



Team LimJung 2:
라이트 API를 사용하여 리그오브레전드
데이터 EDA 및 승리요인 분석

발표자

최수민, 허은민, 윤장원

발표일자

2024-06-21





Team_LimJung_2

역할



운장원



역할군



ETL, 시야 지표 EDA,
머신러닝 모델링



최수민



역할군



ETL, 메타 챔피언 EDA,
머신러닝 모델링



허은민



역할군



ETL, 카운터픽 EDA,
머신러닝 모델링



목차

I . Introduction

II . Data ETL

III. Data Analysis

IV. Result & Limitation



!

프로젝트 소개

- 리그오브레전드: 2008년 라이엇 게임즈가 개발한 멀티플레이어 온라인 배틀 아레나(MOBA) 게임으로, 5명이 한 팀이 되어 상대 넥서스를 파괴하여 승리하는 구조를 가짐.
- 게임의 인기: 2019년 동시 접속자 수 800만 명 돌파, 2021년 롤드컵 결승전 최대 시청자 수 7400만 명 기록.
- 목표: 승리 요인 분석: 천상계(마스터~챌린저) 매치 데이터를 활용하여 승리와 패배의 요인을 분석.
- 기대효과: 일반 플레이어들에게 게임 전략에 대한 중요한 인사이트 제공.

숫자로 본 2023년 롤드컵



The JoongAng



!

ETL 진행 방법

라이엇 OpenAPI에서 API Key 발급



LEAGUE-V4 API에서 summonerID 수집



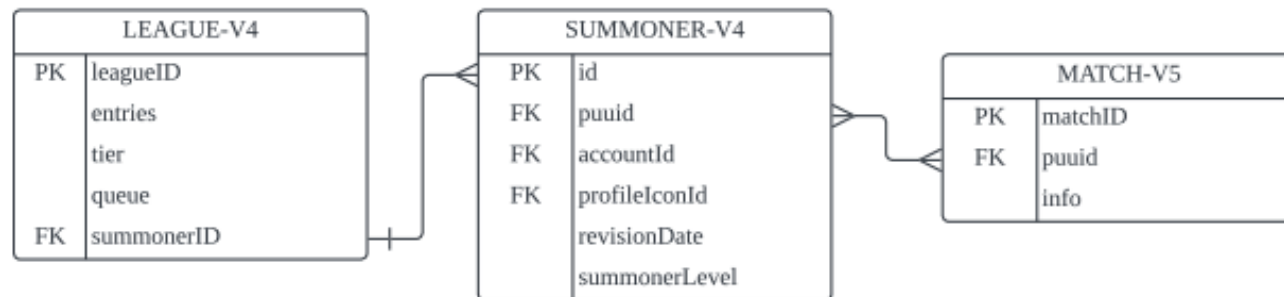
SUMMONER-V4 API에서 puuid 수집



MATCH-V5 API에서 matchId 수집



MATCH-V5 API에서 매치 데이터 수집





데이터 추출 (Extraction)

- 소환사 데이터 샘플링: v14.11 마스터~챌린저 큐에서 각 랭크마다 30명의 소환사를 랜덤 샘플링.
- 소환사 식별자 수집: `summonerId`를 사용하여 `puuid`와 `account_id`를 추출.
- 데이터프레임 생성: 수집된 `summonerId`, `puuid`, `account_id`를 포함하는 데이터프레임 생성.



데이터 변환 (Transformation)

- 매치 데이터 수집: 각 소환사의 `puuid`를 사용하여 최근 20개의 `matchId`를 추출, 중복 제거.
- 매치 데이터 조직: 추출된 `matchId`를 50개 단위로 그룹화하여 딕셔너리에 저장.
- 데이터 병합: 매치 상세 정보와 참가자 정보를 통합하여 데이터프레임으로 병합. 각 랭크별 데이터프레임을 최종 병합.



데이터 로드 (Load)

- CSV 파일 저장: 최종 병합된 데이터프레임을 CSV 파일로 저장하여 분석용 데이터셋 완성.



버전 패치



성능 좋은 신규
챔피언 메타 탄생



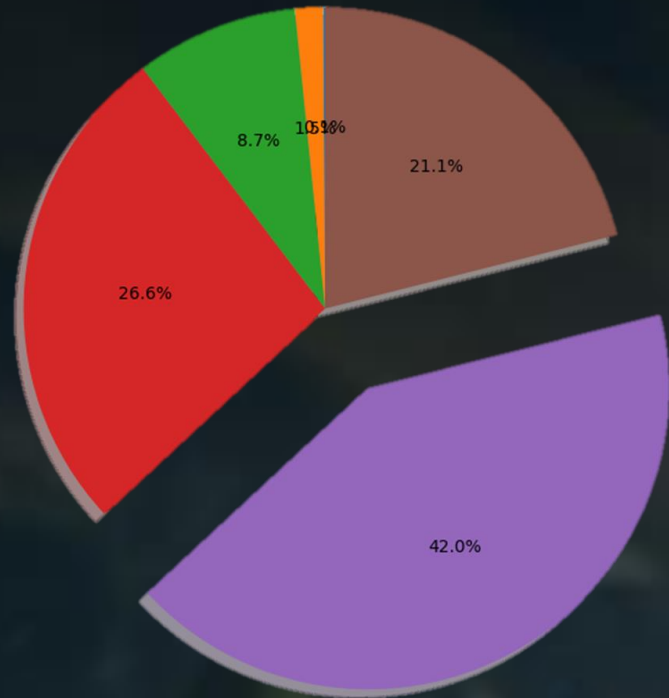
팀 승패 좌우



메타 챔피언 픽은 승리에 영향을 끼치는 요인일까?



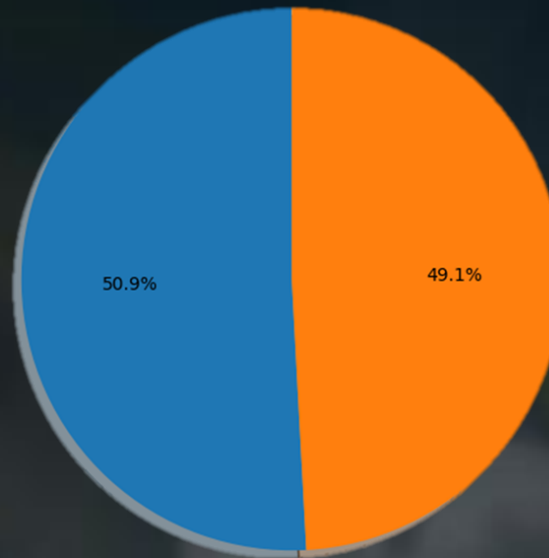
메타 챔피언을 픽한 팀의 수는?



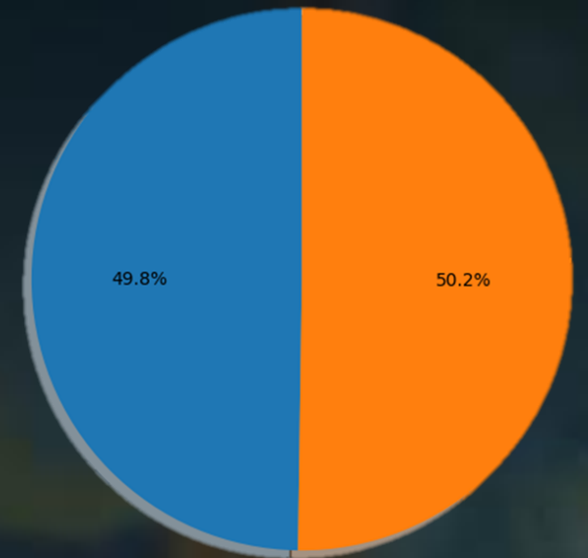
메타 챔피언 픽한 게임 약 80%

메타 챔피언을 픽한 팀과 픽하지 않은 팀의 승률

메타 챔피언 픽하지 않음



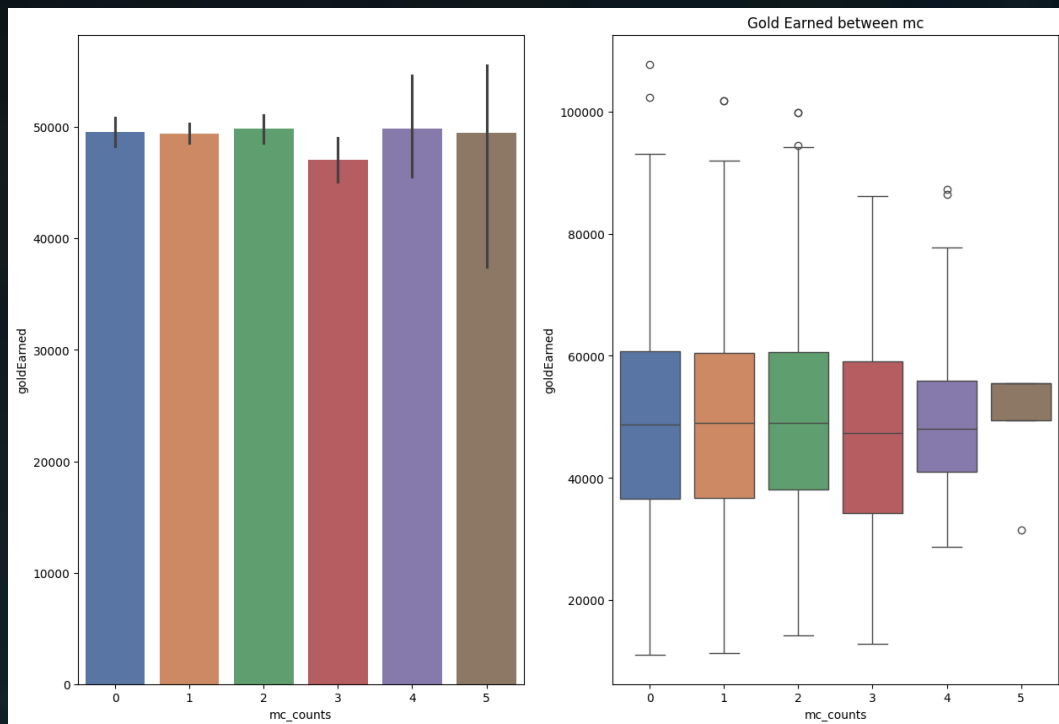
메타 챔피언 픽함



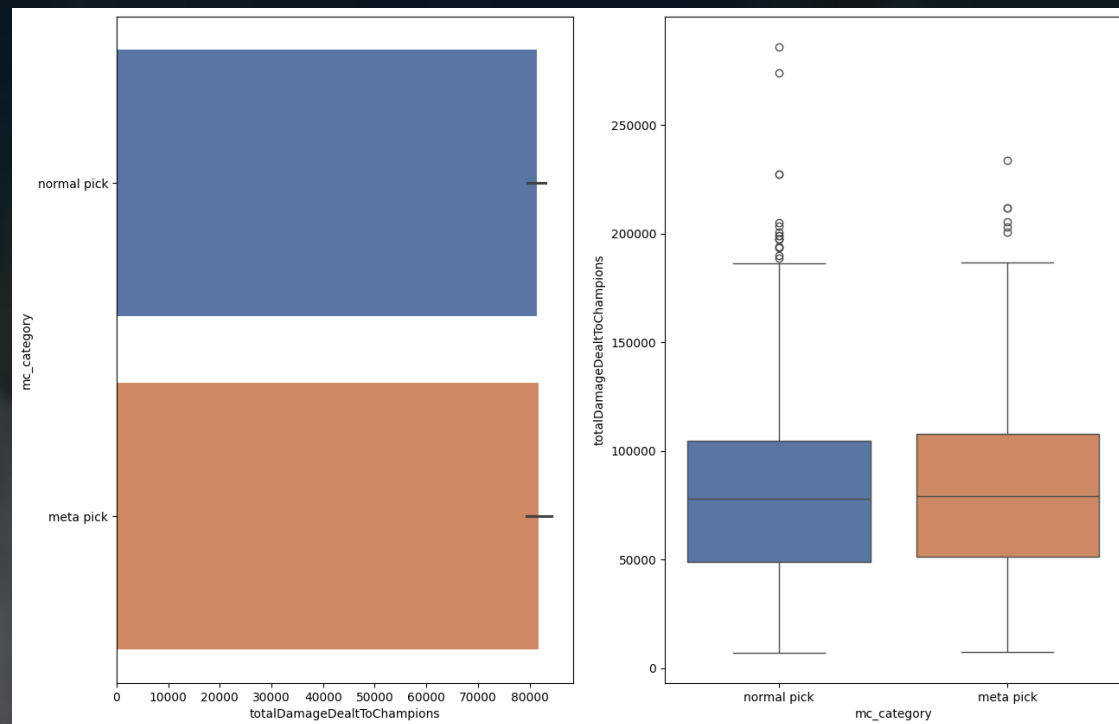
메타 챔피언 선택 -> 50.2% 승률
메타 챔피언 선택 x -> 49.1% 승률



