

Отчет по результатам выполнения тестового задания на позицию Data Analyst

Подготовил
Бычков Владислав

Введение

Вся аналитика для заданий 1.1 и 1.2 была разбита на две логические части:

- Работа с данными , которые не включают записи ботов
- Работа с данными, которые включают как данные ботов, так и данные реальных игроков.

Разделение аналитики на части позволяет посмотреть на данные с разных сторон: общее представление и конкретно по реальным игрокам.

Задание 1.1. Ключевые показатели и важные характеристики

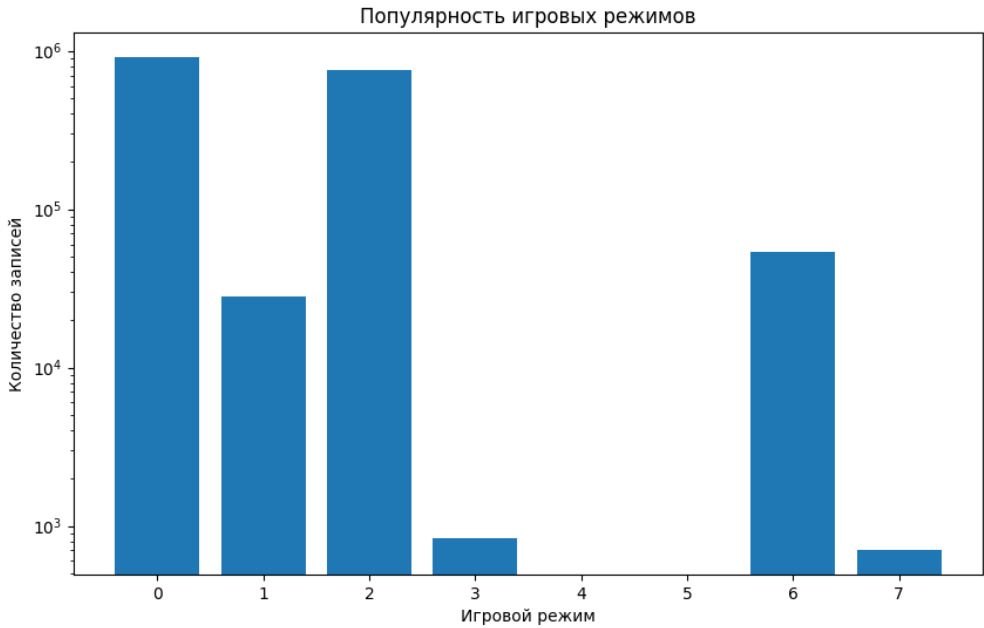
- 1) Количество игровых режимов.
- 2) Количество записей для игровых режимов.
- 3) Среднее количество заработанного опыта, кредитов для каждого игрового режима.
- 4) Средняя доля ботов для каждого игрового режима.
- 5) Средняя продолжительность боя для каждого игрового режима.
- 6) Количество записей для каждого дня недели.
- 7) Количество записей для каждой карты.
- 8) Количество записей для каждого уровня боя.

Задание 1.1

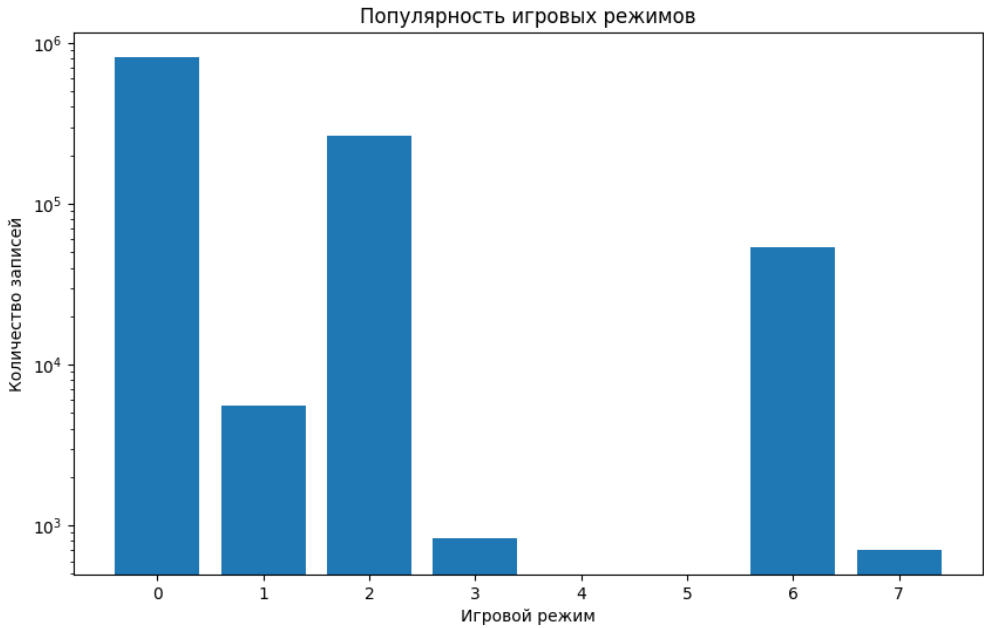
- **Задача:** Проанализировать популярность игровых режимов.
 - Было выявлено 6 различных игровых режимов.

Задание 1.1. Количество записей для игровых режимов.

