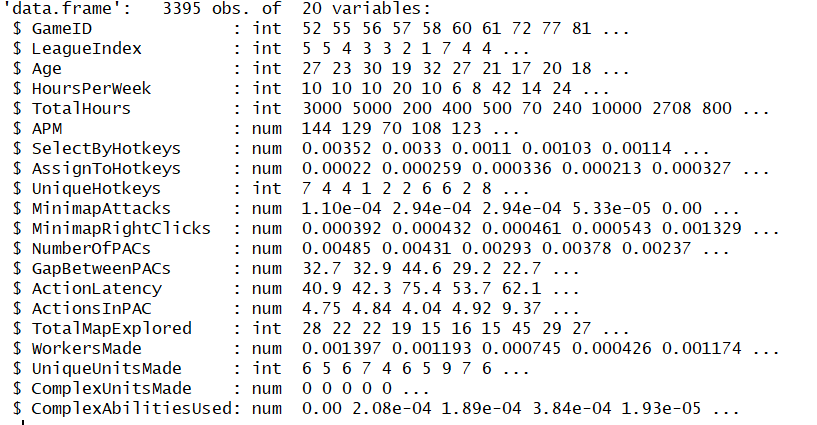
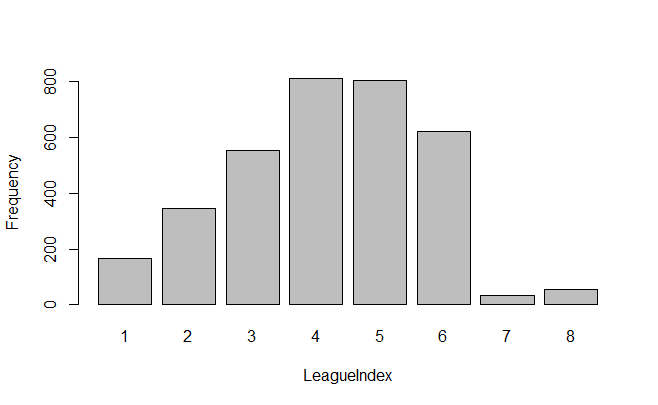
1. **Skillcraft Data**의 등급별 탐색적 자료 분석
2. 데이터 구성



Skillcraft data set은 총 3395개의 관측치와 20개의 변수로 이루어져 있음을 확인할 수 있다.

8개의 값이 integer의 형태로 인식이 되었고, 나머지 12개의 변수들은 numeric 형태로 인식이 되었다. Integer로 인식된 변수 중 ‘GameID’와 ‘LeagueIndex’변수의 경우 숫자가 가지는 값이 숫자로서의 의미가 있는 것이 아니라 하나의 범주라고 인식할 수 있으므로 범주형 형태로 변수의 형태를 바꿔줘야 할 필요성이 있어 보인다.



Group을 나눠주는 변수인 LeagueIndex에 따른 데이터의 분포를 살펴본 결과는 위와 같다. 가장 낮은 등급인 Bronze가 1의 값을 가지고 가장 높은 등급인 Professional이 8의 값을 가지는데, 위의 그래프를 살펴보면 중간 등급인 Gold~Master(3~6) 등급에 대다수의 사용자가 분포하고 있는 것을 알 수 있다.

1. 변수에 따른 등급별 기초통계량 및 Plot
   1. Age (플레이어의 나이)

|  |
| --- |
|  |
| 플레이어들의 나이에 대한 분포를 살펴보면 대다수의 플레이어들이 16~25세에 분포하고 있는 것을 알 수 있다. 플레이어의 나이가 플레이어의 등급에 영향을 주는 변수로 판단되지는 않아 등급별로 나이의 분포를 보는 것은 제외하였다. |

* 1. HourPerWeek (주당 게임시간)

|  |  |
| --- | --- |
|  |  |
| HourPerWeek 변수를 각 등급별로 어떤 차이가 있는지 살펴보았을 때, 위와 같았다. 8번 그룹의 경우 결측치로 처리되어 있어서 분포를 알 수 없었고, 1~5그룹은 서로 비슷한 분포를 띄는 것을 알 수 있었다. 그리고 6,7 그룹은 나머지 5개 그룹과 평균적으로 약간 더 많은 시간을 투자한 것을 알 수 있다. | |