메타 챔피언을 픽 수에 따른
골드 수급량



메타 챔피언 픽 유무에 따른 평균 데미지 차이

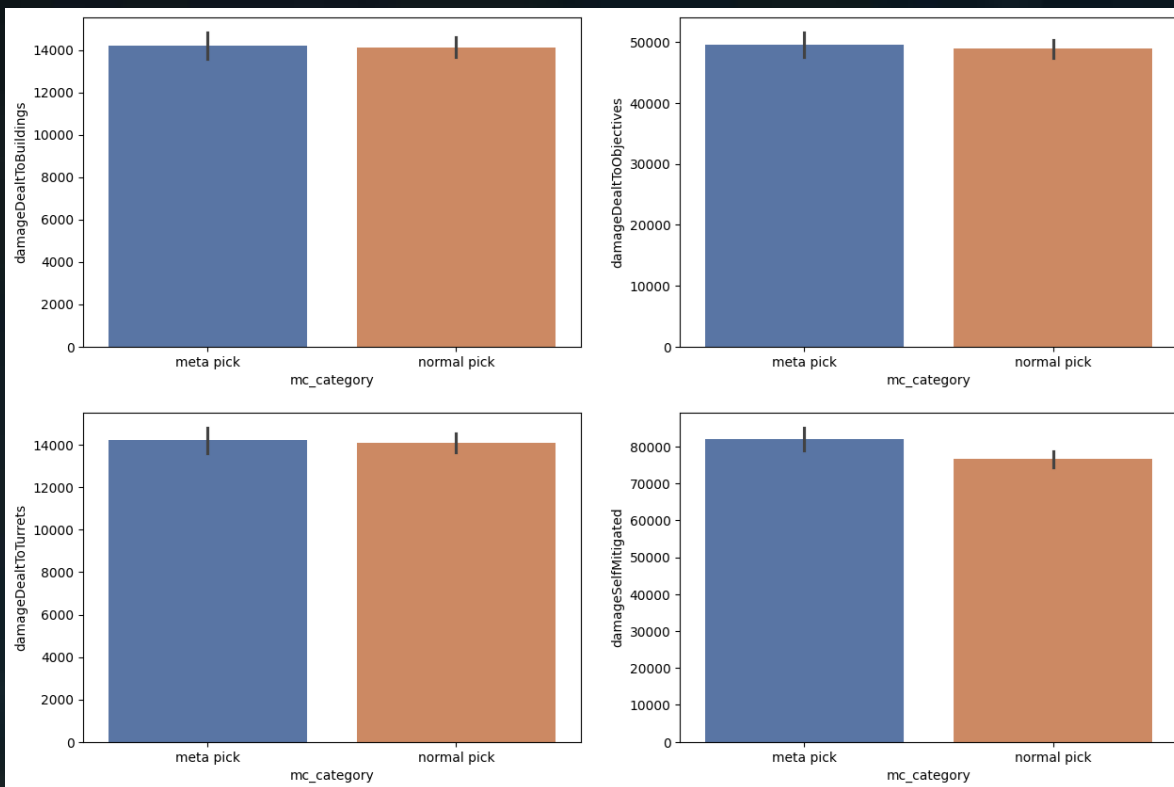


메타 챔피언 픽 수에 따른
골드 수급량 차이 X

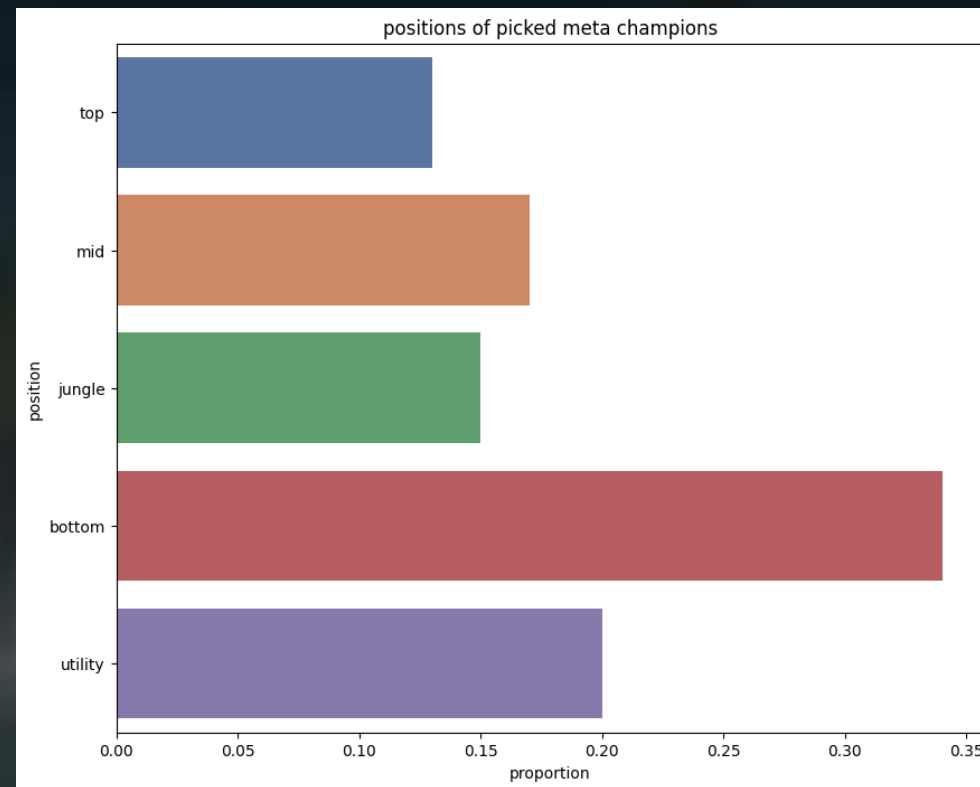
메타 챔피언 픽 유무에 따른
챔피언에게 가한 피해 차이 X



메타 챔피언에 따른
데미지 피해량/감소량



메타 챔피언 별
포지션 선택 빈도 수



가한 피해 차이 X
But, 받는 피해 감소량 차이 O

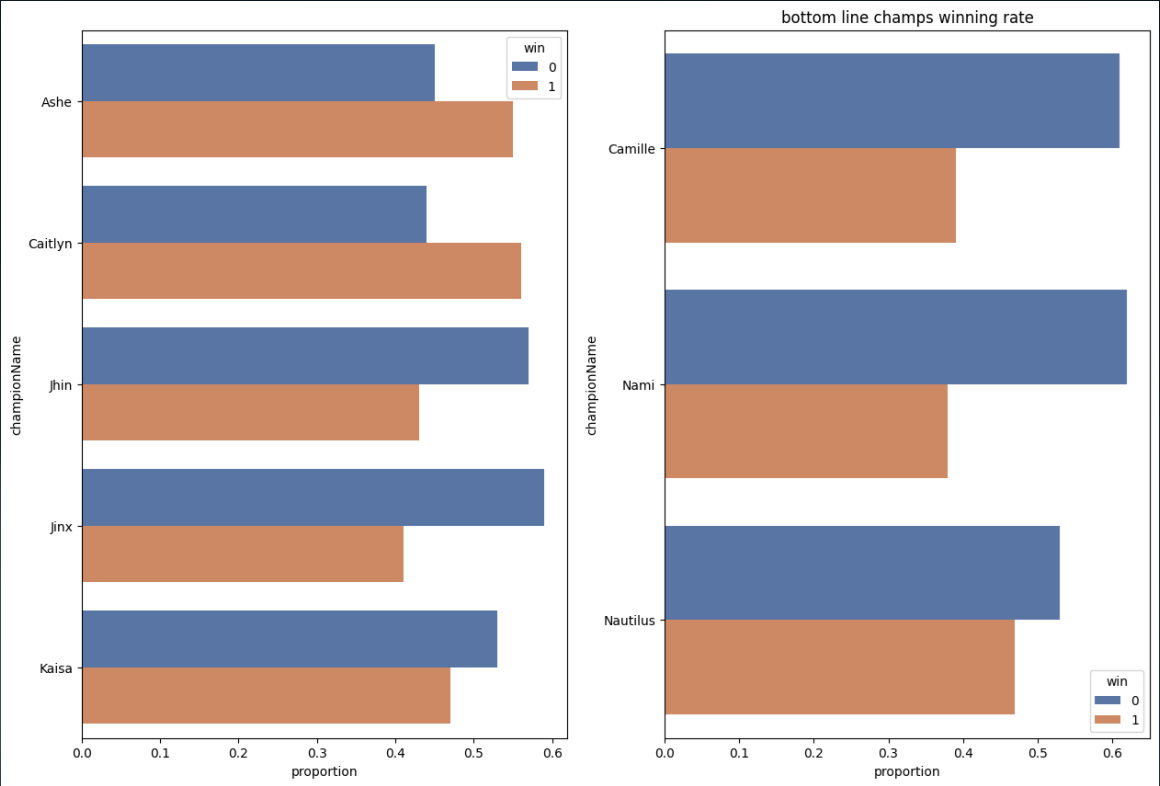
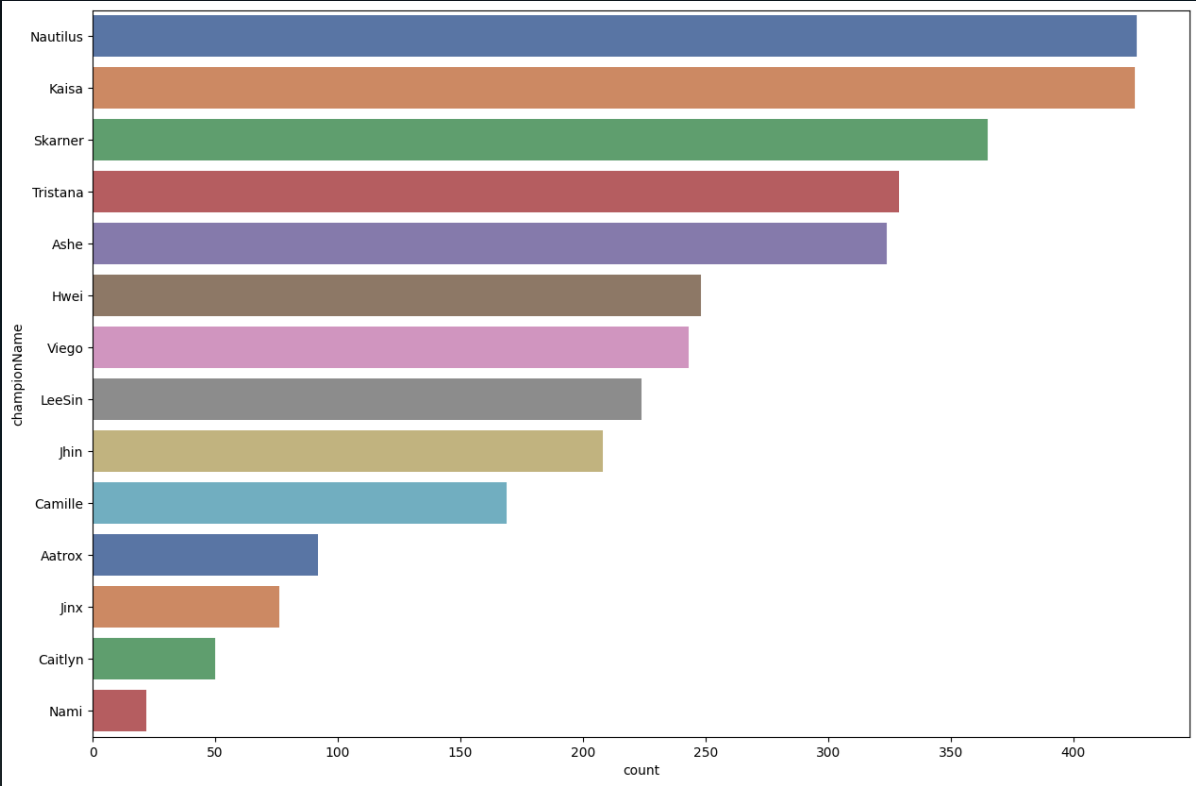


