

크라우드 펀딩 프로젝트

성공요인 분석 및 예측 모델 개발



## 2-1 데이터 선정



### 대상

와디즈 - '리워드형' 프로젝트  
#가장 활성화된 유형 #와디즈 내 최다 데이터

### 기간

2018년 1월 1일~ 2019년 7월 말 기준 '종료된 프로젝트'

### 카테고리 선정

기간 내 종료된 프로젝트가 500개 이상인 카테고리 9개 선정 (전체15개)



테크·가전



패션·잡화



뷰티



푸드



홈리빙



디자인소품



여행·레저



스포츠·모빌리티



반려동물

## 2-2 데이터 수집



메이커	제목	판당시작날짜	판당마감날짜	판당시작일	판당시작요일	목표금액	카테고리	서포터수	종아요수	글업데이트수	댓글수	리워드종류수	배송시작날짜	마감배송차이	이미지수	비디오수	인스타팔로워수	스북팔로워수	sms팔로워수
건강한연구소	[3+1추가] 힐	#####	#####	2018-06	Wed	1000000	푸드	318	193	6	72	6	#####	33	16	0	1826	0	1826
카페 웰리스	홍역백신후	#####	#####	2018-06	Wed	1000000	푸드	46	24								0	47	47
Future Food Lab	미래에서 온	#####	#####	2018-06	Wed	1000000	푸드	559	168								228	235	463
농부대첩	[앵글] 처음	#####	#####	2018-06	Wed	1000000	푸드	543	232								1102	1368	2470
스튜디오와츠	자작나무로	#####	#####	2018-06	Wed	500000	디자인소	3	10	1	0	4	#####	50	20	0	103	35	138
홀로롤루(베네프	프리미엄	#####	#####	2018-06	Wed	500000	패션.잡화	70	72	7	12	2	#####	24	18	0	17	0	17
Run Run Train	기차타고 우	#####	#####	2018-06	Wed	1000000	패션.잡화	65	28	5	18	11	#####	42	46	1	16	33	49
코드퍼블릭	비오는 날	#####	#####	2018-06	Wed	3000000	패션.잡화	804	486	17	282	8	#####	93	80	5	8948	1054	10002
투빅커머스	품질이 곧	#####	#####	2018-06	Wed	1500000	패션.잡화	46	61	2	24	4	#####	41	22	0	0	0	0
원저노트	백화점 담	#####	#####	2018-06	Wed	1000000	패션.잡화	123	101	5	51	5	#####	49	15	0	149	0	149
하이로지	편안함을	#####	#####	2018-06	Wed	1000000	패션.잡화	57	56	2	21	5	#####	42	42	0	51	0	51
루모스 캔들	냄새 잡고	#####	#####	2018-06	Wed	2000000	홈리빙	78	86	18	31	4	#####	55	40	1	0	1206	1206
딩독x원터씨	반려견의 사	#####	#####	Jun-18	Wed	5000000	반려동물	132	71	10	33	1	#####	55	27	1	9820	0	9820
바이봉	즐거운 산책	#####	#####	2018-06	Wed	1000000	반려동물	6	10	0	7	9	#####	63	39	0	4040	0	4040
코코앤심바	코코와 심바	#####	#####	2018-06	Wed	1000000	반려동물	60	32	3	20	6	#####	23	32	3	427	2	429
모아컴퍼니 주식	[최초공개]	#####	#####	2018-06	Wed	1000000	테크.가전	729	400	9	150	7	#####	100	51	0	0	330	330
다오	휴대용 선풍	#####	#####	2018-06	Wed	500000	테크.가전	38	20	1	11	4	#####	37	26	2	0	0	0
이스트인디고	작은 가방	#####	#####	2018-06	Wed	1000000	여행.레저	549	343	13	121	8	#####	26	119	0	2058	29513	31571
소찬호	세상에서 가	#####	#####	2018-06	Wed	1000000	여행.레저	161	137	9	24	5	#####	45	34	0	0	0	0
강북희망협동조합	24시간의 기	#####	#####	2018-06	Tue	1500000	푸드	77	24	5	23	5	#####	50	30	1	0	323	323
설로인(Sir.Loin)	[수요미식회	#####	#####	2018-06	Tue	2000000	푸드	302	125	18	46	8	#####	113	39	2	632	1043	1675
Blackwhale	[남자 필독]	#####	#####	2018-06	Tue	1000000	패션.잡화	122	57	5	12	6	#####	31	28	0	0	655	655
그레이웨일	선스틱에 사	#####	#####	2018-06	Tue	1000000	패션.잡화	122	57	5	12	6	#####	39	32	1	44	101	145
컨퀘스트(CONQ)	색맹디자인	#####	#####	2018-06	Tue	1000000	디자인소	87	57	4	15	4	#####	33	46	1	0	0	0
COODGOODMOI	미니멀을	#####	#####	2018-06	Tue	500000	패션.잡화	56	45	1	7	12	#####	30	53	4	5	3	8
라스트미	특히 받은	#####	#####	2018-06	Tue	2000000	패션.잡화	40	68	7	24	7	#####	90	22	0	8101	10620	18721
(조)레이비글루브	지지 저가	#####	#####	2018-06	Tue	3000000	홈리빙	5	10	0	1	6	#####	47	13	0	33	0	33

