중앙도서관 대출기록 분석을 통한 도서 대출가능성 향상

Business Background & Problem

순위	제목	대출가능여부
1	여자 없는 남자들	N
2	창문 넘어 도망친 100세 노인	X
3	21세기 자본	N
4	어떤 하루	N
5	장하준의 경제학 강의	N

순위	제목	대출가능여부
1	여자 없는 남자들	N
2	비밀의 정원	X
3	버티는 삶에 관하여	N
4	창문 넘어 도망친 100세 노인	N
5	메이즈 러너	N

순위	제목	대출가능여부
1	여자 없는 남자들	N
2	창문 넘어 도망친 100세 노인	X
3	김우중과의 대화	N
4	21세기 자본	N
5	나는 까칠하게 살기로 했다	N

순위	제목	대출가능여부
1	창문 넘어 도망친 100세 노인	X
2	**인생수업**	**Y**
3	멈추면, 비로소 보이는 것들	N
4	원씽	N
5	나는 죽을 때까지 재미있게 살고 싶다	N
6	정글 만리	N
7	그래도 사랑	N
8	총, 균, 쇠	N
9	**흔들리지 않고 피는 꽃이 어디 있으랴**	**Y**
10	삐뽀삐뽀 119 소아과	X

순위	제목	대출가능여부
1	나는 까칠하게 살기로 했다	N
2	**내가 알고있는 걸 당신도 알게 된다면**	**Y**
3	**에드워드 툴레인의 신기한 여행**	**Y**
4	총, 균, 쇠	N
5	칼 비테의 자녀교육 불변의 법칙	X
6	멈추면, 비로소 보이는 것들	N
7	꾸뻬씨의 행복여행	N
8	살아갈 날들을 위한 공부	X
9	아기가 잘 먹는 이유식은 따로 있다	X
10	나미야 잡화점의 기적	N

왼쪽 위부터,
1: 교보문고 월간베스트셀러
2: 반디앤루니스 주간베스트셀러
3: Yes24 월간베스트셀러
4: 교보문고 스테디셀러
5: Yes24 스테디셀러
6. 전공도서

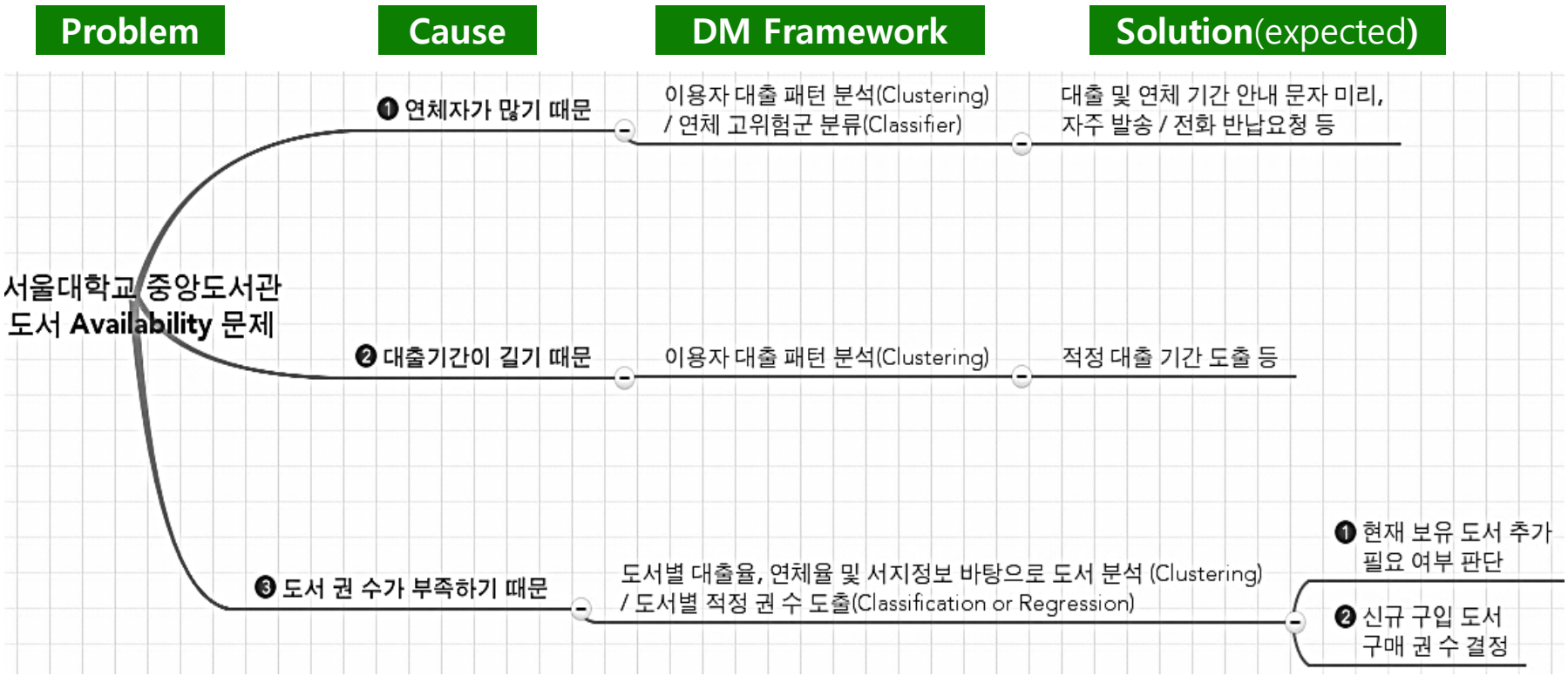
과목	제목	대출가능여부
생산관리	Matching supply with demand	N
데이터마이닝	**Datamining for business intelligence**	**Y**
재료공학개론	Introduction to material science and engineering	N
기술경영	기술과 경영	N
품질경영	Juran's quality planning and analysis for enterprise quality	N