Общие данные



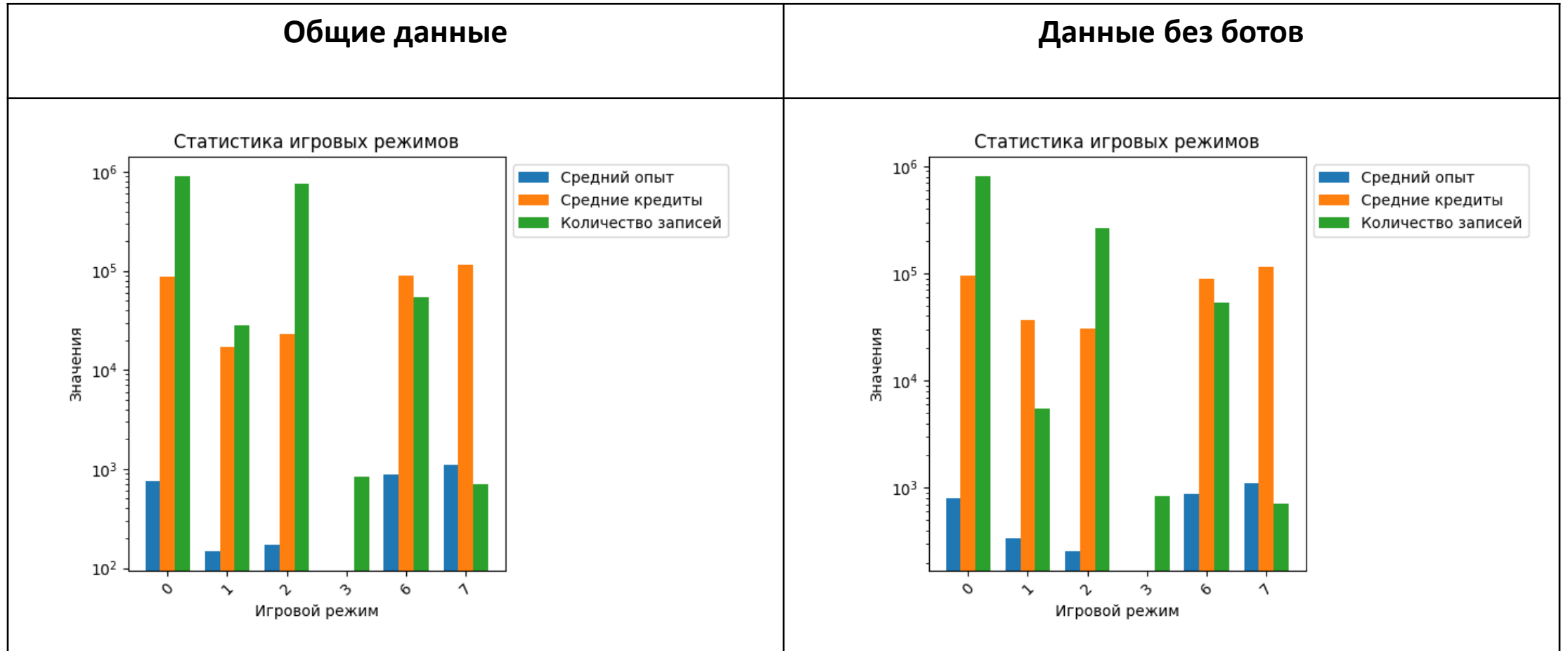
Данные без ботов



Задание 1.1. Количество записей для игровых режимов.

- **Важность рассмотрения** данного аспекта: определение количества записей для каждого игрового режима дает представление о его распространенности и популярности среди игроков. Это может помочь выявить наиболее популярные режимы и использовать эту информацию для улучшения игрового опыта.
- **Вывод:**
 1. Самые популярными режимы являются режимы 0,2.
 2. Самые непопулярные режимы 3,7.
 3. Также отметил то, что в режиме 1 больше всего ботов.

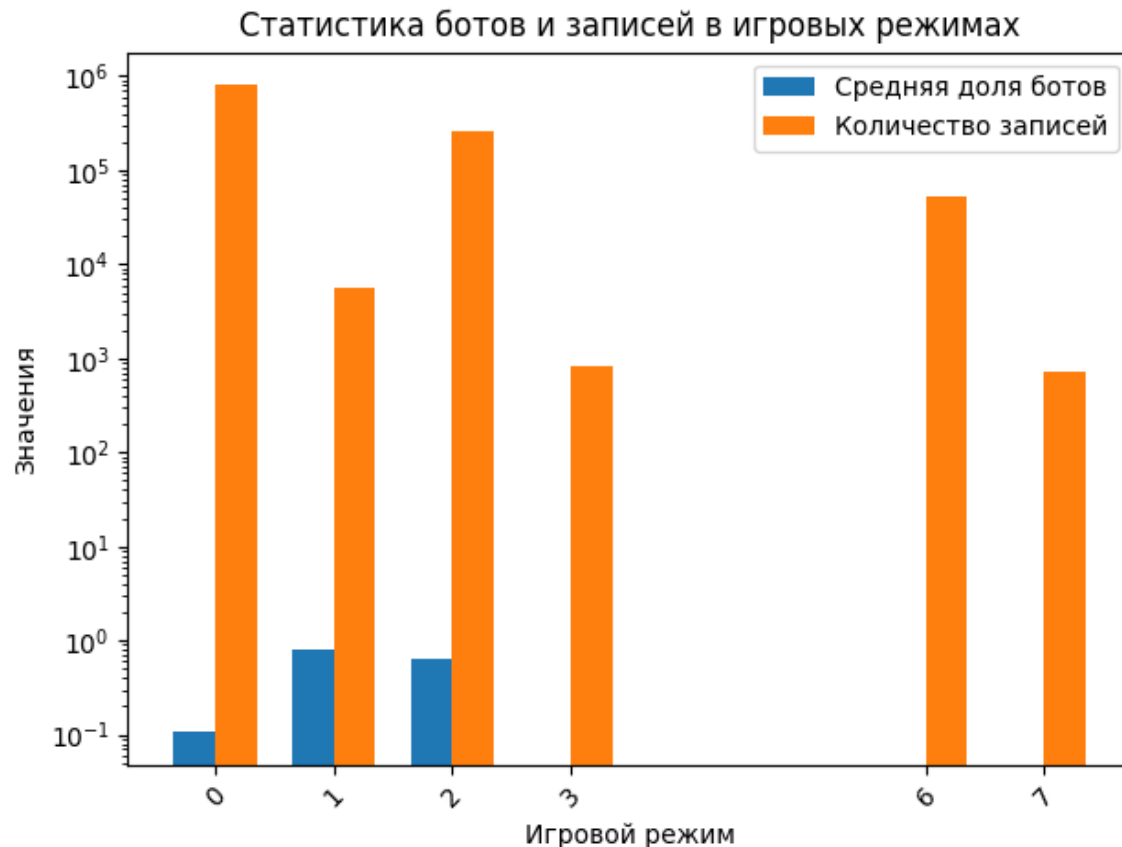
Задание 1.1. Среднее количество заработанного опыта, кредитов для каждого игрового режима.



Задание 1.1. Среднее количество заработанного опыта, кредитов для каждого игрового режима.

- **Важность рассмотрения** данного аспекта: расчет среднего количества заработанного опыта, кредитов и количества записей для каждого игрового режима позволяет определить, какие режимы наиболее прибыльны и пользуются популярностью среди игроков. Также считаю кредиты и опыт являющимися своеобразными показателями эффективности, потому как именно это мотивирует игроков играть лучше. Эти данные могут помочь при принятии решений о балансировке игрового процесса и внесении изменений в игровые режимы.
- **Вывод:**
 1. За 3-ий режим не дают опыта, кредитов.
 2. Самые «прибыльные» режимы: 0,6,7.
 3. Боты «зарабатывают» примерно также, как и реальные игроки.
 4. Популярность игрового режима не зависит от «прибыльности»

Задание 1.1. Средняя доля ботов для каждого игрового режима.