최초 모집된 독립변수 총 32개

총 데이터 개수 6513개

+ SNS팔로워수 / 경제지표 변수 / '와디즈', '크라우드펀딩' 언급 기사 개수  
변수 추가

## 2-3 EDA(1/3)



### 종속변수 - 달성률

성공	78.26%
실패	21.74%

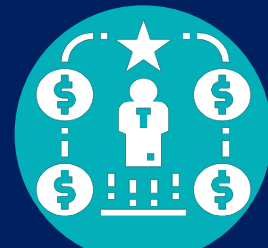
### <분류기준>



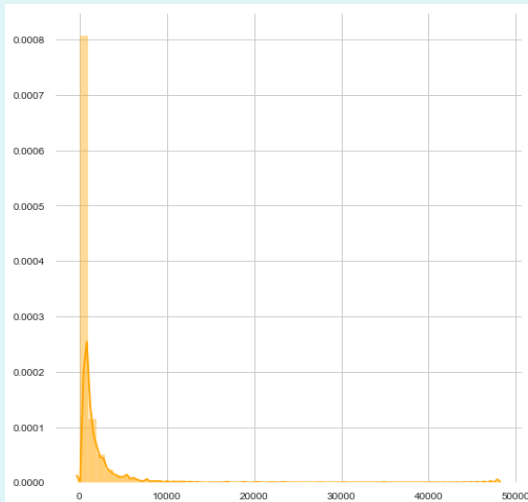
실패  
0 ~ 99%



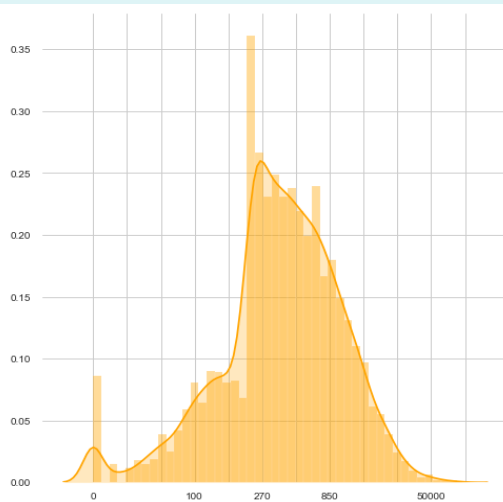
성공  
100% ~ 830%



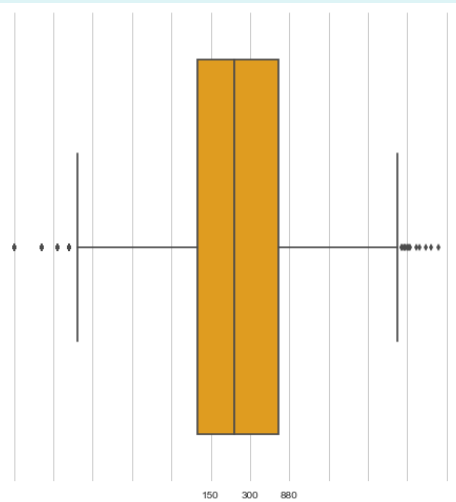
대성공  
831% ~



달성률

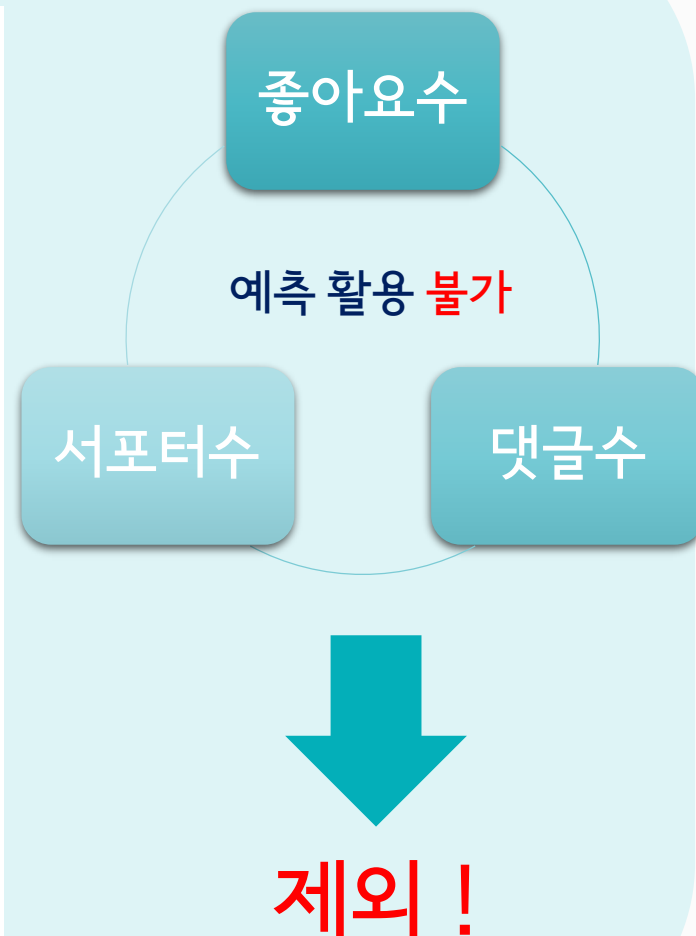
