

수DA쟁이 1기 개인 프로젝트

2020 DAICON CUP

2020 DAICON CUP

지난 2년 동안의 DAICON에 관한 여러가지 데이터를 통해

DAICON 사용자들의 행동 패턴을 예측하기

Contents

- 데이터 소개
- 현재 진행 상황
- 최종 발표까지의 계획

데이터 소개

Train.csv

	DateTime	사용자	세션	신규방문자	페이지뷰
0	2018-09-09 00:00:00	19	19	8	206
1	2018-09-09 01:00:00	20	19	9	259
2	2018-09-09 02:00:00	12	9	1	48
3	2018-09-09 03:00:00	10	10	2	102
4	2018-09-09 04:00:00	6	5	3	18
...
19003	2020-11-08 19:00:00	124	123	19	3128
19004	2020-11-08 20:00:00	166	159	29	4864
19005	2020-11-08 21:00:00	184	173	32	3426
19006	2020-11-08 22:00:00	163	155	34	2845
19007	2020-11-08 23:00:00	160	152	33	3293

19008 rows × 5 columns

훈련시킴 데이터

2018년 9월 9일 ~ 2020년 11월 8일 기간 동안 기록된
한 시간 간격의 사용자 행동 데이터

columns : 사용자 수, 세션 수, 신규 방문자 수, 페이지 뷰 수

데이터 소개

Info_user.csv

	유저 id	아이디 생성 시점	대회 참여 횟수	코드 공유 횟수	토론 횟수	국가	파이썬 수준	데이터 사이언스 수준
1	19195.0	2018-08-08 18:21	1.0	0.0	0.0	82.0	1.0	1.0
2	16339.0	2018-08-08 19:57	1.0	0.0	0.0	82.0	1.0	1.0
3	7290.0	2018-08-08 19:59	0.0	0.0	0.0	82.0	1.0	1.0
4	21287.0	2018-08-08 20:26	0.0	0.0	0.0	82.0	1.0	1.0
5	3828.0	2018-08-08 20:28	0.0	0.0	0.0	82.0	1.0	1.0
...
21687	11584.0	2020-12-08 22:53	1.0	0.0	0.0	82.0	1.0	1.0
21688	6337.0	2020-12-08 22:54	1.0	0.0	0.0	82.0	1.0	1.0
21689	2505.0	2020-12-08 23:07	1.0	0.0	0.0	82.0	1.0	1.0
21690	14182.0	2020-12-08 23:09	1.0	0.0	0.0	82.0	1.0	1.0
21691	8843.0	2020-12-08 23:13	2.0	0.0	0.0	82.0	1.0	1.0

21691 rows × 8 columns

데이터 소개

Info_competition.csv

	대회 id	대회 시작 시점	대회 종료 시점	대회 이름	키워드	참여 자 수	최대 팀 멤버	하루 최대 제출 횟수	상금	상금 정보	팀 병합 데드라인	보여지 는 위너 수
0	136	2018-08-14 0:00	2018-09-13 23:59	대출 상점 총 매출 예측 경진대회	금융 소상공인 신용카드 가맹점 빅데이터와 AI로 매출 예측 시계열, 회귀 ...	303	5	5	850	\$8,500 + 100,000ZPR	2018-09-13 23:59	3
1	9565	2018-09-15 0:00	2018-10-13 23:59	병원 개/폐업 분류 예측 경진대회	금융 병원 재무 데이터와 AI로 개업/폐업 예측 분석 분류 Accuracy	448	10	3	350	\$3,500 + 40,000ZPR	2018-10-13 23:59	3
2	17801	2018-10-18 0:00	2018-12-31 23:59	아파트 경매가격 예측 경진대회	금융 부동산 아파트 경매 빅데이터와 AI로 경매가 예측 분석 회귀 RMSE	316	10	3	1000	\$10,000 + 120,000ZPR	2018-12-31 23:59	3
3	21265	2018-11-13 0:00	2019-01-31 23:59	아파트 실거래가 예측	금융 부동산 빅데이터와 AI를 이용하여 실거래가를 예측 분석 회귀 RMSE	568	10	3	800	\$8,000 + 80,000ZPR	2019-01-31 23:59	3
4	42473	2018-12-25 0:00	2019-01-10 23:59	신용카드 거래 데이터 시각화	금융 신용카드, 시계열, 시각화 Python, R, Tableau, Spot...	252	5	3	100	\$1,000 + 100,000ZPR	2019-01-10 23:59	3
5	62540	2019-02-08 0:00	2019-07-18 23:59	KBO 타자 OPS 예측 경진대회	스포츠 KBO 타자 빅데이터와 AI로 OPS 예측 시계열, 회귀 RMSE...	341	5	3	800	800만원	2019-07-18 23:59	3
6	68346	2019-03-26 0:00	2019-05-20 23:59	KBO 외국인 투수 스카우팅 최적화 경진대회	스포츠 MLB 데이터와 AI를 이용 투수 스카우트 분석 KBO, 회귀 ...	98	1	3	200	200만원	2019-05-20 23:59	3
7	82407	2019-05-06 0:00	2019-07-08 23:59	KCB 금융스타일 시각화 경진대회	금융 구인 개인 신용카드 빅데이터 AI 분석 시각화 Python, R, ...	269	5	3	1000	1,000만원	2019-07-08 23:59	3
8	140472	2019-07-11 0:00	2019-10-21 23:59	상점 신용카드 매출 예측 경진대회	금융 구인 소상공인 가맹점 신용카드 빅데이터와 AI로 매출 예측 분석 시...	624	5	100	200	200만원	2020-10-28 23:59	3
9	196878	2019-10-01 0:00	2019-10-27 23:59	전력 수요량 예측 경진대회	공공 전력 기상 빅데이터와 AI로 수요량 분석 시계열, 회귀 SMAPE	478	4	3	600	600만원	2019-10-27 23:59	3

데이터 소개

Info_submission.csv

	제출 아이디	제출 대회 아이디	팀 아이디	유저 아이디	제출 시점
0	-250918.0	229255.0	5019.0	15880.0	2019-11-24 20:40
1	-250904.0	229611.0	5162.0	19772.0	2019-11-24 20:36
2	-250881.0	235401.0	5042.0	17801.0	2019-11-24 20:24
3	-250873.0	235401.0	5024.0	439.0	2019-11-24 20:23
4	-250860.0	229255.0	4890.0	13722.0	2019-11-24 20:19
...
101582	483653.0	235658.0	44883.0	11628.0	2020-12-08 23:53
101583	483654.0	235671.0	45009.0	10592.0	2020-12-08 23:56
101584	483655.0	235671.0	45013.0	5987.0	2020-12-08 23:56
101585	483656.0	235658.0	42355.0	6586.0	2020-12-08 23:57
101586	483657.0	235658.0	42355.0	6586.0	2020-12-08 23:57