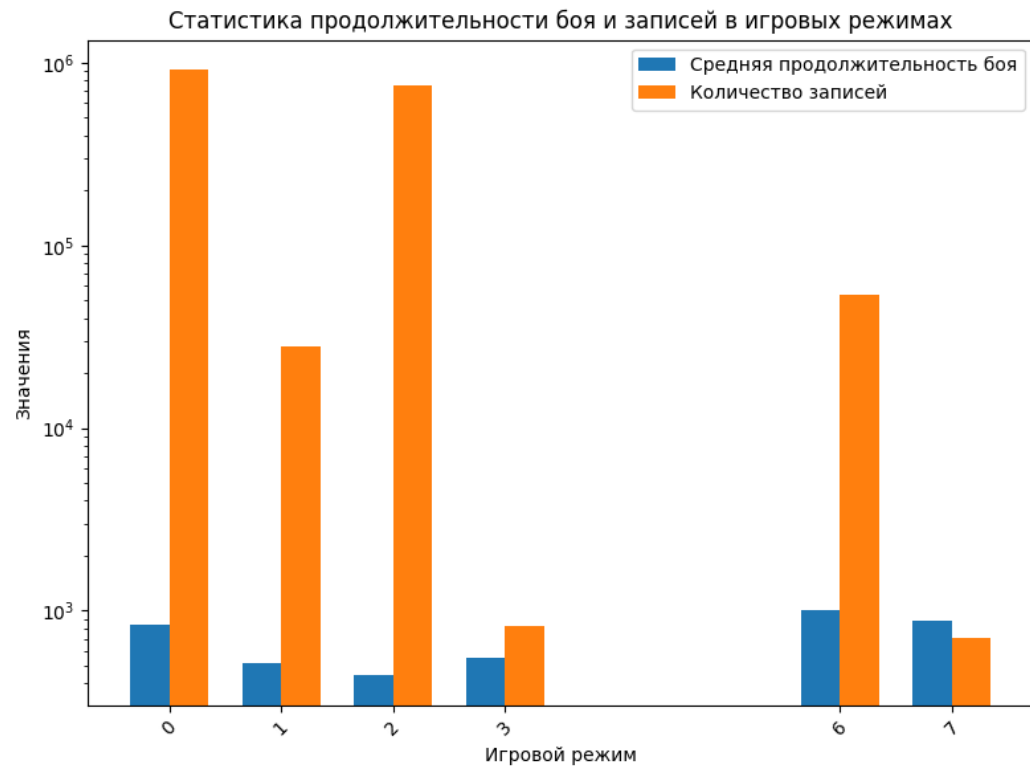
Игровой режим	Средняя доля ботов	Количество записей
0	0,11	814878
1	0,80	5553
2	0,65	263788
3	0,00	831
6	0,00	53620
7	0,00	705

Задание 1.1. Средняя доля ботов для каждого игрового режима.

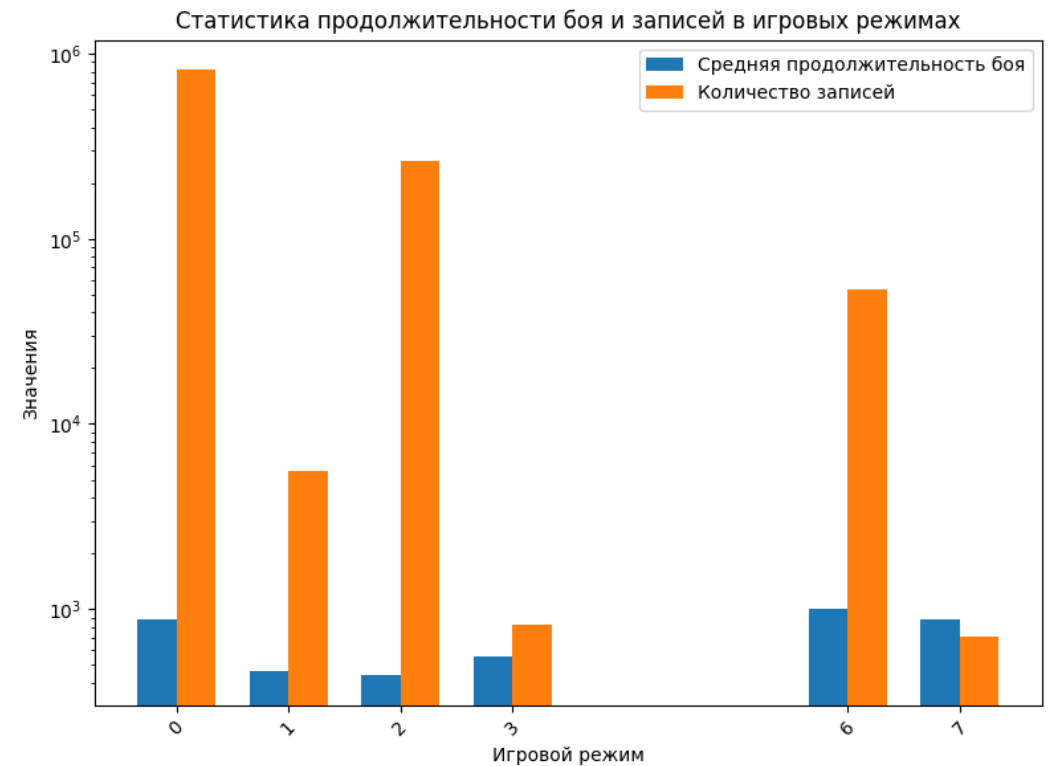
- **Важность рассмотрения** данного аспекта: расчет средней доли ботов для каждого игрового режима дает представление о том, насколько игроки предпочитают играть с другими игроками или с ботами. Это может быть полезной информацией при планировании развития и улучшения мультиплеерной составляющей игры.
- **Вывод:**
 1. Режимы с ботами: 0,1,2. Режимы без ботов: 3,6,7.
 2. Больше всего ботов в режимах 1,2. Меньше всего 0

Задание 1.1. Средняя продолжительность боя для каждого игрового режима.

