

MS SQL과 R Studio를 통한 외국 관광객 신용 카드 사용 데이터 분석과 그에 따른 관광소비동향 파악

김준목, 강수정, 박지민, 이승연
한국외국어대학교 경영정보학과
e-mail: perish92@naver.com

Analysis of Foreign Tourists' Credit Card Data through MS SQL and R Studio and Identifying Tourism Consumption Trends

Jun-Mok Kim, Su-Jeong Kang, Jee-Meen Park, Seung-Yeon Lee
Department of Management Information System

요 약

외국관광객의 신용카드 거래내역을 통해 지역별, 국적별, 업종별 소비 형태와 동향을 파악하고, 이에 따른 여러 이해관계자들의 전략을 수립하는데 필요한 정보를 제공한다.

1. 서론

한국관광공사의 통계자료에 따르면 2005년 600만 명에 불과했던 외국관광객은 2015년 1400만 명에 육박할 만큼 매년 늘어나고 있는 추세이다. 이러한 증가 추세는 현대경제연구원 자료에 의하면 2020년 약 2,300만 명에 달할 것으로 전망된다. 또한 한류의 영향으로 한국에 대한 외국인들의 관심이 높아지면서 국내 관광 비율도 높아짐에 따라 관광산업이 다른 산업에 비해 굉장히 높은 경제기여도와 파급효과를 가지고 있다고 분석하였다. 이에 따라 우리는 MS Access에 데이터를 축적한 뒤 MS SQL을 이용해 유의미한 정보를 추출한 후 RODBC를 통해 MS SQL에 있는 데이터를 R스튜디오와 연동하였다. 이를 통해 외국인이 국내 관광 중에 사용한 비자카드 지출금액과 상세 거래내역을 국가별, 지역별로 분석해서 분야별 지출규모나 선호도, 기타 여러 가지 정보를 파악하고자 한다. 더 나아가 향후 관광소비 증가 전략의 지표로도 사용할 수 있을 것으로 기대된다.

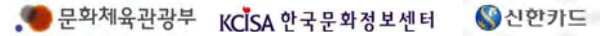
2. 자료 내용 및 자료 분석 범위

- 문화체육관광부, 한국문화정보센터, 신한카드사 간 협업을 통한 자료공유.
 - 신한카드사가 2014년에 매월 국내에서 사용된 외국인 카드(VISA 한정)사의 지출 자료를 매입하여 제공
 - 국가별, 지역별, 업종별 등에 있어 국내사용 지출액을 분석하여 요약자료로 제공함
- *개인정보자료가 들어있는 전표자료는 사용하지 않음

[표1] 자료 범위

부 문	세 부 업종
관광	쇼핑(면세점, 백화점, 대형마트, 전통시장), 숙박(호텔, 게스트하우스, 펜션, 모텔)

자료출처



3. Entity 및 ER Diagram

3-1. Entity 구성

[표2] Entity 구성

Entity Name	Primary Key	Attribute
고객	고객번호	이름, 성별, 나이, *국적번호
국적	국적번호	국적명
지역	거래지역번호	지역명
숙박	숙박번호	숙박업종
쇼핑	쇼핑번호	쇼핑업종
숙박거래	숙박거래번호	*고객번호, *숙박번호, 거래일자, 거래금액, *거래지역번호
쇼핑거래	쇼핑거래번호	*고객번호, *쇼핑번호, 거래일자, 거래금액, *거래지역번호

(* 표시는 Foreign Key를 의미)

- 국적명 : 중국, 일본, 미국, 홍콩, 대만
- 거래일자 : 2014년 1월, 2014년 2월, ..., 2014년 12월
- 거래지역명 : 서울시 중구 명동, 서울시 서초구 반포동, 서울시 강남구 압구정동, 서울시 송파구 잠실동, 서울시 용산구 이태원동, 경상북도 경주시 인왕동, 부산광역시 해운대구 우동, 제주특별자치도 서귀포시 색달동
- 숙박업종 : 호텔, 게스트하우스, 펜션, 모텔
- 쇼핑업종 : 면세점, 백화점, 대형마트, 전통시장

