	0.0336	
<(인천국제공항)(송인1동)>	<(2638)(2396)>	140	0.028	
<(명동)(광희동1가)>	<(2390)(2387)>	140	0.028	
<(명동)(충무로2가)>	<(2390)(2407)>	136	0.0272	
<(충무로1가)(명동)>	<(2406)(2390)>	131	0.0262	
<(인천국제공항)(광장동)>	<(2638)(2040)>	114	0.0228	
<(인천국제공항)(신사동)>	<(2638)(1963)>	114	0.0228	
<(명동)(필동2가)>	<(2390)(2412)>	110	0.022	
<(석촌동)(잠실동)>	<(2256)(2263)>	106	0.0212	
<(인천국제공항)(신당3동)>	<(2638)(2395)>	105	0.021	
<(명동)(과해동)>	<(2390)(2004)>	103	0.0206	
<(석촌동)(명동)>	<(2256)(2390)>	102	0.0204	

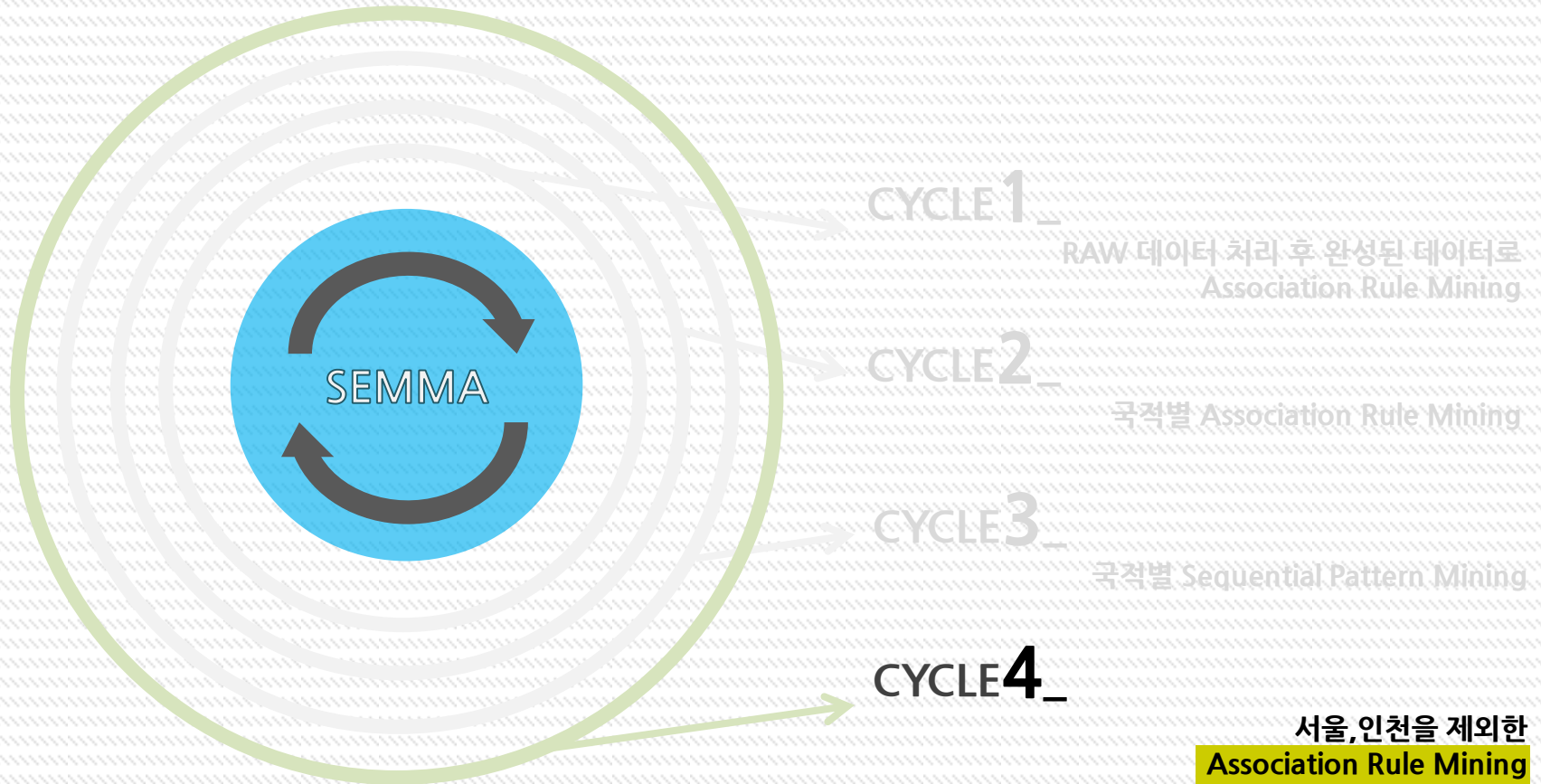
<완성된 N01의 Sequential Pattern>

> 각 국가별 rule이 생성,

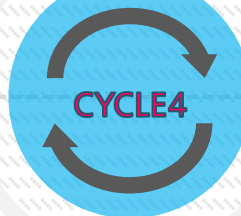
국가	생성된 Pattern 개수	국가	생성된 Pattern 개수
N01	18	N06	48
N02	113	N07	11
N03	10	N08	15
N04	20	N09	16
N05	33	N10	25

비교적 간단한 결과로 해석이 용이함. but,
 모든 방법론이 **서울에 편중되어 룰이 생성됨**
(인천공항이 압도적으로 많아 인천지역도 포함)

서울, 인천지역을 제외한 분석이 필요
 (다음 SEMMA cycle4)



서울, 인천을 제외한
Association Rule Mining



rules		support	confidence	lift
{제주서귀포시색달동}	=> {제주제주시연동}	0.010402	0.742857	5.179279
{제주제주시이도이동}	=> {제주제주시연동}	0.010002	0.675676	4.710882
{제주서귀포시하원동}	=> {제주제주시연동}	0.012002	0.722892	5.040077
{제주제주시구좌읍}	=> {제주제주시연동}	0.010202	0.708333	4.938575
{제주제주시도두일동}	=> {제주제주시연동}	0.013803	0.711134	4.959539
{제주서귀포시성산읍}	=> {제주제주시애월읍}	0.011002	0.591398	7.158348
{제주서귀포시성산읍}	=> {제주제주시연동}	0.014003	0.752688	5.247822
{제주제주시한림읍}	=> {제주제주시연동}	0.015403	0.793814	5.534558
{제주제주시이도일동}	=> {제주제주시연동}	0.021604	0.724832	5.053607
{제주서귀포시중문동}	=> {제주제주시연동}	0.027205	0.829268	5.781746
{부산부산진구부전동}	=> {부산부산진구부전2동}	0.021004	0.719178	12.57053
{대구동구서호동}	=> {광주서구풍암동}	0.020404	0.886957	24.36206
{광주서구풍암동}	=> {대구동구서호동}	0.020404	0.56044	24.36206
{대구동구서호동}	=> {대전중구옥계동}	0.020604	0.895652	24.46648
{대전중구옥계동}	=> {대구동구서호동}	0.020604	0.562842	24.46648
{대구동구서호동}	=> {전북전주시덕진구산정동}	0.020204	0.878261	24.2565
{전북전주시덕진구산정동}	=> {대구동구서호동}	0.020204	0.558011	24.2565
{대구동구서호동}	=> {경기가평균설악면}	0.021404	0.930435	23.73083
{경기가평균설악면}	=> {대구동구서호동}	0.021404	0.545918	23.73083

> 모두 742개의 rule이 생성

→ **마지막 SEMMA cycle 완성**

<완성된 Association Rule>

데이터가 무언가 말하면
분석가는 귀 기울여 듣는다.
그리고 생각한다.

“분명히 뭔가를 숨기고 있구나!”



A_ 과제 파악

B_ 분석 방향 설정

C_ 분석 시행

D_ 추가 분석/ 01_ 밤 시간대 데이터 분석

E_ 결과 해석과 관광마케팅 방안 기획 및 결론

- 밤 시간대 데이터의 프로젝트 적용

> 목적 : “밤 시간대에는 관광객들이 주로 호텔이나 게스트 하우스 등 숙박업소 주위에 머무른다는 가정하에, 어디에 사람들이 많이 모여있는지 파악”

> 여행객이 낮에 여행을 마치고 숙소 주위로 모여드는 시간을 21시 이후라고 가정하고
유흥거리나 상인들이 장사를 마치는 시간을 새벽 3시정도까지
사람들이 이동할 수 있는 시간반경이라고 가정

> 데이터 분석 프로세스





분석한 자료를
우리의 아름다운 지구에
표현해보고 싶지 않으신가요?

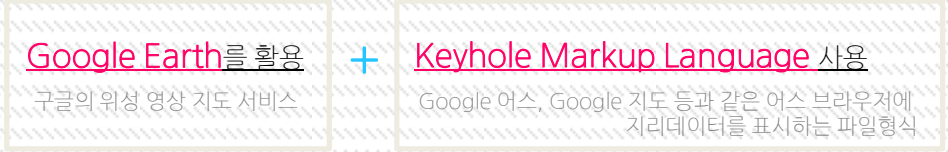
“Visual data mining is a crucial area in explorative data mining,
aimed at enhancing the effectiveness of the overall mining process,
by supporting analytical reasoning.”

P. Compieta et al,

데이터 마이닝의 시각화는 마이닝 프로세스의 이해와 분석에 큰 효과

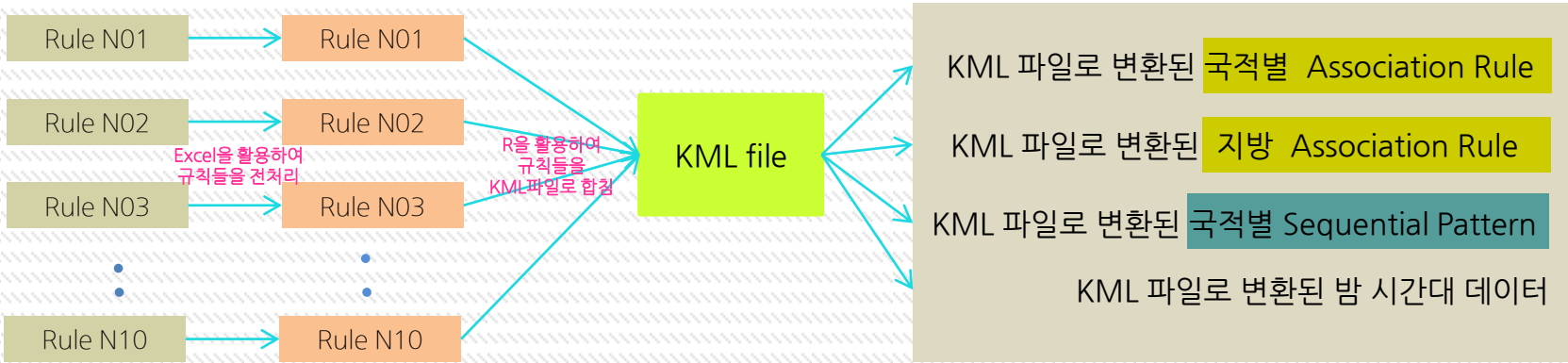


_ 데이터마이닝 결과 특수장소에 특수목적의 방문객 발견

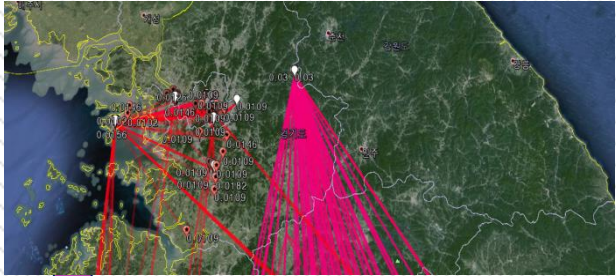


> 목적 : “앞에서 사용된 Mining 기법들을 통해 도출된 규칙들을 지도상에 표현”

> KML 파일 생성 프로세스



데이터마이닝 결과 특수장소에 특수목적의 방문객 발견



> 국가별 Association Ruled에서 N02,N05 나라 관광객들의 경기도 가평군 설악면 특정지역에 몰려있다.

→ 일반사람이 알고 있는 관광지역이 아님에도 불구하고 밀집되어 있는 rule을 보고 더 자세히 검색해 봄

그 결과 검색된 장소는 바로,

가평평화월드센터

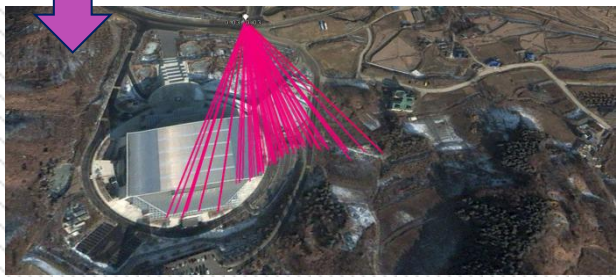


통일교의 성지



> 그러면 항상 우리나라에는 통일교를 믿는 사람들이 성지순례 및 여행을 하러 오는 것일까??

→ 2013년12월 부터 2014년 2월 사이의 인터넷 뉴스를 검색



통일교 2500쌍 합동 결혼식
 (가평=뉴스1)
 최종수정 2014.02.12 19:35:31

세계평화통일가정연합이 12일 오전 경기도 가평 청심평화월드센터에서 천일국 기원절 1주년 기념식 및 2014 천지인참부모 천주 축복식을 진행했다.

천주축복식은 국제합동결혼식으로 현장에 참석하는 2500쌍을 비롯해 인터넷 생중계를 통해 50여개국에서 2만쌍이 결혼식을 올렸다. (세계일보 제공) 2014.02.12/뉴스1

<저작권자 © 뉴스1코리아, 무단전재 및 재배포 금지>

2014년 2월 19일, 전 세계에서 2500쌍이 합동결혼식을 하러 온 것으로 발견,

↓
이 지역에 데이터가 밀집 될 수 밖에 없었음

앞으로의 연구에는 시기별로 분할하여 연구 해야 하는 당위성 발견



news 1

새는 알에서 태어난다_탄생
정보는 데이터에서 태어난다_발생
그리고 지식으로 성장한다_발전



A_ 과제 파악

B_ 분석 방향 설정

C_ 분석 시행

- 01_ SEMMA CYCLE2 해석
- 02_ SEMMA CYCLE3 해석
- 03_ 위성지도를 이용한 빅 데이터 플랫폼 구현
- 04_ SEMMA CYCLE4 해석

D_ 추가 분석

- 05_ 지역관광, 강점에 테마를 더해 체류형 관광을 이끌어낸다.
- 06_ 밤 시간대 데이터 분석
- 07_ 관광객의 발이 머문 시간, 잠들기 직전까지 대한민국에 반하다.

08_ 대한민국 관광 3.0, 새로운 길을 찾다!