Общие данные



Данные без ботов

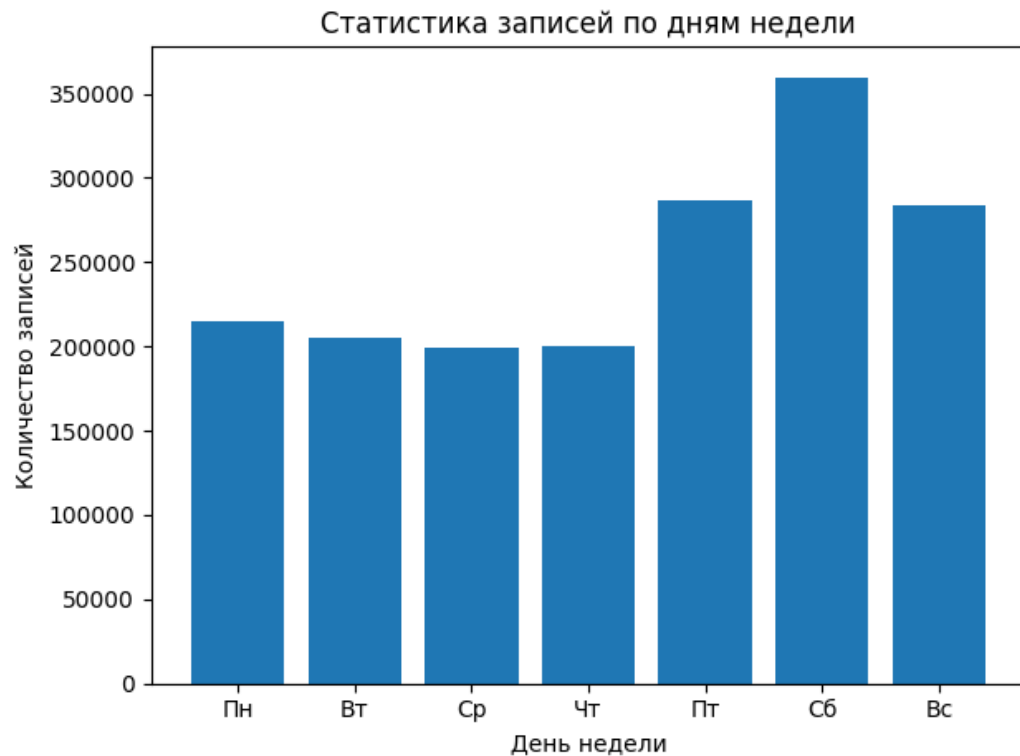


Задание 1.1. Средняя продолжительность боя для каждого игрового режима.

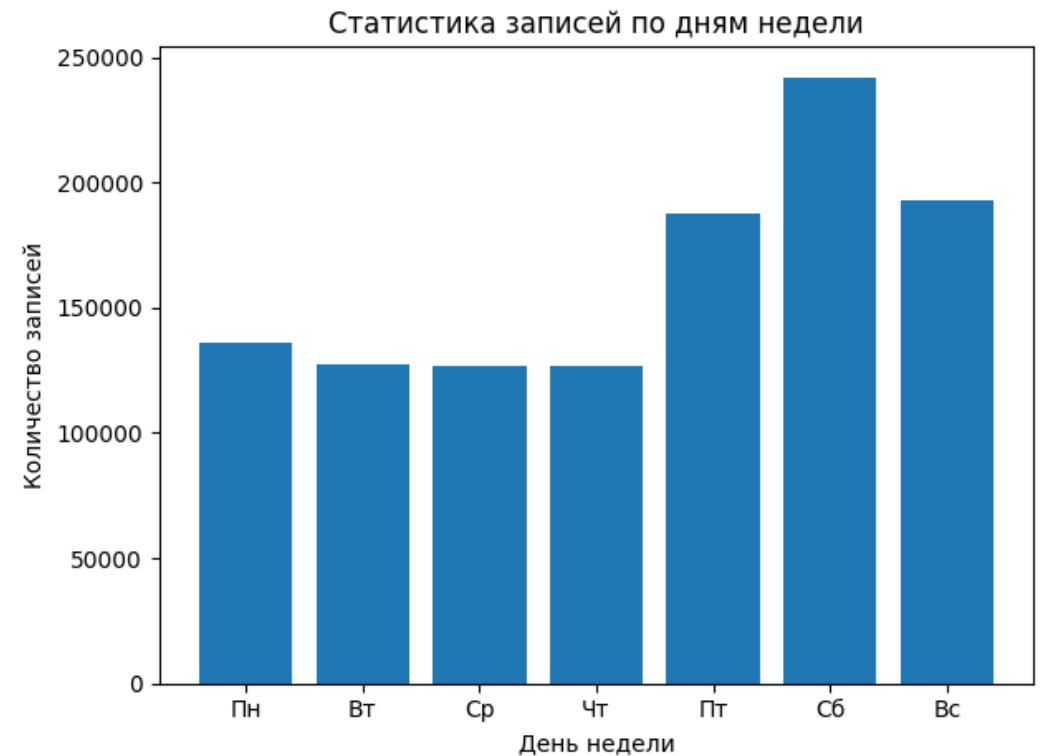
- **Важность рассмотрения** данного аспекта: расчет средней продолжительности боя и количества записей для каждого игрового режима позволяет понять, какие режимы занимают больше времени и наиболее популярны среди игроков. Это может быть полезно для планирования и балансировки игрового времени и ресурсов.
- **Вывод:**
 1. Самые долгие режимы: 6,7,0
 2. Самый короткий режим: 2.
 3. Нет зависимости между продолжительностью боя в режимах и популярностью.

Задание 1.1. Количество записей для каждого дня недели.

