관광 분야에서
빅데이터가 나아가야 할 길

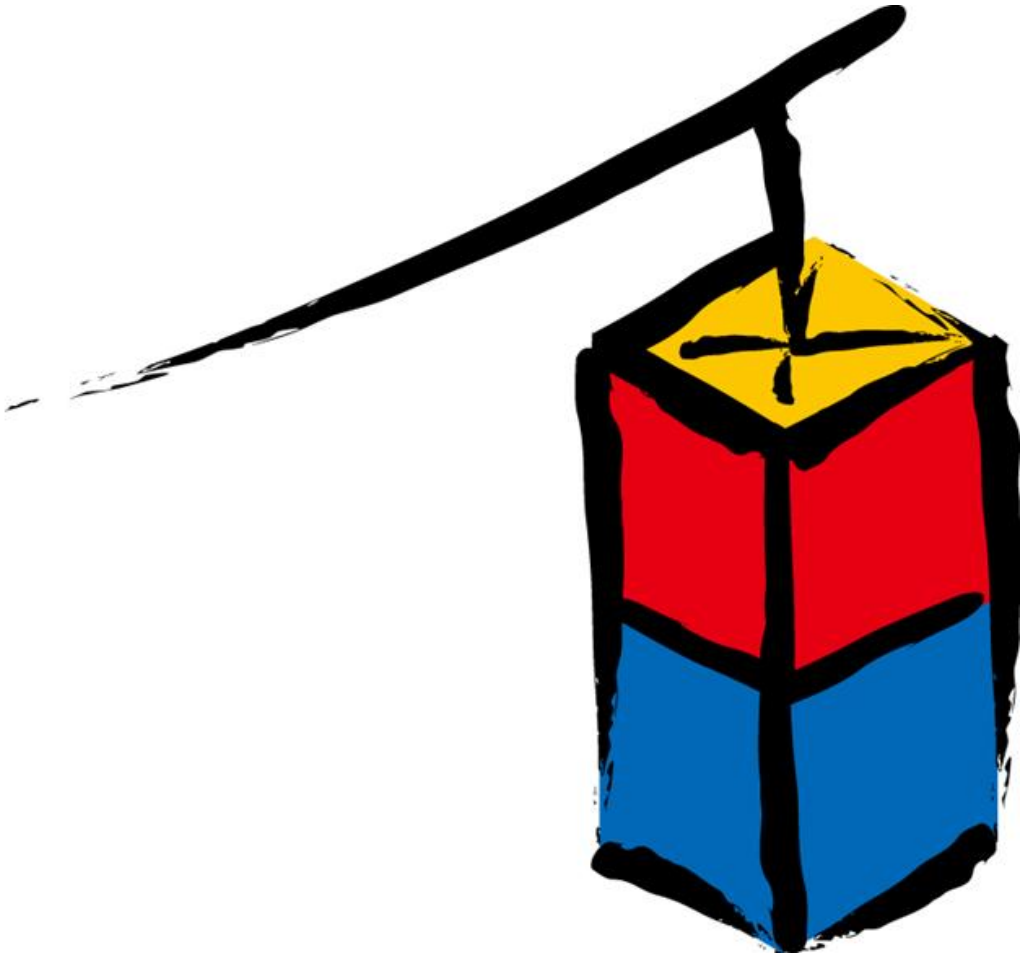
팀 : SA146

목 차

Ⅰ. 문제 도출 및 방향 설정
1) 자료설명
2) 기초분석

Ⅱ. 데이터 분석 및 마케팅 방법
1) 개별, 단체 여행 차이 분석
2) 재방문 예측 모델링
3) 지역별 모델링

Ⅲ. 분석기법의 개선방안
1) 빅데이터 시대에 맞는 데이터 구축
2) GPS, 로그 데이터를 이용한 분석
3) 소셜 네트워크를 이용한 분석



Ⅰ

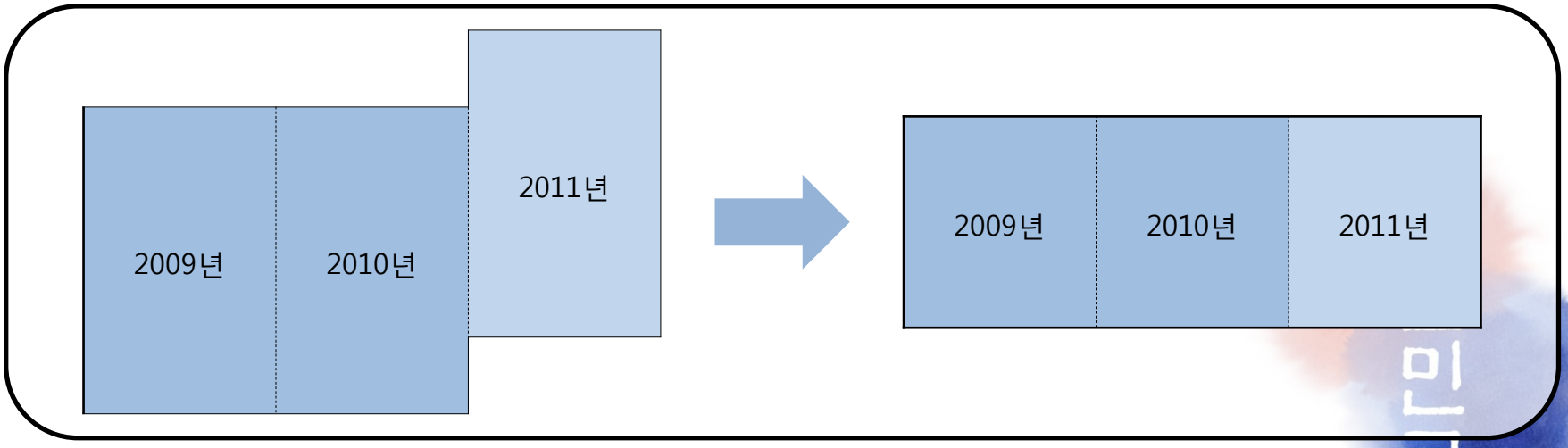
문제 도출 및 방향 설정

I. 문제 도출 및 방향 설정

1) 데이터 구조 및 핸들링

	2009년	2010년	2011년
변수	191개	191개	238개
조사대상 수	11,912명	11,995명	12,038명

- 2009년, 2010년에 사용 된 설문지 동일
- 2011년도에 설문지 내용 변경
  - ✓ 3개 년도 데이터 셋을 하나로 합치는 과정에서 조사된 변수가 달라 문제 발생
  - ✓ 동일한 조건의 새로운 변수들의 생성 필요



I. 문제 도출 및 방향 설정

새로운 변수 - Q2_2. 이전 방문 시 방문한 시기 (분기)

2009년, 2010년 설문지

기존 변수명 – Q1C2

1-2. 최근 한국 방문 시기(이번 여행 제외) : (　　)년 (　　)월

2011년 설문지

기존 변수명 – Q2C

문2-3. **(문2-2에서 '2회 이상'인 경우)**
이번 여행을 제외하고 최근 한국 방문 시기는 언제였습니까? **(20　　)년 (　　)월**

2009년, 2010년, 2011년 통합 설문지

새로 생성 된 변수명 – Q2_2

Q2_1. 이전 방문 시 방문한 시기 (분기)

통합방법

- 2009년과 2010년 설문지는 월 단위 코딩 되어져 있음
- 2011년 설문지에서는 분기 단위로 코딩 되어져 있음
- 3개 년도 통합을 위해 큰 단위인 분기 단위로 통일 함

I. 문제 도출 및 방향 설정

- 앞에서 설명 된 변수 외에도 나머지 변수들에 대해서도 2009년, 2010년, 2011년을 통일 시키기 위한 작업이 필요
- 변경 된 변수들에 대한 자세한 설명은 '코드변환 코드북' 엑셀 파일 참고
- 데이터 정제 과정은 '01. ID 생성 및 데이터 핸들링' SAS 파일 참고

