

Характеристика "скилловых" игроков

В данном мини-исследовании я дам характеристику скилловым игрокам в Доте 2 и попытаюсь проверить некоторые интересные гипотезы

Любимые герои скилловых игроков

В качестве скилловых игроков, берем топ-10% игроков по параметру trueskill_mu. Для этого находим попрогоное значение с помощью перцентиля и отсекаем по этому значению тех, у кого trueskill_mu меньше. Далее смотрим количество игр, с участием тех или иных персонажей среди скилловых игроков.

Запрос

```
with ninety_percent_skill as (
    select percentile_cont(0.90) within group (order by trueskill_mu asc) as skill
    from player_ratings
)
select hn.localized_name, count(*)
from players p left join player_ratings pr on p.account_id = pr.account_id
    left join hero_names hn on p.hero_id = hn.hero_id
where trueskill_mu > (select skill from ninety_percent_skill)
group by p.hero_id, hn.localized_name
order by count(*) desc
limit 10;
```

Результат запроса

	localized_name	count
1	Windranger	83
2	Shadow Fiend	55
3	Invoker	54
4	Dazzle	51
5	Winter Wyvern	46
6	Slardar	42
7	Queen of Pain	42
8	Wraith King	40
9	Juggernaut	38
10	Ancient Apparition	38

Популярность среди всех игроков

	localized_name	count
1	Windranger	2214
2	Shadow Fiend	1762
3	Invoker	1351
4	Earthshaker	1274
5	Queen of Pain	1168
6	Juggernaut	1058
7	Slardar	1033
8	Alchemist	1022
9	Tusk	1014
10	Pudge	1014

Вывод

Лидеры среди популярности в игре: Windranger, Shadow Fiend, Invoker сохраняют свое лидерство и у скилловых игроков. **Но на 4 и 5 месте в топ выбиваются новые герои: Dazzle и Winter Wyvern, популярность которых у скилловых игроков сильно выше, чем в среднем по игре.** Вероятно, эти герои не так просты в освоении

Зависит ли выбор героя от умения играть?

Сравним средний показатель скилла у игрока - trueskill_mu, в разрезе героев, т. е. для каждого героя посчитаем среднее арифметическое параметра trueskill_mu всех игроков, которые играли за этого героя в том или ином матче.

ТОП героев по скиллу игроков

	localized_name	skill_mu	matches
1	Beastmaster	27.1	3062
2	Batrider	26.8	1017
3	Chen	26.7	1024
4	Meepo	26.7	1865
5	Ursa	26.6	4155
6	Vengeful Spirit	26.6	3626
7	Timbersaw	26.6	2586
8	Bane	26.6	6071
9	Death Prophet	26.5	571
10	Terrorblade	26.5	3211
11	Io	26.5	3548
12	Templar Assassin	26.5	6592
13	Earth Spirit	26.5	6593
14	Wraith King	26.5	9450
15	Winter Wyvern	26.4	9801
16	Tiny	26.4	6212
17	Shadow Fiend	26.4	16880

Далее оценим разброс скилла от героя к герою. Для этого посчитаем межквартильный размах для этой выборки и сравним ее с межквартильным размахом скилла среди всех игроков в игре в целом.

	iqr
1	0.6
2	4

1 - размах для выборки по героям 2 - размах в целом по игрокам

Запрос

```
with acc_skill_by_heros as (
    -- герои, по среднему скиллу игроков которые за них играют
    select hn.localized_name,
        round(avg(pr.trueskill_mu), 1) as skill_mu,
        sum(pr.total_matches) as matches
    from players p
        left join player_ratings pr on p.account_id = pr.account_id
        left join hero_names hn on p.hero_id = hn.hero_id
    group by p.hero_id, hn.localized_name
    order by skill_mu desc
)

-- межквартильный размах по среднему скиллу игроков, сгруппированному по героям
select round(cast((percentile_cont(0.75) within group ( order by skill_mu asc ) -
    percentile_cont(0.25) within group ( order by skill_mu asc )) as numeric), 2) as iqr
from acc_skill_by_heros
union all
-- межквартильный размах по скиллу среди игроков
select percentile_cont(0.75) within group (order by trueskill_mu) -
    percentile_cont(0.25) within group (order by trueskill_mu) as iqr
from player_ratings;
```

Вывод

Исходя из маленького межквартильного размаха всего в 0.6 единиц параметра `trueskill_mu` у выборки сгруппированной по героям (при 4 единицах у выборки по всем игрокам в целом), можно сделать вывод, что **выбор героя не зависит от умения играть, как правило за каждого героя играют как умеющие играть игроки, так и те, кто играет не очень.**

Однако, можно выделить исключение из этой тенденции, если изучить результат выборки "ТОП героев по скиллу игроков" фото которой вставлено выше. На нем **можно выделить Beastmaster**, у которого самый высокий средний `trueskill_mu` среди героев, при большом количестве матчей, которые были с ним сыграны

Зависит ли уровень скилла от региона?

