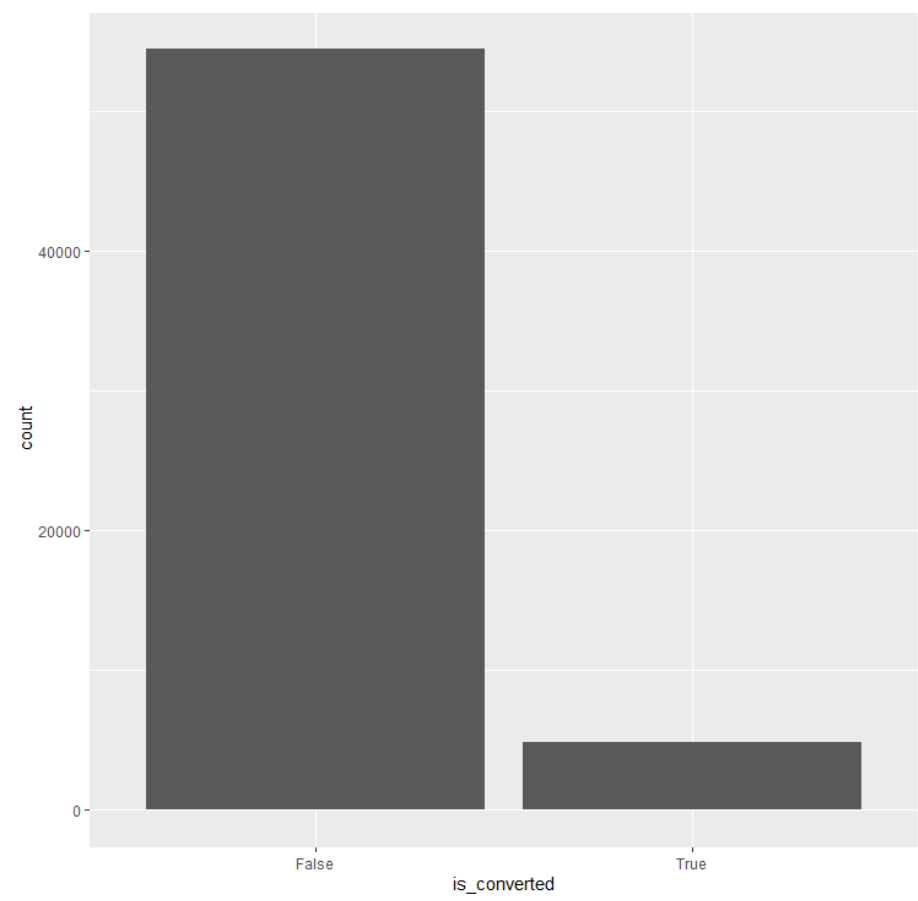


is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

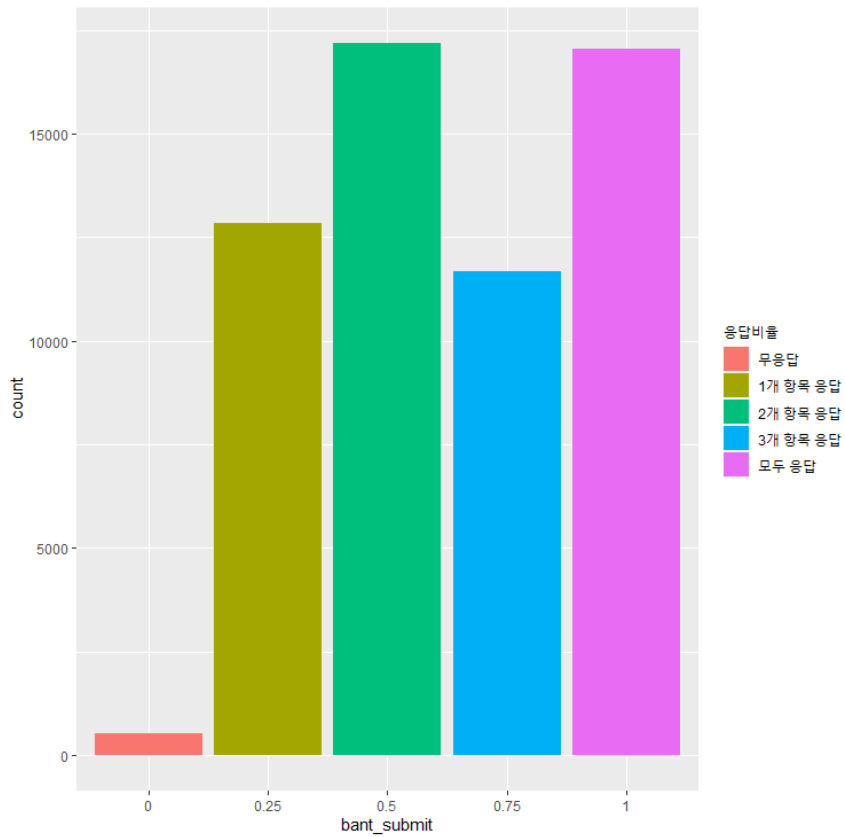
is\_converted (target)



- Target 변수, 영업 전환이 되었는지(True) 되지 못하였는지(False)
- Target 변수의 분포가 심하게 왜곡되어 있다.
- Oversampling 혹은 undersampling 등의 방법을 활용하여 불균형을 해결 해 줄 필요가 있다.
- True의 비율이 매우 작기에 영업전환 자체의 가능성이 낮다 생각된다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## bant\_submit



Cramer V  
0.07042

- LG 홈페이지 가입 시 응답한 질문의 수를 의미하는 변수.
- 응답의 경우의 수는 '무응답', '1개 응답', '2개 응답', '3개 응답', '모두 응답' 으로 총 5 개의 범주를 갖는다.
- 비율을 값으로 갖지만 범주가 존재하기에 category 데이터로 다룰 수 있다.
- 응답한 항목이 많을수록 더 적극적인 고객으로 볼 수 있으며 영업전환가능성이 높을 수 있다.
- Cramer V로 상관관계를 분석하였을 때 유의미하지는 않다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

