

인턴 일지

소비데이터 이상탐지

목차

1. 짧은 시간에 이동거리가 먼 경우
2. 같은 사용자/같은 시점에 다른 상점이 매칭된 경우
3. 거래 금액이 0보다 작은 경우
4. 카드이름이 이상한 경우
5. 그외

1. 짧은 시간에 이동거리가 먼 경우

- 특정 장소 결제 후 그다음 결제 장소 까지 이동 속도가 비정상적 일때
ex. 100km, 250km, 1000km등등
- 온라인 몰, 택시 결제 → 제외
ex. 네이버, gmarket, 11번가 등
- 상점 매칭 오류로 추정
ex. 이전 결제로 보아 인천터미널 부근에 있는데,
cgv인천터미널이 아닌 cgv인천으로 상점이 매칭됨
→ 상점 이름이 짧려와서 그런 게 아닐까 추정

2. 시간 차가 0인 경우

- 동일한 사용자, 동일한 시간(초 단위)에 다른 지점에서 거래가 발생한 경우(금액은 다름)
- 이유 추정: 카드사 정보 등이 넘어올 때 거래 시점 정보가 없으면(초단위 혹은 일 단위로 들어온다면), 해당 정보가 넘어올 때의 시간으로 됨
→ REG_DT를 확인하면 될까?

	index	EPOPCON_ID	COMPANY_ID	CO_NAME	LATITUDE	LONGITUDE	DEAL_DT	CARD_NAME	predict
1021	10249	eppcn157a41ab399bd0366d4be1656	819410	제이엠마트	37.8927981	127.0517055	2015-02-11 15:13:37	현대카드L	1

```
1 test_df[test_df.index == 1022]
```

	index	EPOPCON_ID	COMPANY_ID	CO_NAME	LATITUDE	LONGITUDE	DEAL_DT	CARD_NAME	predict
1022	12129	eppcn157a41ab399bd0366d4be1656	6902514	제일축산물도매센터	37.7405925	127.0509105	2015-02-11 15:13:37	현대카드L	1

시간 차가 0이지만 REG_DT 가 다를 경우

[illegible]

3. 거래 금액이 0보다 작은 경우

- 환불, 할인 등?
- 특정 거래가 발생한 뒤 바로 취소하는 경우

<input type="checkbox"/>	135415	eppcn15755211599	3841117	휴넷 개발	37.5294842931704	126.875174478401	2013-08-15 15:58:00	삼성카드(4760)	생활	생활	18000	0
<input type="checkbox"/>	135416	eppcn15755211599	3841117	휴넷 개발	37.5294842931704	126.875174478401	2013-08-15 20:08:00	삼성카드(4760)	생활	생활	13500	0
<input type="checkbox"/>	135417	eppcn15755211599	5007094	스타벅스 오목교역점	37.5237308	126.87667	2013-08-15 20:16:00	삼성카드(4760)	카페/디저트	카페	8200	0
<input type="checkbox"/>	135418	eppcn15755211599	5007094	스타벅스 오목교역점	37.5237308	126.87667	2013-08-15 20:17:00	삼성카드(4760)	카페/디저트	카페	-8200	0
<input type="checkbox"/>	135419	eppcn15755211599	5007094	스타벅스 오목교역점	37.5237308	126.87667	2013-08-15 20:19:00	삼성카드(4760)	카페/디저트	카페	12200	0

- 그냥 돈금없이 발생한 경우? → 이전/이후 거래 확인

<input type="checkbox"/>	index	EPOPCON_ID	COMPANY_ID	CO_NAME	LATITUDE	LONGITUDE	DEAL_DT	CARD_NAME	CATE	CATE1	PAYMENT	p
<input type="checkbox"/>	39736	eppcn15790cbfe99t	36087	G마켓	37.4998063	127.0366235	2016-09-09 10:32:03	하나카드(5*6*)	온라인쇼핑	온라인쇼핑	0	0
<input type="checkbox"/>	39737	eppcn15790cbfe99t	748870	육대장 대치점	37.497316	127.0549802	2016-09-07 18:10:30	하나카드(5*8*)	외식	한식	8000	0
<input type="checkbox"/>	39738	eppcn15790cbfe99t	306122	롯데마트 송파점	37.4917611	127.1178006	2016-09-07 20:12:56	롯데카드(0*7*)	마트	대형마트	23160	0
<input type="checkbox"/>	39739	eppcn15790cbfe99t	952778	파리바게뜨 가락시장	37.4950385	127.1170868	2016-09-08 19:46:00	하나카드(5*6*)	카페/디저트	베이커리	6100	0
<input type="checkbox"/>	39740	eppcn15790cbfe99t	36087	G마켓	37.4998063	127.0366235	2016-09-09 10:31:40	하나카드	온라인쇼핑	온라인쇼핑	-17000	0
<input type="checkbox"/>	39741	eppcn15790cbfe99t	36087	G마켓	37.4998063	127.0366235	2016-09-09 10:32:03	하나카드	온라인쇼핑	온라인쇼핑	-14500	0