Общие данные



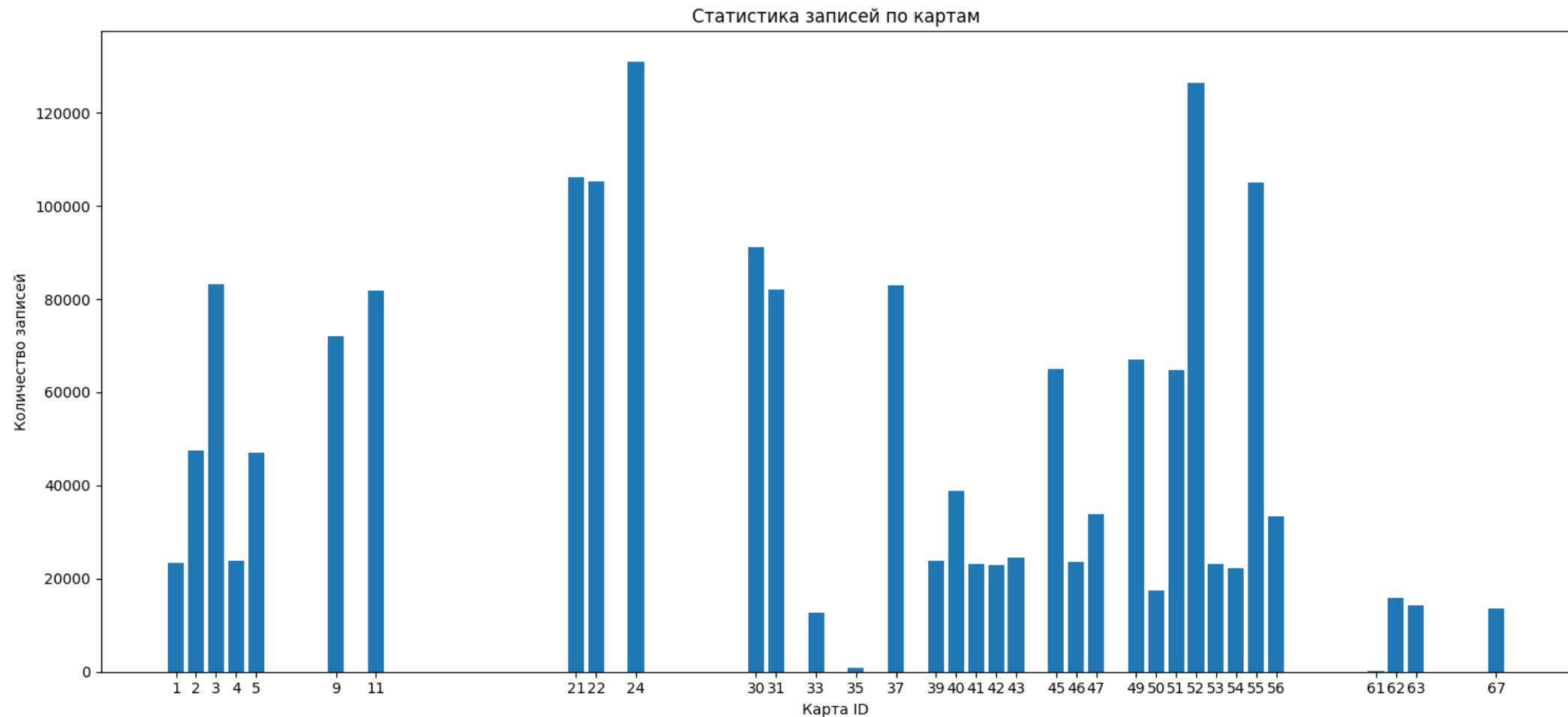
Данные без ботов



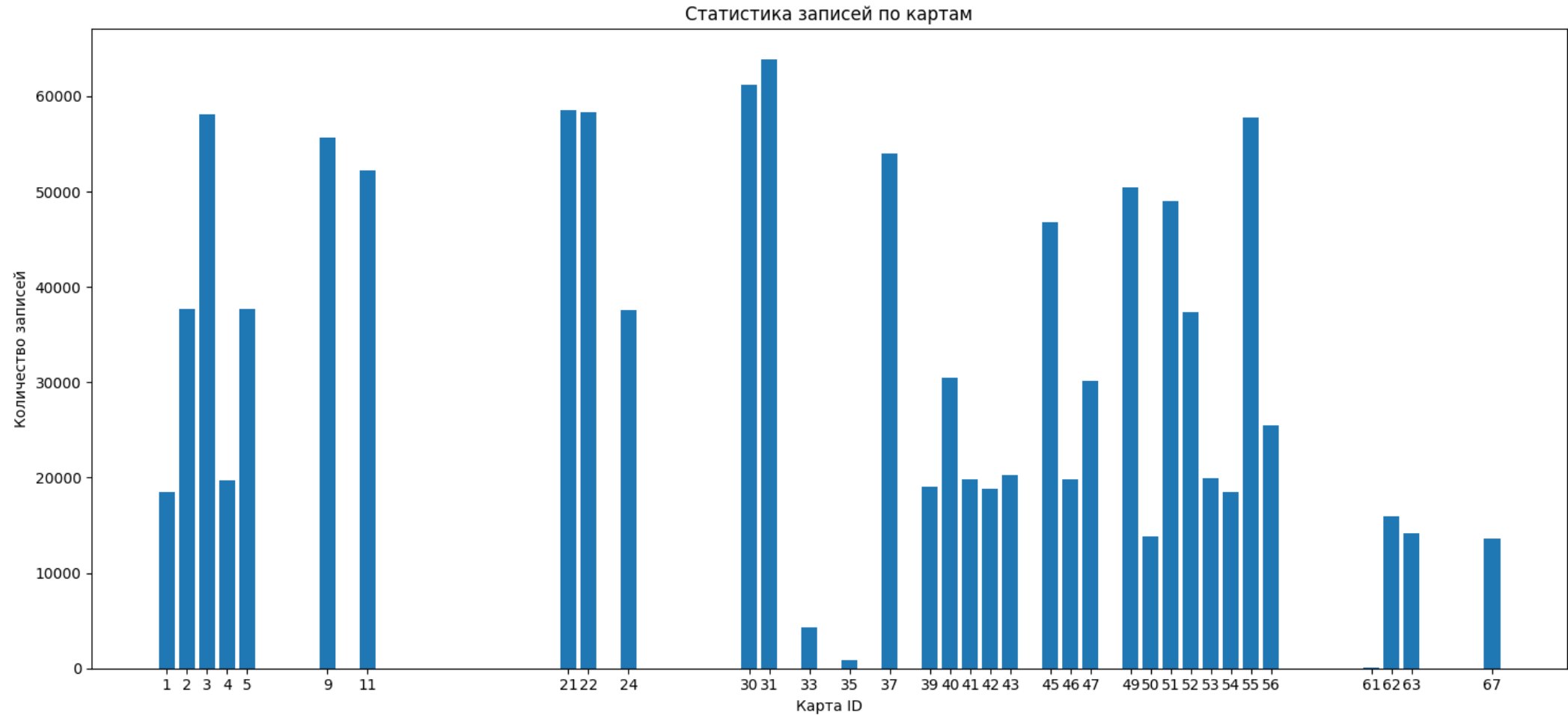
Задание 1.1. Количество записей для каждого дня недели.

- **Важность рассмотрения** данного аспекта: Определение количества записей для каждого дня недели помогает выявить тенденции активности игроков в разные дни. Например, вы можете узнать, в какие дни игровая активность выше, и использовать эту информацию для планирования событий или акций, которые привлекут больше игроков. Кроме того, информация полезна для работников технической сферы, чтобы грамотно управлять загруженностью серверов.
- **Вывод:**
 1. Максимальная активность в пятницу – воскресенье.
 2. Минимальная активность среда – четверг.
 3. Боты «появляются» пропорционально активности игроков.