***대출가능여부 : 2014.10.24 기준**

베스트/스테디셀러 도서 대략 30권 가운데 **오직 4권**만 대출 가능 & 전공도서도 비슷한 상황

중앙 도서관의 낮은 도서 대출가능성에 대한 개선 필요!

Data Mining Problem & Model to Use



Why Data Mining?

- ✓ 기존 도서관 대출기록에 대한 연구는 단편적인 통계 분석 수준에 지나지 않았음. 하지만 **다각도에서 대출 패턴을 파악**하고 수요를 맞추기 위해서는 Data Mining 기법 필요.
- ✓ 도서 Clustering과 도서 별 적정 구매 권수 도출 역시 서지 정보, 대출/연체율 등 **많은 변수들을 동시에 고려**해야 하므로 Domain Knowledge를 바탕으로 한 정성적 접근만으로는 수행하기 어려운 과제.

Expected Result & Business Implication

- ✓ **이용자 중심 도서관 운영** 이용자들의 대출패턴 분석 결과를 중앙도서관 수서 및 장서 관리에 반영함으로써 도서가 부족하여 대출을 하지 못하는 불편 해소.
  - 보유 도서 가운데 추가 구입이 필요한 도서의 추가 구입으로 **현재의 이용 불편** 해소
  - 신규 구입 도서에 대해서도 적정 구입 권 수 도출하여 체계적으로 구입함으로써 **미래의 이용 불편 가능성**까지 해소.
  - 이러한 분석은 추후 **관정도서관의 도서 구입** 시에도 유용하게 활용 가능할 것으로 기대.
- ✓ **연체도서 관리 효율화 :** 연체 고위험군 분류를 통해 연체 이후에 반납을 독촉하는 사후적 대응이 아닌 연체가 생기지 않도록 예방하는 사전적 대응 가능.
  - 연체 고위험군에 대한 관리가 현재 운영되는 시스템이 아니므로 실제 도입 위해서는 타 도서관 및 해외 사례들에 대한 **벤치마킹** 필요.

How to obtain data?

☆ RE: 서울대학교 도서관 대출기록 데이터 열람 관련 문의가 있습니다.

태그를 추가하려면 여기를 클릭하세요.

보낸사람 : "박진만" <praise@snu.ac.kr> | 주소추가 | 수신거부

날짜 : 2014.10.24 09:08

받는사람 : <vluujh@snu.ac.kr> | 주소추가

안녕하세요.
중앙도서관 박진만입니다.

말씀하신 내용은 충분히 제공해 드릴 수 있는 부분입니다.
단, 외부로 나가는 통계는 공문을 통해서만 접수를 하고 있어서
수업 담당 교수님 또는 학과장님 싸인이 들어간 공문이 필요합니다.

