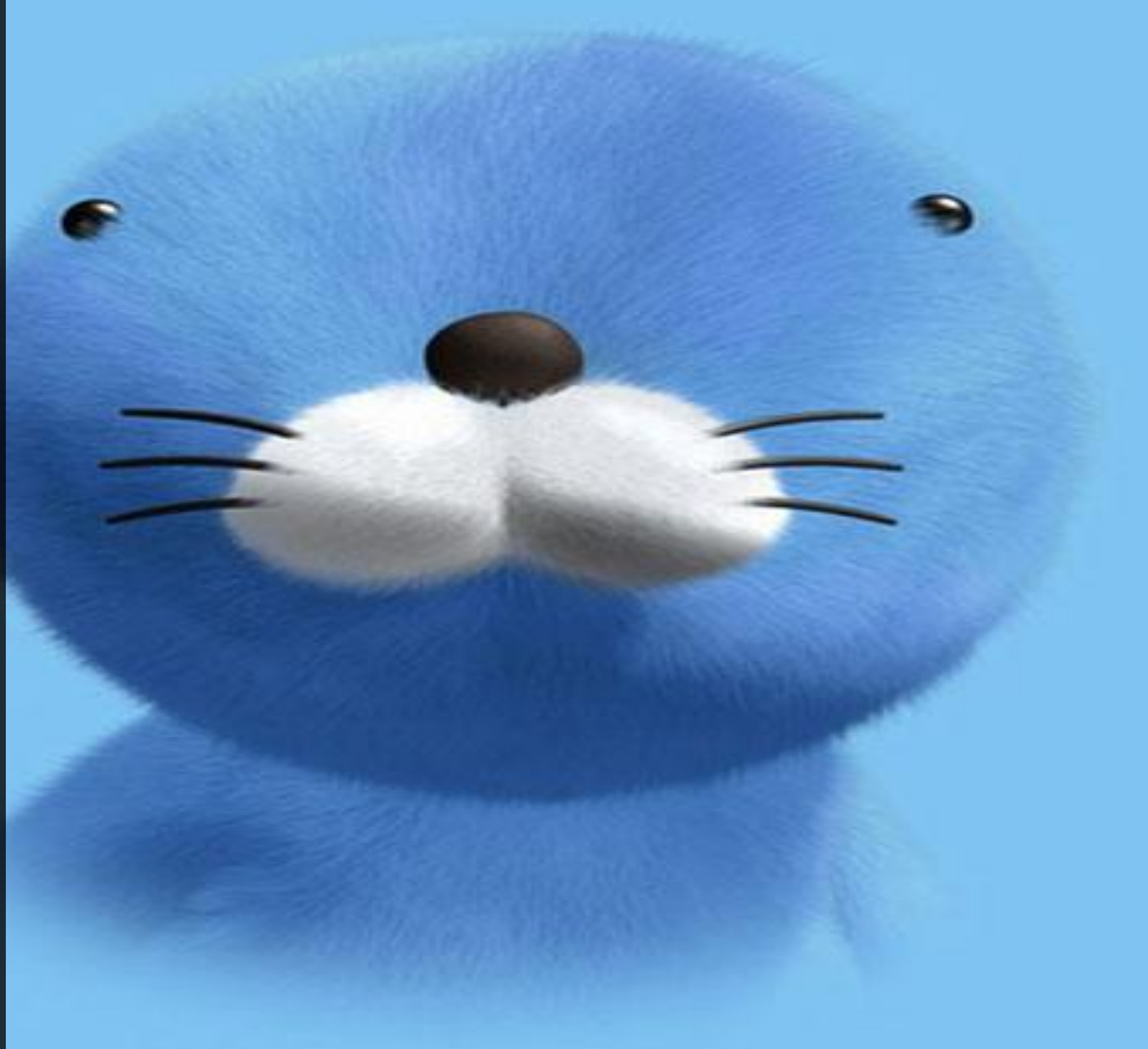


# Human or Robot?

- 소셜 네트워크 분석 통한 Bot 추적

김태관 김석준



# CONTENTS

---

01

프로젝트 배경

프로젝트 소개  
프로젝트 목적

02

데이터

데이터 수집 &  
문제정의 & 데이터 소개

03

EDA(탐색적 데이터분석)

기존데이터 탐색  
파생변수 탐색

04

결과

결론도출

01

프로젝트 배경

# 01.프로젝트 소개

불법 프로그램(메크로)사용자가 증가함에 따라 기존의 고객들의 불만이 증가하고 이탈하는 고객이 많이 생긴다.

게임 아이템이 우수수...불법 프로그램으로 1억4천 채킨 4명

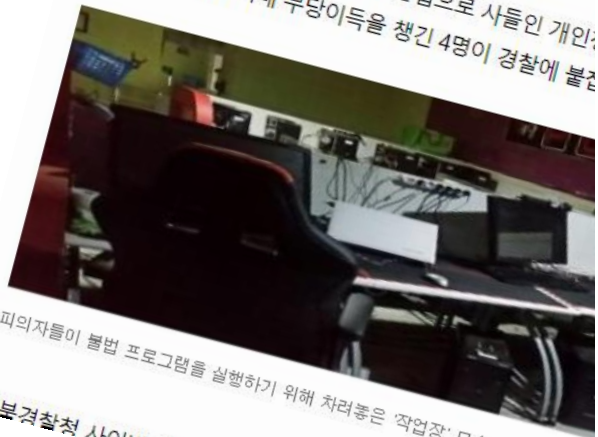
송고시간 | 2017/10/17 10:26



작업장 차려놓고 개인정보 5천300여건 사들여 계정 (전주=연합뉴스) 임재두 기자 = 불법으로 사들인 개인정보를 훔쳐 판매해 역대 부당이득을 채킨 4명이 경찰에 불려

카카오 '배틀그라운드' 국내 서버 따로 운영 "불법프로그램 사용에 따른 불만"

장주영 기자 | 2017-11-13 17:15:31 | 바이오&IT



피의자들이 불법 프로그램을 실행하기 위해 차려놓은 '작업장' 모습. [전북경찰청 제공] 전북경찰청 사이버수사대는 전북특신망 및 게임사업지출에 관하여



카카오 '배틀그라운드' 국내 서버 따로 운영 "불법프로그램 사용에 따른 불만"

- 의사도 놀란, '탈모개선'새까맣게 짝 차..
- 아빠 173cm + 엄마 158cm = '우리 아이는...

카카오게임즈가 '배틀그라운드' 국내 서버를 따로 운영할 예정이다.

뉴스속보 매학기 벌어지는 수강신청 클릭전쟁, 대학교는 '고심중'

기사입력 2016-02-24 07:04

전국의 대학교에선 '수강신청'을 근처 PC방에도 학생들이 몰려 접속자로 폭주해 마비되는 지경에 달 메크로 프로그램도 공공연히 사용 '선착순'이라는 구조적 한계를 극복

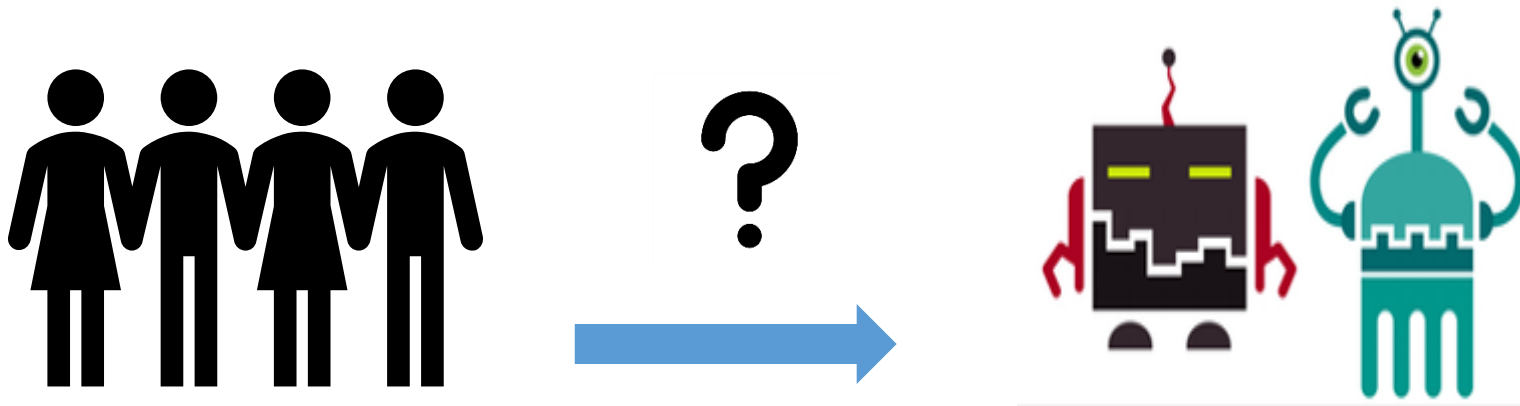


1년해 2학기 '마일리지제'에 기반을 둔 수강신청 전 받은 마일리지를 과목선호도에 따라 배분하는 방식



## 02.프로젝트 목적

Who is a human and a robot?