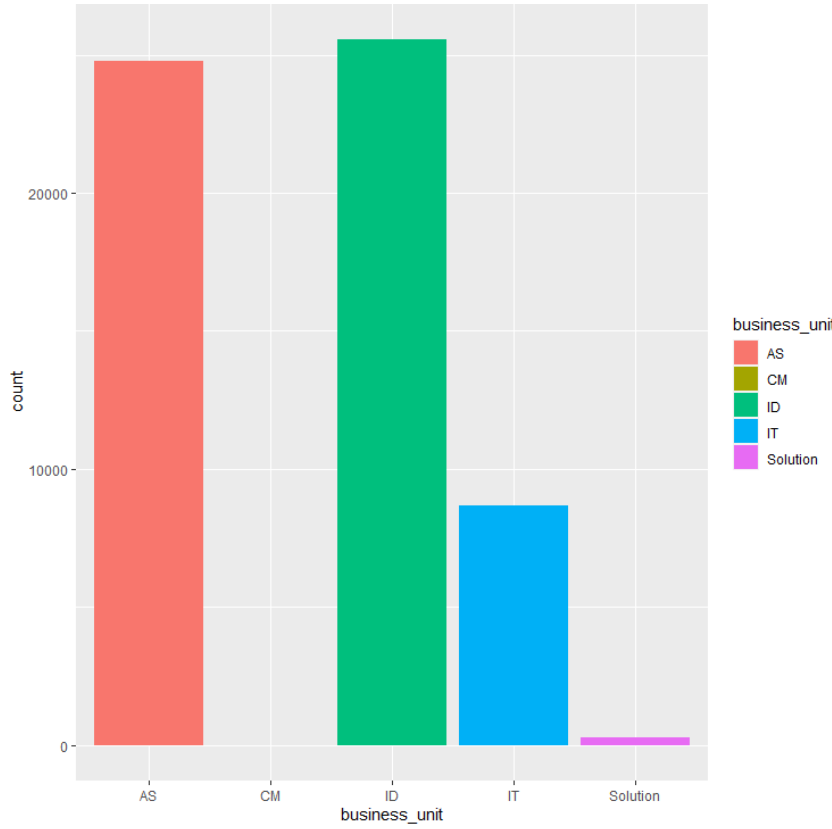
## customer\_country

//India	/São Paulo/Brazil		//United States	//United Kingdom
3037	1344	982	962	805
//Saudi Arabia	//Italy	//Philippines	//U.A.E	//
718	707	705	676	669
/Mumbai/India	//Germany	//Brazil	/Hyderabad/India	/Bangalore/India
594	578	577	508	480
Mexico	//Mexico	/Lima/Peru	/Pune/India	//Colombia
453	441	435	430	403
/Dubai/U.A.E	//Spain	/Delhi/India	/Santiago/Chile	//Hong Kong
397	394	386	386	353
/Chennai/India	//Egypt	/Ahmedabad/India	/Rio de Janeiro/Brazil	/Kolkata/India
288	244	230	214	209
/BOGOTÁ/Colombia	/Patna/India	//France	/Lucknow/India	/Riyadh/Saudi Arabia
194	182	178	178	177
/Surat/India	//Chile	/New Delhi/India	/Indore/India	//Argentina
170	168	162	160	153
//Canada	//Ecuador	/SP/Brazil	//Hungary	//Poland
152	150	149	141	139
/Nagpur/India	//Peru	/Jaipur/India	/CDMX/Mexico	/Bogotá/Colombia
137	135	135	134	132
/São Paulo /Brazil	/India	/Coimbatore/India	/Noida/India	Spain
119	113	110	110	110
United Kingdom	//Singapore	/Bengaluru/India	/Belo Horizonte/Brazil	/sp/Brazil
104	103	103	99	99
/santiago/Chile	/Bogota/Colombia	/Panama/Panama	/Curitiba/Brazil	/Varanasi/India
96	95	95	93	91
/Cairo/Egypt	/Monterrey/Mexico	/São paulo/Brazil	/Aurangabad/India	/Vadodara/India
90	89	89	86	85
/Mexico/Mexico	/Bhubaneswar/India	//Malaysia	/Bangalore /India	/Nashik/India
84	82	81	81	79
/Manila/Philippines	/Santiago /Chile	/Quezon City/Philippines	/Singapore/Singapore	/Gurgaon/India
78	78	77	77	76
/LIMA/Peru	//Australia	/Ghaziabad/India	/Salvador/Brazil	//Thailand
76	75	72	72	71
/Ranchi/India	/Hyderabad /India	/delhi/India	/Makati/Philippines	/mumbai/India
71	70	69	69	67
/Recife/Brazil	/Gwalior/India	/SÃO PAULO/Brazil	//Indonesia	/London/United Kingdom
67	64	63	61	61
/pune/India	/Sao Paulo/Brazil	/Jeddah/Saudi Arabia	/Abu Dhabi/U.A.E	(other)
60	60	59	58	33531

- 고객들의 국가 데이터
- 일반적으로 '구체적인 지역 / 국가명'으로 구성되어 있다.
- 국가 단위로 데이터를 추출할 것인지, 필요에 따라 대도시까지는 포함할 것인지 고려할 필요가 있다.
- 우선적으로 해야할 일은 '/'을 통해 텍스트를 분리하는 것.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

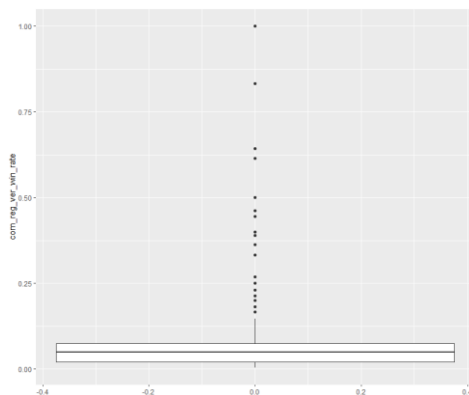
## business\_unit



- 고객이 요청한 업무 유형이다. 자세한 설명은 알 수 없지만 Solution의 빈도가 현저하게 낮다.
- Target 변수의 True비율도 현저하게 낮기 때문에 solution과의 연관성을 살펴보는 것 또한 의미가 있을 수 있다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## com\_reg\_ver\_win\_rate