Задание 1.1. Количество записей для каждой карты. Общие данные



Задание 1.1. Количество записей для каждой карты. Данные без ботов.

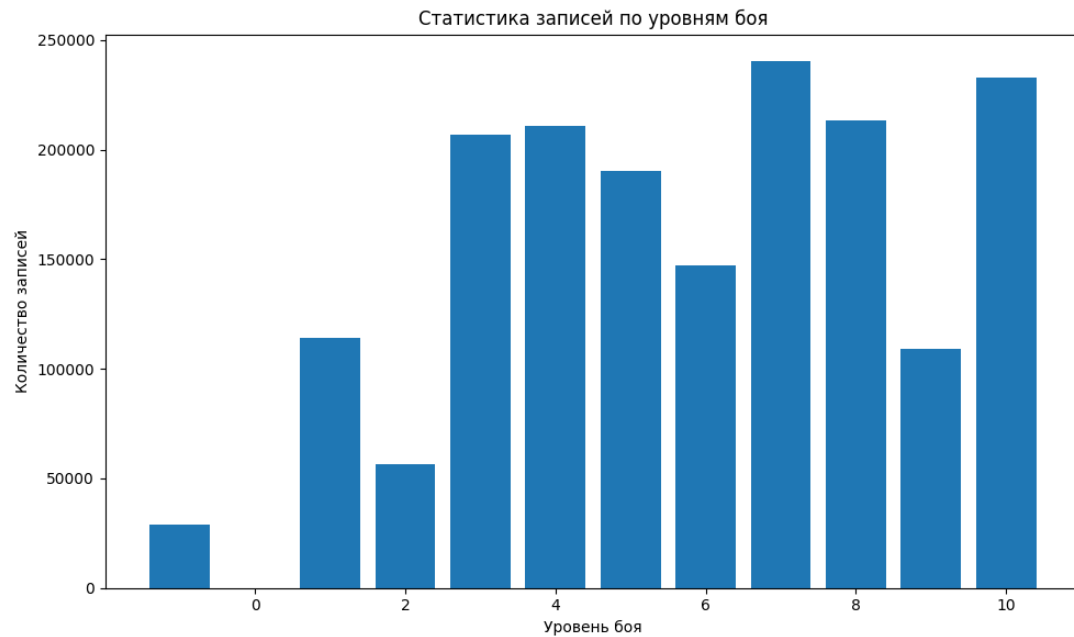


Задание 1.1. Количество записей для каждой карты.

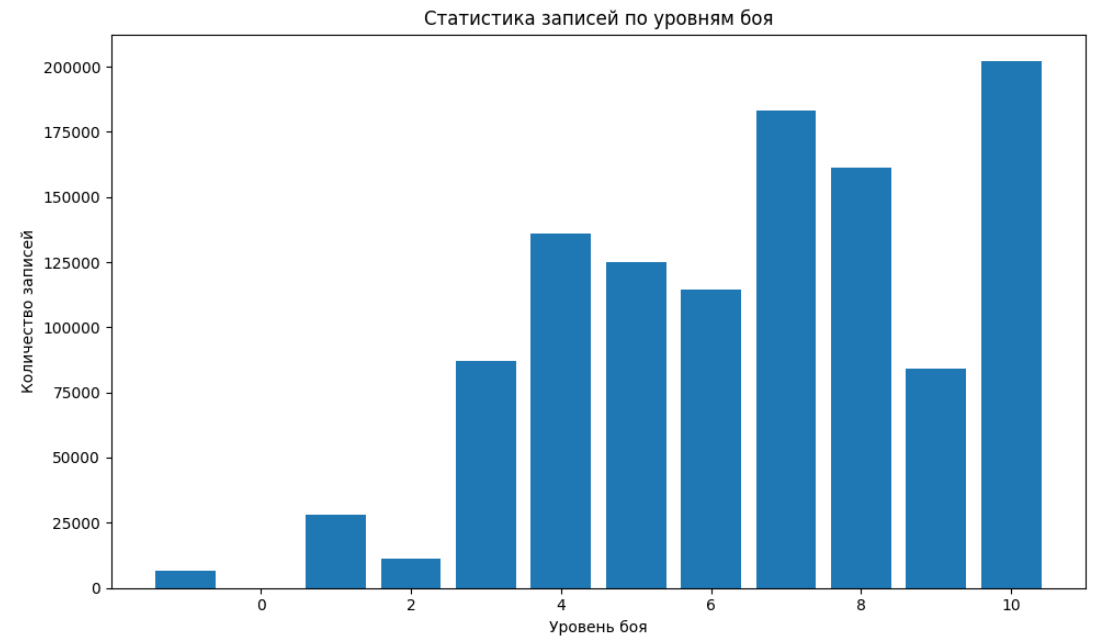
- **Важность рассмотрения** данного аспекта: Определение количества записей для каждой карты позволяет выявить популярность различных игровых карт среди игроков. Это может быть полезным для дальнейшего улучшения и балансировки карт и обеспечения разнообразия в игровом процессе.
- **Вывод:**
 1. Карты имеют разную популярность.
 2. Карты 24,52 имеют самую большую популярность по общей статистике, но на них также много ботов.
 3. Наименьшая популярность у карты 35.
 4. Самые популярные карты среди реальных игроков: 31, 30.

Задание 1.1. Количество записей для каждого уровня боя.

Общие данные



Данные без ботов



Задание 1.1. Количество записей для каждого уровня боя.

- **Важность рассмотрения** данного аспекта: подсчет количества записей для каждого уровня боя помогает понять, какие уровни наиболее популярны среди игроков. Это может быть полезной информацией для разработчиков игры, чтобы обеспечить баланс между различными уровнями и предложить игрокам наиболее интересные и сбалансированные бои.
- **Вывод:**
 1. Выявлен подозрительный уровень = -1
 2. Самые популярные уровни: 7,10.
 3. Самый непопулярный уровень: 2
 4. Количество ботов уменьшается с увеличением уровня.

Задание 1.2. Ключевые показатели и методы / техники анализа.

- **Задача**: Проанализировать эффективность кораблей.
- Выбраны следующие столбцы-показатели эффективности:
Количество уничтоженных вражеских кораблей ,самолетов.
Урон нанесенный вражеским кораблям , нанесенный в бою по союзникам , полученный в бою.
Объем здоровья, восстановленный в бою.
Выживаемость корабля.
Количество заработанных кредитов и опыта.

Задание 1.2. Ключевые показатели и методы / техники анализа.