고객이 일반사용자 인지 메크로(robot) 사용자인지를 탐색한다.

02

데이터 소개

# 01데이터 수집 및 문제정의

데이터수집

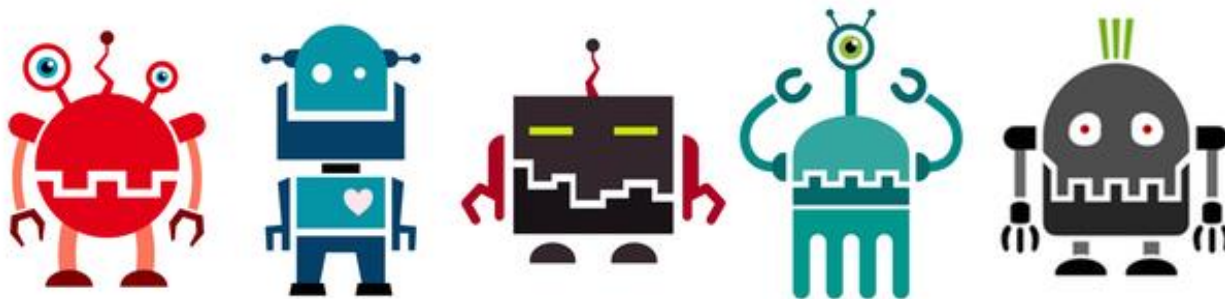
# kaggle

문제정의

## -로봇을 추적한다

온라인 옥션 사이트에서 자동으로 입찰을 하는 봇(robot)이 성행하여 핵심 고객이 급락하고 있다.  
인간 입찰자들은 기계 입찰자들을 이길 수가 없다.

고객 만족을 회복하기 위해 오너들은 봇을 제거할 필요를 느낀다.



## 02.데이터 소개

### 데이터 개괄

- 경매자 Profile 데이터와 경매가 이뤄진 내역이 저장된 Log 데이터로 나뉨

경매자 Profile (약 2013명)

id

Payment 계정

주소

로봇 여부(1/0)

경매 log (약 300만 row)

id

입찰 id

경매 id

상품 종류

device

시간

국가

IP

URL

\* 민감한 개인정보는 정책에 따라 암호화 되어 있음



## 02.데이터 소개

### 데이터 개괄

#### 경매자 Profile

	bidder_id	payment_account	address	outcome
664	46c0ae10d60c4fde2b6301bbafecd7b8dcla1	31b07e6f832c3f123c20c51db4a80bf8oqrrz	000475f9223b613ae0f0939fb7e580ebx9hj	0
1624	74149197c8684977e397ab60a6a264836b0lv	a3d2de7675556553a5f08e4c88d2c228ymxx3	004ea345038ec0fe71240e0a6b440b1dzsg5f	0
15	c0549e76b179912d52a28c3eb9b29376yaxvu	a3d2de7675556553a5f08e4c88d2c228wx3vf	00de21ff3627811075faf2a3a0880053yi40s	0
1814	170a58cf637c08f0818ebdbfdca29c9irkxj	a3d2de7675556553a5f08e4c88d2c228opfyo	00fbee2e7b516ea80cd819debc5b0c9rpj1a	0
1389	2cbacd5b066b7dff7e0a87e1ef00e0b4zpqzw	a3d2de7675556553a5f08e4c88d2c2280m87w	011452b59e47a53fa1818c5d0765dea5pv3f9	0
1395	2d2839fa92d1c8e94bdeb1fbb9e62149x6xc5	b0a4197bc483e81400b84d4d5ada8331mqz7l	01189c5cf2cd2a242ffa812909fb2c9i3f1z	1
228	2918b09c1ed9e9e8602f85175b266aej5v6w	96366bec708b772bd2befe4aef9badb0v5cyh	014c8903c0a7cf8ffdb468f1847b89feix5k9	0
1088	424aed9d807cfe0b2167cfdaa4957fc0heell	7023a9b0ae54fdb8235fa5127ce21f3246s1y	0157f1e009ae2fc517251784bc0c4f0f8mn2d	0

#### 경매 log 데이터

	bidder_id	outcome	bid_id	auction	merchandise	device	time	country	ip	url
1	001068c415025a009fee375a12cff4fcnht8y	0	7179832	4ifac	jewelry	phone561	9.706345e+15	bn	139.226.147.115	vasstdc27m7nks3
2	0030a2dd87ad2733e0873062e4f83954mkj86	0	6805028	obbny	mobile	phone313	9.704554e+15	ir	21.67.17.162	vnw40k8zzokijsv
3	00a0517965f18610417ee784a05f494d4dw6e	0	4574244	gst86	books and music	phone1400	9.642087e+15	ng	160.114.155.198	utvvayooovsqfycf
4	00a0517965f18610417ee784a05f494d4dw6e	0	7525691	b25ls	books and music	phone142	9.708597e+15	in	210.227.213.71	wxmtewnifm8v3kb
5	00a0517965f18610417ee784a05f494d4dw6e	0	6511001	1cbiu	books and music	phone25	9.702977e+15	in	86.245.94.173	q7mzsnjxikn3tq3
6	00a0517965f18610417ee784a05f494d4dw6e	0	7397627	fv0t9	books and music	phone150	9.707885e+15	in	85.250.195.147	rnqfvx02v745bnl
7	00a0517965f18610417ee784a05f494d4dw6e	0	3980517	bpil5	books and music	phone419	9.640088e+15	in	58.209.11.187	jynce1r4nc22vrd
8	00a0517965f18610417ee784a05f494d4dw6e	0	6997002	v772u	books and music	phone150	9.705222e+15	us	76.175.18.78	qx6cww72yw00f1q
9	00a0517965f18610417ee784a05f494d4dw6e	0	4411478	5edd6	books and music	phone25	9.641512e+15	id	185.67.92.189	utvvayooovsqfycf

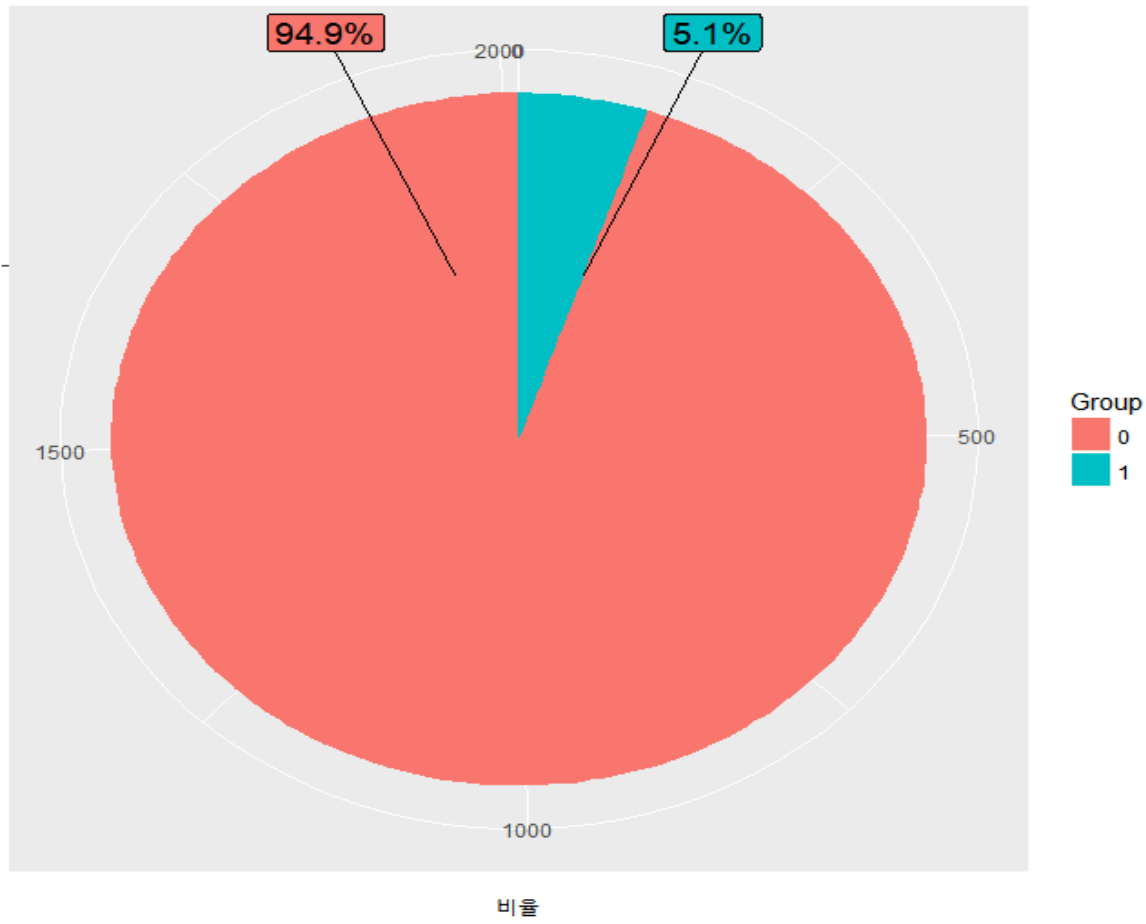
# 03

EDA(탐색적데이터분석)