대형마트-전통시장 순서로 선호도가 나타났다. 명동은 거래횟수가 타 지역에 비해 폭발적으로 많았는데 역시 타 지역과 같은 선호도 순위를 보였다. 특이한 점은 송파구 잠실동인데, 이 지역만 백화점 - 대형마트 - 면세점-전통시장의 선호도 순서를 띄는 것으로 나타났다. 그 이유는 잠실지역에 롯데백화점과 롯데마트가 바로 옆에 연결되어있어 관광객들이 대부분 선물은 백화점에서 사고 먹거리는 마트에서 해결하여 면세점의 이용률이 자연히 떨어지게 되는 것으로 분석했다.

3-2. ER Diagram을 통한 정보도출

[사진1] ER Diagram



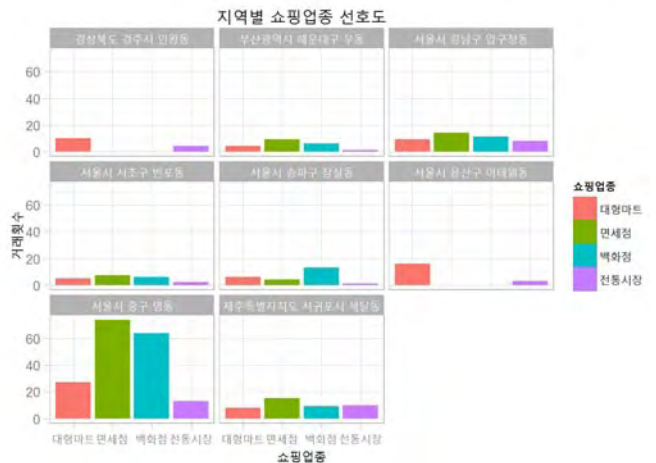
- ① 지역별 인기 쇼핑업종/숙박업종 선호도 파악 가능
- ② 국적별 인기 쇼핑업종/숙박업종 선호도 파악 가능
- ③ 20대 여자/남자에게 가장 인기가 많은 쇼핑업종 파악 가능
- ④ 월별 면세점 지출액에 따른 국적별 성수기/비수기 분석 가능
- ⑤ 국적별 쇼핑물 지출금액에 따른 소비 성향 파악 가능
- ⑥ 국적별/업종별 지출금액 파악 가능

<쿼리문>

```

SELECT 지역명, 쇼핑업종,
COUNT(쇼핑거래.쇼핑번호)
AS 거래횟수, SUM(쇼핑거래.거래금액)
AS 거래금액
FROM 지역, 쇼핑, 쇼핑거래
WHERE 지역.거래지역번호=쇼핑거래.거래지역번호
AND 쇼핑.쇼핑번호=쇼핑거래.쇼핑번호
GROUP BY 지역명, 쇼핑업종
ORDER BY 지역명 ASC, COUNT(쇼핑거래.쇼핑번호)
DESC;
    
```

ER Diagram을 통해 도출해낸 정보들 중 향후 관광소비 증가 전략의 지표로 유용하게 사용될 수 있을만한 ①, ②, ④번 정보들에 대해 분석해 보았습니다.



4. 얻을 수 있는 정보 및 활용

4-1. i) 지역별 인기 쇼핑업종 선호도 파악

먼저 지역별로 인기 있는 쇼핑업종을 파악하기 위해, 지역별로 쇼핑업종의 거래횟수와 거래금액을 추출하였다. 거래횟수는 쇼핑거래번호가 몇 번 있었는지 Count를 하였고, 거래금액은 각 거래금액을 모두 sum으로 더했다.

대부분의 지역은 면세점 - 백화점 - 대형마트 - 전통시장의 순서를 따랐다. 경주시 인왕동과 용산구 이태원동에는 면세점과 백화점이 없기 때문에

4-1. ii) 지역별 인기 숙박업종 선호도 파악

다음은 지역별로 인기 있는 숙박업종을 파악해보기 위해 지역별로 숙박업종의 거래횟수와 거래금액을 파악해보았다. 4-1. i)와 유사한 형태의 쿼리문을 갖는다.