방어 특성 높은 챔프 픽 대세?
탑 포지션 픽 많지 않으려나?

Bottom, Utility 챔피언 픽이 많음.
현재 메타는 바텀 라인에서 메타 챔피언 픽

픽 수가 가장 많은 메타 챔피언은?

원딜/서폿 메타 챔피언의 승률



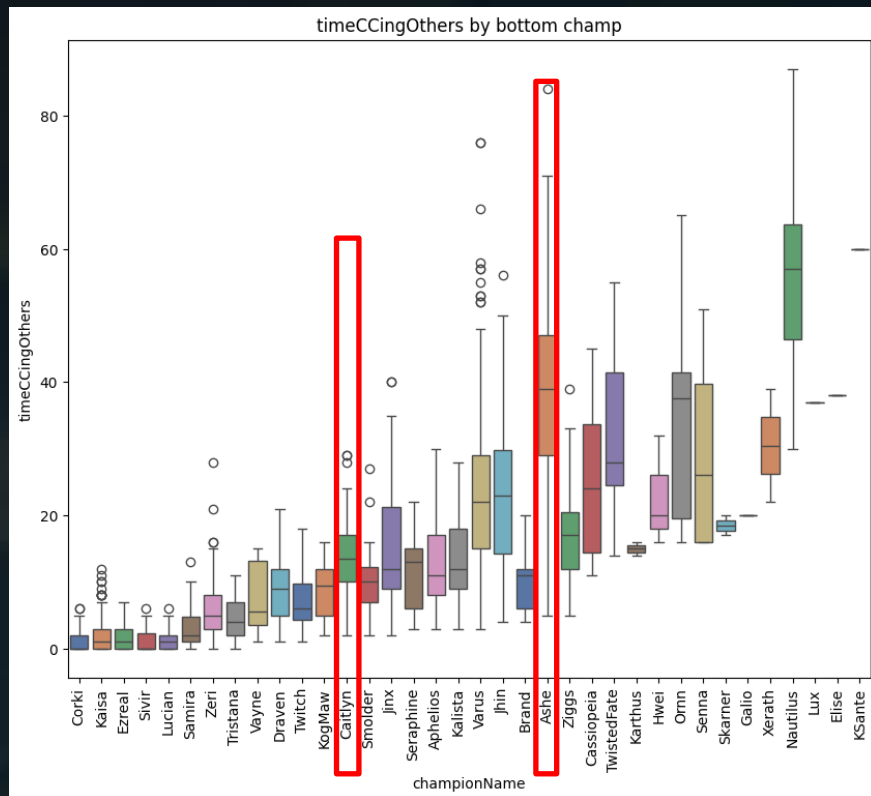
원딜/탱커형 챔피언 픽 많음 특히
원딜 픽이 굉장히 많음.

→ 탱커/원딜 챔피언이 많이 선택
그렇다면, 승률도 높은가?