* 1. TotalHours (게임 총시간)

|  |  |
| --- | --- |
|  |  |
| TotalHours 변수 역시 8번 그룹의 경우 결측치로 처리되어 있어서 분포를 알 수 없었다. 그리고 각 그룹들의 분포가 등급이 올라갈수록 조금씩 더 많은 시간을 게임에 투자하는 것을 알 수 있다. 그리고 각 등급안에서 산포는 대부분 차이가 없었다. | |

* 1. APM (분당 마우스 또는 키보드를 통해 입력한 명령 수)

|  |  |
| --- | --- |
|  |  |
| APM 변수의 등급별 기초통계량과 Boxplot을 살펴보면 등급이 올라갈수록 이 변수의 등급도 비례해서 올라가는 모습을 확인할 수 있다. 따라서 APM은 플레이어의 등급을 그룹별로 구분하는데에 유의한 역할을 할 수 있는 변수라고 판단된다. | |

* 1. SelectByHotkeys (하나의 타임 프레임에서 핫키 사용 평균 횟수)

|  |  |
| --- | --- |
|  |  |
| SelectByHotkeys 변수의 등급별 기초통계량과 Boxplot도 마찬가지로 등급이 올라갈수록 이 변수의 등급도 비례해서 올라가는 모습을 확인할 수 있고 이 변화의 정도가 더 큰 것을 확인할 수 있다. 따라서 핫키의 평균 사용 횟수는 등급이 높을수록 훨씬 더 많이 사용하는 것이므로 플레이어의 등급을 그룹별로 구분하는데에 유의한 역할을 할 수 있는 변수라고 판단할 수 있다. | |

* 1. AssignToHotkeys (핫키에 할당된 건물 또는 유닛 수)

|  |  |
| --- | --- |
|  |  |
| AssignToHotkeys 변수의 등급별 기초통계량과 Boxplot도 SelectByHotkeys에서 처럼 상위 등급으로 올라갈수록 평균이 증가하는 형태를 띄는 것을 알 수 있다. 따라서 핫키의 평균 사용 횟수는 등급이 높을수록 훨씬 더 많이 사용하는 것이므로 플레이어의 등급을 그룹별로 구분하는데에 유의한 역할을 할 수 있는 변수라고 판단할 수 있다.또한 상위 등급일수록 이 변수의 분산이 작은 것으로 보아 하위 등급에서는 편차가 심하지만 상위로 올라갈수록 플레이어들이 비슷한 수준을 유지한다는 것을 알 수 있다. | |

* 1. UniqueHotkeys (유니크 핫키의 사용 수)

|  |  |
| --- | --- |
|  |  |
| UniqueHotkeys 변수의 경우, 상위 등급으로 갈수록 평균이 증가하는 형태를 보이긴 하지만, 가까운 그룹간의 분포 형태의 차이가 크지 않은 것을 확인할 수 있다. Bronze 등급에서부터 Platinum 등급까지 평균과 중앙값이 크게 차이가 나지 않았으며, 이 값은 Diamond 등급부터 변화가 있는 것으로 보였다. 이 변수를 기준으로 등급을 나눈다면 상위그룹과 하위그룹 두 가지 그룹으로 나눌 수 있을 것이라고 판단된다. | |

* 1. MinimapAttacks (미니맵을 통한 공격횟수)

|  |  |
| --- | --- |
|  |  |
| MinimapAttacks 변수의 경우, Boxplot을 살펴보면 그룹별로 분포의 큰 차이가 발생하지 않는 변수라고 판단할 수 있다. 다만, Bronze 등급의 경우 이 값이 0을 가지는 경우가 많아 median 값 또한 0을 가지게 되는 것을 알 수 있고, GrandMaster 그룹과 Professional 그룹에서 다른 그룹에 비해 산포도가 줄어드는 것을 확인할 수 있다. 하지만 확실히 그룹간의 특징을 나눌 수 있는 변수라고 판단할 충분한 근거는 없다고 판단된다. | |

* 1. MinimapRightClicks (미니맵을 이용한 우클릭 횟수)

|  |  |
| --- | --- |
|  |  |
| MinimapRightClick 변수의 경우에도 MinimapAttacks 변수와 마찬가지로 등급이 올라갈수록 평균, 중앙값이 조금 올라가는 경향이 있고, 산포도 또한 줄어드는 경향이 있으나 값의 차이가 크지 않아 그룹 간에 특징을 확실하게 나눌 수 있는 변수라고 판단하기 힘들다.id | |