- business\_area (vertical) – business unit – region 으로 분류된 기준 별 영업 전환율이 다.
- 분류가 되어 있기에 범주로 처리할 수 있다.
- 일반적으로 영업 전환율은 낮다.
- vertical – business unit – region

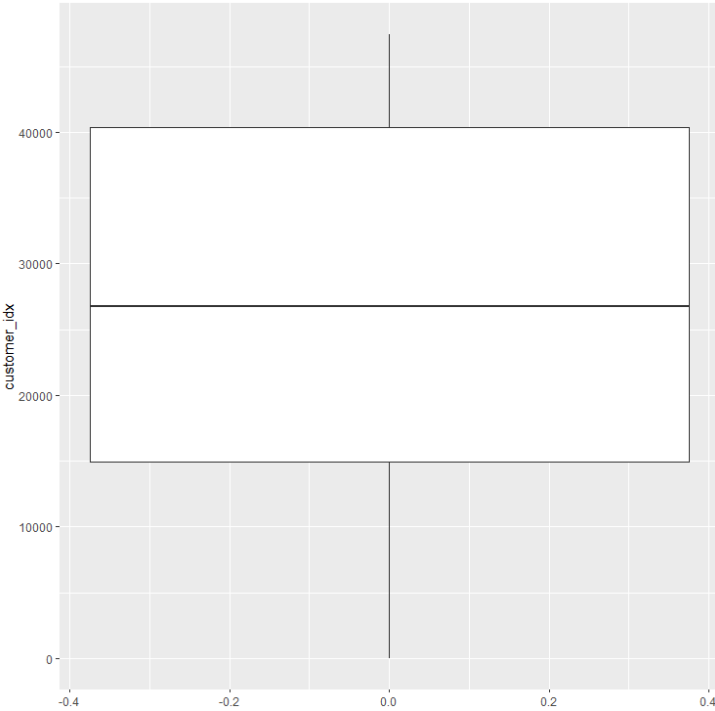
0.0037878787878787	0.0039370078740157	0.004	0.0109890109890109	0.0118577075098814	0.0135135135135135
713	803	434	104	423	242
0.0151515151515151	0.0169491525423728	0.0172413793103448	0.0175438596491228	0.0181818181818181	0.0196078431372549
118	221	63	96	65	107
0.0199004975124378	0.0202020202020202	0.0227272727272727	0.025	0.0289256198347107	0.0289855072463768
316	110	58	54	253	116
0.0311958405545927	0.032258064516129	0.0327868852459016	0.0330578512396694	0.036036036036036	0.037037037037037
609	106	422	128	119	75
0.04	0.0408163265306122	0.0416666666666666	0.0422535211267605	0.043103448275862	0.0434782608695652
276	155	89	100	153	54
0.0446428571428571	0.0476190476190476	0.0485436893203883	0.0491803278688524	0.0496894409937888	0.0531914893617021
292	141	109	214	172	99
0.0538922155688622	0.0544217687074829	0.0555555555555555	0.0575342465753424	0.0666666666666666	0.0677966101694915
193	175	22	399	216	240
0.0681818181818181	0.0695652173913043	0.0714285714285714	0.0732484076433121	0.0749486652977412	0.075
103	251	82	791	1130	274
0.0806916426512968	0.0833333333333333	0.0843373493975903	0.0869565217391304	0.0888888888888888	0.105263157894737
410	17	166	62	120	31
0.113636363636364	0.116279069767442	0.118421052631579	0.11864406779661	0.124121779859485	0.125
57	71	141	118	431	13
0.136363636363636	0.147058823529412	0.166666666666667	0.181818181818182	0.2	0.214285714285714
60	46	17	33	5	60
0.230769230769231	0.25	0.269230769230769	0.333333333333333	0.363636363636364	0.390243902439024
40	16	44	110	13	95
0.4	0.444444444444444	0.461538461538462	0.5	0.615384615384615	0.642857142857143
12	12	16	34	24	782
0.833333333333333	1	NA's			
17	10	44731			

com\_reg\_ver\_win\_rate은  
vertical level 1(사업 분야)  
+ business\_unit (요청 상품 대응 사업 부서)  
+ region (사업 지방(국가-지방))  
을 이용한 비율.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

customer\_idx

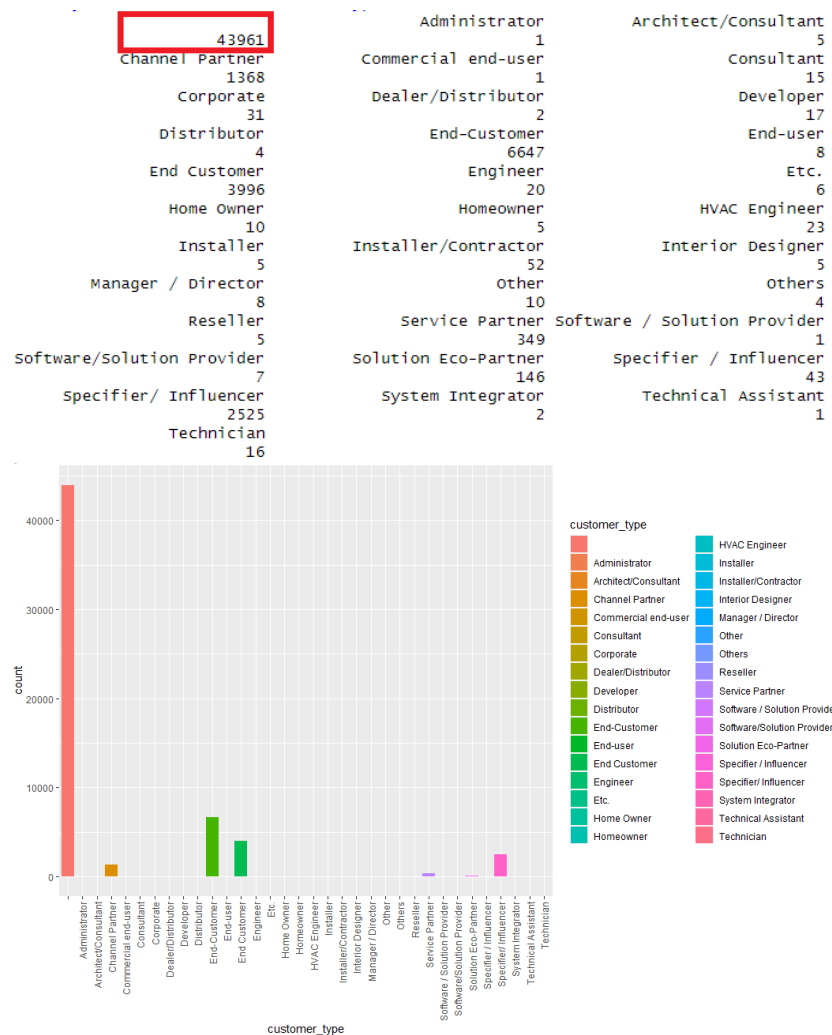
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2	14913	26774	27115	40369	47466