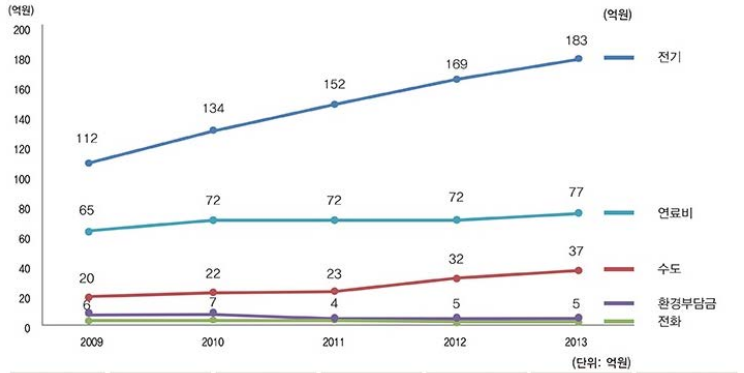
원하는 항목 등이 구체적으로 명시된 공문을 중앙도서관 기획홍보실(02-880-9375)에 제출해 주시면 됩니다.
감사합니다.

Data Mining Term Project Proposal

1. Business Background

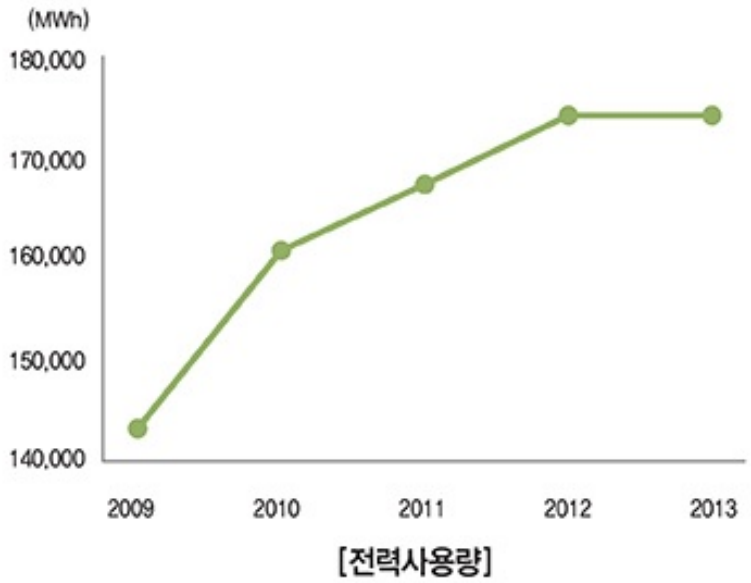
- 신림 발전소(한국 전력 공사 소속)에서 연간 11만 6347톤의 전력을 받고 있음
- 교내에 위치한 파워플랜트 3대를 통해 제공받는 고압을 사용 가능한 저압으로 변환하여 각 건물로 전력을 공급함
  - 제 1 파워플랜트: 캠퍼스 내 전력 총괄적으로 담당
  - 제 2 파워플랜트: 공대 전력 공급
  - 제 3 파워플랜트: 인문대와 음대 일대 전력 공급

2. Business Problem



(단위: 억원)

구분	2009	2010	2011	2012	2013	증가율(%)
전기	111.7	133.8	152.2	169.5	182.8	63.6
수도	20.0	21.6	22.7	32.0	37.3	86.4
전화	3.0	2.9	2.8	2.6	2.4	-20.6
환경부담금	6.4	7.4	4.5	4.7	5.4	-15.0
연료비 (가스)	64.7	71.8	71.8	72.4	76.8	18.7
합계	205.8	237.5	254.0	281.2	304.7	48.0



[전력사용량]

- 지난 4년간 공공요금이 약 100억원 증가하여 2013년 공공요금은 300억원 도달
  - 2013년 서울대학교 운영비 총액의 33%
  - 공공요금 증가분의 약 70%는 전기요금 증가에서 발생
- 냉난방기 순차 운행, 스마트 센서 구축을 통해 2013년 전기사용량 소폭 감소
  - 약 53,000세대(거주인구 약 160,000명) 1년 사용량과 비슷

→ **세밀한 에너지 사용현황 분석 및 추가적인 에너지절약 정책 필요**

Data Mining Term Project Proposal

3. Data Mining Approach

- 시간별 및 계절별 전기 사용패턴이 있을 것으로 추측
- 숨겨진 전기 사용패턴을 통해 세부적인 전기사용 현황분석
- 비슷한 전기 사용패턴과 면적당 사용량을 보이고 있는 건물들을 그룹화할 수 있을 것으로 기대