99203 rows × 5 columns

데이터 소개

Info_login.csv

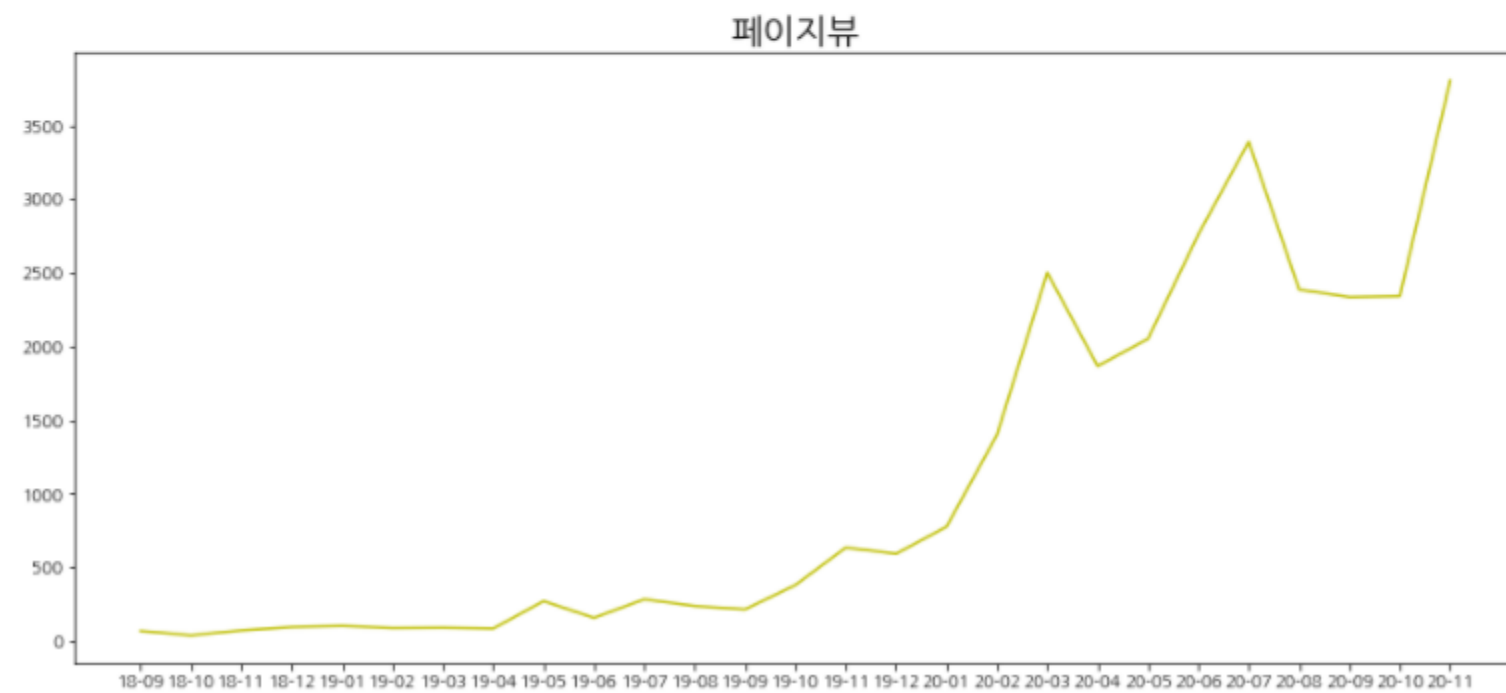
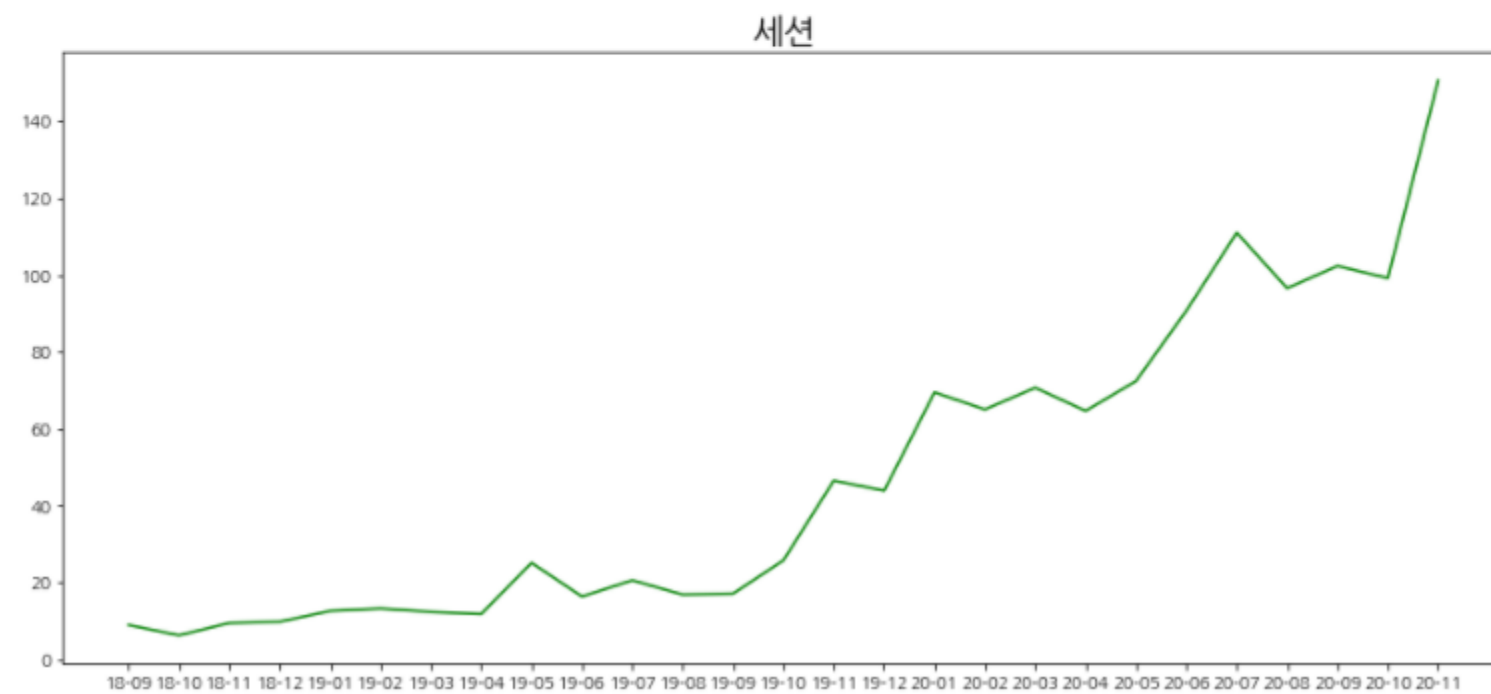
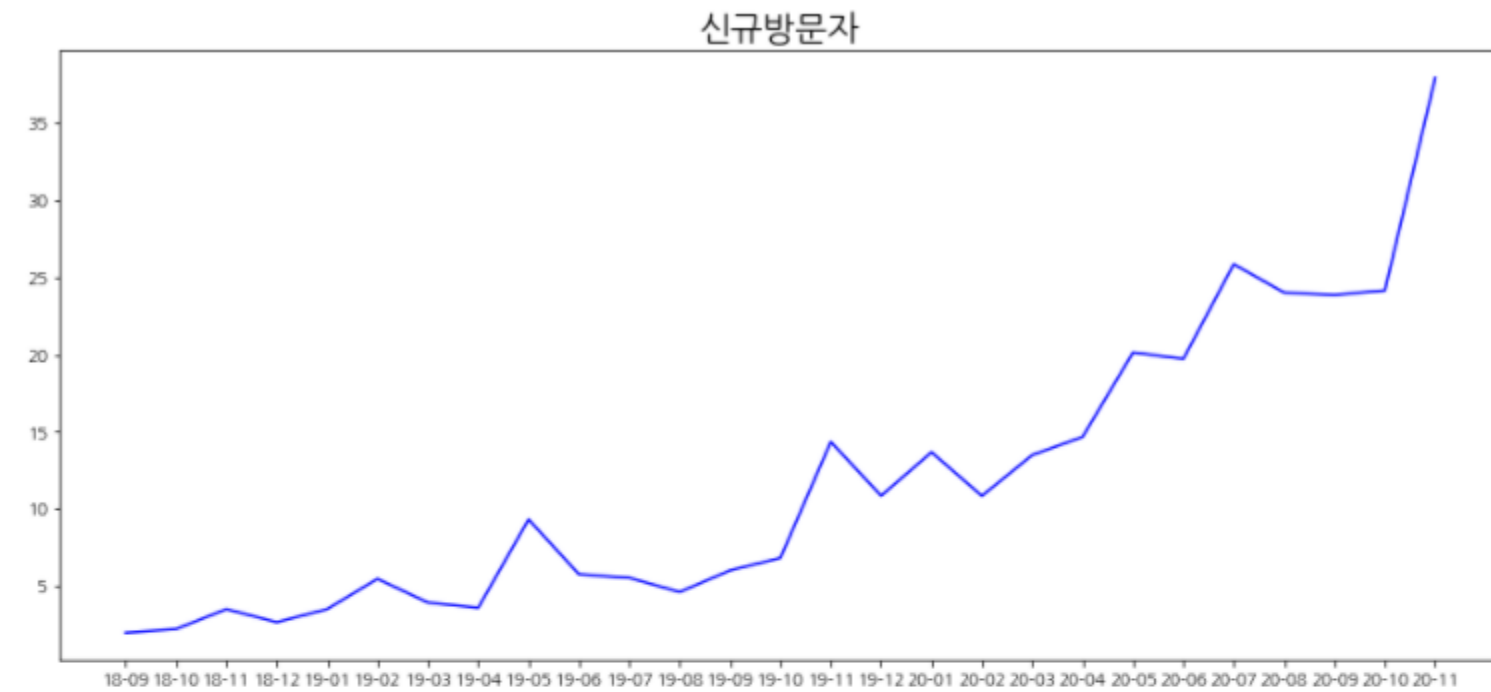
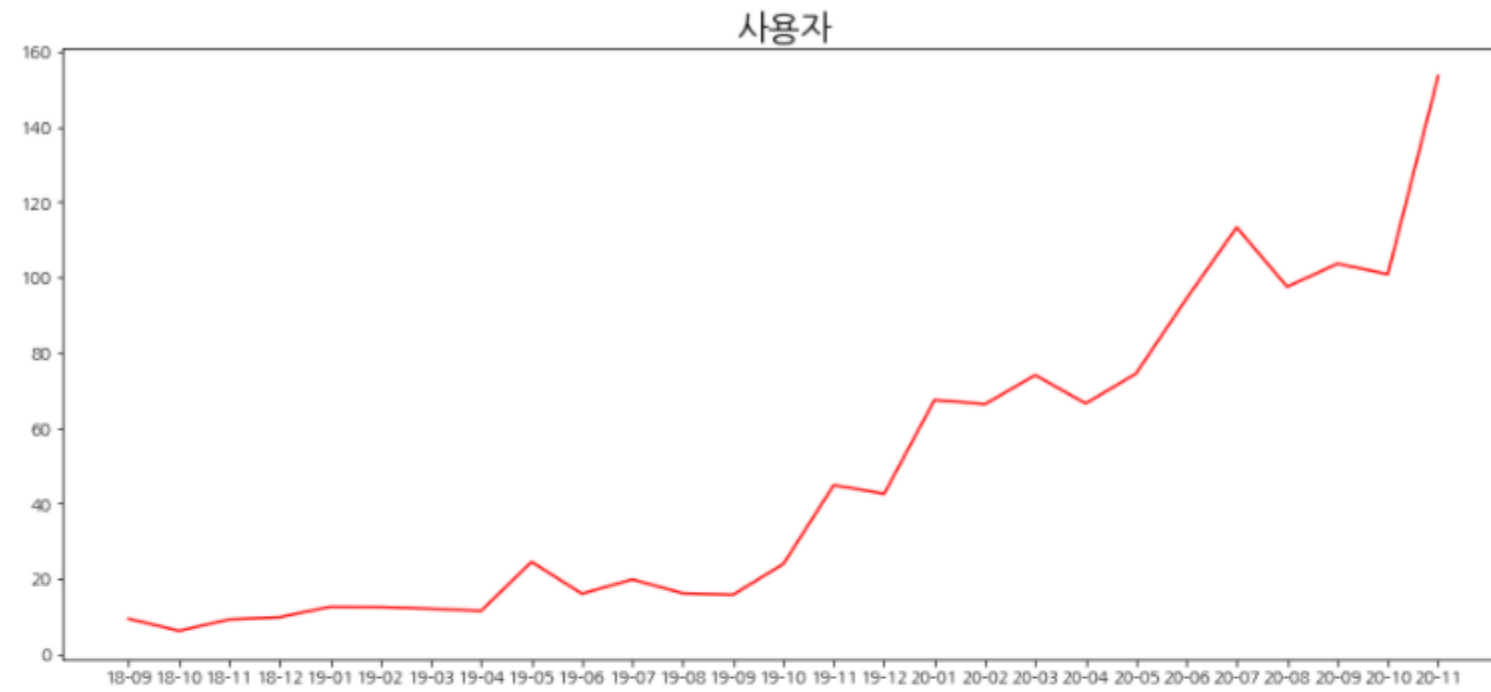
	로그인 id	유저 id	로그인 시점	로그인한 플랫폼	로그인한 브라우저
0	14196.0	19195.0	2018-09-23 1:30	Windows 8	Internet Explorer 10.0
1	14234.0	22045.0	2018-09-23 2:46	Windows	Chrome 67.0.3396.99
2	14256.0	8790.0	2018-09-23 3:35	Windows	Chrome 69.0.3497.100
5	14280.0	10206.0	2018-09-23 13:24	Windows	Chrome 68.0.3440.106
6	14288.0	488.0	2018-09-23 14:05	Windows 7	Chrome 68.0.3440.106
...
64649	328208.0	16088.0	2019-12-30 23:19	Windows	Firefox 71.0
64650	328209.0	14107.0	2019-12-30 23:21	Windows	Chrome 79.0.3945.88
64651	328210.0	1283.0	2019-12-30 23:21	Windows	Chrome 79.0.3945.88
64652	328218.0	4077.0	2019-12-30 23:24	Windows	Chrome 79.0.3945.88
64653	328219.0	21531.0	2019-12-30 23:32	Apple	Chrome 78.0.3904.108

60397 rows × 5 columns

현재 진행 상황

EDA(Exploratory Data Analysis)