- 고객들을 식별하기 위한 id이다. 전체 sample 수와 다른 것으로 보아 일부 중복이 있다.

is\_converted  
 bant\_submit  
 customer\_country  
 business\_unit  
 com\_reg\_ver\_win\_rate  
 customer\_idx  
 customer\_type  
 enterprise  
 historical\_existing\_cnt  
 id\_strategic\_ver  
 ic\_strategic\_ver  
 idit\_strategic\_ver  
 customer\_job  
 lead\_desc\_length  
 inquiry\_type  
 product\_category  
 product\_subcategory  
 product\_modelname  
 customer\_counry.1  
 customer\_position  
 expected\_timeline  
 ver\_cus  
 ver\_pro  
 ver\_win\_rate\_x  
 ver\_win\_ratio\_per\_bu  
 business\_area  
 business\_subarea  
 lead\_owner

## customer\_type

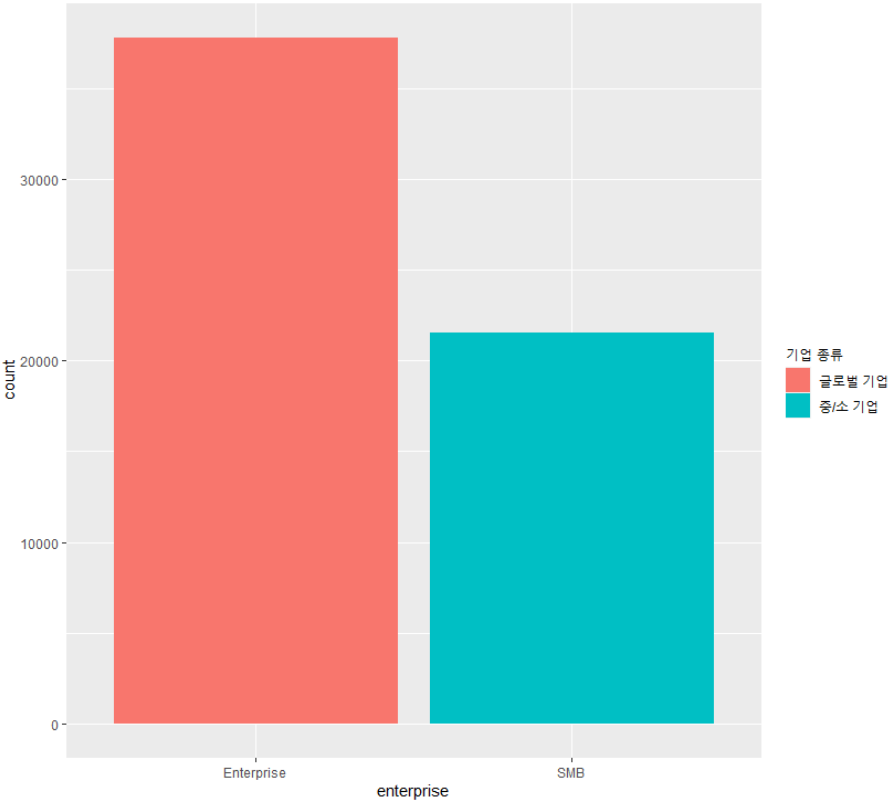


- 고객 유형을 의미한다.
- 공백이 43961개로 너무 크다.
- Unique한 범주가 너무 많다.
- 전체적으로 학습에 사용하기에는 무리가 있어 보여 전처리에 주의를 기울여야 할 것이라 생각한다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## enterprise

```
Enterprise      SMB  
      37783      21516  
  
> phi(enterprise_phi)  
[1] 0.12
```



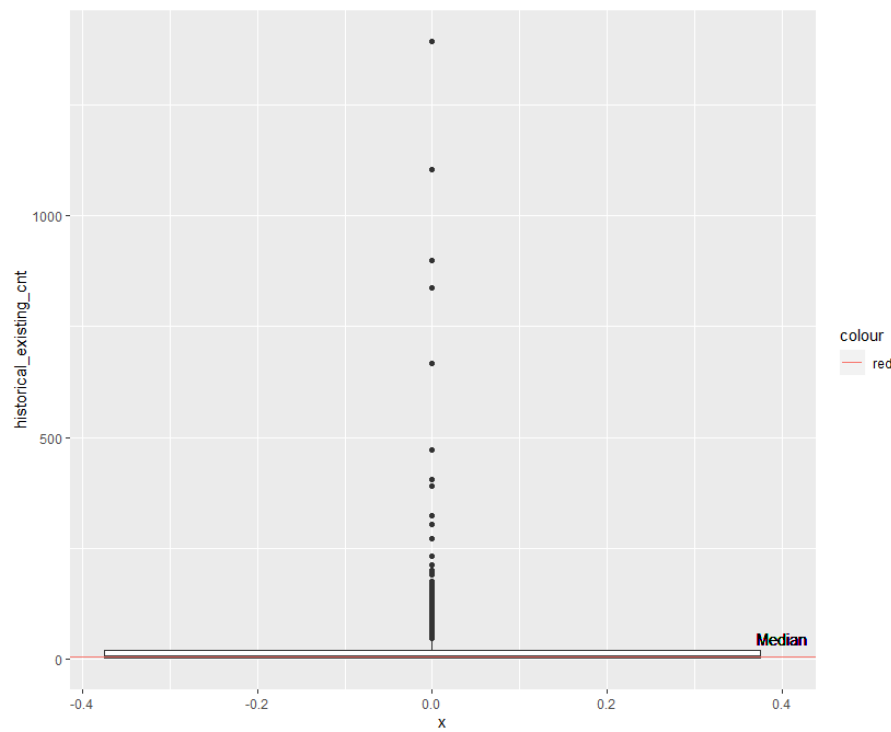
- 고객의 회사가 글로벌 기업(Enterprise)인지 혹은 중·소규모 기업(SMB)인지를 의미한다.
- 결측치가 없다는 특징이 있다.
- 글로벌 기업이 중/소규모 기업보다 더 많다는 정보를 얻을 수 있다.
- Target 변수와의 개별 상관계수는 0.12로 매우 낮은 수준.