* 1. PAC (특정 유닛을 클릭한 후 명령을 완료하는 동작)

|  |  |
| --- | --- |
|  |  |
| PAC 변수의 경우 등급이 올라갈수록 분포 또한 비례적으로 움직이는 것을 확인할 수 있다. 특정 유닛을 클릭한 후 명령을 완료하는 동작이 높은 등급의 플레이어 집단 일수록 높은 것을 의미하며, 이 변수는 그룹 간의 차이를 명확히 나타낼 수 있는 변수라고 판단할 수 있다. | |

* 1. GapBetweenPACs

|  |  |
| --- | --- |
|  |  |
| GapBetweenPACs 변수는 등급이 올라갈수록 각 집단이 가지는 평균, 중앙값, 분산이 작아지는 것을 확인할 수 있다. 특정 유닛을 클릭하고 명령을 완료한 뒤 다음 명령을 완료하기까지의 소요시간을 뜻하는 변수인데, 이러한 소요시간이 상위 등급의 플레이어 일수록 짧다라는 것을 의미한다. | |

* 1. ActionLatency (PAC의 첫번째 동작에서 완료할 동작까지 걸리는 평균시간)

|  |  |
| --- | --- |
|  |  |
| ActionLatency 변수 또한 상위 그룹으로 갈수록 평균, 중앙값, 분산이 줄어드는 것을 확인할 수 있다. 이러한 소요시간이 상위 등급의 플레이어 일수록 짧다라는 것을 의미한다. | |

* 1. ActionsInPAC (PAC 내에서 일어나는 작업의 수)

|  |  |
| --- | --- |
|  |  |
| ActionsInPAC 변수의 Boxplot을 살펴보면 그룹간의 큰 차이가 없는 것으로 판단할 수 있다. 따라서 PAC 내에서 일어나는 작업의 수는 플레이어의 등급 간에 큰 차이가 없는 것이라고 판단할 수 있다. | |

* 1. TotalMapExplored

|  |  |
| --- | --- |
|  |  |
| TotalMapExplored 변수의 Boxplot을 살펴보면 그룹간의 큰 차이는 없으나 상위 그룹으로 갈수록 조금씩 증가하는 형태를 보이고 있다. 한 가지 특이한 점은 Master 그룹보다 Professional 그룹이 더 상위 그룹인데 평균 등의 분포가 더 작다는 것인데, 이를 설명할 수 있는 특정한 패턴은 찾지 못하였다. | |

* 1. WorkersMade (하나의 타임프레임 내의 평균 일꾼 수)

|  |  |
| --- | --- |
|  |  |
| WorkersMade 변수의 Boxplot을 살펴보면 상위 그룹으로 갈수록 조금씩 증가하는 형태를 보이다가 5~8번 그룹에서는 큰 차이가 나지 않는 것으로 판단할 수 있다. 따라서 가장 낮은 그룹인 1번 그룹에서 중간 그룹인 4~5번 까지 올라가는 데에는 유의한 변수로 판단할 수 있지만 그 이상의 상위 그룹에서는 차이를 발생시키는 변수라고 보기 힘들다. | |

* 1. UniqueUnitsMade (하나의 타임프레임 내의 Unique한 유닛의 수)

|  |  |
| --- | --- |
|  |  |
| UniqueUnitsMade 변수의 Boxplot을 살펴보면 각 그룹별로 큰 차이가 나지 않는 것으로 판단할 수 있다. 따라서 특수한 유닛을 많이 만드는 것은 그룹 간의 차이를 구분할 수 있는 변수라고 보기 힘들다. | |

* 1. ComplexUnitMade (다루기 어려운 유닛의 평균 수)

|  |  |
| --- | --- |
|  |  |
| UniqueUnitsMade 변수의 Boxplot을 살펴보면 각 그룹별로 큰 차이가 나지 않고 모든 그룹에 아웃라이어로 판단되는 관측치들이 많다는 것을 판단할 수 있다. 따라서 다루기 어려운 유닛을 많이 만드는 것은 그룹 간의 차이를 구분할 수 있는 변수라고 보기 힘들다. | |

|  |  |
| --- | --- |
|  |  |
| ComplexAbilitiesUsed 변수의 Boxplot을 살펴보면 마찬가지로 각 그룹별로 큰 차이가 나지 않고 모든 그룹에 아웃라이어로 판단되는 관측치들이 많다는 것을 판단할 수 있다. 따라서 어려운 기술을 많이 사용하는 것은 그룹 간의 차이를 구분할 수 있는 변수라고 보기 힘들다. | |

* 1. ComplexAbilitiesUsed (어려운 기술 사용 수)