월별에 따른 사용자 데이터

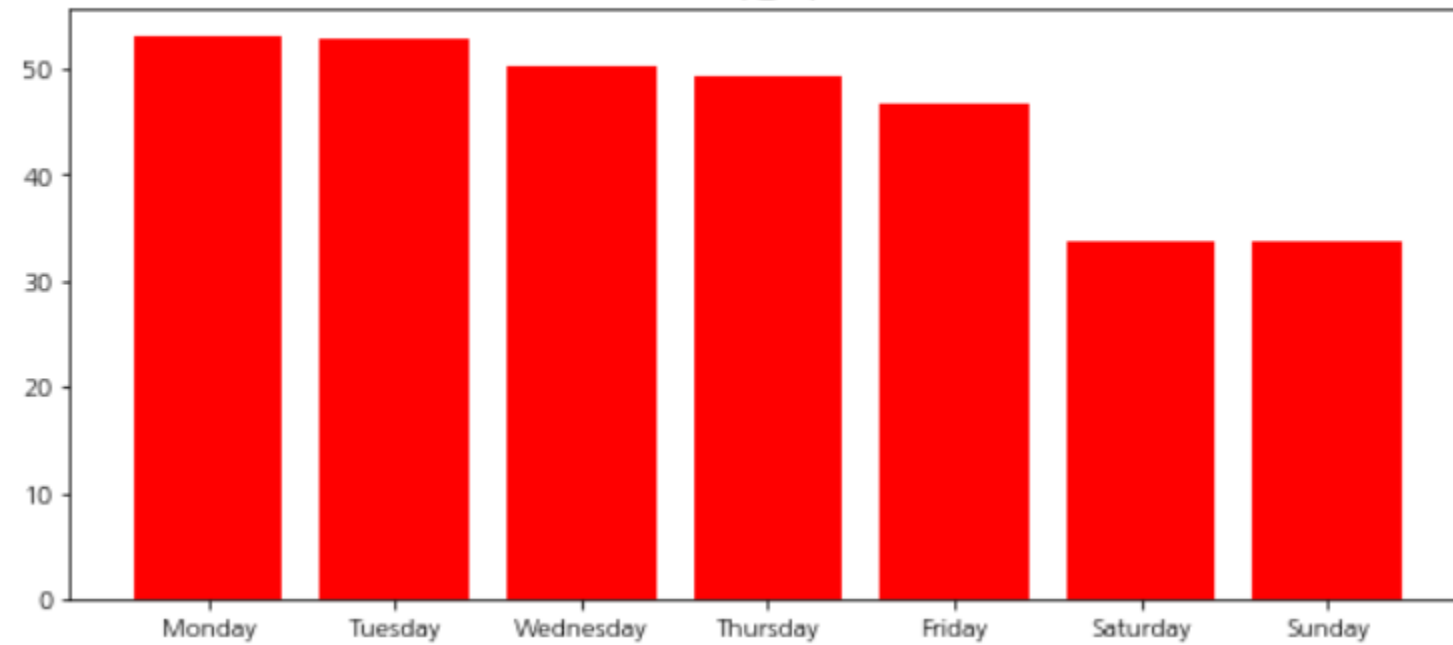


현재 진행 상황

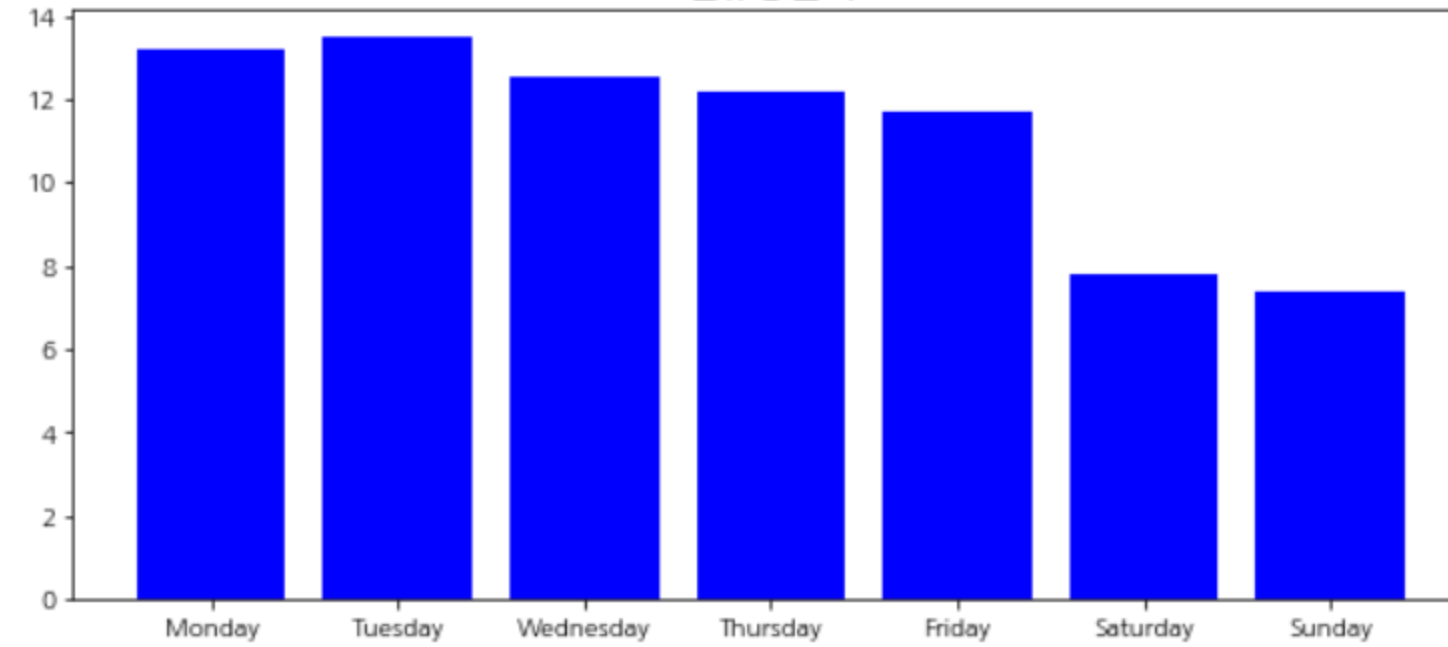
EDA(Exploratory Data Analysis)