- 2009년, 2010년, 2011년 데이터 셋 통합 전

	2009년	2010년	2011년
변수	191개	191개	238개
조사대상 수	11,912명	11,995명	12,038명

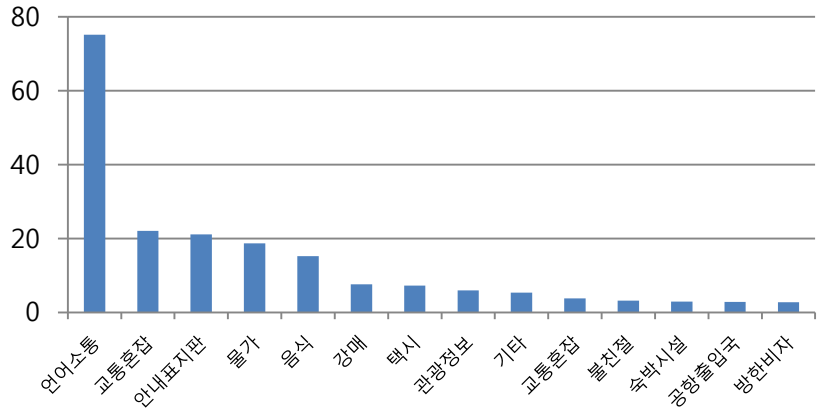
- 2009년, 2010년, 2011년 데이터 셋 통합 후

	통합 데이터 셋
변수	204개
조사대상 수	35,945명

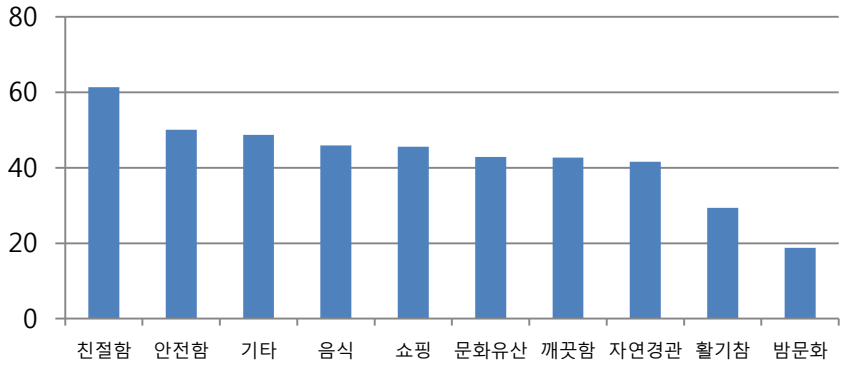
I. 문제 도출 및 방향 설정

2) 기초분석

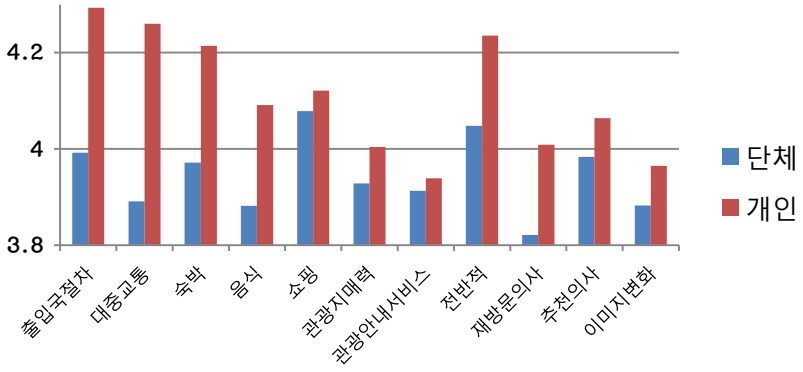
여행시 불편했던 점



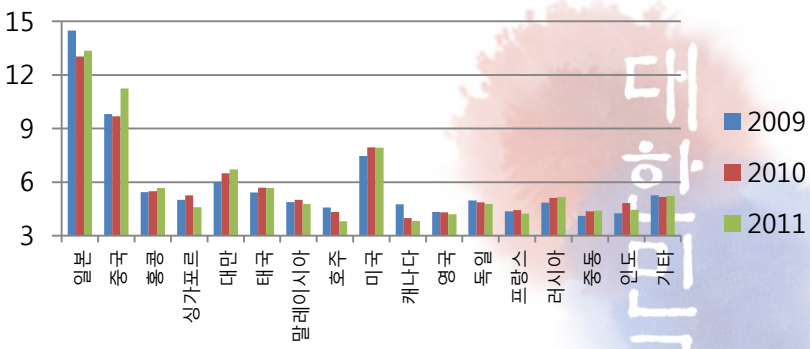
여행시 좋았던 점



한국관광 만족도



연도별 방문자 거주 국

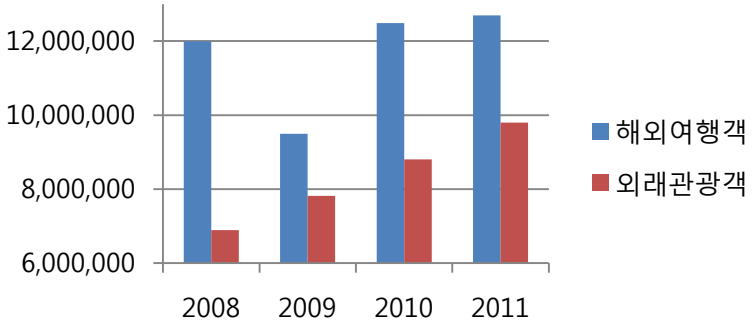


I. 문제 도출 및 방향 설정

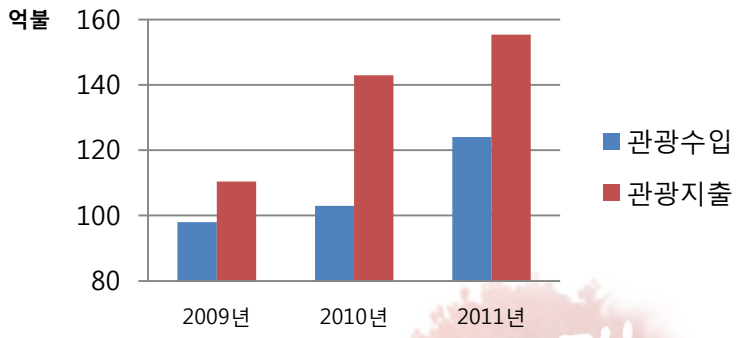
- 2012년 11월 21일 인천공사 입국장에 1,000만 번째 외래 관광객이 입국
- 외래관광객 1,000만 명 돌파는 지난 1978년 100만 명을 기록 후 34년 만에 이룬 쾌거
- 여세를 몰아 5,000만 명, 10,000만 명 시대를 열기 위해서는 해결해야 할 문제

① 여행수지 적자

연도별 입.출국 수



연도별 총 사용금액



— 2000년대 한류열풍 이후 꾸준히 외래관광객이 증가하고 있으나 여행수지 적자는 계속 진행 중

② 관광객들의 서울 집중 현상

③ 재방문 관광객수 감소

Ⅱ

데이터 분석 및 마케팅 방법

1. 여행수지적자

개별방문객 유치

II. 데이터 분석 및 마케팅 방법

1) 여행수지 적자 - 개별 방문 여행객 유치

신문 기사

[문화] 외래관광객 2000만 시대로

게재 일자 : 2012년 11월 21일(水)

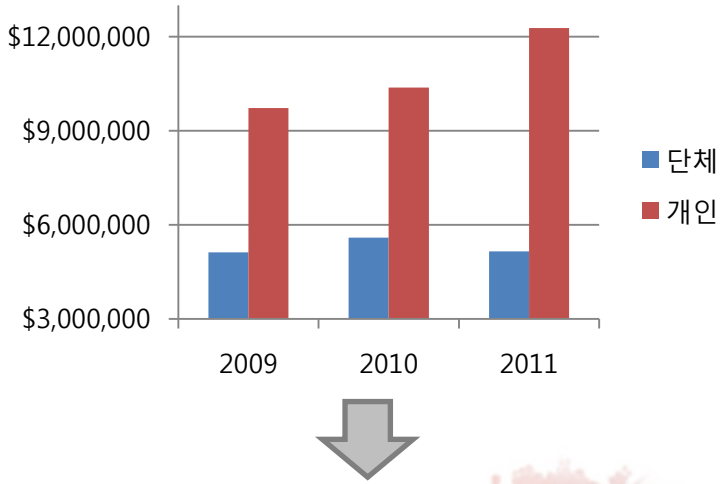
국격 걸맞은 '高품격 상품' 개발… 단체보다 개별여행객 늘려야

〈下〉풀어야 할 숙제

6·25전쟁 직후인 1955년 6월 15일 프랑스 국적의 크루즈선박 '라오스' 호를 타고 부산항으로 입항했던 '첫 공식 외래 관광객'을 유치했을 때 관광수입은 얼마나 됐을까. 이날 외국인 관광객들은 동해남부선 열차로 부산에서 송정까지 해안선을 관광한 뒤 해운대와 동래온천 등의 명승지를 돌아보고 돌아갔다. 그러나 이들을 유치해서 얻은 총 수익은 고작 130달러에 불과했다. 크루즈 승객 138명 중 부산에서 내려 관광에 나선 외국인이 31명에 불과했던데다 이들에게 판매할 만한 관광 상품은 물론이고 변변한 관광안내서나 엽서도 없었다. 그러나 57년이 지난 올해 한국이 외래 관광객 유치를 통해 얻는 관광수익은 150억 달러에 육박할 것으로 기대된다. 한국을 찾아오는 외래 관광객들로부터 매일 400억 원이 넘는 수입을 올린다. 2010년부터 시작된 '한국방문의 해' 사업의 성공적인 마무리를 앞두고 있는 상황에서 이제 앞으로의 과제도 대두되고 있다. 여행상품 고급화, 지방관광 활성화, 재방문 확대 등 외래 관광객 2000만명 시대를 대비하는 전략은 어떻게 마련되고 있을까.

http://www.munhwa.com/news/view.html?no=2012112101033130024002 00

년도 별 개인, 단체 지출금액



- 기사와 같이 분석용 데이터에서도 개별 방문이 단체 여행에 보다 훨씬 많은 돈을 지출하는 것을 볼 수 있음
- 우리나라 관광산업의 여행수지 적자를 개선하기 위하여 개별 방문 관광객 유치에 필요한 마케팅이 필요