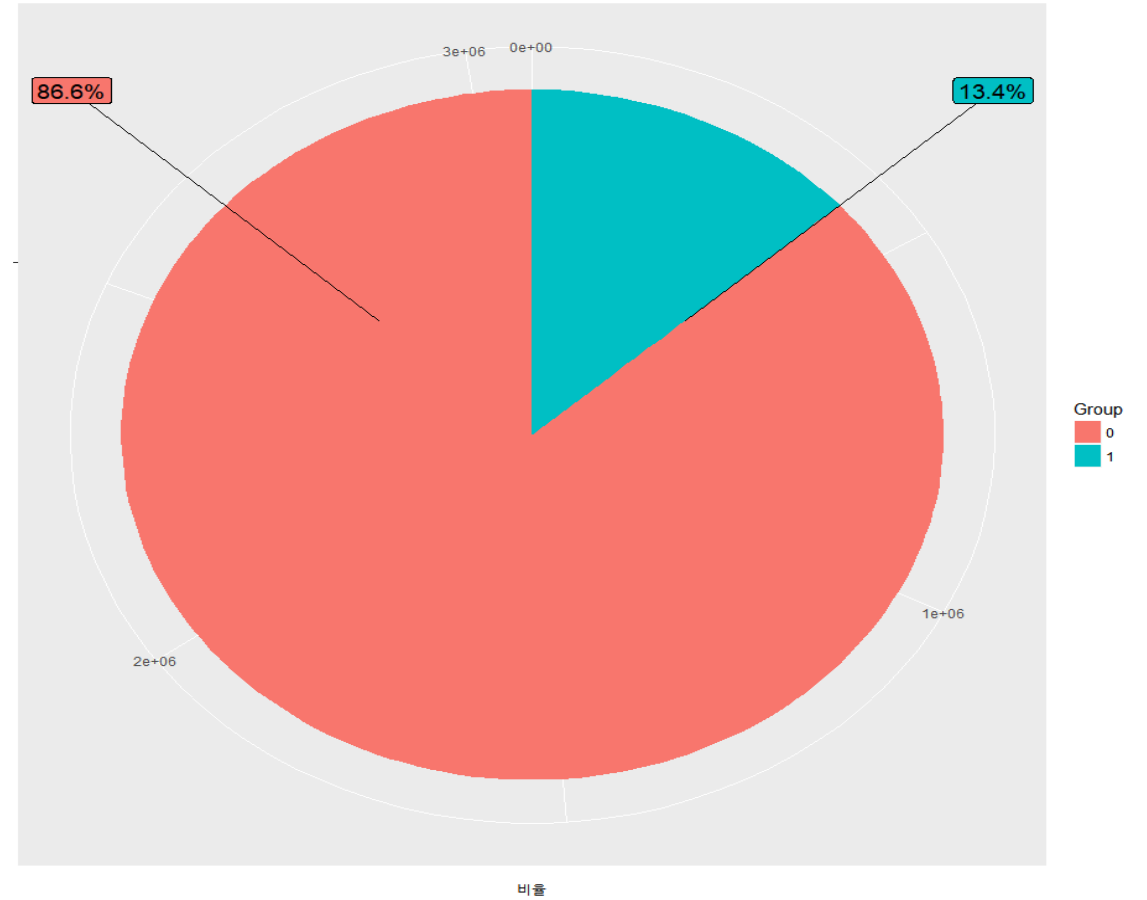
# 01. 기존 데이터탐색

고객 그룹별 비율&총 입찰건수: robot일 경우 입찰건수가 높을 것 이다.