4. Data Description

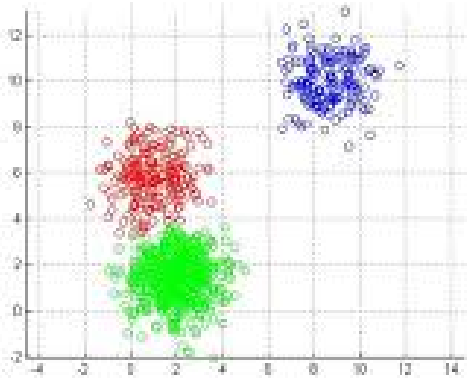
시간		60동(본부)			5동								7동						9동		
		NO.4	NO.6	NO.7	NO.1	NO.2	NO.4	NO.5	NO.6	NO.7	NO.8	NO.8	NO.1	NO.2	NO.3	NO.4	NO.5	NO.6	NO.1	NO.2	NO.3
		1동	2동	109동	5동 LV-1	3동 메인	3동 EHP	5동 3층	5동 1층	5동 1층_1	예비	4동 메인	7동 메인	7동	5,9동	11,14동	76동	11,61동	9동 메인	10동 3층	10동 2층
		전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량	전력량
		kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh	kWh
01시		0.0	0.0	0.0	16.0	9.0	4.0	5.0	3.0	1.0	0.0	11.0	115.0	12.0	22.0	18.0	11.0	30.0	520.0	34.0	47.0
02시		0.0	0.0	0.0	15.0	9.0	3.0	4.0	4.0	0.0	0.0	10.0	107.0	8.0	21.0	19.0	10.0	30.0	19.0	1.0	2.0
03시		0.0	0.0	0.0	16.0	10.0	7.0	4.0	3.0	1.0	0.0	11.0	109.0	8.0	21.0	19.0	11.0	31.0	17.0	2.0	3.0
04시		0.0	0.0	0.0	15.0	10.0	7.0	4.0	3.0	1.0	0.0	9.0	108.0	7.0	22.0	17.0	10.0	30.0	17.0	2.0	2.0
05시		0.0	0.0	0.0	14.0	9.0	4.0	4.0	3.0	0.0	0.0	10.0	100.0	8.0	18.0	16.0	10.0	29.0	17.0	1.0	2.0
06시		0.0	0.0	0.0	15.0	9.0	3.0	4.0	3.0	1.0	0.0	12.0	92.0	10.0	20.0	12.0	9.0	19.0	18.0	2.0	2.0
7시		0.0	0.0	0.0	17.0	9.0	3.0	5.0	4.0	1.0	0.0	16.0	100.0	10.0	27.0	12.0	9.0	21.0	21.0	2.0	2.0
8시		0.0	0.0	0.0	19.0	10.0	5.0	6.0	5.0	0.0	0.0	23.0	117.0	10.0	37.0	13.0	8.0	19.0	28.0	2.0	2.0
9시		0.0	0.0	0.0	20.0	10.0	18.0	7.0	5.0	1.0	0.0	31.0	165.0	13.0	67.0	15.0	16.0	26.0	35.0	4.0	4.0
10시		0.0	0.0	0.0	31.0	14.0	30.0	11.0	10.0	1.0	0.0	43.0	253.0	24.0	111.0	19.0	22.0	44.0	55.0	4.0	6.0
11시		0.0	0.0	0.0	38.0	20.0	42.0	13.0	11.0	1.0	0.0	40.0	304.0	29.0	124.0	23.0	24.0	54.0	69.0	4.0	7.0
12시		0.0	0.0	0.0	43.0	22.0	48.0	15.0	11.0	1.0	0.0	43.0	355.0	33.0	140.0	23.0	39.0	52.0	75.0	5.0	8.0
13시		0.0	0.0	0.0	38.0	19.0	41.0	14.0	10.0	1.0	0.0	42.0	337.0	38.0	137.0	24.0	47.0	46.0	74.0	2.0	7.0
14시		791.0	1161.0	1926.0	44.0	21.0	44.0	16.0	12.0	1.0	0.0	39.0	368.0	40.0	159.0	27.0	47.0	49.0	91.0	2.0	9.0
15시		0.0	0.0	0.0	46.0	22.0	55.0	17.0	13.0	1.0	0.0	40.0	343.0	53.0	142.0	28.0	22.0	47.0	102.0	4.0	9.0
16시		0.0	0.0	0.0	47.0	23.0	57.0	18.0	12.0	1.0	0.0	38.0	338.0	47.0	121.0	27.0	23.0	53.0	105.0	7.0	11.0
17시		0.0	0.0	0.0	43.0	22.0	56.0	17.0	11.0	1.0	0.0	38.0	321.0	47.0	128.0	27.0	21.0	46.0	96.0	7.0	11.0
18시		0.0	0.0	0.0	38.0	18.0	39.0	14.0	10.0	1.0	0.0	38.0	309.0	41.0	116.0	24.0	33.0	46.0	88.0	5.0	8.0
19시		0.0	0.0	0.0	30.0	14.0	29.0	12.0	7.0	1.0	0.0	41.0	252.0	30.0	106.0	21.0	29.0	28.0	74.0	3.0	7.0
20시		0.0	0.0	0.0	27.0	13.0	28.0	11.0	6.0	1.0	0.0	34.0	218.0	31.0	94.0	21.0	12.0	24.0	71.0	3.0	8.0
21시		0.0	0.0	0.0	26.0	12.0	19.0	11.0	6.0	1.0	0.0	39.0	218.0	26.0	88.0	24.0	11.0	35.0	59.0	3.0	7.0
22시		0.0	0.0	0.0	21.0	10.0	16.0	10.0	4.0	0.0	0.0	37.0	198.0	23.0	77.0	22.0	11.0	35.0	42.0	3.0	6.0
23시		0.0	0.0	0.0	19.0	9.0	14.0	7.0	5.0	1.0	0.0	25.0	181.0	22.0	62.0	23.0	10.0	33.0	29.0	1.0	3.0
24시		0.0	0.0	0.0	17.0	9.0	13.0	6.0	4.0	1.0	0.0	10.0	138.0	15.0	38.0	21.0	12.0	32.0	25.0	2.0	3.0