요일별에 따른 사용자 데이터

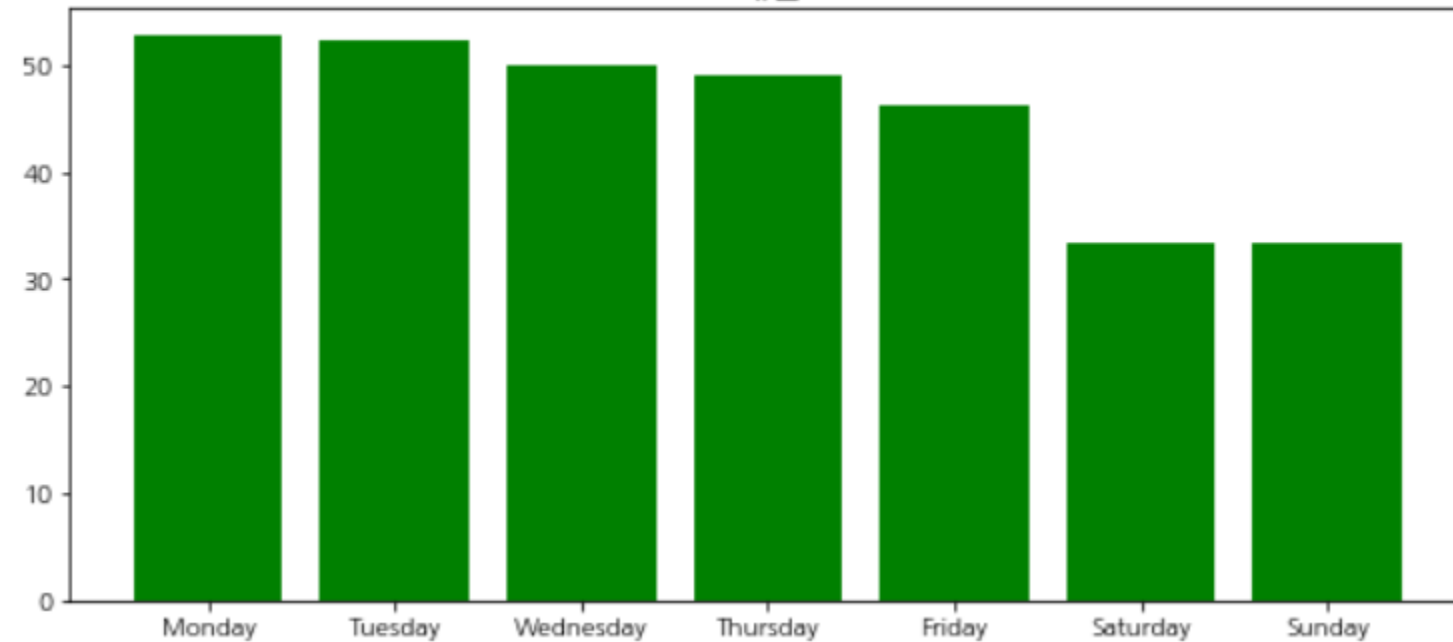
사용자



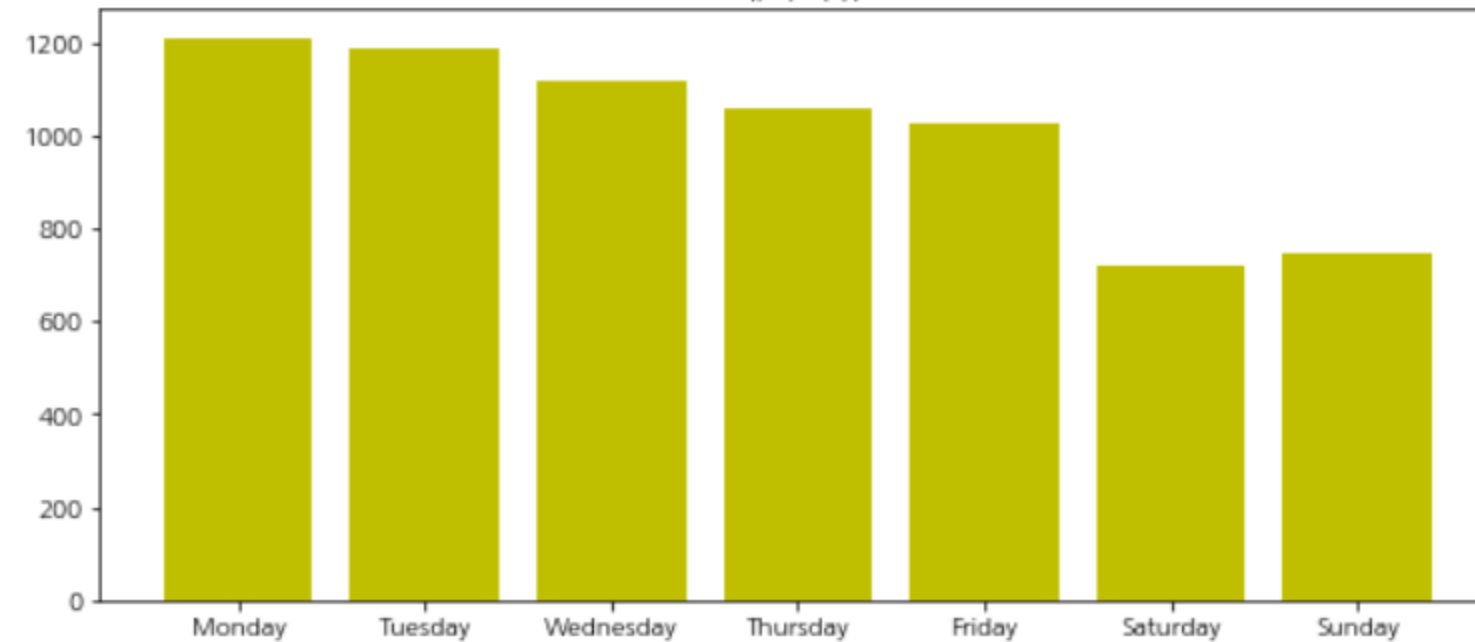
신규방문자



세션



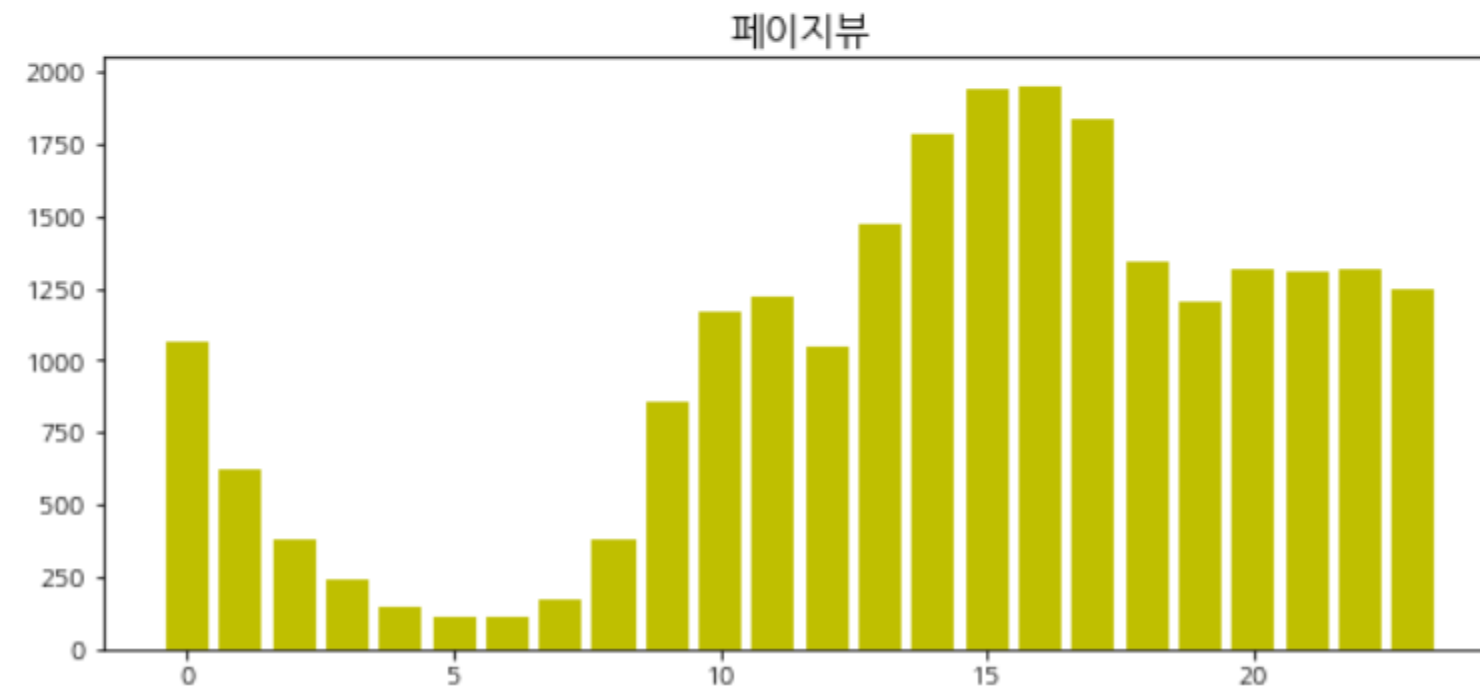
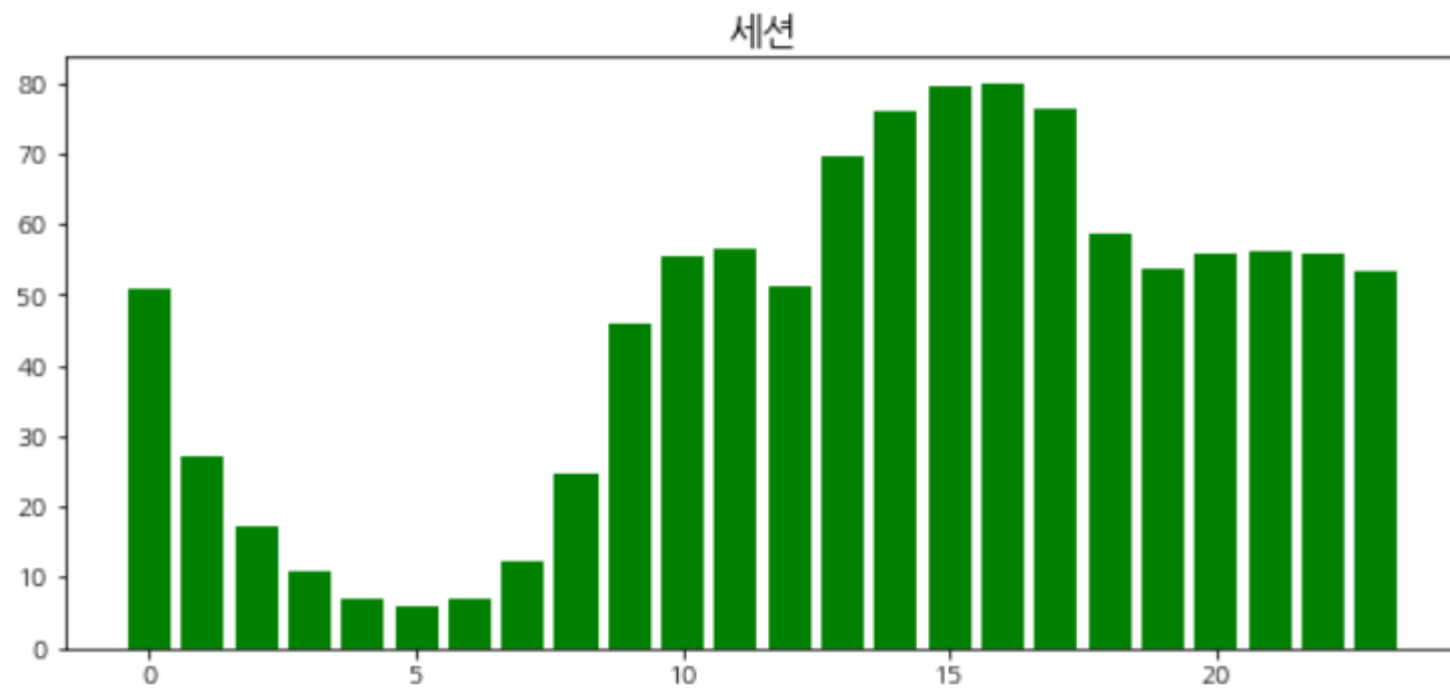
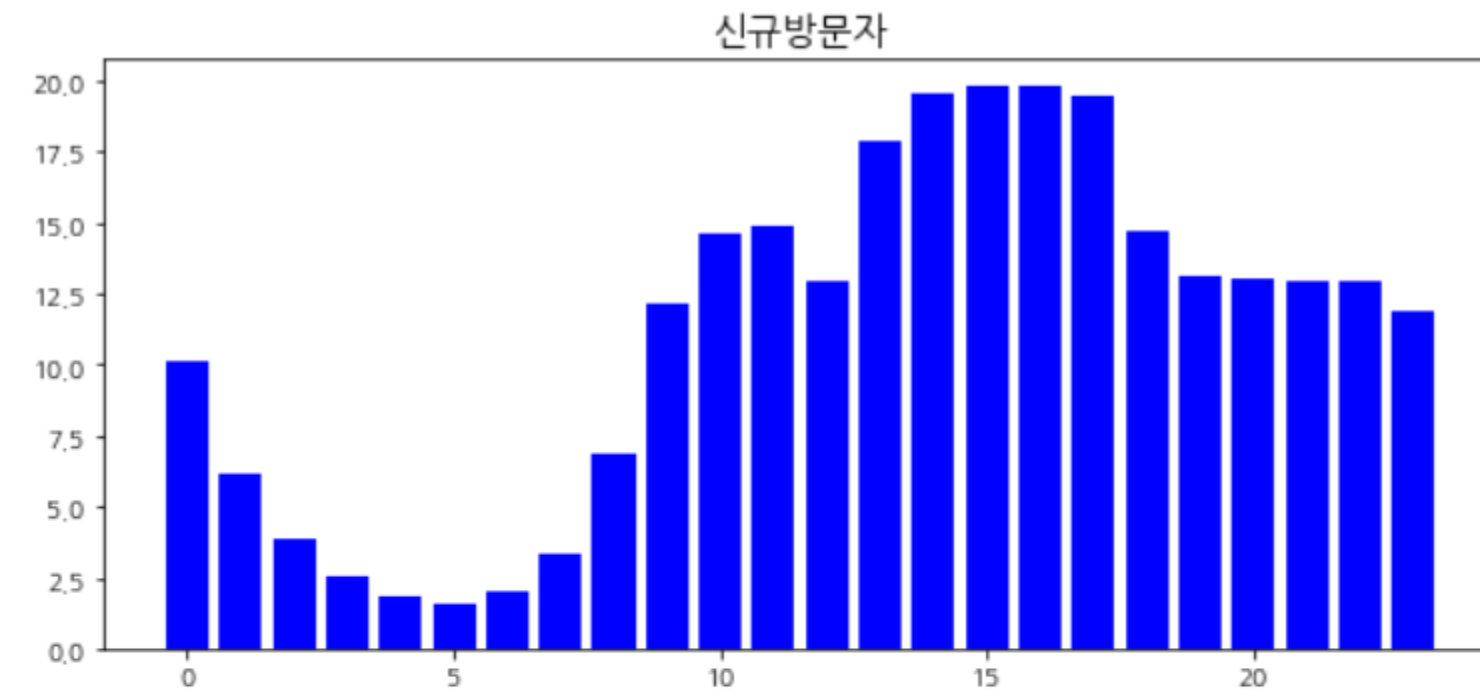
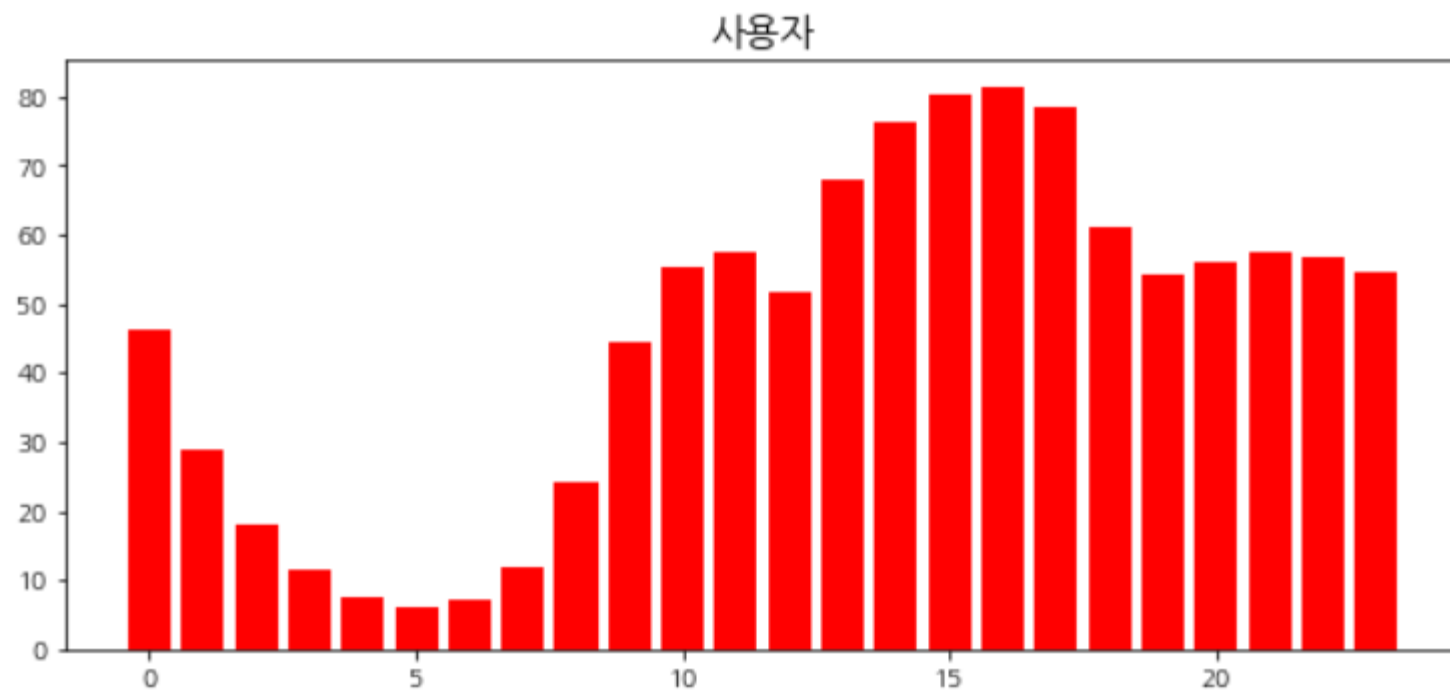
페이지뷰