인간과 로봇의 비율

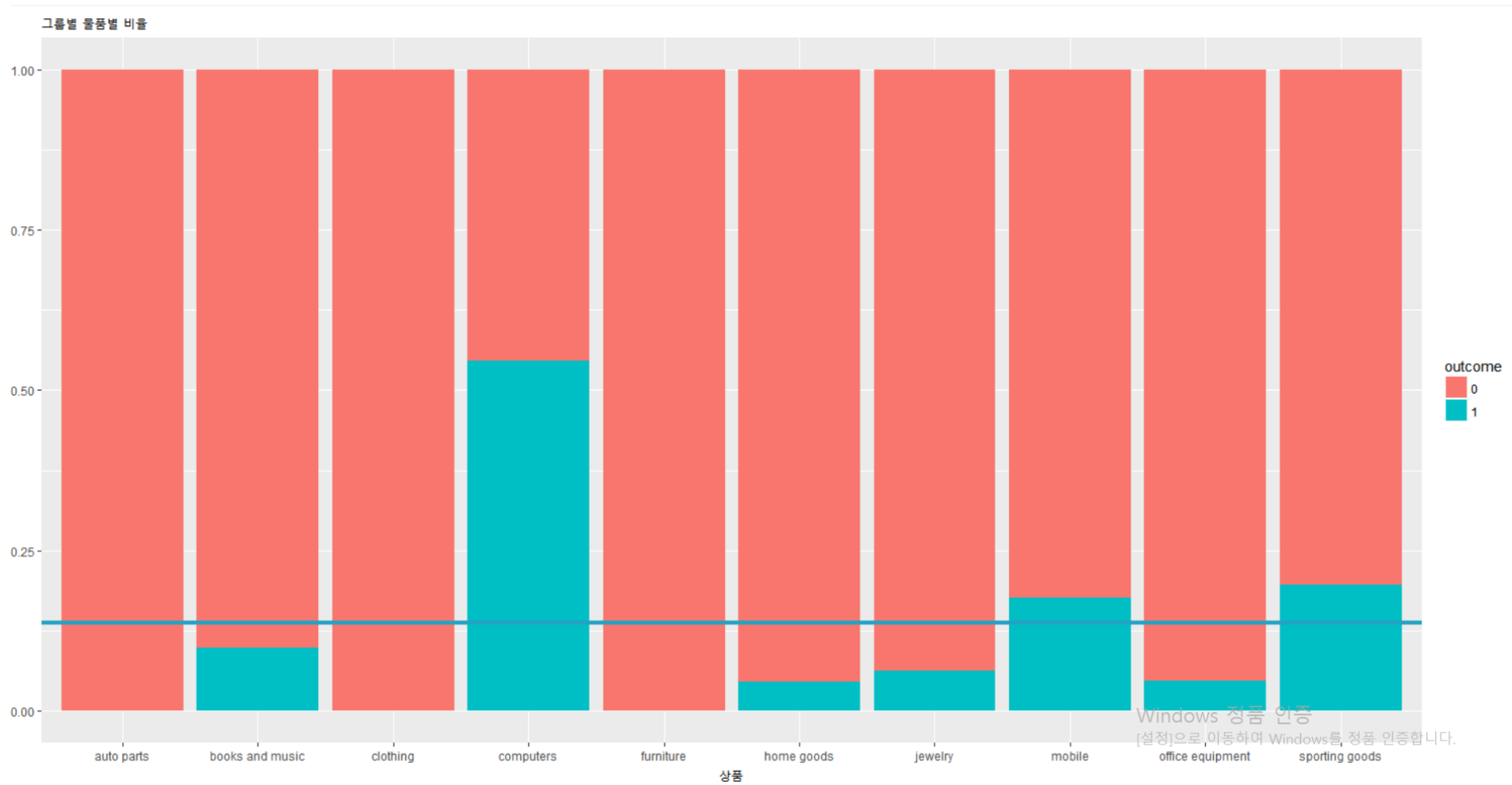


인간과 로봇의 총입찰건수 비율



# 01. 기존 데이터탐색

고객 그룹, 물품별 입찰건수



## 01. 기존 데이터탐색

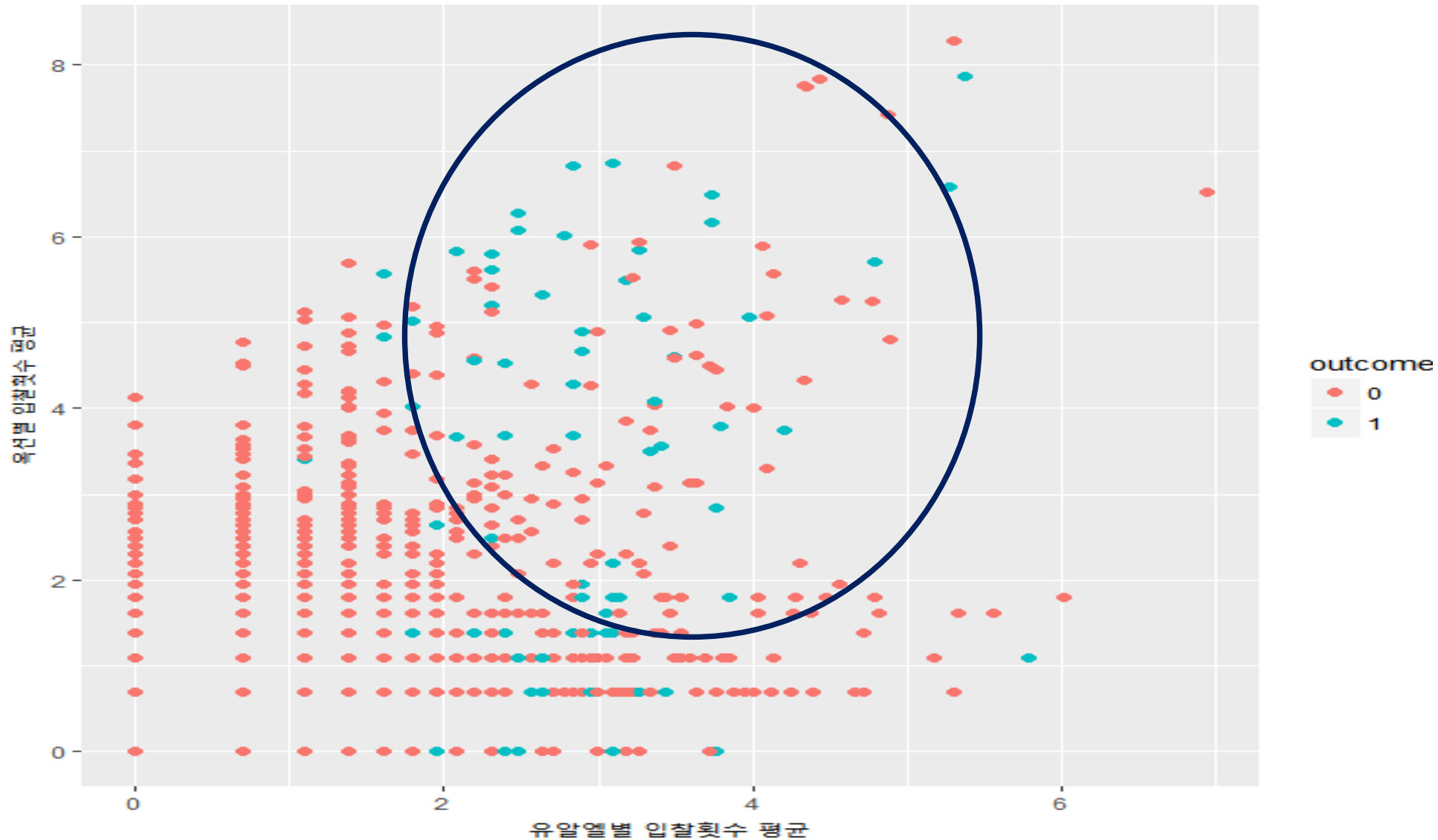
## 고객 그룹, 나라별 입찰건수 : 어느 나라에 robot이 많을 것인가?



## 02. 파생변수 탐색

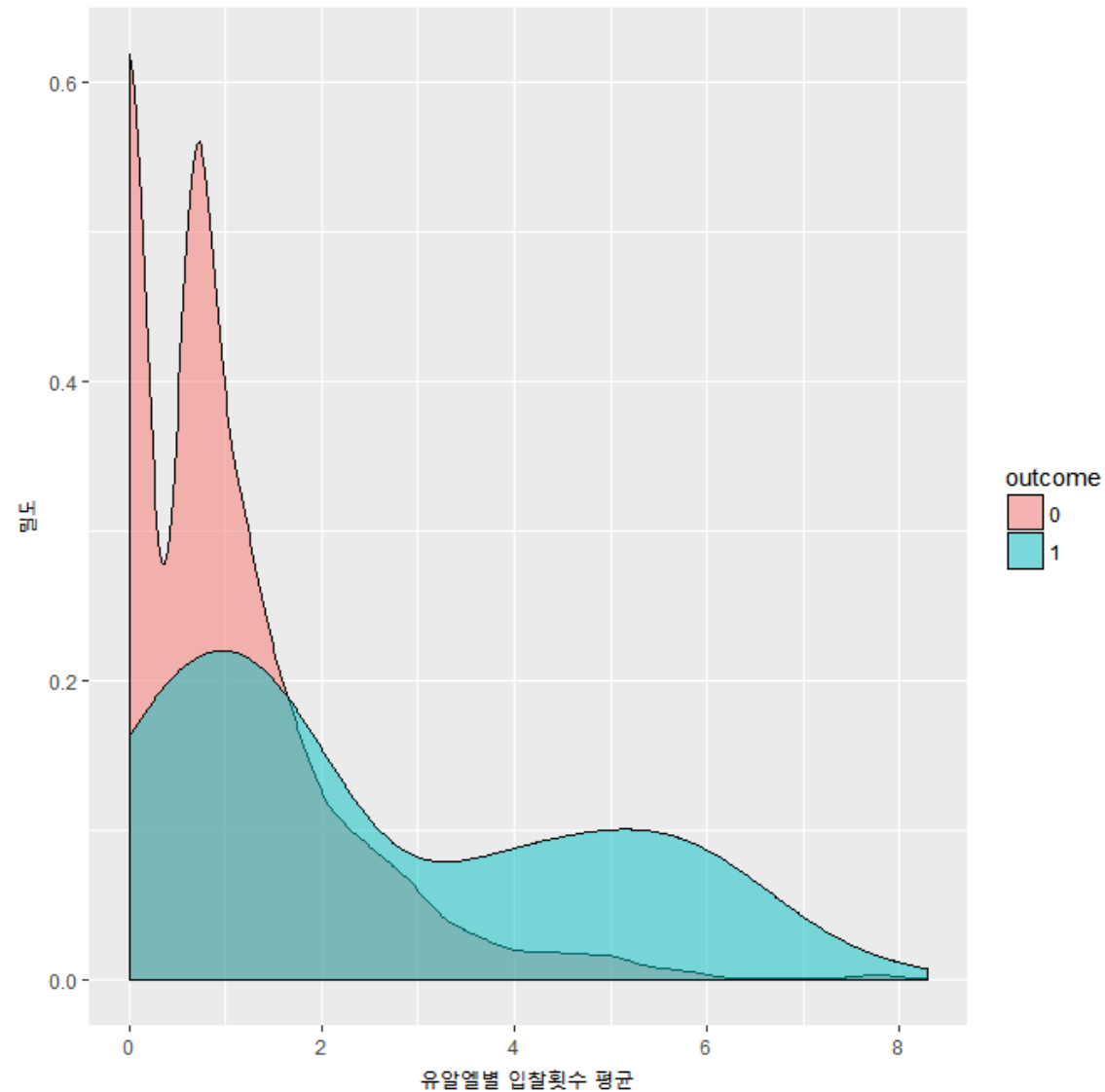
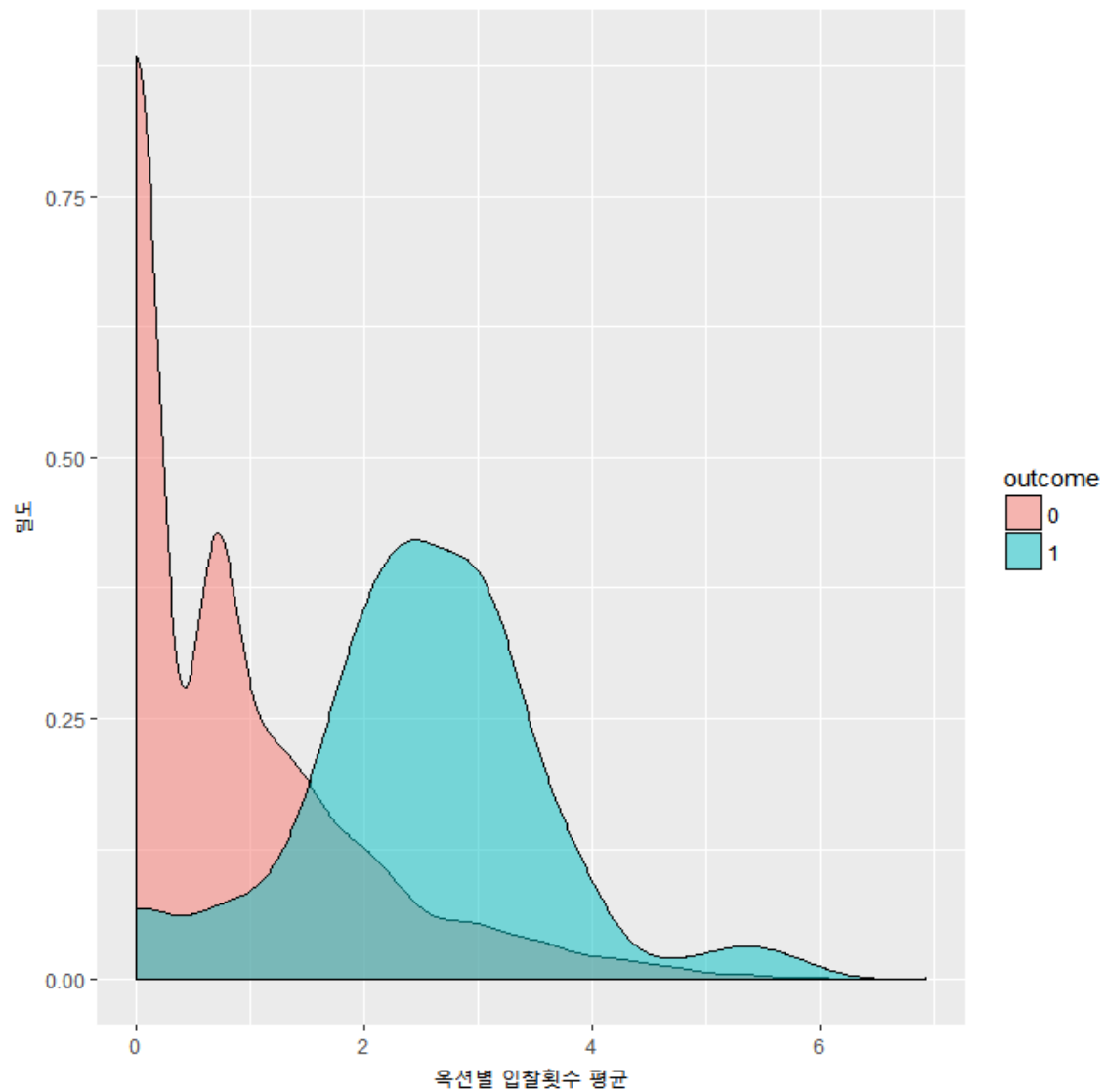
고객 옥션, url 별 평균 입찰건수: robot일 경우 입찰건수가 높을 것 이다.