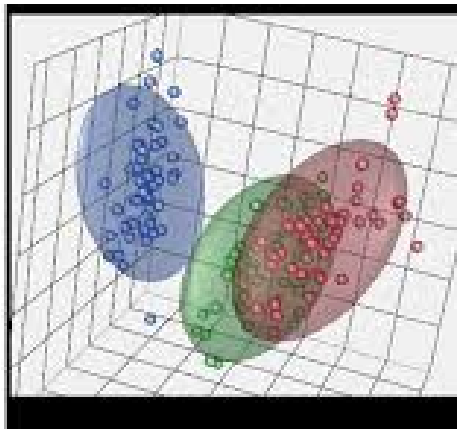
- 2013년 8월 1일부터 2014년 10월 22일까지 건물별 시간당(1시간 단위) 전기 사용량 원본데이터 획득
- 교내 총 143동 390개 센서에서 측정된 자료
- 매월 발행되는 「서울대학교 건물별 전기·도시가스 에너지 사용정보」를 통해 교내 164동 건축면적 및 연면적 획득
  (자료제공: 시설관리국 시설지원과)

Data Mining Term Project Proposal

5. Model to Use

- Data Preprocessing
  - 센서오류로 발생한 비정상 레코드 추출
- K-Means Clustering
  - 비슷한 면적당 사용량, 시간별 및 계절별 사용패턴을 보이는 건물끼리 그룹화
  - 각 건물별 레코드를 plot해보고 다양한 사용패턴 형태를 반영하는 K값 설정





6. Expected Result and Business Implication

- 세부적인 에너지 사용현황 및 시간별 사용패턴 파악가능
  - 현재는 월별 전체 사용량만 보고되고 있음
  - 시간별 사용 패턴 및 사용량 변화를 확인하여 보다 정확한 에너지 사용현황 파악가능
- 에너지 할당제 도입에 앞서 기관별 할당기준 제공
  - 사용패턴 및 사용량 기준에 따른 건물 및 기관 분류를 통해 할당기준 제시 및 타당성 검증
- 건물 및 기관별 맞춤형 에너지 절약 방법 제공
  - 그룹별 증가량 및 사용성향에 대한 원인분석
  - 그룹별 에너지 절약방안 제시

Q. 과연 가요의 흥행은 노래의 퀄리티가 결정할까?

➡ 한국 가요산업에서 음원 판매량 예측

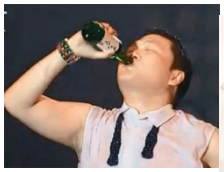
★Why this problem?