전체적으로 호텔의 선호도가 가장 높았는데, 이는 최근 중·저가 호텔, 또는 비즈니스호텔이 많이 생기고 있는 추세 때문으로 보인다. 특히 명동의 호텔 선호도가 압도적인데, 이 또한 명동에 비즈니스호텔이 밀집되어있기 때문으로 보인다. 이태원은 게스트하우스의 비율이 비교적 높는데, 이는 관광객들이 이태원의 이국적 분위기와 잘 어울리는 개성 있는 게스트하우스를 선호하는 경향이 있기

때문으로 보여진다.

<쿼리문>

```
SELECT 지역명, 숙박업종,
COUNT(숙박거래.숙박번호)
AS 거래횟수, SUM(숙박거래.거래금액)
AS 숙박거래금액
FROM 지역, 숙박, 숙박거래
WHERE 지역.거래지역번호=숙박거래.거래지역번호
AND 숙박.숙박번호=숙박거래.숙박번호
GROUP BY 지역명, 숙박업종
ORDER BY 지역명 ASC ,COUNT(숙박거래.숙박번호)
DESC;
```

ORDER BY 국적명 ASC, COUNT(쇼핑거래.쇼핑번호) DESC;



4-2.ii) 국적별 인기 숙박업종 선호도 파악

국적별로 어떤 숙박시설을 선호하는지 알아보기 위해 국적별 숙박업종의 거래횟수를 파악해보았다. 4-2. i)와 유사한 형태의 쿼리문을 갖는다. 전체적으로 호텔 이용횟수가 압도적으로 많다. 특이점으로는 대만은 호텔에 비해 모텔 이용횟수가 많은데, 이는 최근 대만 관광객이 주로 쇼핑 위주의 관광을 하면서 숙소는 모텔같은 저가형을 즐겨찾기 때문으로 분석된다. 또한 대부분 나라들의 펜션숙박 거래횟수가 가장 낮은데, 이는 주로 외국관광객들이 서울에 관광을 오기 때문인 것으로 보인다.

4-2. i) 국적별 인기 쇼핑업종 선호도 파악

국적별로 어떤 쇼핑업종을 선호하는지 알아보기 위해 국적별 쇼핑 거래횟수를 파악해보았다. 쇼핑거래의 쇼핑번호를 COUNT해서 어떤 쇼핑업종을 많이 이용했는지 알아보았다.

전체적으로 면세점-백화점-대형마트-전통시장 순이었다. 특히 중국이 면세점 이용횟수가 많았는데, 면세점 중에서 유포커를 공략하기 위한 '중국인 관광객 대상 다양한 혜택', '중국어 통역 등 중국인 대상 우수한 서비스' 등을 지원하는 곳이 많기 때문인 것으로 분석된다. 그리고 특이점으로는 미국과 대만은 백화점 거래가 더 많은 것으로 보여진다.

<쿼리문>

```
SELECT 국적명, 쇼핑업종,
COUNT(쇼핑거래.쇼핑번호) AS 거래횟수
FROM 국적, 쇼핑, 쇼핑거래, 고객
WHERE 국적.국적번호=고객.국적번호 AND
고객.고객번호= 쇼핑거래.고객번호 AND
쇼핑.쇼핑번호=쇼핑거래.쇼핑번호
GROUP BY 국적명, 쇼핑업종
```

<쿼리문>

```
SELECT 국적명, 숙박업종,
COUNT(숙박거래.숙박번호) AS 거래횟수
FROM 국적, 숙박, 숙박거래, 고객
WHERE 국적.국적번호=고객.국적번호 AND
고객.고객번호= 숙박거래.고객번호 AND
숙박.숙박번호=숙박거래.숙박번호
GROUP BY 국적명, 숙박업종
ORDER BY 국적명 ASC, COUNT(숙박거래.숙박번호)
DESC;
```



4-3. 월별 면세점 지출액에 따른 국적별 성수기 분석

우선 월별 면세점 지출액에 따른 국적별 성수기를 파악하기 위해, 국적별로 거래일자과 거래금액을 파악하였다. 거래일자는 월 단위로 나누어 측정하였고, 거래금액은 각 거래금액을 모두 sum으로 더하였다. group by를 통해 국적별로 1월부터 12월까지의 거래금액을 차례로 보여줄 수 있도록 하였다.