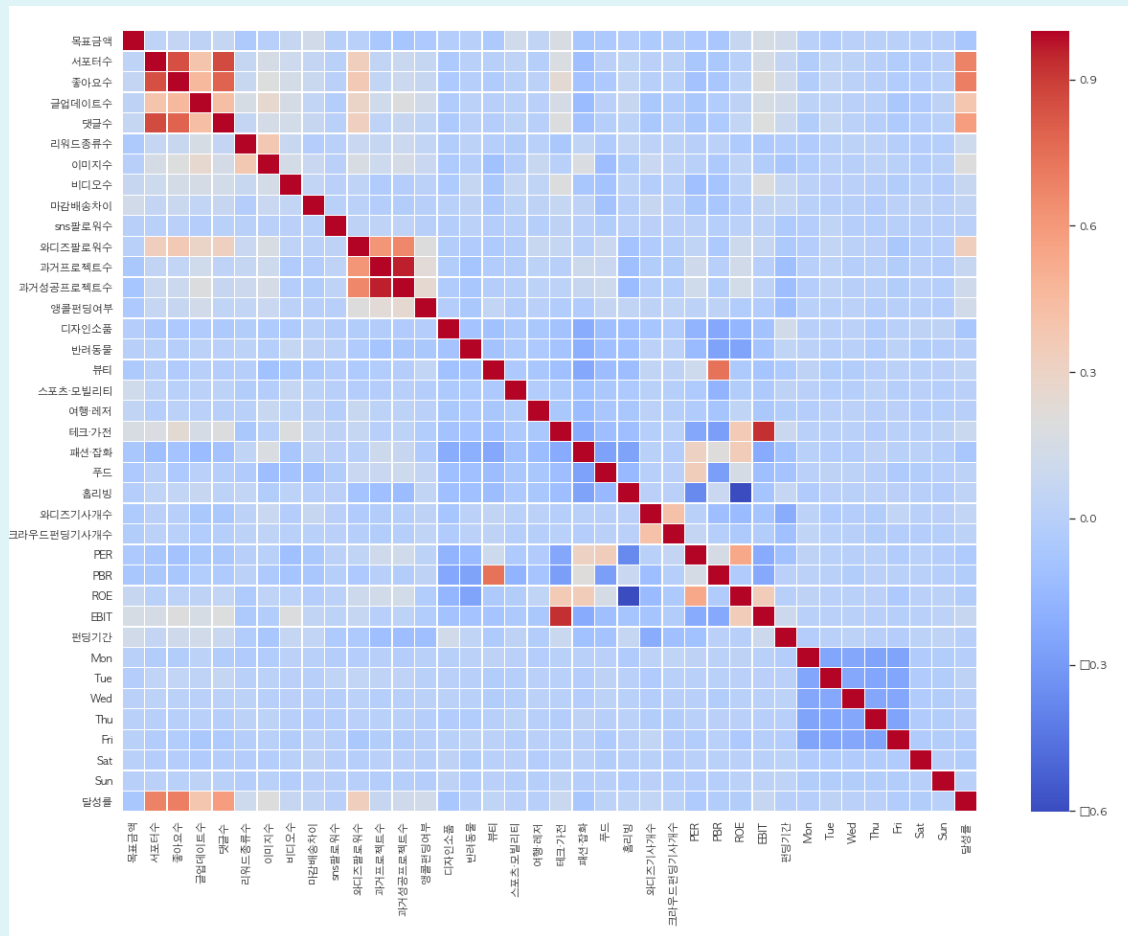


달성률  
Log-Transformation



달성률  
Boxplot

## 변수간 상관관계



## 2-5 모델링(1/8) – 1차 모델링



### 종속변수 - 달성률

- 0 실패 : 달성률 0~99
- 1 성공 : 달성률 100 ~ 830
- 2 대성공 : 달성률 830 이상



### 분류 모델링 – 성능 비교

	Decision Tree	SVM	Random Forest	XGBoost	Perceptron
Train Accuracy	93	70.4	94.2	86.8	82.2
Test Accuracy	69	68.7	77.7	80.2	79.8
Pre-processing	-	Kernel Trick	Scaling	-	Sigmoid/ Softmax/ Scaling