E_ 결과 해석과

관광마케팅 방안 기획 및 결론/



CYCLE2_국적별 Association Rule Mining

> 서울 내에서의 규칙이 자주 발견됨

N01 N02 N03 N04 N06 N08 N10

> 대구, 대전, 전주, 가평 등 지방 단체 패키지 여행으로 추정되는 규칙이 다수 존재함

→ 인천공항이 아닌 부산항이나 부산공항등으로 배를 타고 들어오는 관광객이 생각보다 많을 수 있음

N02 N03 N05 N07 N09 N10

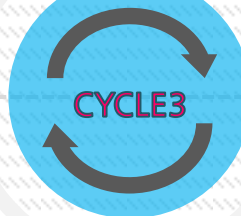
> 부산항에서 입국한 관광객 수가 많을 것으로 추정

N02 N03 N07

> 수도권과 제주도간의 왕래가 자주 관찰됨

N05

Association Rule		Support	Confidence	Life
{제주제주시한림읍}	=> {제주제주시연동}	0.010575	0.726027397	5.859661
{제주서귀포시송문동}	=> {제주제주시연동}	0.015762	0.766990291	6.190266
{제주제주시이도일동}	=> {제주제주시연동}	0.014166	0.702970297	5.67357
{인천서구경서동}	=> {인천중구운서동}	0.011373	0.655172414	1.450408
{서울서초구반포4동}	=> {인천중구운서동}	0.010375	0.634146341	1.403861
{서울용산구후암동}	=> {서울중구명동2가}	0.011173	0.549019608	1.700671
{서울서대문구창천동}	=> {서울중구명동2가}	0.013368	0.609090909	1.886751
{서울서대문구창천동}	=> {인천중구운서동}	0.01257	0.572727273	1.267893
{서울중구우교동}	=> {서울중구명동2가}	0.013966	0.569105691	1.762891
{서울중구우교동}	=> {인천중구운서동}	0.01257	0.512195122	1.133888
{제주제주시건일동}	=> {제주제주시연동}	0.022346	0.670658683	5.412788
{서울종로구종로5가}	=> {인천중구운서동}	0.011772	0.526785714	1.166188
{서울강남구역삼1동}	=> {인천중구운서동}	0.014765	0.637931034	1.41224
{서울강남구논현동}	=> {인천중구운서동}	0.013368	0.540322581	1.196156
{서울금천구가산동}	=> {인천중구운서동}	0.013567	0.561983471	1.244108
{서울마포구합정동}	=> {인천중구운서동}	0.011373	0.527777778	1.168384
{서울용산구한남동}	=> {인천중구운서동}	0.011971	0.5	1.10689
{서울영등포구대림3동}	=> {인천중구운서동}	0.013767	0.575	1.272924
{서울종로구수송동}	=> {서울중구명동2가}	0.013767	0.507352941	1.571603
{제주제주시용담이동}	=> {제주제주시연동}	0.026736	0.666666667	5.380569
{인천중구운북동}	=> {인천중구운서동}	0.020152	0.696551724	1.542013
{서울강남구청담동}	=> {인천중구운서동}	0.018955	0.582822086	1.29024
{제주제주시노형동}	=> {제주제주시연동}	0.034717	0.719008264	5.80301
{서울마포구서교동}	=> {서울중구명동2가}	0.01676	0.506024096	1.567486
{서울중구남대문로5가}	=> {서울중구명동2가}	0.019354	0.5	1.548826
{서울중구남대문로5가}	=> {인천중구운서동}	0.021947	0.567010309	1.255237
{서울용산구동자동}	=> {인천중구운서동}	0.018156	0.579617834	1.283147



CYCLE3_국적별 Sequential Pattern Mining

Sequence_name	Ratio
<(명동)(을지로2가)>	0.0648
<(명동)(충무로1가)>	0.0494
<(명동)(신당3동)>	0.0374
<(명동)(정충동1가)>	0.0364
<(명동)(신사동)>	0.0346
<(충무로1가)(인천국제공항)>	0.0338
<(명동)(태평로2가)>	0.0336
<(인천국제공항)(송인1동)>	0.028
<(명동)(광희동1가)>	0.028
<(명동)(충무로2가)>	0.0272
<(충무로1가)(명동)>	0.0262
<(인천국제공항)(광장동)>	0.0228
<(인천국제공항)(신사동)>	0.0228
<(명동)(필동2가)>	0.022
<(석촌동)(잠실동)>	0.023002
<(인천국제공항)(신당3동)>	0.021852
<(명동)(과해동)>	0.021277
<(석촌동)(명동)>	0.020702

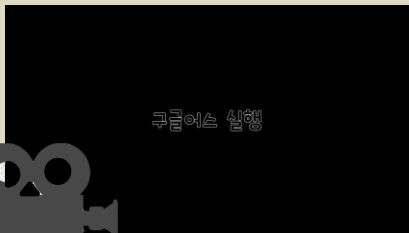
- > N01은 명동일대 (명동/을지로2가)를 시작으로 하는 쇼핑을 포함한 자유여행일 확률이 높음
- > N02,03는 김포공항으로 다수가 들어오고 동대문디자인플라자 부근 쇼핑 할 확률 높음
- > N04는 강북보다는 강남이 다수/ 반포 4동 = 신세계 면세점 백화점 및 고속 터미널 위치 이곳에서 지방으로 이동하는 여행객이 있을 가능성 있음
- > N05는 상일동 선사유적지 등 서울 중심지 외 역사적 중요 지역 방문 가능성 있음
- > N06은 이태원거리지역을 필수 코스로 잡고 있음
- > N07은 대관령과 롯데월드가 비율상 두드러짐
- > N08은 적은 sequence 중에서도 역삼동이 자주나오는 것으로 보아 테헤란로 : 업무목적일 수 있음
- > N09는 동교동(홍대), 대전 용전동, 당주동(광화문, 청계광장, 이순신동상, 대한민국역사박물관) 다수
- > N10은 대관령, 남대문 발견

N01

<완성된 N01의 Sequential Pattern>

E_ 관광마케팅 방안 기획 > 03_ 위성지도를 이용한 빅 데이터 플랫폼 구현

<구글어스 실행 동영상>



(클릭하시면 동영상 재생됩니다)

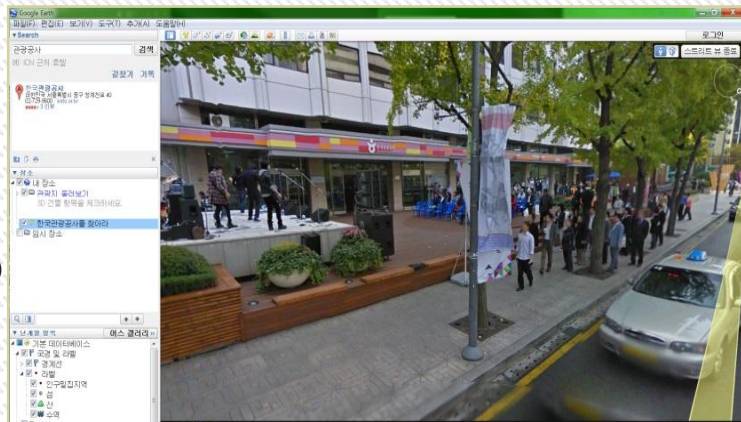
1_ 국가별, 목적별, 지역별 적합장소 추천

> 추천시스템 (item, user matching etc.)

2_ 외국인 관광객을 위한 다양한 언어 지원

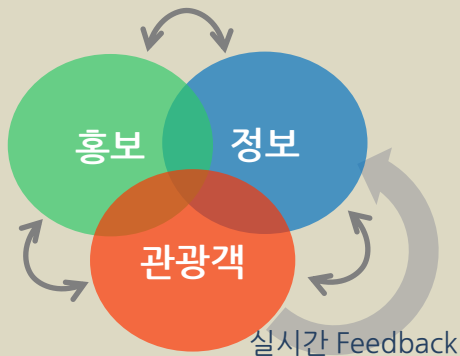
> 외국인 및 외지인의 길 찾기 편의 제공(스트리트 뷰)

3_ 실시간 관광객들의 정보, 의견 공유 가능



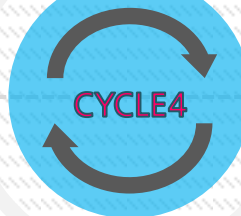
<스트리트 뷰 예시>

기대효과

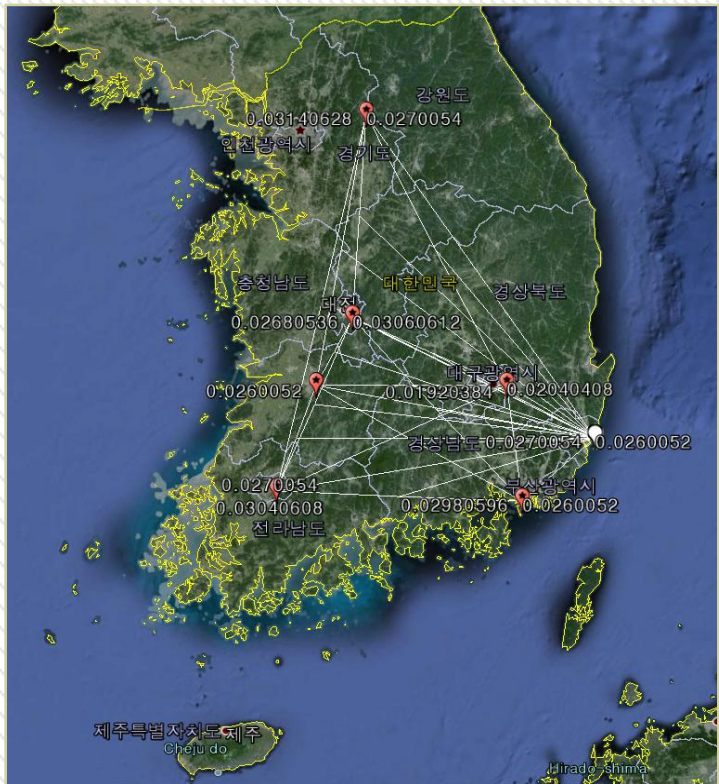


관광산업연계 시스템의 원활한 가동

각 부처간 의사소통의 원활 → “질적” 성장



CYCLE4_서울,인천을 제외한 Association Rule Mining



- 서울, 인천 제외 (출입국전 인천공항에서 통화를 하는 인원이 압도적으로 많음)

불균형 데이터를 고려하지 않음으로써
지역 주요 이동경로 확인

- 다수 빈도 도시 탐색

- > 대전, 광주, 부산, 전주, 가평, 울산, 대구
- > 서로 연관이 있는 도시들은 주요 광역시 + 전주
(한옥마을 등 관광지) + 가평(남이섬, 자라섬 캠핑장)
- > 주요여행사, 안내문, 정보Site에 편중된 정보
= TOP 5 여행사 검색 결과 유사 package 다수 발견
- > 인구밀도가 높을 수록, 도시일수록 인프라 발전
= 특히 숙박시설 주변 발전된 시가지 선호



HanaTour USA
Round Korea 7days

E_ 관광마케팅 방안 기획 > 05_ 지역관광, 강점에 테마를 더해 체류형 관광을 이끌어내다.

1시간이면 명소 어디든 여행할 수 있는 **서울**
뛰어난 **교통, 숙박**인프라, 대한민국의 **심장**의 기능

지역마다 **고유의 매력**이 있다! **체류형** 관광상품 개발로
1박 이상 머물고 싶은 **지역관광**

문제점

- 중소 도시 비방문
혹은 일시적 방문
- 체류시간 부족
근거 - Rule mining 결과
지역 내 이동 rule 회귀
관광수입으로 직결
- 콘텐츠의 한계
인프라 약점 보충 부족

자치시도 시행 현황

- 테마 강화
힐링, 의료, 쇼핑, 교육 등
- 체류형 관광상품개발
예: 담양 시 '700만 관광도시'
1박2일 스테디 투어, 팸 투어
- 한국관광클럽 MOU

제안

1. 연구결과 나온 **동선 이하**
내실이 튼튼한 관광중소도시 중점
매력적 테마적용 - 연계코스제작
 - 경주(문화재 + 전통음식 : 외래객 만족2위 요소)
 - 금산(인삼 + 한의학, 힐링 : 침, 의료, 안마 코스)
 - 다도(多島)지역(휴양 + 해양체험 : 바다낚시 등)
2. **게스트 하우스, 유스호스텔 확대**
 - 개인여행객 증가, 힐링 여행 증가
 - 관광공사공식 안내 매뉴얼, 홈페이지 링크배포 :
전국 공통 서식으로 외국인 접근성 향상



- 밤 시간대 데이터 분석과 발전방향

SIDO	GUGUN	DONG	COUNT_of_date
서울	중구	신당동	10351
인천	제주시	운서동	8631
제주	제주시	연동	8456
서울	중구	명동2가	6552
서울	강남구	삼성1동	5210
서울	광진구	광장동	4021
서울	중구	을지로2가	3757
제주	제주시	건입동	3750
서울	중구	태평로2가	3593
서울	중구	광희동1가	3373
서울	중구	중무로2가	3345
서울	강남구	논현동	3006
서울	강남구	신사동	2939
서울	영등포구	여의도동	2392
서울	강남구	역삼동	2337
서울	용산구	한남동	2335
서울	중구	남대문로5가	2238
서울	강남구	역삼1동	2201
서울	강남구	청담동	2097
서울	중구	중무로1가	2080
서울	종로구	창신1동	2048
경기	수원시 팔	인계동	1969
서울	강서구	과해동	1958
서울	마포구	동교동	1708
서울	서초구	잠원동	1668
서울	중구	필동2가	1621
서울	중구	장충동1가	1428
서울	종로구	서린동	1421
부산	부산진구	부전2동	1325
강원	평창군	대관령면	1301
서울	서초구	반포4동	1286
제주	제주시	이도일동	1283
서울	용산구	동빙고동	1200
서울	강남구	역삼2동	1198
서울	중구	무교동	1173
제주	제주시	애월읍	1101
서울	종로구	낙원동	1091

> 공항주변과 서울 제일 번화가 명동과 동대문디자인플라자, 홍대, 강남역주변과 지방번화가를 제외하고, 관광객들이 밤에 주로 머무르는 장소이지만 주변에 편의시설 및 유흥시설이 위치하고 있지 않은 곳의 활성화 및 콘텐츠 개발이 필요함

W호텔 위치 - 호텔 말고 주변엔 아무것도 없음

주거 지역이므로 대규모 개발보다는 작은 단위의 콘텐츠 개발 필요

제주 항 부근 - 모든 곳으로의 접근성이 뛰어나지만 밤 시간대 편의 시설 및 유흥시설 부족

한강공원을 중심으로 입구 주변에 유흥이나 편의 시설 확충 필요

국회의사당, 63빌딩 및 한강공원, 무역단지 - 낮 시간대에는 유동인구가 꽤 되지만 밤이 되면 조용해지는 지역

이태원거리와는 별개로 대사관이 밀집되어 있다는 특성을 살려 새로운 대사관거리 조성

대사관밀집지역 - 이태원거리와 거리상으로는 인접하지만 길의 경사가 쎈

자연과 밀접하게 연결되어 있는 지역, 단순히 밤거리 개발보다는 자연 속에서 밤에만 할 수 있는 체험적 콘텐츠 개발 필요

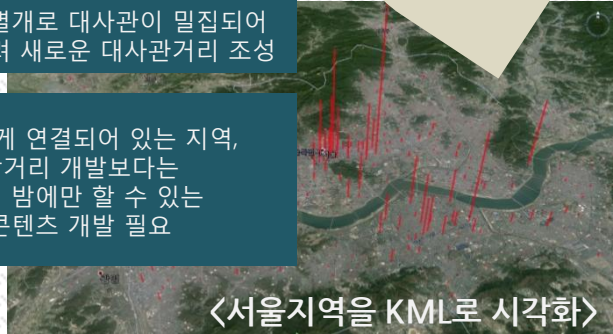
용평스키장 주변 - 12월부터 2월 사이의 데이터라 스키장데이터 꽤 많이 나옴. 그러나, 스키장 외 즐길 거리가 하나도 없음,

왔다가가는 관광지 - 관광지 외 즐길 거리가 하나도 없음, 자동차가 필요

빨간색 막대 = 관광객이 21시부터 3시까지 머무르는 장소의 분포

<내림차순 정리한 데이터>

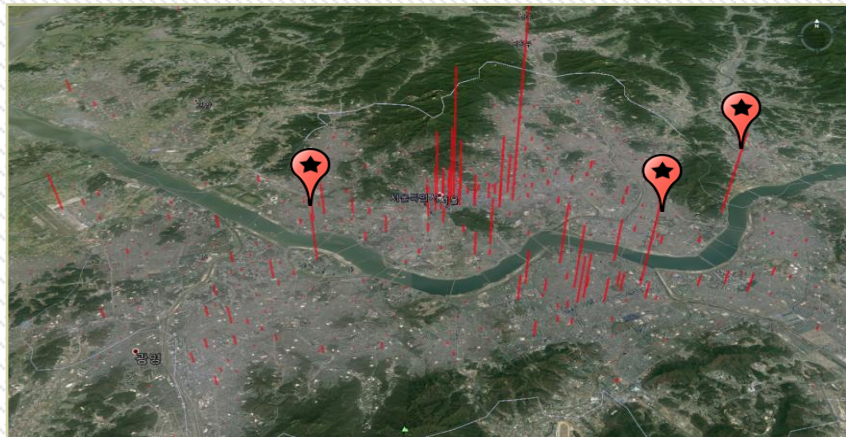
<서울지역을 KML로 시각화>



E_ 관광마케팅 방안 기획 > 07_ 관광객의 발이 머문 시간, 잠들기 직전까지 대한민국에 반하다.