bottom 라인에서 픽이 된 메타 챔피언 승률
애쉬, 케이틀린만 55% 이상 승률, 나머지는
평균 근사 or 패배가 더 많음

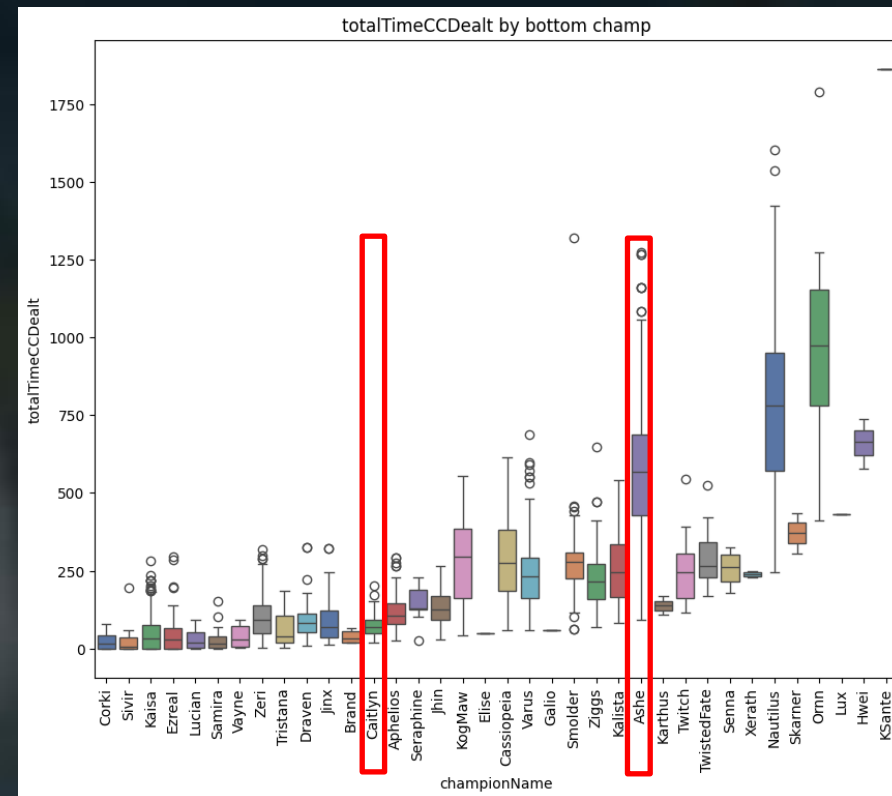


CC기를 입히는 평균 시간 플롯



애쉬 승률 높은 이유 (1)
원딜 챔피언 중에 CC기에 걸리게 하는
시간이 2번째로 높은 원딜

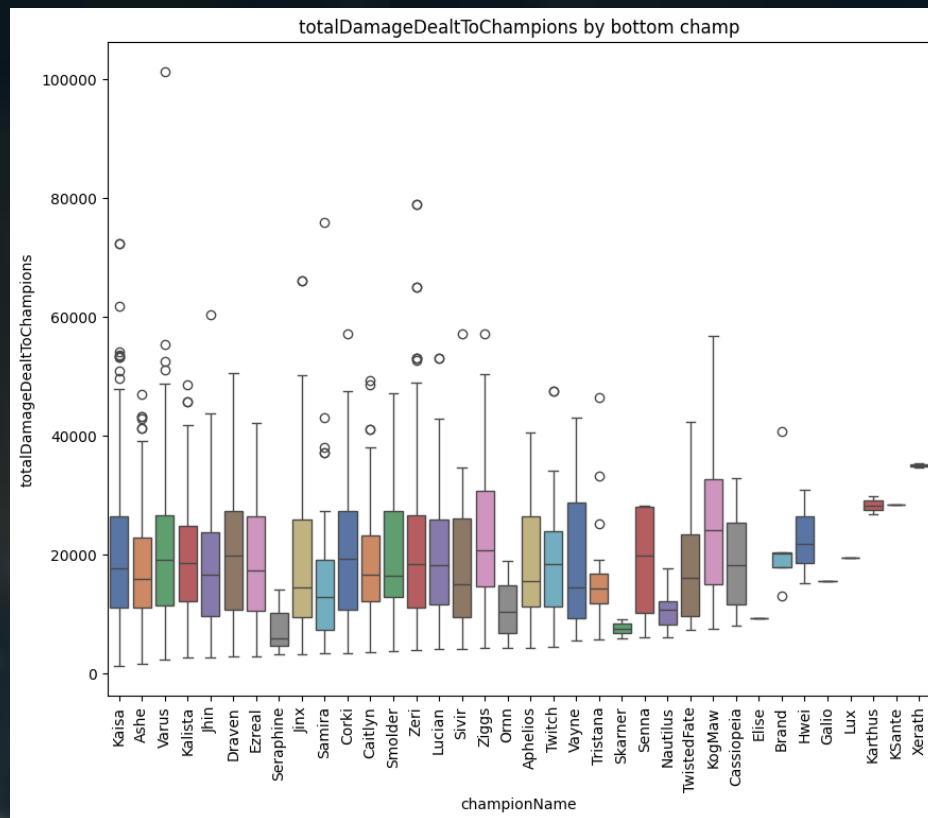
CC기를 동반한 피해량 평균을 나타낸 플롯



애쉬 승률이 높은 이유 (2)
원딜 챔피언 중 CC기 동반한
피해가 2번째로 높은 원딜



애쉬/케이틀린이 챔피언에게 가한 피해량



챔피언에게 가한 피해량에서는
애쉬/케이틀린이 높게 나타나지 않음

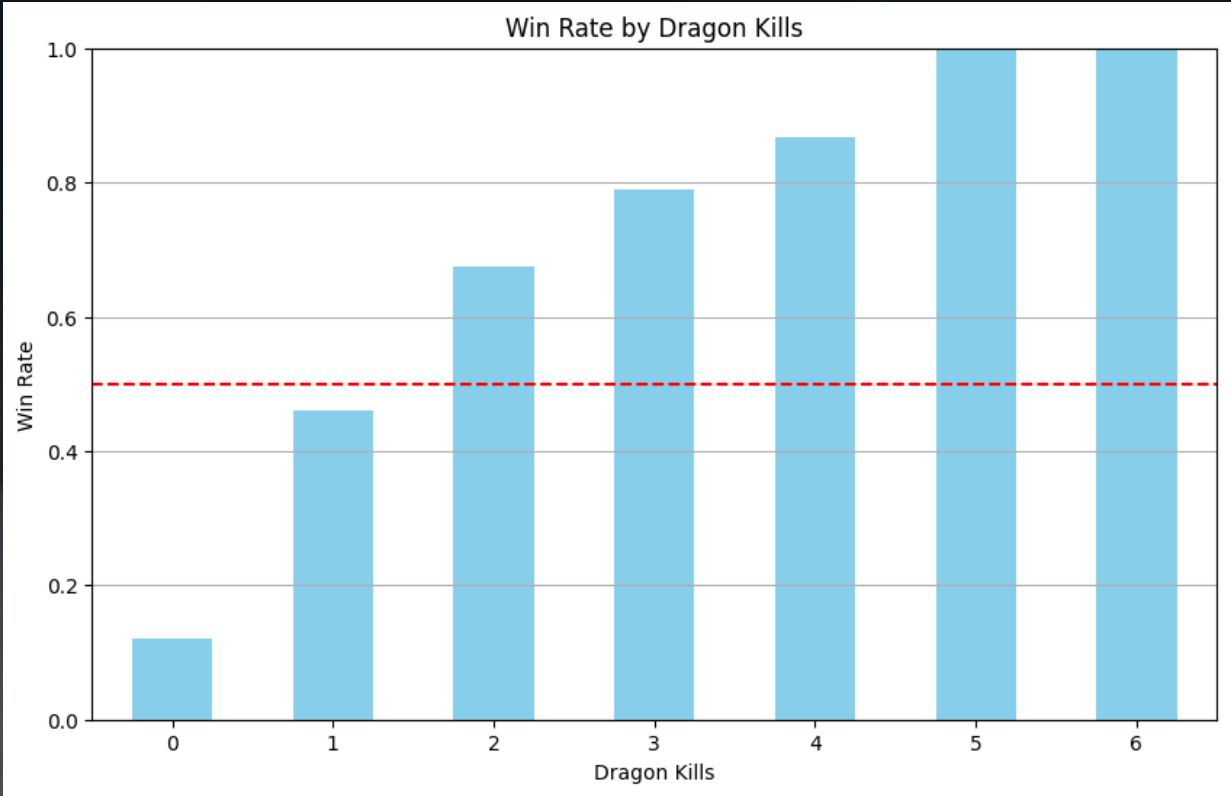
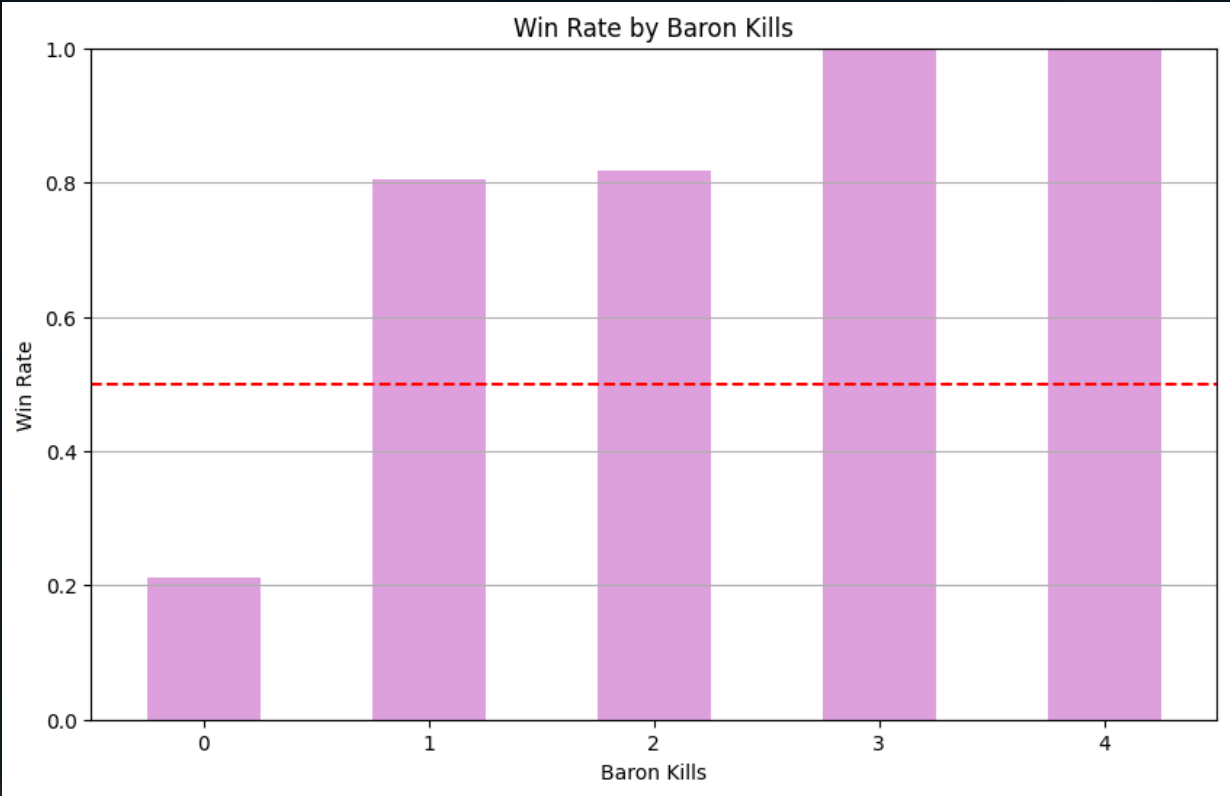


EDA 결론 및 정리

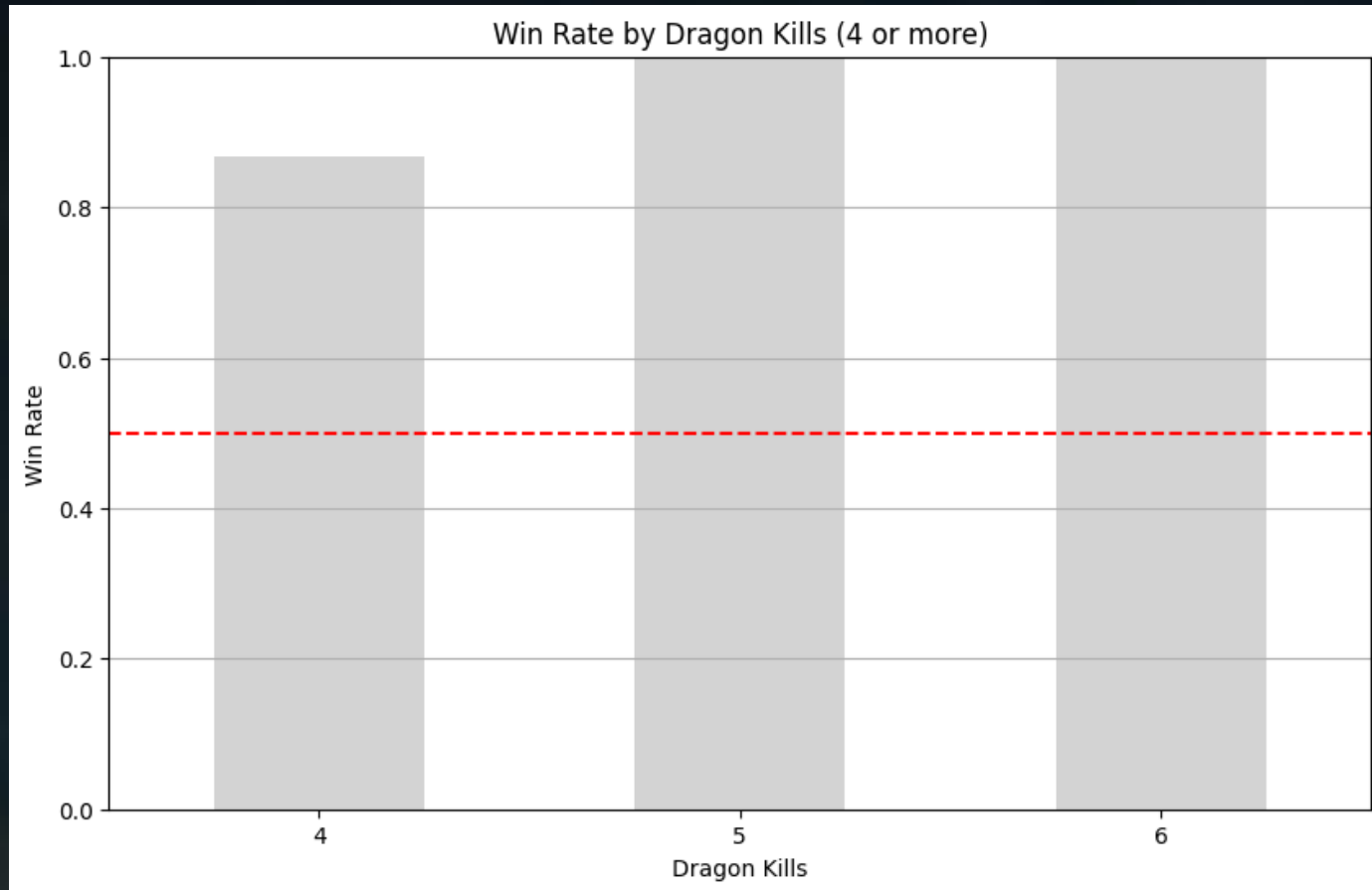
- 매치 챔피언을 픽하는 게임은 약 79%이며, 픽하지 않은 게임은 전체 21%이며 메타 챔피언을 픽한 팀의 승률이 픽하지 않은 팀보다 1.1%정도 더 높게 나타났다.
- 골드 수급량/입히는 피해가 메타 챔피언 픽 유무에 따라 차이가 없다. 그러나 데미지 감소/방어량 부분에서 차이가 존재하여 보았더니 원딜/탱커 계열의 챔피언이 선호되는 경향을 보였다.
- 탱커 계열의 챔피언(노틸러스, 카밀 등)은 서포터(유틸리티)로 많이 선호되고 있지만, 승률은 50% 근사 혹은 50% 이하의 승률을 보였으며, 원딜의 경우 케이틀린/애쉬와 같은 챔피언의 승률이 각각 56%/55%로 나타났다.
- 애쉬/케이틀린이 승률이 높은 이유를 1.사거리, 2. 강력한 CC기, 3. 딜량으로 가정하고 시각화하였고 애쉬의 경우는 강력한 CC기가 있어 승률이 높겠다라고 추정해볼 수 있지만 케이틀린의 경우 승률이 높은 이유를 해당 데이터를 통해 추정할 수 없었다.



오브젝트 획득 횟수는 승리에 영향을 끼치는 요인일까?



바론과 드래곤 모두 많이 획득할 수록 승률이 높아진다.