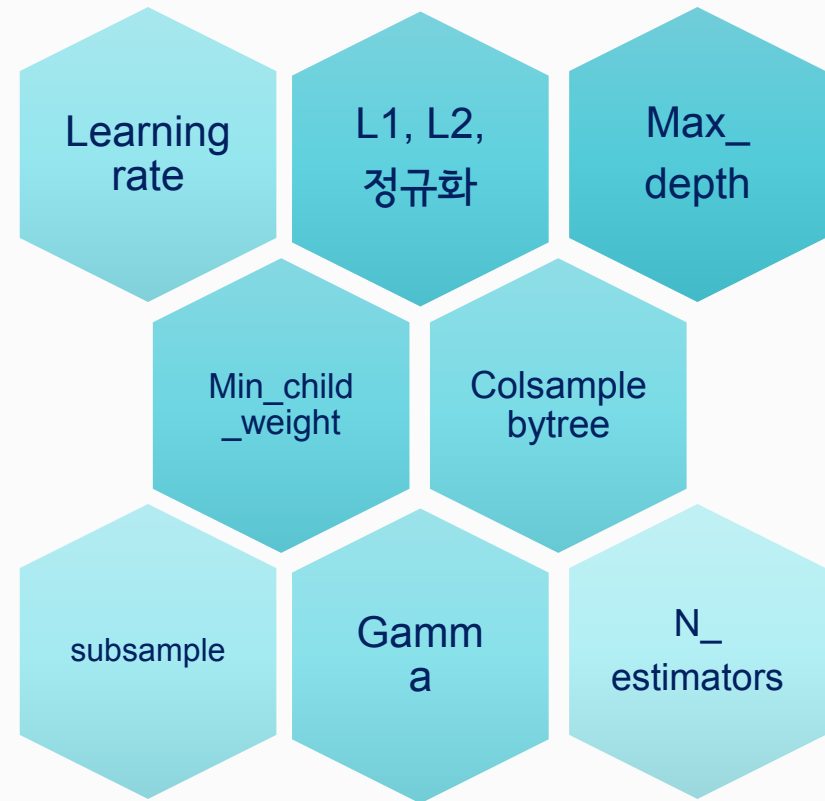
# 2-5 모델링(2/8) – 1차 모델링



## GridSearchCV Parameter Tuning

	colsample_bytree	gamma	learning_rate	max_depth	min_child_weight	n_estimators	reg_alpha	reg_lambda	subsample	mean_test_score
5	0.6	0	0.03	3	5	300	0.01	1	0.8	0.813645
9	0.6	0	0.03	3	5	500	0.01	1	0.8	0.812236
4	0.6	0	0.03	3	5	250	0.01	1	0.8	0.811954
7	0.6	0	0.03	3	5	400	0.01	1	0.8	0.811954
12	0.6	0	0.03	3	5	650	0.01	1	0.8	0.811954
6	0.6	0	0.03	3	5	350	0.01	1	0.8	0.811672
11	0.6	0	0.03	3	5	600	0.01	1	0.8	0.811672
10	0.6	0	0.03	3	5	550	0.01	1	0.8	0.811390
8	0.6	0	0.03	3	5	450	0.01	1	0.8	0.810262
13	0.6	0	0.03	3	5	700	0.01	1	0.8	0.809980
14	0.6	0	0.03	3	5	750	0.01	1	0.8	0.809134
3	0.6	0	0.03	3	5	200	0.01	1	0.8	0.808289
2	0.6	0	0.03	3	5	150	0.01	1	0.8	0.806879
1	0.6	0	0.03	3	5	100	0.01	1	0.8	0.799267
0	0.6	0	0.03	3	5	50	0.01	1	0.8	0.789681

## Parameters



## 2-5 모델링(3/8) – 1차 모델링



### (1) XGBoost - Parameter Tuning

Learning_rate	Gamma	Colsample_bytree	Max_depth	Min_child_weight	Sub-sample	Reg_alpha	Reg_lambda	N_estimators
0.03	0	0.6	3	5	0.8	0.01	1	300

### (2) 성공프로젝트 > 실패프로젝트 → Sampling

	Oversampling		Undersampling	Over+Under Sampling
	SMOTE	ADASYN	TomekLinks	SMOTEENN
Test Accuracy	81.8	81.5	79.5	74.4

