**API 이용하여 데이터 추출하는 전반적인 과정에 대한 대략적 분석**

우리가 원하는 조합 데이터는 게임의 상세 정보 데이터 속에 있으므로 각 게임의 상세 정보 데이터를 가져와야함

1. **‘LEAGUE-V4’ API를 이용하여 각 티어(등급)에 속해있는 모든 유저들의 명단(summonerId)을 가져온다.**

티어 정보: (높음) Challenger – Grandmaster – Master – Diamond1 – Diamond2 – Diamond3 – Diamond4 – Platinum1 - … - Platinum4 – Gold1 - … - Gold4 – Silver1 - … Silver 4 – Bronze1 - … Bronze4 – Iron1 - … Iron4 (낮음)

2월 13일 기준, 챌린저는 현재 약 150명, 그랜드마스터는 약 340명, 마스터는 약 900명, 다이아1은 약 3000명, 다이아2는 약 4700명 정도가 있음. 현재까지 상위 0.4%에 해당하는 대략 9000명 정도(챌린저부터 다이아2까지)의 명단을 추출 완료함.

C: 1 Request, GM: 1 Request, M: 1 Request, D1: 15 Requests, D2: 23 Requests

Total: 41 Requests (C~D2), About 9,000 Data

1. **‘SUMMONER-V4’ API를 이용하여 각 summonerId에 해당하는 유저 고유의 accountId를 불러와 매칭시킨다.**

∵ MATCH-V4 API에 접근하기 위해서는 오직 summonerId로는 불가능하며, 이 summonerId가 encrypted된 accountId가 필요하기 때문. accountId는 SUMMONER-V4에서 제공하기 때문에 2번의 과정을 거쳐야 함.

Total: About 9,000 Requests (C~D2), About 9,000 Data

1. **‘MATCH-V4’ API 중 ‘matchlists’를 이용하여 accountId로 각 유저가 플레이한 최근 100 경기의 고유한 게임 Id 값을 가져온다(gameId).**

Matchlists를 이용하여 기본적으로 최근 100경기만을 가져올 수 있음(한번의 요청으로 가져올 수 있는 경기의 수가 최대 100개).

@논의필요: 물론 QUERY PARAMETERS 중 begin index와 end index를 이용하여 그 이전의 경기도 불러올 수 있으나 유저마다 플레이한 게임의 수가 제각각이라는 점 ㅡ 어떤 유저는 1년 내내 50판도 안하기도 하고 또다른 유저는 1달에 300~400판 이상을 하기도 한다는점 ㅡ 때문에 사실상 힘들다고 여겨짐. (올바르지 않은 parameter로 접근할 시 400: Bad request 뜸)

이러한 문제점 때문에 적은 사람의 모든 경기를 가져오는 것 보다 최대한 많은 사람의 경기를 100경기씩 가져오는 것이 효율적일 것이라고 생각됨.(필요하다면 저티어 유저들의 데이터들까지). 이 방식대로 한다면 우리가 수집하게 될 game 데이터의 수는 ‘(수집한 유저 데이터의 수) \* 100’ 이 될 것임. (현재 랭크게임을 플레이하는 유저는 대략 2,500,000명) 그런데 이는 분석을 진행하기에 충분한 data인지…?

Total: About 9,000 Requests (C~D2), About 900,000 Data

1. **‘MATCH-V4’ API 중 ‘matches’를 이용하여 각 gameId로 해당 게임의 상세 내역을 가져온다. (우리가 분석하고자 하는 raw 데이터)**

전처리 전 데이터 추출 단계 완료

Total: About 900,000 Requests (C~D2), About 900,000 Data

Grand Total: About 920,000 Requests, About 900,000 Data

1 Requests에 넉넉잡아 1.4초가 소요됨. 1,260,000sec (900,000 Data)

Data 추출 속도 : (하나의 API key로 하나의 Device에서) 한 달에 대략 2,000,000개의 Data.

하나의 Data에 2개의 조합(아군의 조합과 적군의 조합)을 얻어낼 수 있기 때문에 조합만 따지면 4,000,000개의 데이터를 얻을 수 있으나 중복되는 데이터들에 대한 고려 필요할 듯(한 게임 내 10명은 하나의 같은 게임을 공유하므로).

주요 논제 : 확실히 request limit 때문에 물리적으로 구할 수 있는 데이터의 수는 처음과 다르게 많이 한정적임 -> 별로 중요하지 않은 조합들(outliers)을 처리해서 가능한 조합의 수를 줄이거나 자체적인 알고리즘을 통해 추천 방식을 변경해야 할수도…?

1. **이긴 팀과 진 팀을 분리하여 승리에 영향을 주는 요인을 분석**

**테이블이(가) 표시된 사진

자동 생성된 설명**

샘플데이터만으로 간단히 보았을때

대체적으로 이긴팀의 통계가 평균적으로 높게 나온다

1. 먼저 firstblood(가장 먼저 킬 을 낸 경우 이 경우 100원의 추가 골드를 얻는다)

이 경우 대략5.8 대 4.2 정도로 생각보다 퍼블은 게임의 큰 영향을 끼치지 못하는 것 같다.

1. 하지만 first tower에 경우 7대 3 정도로 게임의 많은 영향을 미친다.

아무래도 상대 챔피언을 죽여야 타워를 칠 수 있을 뿐 아니라 타워 골드+ 첫포탑 골드까지 해서 상대와의 성장격차를 낼 수 있는 수단이기 때문일 것이다.

1. firstbaron(게임내의 존재하는 에픽 몬스터 20분이후에 나타나며 사냥의 성공할 시 골드와 경험치 몬스터 버프를 얻을 수 있음)

생각보다 첫바론의 경우 5대5로 게임의 영향을 미치지 않는다는 것을 볼수있다.

1. firstdragon(게임 내의 존재하는 에픽 몬스터 바론과 다른 점은 드래곤의 경우 골드와 경험치는 적지만 지속적으로 버프를 얻을 수 있음)

드래곤의 경우 6 대 4 정도로 바론과 드래곤의 경우 선택을 해야 한다면 게임 승리를 위해서는 드래곤이 조금 더 중요해 보인다.

1. riftHeraldKills(게임 내의 존재하는 에픽 몬스터로 사냥의 성공할 시 타워 체력을 상당 부분 깎아주는 몬스터이다)

이 경우 타워의 피를 상당 부분 깎아주기에 first tower랑 연관되어 있을 거라 생각하였지만 거의 5 대 5로서 생각보다 큰 영향을 미치지 않는다