II. 데이터 분석 및 마케팅 방법

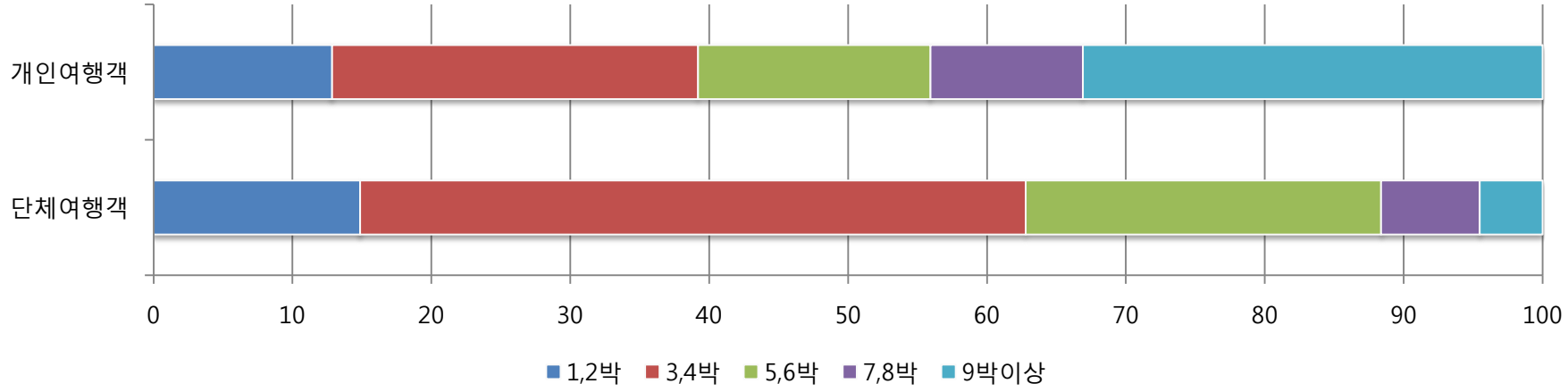
- 개별, 단체 여행 차이 분석

거주국 별 개별, 단체

- 아시아 지역의 경우 단체여행의 비율이 높은 것으로 나타났으나 그 외 지역에서는 개별 방문의 비율이 높은 것으로 나타남
- 여행수지 적자 개선을 위해서는 지출이 많은 개별 방문자를 유치해야 하므로 미국과 유럽지역을 중점으로 개별 방문 시의 장점과 대중교통의 편리함 등을 홍보
- 홍보 시 앞 슬라이드에서 제공한 만족도 조사 결과를 마케팅 대상자들에게 제공하여 개별 여행의 장점을 적극 홍보

II. 데이터 분석 및 마케팅 방법

체류기간 별 개별, 단체 비교



- 개별 여행객의 경우 단체 여행객에 비해 체류기간이 긴 것을 확인 할 수 있음
- 여러 나라에 가능성을 두고 개별 여행을 계획하고 있거나 여러 나라를 방문 할 계획이 있는 외국인들에게 장기 투숙에 대한 할인권을 제공하여 다른 나라가 아닌 우리나라만을 방문하거나 다른 나라에 비해 더 오랜 기간 방문하도록 유도
- 장기적인 방문기간 동안 지속적인 재미를 선사하기 위한 관광지 개발을 시급하며 이를 홍보 할 수 있는 시스템의 구축이 급선무

2. 재방문객 유치

재방문을 위한 모델링

II. 데이터 분석 및 마케팅 방법

2) 재방문 관광객 유치 – 재방문 모델링

신문 기사

[외국인 관광객 1000만 시대] 2000만 '관광부국'으로 가려면

관련이슈 : 세계일보 창간 24주년 특집, 외국인 관광객 1000만 시대

04월 18일자 지면보기

<<< [오늘의 핫이슈] '전의전쟁'출연 "수입 120억 넘어" 깜짝

의료·한류·웨딩… 고급콘텐츠로 외국인 발길 잡아라
고부가 상품으로 선택의 폭 넓혀라
내국인 관광활성화 선결 조건

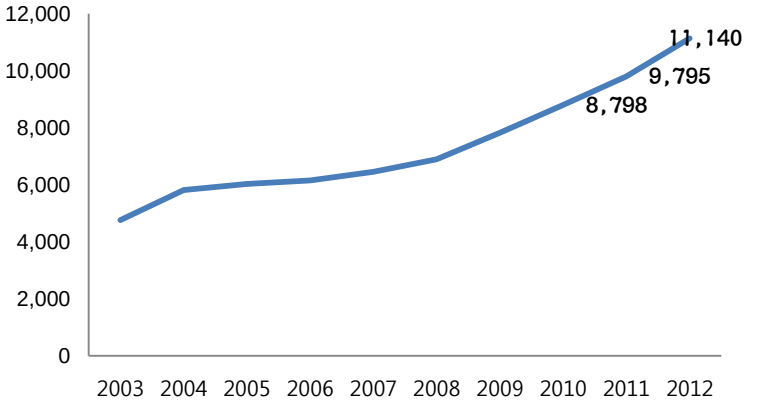
한국男性 성관계 시 평균삽입 시간? [정보]고혈압완치..알고보니 간단해!

지난해 외래관광객 1000만명 달성에 고무된 정부와 여행업계는 2020년 2000만명을 달성하겠다는 의욕적인 목표를 세웠다. 그러나 관광상품의 다양성 부족과 외래관광객의 서울 집중, 저조한 재방문율 등의 난제를 해결해 환골탈태하지 않으면 외래관광객 2000만명 달성은 불가능해 보인다. 전문가들의 견해를 종합해 한국이 명실상부한 관광대국으로 부상하기 위해서는 어떤 대책을 마련하고 어느 분야에 힘을 쏟아야 하는지 살펴본다.

http://www.segye.com/Articles/NEWS/CULTURE/Article.asp?aid=20130131023295&subctg1=&subctg2=&OutUrl=naver

년 단위 방문객 수

출처 : 나라지표



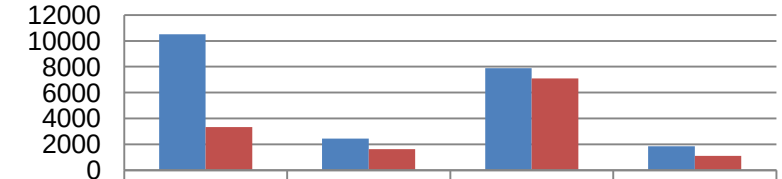
- 방문객의 수는 꾸준하게 증가하고 있지만 재방문을 하는 관광객의 수는 줄어들고 있음
- 재방문 고객을 유치하기 위한 마케팅 방법이 필요함

II. 데이터 분석 및 마케팅 방법

▪ 재방문 예측 모델링_방문목적 : 여가

- 한국 방문의 주된 목적이 여가가 아닌 사업등과 같은 다른 목적으로 방문한 관광객의 빈도가 매우 높게 나타남
- 재방문 예측 모델링에 있어 한국 방문의 주된 목적이 여가인 경우만 따로 뽑아 모델링을 실시할 필요성이 있다고 판단

한국 방문 주된 목적

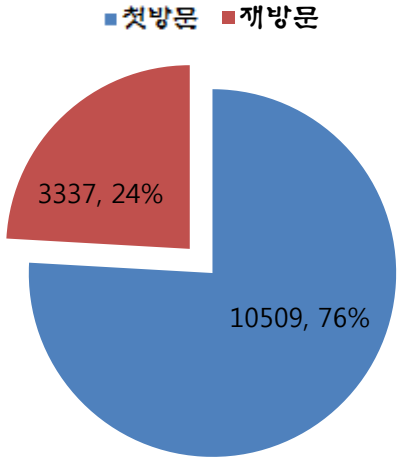


	여가	친구방문	전문활동	기타
첫방문	10509	2432	7889	1843
재방문	3337	1618	7099	1106

재방문 * 한국방문 목적

통계량	자유도	값	확률
카이제곱	3	1698.602	<.0001

재방문 빈도

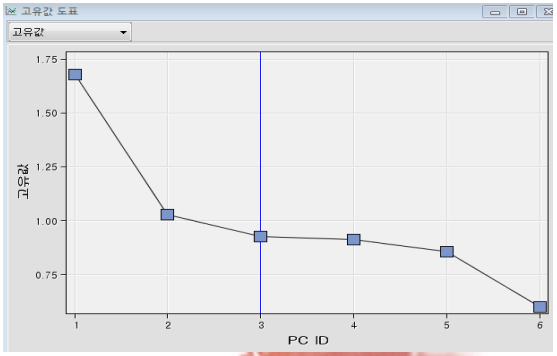


- 한국 방문 주된 목적이 여가인 관광객 중 최근 3년간 한국에 방문 여부로 Target을 설정

II. 데이터 분석 및 마케팅 방법

- 변수선택
  - ✓ 파생변수 & 타겟 변수 생성
  - ✓ 데이터 탐색 & 변수중요도 파악
  - ✓ 변수선택 (교차분석, 상관분석, 변수 선택노드 사용)
  - ✓ 주성분 분석을 통한 파생변수 생성