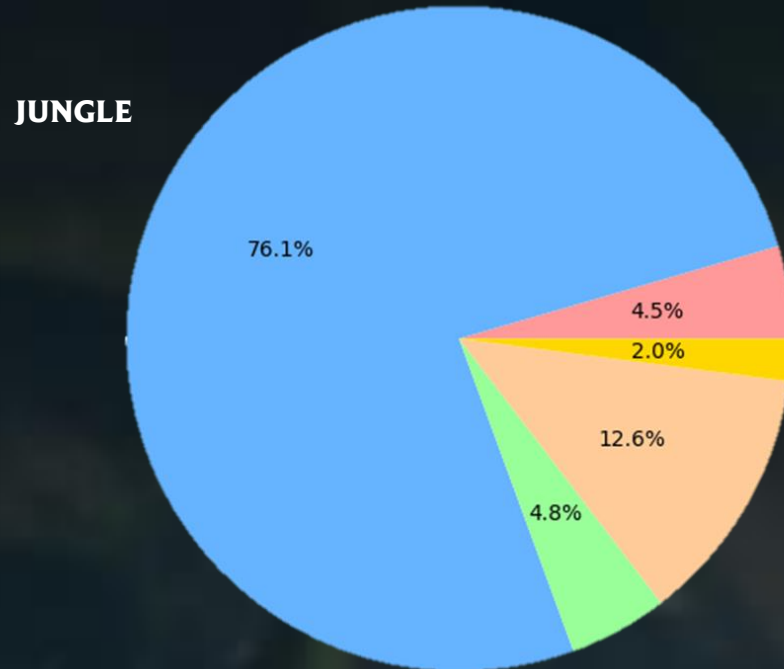
용의 영혼을 획득하거나 장로드래곤을 처치한
팀의 승률이 매우 높게 나왔다.

모든 오브젝트는 승률에
크게 영향을 미친다고 생각하였다.

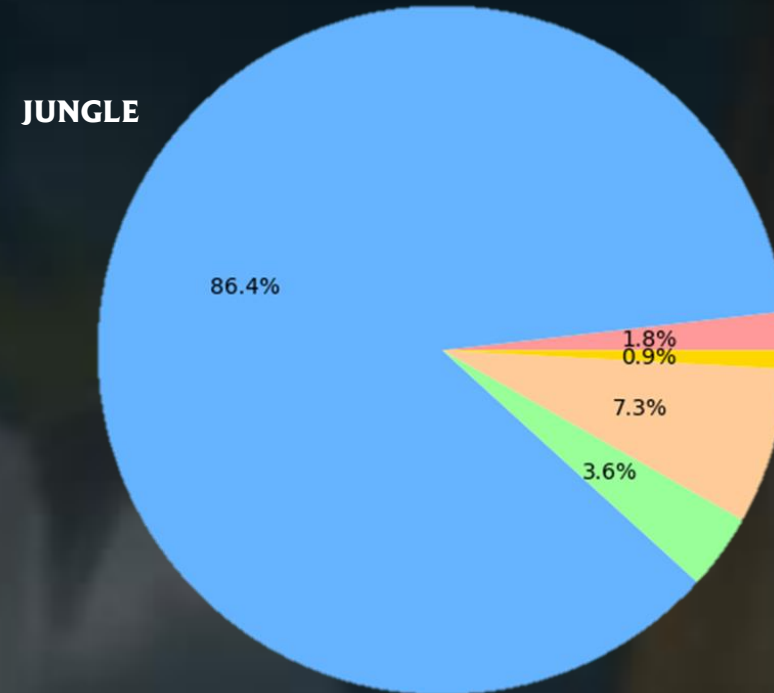
그렇다면 오브젝트를 많이 획득한
챔피언과 승률의 관계는 어떻게 될까?



Baron Kills by Position

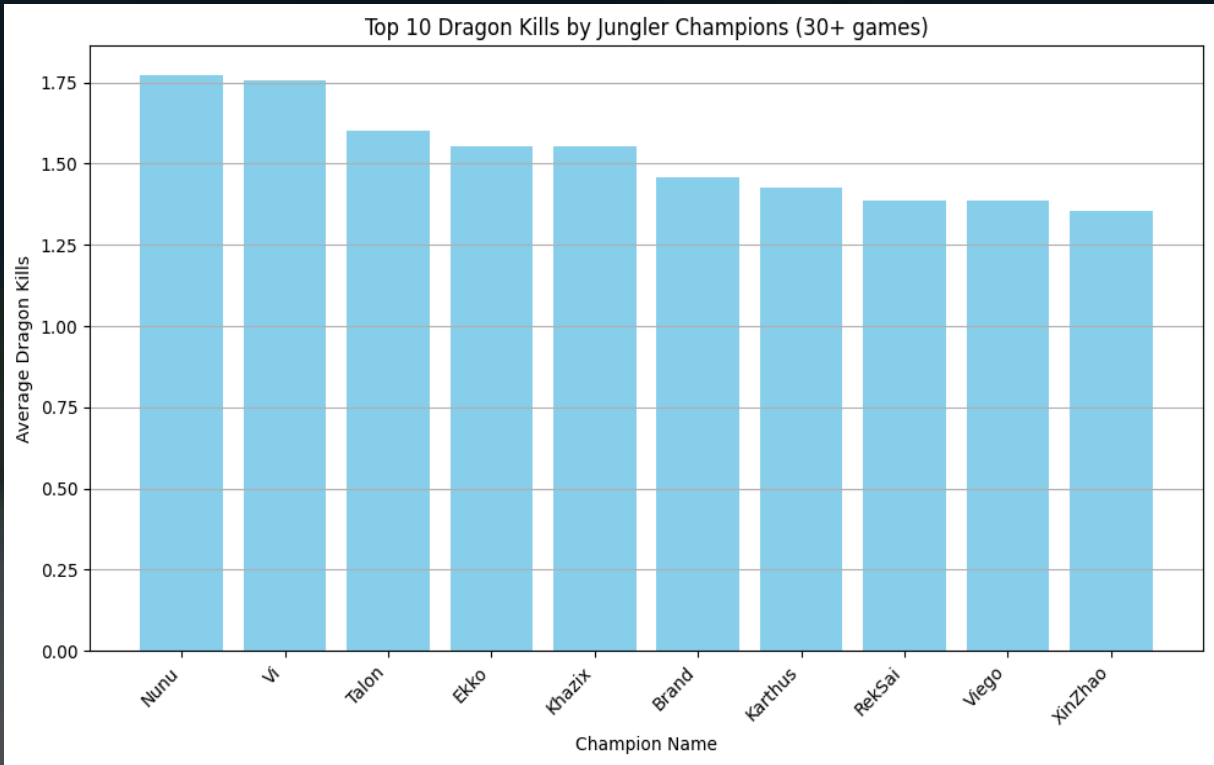
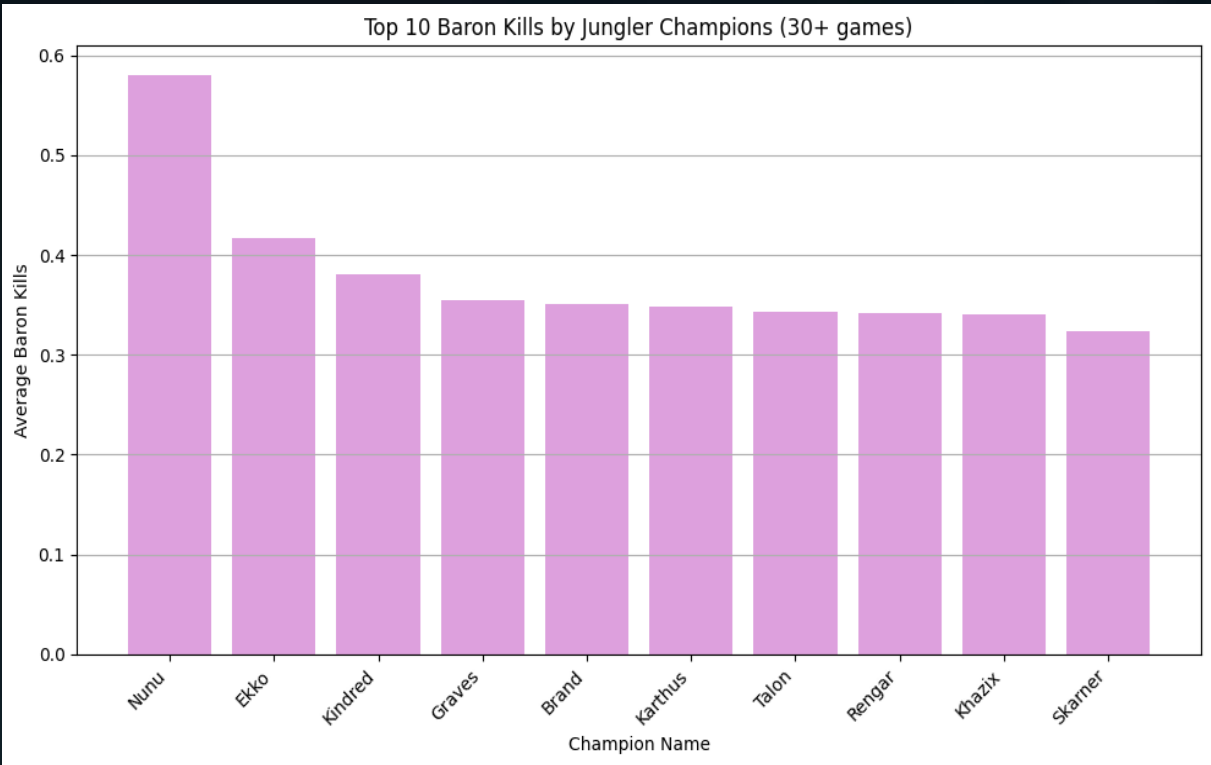


Dragon Kills by Position



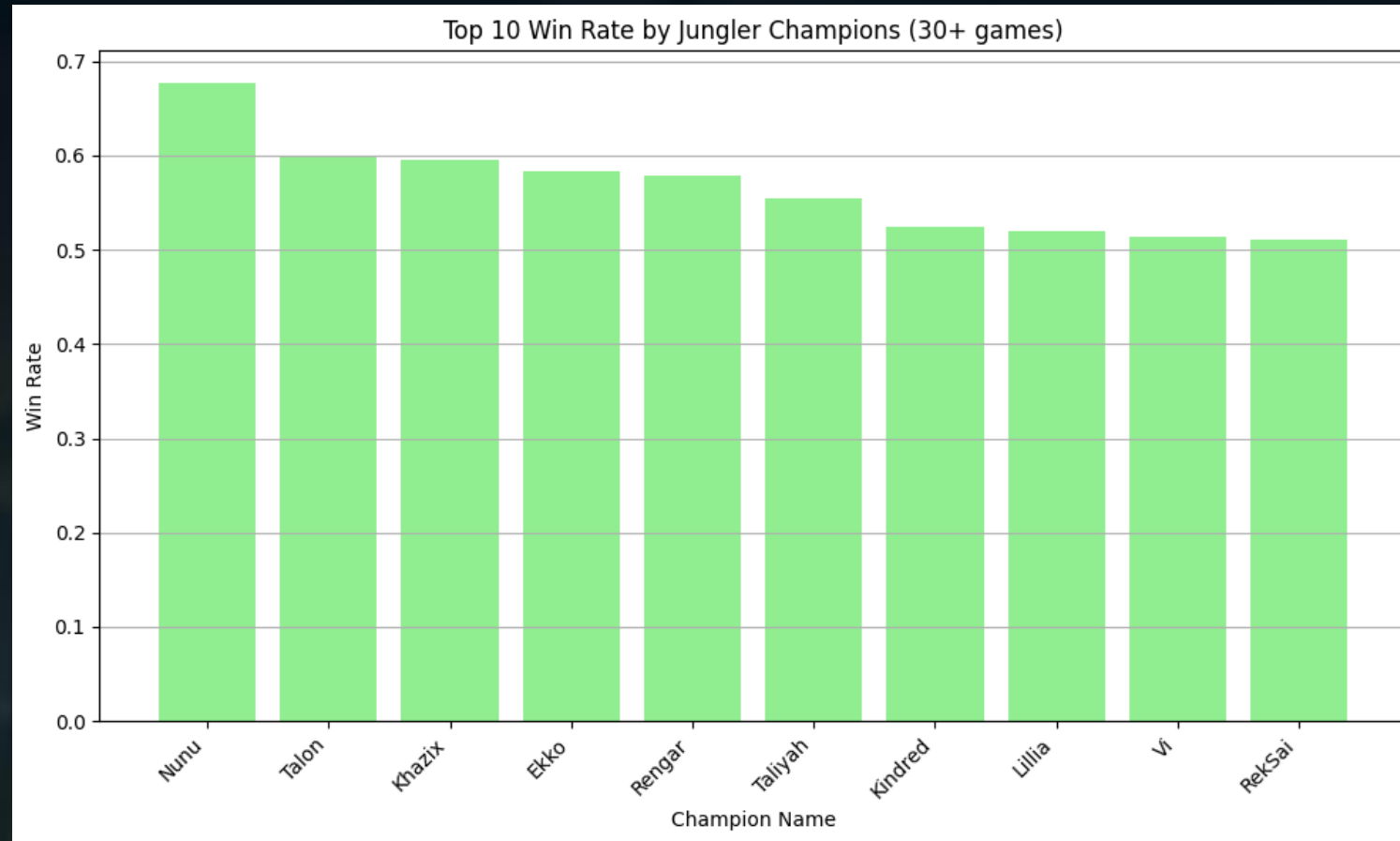
바론과 드래곤을 처치한 챔피언이
어느 포지션이 가장 많은지 살펴보았다.