1. 수도권 숙박시설, 체류장소 부근 인프라 확충

- > 수도권 밤 시간대 데이터분석결과: 인프라 부족 장소 발견
- > 우측 지도 표시 지역 : 여의도, 삼성동, 광진구(광장동)
(좌측부터)
- > 주점, 음식점, 노래방 등 시설 확충, 주요관광번화가 셔틀버스확대

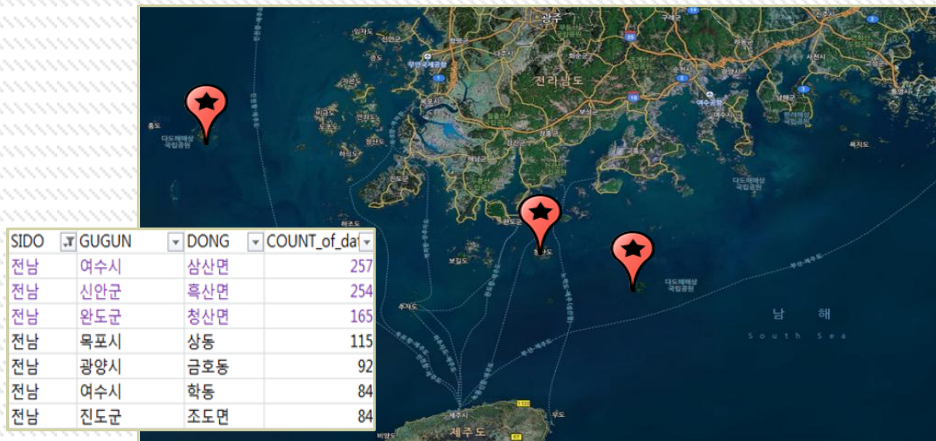


<21:00~03:00 수도권 내 통화량 분포>

관광객 **활동시간** 증가, **신규 시장** 확보

2. 전라남도 섬 관광. 외래관광객의 마음을 훔치다.

- > 비 수도권 밤 시간대 데이터분석결과 : 전남 섬 부근 다수 숙박
- > 좌측 지도 표시 지역 : 흑산도(신안), 청산도(완도), 거문도(여수)
(좌측부터)
- > 섬 주변 크루즈 여행상품 추가 개발 및 섬 관광, 숙박 Package 확충



고부가 관광상품 개발과 **요우커(遊客)** 유치기대

위성지도를 이용한 빅데이터 플랫폼 구현

지역관광에서 체류형 관광을 이끈다

관광객의 발이 머문 시간, 밤시간 관광활용

대한민국 관광3.0, 새로운 길을 찾다!

외래 방문객 로밍데이터 데이터 마이닝 분석보고서

감사합니다!



일신우일신(日新又日新)
관광산업 선순환 생태계를 향해!



A_ 과제 파악

B_ 현재 관광산업의 동향 및 이슈 조사

01_ 무엇이 문제이고 화제일까? 앞으로의 동향은?

C_ 분석 방향 설정

D_ 분석 시행

E_ 추가 분석

F_ 결과 해석

G_ 관광마케팅 방안 기획 및 결론

H_ 참고 문헌

B_ 현재 관광산업의 동향 및 이슈조사 > 01_ 무엇이 화제이고 문제일까? 앞으로의 동향은?

관광의 질 향상

Increase the quality of tour



한국관광공사
KOREA TOURISM ORGANIZATION

2014.05.20. 비전 나눔식

> “고부가 관광상품 개발” : 크루즈 관광, 고급 쇼핑 코스 개발 등

> “여성전용 관광상품 출시” : 명품과 프리미엄 백화점 아울렛, 미식, 쇼핑, 의료, 숙박, 미용 연계

Ex) 매해 4월 한국관광공사, 강원도, 용평 리조트 공동 눈썰매 대회 개최

- April Snow Festival, 한국의 눈과 꽃을 동경하는 동남아 수요 겨냥
- 태국 등 동남아관광객 1천여 명 참가(2013), 500여 명(2014)
- 스키장 ‘잔설(殘雪)’을 활용하여 고부가 관광상품 개시



© 한국관광공사

국내 관광 육성

Upgrade the domestic tour level

> 지역 관광 활성화

- 통계상 서울을 비롯한 수도권, 제주, 부산과 이외 지역과의 관광산업 불균형
- 10대 실천 과제 중 1 : “촌장님에게 든든한 관광공사”

Ex) 생태관광, 공정여행, 농어촌마을 체험, 근로자 휴가 지원사업 및 관광소외계층 여행기회 확대, 관광두레 등

- 창조관광기업과 과련, 발굴 및 육성체계 전면 개편

> 국내 관광을 내수에서부터 튼튼히

- 효과적 성공 사례 : 내일로(Korail, 일정기간 전국 기차여행 장려제도)

세부 타겟 관광객을 겨냥한 콘텐츠 개발

New contents aimed to target tourists

- > 한류(韓流) : 드라마, 영화, 가요 등 아시아를 중심으로 대한민국의 대중 문화의 인기
- > 요우커(遊客) : 중국인 관광객의 파워가 커짐에 따라 전략적인 접근이 필수적

Trend 과거 드라마, 영화 촬영지가 관광객의 대부분을 이루었다면 최근 국내 가수, 연기자의 공연이나 팬 미팅 등의 목적으로 방문하는 관광객이 증가하는 추세이다.

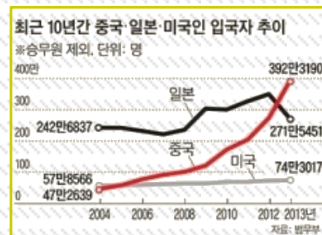
- > 요우커(遊客)에 증가에 따른 유치전략

인터컨티넨탈 서울 코엑스 호텔

- : 2011, 2012, 2013년에 해당 10%, 22%, 29%씩 요우커 수가 증가함에 따라
 1. 콘지(죽), 담삼, 티에그(계란 찻물로 삶는 것) 등 중국인 선호 메뉴 추가
 2. 가격 8만원, 8000원 등 중국인이 좋아하는 숫자 '8'이 들어가게 조정
 3. 중국인 고객 대상 세일즈 매니저, 웨딩 리허설 서비스 운영

롯데 면세점

- : 2012년 대비 2013년 중국인 매출 비중 70% 증가
 1. 매장 각 층에 중국어 통역원 5명씩 배치
 2. 중국인 인기 물품 BB크림존 4배로 확대
 3. 슈퍼주니어, 2PM 등 한류스타 모델 현지 홍보



- 한국 관광 상권에 영향력이 급증함에 따라 주 방문 지, 쇼핑 지 중국어 간판, 안내문 기본
- 명동, 남대문 등 주 방문 지에 있는 음식점, 호텔, 백화점의 관광객 유치 전략 변화

Ex) 쉬어 쩌개 : 이카나베(일본) -> 직화구이(중국)

- > 지방까지 요우커(遊客) 방문 확대

2014년도 7월~10월 요우커 대거 방문 예상

- ~4월 대구 방문 중화권 관광객 수 25,729명 - 지난해 대비 45.4% 증가
- 7월 치맥 페스티벌 연계 중국 관광객 200여명
- 7~8월 중국 청소년 수학 여행단 2,000여명
- 7~10월 중국 하얼빈, 닝보 등지 직항 전세계로 관광객 1만 여명
 1. 드라마 '별에서 온 그대'의 인기영향
 2. 중국 전세기의 본격 유행
 3. 대구시 동성로 상점가 중국 은행카드 가맹점 확대, 음식점 발굴 전세버스 주, 청차 공간 마련, 기본 회화 책자 제공 등

관광 3.0 을 통한 창조적 관광 가치 극대화

Maximize the creative tour-value by TOUR 3.0

추가 보충 자료

> 브리콜레르(Bricoleur)적 사고와 편집학(Editology)으로 접근하여 관광 3.0 개념 구성

> 정부 3.0 및 창조 경제가치 활성화

- 관광 빅 데이터 플랫폼 접근성 개선 및 활용 촉진
- ICT 기반의 관광객 활동 행태 피드백, 개선
- 관광 벤처 육성을 통한 일자리 창출 지원

> 지역 관광 활성화 모델 구축

- 지역사회 주도 하의 지역 관광 활성화
- 지역의 '일상'을 관광상품 및 콘텐츠로 기획, 개발
- 관광과 지역 주민이 상호보완 관계 형성

> 관광산업 융, 복합화

- 관광을 매개로 산업간 융, 복합화 추구
- 신규 상품 및 서비스 개발, 신규 시장 계획
- 산업간 융, 복합을 통한 새로운 부가가치 창출

> 핵심 가치



	공항						항구				
	계	인천	김해	김포	제주	기타	소계	부산	인천	기타	소계
중국	276,234	141,416	8,250	13,739	39,347	8,271	211,023	2,748	47,047	15,416	65,211
일본	215,050	104,802	27,455	68,854	4,726	12	205,849	8,965	17	219	9,201
미국	52,852	46,570	2,503	1,769	177	975	51,994	562	53	243	858
대만	43,526	33,845	3,545	4,452	1,283	6	43,131	162	52	181	395
홍콩	38,677	35,232	2,563	258	566	11	38,630	11	35	1	47
태국	37,679	35,418	1,162	107	547	111	37,345	174	24	136	334
싱가포르	32,725	31,026	601	266	769	7	32,669	26	6	24	56
말레이시아	29,985	26,444	3,056	97	259	2	29,858	43	19	65	127
필리핀	25,925	8,777	1,572	60	1,172	5	11,586	4,612	2,304	7,423	14,339

> 2013년도 12월 외래객 입국교통수단 통계자료

> 공항, 항구 이용 빈도, 주 목적지, 전체 관광객 수를 기준으로 N## 국가 탐색 가능

> 데이터 마이닝 결과와 비교하여 전략적 접근 가능

SEMMA, Data Mining,

For Tourist Pattern



A_ 과제 파악

B_ 현재 관광산업의 동향 및 이슈 조사

C_ 분석 방향 설정

D_ 분석 및 실행/ 01_ RAW 데이터 처리
02_ Explore
03_ SEMMA 방법론(cycle1,cycle2,cycle3,cycle4)

E_ 추가 분석

F_ 결과 해석

G_ 관광마케팅 방안 기획 및 결론

H_ 참고 문헌

D_ 분석 및 실행 > 01_RAW 데이터 처리

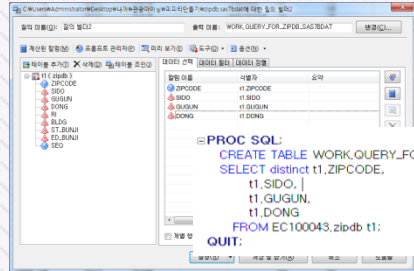
- 우편번호 DB를 이용하여 데이터에 행정구역 추가

ZIPCODE	SIDO	GUGUN	DONG	RI	BLDG	ST_BUNJI	ED_BUNJI
1	13006	서울	영등포구	영등포동	영등포로		
2	13007	서울	영등포구	영등포동	우암동1가	1	5
3	13006	서울	영등포구	영등포동	우암동1가	901	902
4	13070	서울	영등포구	영등포동	주방1가	1	16
5	13066	서울	영등포구	영등포동	주방1가	41	40
6	13007	서울	영등포구	영등포동	주방1가	66	65
7	13006	서울	영등포구	영등포동	주방1가	194	193
8	13007	서울	영등포구	영등포동	주방1가	194	195
9	13005	서울	영등포구	영등포동	주방1가	194	196
10	13005	서울	영등포구	영등포동	주방1가	194	197
11	13006	서울	영등포구	영등포동	주방1가	194	198
12	13007	서울	영등포구	영등포동	주방1가	194	199
13	13010	서울	영등포구	영등포동	주방1가	194	200
14	13041	서울	영등포구	영등포동	우암동1가	801	803
15	13000	서울	영등포구	영등포동	우암동1가	801	802
16	13071	서울	영등포구	영등포동	우암동1가	801	804
17	13000	서울	영등포구	영등포동	우암동1가	801	805

우편번호 DB

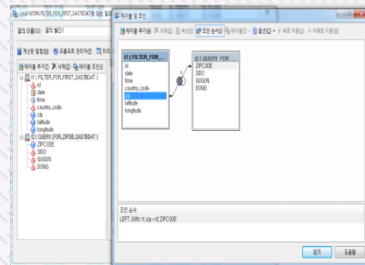
- > 행정구역은 "동"까지만 사용
- > 출처 : www.ZipFinder.co.kr

- 우편번호 DB + 기존 데이터 통합 과정_1



- > 같은 우편번호에 다른 주소가 있을 수 있으므로 중복을 제거 → (distinct)
- > 같은동, 다른건물 = 같은 우편번호로 포함

- 우편번호 DB + 기존 데이터 통합 과정_2



- > 우편번호DB에 없는 ZIP_code가 존재 → 조인속성을 LEFT JOIN으로 지정
- > 조인속성 처리 후, 우편번호DB에 없는 ZIP_code는 직접 검색하여 입력

- 우편번호 DB + 기존 데이터 통합 과정_3

id	date	time	country_code	zip	latitude	longitude	SIDO	GUGUN	DONG
2e4f8e88a0c...	2013-12-07	09:31	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
25610c54320...	2014-05-13	12:01	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
a5e36e8236c...	2014-02-19	12:30	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
8f8c a88aa8d...	2014-01-27	17:43	NO2	10001	37.56075	126.9821972	서울	영등포구	영등포동
529296c53681...	2014-01-08	23:41	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
284c-d67c600...	2014-02-09	11:14	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
1d898f84c23ad...	2014-03-05	10:04	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
4565d7681f4...	2014-02-05	11:14	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
e1201301b11...	2013-10-13	20:48	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
0a48c11a0b5a...	2014-01-01	04:42	NO2	10001	37.56075	126.9821972	서울	영등포구	영등포동
201ba6418a3...	2014-05-05	04:42	NO2	10001	37.56075	126.9821972	서울	영등포구	영등포동
273c55a50a11...	2014-01-29	00:07	NO2	10001	37.56075	126.9821972	서울	영등포구	영등포동
2861fa9a601...	2013-12-26	08:51	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
6656c2c465a...	2014-01-13	20:48	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
c33644e4699c...	2014-02-17	12:23	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
e7e421ba4a5...	2013-12-04	22:23	NO3	10001	37.56075	126.9821972	서울	영등포구	영등포동
1786c4e1343...	2014-02-24	17:55	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
6656c2c465a...	2014-01-11	15:44	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
9f76764167c...	2014-09-08	17:03	NO1	10001	37.56075	126.9821972	서울	영등포구	영등포동
9a33394463a...	2013-12-13	00:07	NO2	10001	37.56075	126.9821972	서울	영등포구	영등포동