고객의 옥션별 입찰횟수평균과 유알엘별 입찰횟수 평균의 pointplot



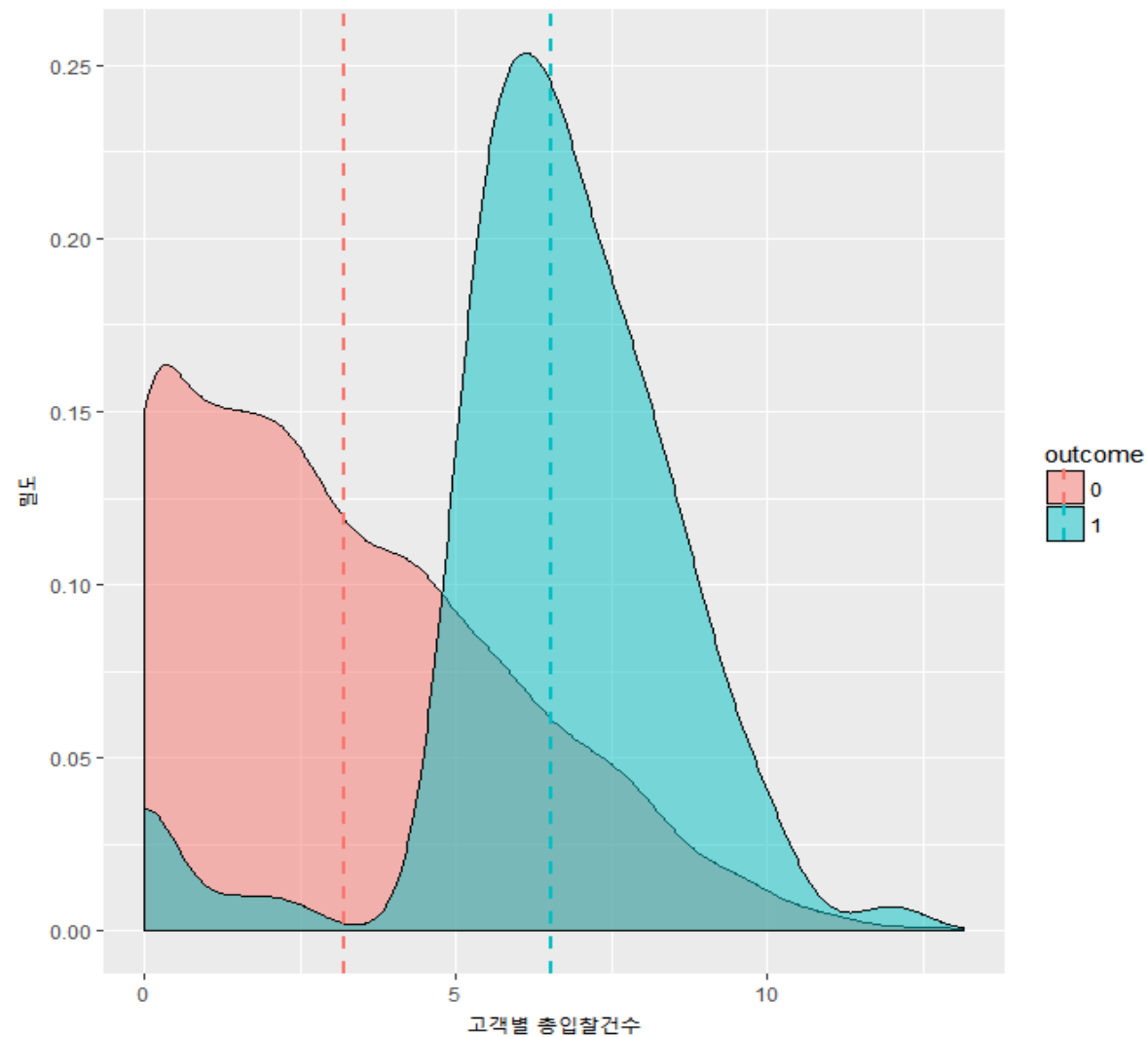
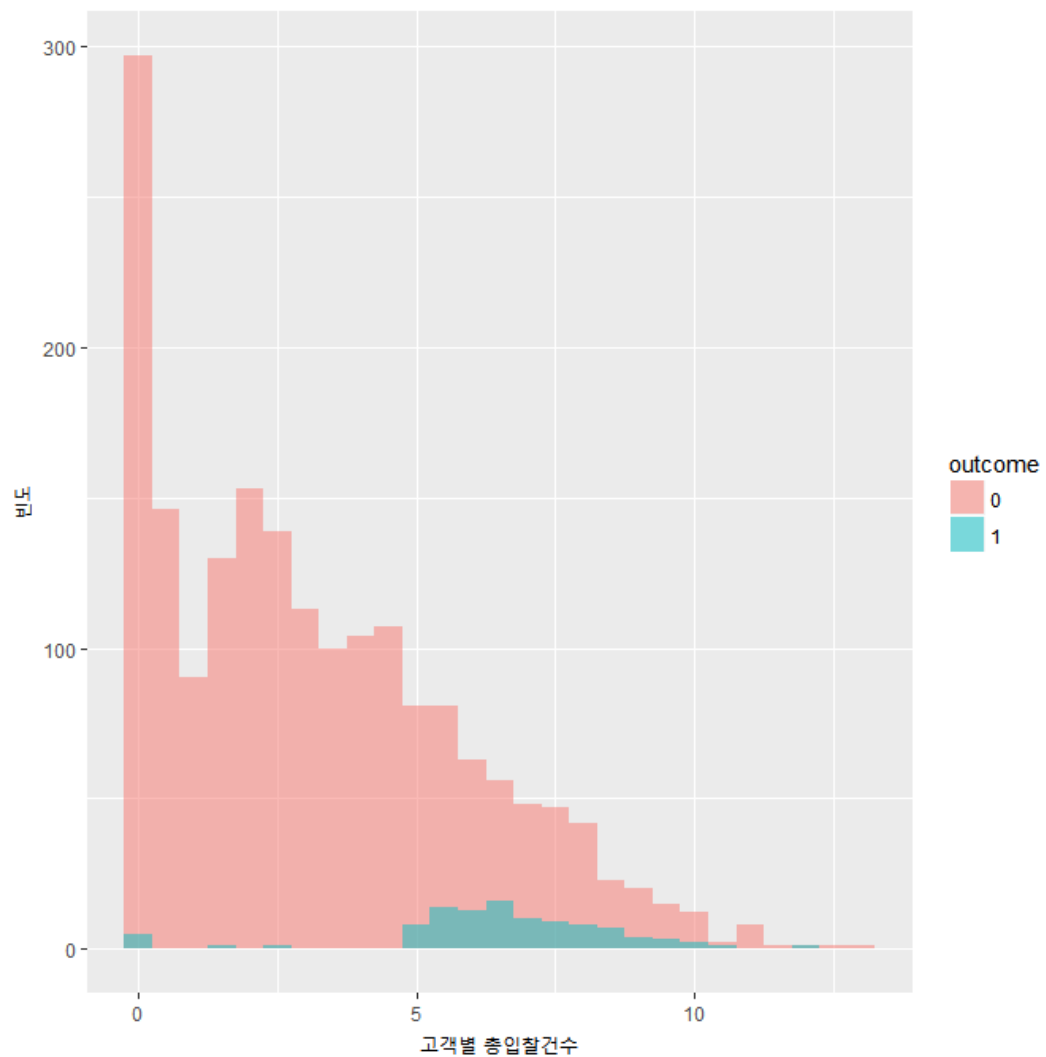
## 02. 파생변수 탐색