이전/이후 거래 확인(이상한 data)

index	ID	EPOPCON_ID	ES_ID	COMPANY_ID	TIMESTAMP	DEAL_DT	CARD_NAME	CARD_ID	CARD_APPR_NUM	PAYMENT	ISS
326052	170634932	cjcis15a39cc1ae946d9170e4d2d00		8116966	1521683034000	2018-03-22 10:43:54	신한카드	0	-1	-186700	
269483	156508974	cjcis15a39cc1ae946d9170e4d2d00		8116966	1523069261000	2018-04-07 11:47:41	신한카드 (5168)	0	-1	9600	
326058	170635665	cjcis15a39cc1ae946d9170e4d2d00		8116966	1523583834000	2018-04-13 10:43:54	신한카드	0	-1	-681550	
326054	170635301	cjcis15a39cc1ae946d9170e4d2d00		8116966	1523929434000	2018-04-17 10:43:54	신한카드	0	-1	-540750	

- Company_id를 통해 확인 결과 (ex.제일의료재단)
- 비정상적인 환불?

이전/이후 거래가 존재 하지 않을경우

EPOPCON_ID	ES_ID	COMPANY_ID	TIMESTAMP	DEAL_DT	CARD_NAME	CARD_ID	CARD_APPR_NUM	PAYMENT	I
cjcis15a39b152e946d9170e4ce500		4983330	1524705258000	2018-04-26 10:14:18	삼성카드 (7529)	0	-1	-3329000	

- Company_id 확인 결과 공영홈쇼핑
- 이전 이후 거래가 없지만 – 거래 존재

4.카드 이름이 불완전하거나 이상한 경우

- 재전송(거래기록 확인)

102326875	cjcis15a392210b267560d56e3c100		950788	1.51559E+12	2018-01-10 오후 10:55:59	롯데카드(2*5*)	0	-1	258950	100.0,0.0	-1	2018-01-010 오후 10:56:14	2018-01-10 오후 10:56:14
102375637	cjcis15a392210b267560d56e3c100		950788	1.51562E+12	2018-01-11 오전 6:02:15	재전송(2*5*)	0	-1	17600	100.0,0.0	-1	2018-01-011 오전 6:10:29	2018-01-11 오전 6:10:29

- RE)가 붙어 있는 경우
- 복지 카드는 체크카드? or 신용카드?
- 신한 이라 표시 되어있을때 (신한 신용카드 인지, 체크카드인지)
해당 id의 카드 사용 내역으로 판단?