1. Проведена кластеризация боев – позволяет определить, на какие группы относительно выбранных столбцов могут делиться записи.
2. С помощью метода кластеризации была построена классификация кораблей на 2 класса (возможны другие варианты). Появилась возможность причислять корабль к определенному классу эффективности.
3. Построена линейная регрессия (количество уничтоженных вражеских кораблей, как основной показатель эффективности) на остальные показатели. Позволяет учесть вклад других показателей в способность уничтожать корабли. Точно также построена для количества кредитов и опыта (как показатель эффективности от разработчиков игры)
4. Построена линейная регрессия для каждого класса (Cruiser, Battleship...) кораблей.
5. Найдены различные показатели, основанные на средних значениях.

Задание 1.2. Кластеризация.

Общая Кластеризация

Cluster Label	ships_killed	planes_killed	damage	team_damage	received_damage	regen_hp	is_alive	credits	exp
Cluster 0	0,71	1,42	28852,70	51,56	27720,96	4315,07	0,35	59211,95	497,13
Cluster 1	0,96	2,44	82003,24	198221,40	31790,84	3233,04	0,64	39927,88	321,16
Cluster 2	0,85	0,96	24559,95	18261,62	21398,05	1521,61	0,27	41216,07	332,27
Cluster 3	0,63	2,16	30265,22	47085,65	37133,15	3502,62	0,22	48672,29	363,75
Cluster 4	0,87	1,00	22852,03	6189,10	19103,12	1625,62	0,31	40263,95	351,88

Кластеризация без ботов

Cluster Label	ships_killed	planes_killed	damage	team_damage	received_damage	regen_hp	is_alive	credits	exp
Cluster 0	0,83	1,89	37247,11	29,19	30978,82	6004,91	0,47	79242,12	670,95
Cluster 1	0,81	1,77	32025,48	25933,83	26015,94	2927,93	0,39	56639,16	467,20
Cluster 2	0,92	0,69	78679,62	254975,62	29946,69	6217,38	0,69	2650,54	16,31
Cluster 3	1,00	1,72	32921,43	7688,33	20837,63	3127,91	0,51	59462,95	509,43
Cluster 4	0,46	3,54	30706,66	65918,55	45375,11	3918,37	0,18	48707,88	373,85

Задание 1.2. Кластеризация.

- **Важность рассмотрения** данного аспекта: кластеризация боев: кластеризация боев позволяет группировать бои на основе их характеристик, что может помочь выделить различные паттерны и типы боев.
- **Вывод:**
 1. Путем различных методов был выбрано оптимальное количество кластеров для боев = 5. В каждом из кластеров есть свои особенности: у представителей одного кластера больше уничтоженных кораблей, у другого больше полученного опыта и т.д.
 2. В дальнейшем идею этого метода используем для классификации эффективности кораблей.

Задание 1.2. Классификация на 2 класса

Общая Классификация

Cluster Label	AVG(ships_killed)	AVG(planes_killed)	AVG(damage)	AVG(team_damage)	AVG(received_damage)	AVG(regen_hp)	AVG(is_alive)	AVG(credits)	AVG(exp)
Cluster 0	0,84	15,45	46288,05	262,91	20124,19	150,82	0,65	83951,23	613,71
Cluster 1	0,76	0,84	31578,34	195,57	28573,93	4631,25	0,34	72068,60	570,33

Классификация без ботов

Cluster Label	AVG(ships_killed)	AVG(planes_killed)	AVG(damage)	AVG(team_damage)	AVG(received_damage)	AVG(regen_hp)	AVG(is_alive)	AVG(credits)	AVG(exp)
Cluster 0	1,01	18,73	55411,26	298,89	14913,70	192,58	0,78	104903,49	779,83
Cluster 1	0,87	0,90	35664,87	131,69	27627,97	5022,09	0,45	83638,50	685,25

Задание 1.2. Классификация на 4 класса

Общая Классификация

Cluster Label	AVG(ships_killed)	AVG(planes_killed)	AVG(damage)	AVG(team_damage)	AVG(received_damage)	AVG(regen_hp)	AVG(is_alive)	AVG(credits)	AVG(exp)
Cluster 0	0,809702885	1,855676538	58235,41286	139,0442452	64191,75031	17951,422	0,445681083	113117,1215	814,8998869
Cluster 1	0,864447022	0,929863009	33993,22159	126,2340828	25601,1669	3043,057432	0,381048842	86141,11817	666,8833793
Cluster 2	0,838568946	15,45095086	46288,05286	262,9050368	20124,19212	150,8207015	0,651126304	83951,22712	613,7096238
Cluster 3	0,623630927	0,285097364	16878,22835	296,1839814	15636,39984	308,1721678	0,260107664	38235,62425	355,0141657