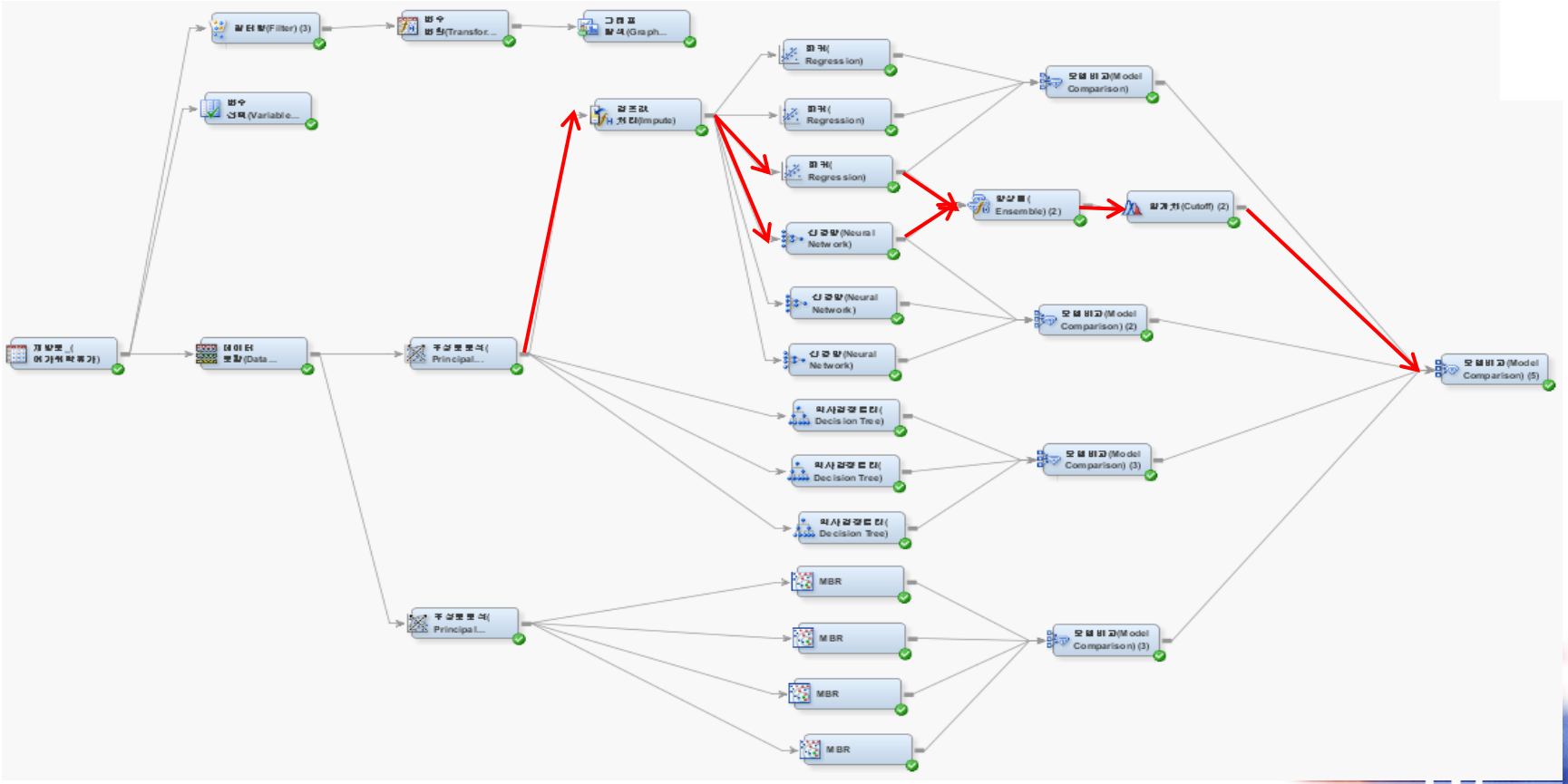
변수명	변수 설명	Prin1	Prin2	Prin3
Q39	쇼핑비	0.2103	0.7666	0.2389
Q40	식음료비	0.5736	0.2057	-0.0998
Q41	교통비	0.5163	-0.1226	-0.2345
Q42	오락관련지출	0.3294	-0.2642	-0.5741
Q44	문화관련지출	0.4256	0.0300	0.3322
Q45	운동관련지출	0.2656	-0.5332	0.6619
		식도락 소비	레져형소비	쇼핑형 소비



- 주성분 분석을 실시한 결과 주성분의 개수는 3개로 정하고, 의미를 파악해보면 제 1 주성분은 "전반적인 소비" 제 2 주성분은 쇼핑소비 관련지출이 높은 "쇼핑형 소비"형태를 보이고, 제 3주성분은 운동과 문화에 지출하는 "레져형 소비"로 구분됨을 볼 수 있다.

II. 데이터 분석 및 마케팅 방법

- 최종 선택 모형



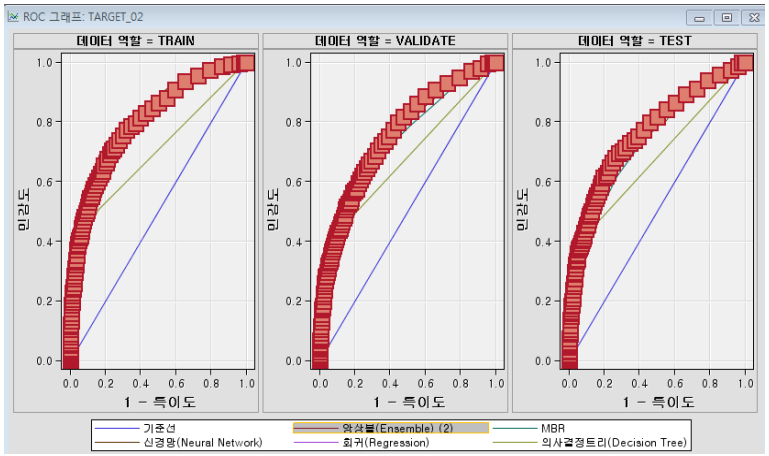
✓전체적인 모형에서 회귀분석 노드가 가장 높은 예측율을 보였다.

II. 데이터 분석 및 마케팅 방법

- 최종 선택 모델

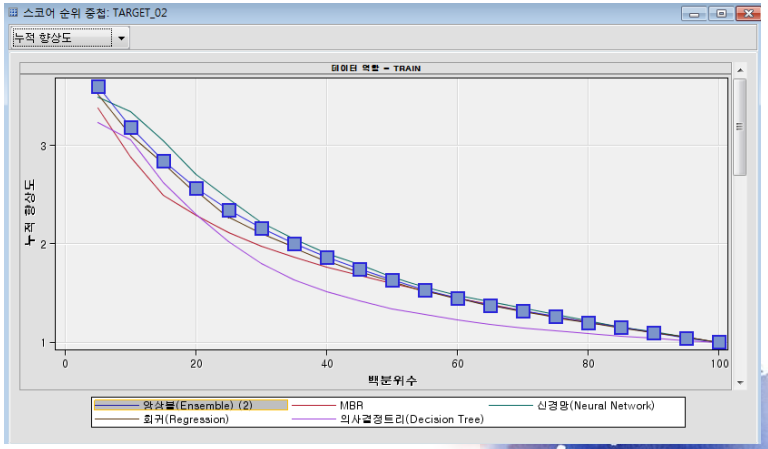
적합통계량

선택된 모델	선행 노드	모델 노드	모델 설명	분석: 오분류 비율	검증: 오분류 비율	평가: 오분류 비율	분석: Roc 인덱스	검증: Roc 인덱스	평가: Roc 인덱스
Y	MdlComp18	Tree10	의사결정트리(Decision Tree)	0,192669	0,199567	0,206307	0,69	0,668	0,666
	Ensmbl2	Ensmbl2	앙상블(Ensemble) (2)	0,182196	0,193548	0,20727	0,805	0,761	0,758
	MdlComp17	Neural11	신경망(Neural Network)	0,173709	0,206548	0,20727	0,824	0,737	0,744
	MdlComp16	Reg11	회귀(Regression)	0,187071	0,201011	0,211603	0,796	0,758	0,755
	CUT2	Reg12	회귀(Regression)	0,185988	0,197882	0,212325	0,796	0,759	0,756
	MdlComp19	MBR13	MBR	0,200072	0,215936	0,219066	0,78	0,723	0,716



적합통계량

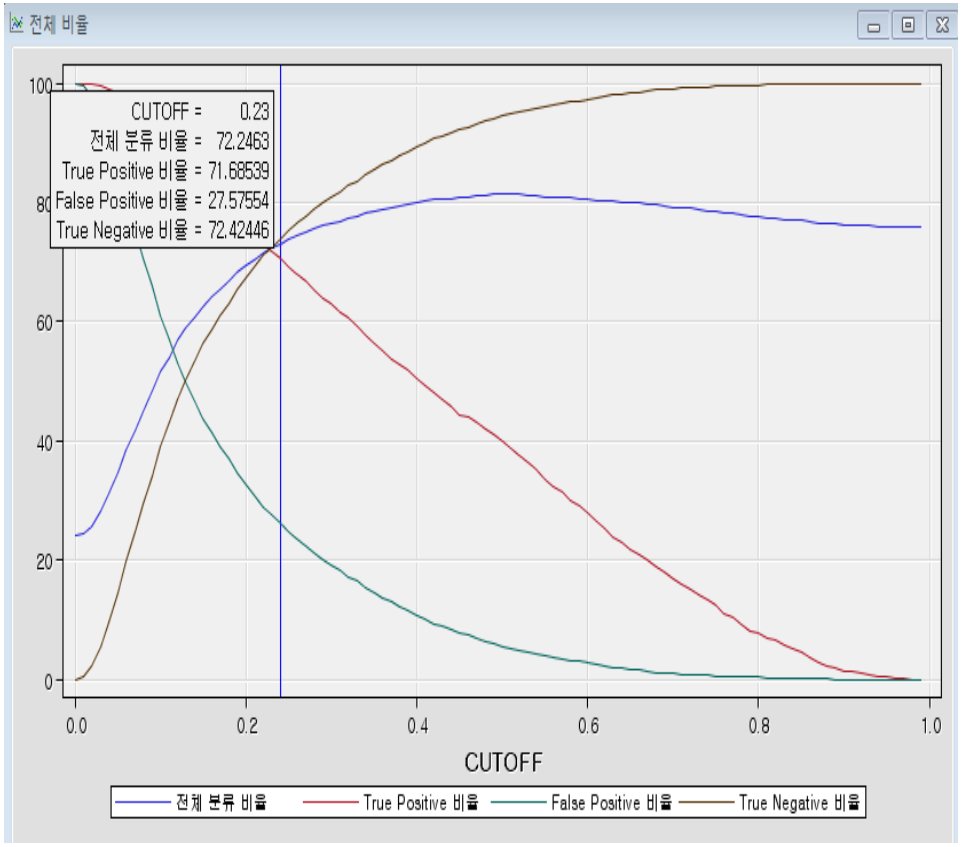
타겟	적합통계량	통계량 레이블	분석	평가	검증
TARGET_02	_ASE_	평균제곱오차	0,13537	0,14992	0,145644
TARGET_02	_DIV_	ASE의 분모	11076	8308	8308
TARGET_02	_MAX_	최대절대오차	0,965742	0,970476	0,972996
TARGET_02	_NOBS_	빈도 합계	5538	4154	4154
TARGET_02	_RASE_	제곱근평균제곱오차	0,367927	0,387196	0,381633
TARGET_02	_SSE_	오차제곱합	1499,363	1245,539	1210,009
TARGET_02	_DISF_	분류된 관측치 빈도	5538	4154	4154
TARGET_02	_MISC_	오분류 비율	0,182196	0,20727	0,193548
TARGET_02	_WRONG_	잘못된 분류 수	1009	861	804