is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
ic\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## historical\_existing\_cnt

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
0.00	1.00	4.00	19.91	19.00	1394.00	45543



- 이전에 영업 전환된 횟수를 의미한다.
- 1000건이 넘는 경우는 1000번이나 영업전환이 되었다는 것인데 noise일 가능성을 염두해야 한다.
- Median은 4로 분포를 통해 조심히 접근해볼 필요가 있다.

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## id\_strategic\_ver / it\_strategic\_ver / idit\_strategic\_ver

id_strategic_ver	1 NA's
	3444 55855
it_strategic_ver	1 NA's
	1121 58178
idit_strategic_ver	1 NA's
	4565 54734

- 후에 나오는 ver\_cus 가중치 칼럼을 확인하면 1과 0으로 구성된 것을 알 수 있다.
- 아마 동일한 방법으로 가중치를 부여 했다면 Na 값은 0이 아닐까 생각

is\_converted  
 bant\_submit  
 customer\_country  
 business\_unit  
 com\_reg\_ver\_win\_rate  
 customer\_idx  
 customer\_type  
 enterprise  
 historical\_existing\_cnt  
 id\_strategic\_ver  
 it\_strategic\_ver  
 idit\_strategic\_ver  
 customer\_job  
 lead\_desc\_length  
 inquiry\_type  
 product\_category  
 product\_subcategory  
 product\_modelname  
 customer\_counry.1  
 customer\_position  
 expected\_timeline  
 ver\_cus  
 ver\_pro  
 ver\_win\_rate\_x  
 ver\_win\_ratio\_per\_bu  
 business\_area  
 business\_subarea  
 lead\_owner

18733  
 other  
 4852  
 education  
 2287  
 operations  
 1961  
 information technology  
 1937  
 consulting  
 1349  
 entrepreneurship  
 1192  
 program and project management  
 948  
 media and communication  
 709  
 product management  
 573  
 accounting  
 396  
 graphic/color art  
 331  
 film production  
 241  
 clinical specialist  
 227

## customer\_job

engineering  
 6342  
 administrative  
 3359  
 sales  
 2218  
 purchasing  
 1938  
 business development  
 1801  
 arts and design  
 1276  
 marketing  
 1161  
 others  
 763  
 healthcare services  
 632  
 finance  
 463  
 support  
 358  
 real estate  
 295  
 3d/vfx art  
 235  
 radiology professional  
 207

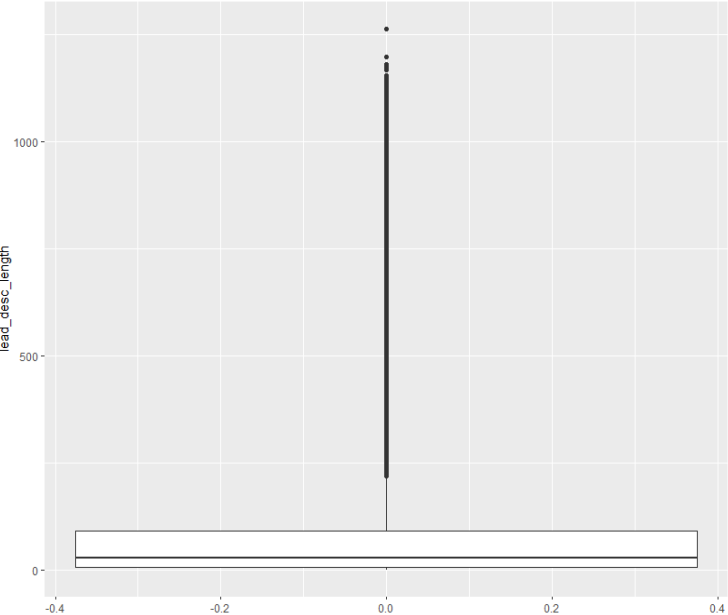
- 고객들의 직업을 나탄낸다.
- 공백값이 어느 정도 있다.
- 'other'와 'others' 등 모호한 범주가 존재  
이를 별도로 매핑하는 등의 처리가 필요하다.
- 세부 분류된 업종이 있는 것으로 보이며  
이 또한 매핑이 필요할 수 있음을 주의해  
야 한다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## lead\_desc\_length

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.00	7.00	29.00	79.27	92.00	1264.00