:한국 가요산업에서 곡 수입은 가수에게 매우 적은 부분만 돌아옴

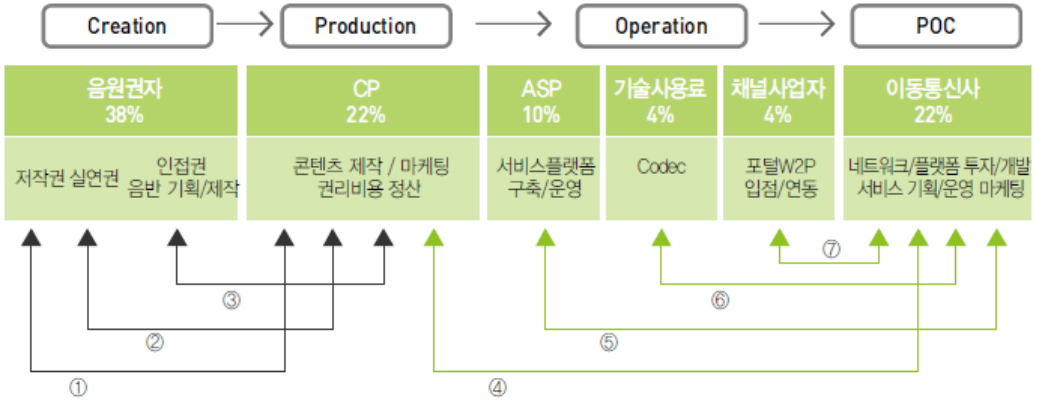
음반 산업은 1999년부터 하락세, 즉 곡 단위 판매로 시장구조의 변화

가수지망 청소년의 증가/ '연습생'의 증가

한류, K-pop의 실체 파악



- 2010년 3월 16일 조선일보에 가수지망생의 수면부족, 다이어트 강요, 술자리 강요, 학습권 침해 등 인권침해 사례 보도
- 2010. 4. 인권위에서 여성연예인 인권조사 발표
- 2010. 8. 23. 여성가족부에서 청소년 연예인의 생활실태조사 발표(청소년 연예인과 연예인지망생을 대상으로 조사하였음)
- 2010. 8. 26. 공정거래위원회에서 '청소년 연예인의 전속계약서 인권 침해' 개선 검토 발표



CP계약 관계

① : 음원 사용 및 저작권료 납부 계약 (음저협)
② : 음원 사용 및 실연권료 납부 계약 (음실연)
③ : 음원 사용 및 인접권료 정산 계약 (음제협 외)

이동 통신사 계약 관계

④ : CP 콘텐츠 공급 및 권리비용 정산 계약
⑤ : ASP 운영 계약
⑥ : 기술사용 수수료 계약
⑦ : WEB Portal 내 W2P 서비스 인접 계약

한국 가요산업에서 음원 판매량 예측

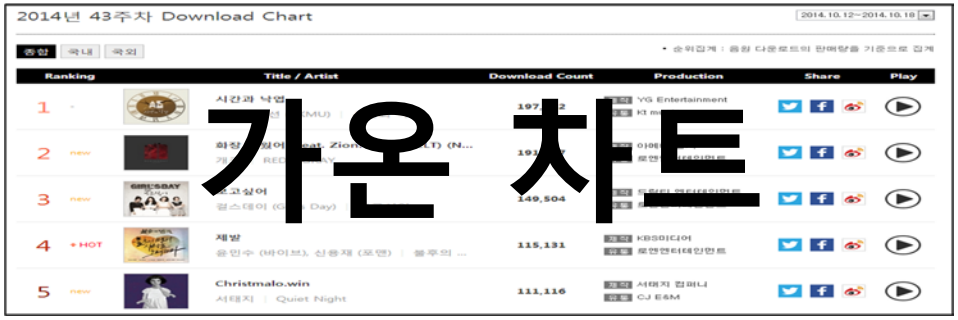
★어떤 Data?

- √ 해당 가수의 SNS에서 부정적 평가량
- √ 해당 가수의 SNS에서 긍정적 평가량
- √ 소속사
- √ 유통사
- √ 뉴스 노출 횟수
- √ 해당 가수의 검색량
- √ 판매량(다운로드, 스트리밍, bgm)

SOCIAL metrics™

-가수별 SNS에서의 긍정적/부정적 평가량 분석가능

★How to get data?

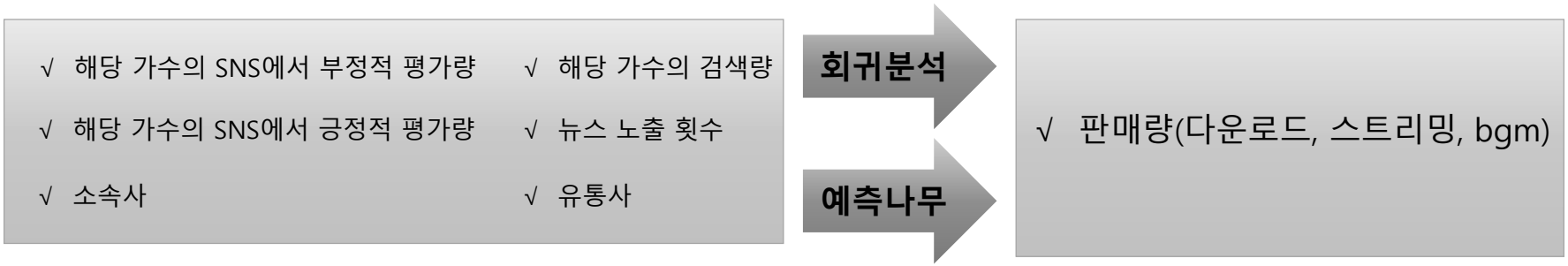


-2010년부터 종목별 1~100위까지 판매량을 주단위로 제공
+ 해당 음원의 아티스트, 제작사, 유통사 정보 제공

-과거 기간별 검색량 추이 제공

한국 가요산업에서 음원 판매량 예측

★How to 분석?