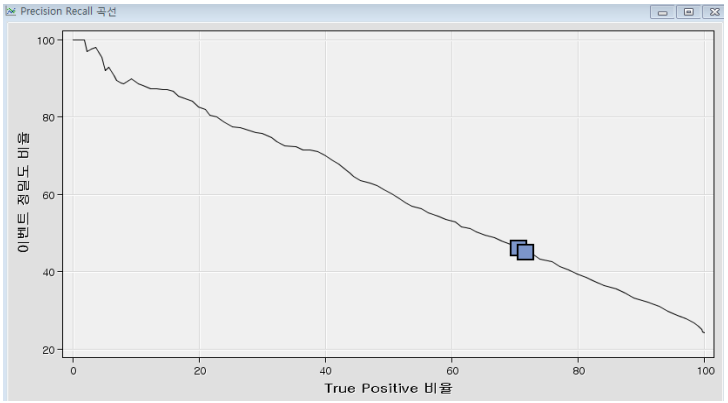


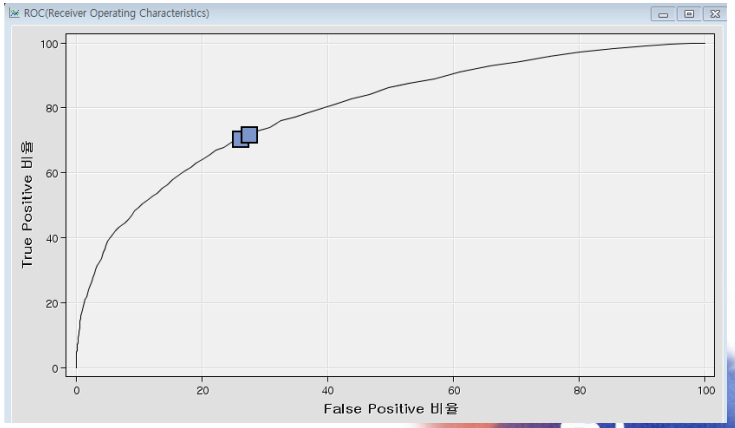
✓모델선택 노드에서 오분류 비율과 ROC 인텍스값 등을 통해 앙상블 노드를 선택하였다.

II. 데이터 분석 및 마케팅 방법

- Cut off 설정







✓ K-S통계량 값을 이용하여 Cut off(0.23)를 설정 하였다.

II. 데이터 분석 및 마케팅 방법

- Score

	Impute...	Warnings	Residu...	Unnor...	Residu...	Into: TARGET_02	From: TARGET_02	Predicted: TARGET_02=1	Predicted: TARGET_02=0	EM_CUTOFF
6	4		0.094766	1	-0.094771		1	0.905234	0.094766	1
3	5		0.240503	1	-0.24051		1	0.759497	0.240503	1
2	3.285714		-0.89772	1	0.8977221		0	0.897722	0.102278	1
3	3.571429		0.260361	1	-0.260361		1	0.739639	0.260361	1
2	4.166667		-0.36482	0	0.3648170		0	0.364817	0.635183	1
2	3.857143		-0.35536	0	0.3553610		0	0.355361	0.644639	1
0	3.428571		-0.03877	0	0.0387660		0	0.038766	0.961234	0
0	3.142857		-0.07482	0	0.0748150		0	0.074815	0.925185	0
1	3.428571		-0.05532	0	0.0553230		0	0.055323	0.944677	0
1	4.142857		-0.21605	0	0.2160520		0	0.216052	0.783948	0
1	3.571429		-0.02963	0	0.0296290		0	0.029629	0.970371	0
2	5		-0.30702	0	0.3070180		0	0.307018	0.692982	1
1	4.714286		0.890285	0	-0.890290		1	0.109715	0.890285	0
1	3.571429		-0.06868	0	0.0686790		0	0.068679	0.931321	0
1	4.857143		-0.08058	0	0.0805770		0	0.080577	0.919423	0
3	4		-0.76185	1	0.7618451		0	0.761845	0.238155	1
1	3.714286		-0.22952	0	0.2295190		0	0.229519	0.770481	0
2	4		0.322926	1	-0.322931		1	0.677074	0.322926	1
1	3.714286		-0.57276	1	0.5727631		0	0.572763	0.427237	1
1	4		-0.19985	0	0.199850		0	0.19985	0.80015	0
0	4.857143		-0.08973	0	0.0897280		0	0.089728	0.910272	0
2	4.428571		0.339237	1	-0.339241		1	0.660763	0.339237	1
1	3.833333		0.513602	0	-0.51360		1	0.486398	0.513602	1
1	4.571429		-0.36097	0	0.3609740		0	0.360974	0.639026	1
1	3.833333		-0.10818	0	0.1081830		0	0.108183	0.891817	0
1	4.142857		-0.03053	0	0.030530		0	0.03053	0.96947	0
2	4.428571		-0.01388	0	0.013880		0	0.01388	0.98612	0
0	3.857143		-0.2237	0	0.2237030		0	0.223703	0.776297	0
2	4		-0.15995	0	0.1599470		0	0.159947	0.840053	0
2	4.142857		-0.14726	0	0.1472640		0	0.147264	0.852736	0
1	4		-0.07689	0	0.0768910		0	0.076891	0.923109	0
1	4.428571		-0.19022	0	0.1902150		0	0.190215	0.809785	0
9	4.333333		0.931222	0	-0.931220		1	0.068778	0.931222	0
1	4.285714		-0.18007	0	0.1800680		0	0.180068	0.819932	0
1	4		-0.38425	0	0.3842530		0	0.384253	0.615747	1
1	3.428571		0.676445	0	-0.676450		1	0.323555	0.676445	1
2	4		0.540844	0	-0.540840		1	0.459156	0.540844	1
7	3.857143		0.417118	1	-0.417121		1	0.582882	0.417118	1
1	3.714286		-0.10055	0	0.1005550		0	0.100555	0.899445	0
1	4		-0.11803	0	0.1180270		0	0.118027	0.881973	0
0	3.714286		-0.08534	0	0.0853390		0	0.085339	0.914661	0
2	3.428571		-0.14146	0	0.1414570		0	0.141457	0.858543	0
1	3.142857		-0.12028	0	0.1202790		0	0.120279	0.879721	0
0	2.857143		-0.08464	0	0.0846430		0	0.084643	0.915357	0
0	3.571429		-0.11202	0	0.1120160		0	0.112016	0.887984	0
7	3.571429		-0.01387	0	0.0138690		0	0.013869	0.986131	0
0	3.714286		-0.0364	0	0.03640		0	0.0364	0.9636	0
0	4.428571		-0.06168	0	0.061680		0	0.06168	0.93832	0
0	4.428571		-0.06765	0	0.0676510		0	0.067651	0.932349	0

✓ Cut-off를 사용하여 23% 이상의 확률을 가지면 재방문자라고 선정하였다.

- 이벤트 분류

FALSE Negative	TRUE Negative	FALSE Positive	TRUE Positive
357	2436	644	717

II. 데이터 분석 및 마케팅 방법

등급	빈도			반응률		
	관측치 수	이벤트 수	비이벤트 수	%Response	Lift	%Captured
1	416	301	115	72.4%	3.00	30.1%
2	831	479	352	57.6%	2.39	47.9%
3	1247	610	637	48.9%	2.03	60.9%
4	1662	709	953	42.7%	1.77	70.8%
5	2077	779	1298	37.5%	1.56	77.8%
6	2493	832	1661	33.4%	1.38	83.1%
7	2908	891	2017	30.6%	1.27	89.0%
8	3324	933	2391	28.1%	1.16	93.2%
9	3739	968	2771	25.9%	1.07	96.7%
10	4154	1001	3153	24.1%	1.00	100.0%
전체	Base Line %Response =24.1%					

✓ 외국인 관광객 중 관광목적이 여가인 관광객들의 상위 30%를 목표고객으로 하여 캠페인을 수행할 경우, 반응률이 48.9%이며 따라서 포획 가능한 반응고객의 60.9%에 이를 것으로 기대된다.