우편번호 DB

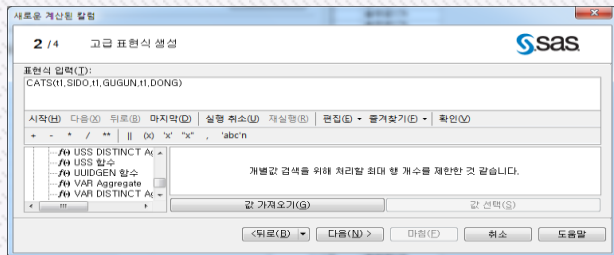


기존 데이터

- > 행정구역별로 칼럼이 나뉘면 → 잠바구니 분석에 적합하지 못함
- > 따라서, (시도 + 구군 + 동)을 합친 새로운 칼럼이 필요

D_ 분석 및 실행 > 01_RAW 데이터 처리

- 3개의 행정구역 칼럼 하나로 통일



- > 개별 값을 합치고, 선후 행 공백을 제거하는CATS함수를 사용
- > 그래도 제거되지 않은 공백은 엑셀을 통해서 제거

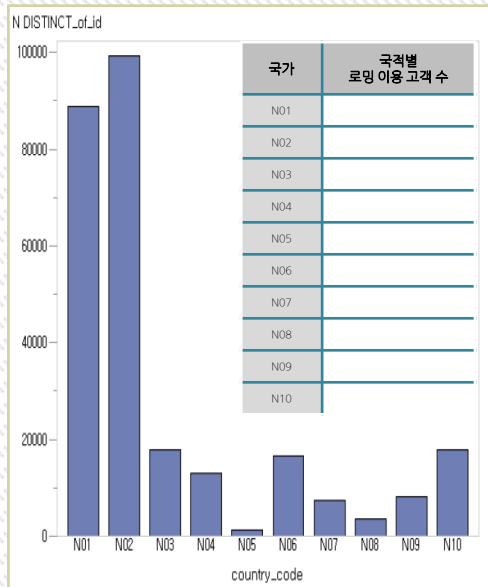
- RAW 데이터 처리 후 완성된 데이터

id	date	time	country	zip	latitude	longitude	SIDO	GUGUN	DONG	Location
2e49e166bc	2013-12-20	20:21 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
2591015420	2013-12-13	19:59 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
e19e266a23c	2014-02-19	13:30 N09	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
891ca86a6d	2014-01-07	17:43 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
9209e22959	2014-01-06	25:41 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
264cde81e0	2014-02-06	11:14 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
b8994a2a9d	2014-02-04	16:54 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
d85b076814	2014-02-14	17:46 N06	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
e1201381b1h	2013-12-07	16:47 N09	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
0e8bc1aa85a	2014-01-21	8:35 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
201be6418d3	2014-02-05	14:42 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
b73c59a6af	2014-01-25	12:24 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
26c95a9601	2013-12-25	19:57 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
6699c2c46a	2014-01-13	20:46 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
c35c4e9f59c	2014-02-17	12:23 N03	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
c7e420be46	2013-12-24	22:23 N03	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
196ca41343	2013-12-24	17:59 N10	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
6699c2c46a	2014-01-11	15:44 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
9176741b7c	2014-01-20	17:37 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
9d1393a4a3e	2013-12-13	22:05 N02	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	
3704dea4fb	2014-01-15	9:14 N01	100011	315015	126.902192	서울	영등포	영등포1가	서울 영등포구영등포	

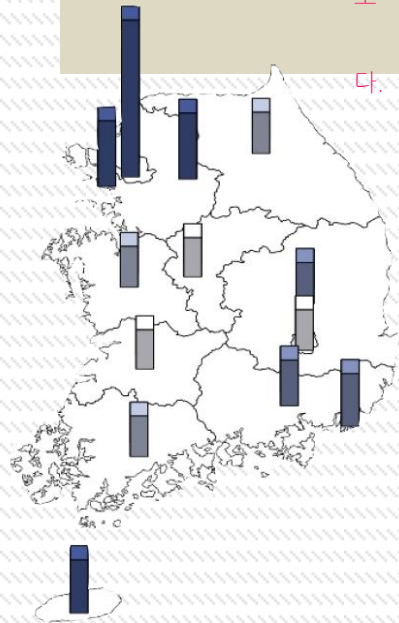
- RAW 데이터 처리 후 완성된 데이터의 변수 설명

id	date	time	Country_code	zip	latitude	longitude	SIDO	GUGUN	DONG	Location
각각의 고객의 암호화된 ID	통합가 발생한 날짜기록 ex)2014-02-21	로밍이 발생한 시간기록 ex)12:40	고객의 국적정보 (10개국) ex)N01~N10	우리나라 우편번호	기지국의 위도 정보	기지국의 경도 정보	기지국의 시,도 정보	기지국의 구,군 정보	기지국의 동 정보	SIDO+GUGUN+DONG
	분석에 용이한 단위로 변환						새로 추가된 변수들			

- Sampling 전, 데이터의 시각화



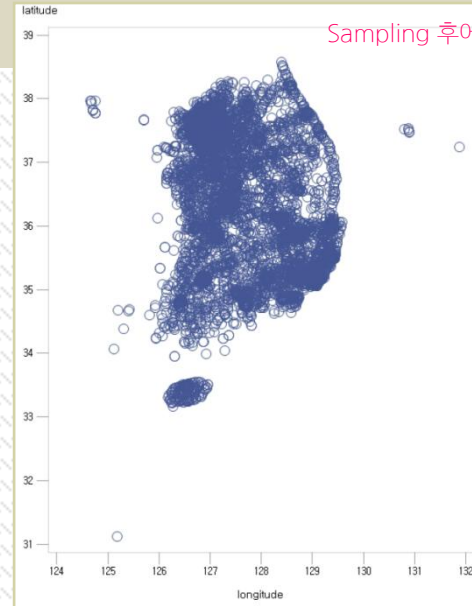
〈국적별 로밍 이용 고객 수〉
Imbalanced data
N01, N02 나라에 집중



〈지역별 로밍 사용횟수〉
서울과 경기,
인천, 제주도에 집중

* 기존 SEMMA cycle에서 2번째 단계인 Explore,
현재의 Explore만으로도 충분히, 데이터의 분포나 특징을 보여줄 수 있으며

로
다.



Sampling 후에는 첨가 하지 않았습니다

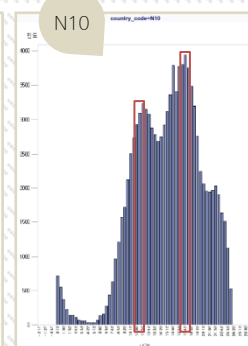
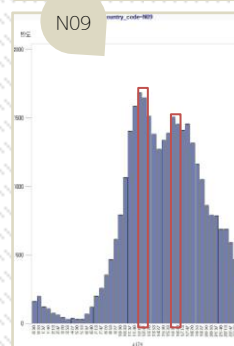
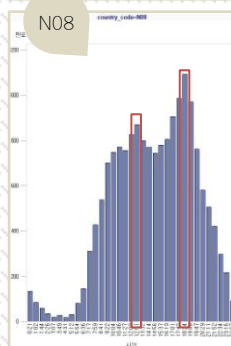
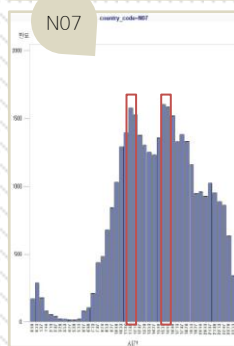
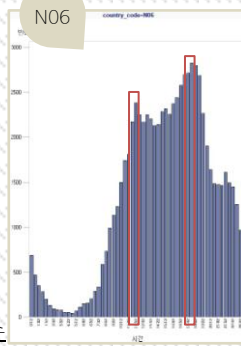
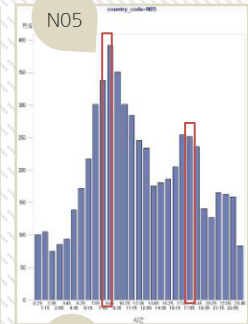
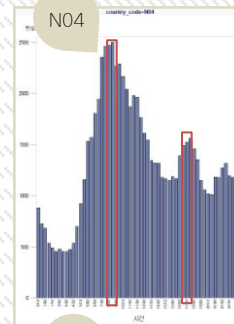
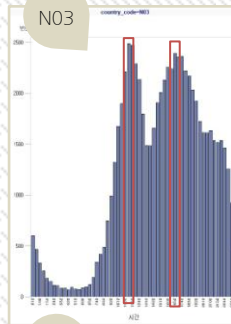
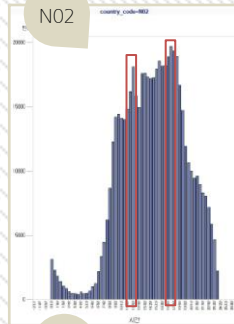
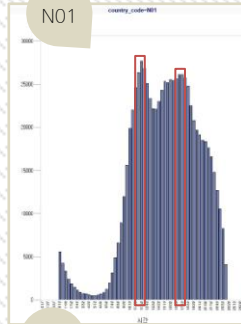
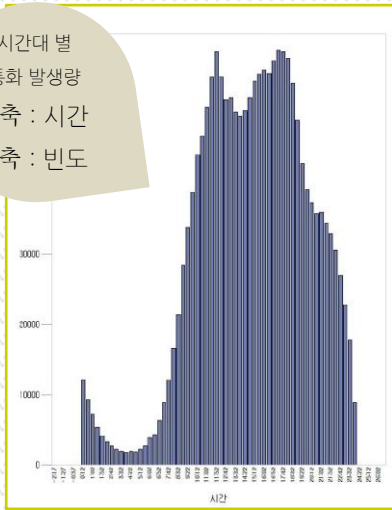
〈로밍이 발생한 장소〉

= 기지국의 위치

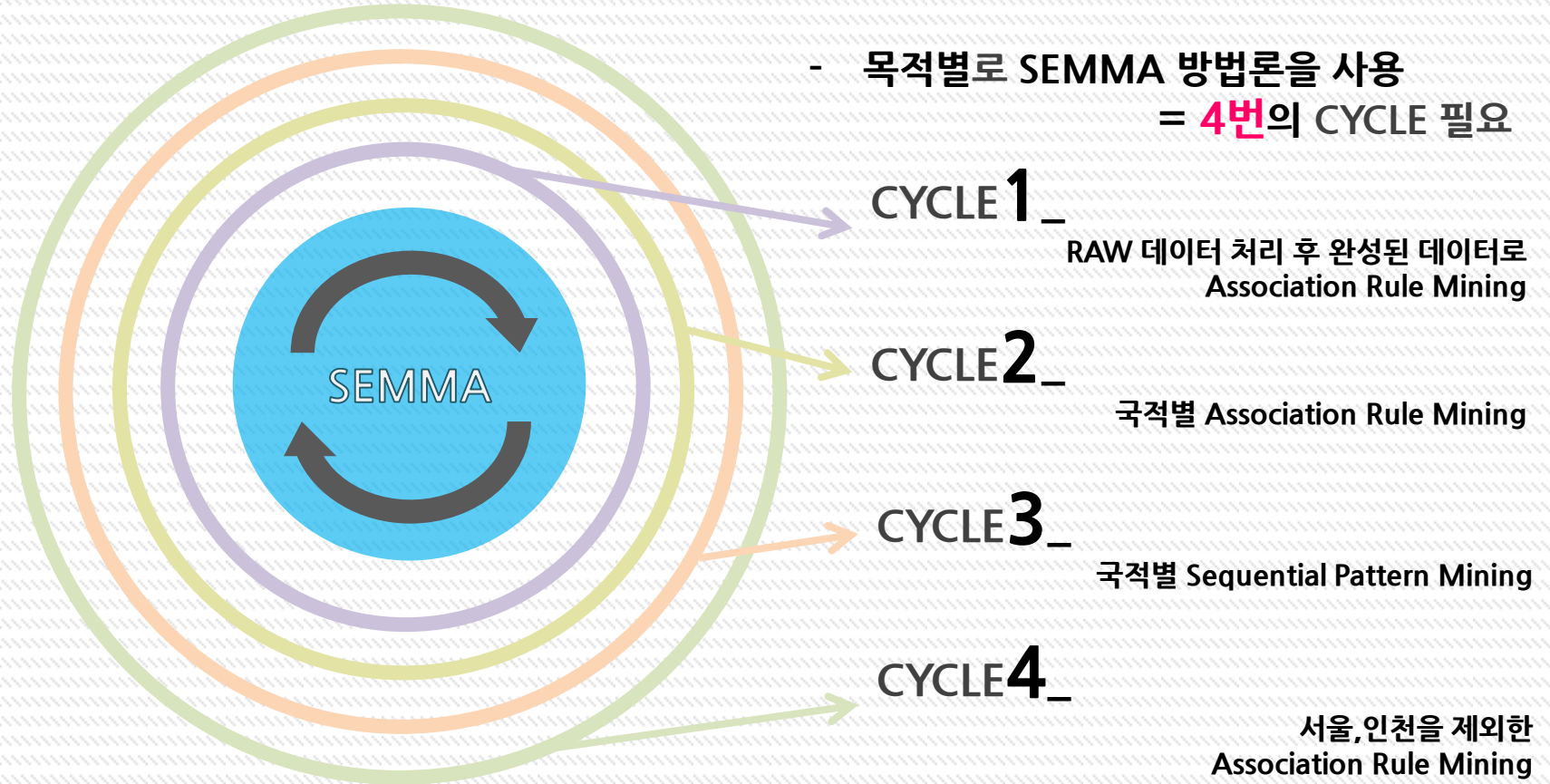
+
진한부분은 높은 빈도로 로밍 발생

- Sampling 전, 데이터의 시각화

시간대 별
동화 발생량
x 축 : 시간
y 축 : 빈도



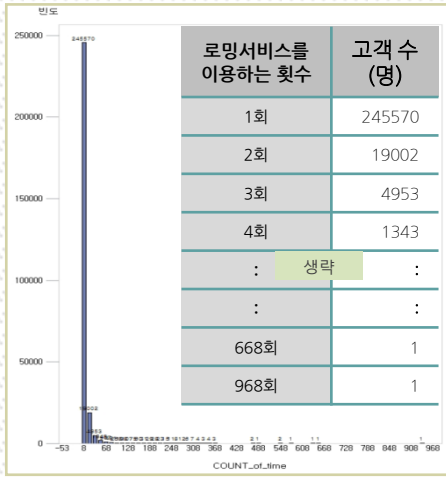
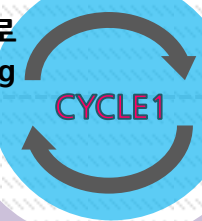
- > N01 ~ N10 이상 10개국 공통적으로 일상 생활을 유지하는 시간 = 큰 빈도수
- > N01 ~ N10 이상 10개국마다 local maximum 값을 갖는 횟수나 시간-빈도 그래프 모형 차이 발견
- > 이동 경로 연구와 더불어 위치 정보를 확인하여 N## 나라 예측, 나라 별 선호도 예측, 설문과 실제 관광 경로와의 차이 발견 가능



- 목적별로 SEMMA 방법론을 사용
= 4번의 CYCLE 필요

D_ 분석 및 실행 > 03_SEMMA방법론 > CYCLE1_Sample

RAW 데이터 처리 후 완성된 데이터로
Association Rule Mining



> 1번만 로밍서비스를 이용하는 고객
= 약 25만 명으로 과반수이상을 차지

→ Association Mining Rule방법에 적합하지 않은 imbalanced 데이터
→ 분석에 적합한 Sampling 실시 필요

Rule을 파악할 수 있는 **최소 3번 이상 통화한 사람 중**
N01~N10상관없이
랜덤 5000명을 표본 추출

<개인당 로밍 횟수>

id	country_code	Location
22a4aba2e3dbf62...	N02	영북구미시비산동
6dd8a3bfa29c505...	N01	영북구미시비산동
07aa452c96da9c0...	N01	영북구미시비산동
eefdf6a442880e2...	N02	영북구미시비산동
ba860d73e19828a...	N09	영북구미시비산동
6dd8a3bfa29c505...	N01	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
22a4aba2e3dbf62...	N02	영북구미시비산동
b11c35e460c3d72...	N01	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
07aa452c96da9c0...	N01	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
07aa452c96da9c0...	N01	영북구미시비산동
e50697d7eb04b01...	N02	영북구미시비산동
22a4aba2e3dbf62...	N02	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
38857bdd15e03b7...	N02	영북구미시비산동
c9884c0c571396...	N02	영북구미시비산동
ba860d73e19828a...	N09	영북구미시비산동
f2a2109115dc1c5...	N07	영북구미시비산동
07aa452c96da9c0...	N01	영북구미시비산동
c657750d9d1102e...	N02	영북구미시비산동
80b93f8e2b3a811d...	N09	영북구미시비산동
0d7b324cf70363b...	N06	영북구미시비산동
ddcb3137f9b25e...	N03	영북구미시비산동
38857bdd15e03b7...	N02	영북구미시비산동
fec0e41035e56c8...	N02	영북구미시비산동
ba860d73e19828a...	N09	영북구미시비산동
14eafcf24dec191d...	N03	영북구미시비산동
b7b7522917dacc...	N02	영북구미시비산동
6dd8a3bfa29c505...	N01	영북구미시비산동