★Expected Result and Business Implication

- ❖ 판매량에 소속사, 유통사가 미치는 영향 파악 → 현재 수익구조의 정당성 평가
- ❖ 연예 기획사의 효율적 신인 육성 전략 수립
- ❖ 가수지망생, 연습생들의 정당성 평가
- ❖ 음원 시장의 dominate class 파악

공과대 체력단련실의 고객 차별화 시스템 제안

1. Business background

- 1개월/3개월/6개월/1년 Program
- 유료 서비스인 신발장 서비스
- 기간 중단 기회(3개월 1번, 6개월 2번)
- 개강맞이 EVENT(3개월/6개월 대상)
  . 3명 이상 10% 할인, 6명 이상 20% 할인
  . 선착순 30명 사물함 2개월 무료
  . 6개월 등록 고객 사물함 2개월 무료
  ※ 기존 회원은 환급 후 재 신청 가능
  결제금액 10% + 사용일수 공제 후 환급
- 대다수의 수요가 개강맞이 EVENT에 존재

2. Business Problem

- 계층② : 다수의 회원들이 이에 속하는데, 만족도가 낮고, 충성도도 낮아 이탈 확률이 크며 이는 매출에 직접적인 영향

- 계층④ : 할인 혜택 없이 등록하는 회원 중에서 다수가 이에 속하는데, 이 고객들을 계층①로 유인했을 때 얻을 수 있는 부가가치가 존재함

※ 사전 고객 계층 분류

(i) 할인 혜택 기간에만 등록하는 회원 (다수)

① 운동 열심히 하는 회원 (Good)
- 할인 혜택을 받기 위해 운동 열심히 하지 않는 회원들까지 모아오는 접점 역할을 함
- 만족도가 높아 이탈 확률도 작음

② 운동 열심히 하지 않는 회원 (Bad)
- 만족도가 낮아 이탈 확률이 큼

(ii) 할인 혜택 없어도 등록하는 회원 (소수)

③ 1개월 이용 고객 (관심 대상x)
- 3개월이나 6개월 프로그램이 단위 비용이 저렴하지만 1개월을 이용한다는 것은 스케줄 상의 이유로 단기간 잠깐 이용하고자 하는 고객
- 매출에 큰 영향을 미치지 않고 이탈 확률 낮음

④ 3개월/6개월 이용 고객 (Not bad)
- 할인 혜택 없이 장기간을 이용하는 고객의 경우 운동을 열심히 할 사람일 확률이 큼
- 계층 ①로 유인하면 부가가치 창출 가능

⑤ 1년 고객 (극소수로 별도 관리)
- 헬스를 꾸준히 이용하는 계층
- 신학기 프로그램 사용 x -> 별도 관리 필요

3. Datamining Problem

- 계층 ②의 고객이 계층 ①의 고객으로 변화하도록 유도하고,
  계층 ④의 고객이 계층 ①의 고객으로 변화하도록 유인하기 위하여,
  헬스장 회원을 계층화하여 차등 관리를 하는 게 필요함
- 이를 위해 헬스장 회원들을 data를 바탕으로 계층화하는 작업이 필요함

4. How to Obtain Data

- 작년 1년(2013.03.04~2014.03.02)간 구매가 이뤄진 전체 프로그램에 대한 고객 이용 정보
  회원 고유 번호, 구매한 프로그램(+이벤트 유무), 환급 기록, 기간 내 헬스장 이용횟수
  일별 헬스장 이용 시간, 프로그램 기간 내 중단 기간, 기간 내 헬스장 이용 날짜 분포
- 공과대 체력단련실에 문의한 결과, 개인 정보를 포함한 data이기 때문에 상위 기관인
  포스코 스포츠센터에 공식적인 요청이 필요할 것 같다는 답변을 받았음.
  이에 포스코 스포츠센터 서비스 팀에 이를 요청한 상황.