현재 진행 상황

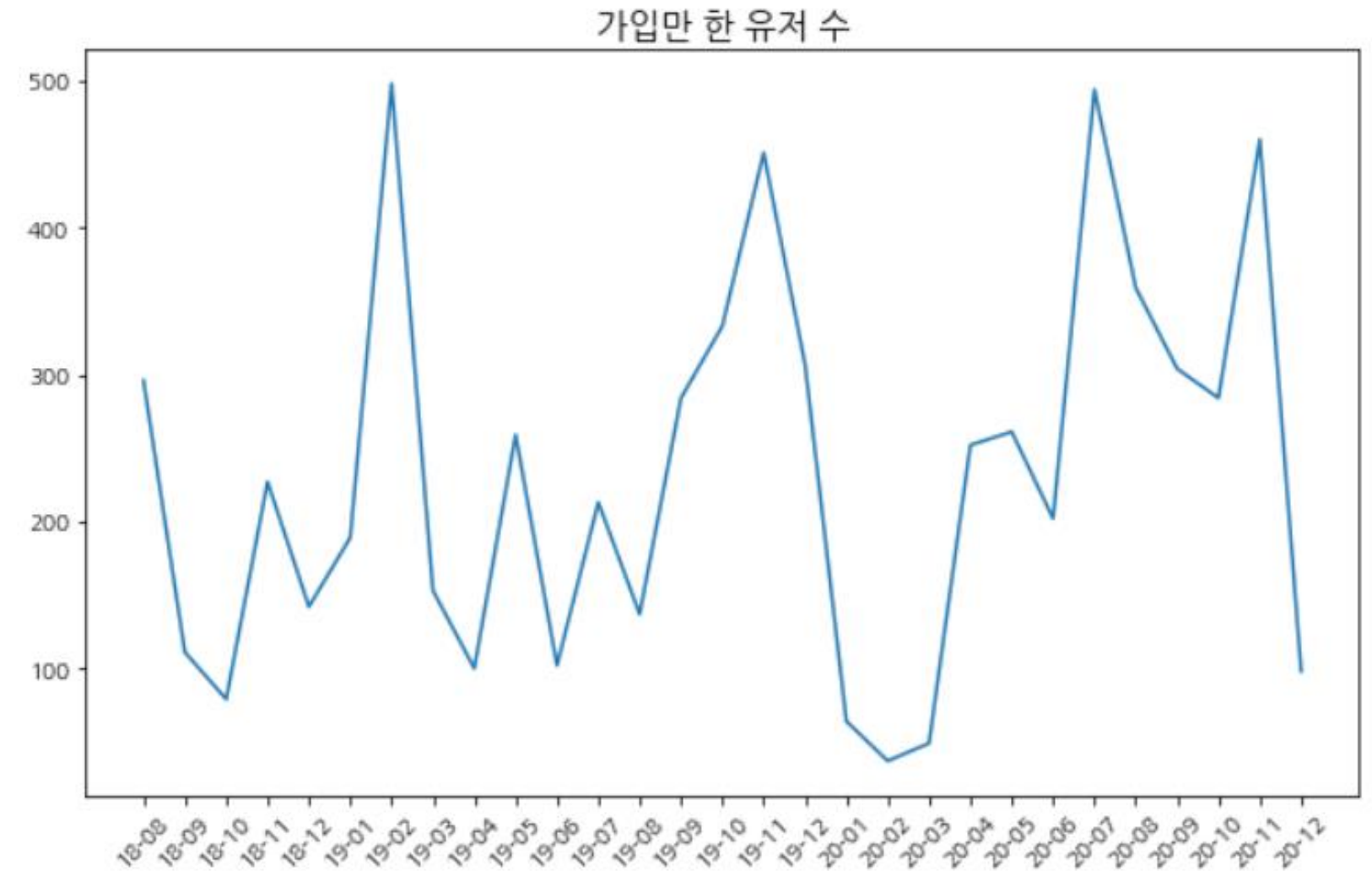
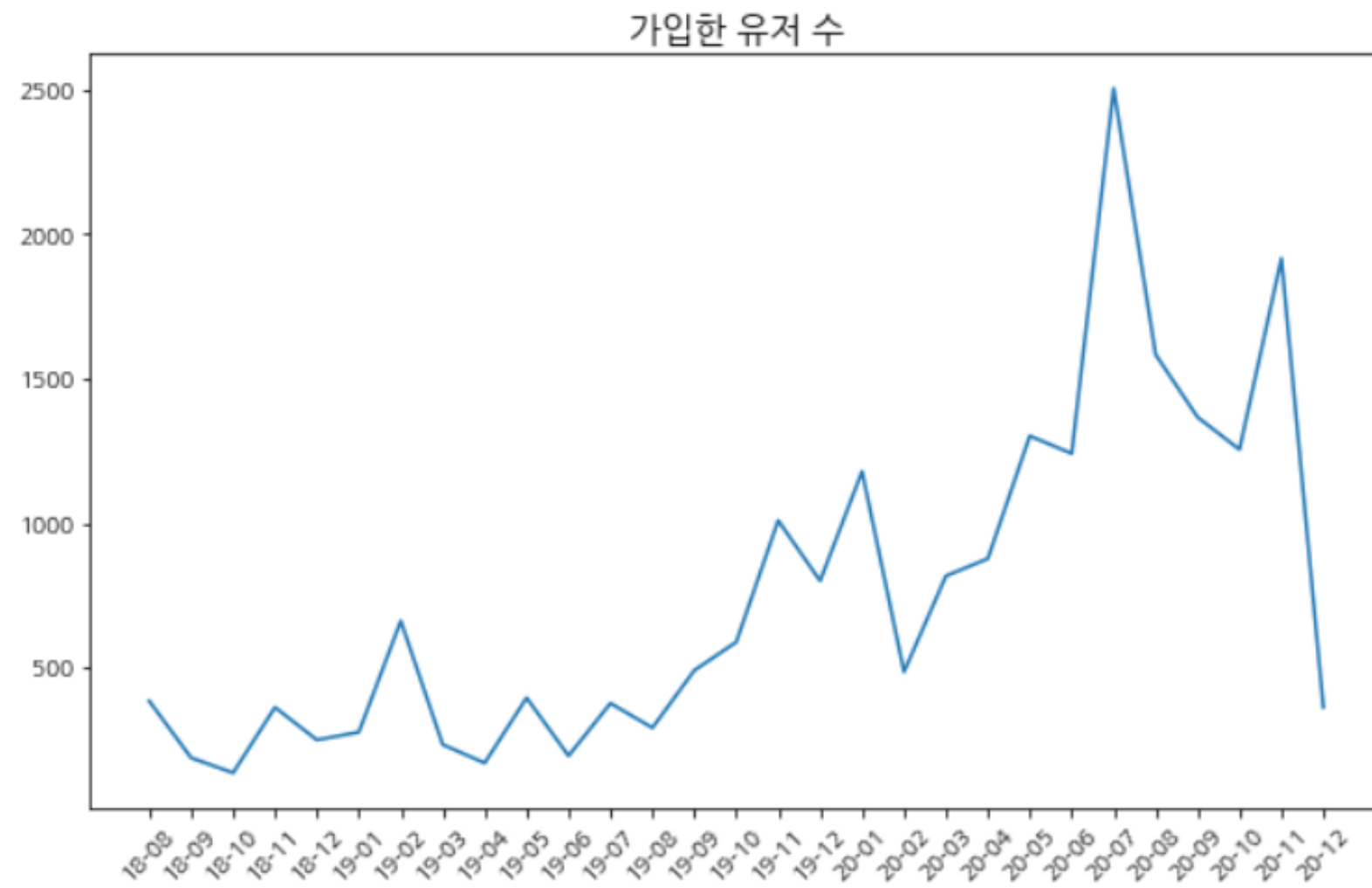
EDA(Exploratory Data Analysis)