Local WORK QUERY FOR SECOND SAS7BDAT에 대한 실행 결과

Local WORK QUERY FOR SECOND SAS7BDAT_0000에 대한 필터 및 조건

필터 설정 (F):
COUNT DISTINCT_of_Location > Greater than or equal to 3

Local WORK QUERY FOR SECOND SAS7BDAT에 대한 실행 결과

Local WORK QUERY FOR SECOND SAS7BDAT_0000에 대한 필터 및 조건

필터 설정 (F):
COUNT DISTINCT_of_Location > Greater than or equal to 3

Distinct 명령어 사용
= 관광객 별 중복된 위치 제거

<완성된 Sampling>



- 추가 분석 여부 결정

lhs	rhs	support	confidence	lift
1 {제주서귀포시중문동}	=> {제주제주시연동}	0.01020204	0.796875	12.371361
2 {제주서귀포시중문동}	=> {N01}	0.01180236	0.921875	2.179968
3 {광주서구풍암동}	=> {N02}	0.0110022	0.9821429	3.149283
4 {대전중구옥계동}	=> {경기가평균설악면}	0.010002	0.8333333	69.430556
5 {경기가평균설악면}	=> {대전중구옥계동}	0.010002	0.8333333	69.430556
6 {대전중구옥계동}	=> {부산강서구송정동}	0.010002	0.8333333	59.511905
7 {부산강서구송정동}	=> {대전중구옥계동}	0.010002	0.7142857	59.511905
8 {대전중구옥계동}	=> {N02}	0.01160232	0.9666667	3.099658
9 {전북전주시덕진구산정동}	=> {N02}	0.01020204	0.9272727	2.97334
10 {울산동구전하동}	=> {부산강서구송정동}	0.010002	0.8196721	58.5363
11 {부산강서구송정동}	=> {울산동구전하동}	0.010002	0.7142857	58.5363
12 {울산동구전하동}	=> {N02}	0.01120224	0.9180328	2.943711
13 {경기가평균설악면}	=> {N02}	0.01160232	0.9666667	3.099658
14 {부산부산진구부전2동}	=> {N02}	0.0120024	0.7407407	2.375217
15 {서울구로구구로2동}	=> {N01}	0.010002	0.8064516	1.907025
16 {인천중구항동7가}	=> {N01}	0.01060212	0.9298246	2.198767
17 {부산강서구송정동}	=> {N02}	0.01320264	0.9428571	3.023312

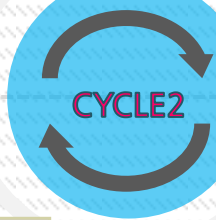
<완성된 Association Rule>

> 모두 140개의 rule이 생성, but,

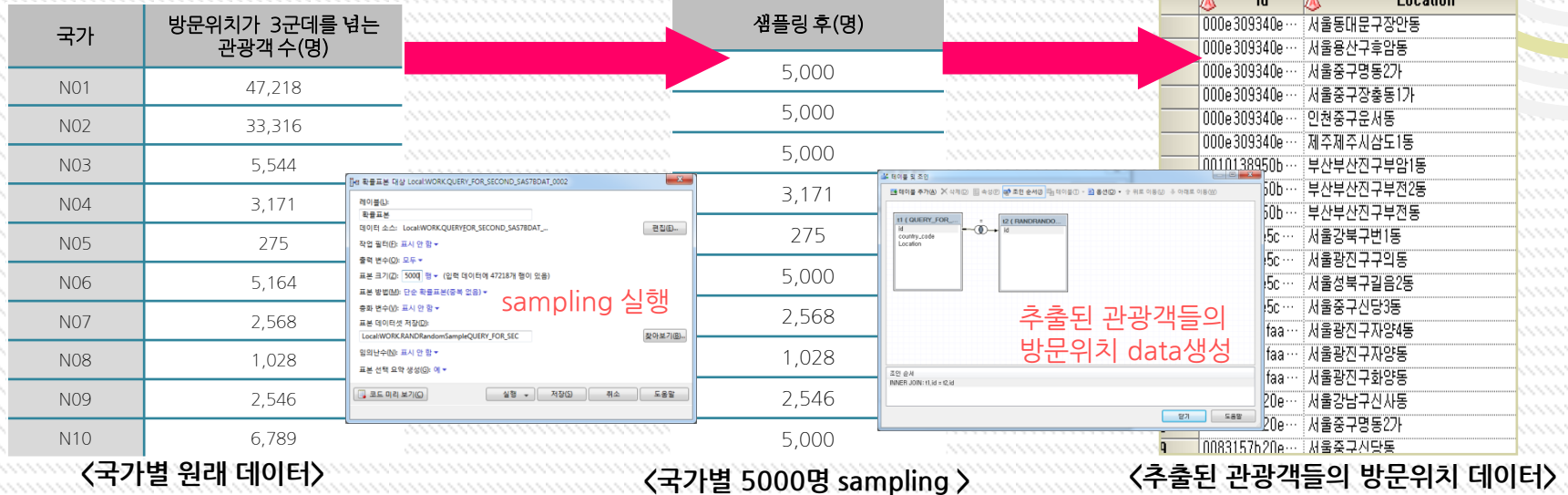
Rule을 파악할 수 있는 최소 3번 이상 통화한 사람 중
N01~N10상관없이
랜덤 5000명을 표본 추출

이 조건 때문에,
로밍 이용 고객수가 많은 나라에
Rule이 집중될 수 밖에 없음

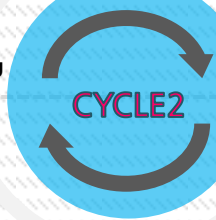
국가별 분석이 필요
(다음 SEMMA cycle2)



- 표본 추출 과정



- > N01의 방문객들만을 대상으로 5,000명의 확률표본을 추출
- > 추출된 관광객들의 방문위치의 데이터 생성 → .csv 파일로 변환
- > 나머지 N02~N10 국가에 대해서도 반복하여 데이터 생성



- Modify_R을 이용한 데이터 변환

```
dataCh <- function(filename){
  colname <- read.csv("ids.csv", header = T)
  temp <- matrix(0, nrow = 1, ncol = nrow(colname))
  dimnames(temp) <- list(c(), c(as.character(c(colname)$Location)))
  write.table(temp, paste(filename, "data.csv"), row.names = F)
  plain <- read.csv(paste(filename, "data.csv"), sep = ";")
  temp <- plain
  t <- as.matrix(read.csv(filename, header = T))
  row <- nrow(t)
  current <- t[1,1]
  for(i in 1:row){
    if(current != t[i,1]){
      write.table(temp, paste("data.csv",filename), append = T, row.names = F, col.names = F)
      temp <- plain
      current <- t[i,1]
      print(paste(filename,"",i / row))
    }
    temp[t[i,2]] <- 1
  }
}

Allchange <- function(){
  files <- c("N01.csv",
            "N02.csv",
            "N03.csv",
            "N04.csv",
            "N05.csv",
            "N06.csv",
            "N07.csv",
            "N08.csv",
            "N09.csv",
            "N10.csv")

  for(i in 1:10){
    dataCh(files[i])
  }
}
```

R 스크립트

모든 나라에 대해 변환

- > SAS E-Guide를 통해 추출된 관광객들의 방문위치 데이터를 R script를 활용하여 CYCLE1에서 했던 방법과 동일한 방법으로 모든 나라에 대해서 변환한다.

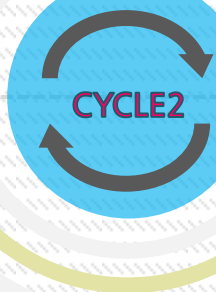
- Model_모델링 방식 선택

Association Rule Mining

```
Allrules <- function(){
  write("", "rule.txt")
  files <- c("data.csv N01.csv",
            "data.csv N02.csv",
            "data.csv N03.csv",
            "data.csv N04.csv",
            "data.csv N05.csv",
            "data.csv N06.csv",
            "data.csv N07.csv",
            "data.csv N08.csv",
            "data.csv N09.csv",
            "data.csv N10.csv")

  for(i in 1:10){
    data <- as(as.matrix(read.csv(files[i])), "itemMatrix")
    rule <- apriori(data, parameter = list(supp = 0.01, conf = 0.5, target="rules"))
    write("", "rule.txt", append = T)
    write("", "rule.txt", append = T)
    write(files[i], "rule.txt", append = T)
    write(rule, "rule.txt", append = T)
  }
}
```

- > R을 이용
- > R의 “Apriori”알고리즘을 통해 minsup = 0.01, minconf = 0.5의 규칙을 찾는다
- > 모든 국가를 대상으로 각각 실시



- 추가 분석 여부 결정

```

data.csv N01.csv
rules support confidence lift
1 {제주제주시한민} => {제주제주시연도} 0.0105746209098164 0.726027397260274 5.85966073279951
2 {제주제주포시} => {제주제주시연도} 0.0157621707901038 0.766990291262136 6.1902662476744
3 {제주제주시미} => {제주제주시연도} 0.0141660015961692 0.702970297029703 5.6735702555762
4 {인천서구경서} => {인천중구문서동} 0.0113727055067837 0.6955172413793103 1.45040818813208
5 {서울서초구반포4동} => {인천중구문서동} 0.0103750997605746 0.634146341463415 1.40386107041282
6 {서울송파구구로동2가} => {서울송파구구로동2가} 0.0111731843575419 0.549019607843137 1.70067136867108
7 {서울서대문구구로동2가} => {서울서대문구구로동2가} 0.0133679169992019 0.609090909090909 1.886751320373
8 {서울서대문구구로동2가} => {인천중구문서동} 0.0125698924022346 0.572727272727273 1.2678927079987
9 {서울서대문구구로동2가} => {서울중구영등포동2가} 0.0139664804469274 0.569105691056911 1.76289105289075
10 {서울서대문구구로동2가} => {인천중구문서동} 0.0125698924022346 0.51219512195122 1.13388778764113
11 {제주제주시연도} => {제주제주시연도} 0.0223463687150838 0.67065868263473 5.41278795066871
12 {서울서대문구구로동2가} => {인천중구문서동} 0.0117717478052674 0.526785714285714 1.1661881625441
13 {서울서대문구구로동2가} => {인천중구문서동} 0.0147645650438947 0.637931034482759 1.4122395516022
14 {서울서대문구구로동2가} => {인천중구문서동} 0.0133679169992019 0.540322580645161 1.19615581899008
15 {서울서대문구구로동2가} => {인천중구문서동} 0.0135674381484437 0.56198347107438 1.24410828490494
16 {서울서대문구구로동2가} => {인천중구문서동} 0.0113727055067837 0.527777777777778 1.168884373737307
17 {서울서대문구구로동2가} => {인천중구문서동} 0.0119712689545092 0.5 1.106889045936396
18 {서울서대문구구로동2가} => {인천중구문서동} 0.0137669592976856 0.575 1.27292402826855
19 {서울서대문구구로동2가} => {서울서대문구구로동2가} 0.0137669592976856 0.507352941176471 1.5716025594415
20 {서울서대문구구로동2가} => {서울서대문구구로동2가} 0.0267358339994038 0.666666666666667 5.380568974771
21 {서울서대문구구로동2가} => {인천중구문서동} 0.0201516360734298 0.696551724137931 1.54201291580358
22 {서울서대문구구로동2가} => {인천중구문서동} 0.0189545091779729 0.582822085889571 1.29024041275553
23 {서울서대문구구로동2가} => {서울서대문구구로동2가} 0.0347166799680766 0.71900826446281 5.80301034055975
24 {서울서대문구구로동2가} => {서울서대문구구로동2가} 0.0167597765363128 0.506024096385542 1.5674862614859
25 {서울서대문구구로동2가} => {인천중구문서동} 0.0193535514764565 0.5 1.54882571075402
26 {서울서대문구구로동2가} => {인천중구문서동} 0.0219473264166002 0.56701030927895 1.255236603402
27 {서울서대문구구로동2가} => {인천중구문서동} 0.0181564245810066 0.579617834394904 1.28314690193784
    
```

<완성된 N01의 Association Rule>

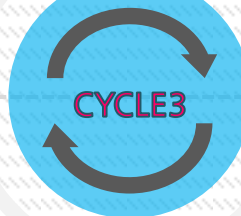
> 각 국가별 rule이 생성, but,

국가	생성된 Rule 개수	국가	생성된 Rule 개수
N01	138	N06	73
N02	450	N07	459
N03	282	N08	31
N04	32	N09	1,698
N05	208	N10	60

너무 많은 룰의 개수,
해석이 복잡

+ 순서의 개념의 부재

순서를 고려한 분석이 필요
(다음 SEMMA cycle3)



- 표본 추출 과정

id	date	country_code	Location
000025e56ff7...	2014-01-29	N06	서울강남구신...
000025e56ff7...	2014-01-29	N06	인천중구운서동
00002931bea...	2013-12-03	N09	서울영등포구...
00002931bea...	2013-12-03	N09	서울용산구...
00002931bea...	2013-12-04	N09	서울마포구...
00002931bea...	2013-12-04	N09	서울영등포구...
00009b90331...	2014-02-10	N10	서울중로구인...
00009b90331...	2014-02-11	N10	서울중로구...
00009b90331...	2014-02-11	N10	서울중로구인...
00009b90331...	2014-02-12	N10	경기화성시양...
00009b90331...	2014-02-12	N10	서울마포구...
00009b90331...	2014-02-12	N10	서울용산구...
00009b90331...	2014-02-13	N10	경남거제시...
00009b90331...	2014-02-13	N10	경남거제시...
00009b90331...	2014-02-14	N10	경남거제시...
00009b90331...	2014-02-14	N10	인천중구운서동
0000c3f9			용산구...
0000c3f9			중구영등...
0000eb5e			영등포구...
0000eb5ac31...	2014-01-12	N04	인천중구운서동
0001040080b...	2013-12-12	N02	서울광진구...
0001040080b...	2013-12-12	N02	서울마포구...
0001040080b...	2013-12-13	N02	서울중로구...
0001040080b...	2013-12-13	N02	서울중로구...