Cramer V  
0.2491



- 고객들의 리뷰와 같은 평가인데 텍스트 데이터는 주어지지 않았다.
  - 텍스트의 길이가 길다면 적극적인 고객으로 해석하고 영업전환 가능성이 높을 수 있다고 추론할 수 있다.
- 따라서 문장의 길이와의 관계를 살펴볼 필요가 있다.
- Cremer's V 상관계수는 0.24이다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## inquiry\_type

```
[1] ""  
[2] "(select ID_Needs)"  
[3] "AIO"  
[4] "Customer Suggestions"  
[5] "Digital platform"  
[6] "display product"  
[7] "Display Textbook and photos"  
[8] "EDUCATIONAL EQUIPMENTS"  
[9] "estoy buscando para Ecuador este producto LG MAGNIT micro LED, para un cliente de 138 pulgadas, con envío marítimo."  
[10] "Etc."  
[11] "ETC."  
[12] "Event Inquiry"  
[13] "Evento_SdeleStero"  
[14] "first Info and pricing"  
[15] "for school"  
[16] "High inch 86 / 98 or 110"  
[17] "Hola me pueden cotizar 19 pantallas interactivas de 100 pulgadas entregadas en Guayaquil -Ecuador."  
[18] "Hospital TV"  
[19] "Hotel TV products"  
[20] "i want to know the details about it"  
[21] "IDB"  
[22] "Intégrateur historique du George V"  
[23] "LED Signage"  
[24] "Media Inquiry"  
[25] "Needs"  
[26] "Not specified"  
[27] "OEM/ODM Request"  
[28] "One Quick:Flex"  
[29] "other"  
[30] "other"  
[31] "other_"  
[32] "others"  
[33] others  
[34] "Pantallas Interactivas para Clinicas"  
[35] "Preciso de um monitor médico para radiografia convencional e tomografia."  
[36] "Probeam precio"  
[37] "Product Information"  
[38] "Purchase"  
[39] "Purchase or Quotation"  
[40] "Quotation or purchase consultation"  
[41] "Quotation or Purchase consultation"  
[42] "Quotation or Purchase Consultation"  
[43] "quotation_"  
[44] "quotation_or_purchase_consultation"  
[45] "Request a Demo"  
[46] "Request for Distributorship"  
[47] "Request for Partnership"  
[48] "Request for quotation or purchase"  
[49] "Request for technical consulting"
```

- 고객들의 문의 유형일 것이라 생각된다.
- 대·소문자와 공백 등으로 의미 없는 범주가 많다. 이를 처리할 방법을 생각해 볼 필요가 있다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## product category

```
[235] "services"  
[236] "signage"  
[237] "signage care solution"  
[238] "signage care solutions"  
[239] "single-split"  
[240] "single cac"  
[241] "single package"  
[242] "single split"  
[243] "smart tv"  
[244] "smart tv signage"  
[245] "software solution"  
[246] "soğutucu"  
[247] "solar"  
[248] "solar aircare"  
[249] "solar built-in/cooking"  
[250] "solar chiller"  
[251] "solar dryer"  
[252] "solar energy storage system"  
[253] "solar ess"  
[254] "solar ess,ems"  
[255] "solar ess,system ac"  
[256] "solar monior/monitor tv"  
[257] "solar monior/monitor tv,pc,tv,refrigerator,washing machine,dryer,home beauty"  
[258] "solar projector"  
[259] "solar refrigerator"  
[260] "solar robot"  
[261] "solar system ac"  
[262] "solar system ac,aircare"  
[263] "solar system ac,water care"  
[264] "solar tv"  
[265] "solar vacuum cleaner"  
[266] "solar water care"
```

- 문의 제품의 유형이다.
- 이 경우에도 많은 범주가 존재하는데 'solar'와 같은 부분을 보면 제품별로 공통 분모가 존재할 수 있음을 추론할 수 있다.

경우에 따라서는 그룹화도 가능할 것이라 생각한다.

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
er_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

product\_subcategory

[233]	"TV Monitors"	"TV Signage"
[235]	"UH5E Series"	"UH5F-H Series"
[237]	"UH5F Series"	"UH7F-H Series"
[239]	"UH7F Series"	"UH7F 系列"
[241]	"UHD 4K Monitors"	"UHD IR-type Touch CreateBoard"
[243]	"UHD Large Screen Signage Display"	"UHD Signage"
[245]	"UHD TV Signage"	"UL3G Series"
[247]	"UL3J"	"UL3J Series"
[249]	"Ultra-Light Series"	"Ultra High Brightness Series"
[251]	"Ultra Light"	"Ultra Light Series"
[253]	"Ultra Narrow Bezel Video wall"	"Ultra Slim"
[255]	"Ultra Slim Series"	"Ultra Stretch Series"
[257]	"Ultra Stretch Signage"	"UltraFine Display OLED Pro"
[259]	"UltraFine Monitors"	"UltraGear Monitors"
[261]	"Ultrawide Monitors"	"Ultrawide™ Monitors"
[263]	"UM3DG-H Series"	"UM3DG Series"
[265]	"UM3E Series"	"UM3F Series"
[267]	"UM5J Series"	"UQ801"
[269]	"UR640S Series"	"US342H Series"
[271]	"US660H Series"	"US662H Series"
[273]	"US665H Series"	"US670H Series"
[275]	"US760H Series"	"US761H Series"
[277]	"US762H Series"	"US765H Series"
[279]	"US770H Series"	"US772M Series"
[281]	"UT340H Series"	"UT567H Series"
[283]	"UT570H Series"	"UT640S"
[285]	"UT640S Series"	"UT660H Series"
[287]	"UT662M Series"	"UT665H Series"
[289]	"UT670H Series"	"UT672M Series"
[291]	"UT672M Series"	"UT770H Series"
[293]	"UT781H"	"UT781H Series"
[295]	"UT781H 系列"	"UT782H Series"
[297]	"Versatile"	"VH7B Series"
[299]	"VH7E-H Series"	"VH7E Series"
[301]	"Video"	"Video-wall OLED Signage"
[303]	"Video wall"	"Video wall OLED"
[305]	"Virtual Production Series"	"VL5D Series"
[307]	"VL5D 系列"	"VL5F Series"

- 제품별 subcategory 데이터이다.
- 331 개의 범주를 갖는다
- product category와 유사한 상황이며 따라서 동일하게 처리 방법을 생각해 보아야 한다.
- 전체 59299 데이터 중 331개의 범주는 괜찮은가?