시간대별에 따른 사용자 데이터



현재 진행 상황

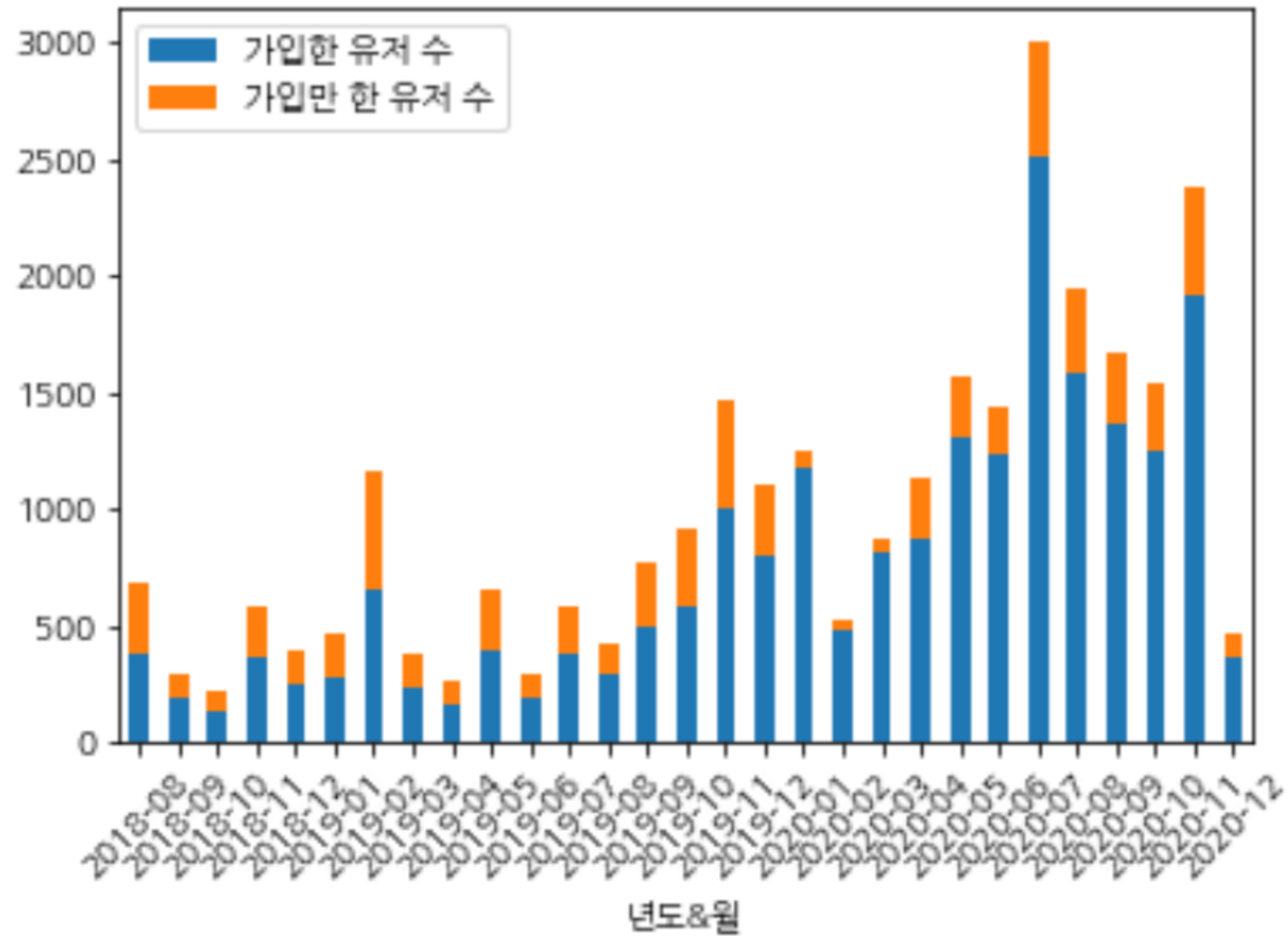
EDA(Exploratory Data Analysis)



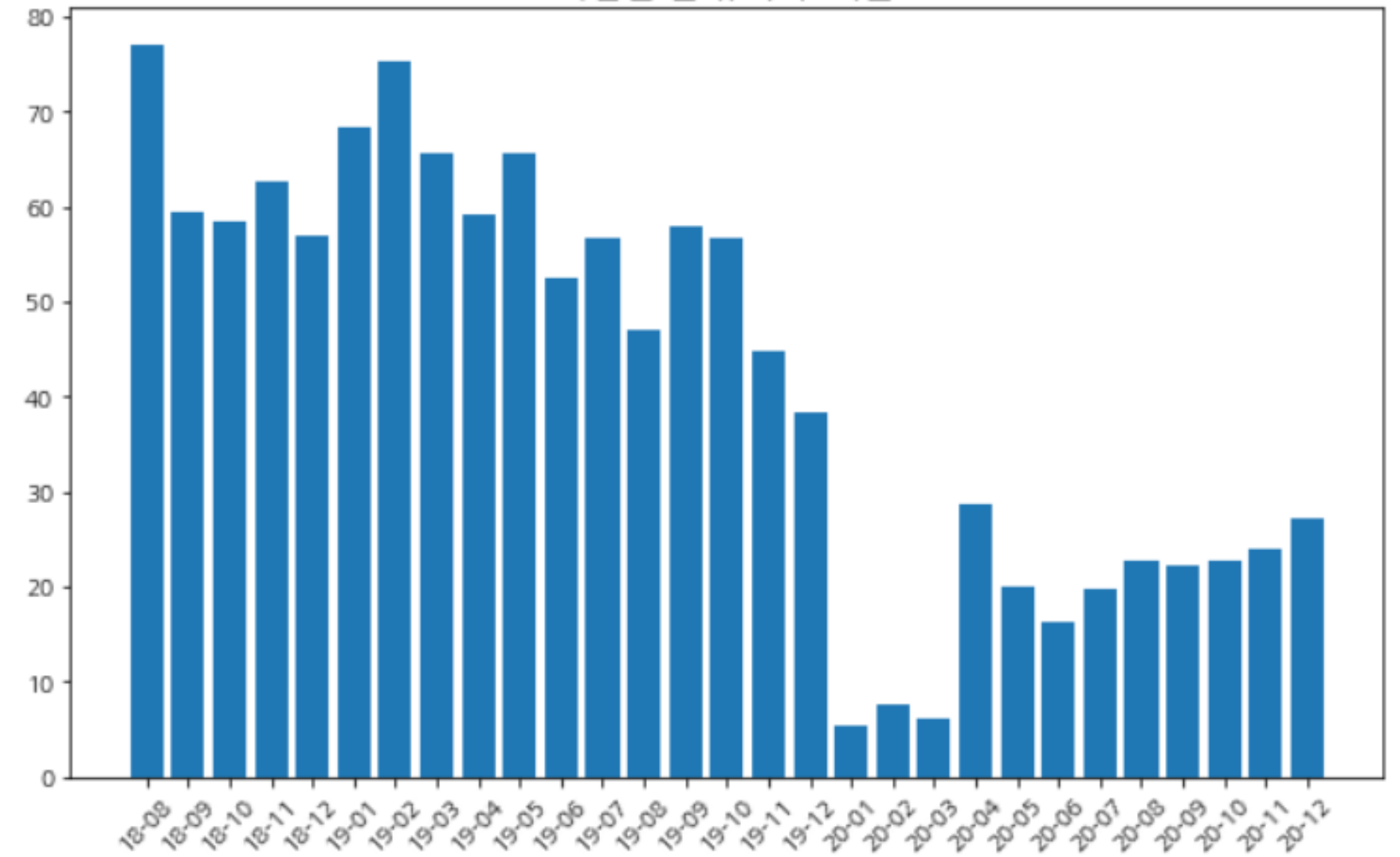
현재 진행 상황

EDA(Exploratory Data Analysis)