한글로 표시된 위치

Location	index
강원강릉시강동면	1
강원강릉시견소동	2
강원강릉시교1동	3
강원강릉시교2동	4
강원강릉시교동	5
강원강릉시구정면	6
강원강릉시노암동	7
강원강릉시대전동	8
강원강릉시사천면	9
강원강릉시성산면	10
강원강릉시안현동	11
강원강릉시연곡면	12
강원강릉시옥계면	13
강원강릉시옥천동	14
강원강릉시왕산면	15
강원강릉시임암동	16
강원강릉시입암동	17
강원강릉시주운면	18
강원강릉시지변동	19
강원강릉시초암동	20
강원강릉시초남1동	21
강원강릉시포남2동	22
강원강릉시포남동	23
강원강릉시홍계동	24
강원강릉시회산동	25
강원고성군간성읍	26
강원고성군겨례동	27

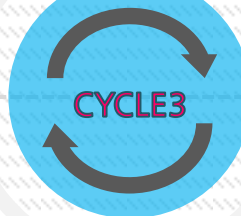
<국가별 원래 데이터> <Location DB>

id	date	country_code	Location
000e309340e...	2014-02-05	N01	
000e309340e...	2014-02-06	N01	
000e309340e...	2014-02-06	N01	
000e309340e...	2014-02-07	N01	
000e309340e...	2014-02-07	N01	
000e309340e...	2014-02-08	N01	
0010138950b...	2014-01-27	N01	
0010138950b...	2014-01-27	N01	
0010138950b...	2014-01-28	N01	
0010138950b...	2014-01-28	N01	
0024d6f5e5c...	2014-02-01	N01	
0024d6f5e5c...	2014-02-02	N01	
0024d6f5e5c...	2014-02-02	N01	
0024d6f5e5c...	2014-02-04	N01	
006638f0afaa...	2014-02-19	N01	
006638f0afaa...	2014-02-25	N01	
006638f0afaa...	2014-02-27	N01	

숫자 코드

<국가별 완성된 Sampling>

- > 마찬가지로 국가별 5,000명 이하로 추출된 관광객을 대상, data를 포함, 같은 날에 중복된 장소를 제거
- > 알고리즘을 사용할 수 있게 Location DB에 숫자 index 붙인 데이터를 더해 한글로 표시된 위치숫자코드로 변환
- > 모든 국가(N01-N10)을 대상으로 데이터 set 생성 → .csv 파일로 변환



- 추가 분석 여부 결정

sequence_name	sequence_value	support	ratio	5000
<(명동)(을지로2가)>	<(2390, 2398)>	324	0.0648	
<(명동)(충무로1가)>	<(2390)(2406)>	247	0.0494	
<(명동)(신당3동)>	<(2390)(2395)>	187	0.0374	
<(명동)(장충동1가)>	<(2390)(2402)>	182	0.0364	
<(명동)(신사동)>	<(2390)(1963)>	173	0.0346	
<(충무로1가)(인천국제공항)>	<(2406)(2638)>	169	0.0338	
<(명동)(태평로2가)>	<(2390)(2411)>	168	0.0336	
<(인천국제공항)(송인1동)>	<(2638)(2396)>	140	0.028	
<(명동)(광희동1가)>	<(2390)(2387)>	140	0.028	
<(명동)(충무로2가)>	<(2390)(2407)>	136	0.0272	
<(충무로1가)(명동)>	<(2406)(2390)>	131	0.0262	
<(인천국제공항)(광장동)>	<(2638)(2040)>	114	0.0228	
<(인천국제공항)(신사동)>	<(2638)(1963)>	114	0.0228	
<(명동)(필동2가)>	<(2390)(2412)>	110	0.022	
<(석촌동)(잠실동)>	<(2256)(2263)>	106	0.0212	
<(인천국제공항)(신당3동)>	<(2638)(2395)>	105	0.021	
<(명동)(과해동)>	<(2390)(2004)>	103	0.0206	
<(석촌동)(명동)>	<(2256)(2390)>	102	0.0204	

<완성된 N01의 Sequential Pattern>

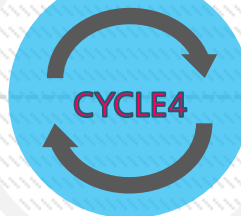
> 각 국가별 rule이 생성,

국가	생성된 Pattern 개수	국가	생성된 Pattern 개수
N01	18	N06	48
N02	113	N07	11
N03	10	N08	15
N04	20	N09	16
N05	33	N10	25

비교적 간단한 결과로 해석이 용이함. but,
 모든 방법론이 **서울에 편중되어 룰이 생성됨**
(인천공항이 압도적으로 많아 인천지역도 포함)

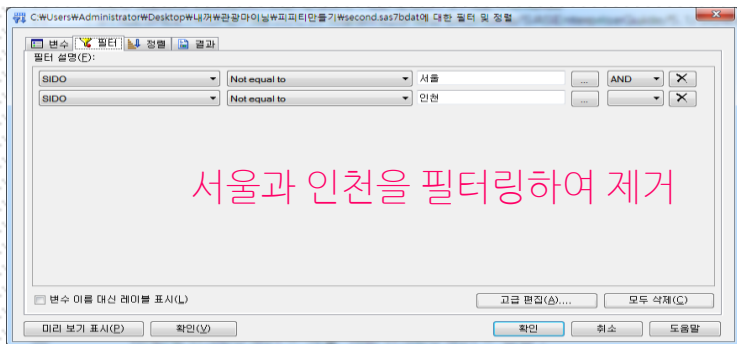
서울, 인천지역을 제외한 분석이 필요

(다음 SEMMA cycle4)



* 샘플링을 제외한 모든 과정이 CYCLE 1과 똑같음

- 표본 추출 과정



- > 앞선 cycle1,2,3에서 서울지역의 rule 및 Pattern이 지배적으로 나와 지방에 대한 데이터가 모자라다.
- > 마찬가지로 국가별 5,000명 이하로 추출된 관광객을 대상, data를 포함, 같은 날에 중복된 장소를 제거

- 추가 분석 여부 결정

rules		support	confidenc	lift
{제주서귀포시색달동}	=> {제주제주시연동}	0.010402	0.742857	5.179279
{제주제주시이도이동}	=> {제주제주시연동}	0.010002	0.675676	4.710882
{제주서귀포시하원동}	=> {제주제주시연동}	0.012002	0.722892	5.040077
{제주제주시구좌읍}	=> {제주제주시연동}	0.010202	0.708333	4.938575
{제주제주시도두일동}	=> {제주제주시연동}	0.013803	0.71134	4.959539
{제주서귀포시성산읍}	=> {제주제주시애월읍}	0.011002	0.591398	7.158348
{제주서귀포시성산읍}	=> {제주제주시연동}	0.014003	0.752688	5.247822
{제주제주시한림읍}	=> {제주제주시연동}	0.015403	0.793814	5.534558
{제주제주시이도일동}	=> {제주제주시연동}	0.021604	0.724832	5.053607
{제주서귀포시중문동}	=> {제주제주시연동}	0.027205	0.829268	5.781746
{부산부산진구부전동}	=> {부산부산진구부전2동}	0.021004	0.719178	12.57053
{대구동구서호동}	=> {광주서구풍암동}	0.020404	0.886957	24.36206
{광주서구풍암동}	=> {대구동구서호동}	0.020404	0.56044	24.36206
{대구동구서호동}	=> {대전중구옥계동}	0.020604	0.895652	24.46648
{대전중구옥계동}	=> {대구동구서호동}	0.020604	0.562842	24.46648
{대구동구서호동}	=> {전북전주시덕진구산정동}	0.020204	0.878261	24.2565
{전북전주시덕진구산정동}	=> {대구동구서호동}	0.020204	0.558011	24.2565
{대구동구서호동}	=> {경기가평군설악면}	0.021404	0.930435	23.73083
{경기가평군설악면}	=> {대구동구서호동}	0.021404	0.545918	23.73083

<완성된 Association Rule>

- > 모두 742개의 rule이 생성

- Model_모델링 방식 선택

Association Rule Mining



마지막 SEMMA cycle 완성

데이터가 무언가 말하면
분석가는 귀 기울여 듣는다.
그리고 생각한다.

“분명히 뭔가를 숨기고 있구나!”



A_ 과제 파악

B_ 현재 관광산업의 동향 및 이슈 조사

C_ 분석 방향 설정

D_ 분석 시행

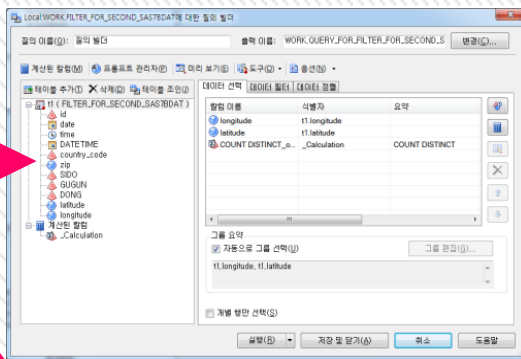
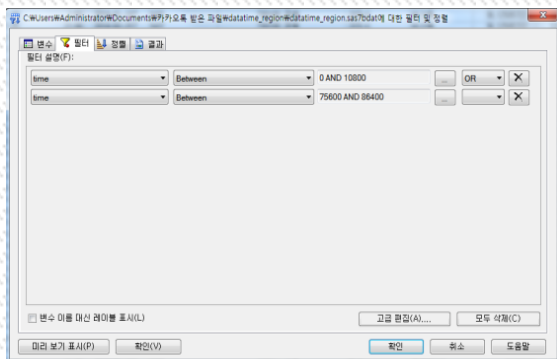
E_ 추가 분석/ 01_ 밤 시간대 데이터 분석

F_ 결과 해석

G_ 관광마케팅 방안 기획 및 결론

H_ 참고 문헌

- 밤 시간대 데이터 조사



21:00부터 3:00까지의 데이터 필터링 후
중복되는 고객 제거

longitude	latitude	COUNT DISTINCT_of_k
124.6631472	37.96430555	1
124.6649333	37.96481388	1
124.687411	37.81811527	2
124.7071583	37.83226111	1
124.709695	37.83195666	1
124.7441	37.95343888	4
124.7443888	37.77235277	2
124.749666	37.77421388	1
125.1234125	34.06342666	6
125.181944	31.1225	1
125.189288	34.66110194	2
125.3011583	34.39010277	1
125.404388	34.66523877	6
125.4174441	34.69293277	1
125.7036833	37.67298333	1
125.7041361	37.66711388	2
125.8149186	34.59915833	1
125.93006	34.72274833	2
125.9330444	34.24464444	4
125.9333333	34.23944444	2
125.9493288	37.08340472	7
125.9718694	37.18668611	1
125.9736625	36.12074855	1
125.9743066	37.15167666	1
126.0048405	34.32837944	4
126.0169722	35.34341388	1
126.0170027	35.34393611	1
126.0491222	34.59971055	1

각 그룹별 date를 Count

같은 SIDO,GUGUN, DONG은
latitude,longitude 평균 Mean을 활용하여
하나의 그룹으로 통일

SIDO, GUGUN, DONG 데이터 추가 변환

SIDO	GUGUN	DONG	MEAN_of_longitude	MEAN_of_latitude	COUNT_of
서울	중구	신상동	37.56498	127.013	10351
인천	중구	운서동	37.45456	126.4531	8631
제주	제주시	연동	33.48798	126.4906	8456
서울	중구	명동2가	37.56351	126.9845	6552
서울	강남구	삼성1동	37.51061	127.059	5210
서울	광진구	광장동	37.55504	127.1102	4021
서울	중구	율지로2가	37.56522	126.9835	3757
제주	제주시	건입동	33.509	126.5242	3750
서울	중구	태평로2가	37.56466	126.9767	3593
서울	중구	광희동1가	37.56478	127.0068	3373
서울	중구	충무로2가	37.56125	126.9873	3345
서울	강남구	논현동	37.50953	127.0316	3006
서울	강남구	신사동	37.52354	127.0284	2939
서울	영등포구	여의도동	37.52592	126.9244	2392
서울	강남구	역삼동	37.50311	127.0311	2337
서울	용산구	한남동	37.53493	126.9996	2335
서울	중구	남대문로5	37.5575	126.9758	2238
서울	강남구	역삼1동	37.50388	127.0341	2201
서울	강남구	청담동	37.52449	127.0517	2097
서울	중구	충무로1가	37.56075	126.9822	2080
서울	종로구	창신1동	37.57076	127.0113	2048
경기	수원시 팔달구	인계동	37.26691	127.031	1969
서울	강서구	과해동	37.56119	126.7977	1958
서울	마포구	동교동	37.55629	126.9234	1708
서울	서초구	잠원동	37.51224	127.012	1668
서울	중구	필동2가	37.56091	126.9944	1621
서울	중구	창충동1가	37.55972	127.0062	1428
서울	종로구	서린동	37.56991	126.9797	1421
부산	부산진구	부전2동	35.15537	129.058	1325
강원	평창군	대관령면	37.6563	128.6741	1301
서울	서초구	반포4동	37.50341	127.0079	1286
제주	제주시	이도일동	33.50647	126.5232	1283
서울	용산구	동빙고동	37.52875	126.9923	1200
서울	강남구	역삼2동	37.50186	127.0419	1198

<완성된 밤 시간대 데이터>

새는 알에서 태어난다_탄생
정보는 데이터에서 태어난다_발생
그리고 지식으로 성장한다_발전



A_ 과제 파악

B_ 현재 관광산업의 동향 및 이슈 조사

C_ 분석 방향 설정

D_ 분석 시행

E_ 추가 분석

F_ 결과 해석/

01_ SEMMA CYCLE2 해석

02_ SEMMA CYCLE3 해석

03_ SEMMA CYCLE4 해석

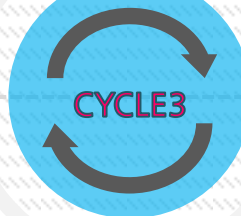
04_ 밤 시간대 데이터 해석

05_ 구글어스에 시각화

06_ 데이터마이닝 결과 특수장소에 특수목적의 방문객

G_ 관광마케팅 방안 기획 및 결론

H_ 참고 문헌



N01	Sequence_name	Ratio
<(명동)(을지로2가)>		0.0648
<(명동)(충무로1가)>		0.0494
<(명동)(신당3동)>		0.0374
<(명동)(강충동1가)>		0.0364
<(명동)(신사동)>		0.0346
<(충무로1가)(인천국제공항)>		0.0338
<(명동)(태평로2가)>		0.0336
<(인천국제공항)(송인1동)>		0.028
<(명동)(광희동1가)>		0.028
<(명동)(충무로2가)>		0.0272
<(충무로1가)(명동)>		0.0262
<(인천국제공항)(광장동)>		0.0228
<(인천국제공항)(신사동)>		0.0228
<(명동)(필동2가)>		0.022
<(석촌동)(잠실동)>		0.023002
<(인천국제공항)(신당3동)>		0.021852
<(명동)(과해동)>		0.021277
<(석촌동)(명동)>		0.020702

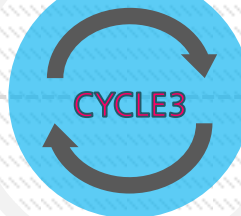
- > 명동일대 (명동/을지로2가)를 시작으로 하는 경우가 많음
- > 기타지역 포함된 상태로 특히 명동부근에서 통화량이 많음
 - **쇼핑을 포함한 자유여행**일 확률이 높음
 - 개인이든 단체든 schedule 있는 여행을 할 경우 전화가 필요한 경우가 거의 없기 때문(여행객 상호 간)
- > 인천국제공항, 여객선을 통해 출입국 하는 경우 대다수

N02	Sequence_name	Ratio
<(을지로2가)(명동)>		0.05968
<(송인1동)(명동)>		0.057054
<(을지로2가,명동)>		0.056815
<(명동)(과해동)>		0.05419
<(신사동)(과해동)>		0.053712
<(과해동)(방화2동)>		0.049176
<(명동)(을지로2가)(인천국제공항)>		0.04846
<(신사동)(명동)>		0.047267
<(명동)(충무로1가)>		0.046312
<(신사동)(인천국제공항)>		0.045596
<(명동)(인천국제공항)>		0.045357
<(신당3동)(명동)>		0.043447
<(광장동)(명동)>		0.043447
<(과해동)(신사동)>		0.042731
<(을지로2가)(과해동)>		0.042254
<(방화2동)(과해동)>		0.042254
<(송인1동)(을지로2가)>		0.042015
<(신당3동)(강충동1가)>		0.040582
<(광희동1가)(명동)>		0.040344
<(삼성1동)(과해동)>		0.039866
<(명동)(충무로2가)>		0.038434
<(신당3동)(인천국제공항)>		0.038434
<(광희동1가)(인천국제공항)>		0.037957
<(인천국제공항)(명동)(을지로2가)>		0.033421

- > 명동관련 : N01과 동일
- > 앞선 데이터에 비해 **신당동, 신당3동 데이터 다수**
 - **신라호텔, 동대문디자인플라자 부근**
 - 여행의 목적 중 쇼핑이 꼭 잡혀있음
- > 이상 3개 이상의 sequence를 보인 항목은 2개, 명동-을지로2가-인천국제공항-인천국제공항-명동-을지로2가
 - 이는 순서의 차이가 있을 뿐 대부분의 **tour의 정석** / (완전한 개인 목적이 아닌 이상 관광객 필수코스)
- > **김포공항(과해동)을 이용해서 입국** 다수 관찰가능.