Before Tuning  
80.2

(1) After Tuning  
81.4

(2) After Sampling  
81.8

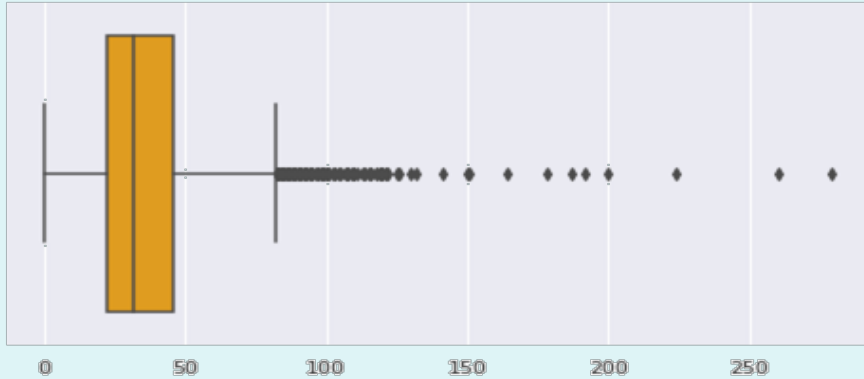


# 2-5 모델링(4/8) - 2차 모델링 준비

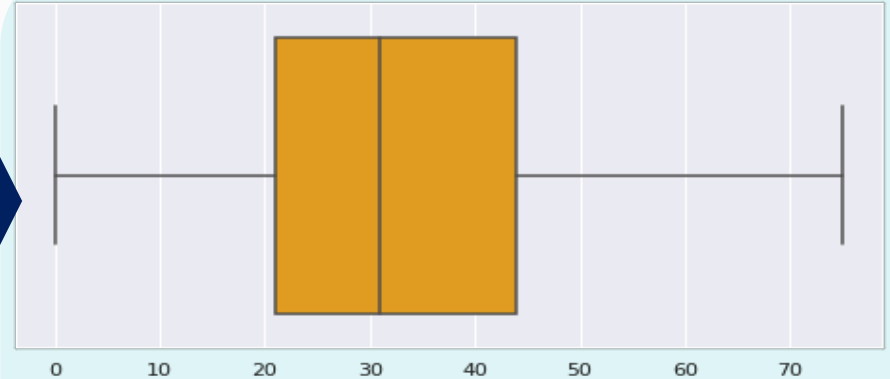


파생변수 생성 및 변수 범주화

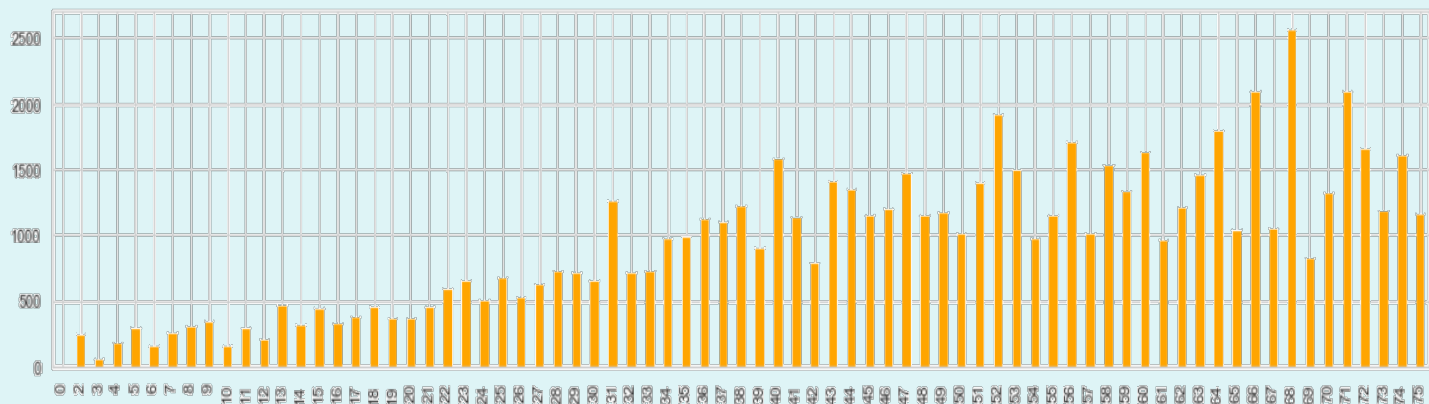
이미지 수



이상치 제거 전 이미지 개수 분포



이상치 제거 후 이미지 개수 분포



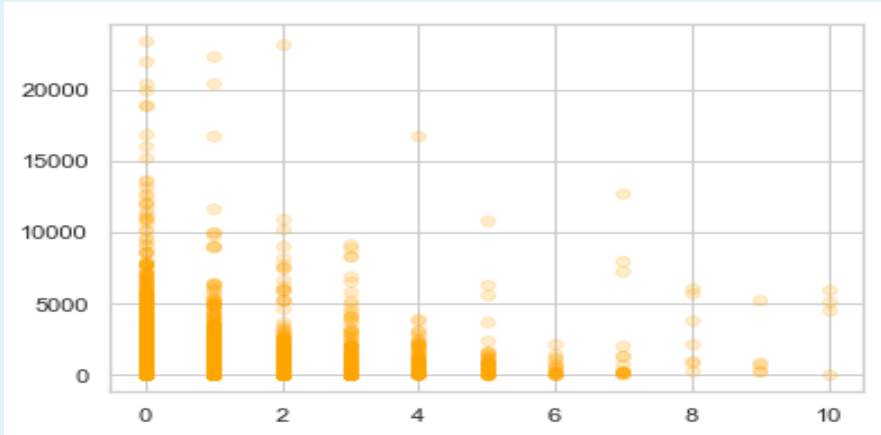
이미지 개수에 따른 달성률