5. Model to use -> Clustering

- Model을 통해 우선 Outlier로서 ③, ⑤ 계층을 분리해 낸 뒤, 남은 대다수 고객(3개월/6개월)
  을 대상으로 Clustering -> 여러 차원(기간 내 월평균 이용횟수, 기간 내 월평균 이용 시간,
  프로그램 기간 내 중단 기간 비율, 헬스장 이용 날짜 분포)의 Data를 바탕으로 2가지 계층
  (Gold/Platinum)으로 그룹화
- Gold/Platinum 기준 설정의 문제 -> 두 계층간의 그룹 설정은 기업 매출에 상당한 영향
  · 따라서 기업이 통계적 분석을 통해 자체적으로 설정한 기준을 Domain Knowledge로 활
  용
  · 얻어진 Data의 통계치 분석에 바탕을 두고, 정책적인 기준(ex Gold : Platinum = 7 :3)을
  이용하여 분류 기준 설정 가능

※ 해결방안 -> '회원 등급제' – Gold, Platinum, Diamond

- Default는 Gold이며, 회원 등급을 프로그램 기간이 종료되었을 때 수행하는 기말 평가로 조정하고, 이를 프로그램 재 구매시에 반영하여 회원 등급에 따라 차별화된 혜택을 제공
- 1개월 이용 고객 -> 평가 없이 기존의 등급이 유지
- 3/6개월 이용 고객 -> 기말 평가로 열심히 했느냐, 열심히 하지 않았느냐에 따라 등급 조정
- 1년 이용 고객 -> Diamond 계급 부여(평가x), 재 구매할 때 1년 program 구매하지 않을 경우 구매 당일 Platinum으로 downgrade 되며, 기말 평가 수행하여 등급 조정
- 등급 조정 결과는 메일을 통해 통보되며, 개강맞이 EVENT 안내문은 등급별로 다르게 전송
- 등급별 혜택


Gold -> 기존의 규정대로


Platinum -> 신발장 할인 / 중단 기회 +1회
개강맞이 EVENT 이용 시 그룹 내 Gold회원 포함 정도에 따라 할인% 증가
(ex Gold 4명, Platinum 2명 -> Platinum 2명이 각각 할인율 4/2=2% 증가)


Diamond -> 신발장 연중 무료 / 중단 기회 총 4회

6. Expected Result and Business Implication

- 기존에 이탈 확률이 컸던 계층② 고객의 이탈을 방지하는 효과 뿐만 아니라, 이들이 기업의 주요 부가가치창출원인 계층①로 성장할 수 있게 되어 매출 상승 효과가 상당할 것으로 기대됨
- 또한, 계층④의 고객이 계층①의 고객이 될 유인이 증가함으로써, 이로 인한 부가가치 창출도 기대됨
- 전반적으로 운동을 열심히 하는 사람이 증가함에 따라 고객 만족도가 증가함으로써 고객의 충성도가 높아지고, 이탈 가능성이 크게 줄어들게 될 것으로 예상됨.

1. Business Overview

- 2013년 350억 매출 달성, 국내 1위 결혼정보회사

- 2014년 10월 한달 간 성혼인원 약 300명 (150쌍)

- 서비스진행절차 :
  회원가입 – 160여개 데이터폼 작성 – 듀오매칭시스템을 통한 커플매칭

  However, 점수/이상형조사에 의한 매칭커플시스템은 실제 만남과는 괴리가 있음.
  Therefore, 실제 성혼수와 성혼율을 증가시키는 것이 제한적이라 판단됨.

  ➡ Business Problem!!

- 성혼커플의 특성이 조사된다면, 보조적인 매칭시스템을 개발하여
  신규회원이 성혼에 이르는 시간과 비용을 단축시킬 수 있을지도.....

2. Data Mining Problem

- Data to be obtained : 성혼커플 1000명의 소득, 외모, 키 등 여러 범주의 데이터
- Data to be needed : 1000명 이상의 데이터, 성혼에 실패한 경우의 데이터
- Data Mining Framework : Clustering
  (k-means clustering using Gower's similarity measure)
- 1. 남성과 여성을 각각 Clustering 해서, 각 Cluster별 특성 파악

  2. 임의의 Cluster-Wn 과 Cluster-Mn 사이에 연관관계가 있는지 확인

  3. 만약, Clusters끼리 유의미한 관계가 없다면, 한 Cluster의 여성들이 결혼한 남성들이 공통적으로 어떤 특성을 갖고 있는지 확인

3. Expected Result & Business Implication

- 개인별 조건으로 묶은 Cluster가 선호하는 이성의 특정 모습이 있다.

  Ex) 가족이 네명, 서울에 사는, 연봉 4000만원,,,의 여성들은 외동, 경기도 남성과??

- 비즈니스 문제 해결 :

  1. 기존 매칭시스템에 대한 보완
  2. 부수적으로 Cluster별로 마케팅기법을 적용