고객 옥션, url 별 평균 입찰건수: robot일 경우 입찰건수가 높을 것 이다.



## 02. 파생변수 탐색

고객별 총 입찰건수 : robot일 경우 총입찰건수가 높을 것 이다.





## 02. 파생변수 탐색

고객별 총 입찰시간평균 : 한고객의 입찰과 입찰사이의 시간의 평균. Robot일 경우 입찰과 입찰사이의 시간이 짧은 것이다.

	bidder_id	outcome	bid_id	auction	merchandise	device	time	country	ip	url
1	001068c415025a009fee375a12cff4fcnht8y	0	7179832	4ifac	jewelry	phone561	9.706345e+15	bn	139.226.147.115	vasstdc27m7nks3
2	0030a2dd87ad2733e0873062e4f83954mkj86	0	6805028	obbny	mobile	phone313	9.704554e+15	ir	21.67.17.162	vnw40k8zzokijsv
3	00a0517965f18610417ee784a05f494d4dw6e	0	4574244	gst86	books and music	phone1400	9.642087e+15	ng	160.114.155.198	utvvayoovsqfycf
4	00a0517965f18610417ee784a05f494d4dw6e	0	7525691	b25ls	books and music	phone142	9.708597e+15	in	210.227.213.71	wxmtewnifm8v3kb
5	00a0517965f18610417ee784a05f494d4dw6e	0	6511001	1cbiu	books and music	phone25	9.702977e+15	in	86.245.94.173	q7mzsnjxikn3tq3
6	00a0517965f18610417ee784a05f494d4dw6e	0	7397627	fv0t9	books and music	phone150	9.707885e+15	in	85.250.195.147	rnqfvx02v745bnl
7	00a0517965f18610417ee784a05f494d4dw6e	0	3980517	bpil5	books and music	phone419	9.640088e+15	in	58.209.11.187	jynce1r4nc22vrd
8	00a0517965f18610417ee784a05f494d4dw6e	0	6997002	v772u	books and music	phone150	9.705222e+15	us	76.175.18.78	qx6cww72yw00f1q
9	00a0517965f18610417ee784a05f494d4dw6e	0	4411478	5edd6	books and music	phone25	9.641512e+15	id	185.67.92.189	utvvayoovsqfycf

```
merge<-merge[order(merge$time),]
```

```
E<-merge %>%
  group_by(bidder_id) %>%
  summarise((mean(diff(time))))
E$` (mean(diff(time)))`<-E$` (mean(diff(time)))`/100000000000
```

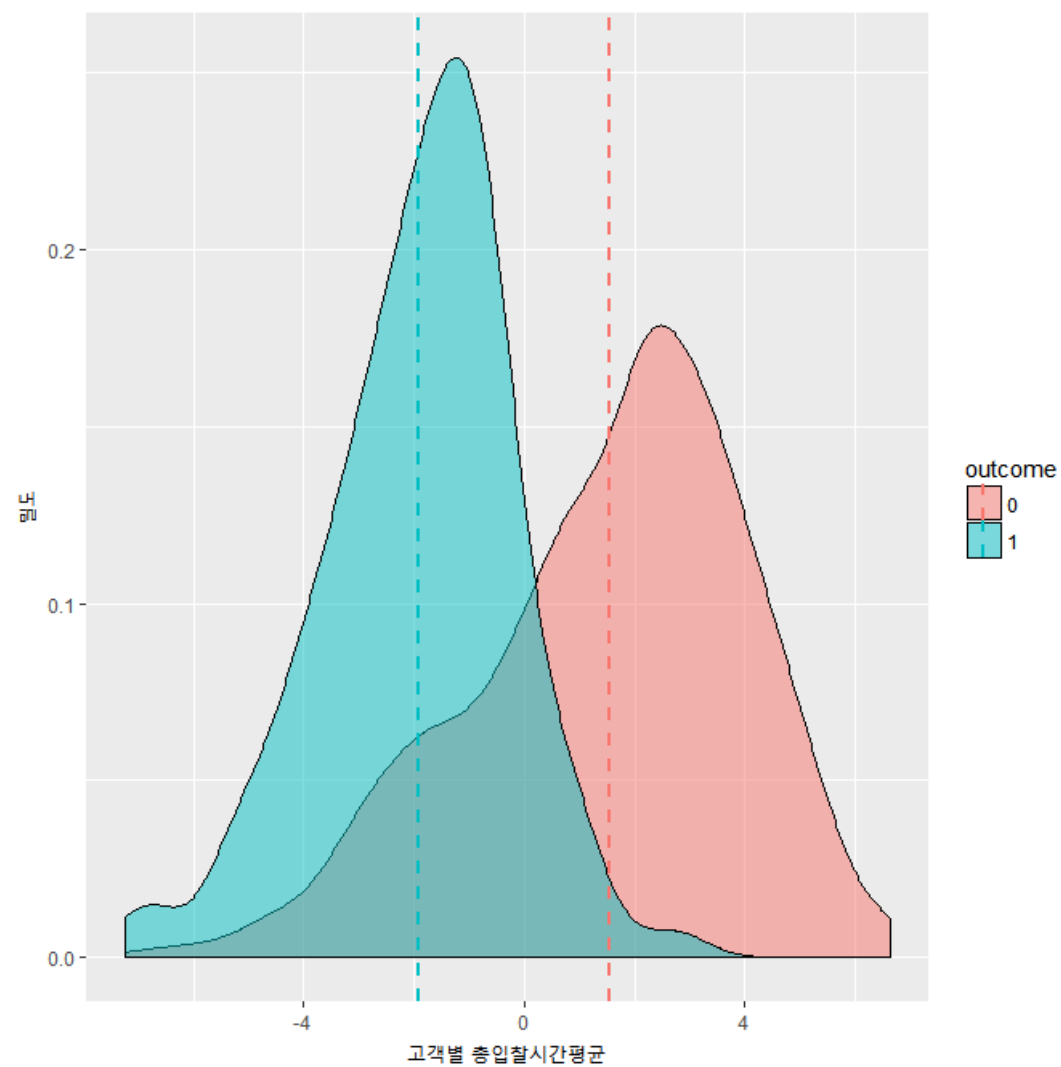
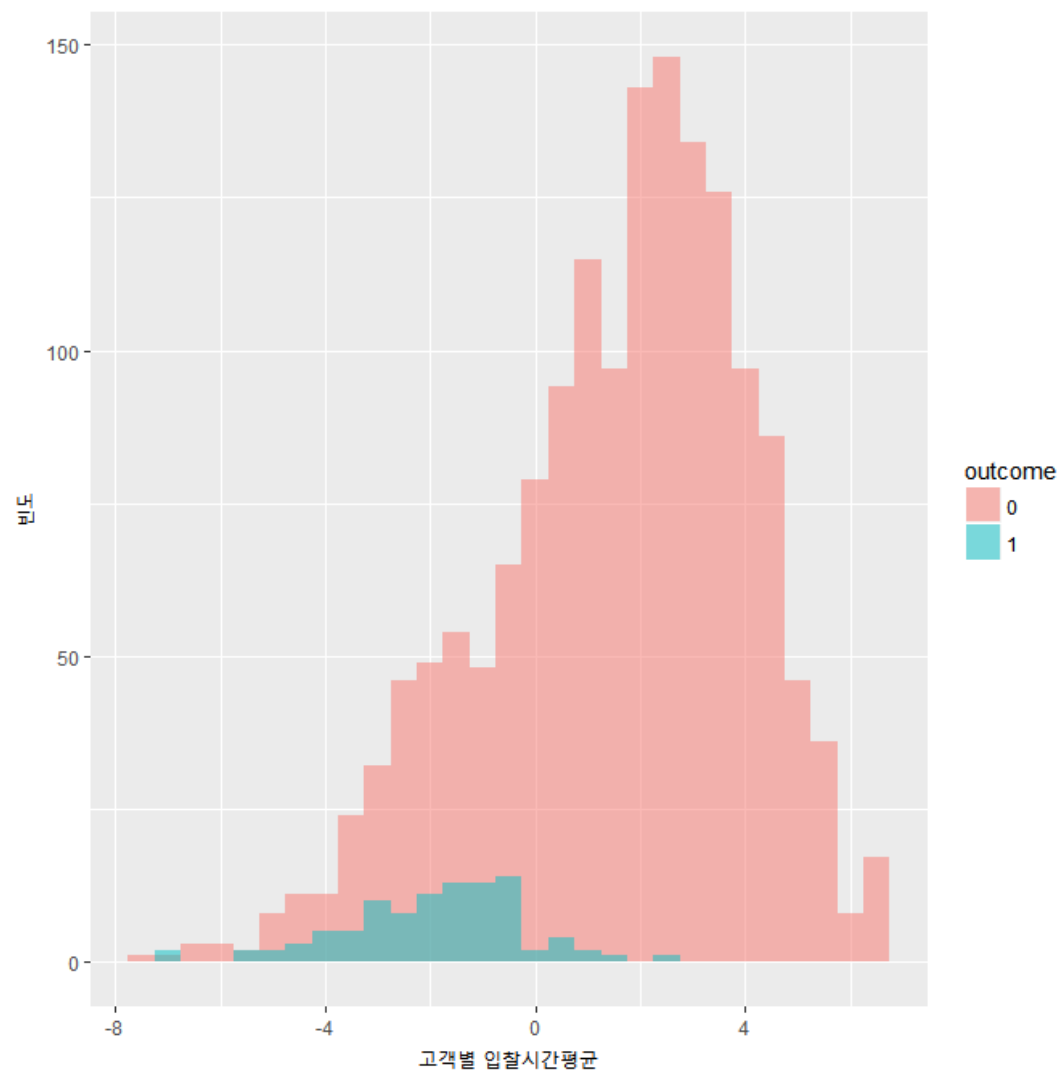