게임 특성상 정글 몬스터에게 큰 피해를 줄 수 있는
JUNGLE 포지션이 가장 많이 처치하는 것으로 나왔다.



정글 포지션으로 사용된 챔피언 중 30게임 이상 플레이되었고 정글 포지션 선호도가 10% 이상인 챔피언을 대상으로 분석하였다.

바론과 드래곤을 많이 처치한 정글 챔피언 TOP10을 각각 알아보았다.



정글 챔피언 중 승률 TOP10인 챔피언들과 비교하였을 때
바론 처치 TOP10에서는 6명의 챔피언 중복,
드래곤 처치 TOP10에서는도 6명의 챔피언이 중복된 것으로 보아
오브젝트 획득 수는 승리에 영향을 준다는 것을 알게 되었다.



EDA 결론 및 정리

- 바론을 획득한 경우에는 승률이 80%를 넘어가며, 용을 2개 이상 획득한 경우 승률이 60% 이상 넘어간다.
- 용의 영혼을 획득한 경우 승률이 80%를 넘어가며, 장로드래곤을 획득한 경우 매우 높은 승률을 보인다.
- 정글 포지션이 다른 포지션보다 오브젝트 획득 확률이 높다.
- 정글 포지션 중 오브젝트 획득률이 높은 챔피언이 승률이 높다.



시야는 승리에 가장 중요한 요인일까?

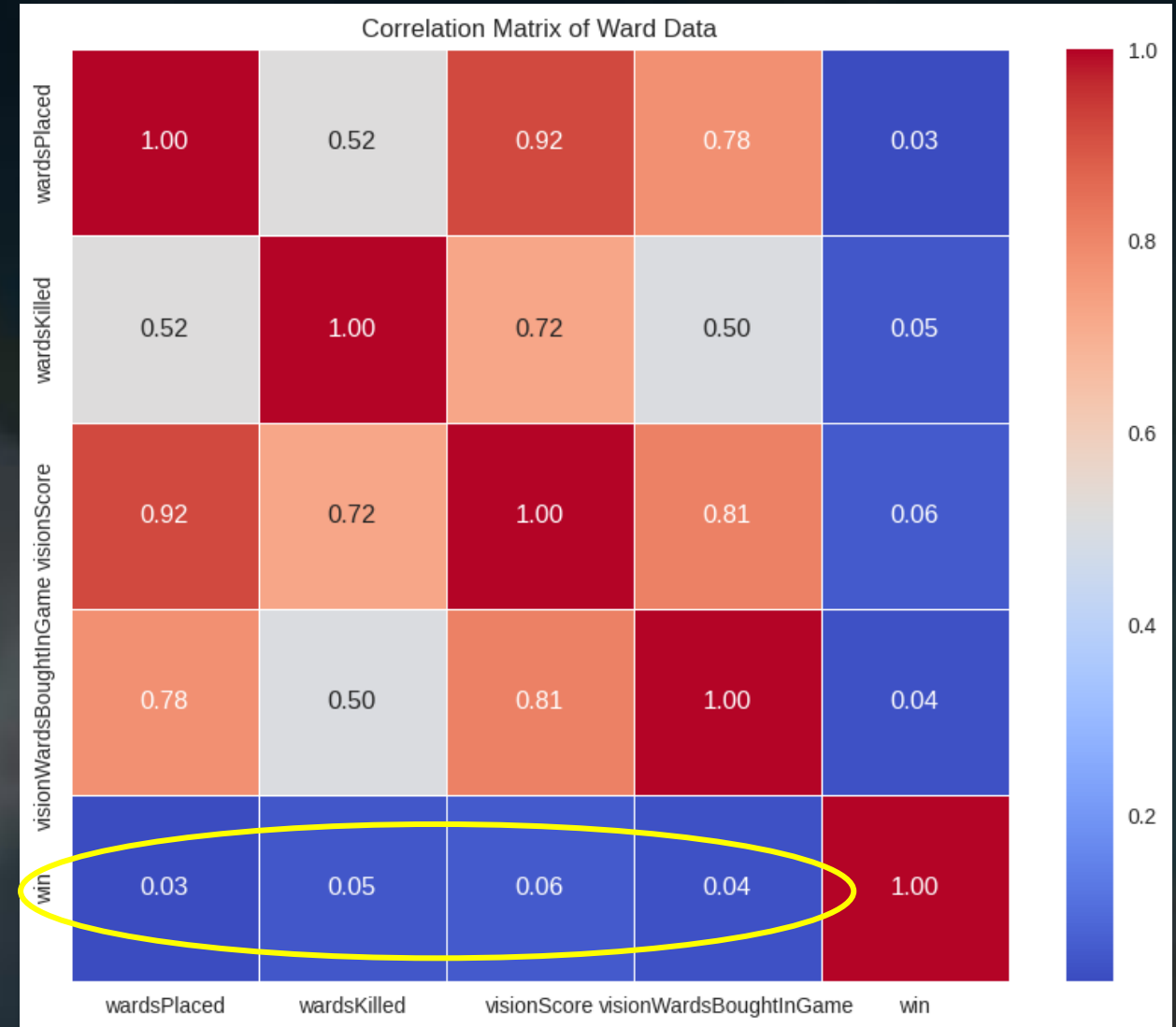


- 와드 관련 지표 & 승리 상관관계 계산
- 승리와 낮은 상관관계(0.03~0.06)

//

천상계에서 와드 관련 지표는 승리와 관계없다

//





그렇다면 어떤 요인이 승리와 가장 연관있을까?

데이터 전처리

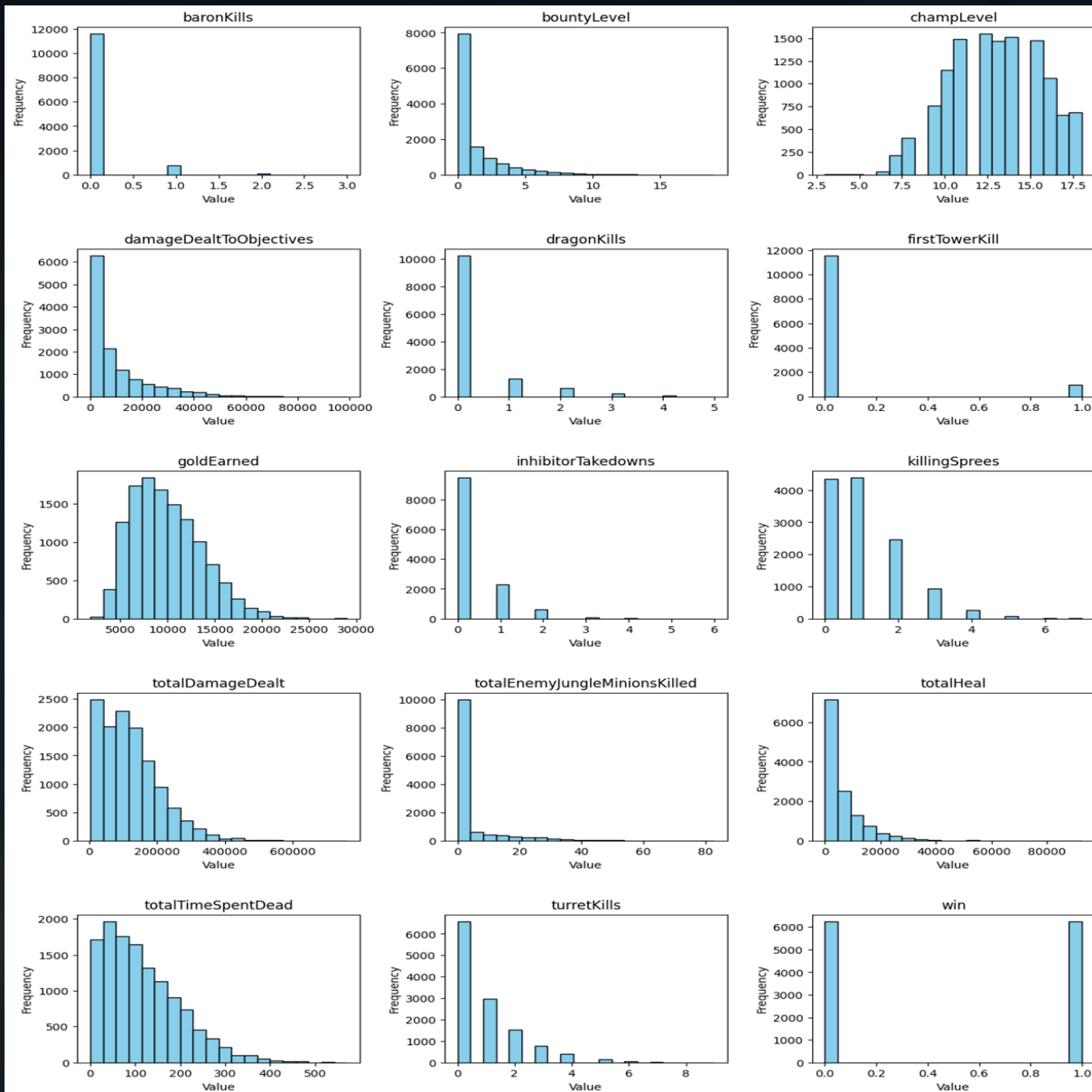
- **다중공선성 방지**

- 연관있어 보이는 변수 상관관계 시각화
- 상관관계 $0.7 \wedge$ 이면 1,
 $0.7 \vee$ 이면 0
- 관계있는 변수 중 1개 선택
- 도메인 지식 활용, 변수 선택

- 승리과 관계있는 변수 선택

- 상관계수 $0.1 \wedge$ 이면 선택, 아니면 제외

[illegible]



- 데이터 분포도 시각화
 - 대부분이 Right-Skewed
- 데이터 스케일링 필요
 - Standard Scaling 적용
 - 평균을 0, 분산을 1로 표준화