# 2-5 모델링(5/8) - 2차 모델링 준비



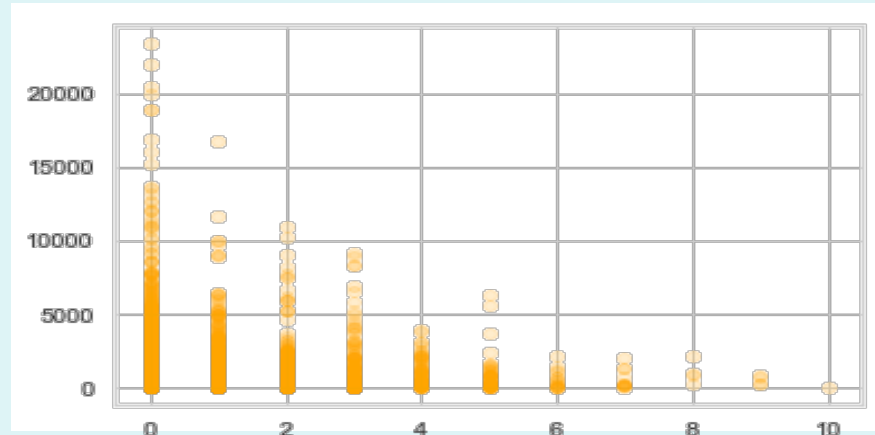
## 비디오 수

< 이상치 제거 전 >



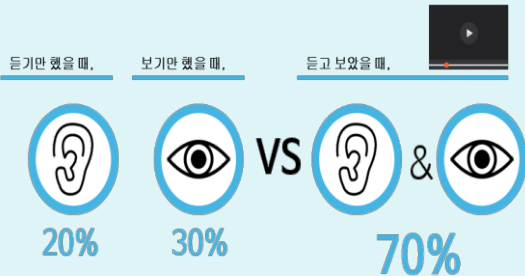
비디오수에 따른 달성률

< 이상치 제거 후 >



비디오수에 따른 달성률

## 이미지/비디오수 - 시각화변수



‘듣기만 하거나, 보기만 하는 콘텐츠보다 보고&듣고 둘 다 활용한 콘텐츠를 소비자들이 더 오래 기억한다.’

< “2016 콘텐츠 마케팅 트렌드”, IDG >

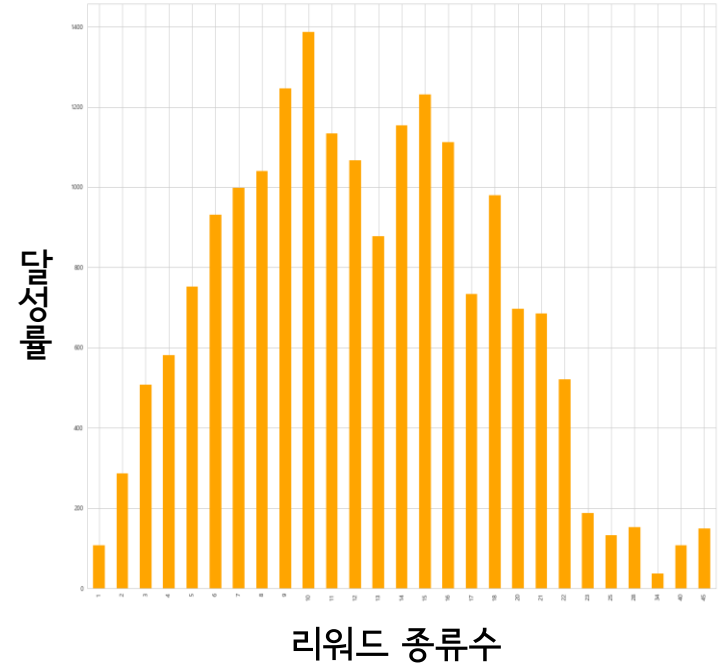
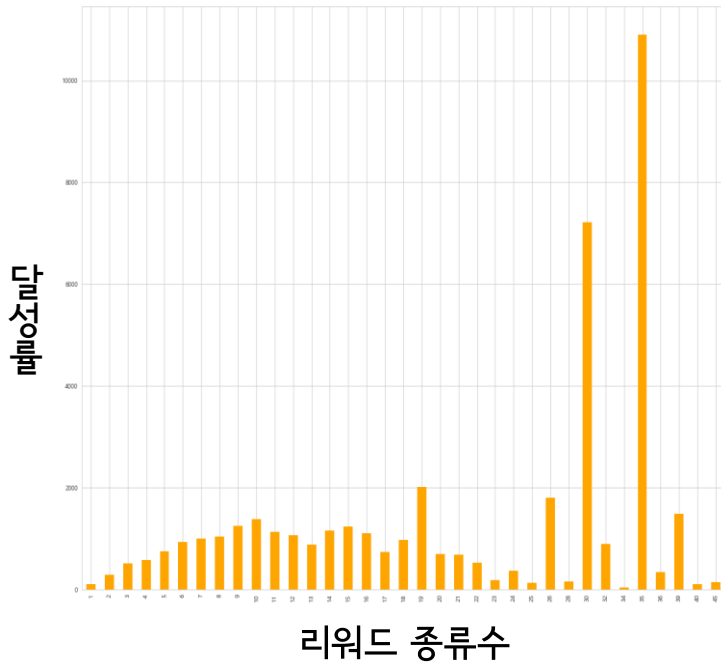
시각화 변수 생성