	bidder_id	(mean(diff(time)))
1	001068c415025a009fee375a12cff4fcnht8y	NaN
2	0030a2dd87ad2733e0873062e4f83954mkj86	NaN
3	00a0517965f18610417ee784a05f494d4dw6e	5.486831e+11
4	00cc97158e6f4cb8eac3c0075918b7ffi5k8o	2.748400e+13
5	01067975436d123f717ee5aba0dd4bbfa0937	2.507283e+10
6	012441119bcf83b23d4768bb72cea6d6carua	5.622344e+11
7	01255c2c7c5578c186873422fc00fd7afwk8k	2.074860e+11
8	01349288df20199905e719f0ff7ee771nwryj	1.909924e+11
9	0141844fc725a85e38e12cc02fcdbbe8ofqvx	7.225108e+11

⇒ NaN값은 입찰건수가 1개인 고객

## 02. 파생변수 탐색

고객별 총 입찰시간평균 : 한고객의 입찰과 입찰사이의 시간의 평균. Robot일 경우 입찰과 입찰사이의 시간이 짧은 것이다.

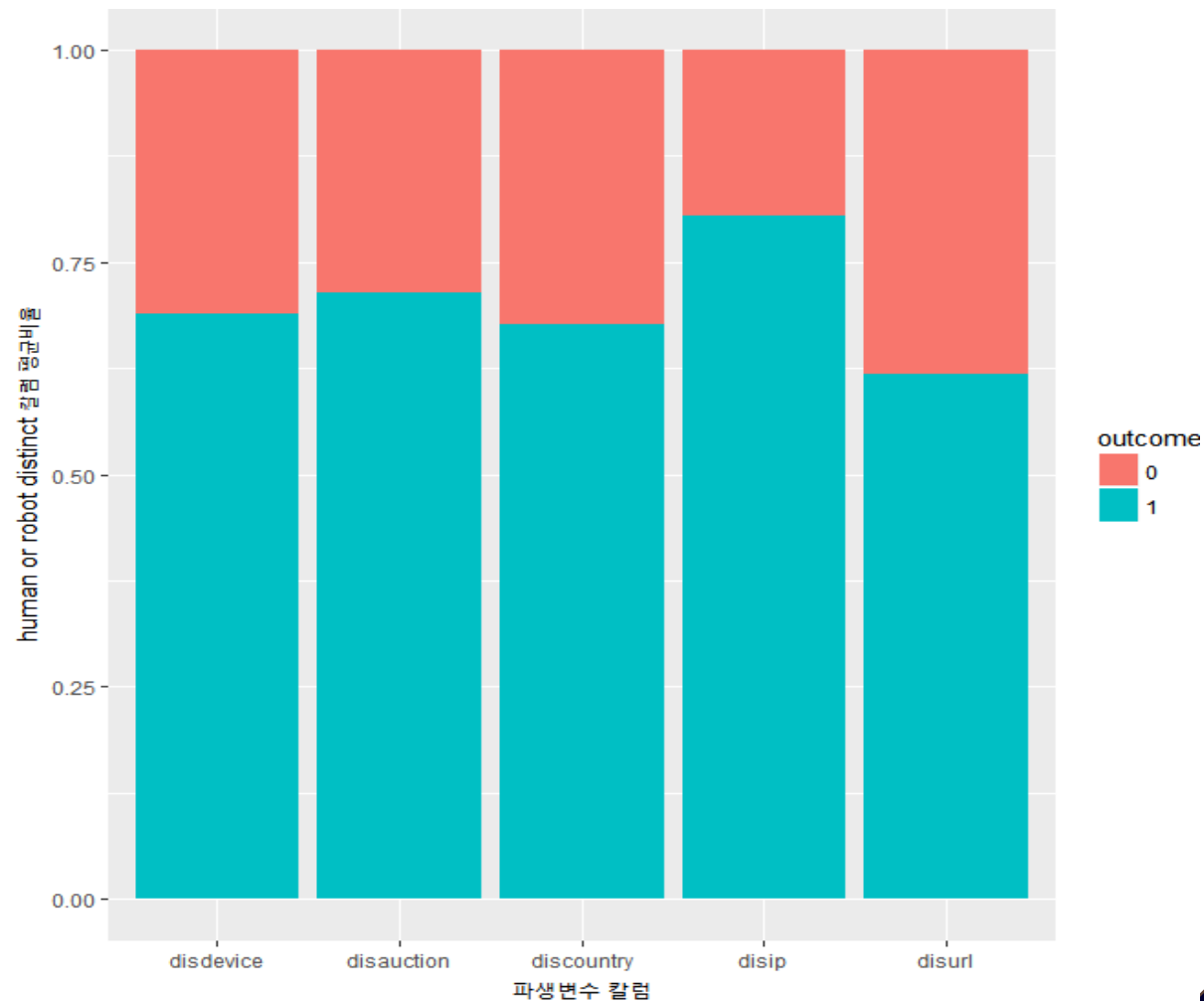


## 02. 파생변수 탐색

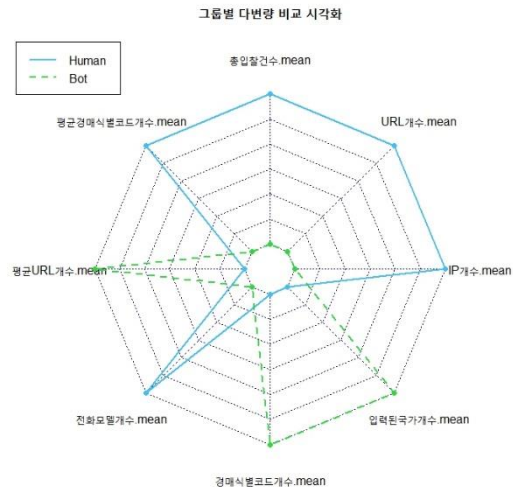
고객별 url, device, country, auction, ip 갯수: 한고객의 입찰과 입찰사이의 시간의 평균. Robot일 경우 입찰과 입찰사이의 시간이 짧은 것이다.