전체적으로 겨울에 면세점 거래가 많은 성수기이며 특이점으로는 미국과 중국이 있다. 미국의 성수기는 방학시즌으로 분석되어진다. 중국의 성수기는 중국명절인 춘절로 분석되며, 1~2월에 압도적으로 면세점 지출액이 많음을 알 수 있다.

<쿼리문>

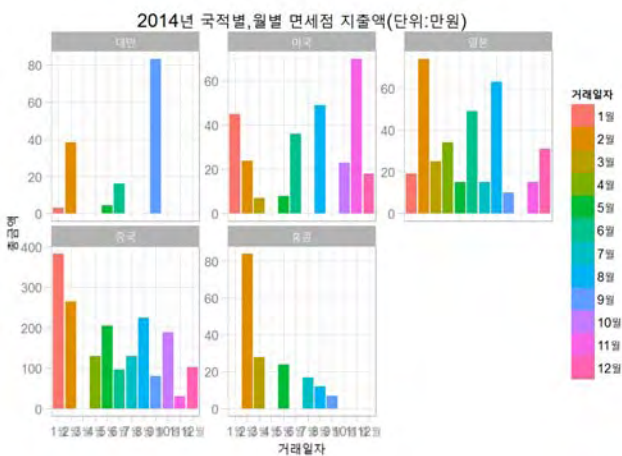
```
SELECT 국적명, 거래일자, SUM(거래금액) AS
총금액
FROM 국적, 쇼핑거래 ,고객
WHERE 국적.국적번호=고객.국적번호 AND
쇼핑거래.고객번호=고객.고객번호 AND
쇼핑번호=201
GROUP BY 국적명, 거래일자;
```

관광산업 및 관광 상품을 제시할 수 있을 것으로 기대되며 정부 또한 데이터의 축적과 의미 있는 정보를 추출하여 관광정책을 수립할 수 있을 것이다.

향후 연구 과제로는 더 다채로운 데이터를 수집하여 시간대별 외국인 관광객 쇼핑 행동 등을 관찰해보고, R의 지리정보 패키지를 활용하여 외국인 관광객의 지역별 행동 특성을 더욱 더 뚜렷하게 살펴봄으로써 좀 더 실질적인 상업적 활용 자료를 제공하는 것이다.

6. 출처

David Kroenke and David Auer, 「Database Concepts」 (Pearson, 2007)
 N. Umanath and R. Scamell, 「Data Modelling and Database Design」 (Thomson, 2007)
 Jeffrey Ullman, 「Database Systems」 (Pearson Prentice Hall, 2001)
 Silberschatz et al., 「Database System Concepts」 (McGraw-Hill, 2010)
 성윤정, 「MS SQL Server 2012」 (인피니티 북스 ,2013)
 김연희, 「데이터베이스 개론」 (한빛미디어, 2013)
 문화체육관광부, 「2014 외국인 신용카드 사용 지출액 현황 분석 보고서」



5. 결론

본 논문에서는 MS SQL과 R 을 통해 외국관광객의 신용카드 거래내역을 업종, 국적, 시기 등으로 세분화하여 분석한 내용에 대해서 기술하였다. 신용카드 거래내역을 관계형 데이터베이스에 저장함으로써 여러 이해관계자들이 다양한 방면에서 자료를 활용할 수 있도록 하였으며, 본 논문에서 다룬 예시 이외에도 자료 추출을 통해 상업적으로 활용할 수 있는 방안이 무궁무진할 것으로 예상된다.

이를 통한 기대효과로는 외국인 관광과 관련된 다양한 업체들이 효율적인 관광마케팅을 시도해