$$3 * \text{number of image} + \frac{7}{\text{number of video}}$$

## 2-5 모델링(6/8) - 2차 모델링 준비



### 리워드 종류수 이상치 제거



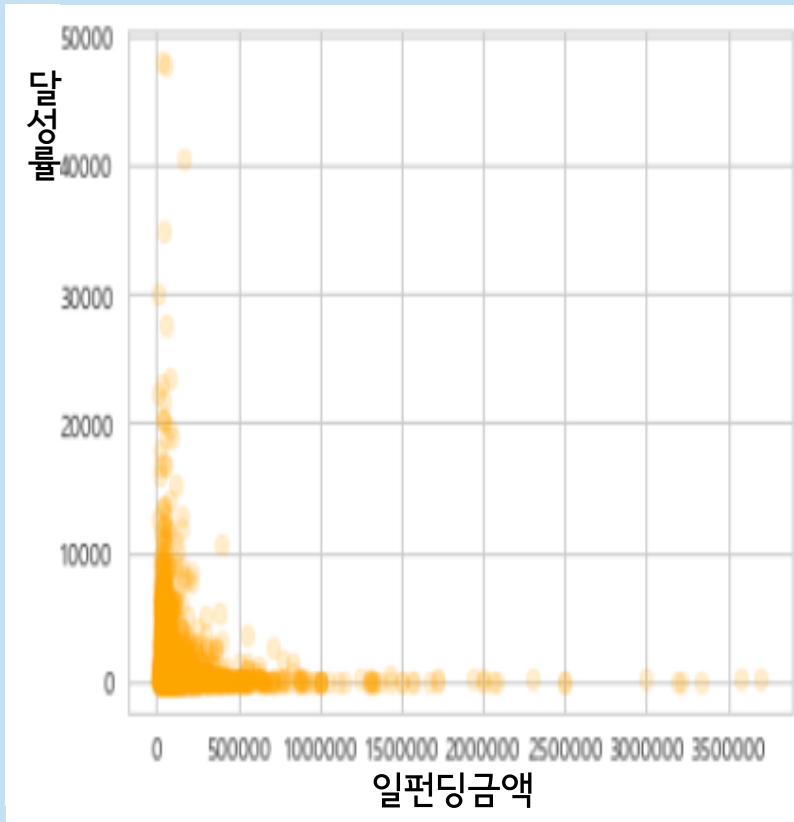
-> 특징에서 매우 높은 값들은 해당 값이 하나밖에 없어서 편중된 결과를 나타냄

- > 리워드 종류수에 따라 증가하다가 감소하는 경향
- > 범주화의 필요성
- > 범주를 나눠서 리워드종류수\_분류 변수 생성

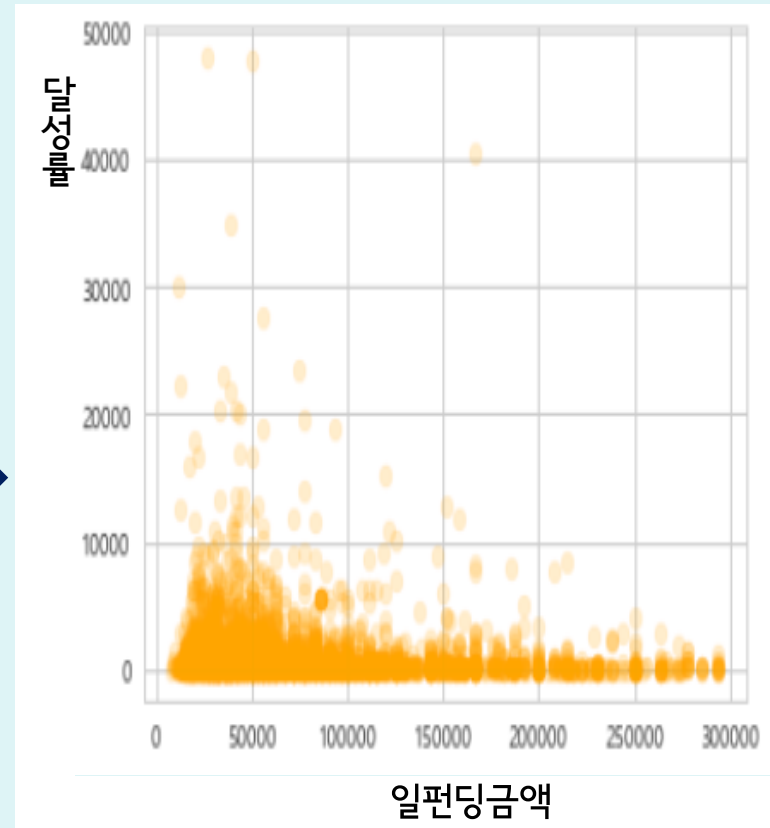
## 2-5 모델링(7/8) - 2차 모델링 준비



### 일편당금액 이상치 제거



-> 앞 쪽에는 큰 차이가 없지만 금액이 커지면 실패로 간다.