고객별 url, device, country, auction, ip 갯수

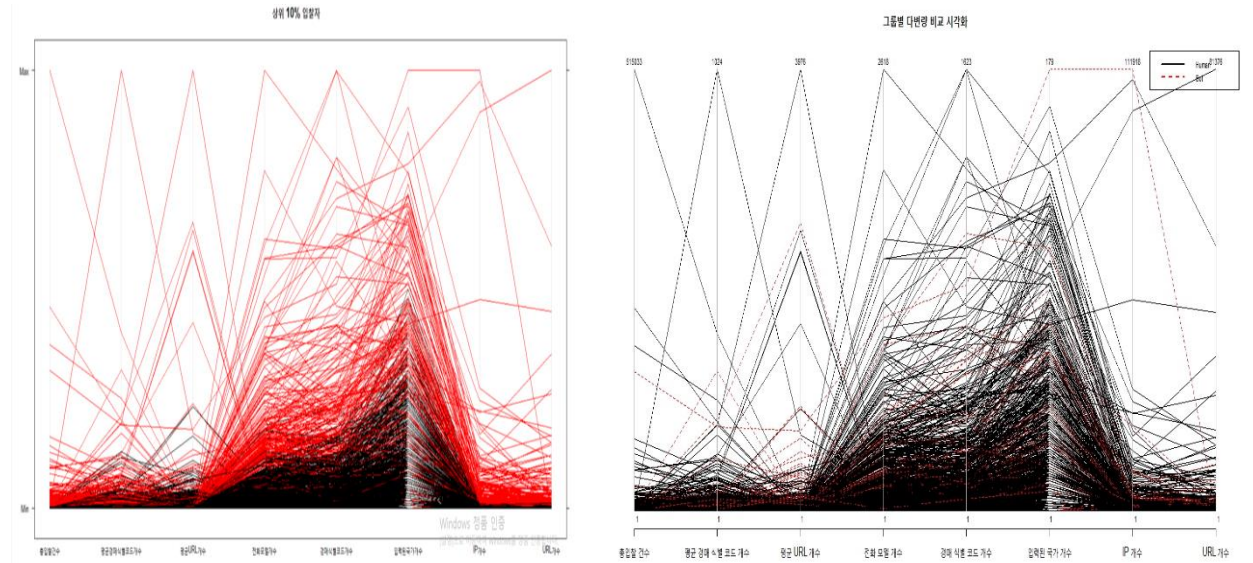
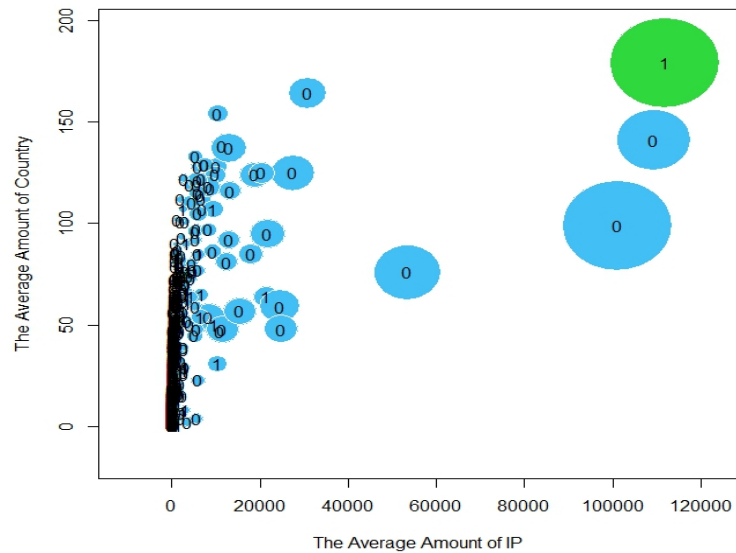
	bidder_id	outcome	disdevice	disauction	discountry	disip	disurl
1	001068c415025a009fee375a12cff4fcnht8y	0	1	1	1	1	1
2	0030a2dd87ad2733e0873062e4f83954mkj86	0	1	1	1	1	1
3	00a0517965f18610417ee784a05f494d4dw6e	0	67	74	16	103	84
4	00cc97158e6f4cb8eac3c0075918b7ffi5k8o	0	3	3	2	3	1
5	01067975436d123f717ee5aba0dd4bbfa0937	0	165	17	72	398	4
6	012441119bcf83b23d4768bb72cea6d6carua	0	16	19	10	19	3
7	01255c2c7c5578c186873422fc00fd7afwk8k	0	34	13	6	55	29
8	01349288df20199905e719f0ff7ee771nwryj	0	28	31	4	49	23
9	0141844fc725a85e38e12cc02fcdbbe8ofqvx	0	13	14	6	16	2
10	0181b91f58e6cbaa832235e2879c9c00ovrzi	0	7	9	3	14	9



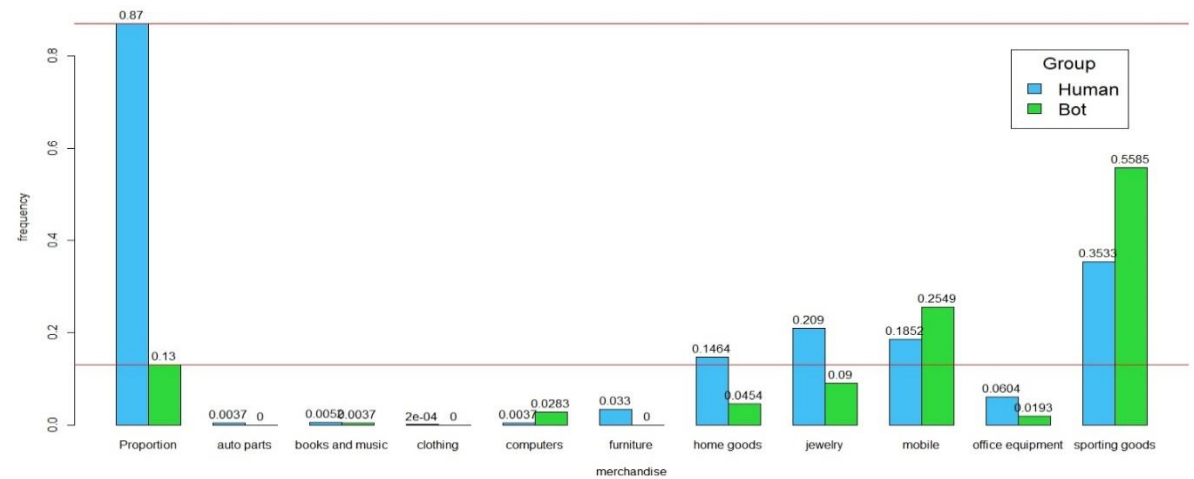
# 다각적 EDA



유동IP를 사용한 우회 입찰자



인간과 로봇의 물품별 구매빈도의 분포



04

결론

## 01. 결론



로봇은 어떤 특징들을 가지고 있을까

## 01. 결론

일반 및 핵심 입찰자들이

- 총 입찰건수가 상대적으로 많은 사람
- 품선별 입찰건수의 평균이 높은 사람

- Url별 입찰건수의 평균이 높은 사람

- 고객별 distinct한 auction이 많은 사람

- 고객별 distinct한 url이 많은 사람

- 고객별 distinct한 ip가 많은 사람

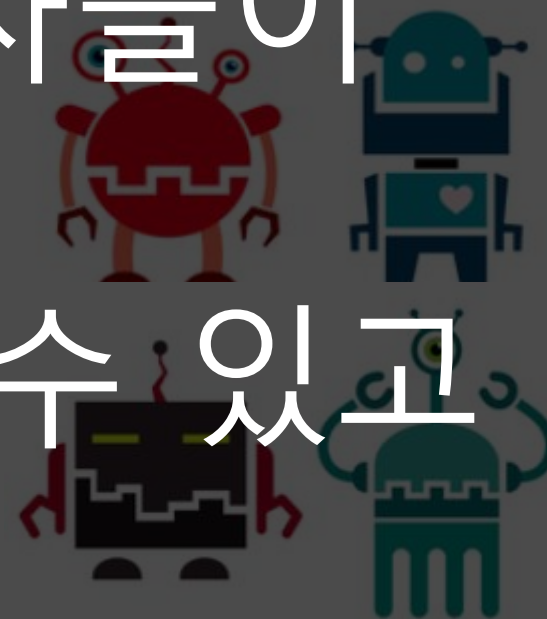
- 고객별 distinct한 country가 많은 사람

- 고객별 distinct한 devices가 많은 사람

- 입찰과 입찰사이의 시간이 짧은 사람

더욱 기회를 얻을 수 있고

고객들의 이탈을 막을 수 있다.





# Thank you

감사합니다