지표 설정

- 중요한 지표: F1-Score
 - 전반적인 모델 예측 성능이 중요하기 때문에!
 - Precision과 Recall을 잘 유지하는 모델 선택

Precision: 모델이 승리라고 예측한 경기들 중 실제로 승리한 경기의 비율

Recall: 모든 실제 승리 경기 중 모델이 승리로 예측한 경기의 비율

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

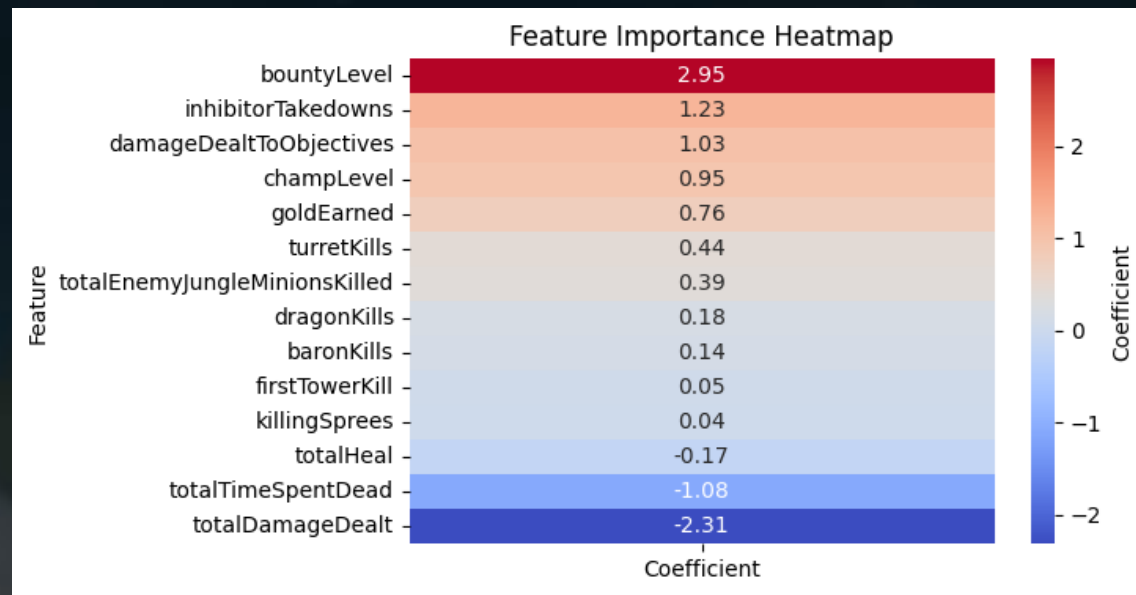
$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



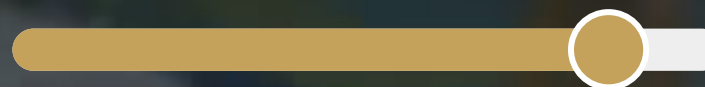
!

Logistic Regression

- F1-Score = 86% 달성
- 'bountyLevel' 이 2.95 계수
 - 1 증가시 승리 log(odds) 2.95 증가
- 'totalDamageDealt' 이 -2.31 계수
 - 1 증가시 승리 log(odds) 2.31 감소
- 다중공선성 존재, 따라서 Coefficient 신뢰불가

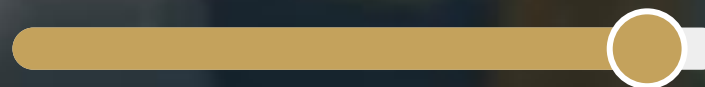


Accuracy



86%

Precision



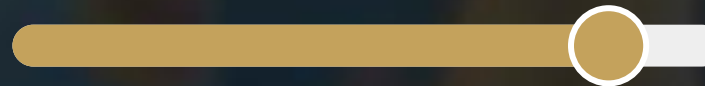
91%

Recall



82%

F1-Score



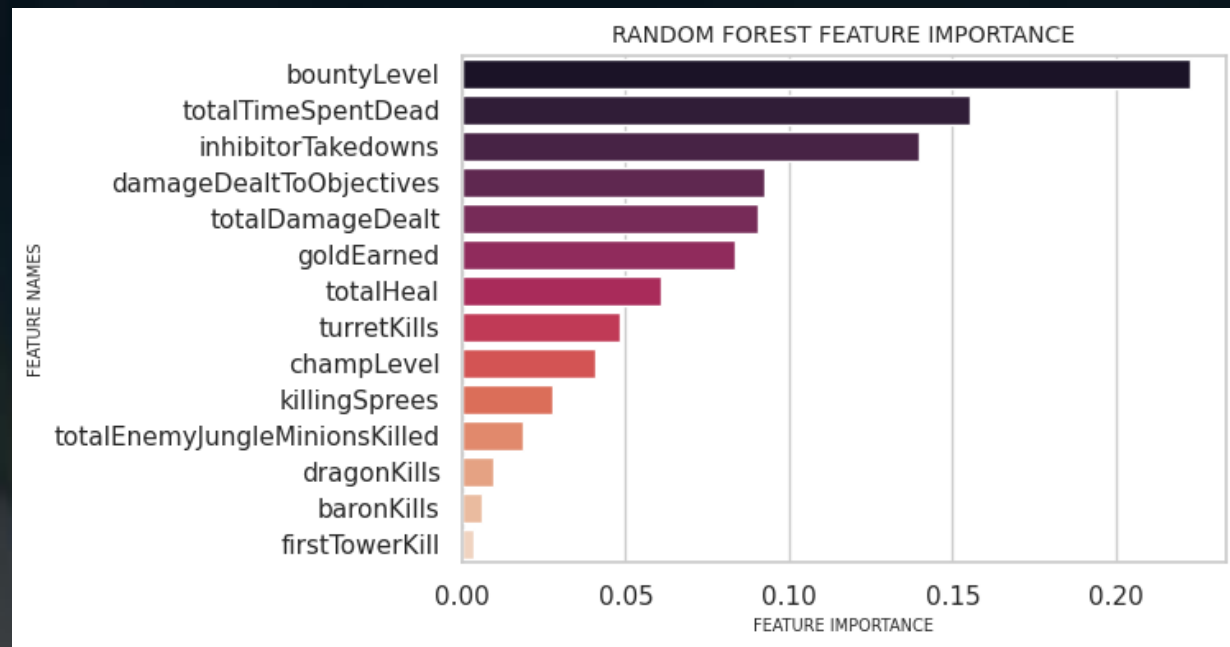
86%



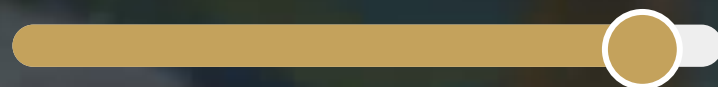
!

Random Forest

- F1-Score = 90% 달성
- 'bountyLevel' 이 가장 중요한 요인으로 도출
- 다른 중요한 요인:
 - totalTimeSpentDead
 - inhibitorTakedowns
 - damageDealtToObjectives,
 - totalDamageDealt
 - goldEarned

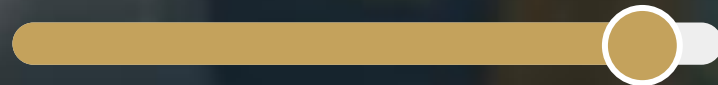


Accuracy



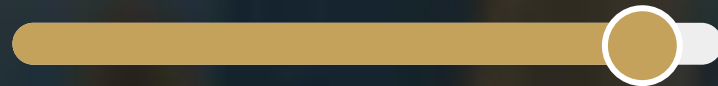
90%

Precision



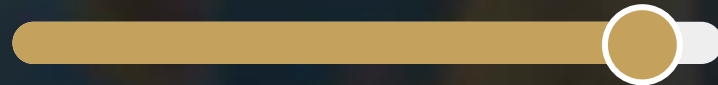
90%

Recall



90%

F1-Score



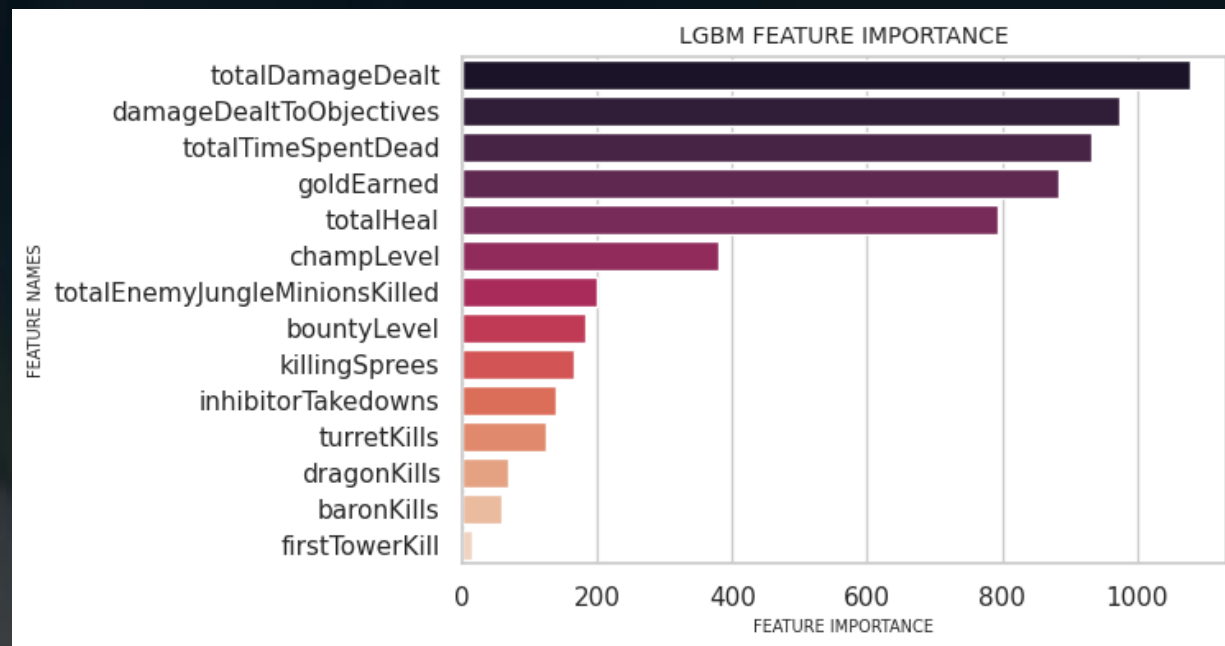
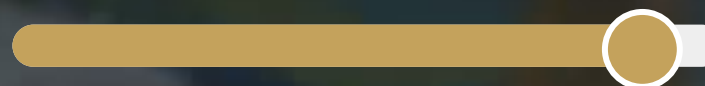
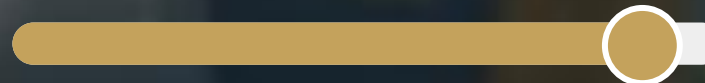
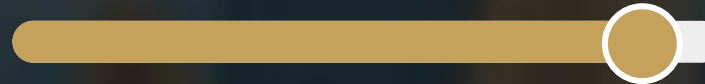
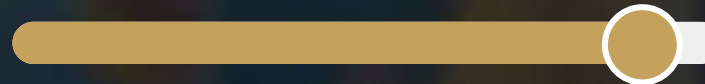
90%



!