-> 범주형 변수를 만들어 진행

## 2-5 모델링(8/8) - 2차 모델링



독립변수 : 50개 -> 53개

Before  
Tuning  
80.2



After  
Tuning  
84.5

Accuracy  
84.5

Precision  
84.3

성능평가

Recall  
84.5

F1  
84.5

[[오차행렬]]

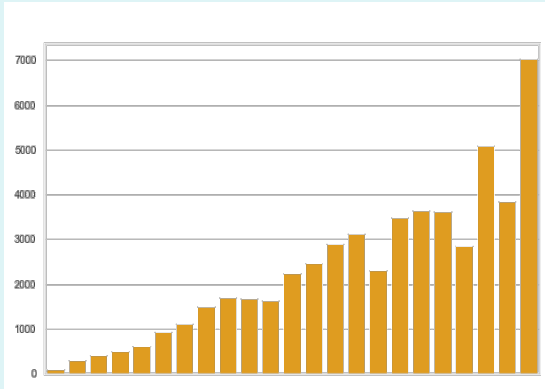
286	21	0
40	<b>697</b>	88
0	86	<b>303</b>

속성명	중요도
성공비율	1
글업데이트수	2
리워드종류수_분류	3
시각화변수	4
와디즈팔로워수	5
패션 잡화	6
홈리빙	7
일편딩_분류	8
푸드	9
뷰티	10

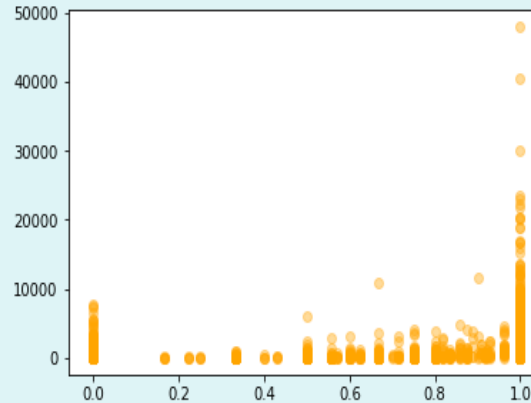
# 3-1 성공요인 분석



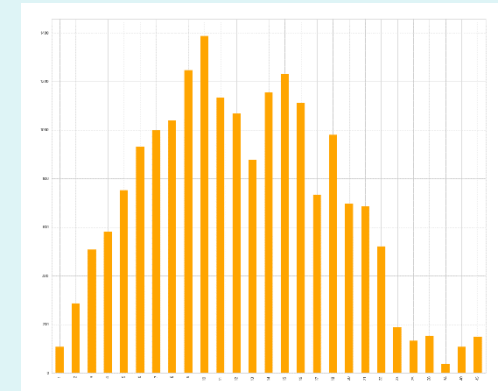
<글 업데이트 수>



<성공비율>



<리워드 종류 수 분류>



“메이커의 고객과의 소통력”

프로젝트 관련 이벤트 및 진행상황 공유



“메이커의 신뢰도”

과거의 성공경험을 통해 메이커에 대한 신뢰감 형성



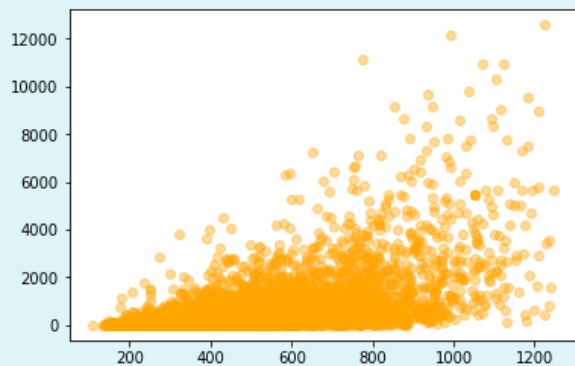
“펀딩옵션의 다양성(금액)”

다양한 가격과 옵션을 제공

# 3-1 성공요인 분석

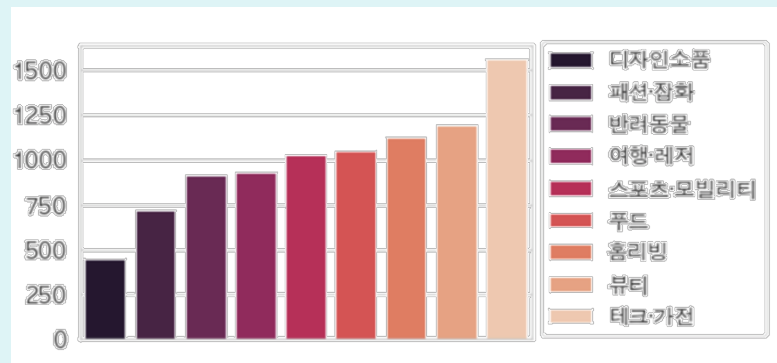


<시각화변수>



“컨텐츠 마케팅 능력“

<카테고리>



“카테고리별 달성률 차이“

## 3-3 기대효과 및 활용방안(1)



### 메이커

- 자신의 프로젝트에 대한 진단 / 마케팅 보완

### 서포터

- 성공할 프로젝트를 선별, 서포터의 시간과 비용 절약

### 와디즈

- 메이커들에게 프로젝트 성공, 달성을 향상을 위한 구체적 가이드 라인 제시



# 3-3 기대효과 및 활용방안(3)



Wadiz를 넘어

E-Commerce 상품 판매 게시물의 효과성을 높이는 모델로 확장



wadiz

**1**

아이디어(창의성) 객관적 측정 한계

**2**

데이터셋 부족

**3**

썸네일 영향력 분석의 한계