N03	Sequence_name	Ratio
<(신당동)(인천국제공항)>		0.028227
<(명동)(인천국제공항)>		0.02791
<(명동)(을지로2가)>		0.026958
<(창신1동)(신당동)>		0.026641
<(광희동1가)(신당동)>		0.024421
<(인천국제공항)(명동)>		0.024421
<(신당동)(창신1동)>		0.022835
<(광희동1가)(인천국제공항)>		0.022201
<(신당동)(광희동1가)>		0.020932
<(인천국제공항)(신당동)>		0.020615

- > 인천국제공항 입국
- > **창신1동** 다수발견 (동대문시장)
- > 신당동
 - **대부분 동대문디자인플라자 주변**으로 관광 필수 코스



N04

Sequence_name	Ratio
<(삼성1동)(인천국제공항)>	0.038528
<(역삼동)(인천국제공항)>	0.035653
<(반포4동)(인천국제공항)>	0.031627
<(반포4동)(잠원동)>	0.030477
<(인천국제공항)(삼성1동)>	0.027027
<(논현동)(역삼동)>	0.027027
<(용산동2가)(한남동)>	0.027027
<(역삼1동)(역삼동)>	0.025877
<(한남동)(용산동2가)>	0.025877
<(잠원동)(인천국제공항)>	0.025302
<(길월동)(후암동)>	0.025302
<(길월동)(용산동3가)>	0.024727
<(잠원동)(반포4동)>	0.024152
<(역삼1동)(인천국제공항)>	0.023577
<(논현동)(인천국제공항)>	0.023002
<(용산동3가)(용산동6가)>	0.021852
<(용산동3가)(길월동)>	0.021277
<(신사동)(인천국제공항)>	0.020702
<(역삼동)(논현동)>	0.020127
<(한남동)(인천국제공항)>	0.020127

- > 인천국제공항 입국
- > 용산 이태원거리 및 힐튼호텔 주변 다수
- > 강북보다는 **강남이 다수**
- > 반포 4동 = 신세계 면세점 백화점 및 **고속 터미널 위치**
이곳에서 **지방으로 이동하는 여행객이 있을 가능성** 있음
- > 역삼동, 신사동, 삼성1동 등 소위 강남이 다수
모두 다음 sequence가 인천국제공항
→ **마지막으로 서울에서 마무리는 잠소일** 가능성 있음

- > 인천국제공항 입국
- > 명동, 을지로2가, 낙원동(인사동), 관철동(보신각 및 종로 피아노 거리), 무교동, 태평로 2가(서울시청광장) 필수 코스
- > **잠실동-상일동-석촌동** → 잠실 롯데월드, 석촌백제고분 지역과

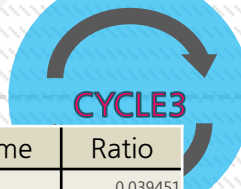
상일동 선사유적지 등

서울 중심지 외 역사적 중요 지역 방문 가능성 있음

- > **도선동-마장동** → 왕십리 굽창골목, 축산물마장시장 등 식도락 여행

N05

Sequence_name	Ratio	Sequence_name	Ratio
<(삼성1동)(인천국제공항)>	0.052239	<(충무로1가)(인천국제공항)>	0.022388
<(태평로2가)(을지로2가)>	0.044776	<(서초3동)(인천국제공항)>	0.022388
<(반포4동)(인천국제공항)>	0.037313	<(삼성1동)(삼성2동)>	0.022388
<(을지로2가)(인천국제공항)>	0.029851	<(수송동)(을지로2가)>	0.022388
<(잠원동)(인천국제공항)>	0.029851	<(도선동)(마장동)>	0.022388
<(양재2동)(한남동)>	0.029851	<(을지로1가)(을지로2가)>	0.022388
<(역삼2동)(역삼동)>	0.029851	<(무교동)(을지로2가)>	0.022388
<(인천국제공항)(신사동)>	0.029851	<(무교동)(명동)>	0.022388
<(장충동1가)(인천국제공항)>	0.029851	<(을지로2가)(낙원동)(관철동)>	0.022388
<(을지로2가)(무교동)>	0.022388	<(을지로2가)(명동)(을지로1가)>	0.022388
<(반포4동)(잠원동)>	0.022388	<(낙원동)(을지로2가)(관철동)>	0.022388
<(과해동)(인천국제공항)>	0.022388	<(낙원동)(관철동)(을지로2가)>	0.022388
<(역삼2동)(인천국제공항)>	0.022388	<(잠실동)(상일동)(석촌동)>	0.022388
<(역삼동)(역삼2동)>	0.022388	<(신사동)(반포1동)(인천국제공항)>	0.022388
<(역삼동)(인천국제공항)>	0.022388	<(명동)(을지로2가)(태평로2가)>	0.022388
<(인천국제공항)(을지로2가)>	0.022388	<(명동)(을지로2가)(을지로1가)>	0.022388
<(태평로2가)(명동)>	0.022388		



Sequence_name	Ratio	Sequence_name	Ratio
<(신사동)(인천국제공항)>	0.056417	<(신사동)(여의도)>	0.024905
<(명동)(인천국제공항)>	0.046506	<(무교동)(태평로2가)>	0.024651
<(명동)(을지로2가)>	0.045743	<(신사동)(명동)>	0.023888
<(태평로2가)(인천국제공항)>	0.043202	<(청담동)(인천국제공항)>	0.023634
<(을지로2가)(인천국제공항)>	0.039898	<(신사동)(잠실동)>	0.023634
<(삼성1동)(인천국제공항)>	0.03939	<(청담동)(신사동)>	0.02338
<(인천국제공항)(신사동)>	0.038882	<(잠원동)(인천국제공항)>	0.022872
<(삼성1동)(신사동)>	0.034562	<(신사동)(태평로2가)>	0.022872
<(신사동)(한남동)>	0.034053	<(여의도)(태평로2가)>	0.022618
<(한남동)(인천국제공항)>	0.034053	<(태평로2가)(명동)>	0.022618
<(신사동)(삼성1동)>	0.033545	<(명동)(충무로1가)>	0.022109
<(인천국제공항)(태평로2가)>	0.03202	<(인천국제공항)(역삼동)>	0.020839
<(명동)(태평로2가)>	0.031766	<(충무로1가)(인천국제공항)>	0.020839
<(신사동)(청담동)>	0.031258	<(신사동)(충무로1가)>	0.020584
<(여의도)(인천국제공항)>	0.031258	<(명동)(충무로2가)>	0.020584
<(인천국제공항)(명동)>	0.030496	<(용산동2가)(한남동)>	0.02033
<(한남동)(신사동)>	0.029479	<(신사동)(역삼동)>	0.02033
<(인천국제공항)(여의도)>	0.028463	<(태평로2가)(여의도)>	0.020076
<(인천국제공항)(한남동)>	0.028208	<(논현동)(인천국제공항)>	0.020076
<(을지로2가)(태평로2가)>	0.027954	<(명동)(신사동)>	0.020076
<(인천국제공항)(삼성1동)>	0.0277	<(신사동)(을지로2가)>	0.020076
<(인천국제공항)(을지로2가)>	0.026938	<(충무로1가)(태평로2가)>	0.020076
<(을지로2가)(명동)>	0.026175	<(태평로2가)(을지로2가)>	0.020076
<(역삼동)(인천국제공항)>	0.025413	<(서린동)(태평로2가)>	0.020076

N06

- > 인천국제공항 입국
- > 태평로: 서울시청, 덕수궁, 승례문
- > 명동, 을지로2가, 낙원동, 관철동, 태평로 2가 등 **서울 강북의 중심부**, 신사동, 삼성동, 역삼동, 청담동 등 **강남의 중심부**, 용산동 한남동 **이태원거리지역을 필수 코스로 잡고 있음**
- > 여의도가 다수 나옴
- > 인천국제공항에 도착하고 나서
→ 삼성1동이나 을지로2가, 역삼동, 태평로 2가 쪽으로 이동
짐을 풀어 놓을 **숙소가 많은 장소로 추측** 가능

N07

- > 4.17% 대관령 - 명동
- 2.41% 대관령-인천국제공항
- 2.29% 대관령-을지로2가
- > 2.23% + 2.21% + 2.05% 롯데월드 - 명동
- > **대관령과 롯데월드가**

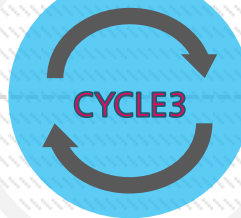
비율상 두드러짐

- > 명동 동일

N08

Sequence_name	Ratio
<(역삼동)(인천국제공항)>	0.039451
<(인천국제공항)(삼성1동)>	0.034305
<(역삼1동)(역삼동)>	0.030875
<(을지로2가)(명동)>	0.02916
<(삼성1동)(인천국제공항)>	0.02916
<(인천국제공항)(역삼동)>	0.027444
<(인천국제공항)(신사동)>	0.025729
<(명동)(을지로2가)>	0.024014
<(명동)(인천국제공항)>	0.022298
<(신사동)(역삼동)>	0.022298
<(역삼동)(역삼1동)>	0.020583
<(인천국제공항)(을지로2가)>	0.020583
<(삼성1동)(역삼동)>	0.020583
<(역삼2동)(역삼동)>	0.020583
<(신사동)(인천국제공항)>	0.020583
<(논현동)(역삼동)>	0.020583

- > 역삼동 다수 발견
삼성동 논현동 등으로 주로 **강남지역**
- > 적은 sequence 중에서도
역삼동이 자주 나오는 것으로 보아
→ 테헤란로 : **업무목적**일 수 있음



N09

Sequence_name	Ratio
<(명동)(인천국제공항)>	0.061283
<(명동)(을지로2가)>	0.046143
<(을지로2가)(인천국제공항)>	0.036049
<(동교동)(인천국제공항)>	0.031002
<(대전용전동)(명동)>	0.027397
<(대전용전동)(동교동)>	0.026676
<(충무로1가)(인천국제공항)>	0.026676
<(동교동)(명동)>	0.026676
<(당주동)(인천국제공항)>	0.025955
<(명동)(충무로2가)>	0.024513
<(명동)(당주동)>	0.024513
<(대전용전동)(인천국제공항)>	0.023071
<(을지로2가)(명동)>	0.023071
<(명동)(충무로1가)>	0.02235
<(대전용전동)(당주동)>	0.020187
<(신사동)(인천국제공항)>	0.020187

- > 인천국제공항 입국
- > 명동 동일
- > 동교동(홍대) 3.1%
- > 대전 용전동을 들렀다 오는 경우 다수
 - 대전 터미널이 있는 곳 (주변 : 한남대, 신학대 및 산업단지)
- > 당주동(광화문, 청계광장, 이순신동상, 대한민국역사박물관) 다수
 - 다른 나라도 분명히 갔을 것이라 예상되지만 비율상 등장하지 않음 (0.2% 미만 일듯)

- > 인천국제공항 입성
- > 명동 동일
- > 대관령, 남대문 발견
- > 강남과 강북이 비례하게 발견됨

N10

Sequence_name	Ratio
<(명동)(을지로2가)>	0.060761
<(명동)(인천국제공항)>	0.042915
<(을지로2가)(인천국제공항)>	0.038666
<(태평로2가)(인천국제공항)>	0.036116
<(명동)(태평로2가)>	0.033567
<(신사동)(인천국제공항)>	0.03208
<(을지로2가)(명동)>	0.03208
<(삼성1동)(인천국제공항)>	0.030805
<(역삼동)(인천국제공항)>	0.03038
<(을지로2가)(태평로2가)>	0.028681
<(태평로2가)(명동)>	0.028043
<(여의도동)(인천국제공항)>	0.027194
<(인천국제공항)(명동)>	0.025281
<(인천국제공항)(삼성1동)>	0.024219
<(인천국제공항)(태평로2가)>	0.022732
<(역삼1동)(역삼동)>	0.022732
<(인천국제공항)(을지로2가)>	0.022095
<(무교동)(태평로2가)>	0.022095
<(논현동)(역삼동)>	0.02167
<(태평로2가)(을지로2가)>	0.021457
<(한남동)(인천국제공항)>	0.02082
<(서린동)(인천국제공항)>	0.020608
<(강원도대관령)(명동)>	0.020395
<(남대문로5가)(인천국제공항)>	0.020183
<(충무로1가)(인천국제공항)>	0.020183



E_ 결과 해석 > 05_ 구글어스에 시각화

- Google Earth란?

- > 구글이 제공하는 서비스
- > 위성 이미지, 지도, 지형 및 3D 건물 정보 등
전 세계의 지역 정보를 제공, 세계의 여러 지역들을
볼 수 있는 구글의 위성 영상 지도 서비스
- > 2005년6월 28일부터 배포, 30개 이상의 언어로 제공
- > 개발자 도구 중 KML파일이 제공

- KML의 프로젝트 적용

- > 목적 : “앞에서 사용된 Mining 기법들을 통해 도출된 규칙들을 지도상에 표현”
- > P. Compieta et al[4], 데이터 마이닝의 시각화는 마이닝 프로세스의 이해와 분석에 큰 효과

“Visual data mining is a crucial area in explorative data mining, aimed at enhancing the effectiveness of the overall mining process, by supporting analytical reasoning.”

- Keyhole Markup Language (KML)

- > Google 어스, Google 지도 등과 같은 어스 브라우저에 지리데이터를 표시하는 파일형식
- > XML(Extended Markup Language)와 비슷한 구조를 가짐

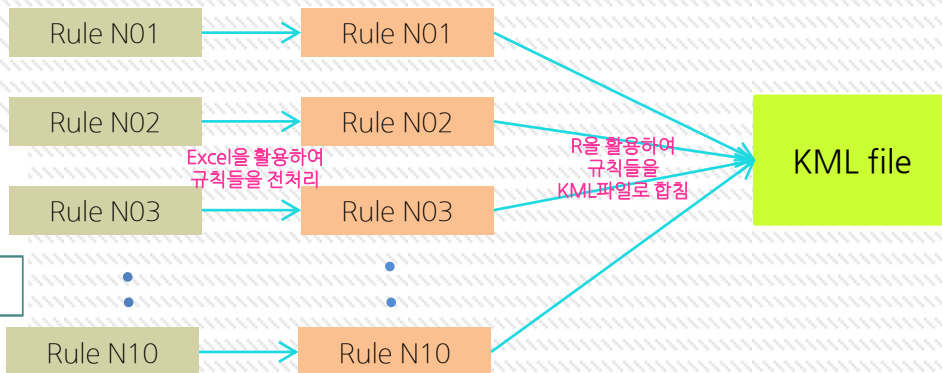
```

<?xml version="1.0" encoding="UTF-8"?>
<kml xmlns="http://www.opengis.net/kml/2.2"> XML 및 KML 선언
  <Placemark>
    <name>New York City</name>
    <description>New York City</description>
    <Point>
      <coordinates>-74.006393, 40.714172, 0</coordinates>
    </Point>
  </Placemark>
</kml>
  