LGBM Classifier

- F1-Score = 90% 달성
- 'totalDamageDealt' 가 가장 중요한 요인으로
도출
- 다른 중요한 요인:
 - damageDealtToObjectives
 - totalTimeSpentDead
 - goldEarned
 - totalHeal
 - champLevel

**Accuracy****90%****Precision****90%****Recall****90%****F1-Score****90%**



그렇다면 실제 데이터에선 어떨까?

챌린저와 마스터가 둘 다 있을 수 있는
그랜드 마스터 큐 솔랭 데이터로 실험해보자

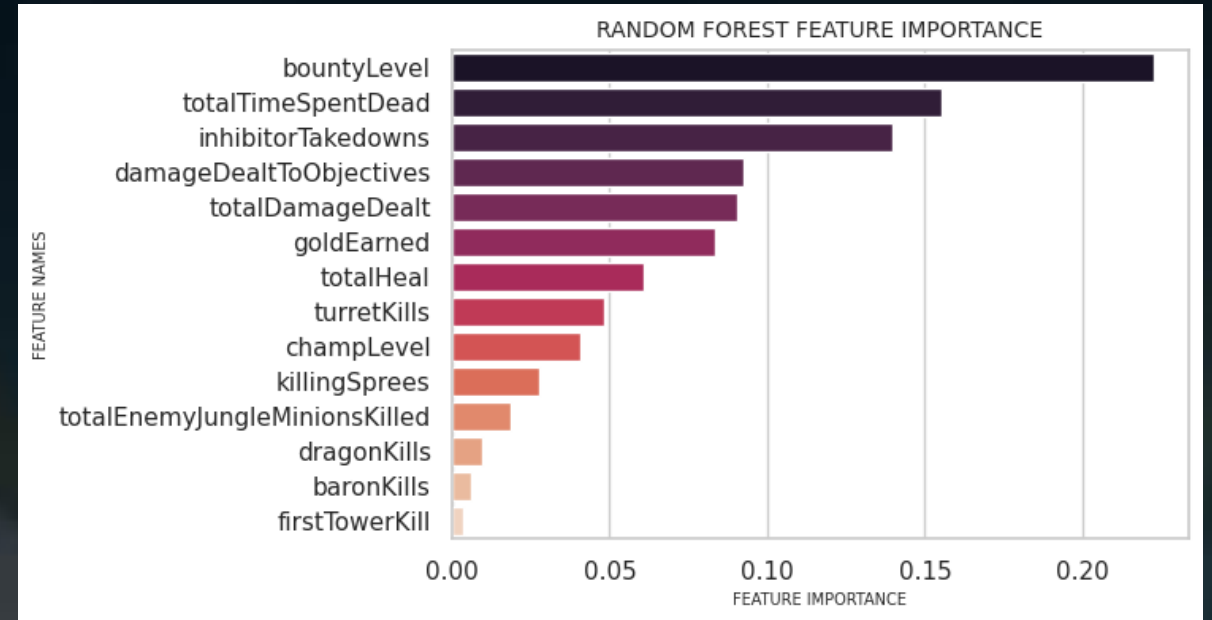
사용할 모델: Random Forest & LGBM



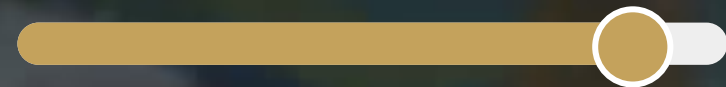
!

Random Forest

- F1-Score = 88% 달성
- 'bountyLevel' 이 가장 중요한 요인으로 도출
- 다른 중요한 요인:
 - totalTimeSpentDead
 - inhibitorTakedowns
 - damageDealtToObjectives,
 - totalDamageDealt
 - goldEarned

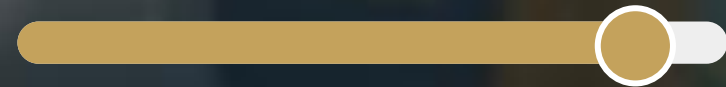


Accuracy



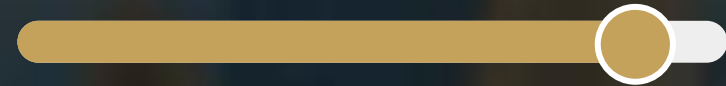
88%

Precision



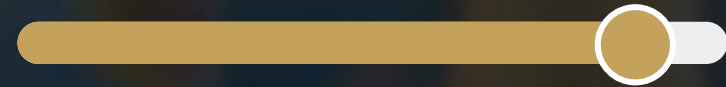
88%

Recall



88%

F1-Score



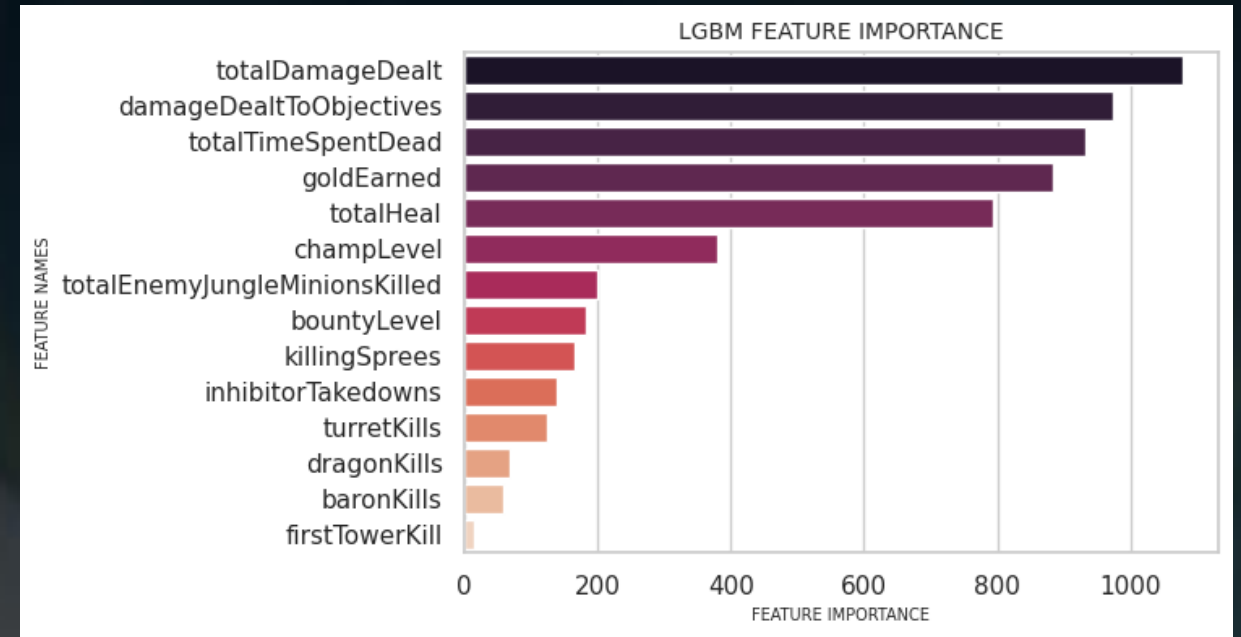
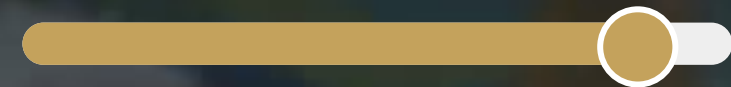
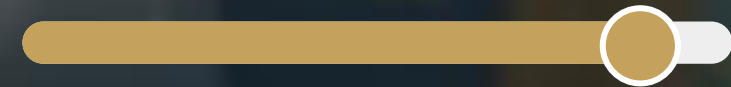
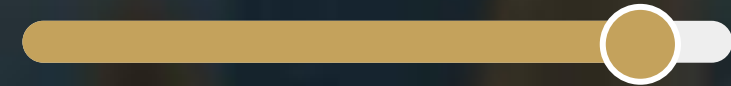
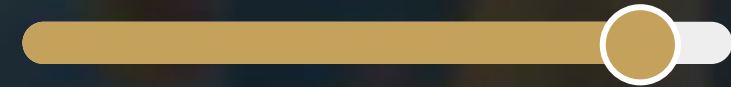
88%



!

LGBM Classifier

- F1-Score = 88% 달성
- 'totalDamageDealt' 가 가장 중요한 요인으로
도출
- 다른 중요한 요인:
 - damageDealtToObjectives
 - totalTimeSpentDead
 - goldEarned
 - totalHeal
 - champLevel

**Accuracy****88%****Precision****88%****Recall****88%****F1-Score****88%**



모델 평가 요약

- 높은 성능 모델: Random Forest와 LGBM
- 평가 지표 F1-score: 0.9
- 실제 데이터 예측 정확도: 약 0.88

공통적으로 중요한 요인

- **totalTimeSpentDead**: 플레이어의 생존 시간이 짧을수록 승리 확률이 높음
- **damageDealtToObjectives**: 목표물에 대한 피해가 많을수록 승리에 크게 기여
- **totalDamageDealt**: 게임에서 입힌 전체 피해량
- **goldEarned**: 획득한 골드량이 높을수록 승리에 기여

특정 상황에서 중요한 피쳐

- **bountyLevel**: 보상 레벨이 높을수록 플레이어의 성과에 영향
- **champLevel**: 챔피언 레벨이 높을수록 경기에 유리
- **totalHeal**: 치유량이 많을수록 게임 내에서 유리한 위치 유지 가능



결론

승리의 핵심 요인

- 생존 시간과 목표물 피해: 게임 진행에 직접적인 영향을 미침
- 승리에 중대한 영향을 미치는 결정적 요소:
 - **totalTimeSpentDead** (총 죽은 시간)
 - **damageDealtToObjectives** (오브젝트에 가한 피해량)

전략적 권장 사항:

효율적인 오브젝트 관리:

오브젝트에 대한 피해 증대
시간 관리 및 자원 배분

생존성 강화:

사망 시간 최소화
안전한 플레이 유도

자원 획득의 최적화:

골드 획득에 집중
아이템 구매 및 업그레이드 강화



한계점



데이터의 한계

데이터 샘플링과 선택 편향

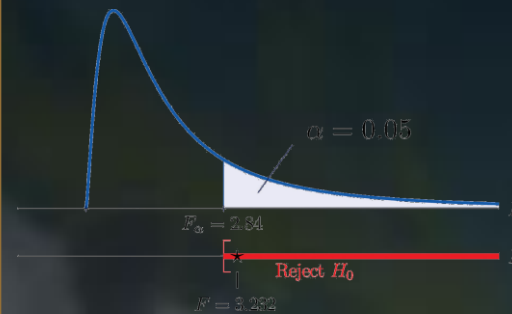
데이터 양과 질



모델의 한계

모델의 일반화 문제

피쳐 중요도의 해석과 측정편향



분석 방법의 한계

통계적 검증과 변수 처리



외부요인의 무시

팀 상호작용과 전략

게임 패치와 업데이트



출처

- Riot Games. (2024). League of Legends Developer API. Retrieved June 10, 2024, from <https://developer.riotgames.com/>
- 고봉준. (2023, November 20). 중앙일보. <https://www.joongang.co.kr/article/25208471#home>
- OpenAI. (2024). Interaction with ChatGPT. Accessed on June 18, 2024, via OpenAI's ChatGPT platform.
- 챌린저 마크, 티모, 노틸러스, 트리스타나, 애쉬 일러스트 PNGWings