

안녕하세요 이번 하계 인턴을 하고 있는 박지훈 입니다.

저번 주 회사 내 meums_comp_revisit_his(거래 기록)와 meums_company table(상점 정보)의 데이터를 살펴 보면서 **이상이 의심되는 데이터**에 대해 회사 사람들과 공유하고자 정리 하였습니다.

1. 속도와 시간

우선 저는 data 를 살펴보기 위해 각 epopcon_id 별 한 결제와 다음 결제 사이의 이동 속도에 대해 주목 했습니다. 그러다 크게 두 가지의 이상 경우가 발생 했습니다.

1) 속도

첫 번째 짧은 시간 동안 상식적으로 이동 할 수 없는 거리를 이동하는 경우 입니다. 현재 결제가 일어난 장소를 A 그 다음 결제가 이루어진 장소를 B 라고 할 때 A 와 B 의 물리적인 거리 차에 비해 결제 시간의 차이는 1~2 분 정도로 비정상적 이였습니다.

data 확인 결과 온라인 결제 일 경우 현재 위치와 상관없이 온라인 물의 등록 위치로 결제 정보가 표시 되었습니다. 택시 결제 역시 현재 위치가 아닌 택시 본사의 위치가 찍히는 것을 발견 하였습니다.

온라인 물 과 택시 결제를 지우고 다시 살펴 보았지만 여전히 문제가 발견되었습니다.

이때 발견된 문제의 경우 예를 들어 CGV 인천터미널 점에서 결제 후 1 분 뒤 CJCGV 결제가 이루어 지는 경우였습니다. 이는 자주와 어플을 통해 받아오는 상점 이름 정보가 누락 되어 디비에서도 이름이 잘못 매칭된 결과인 것 같습니다.

2) 시간

두 번째 현재 결제 시간과 다음 결제 시간이 같은 시간 차가 0 인 경우였습니다.

이 경우엔 두 가지로 나뉩니다.

- 1) 시간이 같고 결제 장소 역시 같은 경우,
- 2) 시간은 같지만 결제 장소가 다른 경우

로 나뉩니다.

첫 번째의 경우 자주와 어플을 사용 시 카드사 연동을 했을 때 자주와 어플을 통해 회사로 들어오는 정보가 문자와 카드사의 정보 두 가지 이기 때문에 발생 할 수 있다고 들었습니다.

두 번째 경우엔 회사 이름과 장소가 다르지만 같은 시간에 결제가 이루어 진 경우 입니다.

이 경우 정보가 넘어올 때 거래 시점 정보가 없거나 누락될 경우 해당 정보가 넘어 올 때

시간(REG_DT)으로 설정 될 수 있다 들었습니다. 그러나 이를 감안 하고도 시간(DEAL_DT)이 같고 다른 REG_DT 인 경우가 존재 합니다. 추후 확인이 필요할 것 같습니다.

In [35]: 1 extra[0]

Out [35]:

DEAL_DT	CARD_NAME	CARD_ID	CARD_APPR_NUM	PAYMENT	ISSUED_COMPANY_ID	DEAL_LOCATION	LIFEZONE_HIS_ID	TRANS_YN	REG_DT	UPT_DT
2018-01-12 10:06:03	롯데카드 (2*2*)	0	-1	192540	10	0.0,0.0	-1	0	2018-01-15 10:06:45	2018-01-15 10:11:28

In [36]: 1 extra[1]

Out [36]:

DEAL_DT	CARD_NAME	CARD_ID	CARD_APPR_NUM	PAYMENT	ISSUED_COMPANY_ID	DEAL_LOCATION	LIFEZONE_HIS_ID	TRANS_YN	REG_DT	UPT_DT
2018-01-12 10:06:03	롯데카드 (2*2*)	0	-1	29660	10	0.0,0.0	-1	0	2018-01-15 10:07:15	2018-01-15 10:07:15

2.거래 금액 및 결제 수단(카드 정보)

속도를 위해 거리와 시간을 확인 해보는 중 거래 금액이 0 보다 작은 경우도 발견 되었습니다. 이는 환불이나 할인의 상황이 발생 할 경우 일 듯 합니다. 그러나 몇몇의 EPOPCON_ID 의 경우 종 결제 금액의 합이 0 보다 작는데 이는 결제 기록보다 환불이나 할인 기록이 많다는 것을 의미 하는 것이라 생각하니 이상하다 여겨 이전과 이후의 거래를 살펴보았습니다.

첫 번째 이전과 이후의 기록 역시 가격이 - 인 경우가 존재 했습니다. (EX, 제일의료 재단에서의 거래 기록) 이는 아마도 보험금 수령일 수도 있겠다 생각 합니다.

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	1523829434000	2018-04-17 10:43:54	신한카드	0	-1	-540750	

두 번째 이전과 이후의 기록이 없고 오직 한 건의 거래만 이루어 졌고 이 거래의 가격은 -인 경우였습니다. 더 많은 경우를 살펴봐야 할 거 같습니다.

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

또한 가격을 살펴보면서 카드 이름 역시 정규화가 이루어지지 않는 것을 발견 하였습니다.

이름에 RE 가 붙어있거나 단순히 재전송이라고만 표시되어있는 경우, 이름이 단순히 신한카드이거나 국민 카드 등 체크카드인지 신용카드인지 정확하지 않은 경우 등이 있습니다.

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