유저 수 집계



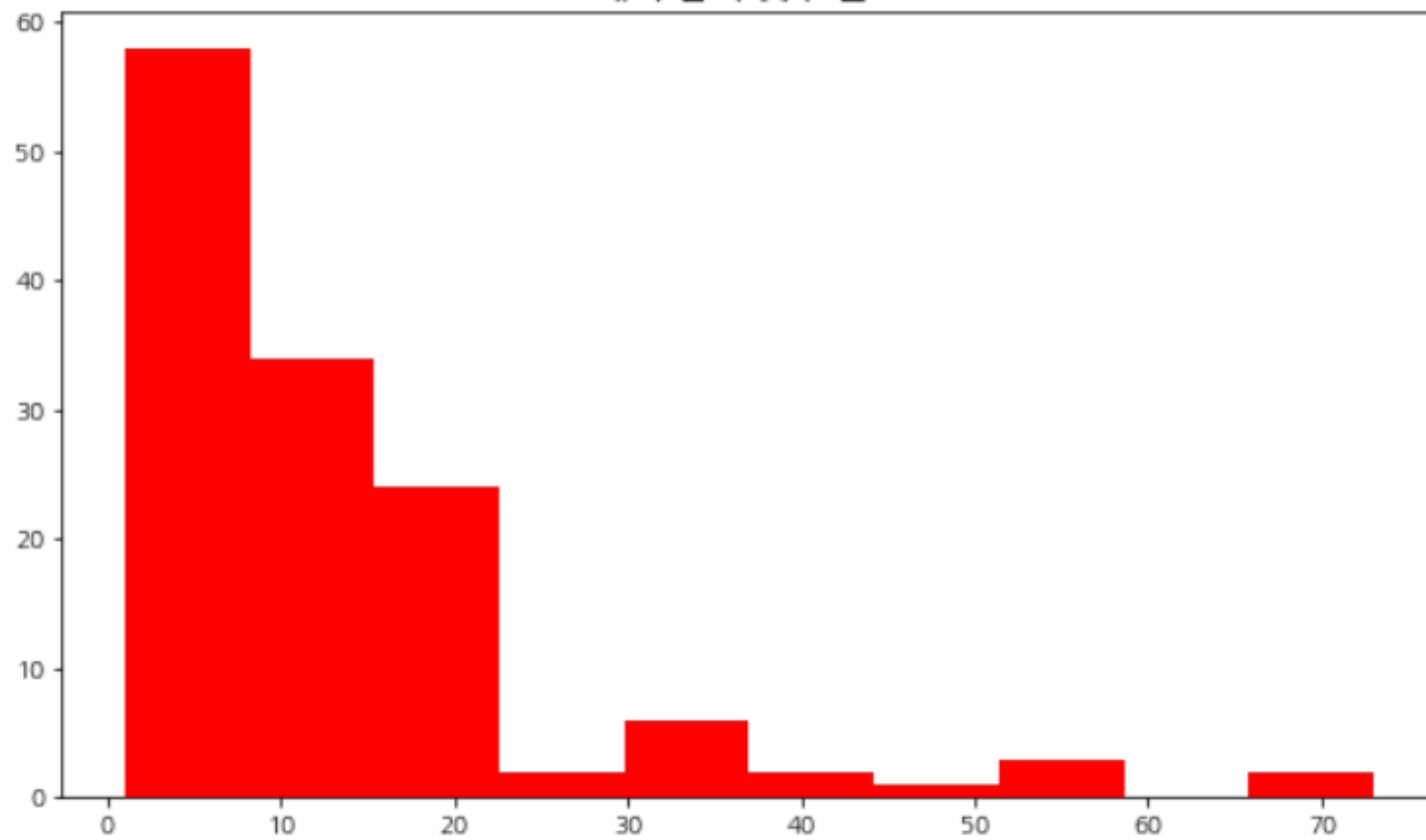
가입만 한 유저 수 비율



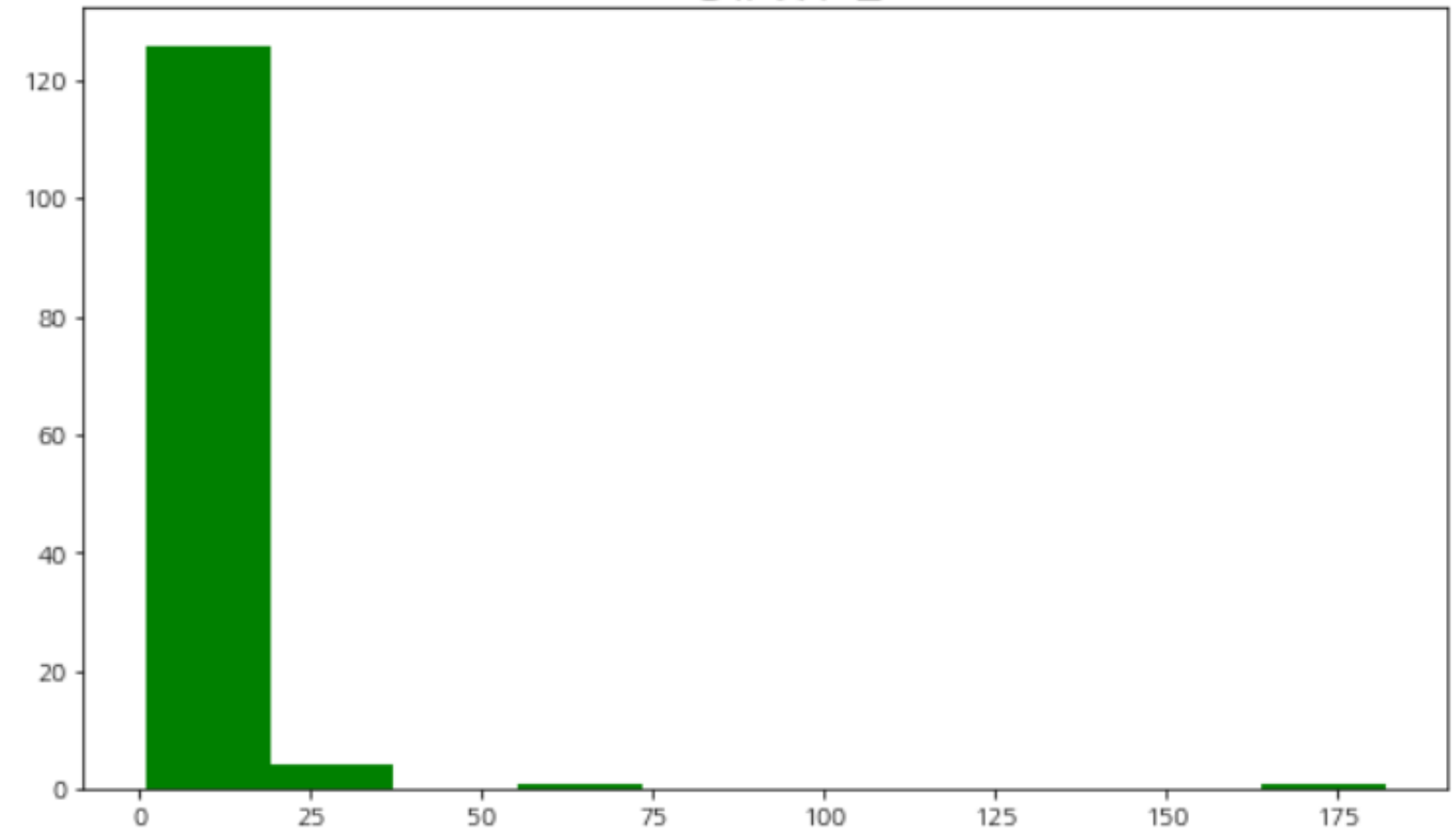
현재 진행 상황

EDA(Exploratory Data Analysis)