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## product\_modelname

```
[626] "LTAK140"
[627] "LTAK140-GV"
[628] "LTAK140-GW"
[629] "LWBC026"
[630] "LWBC029"
[631] "LWBC039"
[632] "Model Name"
[633] "One:Quick"
[634] "Only Installation Need"
[635] "OPSJ-5LDJA"
[636] "Passed on to Fixxy distribution"
[637] "Required After 3 Months"
[638] "SC-00DA"
[639] "Solution"
[640] "ST-43HF"
[641] "STB-5500 (ASIA)"
[642] "STB-5500 (EU/CIS)"
[643] "STB-5500 (NA)"
[644] "STB-6500 (ASIA)"
[645] "STB-6500 (EU)"
[646] "STB-6500 (MEA)"
[647] "STB-6500 (NA)"
[648] "Supersign CMS"
[649] "Supersign Media Editor"
[650] "Supersign WB"
[651] "Surgical"
[652] "Surgical Monitors"
[653] "T-OLED"
[654] "This is being dealt with by LG Germany."
[655] "Total Care Thru one-stop Service"
[656] "UltraFine Ergo(32UN880)"
[657] "Ultrawide Ergo(34WN780)"
[658] "UR Series"
[659] "UR340C9UD"
[660] "VerticalBrochure"
[661] "video"
[662] "want Split AC"
[663] "whitePaper"
[664] "WP400"
[665] "WP401"
[666] "WP402"
```

- 666 개의 model name
- product\_category - product\_subcategory  
– product\_modelname 세 개의 칼럼은 관계가 있을 것이라는 생각이 든다. 특정한 처리가 필요할수도?



is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
er_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## customer\_country.1

```
[941] "/Aizawl/India"
[942] "/Ajah - Lagos/Nigeria"
[943] "/Ajdovščina/Slovenia"
[944] "/ajman/U.A.E"
[945] "/Ajman/U.A.E"
[946] "/Ajmer district/India"
[947] "/Ajmer road nisangan/India"
[948] "/Ajmer/"
[949] "/ajmer/India"
[950] "/Ajmer/India"
[951] "/AJMER/India"
[952] "/ajna/India"
[953] "/Akbarpur/India"
[954] "/Aklan/Malay/Philippines"
[955] "/Aklan/Philippines"
[956] "/AKLAN/Philippines"
[957] "/Akluj/India"
[958] "/Akola/India"
[959] "/Ako/India"
[960] "/Akot/India"
[961] "/Akwa Ibom /Nigeria"
[962] "/Al-Ain/U.A.E"
[963] "/Al-Kharj/Saudi Arabia"
[964] "/al-Riyad/Saudi Arabia"
[965] "/AL / maceio/Brazil"
[966] "/Al ain/U.A.E"
[967] "/Al Ain/U.A.E"
[968] "/Al Ain/U.A.E"
[969] "/AL Ain/U.A.E"
[970] "/Al Barsha 1/U.A.E"
[971] "/Al Fujairah/U.A.E"
[972] "/Al Gudaibiys/Bahrain"
[973] "/Al Jubail/Saudi Arabia"
[974] "/Al Kharj/Saudi Arabia"
[975] "/Al Khawaneej/U.A.E"
[976] "/Al Khobar - Saudi Arabia/Saudi Arabia"
[977] "/Al Khobar/Saudi Arabia"
[978] "/AL KHOBAR/Saudi Arabia"
[979] "/Al Madina/Saudi Arabia"
[980] "/Al Madinah /Saudi Arabia"
[981] "/Al Medhatiya/U.A.E"
[982] "/al minya/Egypt"
[983] "/Al nadha /U.A.E"
[984] "/Al Najaf/Mushkab city/Iraq"
[985] "/Al Nakheel, Ras Al Khaimah/U.A.E"
[986] "/Al Qouz 4/U.A.E"
```

- customer\_country가 고객의 국적을 의미한다.
- 이번 주제는 B2B 고객을 대상으로하기엔 customer\_country보다 customer\_counry.1만을 이용하는 게 더 괜찮지 않을까 가능성.
- 문제는 데이터 자체에 오타와 누락된 값으로 인해 불필요하게 범주의 수가 증가하였다. 이부분을 처리해 주어야 할 듯하다.
- Customer\_countr와 둘 중하나만 사용하는 방향으로 진행하는 것이 좋을 듯하다.

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## customer\_position

	none	Cramer V
	19680	0.2491
	manager	
	8217	
	ceo/founder	
	7984	
	other	
	6499	
	director	
	4843	
	associate/analyst	
	2597	
	partner	
	2536	
	entry level	
	2216	
	others	
	964	
	c-level executive	
	851	
	trainee	
	849	
	vice president	
	509	
	intern	
	446	
	end-user	
	424	
	hospital	
	201	
	installer	
	60	
	government	
	29	
	cargo	
	27	
	software /solution provider	
	25	
	-	

- 고객의 직책을 의미한다.
- 총 117개의 범주가 존재한다.
- 직책(position)은 B2B 거래에 있어서 의사 결정권과 관련하여 중요한 정보를 제공한다.
- 결측값이 상당히 많다.
- 아무런 처리를 하지 않는 경우 Cramer's V 는 0.24이다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

expected\_timeline

30863

less than 3 months

17250

3 months ~ 6 months

5026

more than a year

3023

9 months ~ 1 year

1101

6 months ~ 9 months

1098

less than 6 months

108

etc.

95

less\_than\_3\_months

76

being followed up

66

being followed up.

24

idb are mailed to client.

21

the client is not having any requirement hence closig in system. although the details of

- 예상되는 서비스 소요 시간이다.
- 공백값 30863개 이다.
- 오타로 인한 무의미한 범주가 존재한다.
- 공백 값과 오타로 인한 문제를 우선적으로 해결할 필요가 있다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

## ver\_cus / ver\_pro

```
> summary(train_df_feature$ver_cus)
  0      1 
56832 2467 

> summary(train_df_feature$ver_pro)
  0      1 
56286 3013
```

- 가중치를 의미하는 feature로 이번에는 결측치가 아니라 0으로 되어 있다.
- 결측치가 아닌 0으로 된 것을 기반으로 이전의 id, it, idit 칼럼 또한 결측치를 0으로 처리할 수 있을 가능성이 있다.

is\_converted  
bant\_submit  
customer\_country  
business\_unit  
com\_reg\_ver\_win\_rate  
customer\_idx  
customer\_type  
enterprise  
historical\_existing\_cnt  
id\_strategic\_ver  
it\_strategic\_ver  
idit\_strategic\_ver  
customer\_job  
lead\_desc\_length  
inquiry\_type  
product\_category  
product\_subcategory  
product\_modelname  
customer\_counry.1  
customer\_position  
expected\_timeline  
ver\_cus  
ver\_pro  
ver\_win\_rate\_x  
ver\_win\_ratio\_per\_bu  
business\_area  
business\_subarea  
lead\_owner

ver\_win\_rate\_x

$$\frac{\text{vertical 수 비율} * \text{vertical 별 Lead 수 대비 영업 전환 성공 비율}}{100}$$

① vertical 수 비율 =  $\frac{\text{신규 고객}}{\text{기존 고객}}$

② vertical 별 Lead 수 대비 영업 전환 성공 비율

=  $\frac{\text{vertical 별 Lead 수}}{\text{전체 Lead 수}}$  의 영업 전환 성공 비율?