```

지도에 객체표현

점 객체 표시

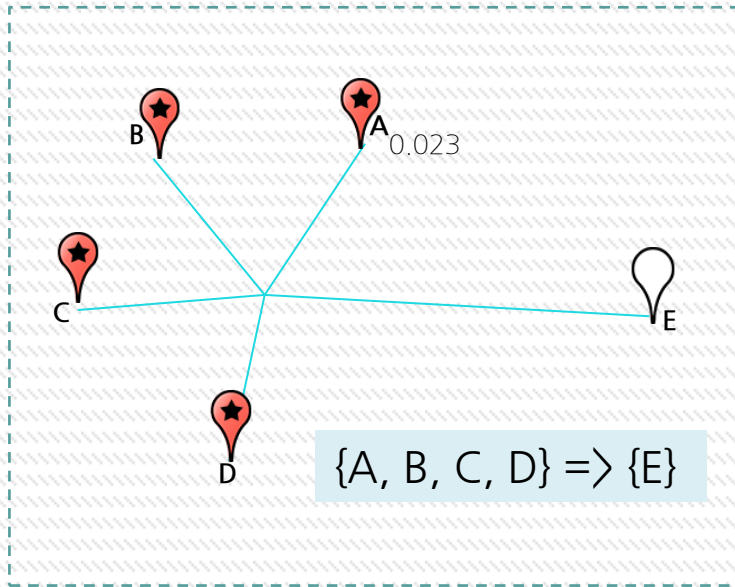
- > KML 파일 생성 프로세스





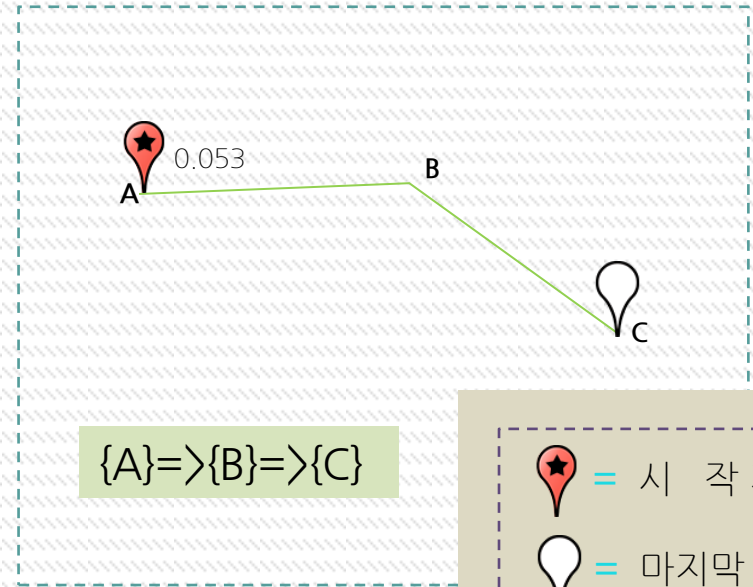
- Association Rule의 표현

> 지도상 위치에 support 값을 표시



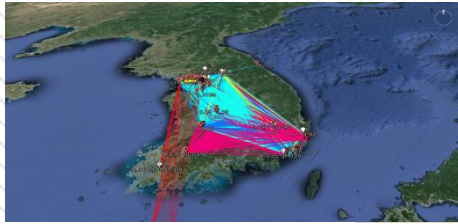
- Sequential Pattern의 표현

> 지도상 위치에 support 값을 표시





- 완성된 KML 파일을 구글어스에 적용 KML 파일로 변환된 국적별 Association Rule



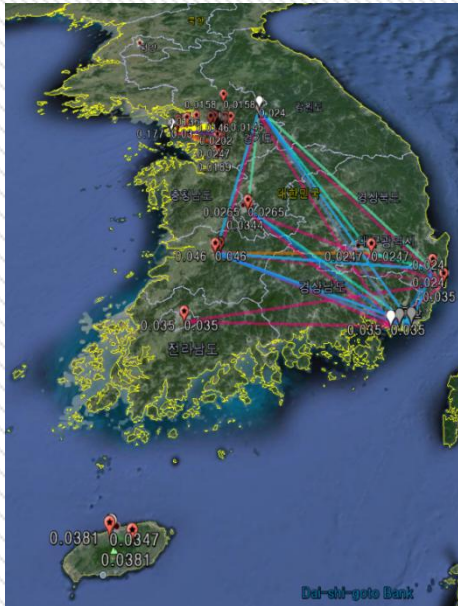
그러나,

너무 많은 규칙으로 알아보기가 힘들

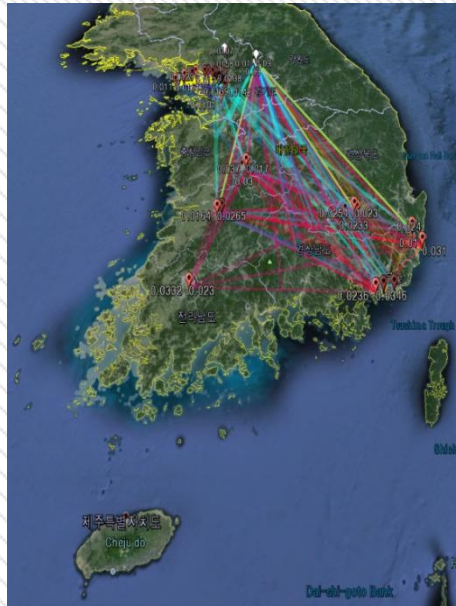
→ 각각 요소 별

(support, confidence, lift)

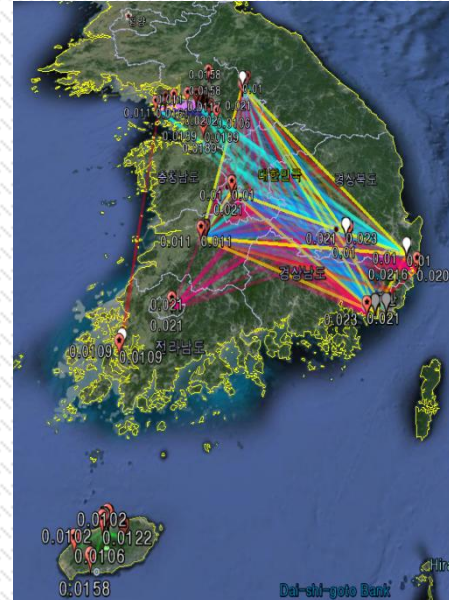
상위규칙 20개만 적용함



<Support 상위 20개 규칙>



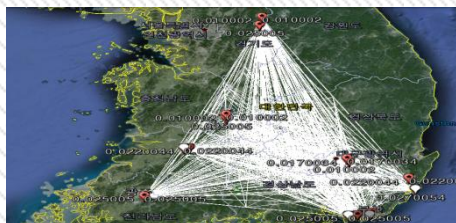
<Confidence 상위 20개 규칙>



<Lift 상위 20개 규칙>



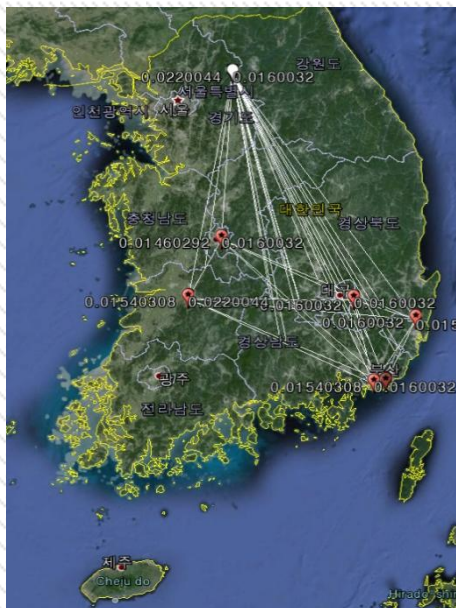
- 완성된 KML 파일을 구글어스에 적용 KML 파일로 변환된 지방 Association Rule



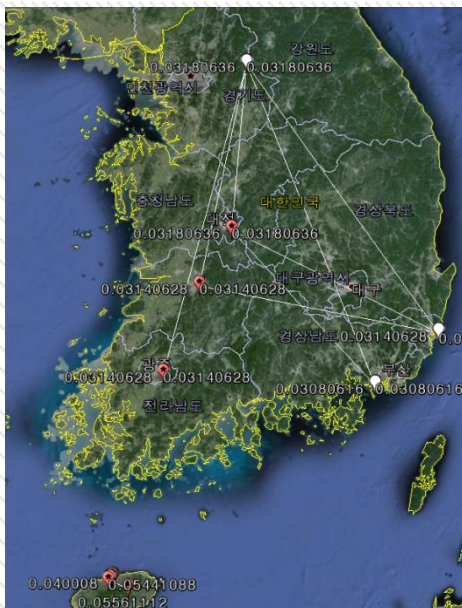
그러나,

너무 많은 규칙으로 알아보기가 힘들

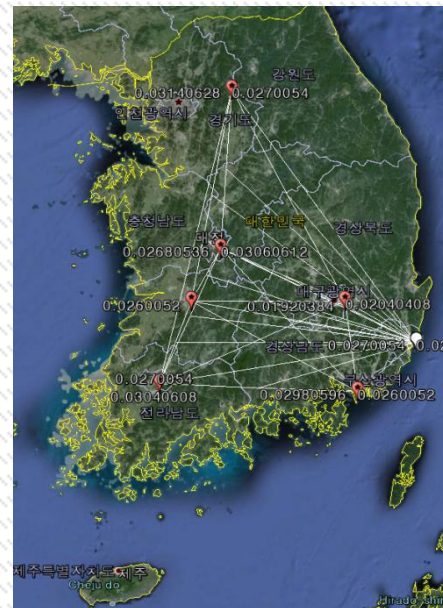
→ 각각 요소 별
(support, confidence, lift)
상위규칙 20개만 적용함



<Support 상위 20개 규칙>



<Confidence 상위 20개 규칙>



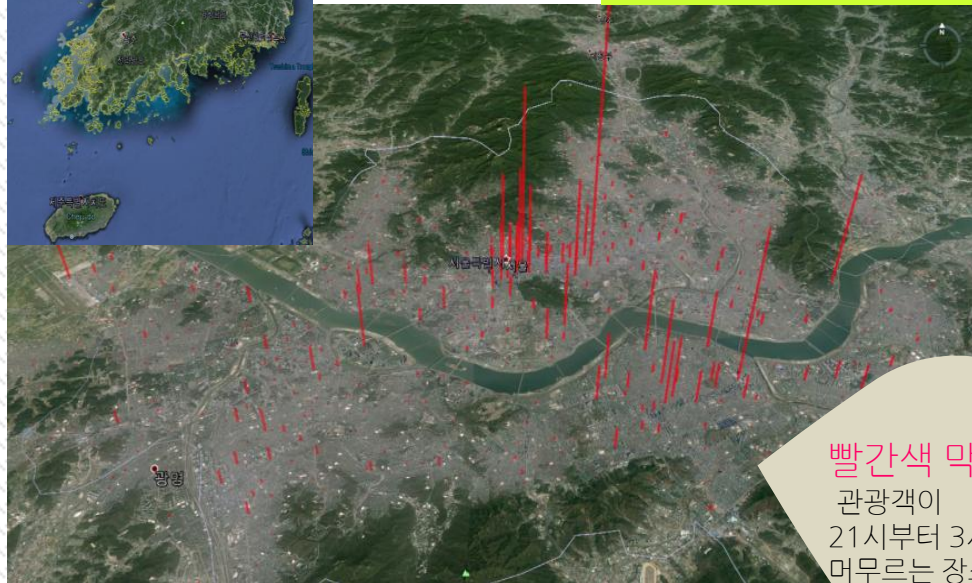
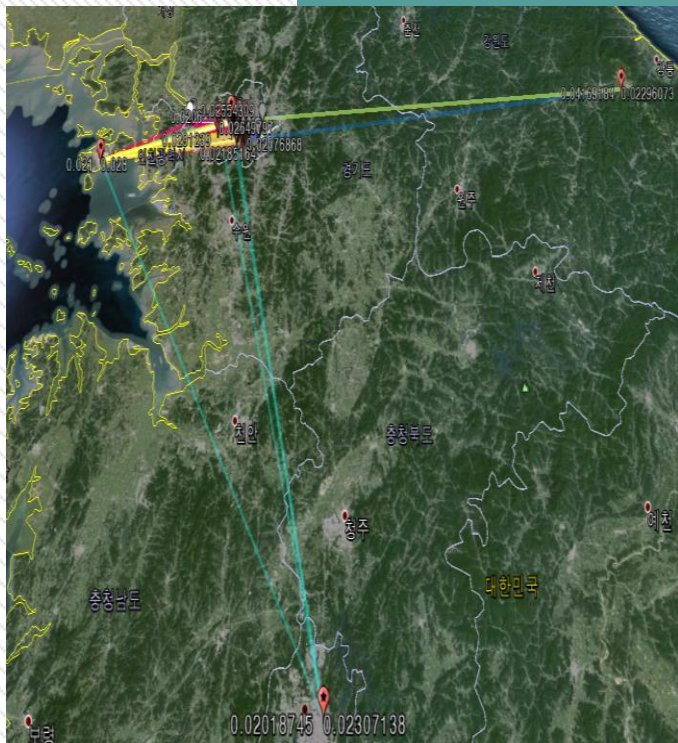
<Lift 상위 20개 규칙>



- 완성된 KML 파일을 구글어스에 적용

KML 파일로 변환된 국적별 Sequential Pattern

KML 파일로 변환된 밤 시간대 데이터



<서울지역만 확대하여 캡처>

빨간색 막대 =
관광객이
21시부터 3시까지
머무르는 장소의 분포

대한민국 관광 3.0

질적 도약 및 균형 발전
솔루션 제안



A_ 과제 파악

B_ 현재 관광산업의 동향 및 이슈 조사

C_ 분석 방향 설정

D_ 분석 시행

E_ 추가 분석

F_ 결과 해석

- 01_ 위성지도를 이용한 빅 데이터 플랫폼 구현
- 02_ 지역관광, 강점에 테마를 더해 체류형 관광을 이끌어낸다.
- 03_ 관광객의 발이 머문 시간, 잠들기 직전까지 대한민국에 반하

G_ 관광마케팅 방안 기획 및 결론/

H_ 참고 문헌

H_ 참고 문헌 > 01_참고 문헌

- [1] John F. Roddick and Brian G. Lees, “Paradigms for Spatial and Spatio-Temporal Data Mining”, Geographic Information Systems, 2001
- [2] Rakesh Agrawal and Ramakrishnan Srikant, “Fast Algorithms for Mining Association Rules”, Proceeding of 20th international conference of large data bases, VLDB, 1994
- [3] Rakesh Agrawal and Ramakrishnan Srikant, “Mining Sequential Patterns”, Proceedings of the 11th International Conference of Data Engineering, 6-10 Mar 1995.
- [4] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci and T. Kechadi, “Exploratory spatio-temporal data mining and visualization”, Journal of Visual Languages and Computing vol18, 255-279, 2007
- [5] 김홍식 외, “도심활성화를 위한 외국인 개별관광객 숙박인프라 확충 방안
Implementation plan of Lodging Facilities for Foreign Independent Travelers”, 정책연구 2012-54, 경기개발 연구원, 2012
- [6] 한국문화관광연구원, 밀워드브라운 미디어리서치, 문화체육관광부, “외래관광객 실태조사 2013(International Visitor Survey)”, 발간번호 11-1371000-000194-10, 승인번호 제 31402호, 문화체육관광부, 2014
- [7] P. Compieta, S. Di Martino, M. Bertolotto, F. Ferrucci and T. Kechadi, “Exploratory spatio-temporal data mining and visualization”, Journal of Visual Languages and Computing vol18, 255-279, 2007

[8] 최영수 저, “쇼핑, 관광, 한류의 최전선 면세점 이야기”, 서울: 미래의 창, 2013

[9] 최정규, 박성원, 정민용, 박정현 저, “죽기 전에 가봐야 할 국내 여행 1001”, 마로니에북스, 2010

[10] 유혜준 저, “남도 섬길 여행”, 미래의 창, 2010

[11] 이하 한국관광공사 제공 데이터, 통계자료

- 2014년 1월, 2월 외래객입국, 국민해외여행객 및 관광수입, 지출 동향
- 2013.12, 2014.01, 2014.02 외래객 입국교통수단별 통계
- 2013.12, 2014.01, 2014.02 주요국가 통계, 세부국가통계
- 2013 외래관광객 실태조사 유저가이드
- 2013 외래관광객 실태조사 코드북
- 2013 관광수입