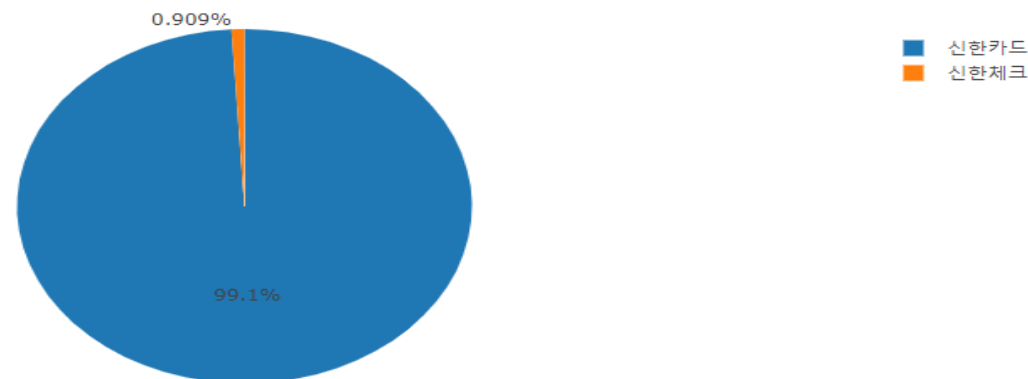
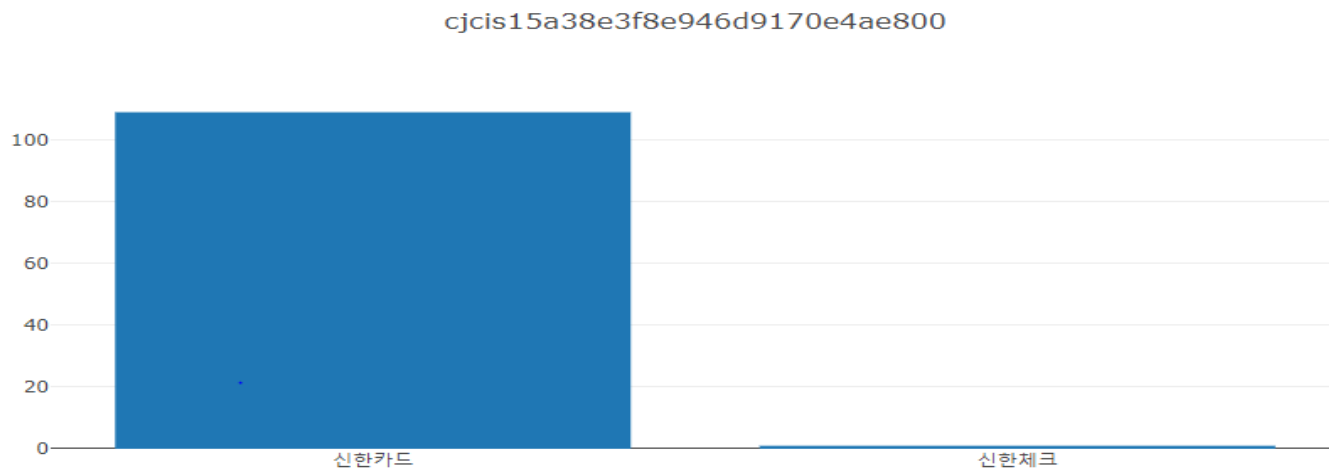
신한 이라 표시 되어있을때 (신한 신용카드 인지, 체크카드인지)

• 해당 id의
카드 내역으로 판단?

이경우

신한카드 : 109건

신한체크 : 1건



5. 그외

- DEAL_DT가 덜 차있는 경우 (00:00:00)
→ (비율 : 0.00168% 35만개 중 6개)

EPOPCON_ID	COMPANY_ID	CO_NAME	LATITUDE	LONGITUDE	DEAL_DT	CARD_NAME	CATE	CATE1	PAYMENT	predict
eppcn15702790e99	1017006	한솔도시락 국기원점	37.5004619	127.030865	2015-06-21 17:39:0	우리카드(5*3*)	외식	도시락	5000	0
eppcn15702790e99	328770	맛집골창구이	37.4780056	126.983744	2015-06-21 21:43:0	우리카드(5*3*)	외식	한식	58000	0
eppcn15702790e99	518258	세븐일레븐 양재점	37.47154366169184	127.0450604275348	2015-06-22 07:52:0	우리카드(5*3*)	마트	편의점	4400	0
eppcn15702790e99	918472	쿠팡	37.5073967	127.056841	2015-06-23	우리카드(5*3*)	온라인쇼핑	온라인쇼핑	-19800	0
eppcn15702790e99	5516696	하이베스트문구 교대점	37.4907467	127.0102191	2015-06-23 08:33:0	우리카드(5*3*)	교육	사무용품	11500	0
eppcn15702790e99	187237	소풍엔	37.5042471	127.0357009	2015-06-23 20:00:0	우리카드(5*3*)	생활	편의	5500	0

단순히 00:00:00 일때 거래 기록일 가능

- 데이터 중복
 - 문자 메시지 결제 후 카드사 연동 시 데이터가 중복될 수 있음
 - 탈퇴 후 재가입한 케이스를 걸러야 함