II. 데이터 분석 및 마케팅 방법

- 우리나라에 재 방문하는 관광객의 특성
  - ✓ 40~60세
  - ✓ 대졸
  - ✓ 1~2박
  - ✓ 개인관광객
  - ✓ 서울방문
  - ✓ 일본 중국 홍콩 싱가포르 미국 캐나다 프랑스
  - ✓ 지출금액 높음

- 한 번의 재 방문은 모델링 결과로 이끌어 낼 수 있음
- 하지만 관광객의 2차, 3차 재방문으로 이끌기 위해서는 다양한 문화행사 및 관광지 개발이 필요
- 매년 초대가수의 명단만 바뀌는 문화행사가 아닌 다양한 스토리가 필요

3. 서울집중현상

지역별 모델링

II. 데이터 분석 및 마케팅 방법

3) 관광객들의 서울 집중 현상 해결 – 지역별 모델링

신문 기사

메일보내기 오류신고 프린트 가 가

외국인 관광객 80% 서울 집중, 인프라 확충 시급

서울 방문 10년새 2배, 소규모 비즈니스호텔 건립 우후죽순

2012년 06월 05일 (화) 10:05:14 이계덕 기자 dirpejr@seoultimes.net

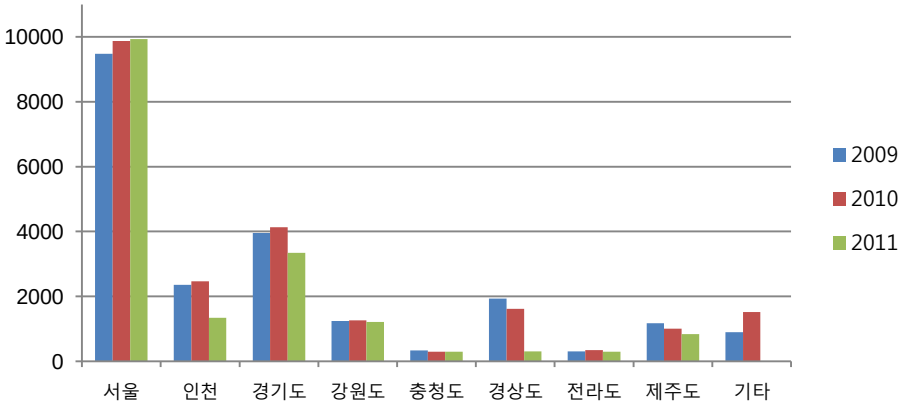
▲세종로 광화문광장에서 나들이 나온 외국인 관광객들이 경찰기마대와 함께 사진촬영을 하며 즐거워 하고 있다.[사진=뉴시스]

지난 2002년부터 2011년까지 서울시를 방문한 외국인 관광객의 수 방문동향을 조사한 결과 10년 전보다 서울을 방문하는 관광객 수가 두 배 가까이 늘어난 것으로 나타났다. 하지만 급증하는 외국인 관광객을 수용할 인프라는 수요에 미치지 못해 대책 마련이 시급하다.

한국관광공사 '방한 외래관광객 통계' 및 한국문화관광연구원 '외래관광객 실태조사'에 따르면 지난 2002년 449만 명이던 외국인 관광객이 지난해 780만 명으로 두 배 가까이 늘었다. 지난해 우리나라를 찾은 외국인 관광객은 979만 명으로 이 중 79.6%가 서울에 집중됐다.

http://www.seoultimes.net/news/articleView.html?idxno=19020

지역별 방문객 수

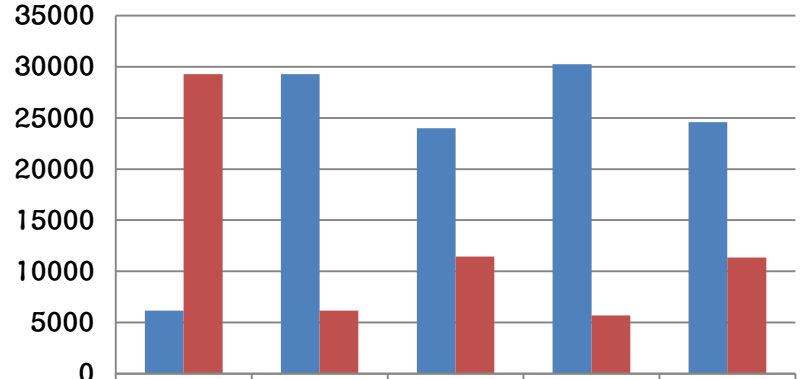


- 기사와 같이 분석용 데이터에서도 관광객들의 방문지역이 서울에 편중되어짐을 확인
- 우리나라의 장기적인 관광산업 발전을 위해서는 서울 외 지역에서의 관광객 유치가 필요함

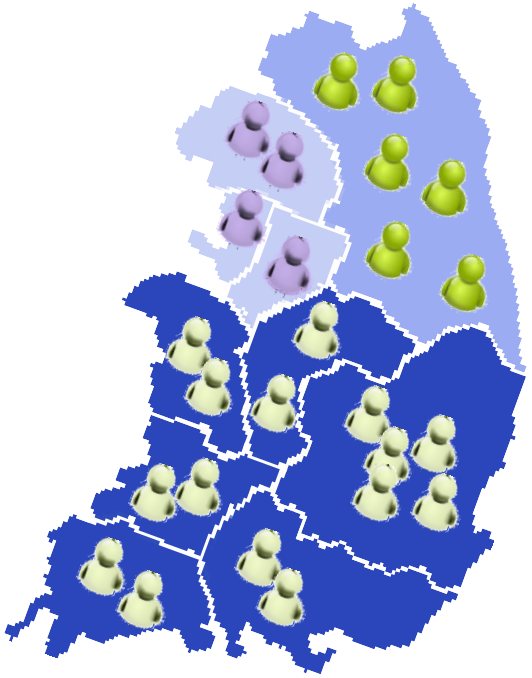
II. 데이터 분석 및 마케팅 방법

- 방문 여부에 따른 지역별 빈도표

지역별 방문 여부



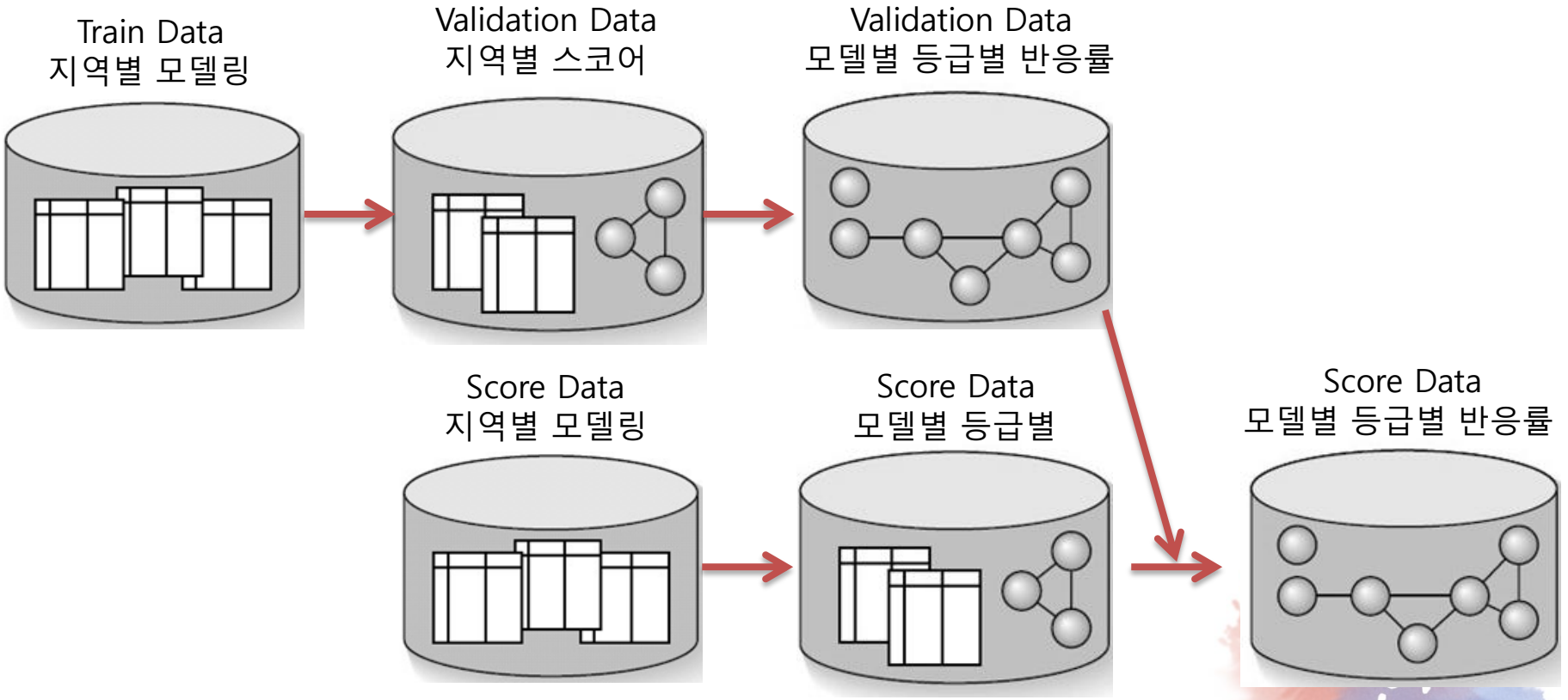
	서울	인천	경기도	강원도	남부지역
미방문	6149	29280	23996	30250	24600
방문	29286	6154	11438	5695	11345