→ 위의 해석보다는 vertical (사업 분야) 별 Lead (고객군) 대비 영업 전환 성공 비율이 맞는듯?

→ 위의 가정을 전제하면 연속형 feature가 아니라 각 vertical에 속하는 lead 별로 동일한 영업 전환율을 갖을 것임.



실제로 factor로 type을 변경하고 확인하니 연속형 값이 아니라 범주화 되어 있음.

- vertical은 business area
- vertical별로 영업 전환율을 계산한다는 점은 LG가 business area 별로 데이터를 저장하고 업무를 진행하는 건가?
- vertical에는 특정 사업 영역이라는 의미도 있다? → bard가 잘 설명해 주었다.

2. 31593813372328e-06	1. 27659028834503e-05	2. 58895523078822e-05	6. 04403366605833e-05	9. 65915660650443e-05	0. 0002153634176709
113	306	519	1199	650	1035
0. 0002983104051378	0. 0005432224318428	0. 0005719551277132	0. 0007167734380046	0. 0011827288932506	0. 0030792876608617
1818	1929	1727	2013	3011	4097
NA's					
40882					

Vertical Marketing 이란? - <https://www.investopedia.com/terms/v/verticalmarket.asp>

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

ver\_win\_ratio\_per\_bu

0.0115830115830115	0.0201207243460764	0.022633744855967	0.0227272727272727	0.0268456375838926	0.0287769784172661
788	1365	713	173	1278	200
0.031578947368421	0.0344827586206896	0.0354838709677419	0.0369127516778523	0.048629531388152	0.0498402555910543
271	29	385	403	1356	2028
0.0514705882352941	0.0535714285714285	0.0609243697478991	0.0640703517587939	0.0645661157024793	0.071345029239766
210	216	540	1075	2460	984
0.0794117647058823	0.128571428571429	0.131147540983607	0.227272727272727	0.285714285714286	NA's
421	156	148	62	43	43995

→ 결국 vertical 에 속하는 business\_unit 별로 영업 전환율을 계산한 것이기에 연속값이 아닌 범주형 데이터일 수 있다고 가정함.

→ 결과적으로 범주화 가능.

- 이전 feature인 ver\_win\_rate\_x가 Lead(고객 군)별 vertical 영업 전환율이라면 현재 featur는 business\_unit 별 영업 전환율을 의미한다.

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## business\_area

	40882	corporate / office	4097	education	1727	factory	1035
government department	650	hospital & health care	1199	hotel & accommodation	2013	power plant / renewable energy	113
public facility	519	residential (home)	1818	retail	3011	special purpose	1929
transportation	306						

- 해당 칼럼이 바로 Vertical로 불리는 칼럼
- 공백값이 40882 개 존재.
- level이라는 단어를 사용한 것으로 보면 LG는 DB 내에서 Vertical을 가장 상위 계층으로 계층형 쿼리를 구축한 것? (데이터 마이닝에서는 큰 의미없이 데이터 구조가 어떤지 정도만 판단할 수 있을듯)
- 현재로서는 대충 사업 대상 영역 별로 segmentation 해둔 듯하다.
- 계층형 쿼리가 맞다면 가장 상위 계층의 level 이며 다른 칼럼들이 여기에 속해 있을 가능성이 있다.
- 공백값을 제외하면 빈도가 급격하게 낮은 범주는 없기에 양호한듯. 그래도 마이닝은 필요해 보임.

is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

## business\_subarea

- 세부 사업 영역으로 들어오면 공백값이 53773으로 늘어남 아마도 사업 영역은 있지만 세부 사업 영역은 없는 데이터가 있는 듯함.
- 이미 사업 영역이 나쁘지 않게 존재하는데 너무 많은 범주를 갖는 feature가 필요한가?



is_converted
bant_submit
customer_country
business_unit
com_reg_ver_win_rate
customer_idx
customer_type
enterprise
historical_existing_cnt
id_strategic_ver
it_strategic_ver
idit_strategic_ver
customer_job
lead_desc_length
inquiry_type
product_category
product_subcategory
product_modelname
customer_counry.1
customer_position
expected_timeline
ver_cus
ver_pro
ver_win_rate_x
ver_win_ratio_per_bu
business_area
business_subarea
lead_owner

lead\_owner

174	175	97	23	194	42	96	185	363	115	38	125	402	108	124
4627	4120	2499	2453	2208	1215	981	880	839	734	693	690	490	487	468
816	341	167	100	0	68	16	487	176	40	342	662	422	27	359
464	462	459	457	429	427	413	405	374	355	346	327	326	317	305
39	205	15	279	441	25	155	103	233	374	405	393	589	46	21
304	304	301	294	281	272	271	270	265	257	246	238	237	235	228
831	78	436	148	153	642	694	654	384	187	45	788	638	717	29
227	223	219	216	215	215	213	208	204	200	198	195	180	171	168
414	391	81	214	285	71	344	225	388	166	480	435	798	20	234
165	163	153	151	151	150	150	148	148	146	146	145	145	144	143
268	423	580	437	820	13	337	679	392	980	51	33	189	192	211
142	142	142	141	140	139	138	138	137	136	134	133	132	130	130
494	56	179	254	781	426	190	730	819	(other)					
130	127	126	126	126	121	120	119	117	17580					

Cramer V  
0.619

- 담당자 이름을 고유 숫자로 표현한 변수이다.
- 동일한 번호를 갖는다면 동일한 이름인 것 인지 생각해 볼 필요가 있다.
- 담당자의 이름이 영업전환과 관련이 있는지 생각해볼 필요가 있다.
- 이유는 모르겠으나 Cramer's V 상관계수가 약 0.62 높게 나타난다.