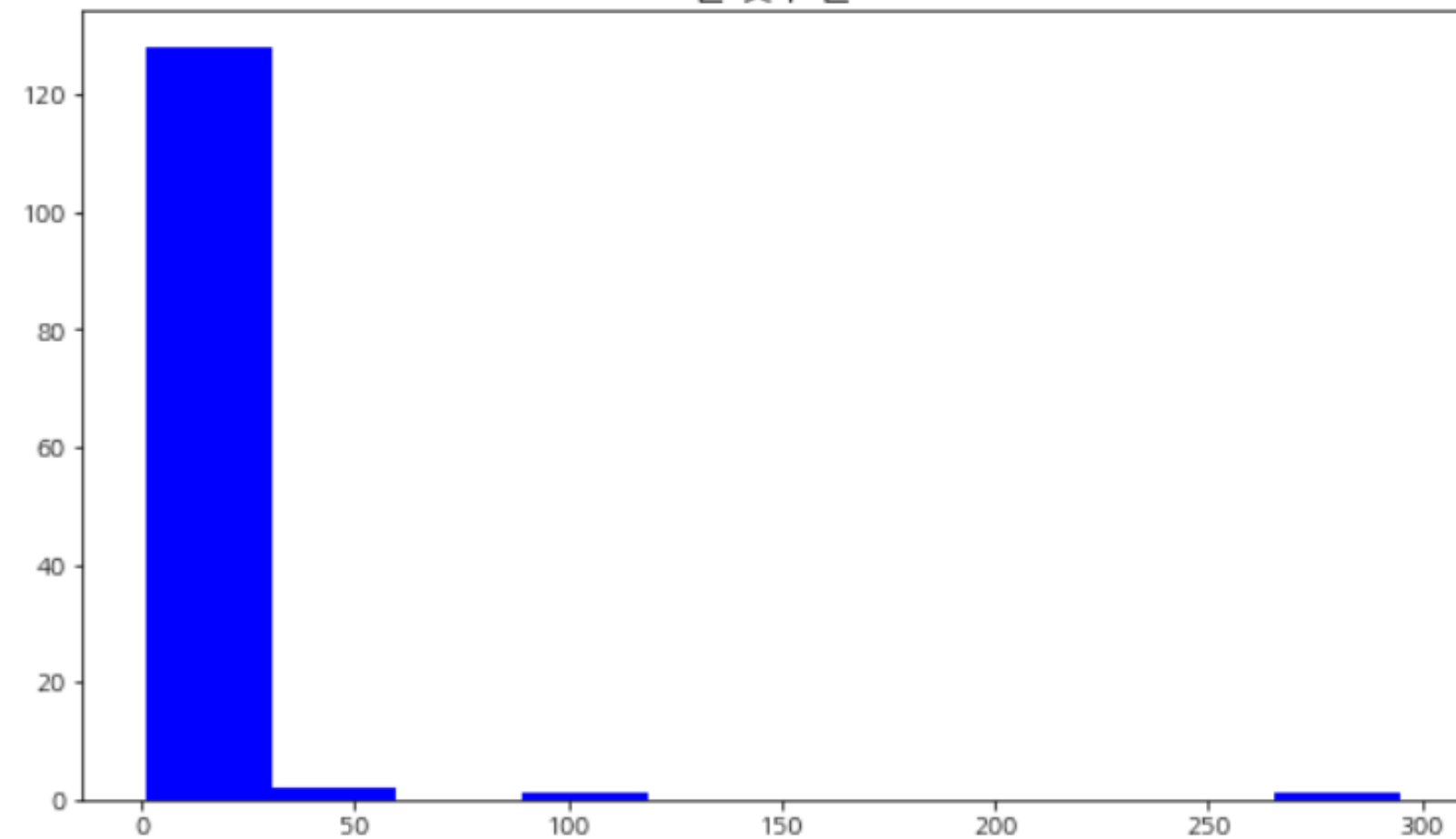
대회 참여 횟수 분포



코드 공유 횟수 분포



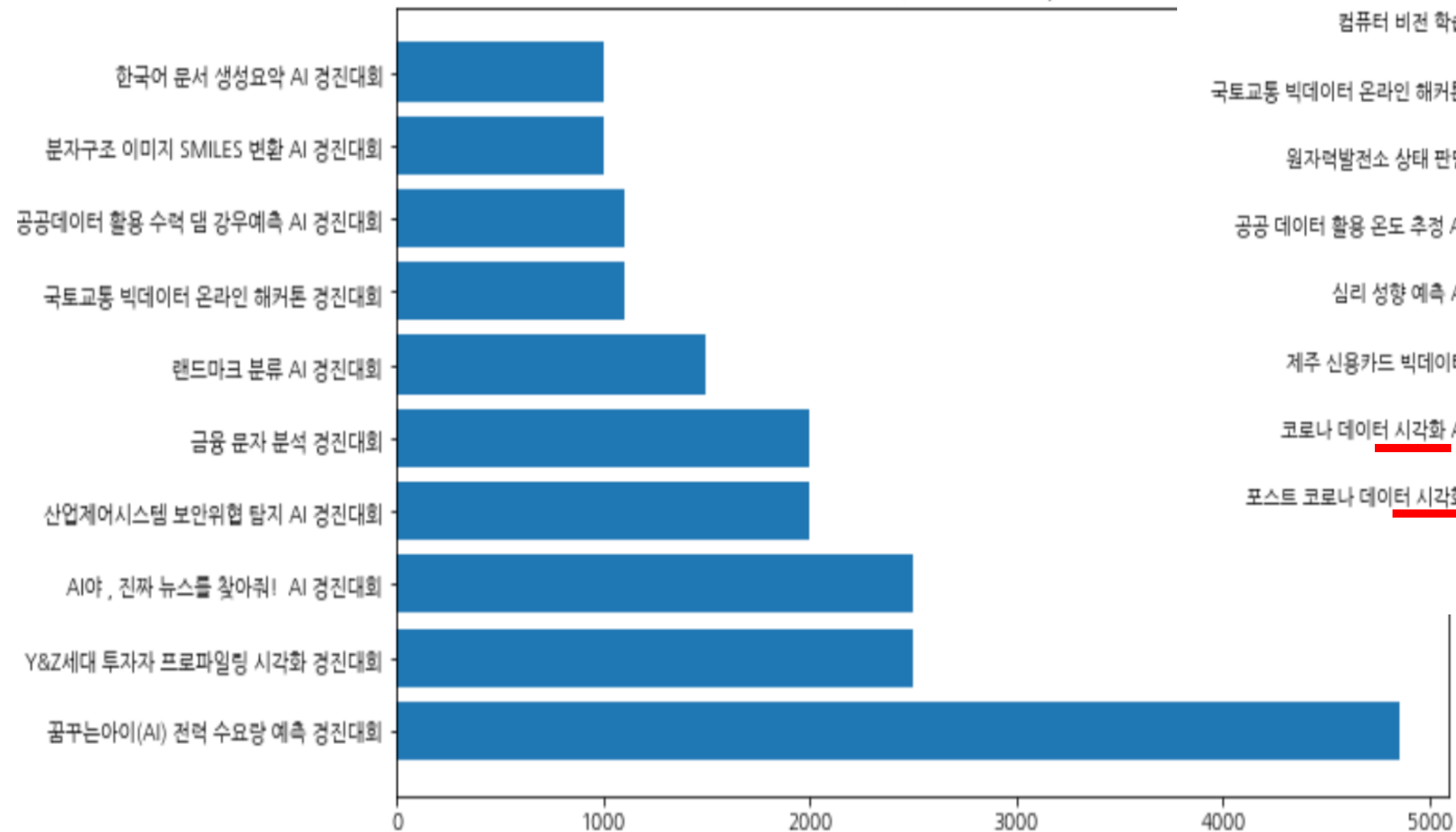
토론 횟수 분포



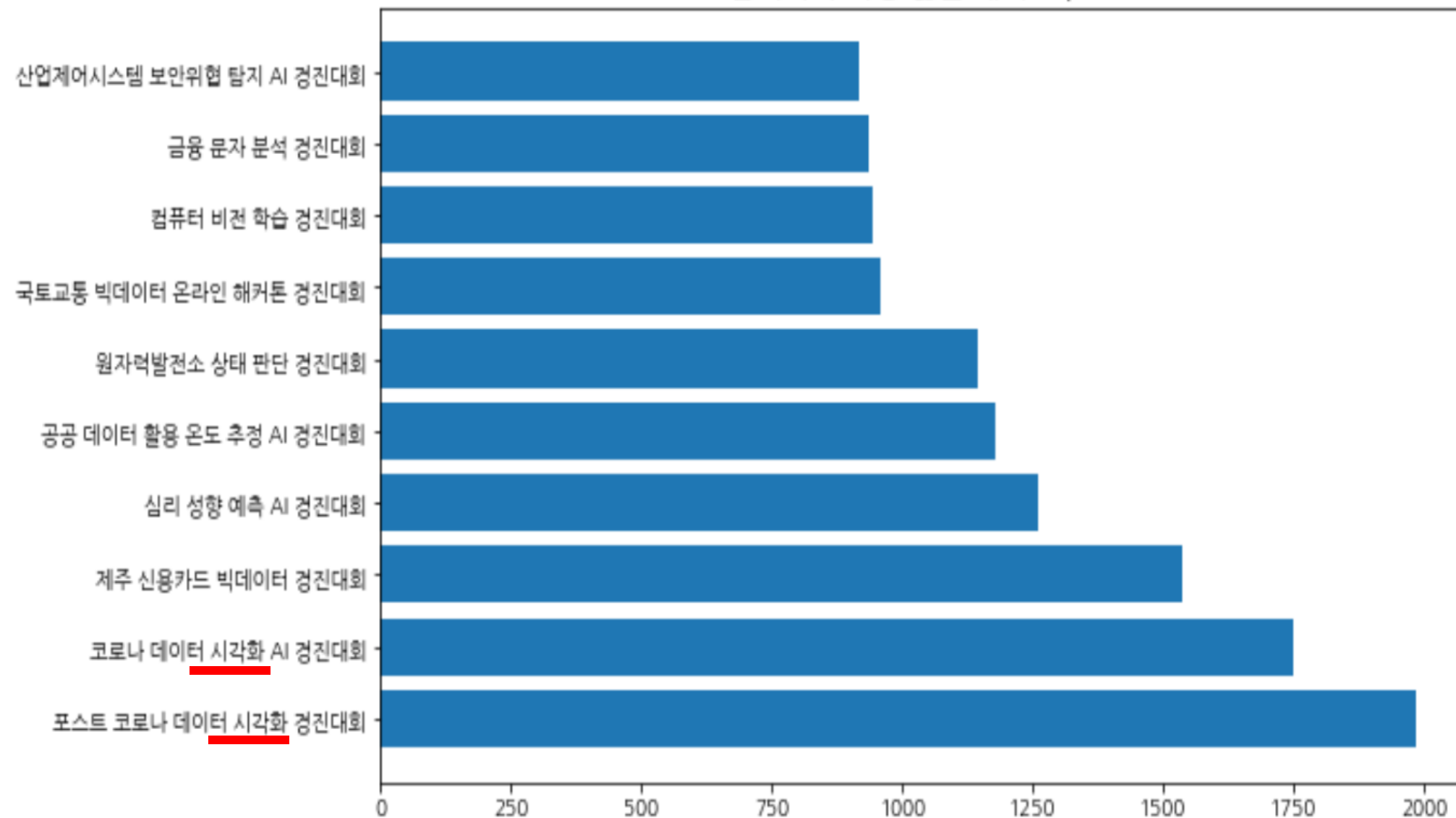
현재 진행 상황

EDA(Exploratory Data Analysis)

상금이 가장 많은 대회 top10



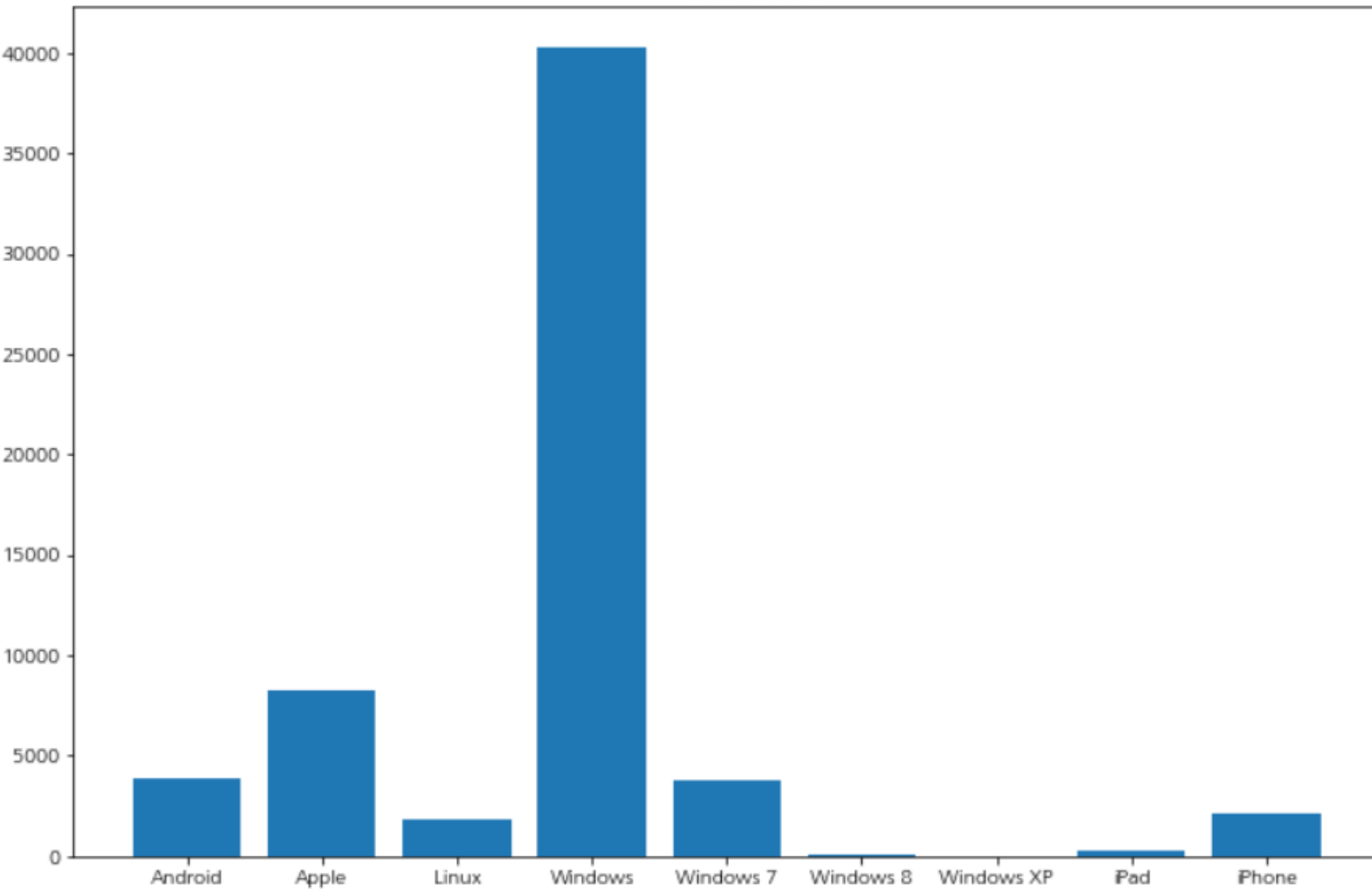
참여자가 가장 많은 대회 top10



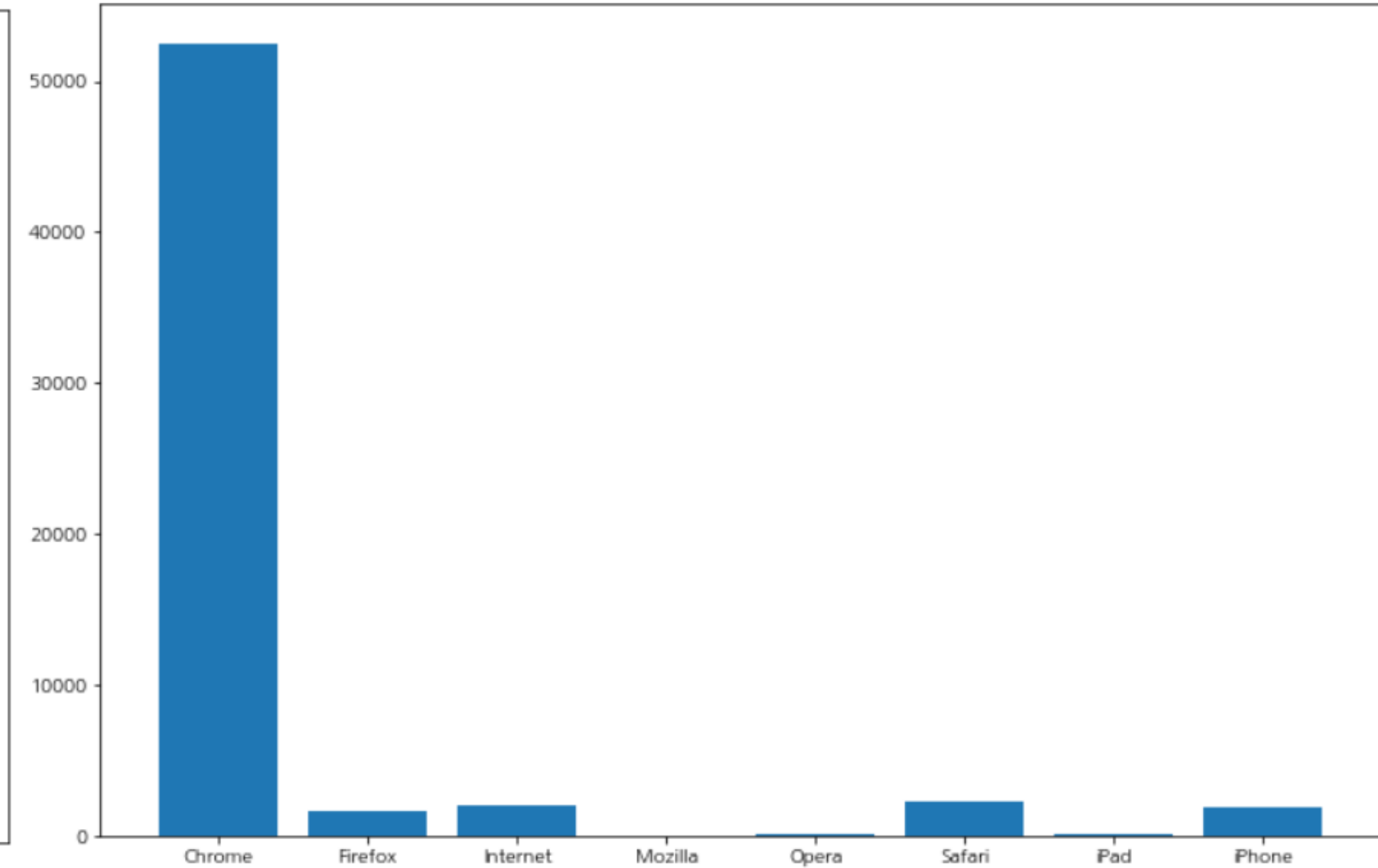
현재 진행 상황

EDA(Exploratory Data Analysis)

로그인한 플랫폼



로그인한 브라우저



최종 발표까지의 계획



감사합니다