- ✓ 지역별 방문 여부를 확인해보면 서울지역을 방문한 여부가 매우 높음을 볼 수 있으며 서울지역을 제외한 타 지역은 방문여부가 매우 낮음
- ✓ 외국인 관광객들을 지방으로 분산시켜 서울편중을 해소하고, 관광객들에게 만족도를 높이기 위해 지역별 모델링을 통해 개인 맞춤형 관광 추천 시스템 개발

II. 데이터 분석 및 마케팅 방법

각 모델 별 반응률을 고려한 지역 추천

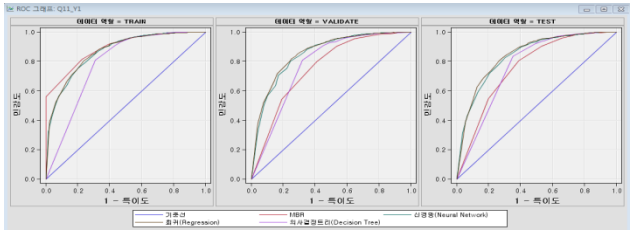


- 지역별 스코어를 비교하기 위해 등급별 반응률을 고려한 스코어 값 생성

II. 데이터 분석 및 마케팅 방법

- 모델링 결과

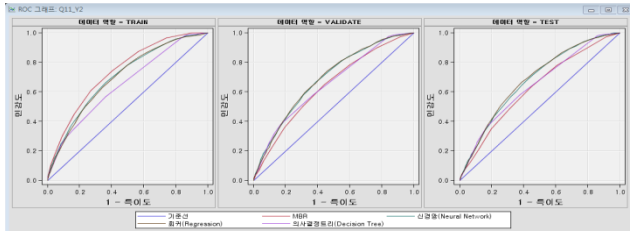
✓ 서울



적합통계량

선택된 모델	선행 노드	모델 노드	모델 설명	타겟 변수	분석: 오분류 비율	평가: 오분류 비율	검증: 오분류 비율	분석: Roc 인덱스	검증: Roc 인덱스	평가: Roc 인덱스
Y	MdlComp7	Neural5	신경망(Ne,,,	Q11_Y1	0,058813	0,061761	0,064609	0,867	0,838	0,85
	MdlComp6	Reg5	회귀(Regre,,,	Q11_Y1	0,062251	0,064415	0,063404	0,868	0,842	0,857
	MdlComp8	Tree5	의사결정트,,,	Q11_Y1	0,059356	0,064897	0,064609	0,785	0,765	0,775
	MdlComp9	MBR6	MBR	Q11_Y1	0,061165	0,072237	0,067356	0,892	0,759	0,755

✓ 인천



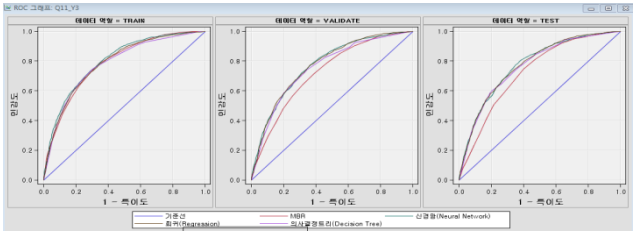
적합통계량

선택된 모델 ▼	선행 노드	모델 노드	모델 설명	타겟 변수	분석: 오분류 비율	평가: 오분류 비율	검증: 오분류 비율	분석: Roc 인덱스	검증: Roc 인덱스	평가: Roc 인덱스
Y	MdlComp13	Tree8	의사결정트,,,	Q11_Y2	0,172241	0,173001	0,174944	0,649	0,645	0,652
	MdlComp12	Neural8	신경망(Ne,,,	Q11_Y2	0,172806	0,173754	0,17532	0,7	0,675	0,686
	MdlComp11	Reg10	회귀(Regre,,,	Q11_Y2	0,174358	0,173942	0,174003	0,696	0,679	0,687
	MdlComp14	MBR9	MBR	Q11_Y2	0,172664	0,187628	0,186428	0,736	0,623	0,637

II. 데이터 분석 및 마케팅 방법

- 모델링 결과

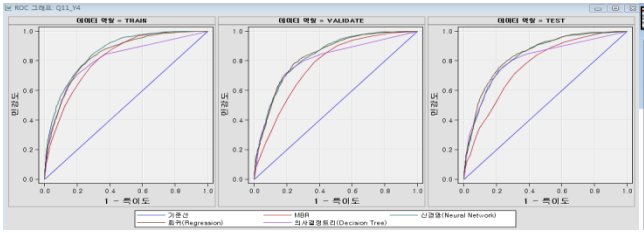
✓ 경기도



적합통계량

선택된 모델	선행 노드	모델 노드	모델 설명	타겟 변수	분석: 오분류 비율	평가: 오분류 비율	검증: 오분류 비율	분석: Roc 인덱스	평가: Roc 인덱스	검증: Roc 인덱스
Y	MdlComp12	Reg5	회귀(Regre,,,	Q11_Y3	0,278813	0,289677	0,295224	0,787	0,769	0,77
	MdlComp14	Tree5	의사결정트,,,	Q11_Y3	0,273747	0,289918	0,286541	0,78	0,755	0,763
	MdlComp13	Neural5	신경망(Ne,,,	Q11_Y3	0,271395	0,292812	0,297154	0,799	0,769	0,771
	MdlComp15	MBR8	MBR	Q11_Y3	0,28207	0,33221	0,335981	0,785	0,72	0,722

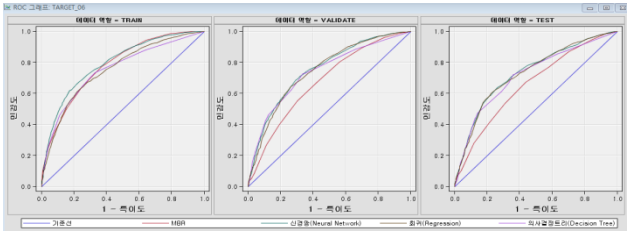
✓ 강원도



적합통계량

선택된 모델	선행 노드	모델 노드	모델 설명	타겟 변수	분석: 오분류 비율	평가: 오분류 비율	검증: 오분류 비율	분석: Roc 인덱스	평가: Roc 인덱스	검증: Roc 인덱스
Y	MdlComp4	Tree8	의사결정트,,,	Q11_Y4	0,203864	0,204863	0,210159	0,807	0,8	0,806
	MdlComp3	Neural8	신경망(Ne,,,	Q11_Y4	0,191405	0,208233	0,21064	0,858	0,832	0,829
	MdlComp2	Reg	회귀(Regre,,,	Q11_Y4	0,20585	0,214492	0,197882	0,839	0,833	0,84
	MdlComp5	MBR12	MBR	Q11_Y4	0,228602	0,259509	0,248435	0,816	0,767	0,761

✓ 남부지역(경상도,전라도,제주도)



적합통계량

선택된 모델	선행 노드	모델 노드	모델 설명	타겟 변수	분석: 오분류 비율	평가: 오분류 비율	검증: 오분류 비율	분석: Roc 인덱스	검증: Roc 인덱스	평가: Roc 인덱스
Y	MdlComp10	Tree2	의사결정트,,,	TARGET_06	0,241691	0,248254	0,257638	0,751	0,717	0,736
	MdlComp9	Neural2	신경망(Ne,,,	TARGET_06	0,230672	0,25596	0,262449	0,792	0,733	0,75
	MdlComp8	Reg4	회귀(Regre,,,	TARGET_06	0,256864	0,269685	0,274477	0,752	0,728	0,744
	MdlComp11	MBR5	MBR	TARGET_06	0,253613	0,300987	0,302382	0,775	0,656	0,673

II. 데이터 분석 및 마케팅 방법

각 모델 별 반응률을 고려한 지역 추천

- Validation Data 모델별 등급별 반응률

등급	인천	남부	경기도	강원도
1	9.78	10.34	9.59	13.58
2	10.41	10.86	8.60	13.11
3	9.40	10.70	9.47	10.33
⋮	⋮	⋮	⋮	⋮
19	0.79	3.13	1.29	1.66
20	0.79	2.78	0.88	1.65

- Score Data 모델별 등급

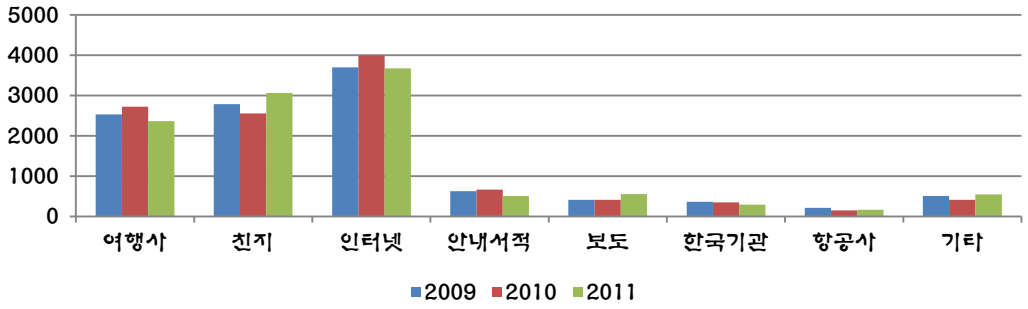
ID	인천	남부	경기도	강원도
10	19	17	1	1
112	9	3	20	20
3034	3	8	17	9