Информацию о регионе берем из таблицы cluster_regions и смотрим среднее арифметическое параметра trueskill_mu для разных регионов. Отсекаем регионы с маленькой совокупной активностью

	region	matches_count	avg_player_skill
1	BRAZIL	1540	26.2
2	EUROPE	23510	26
3	US EAST	6590	26
4	AUSTRALIA	2720	25.9
5	STOCKHOLM	2272	25.9
6	US WEST	2190	25.9
7	SINGAPORE	10260	25.8
8	AUSTRIA	2960	25.6

Запрос

```
with cte as (
    select c.region,
        count(*) as matches_count,
        round(avg(pr.trueskill_mu), 1) as avg_player_skill
    from players p left join match m on p.match_id = m.match_id
        left join cluster_regions c on m.cluster = c.cluster
        left join player_ratings pr on p.account_id = pr.account_id
    group by c.region
)

select * from cte
where matches_count > 500
order by avg_player_skill desc;
```

Вывод

Уровень скилла не зависит от региона к региону, имея размах всего 0.6.

В разных регионах предпочитают разных героев?

Группируем матчи по регионам и героям, считаем количество сыгранных матчей. Берем первые 3 героя по популярности среди каждого региона.

AUSTRALIA	Windranger	115
AUSTRALIA	Shadow Fiend	65
AUSTRALIA	Juggernaut	63
AUSTRIA	Windranger	127
AUSTRIA	Invoker	79
AUSTRIA	Shadow Fiend	79
BRAZIL	Invoker	61
BRAZIL	Windranger	59
BRAZIL	Shadow Fiend	58
EUROPE	Windranger	988
EUROPE	Shadow Fiend	750
EUROPE	Invoker	623
SINGAPORE	Windranger	461
SINGAPORE	Shadow Fiend	370
SINGAPORE	Earthshaker	356
STOCKHOLM	Windranger	96
STOCKHOLM	Shadow Fiend	85
STOCKHOLM	Invoker	68
US EAST	Windranger	250
US EAST	Shadow Fiend	238
US EAST	Invoker	193
US WEST	Windranger	87
US WEST	Shadow Fiend	83
US WEST	Doom	53

Запрос

```
with region_group as (
    select cr.region as region,
           hn.localized_name as name,
           count(p.hero_id) as hero_count
  from players p left join hero_names hn on p.hero_id = hn.hero_id
                  left join match m on p.match_id = m.match_id
                  left join cluster_regions cr on m.cluster = cr.cluster
 group by cr.region, hn.localized_name
), region_hero_popularity_ranking as (
    select *,
           row_number() over (partition by region order by hero_count desc) as popularity_rank
  from region_group
)
select region, name, hero_count
  from region_hero_popularity_ranking
 where popularity_rank <= 3 and hero_count > 50
 order by region, hero_count desc;
```

Вывод

Нет, картина популярности героев практически не меняется от региона к региону.

Скилловые игроки в среднем играют более долгие матчи?

Делим матчи на 10 групп, каждая группа будет содержать матчи, длительность которых попадает в определенный интервал (длина интервала = (макс. длительность матча - мин. длительность матча) / 10). Для каждой группы считаем средний опыт в минуту у игроков в матче (параметр xp_per_min). Для наглядности показываем динамику, в процентах от предыдущего значения.

	duration_group	avg_duration	avg_xp_per_min	xp_per_min_diff_percent
1	1	3709	466	-5
2	2	3128	488	1
3	3	2875	483	2
4	4	2665	475	1
5	5	2495	472	3
6	6	2340	458	2
7	7	2184	448	3
8	8	2028	433	5
9	9	1846	412	13
10	10	1486	364	<null>

Здесь прослеживается четкая тенденция - чем дольше матч, тем больше опыта в минуту у игроков, участвующих в этом матче. Это можно было бы интерпретировать как ответ на нашу гипотезу, однако эта тенденция не сохраняется, если также для каждой группы посчитать среднее значение параметра trueskill_mu для всех игроков в матчах:

	duration_group	avg_duration	avg_xp_per_min	avg_acc_skill
1	1	3709	466	25.7
2	2	3128	488	25.51
3	3	2875	483	25.85
4	4	2665	475	25.83
5	5	2495	472	25.9
6	6	2340	458	25.81
7	7	2184	448	25.96
8	8	2028	433	26.18
9	9	1846	412	26.22
10	10	1486	364	26.41

Запрос

```
with duration_groups as (
    select
        match_id,
        duration,
        ntile(10) over (order by duration desc) as duration_group
    from match
),
aggregated_groups as (
    select duration_group,
        round(avg(duration), 0) as avg_duration,
        round(avg(pp.xp_per_min), 0) as avg_xp_per_min,
        round(avg(trueskill_mu), 2) as avg_acc_skill
    from duration_groups p left join players pp on p.match_id = pp.match_id
        left join player_ratings pr on pp.account_id = pr.account_id
    group by duration_group
)
select *,
    round(
        (avg_xp_per_min - lead(avg_xp_per_min) over (order by avg_duration desc)) /
        lead(avg_xp_per_min) over (order by avg_duration desc), 2
    ) * 100 as xp_per_min_diff_percent
from aggregated_groups
order by avg_duration desc;
```

Вывод

Несмотря на то, что мы нашли корреляцию опыта в минуту и длительностью матча, ее нельзя интерпретировать как связь скилла игроков и длительностью матча. Вероятно, есть другие факторы, влияющие на то, что с увеличением длительности матчей, опыт в минуту у игроков тоже увеличивается (например, дается какой-то множитель очков), сложно предположить, так как давно не играла в доту.