- 모델별 등급별 반응률을 비교하여 반응률이 가장 높은 지역을 추천

ID	인천	남부	경기도	강원도
10	0.786	3.198	9.591	13.575
112	4.352	10.699	0.877	1.651
3034	9.403	3.198	1.462	3.729

II. 데이터 분석 및 마케팅 방법

- 개인 맞춤형 관광 코스 추천 시스템의 예 – 인터넷 사이트 이용

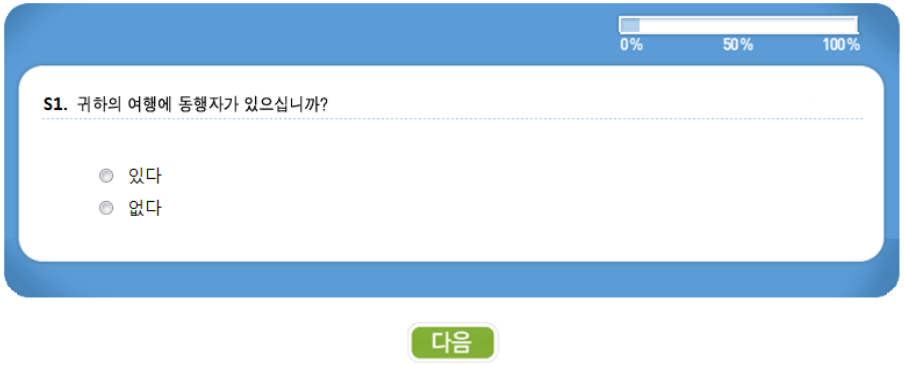


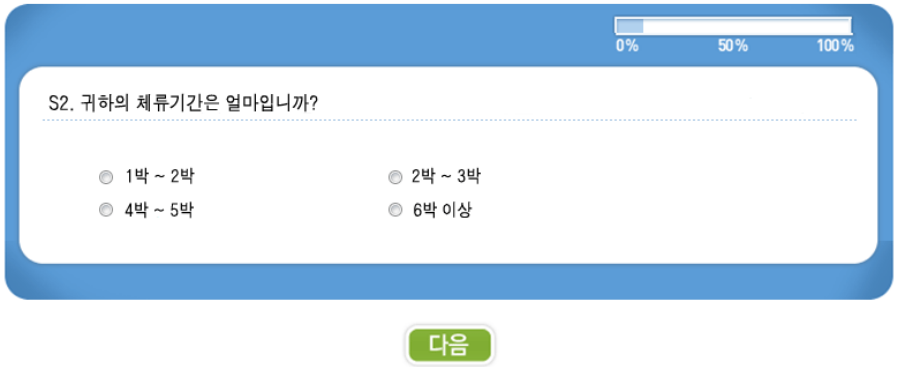
✓ 여행 정보 입수경로 설문에 대한 답변에서 인터넷이 많음
✓ 그에 따른 인터넷 여행지 추천 시스템 제안

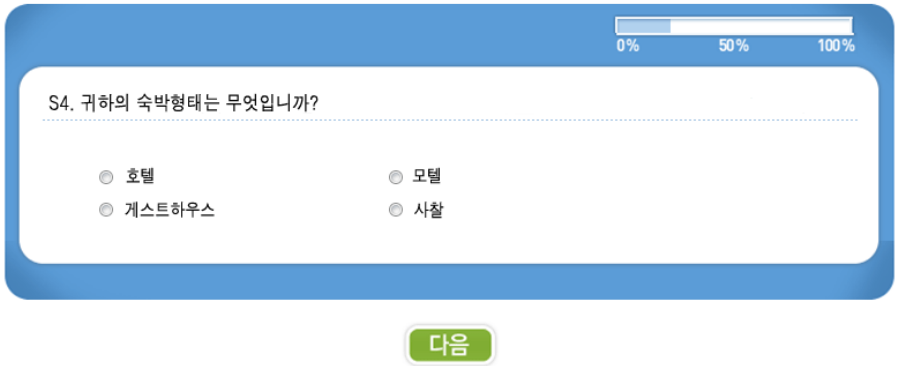
대한민국

II. 데이터 분석 및 마케팅 방법

- 앞에서 맞춤 여행지 추천 클릭 후 아래와 같은 정보들 외에도 모델링에 사용 된 모든 변수들을 조사







- 조사 된 정보를 이용하여 앞에서 만들어진 모델에 적용시켜 관광지 별 스코어 점수 생성 후 상위 점수의 관광지, 숙박 등을 추천

II. 데이터 분석 및 마케팅 방법

- 추천이 완료 된 화면

✓ 스쿼어 점수가 좋은 관광지 추천 이외에도 조사 내용들을 이용하여 숙박 등을 추천하는 시스템을 만들 수 있다.

✓ 더 나아가 소비가 이루어질 것으로 예상되는 부분의 쿠폰 제공함으로써 소비를 촉진 시킨다.

II. 데이터 분석 및 마케팅 방법

▪ 개인 맞춤형 관광 코스 추천 시스템

✓ 위와 같이 지역별 모델링을 통하여 관광객에게 관광지역을 추천할 수 있는 모델을 만들었다. 모델에서 사용한 변수는 관광 전에 얻을 수 있는 정보로 이를 이용하여 스코어 값이 높은 지역에 관광 및 숙박, 맛 집 정보 지원과 그 지역에 관광지 입장료 할인권을 지원함으로써 지역을 방문 가능성을 높인다.

✓ 하지만 개인맞춤형 관광코스 추천시스템 개발을 위해서 지역별 모델링보다는 관광지 별 모델링이 필요로 한다. 즉 개인맞춤형 관광코스 추천시스템을 개발하기 위해서는 관광지 별 방문 여부 데이터가 필요로 한다.

✓ 모델링에 필요한 설문문항

➢ 이름, 나이, 성별, 가족형태, 거주도시, 결혼여부, 자녀 수, 학력, 직업, 운전면허 여부, 소득 수준, 성격, 취미, 특기, 가족형태, 여행경험, 해외여행건수, 여행 목적 체류기간, 동반자수, 한국에 대한 정보 출처, 관심 있는 체험 외 빅데이터 기반 자료 요구

Ⅲ

분석 기법의 개선방안

III. 분석 기법의 개선방안

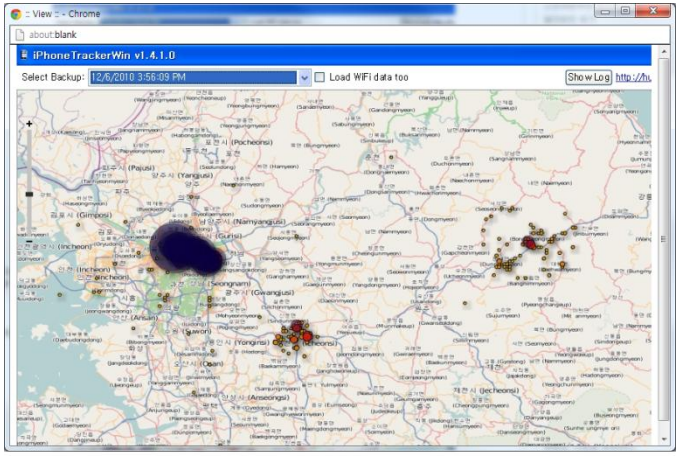
1) 빅데이터 시대에 맞는 데이터 구축

- ✓ 기존의 데이터베이스나 아키텍처가 저장, 관리, 분석할 수 있는 범위를 초과하는 거대한 규모의 데이터 집합 또는 이를 분석하는 기법을 뜻함
- ✓ 거대 규모의 데이터를 분석, 저장이 가능한 컴퓨터 기술의 발전과 대용양 데이터 분석을 위한 분산 처리 시스템 하둡의 등장으로 빅데이터 분석이 가능
- ✓ 빅데이터 시대에 문화관광부의 데이터 또한 빅데이터 기반을 통한 모집단 전체에 대한 정확한 분석 필요
- ✓ 문화 관광부에 필요한 빅데이터의 구조
  - ➢ 소셜 네트워크
  - ➢ 스마트폰, 컴퓨터 로그 자료 분석
  - ➢ GPS 기반 자료 분석

III. 분석 기법의 개선방안

2) GPS, 로그 데이터를 이용한 분석

- ✓ 스마트폰으로부터 얻은 개인별 GPS 정보를 이용한 관광객들의 방문 관광지 추적이 가능하며 이를 통한 개인별 관광객 추천 시스템 모델링 분석이 가능하며 실시간 추천 가능
- ✓ 스마트폰, 컴퓨터로부터 얻은 로그 기반의 데이터 분석으로 관심도가 높은 관광지를 파악, 실시간 추천이 가능해짐
  - ➢ 클릭 정보
  - ➢ 화면 별 정지 시간
- ✓ 민감한 개인 정보로 조사 이전 쿠폰 제공 등의 혜택을 통한 자발적인 동의를 얻는 것이 중요



III. 분석 기법의 개선방안

3) 소셜 네트워크를 이용한 분석

TEXT Mining

- ✓ 소셜 네트워크는 대표적인 빅데이터 자료로서 텍스트 마이닝 기법을 이용한 분석 가능
- ✓ 텍스트 마이닝 기법을 이용하여 실제 예비 관광객들의 Needs를 알아낼 수 있음
- ✓ 출국 후 한국에 대해 평가 한 소셜 데이터를 이용하여 전반적인 만족도 조사 가능

Busan ??

Like, nice, ...

대한민국

감사합니다.