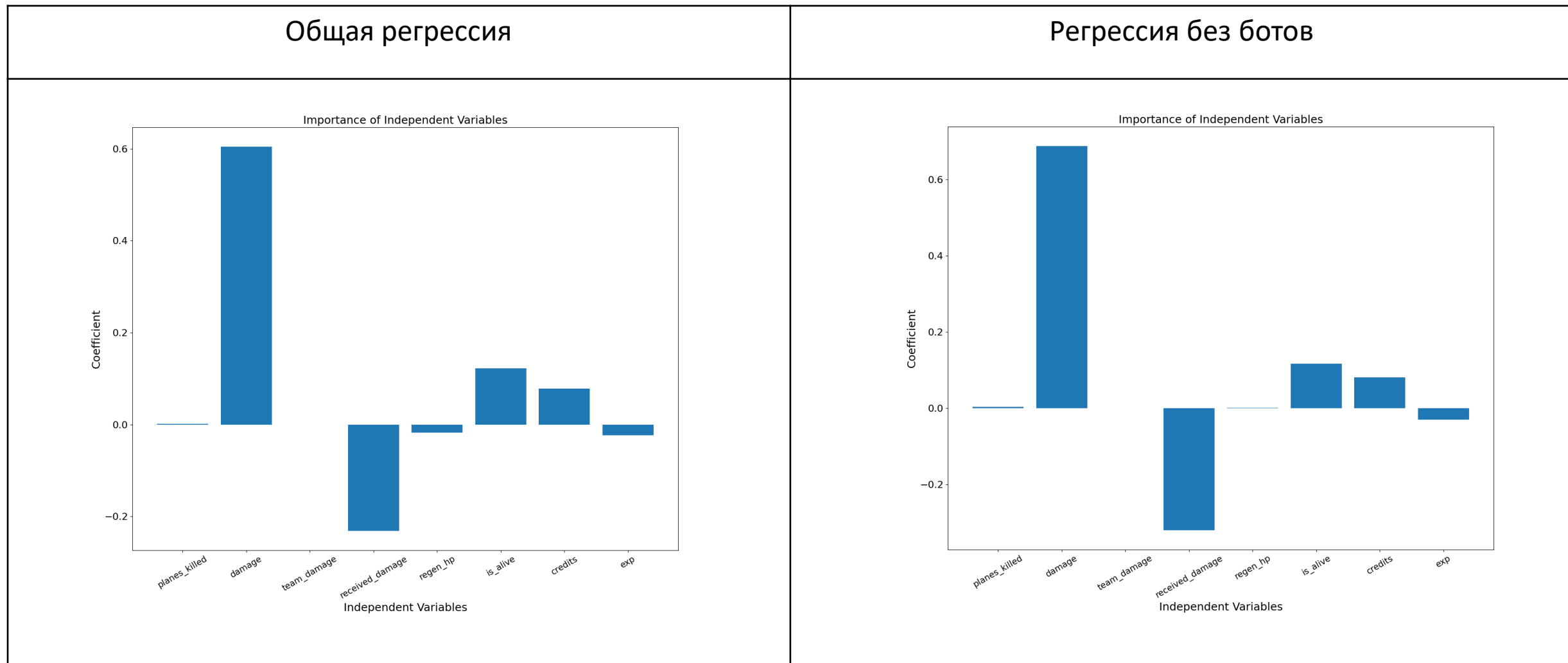
Классификация без ботов

Cluster Label	AVG(ships_killed)	AVG(planes_killed)	AVG(damage)	AVG(team_damage)	AVG(received_damage)	AVG(regen_hp)	AVG(is_alive)	AVG(credits)	AVG(exp)
Cluster 0	1,006101978	18,72742645	55411,25956	298,8884862	14913,70014	192,5785258	0,780269362	104903,4879	779,8278969
Cluster 1	0,863451086	1,835908482	56902,08449	96,06845298	56924,67415	16441,23911	0,495167452	115056,9514	865,7980448
Cluster 2	0,937737987	0,276538599	22144,7438	498,7150272	11174,66785	88,44322775	0,395092012	48078,42814	448,7330065
Cluster 3	0,867588992	0,601818624	28854,60093	102,4567556	17973,7629	1084,451505	0,433129805	75400,09637	641,7097524

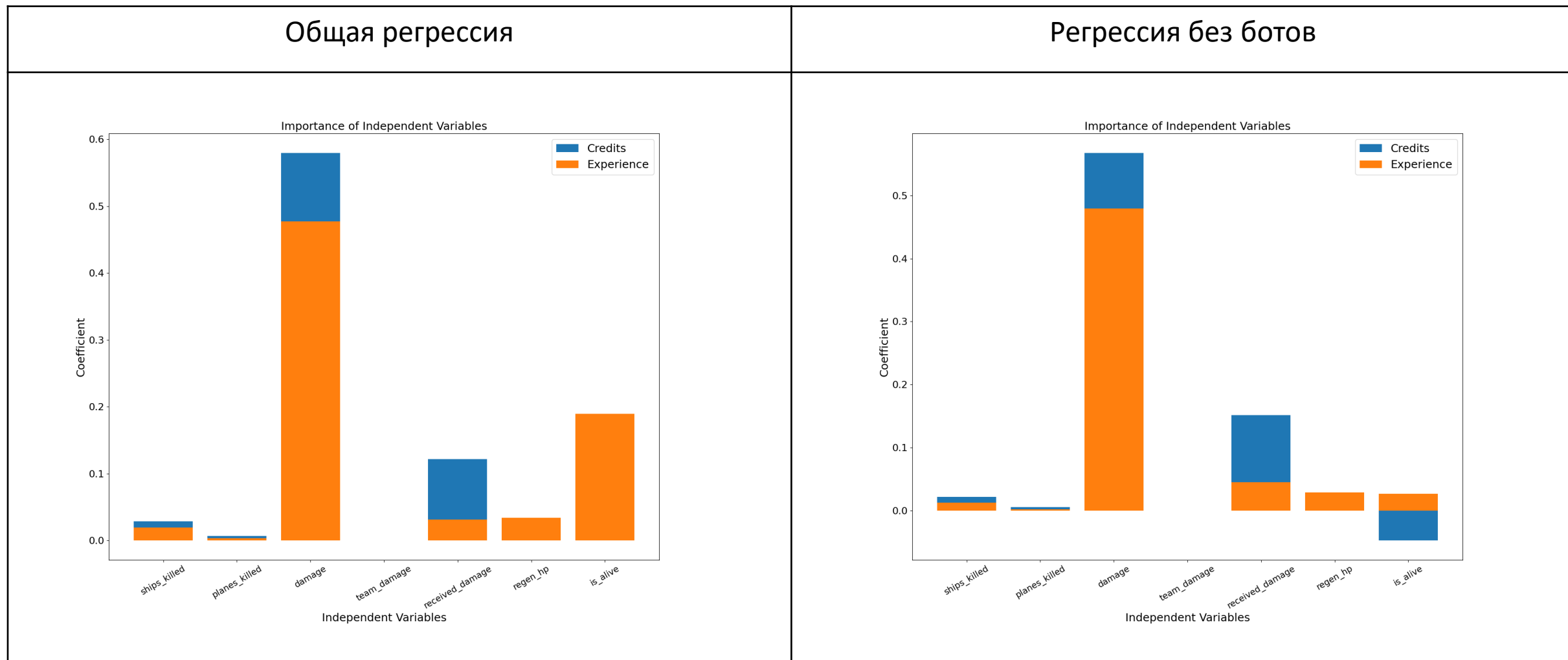
Задание 1.2. Классификация.

- **Важность рассмотрения** данного аспекта: Классификация кораблей позволяет разделить корабли на группы схожих характеристик и определить их классы эффективности, что полезно при анализе и сравнении их эффективности. Выбор методологии классификации обусловлен ее простотой в реализации, возможностью работы с многомерными данными и интерпретируемостью результатов.
- **Вывод:**
 1. Путем использования различных методов было выбрано оптимальное количество кластеров для кораблей = 2. Но мы можем посмотреть разделение на любое количество классов. Разделение на классы позволяет проследить, чему равны «второстепенные» показатели у классовой с высокой и низкой эффективностью.
 2. Классификация кораблей реальных игроков отличается.

Задание 1.2. Линейная регрессия количества уничтоженных вражеских кораблей.



Задание 1.2. Линейная регрессия количества опыта и кредитов

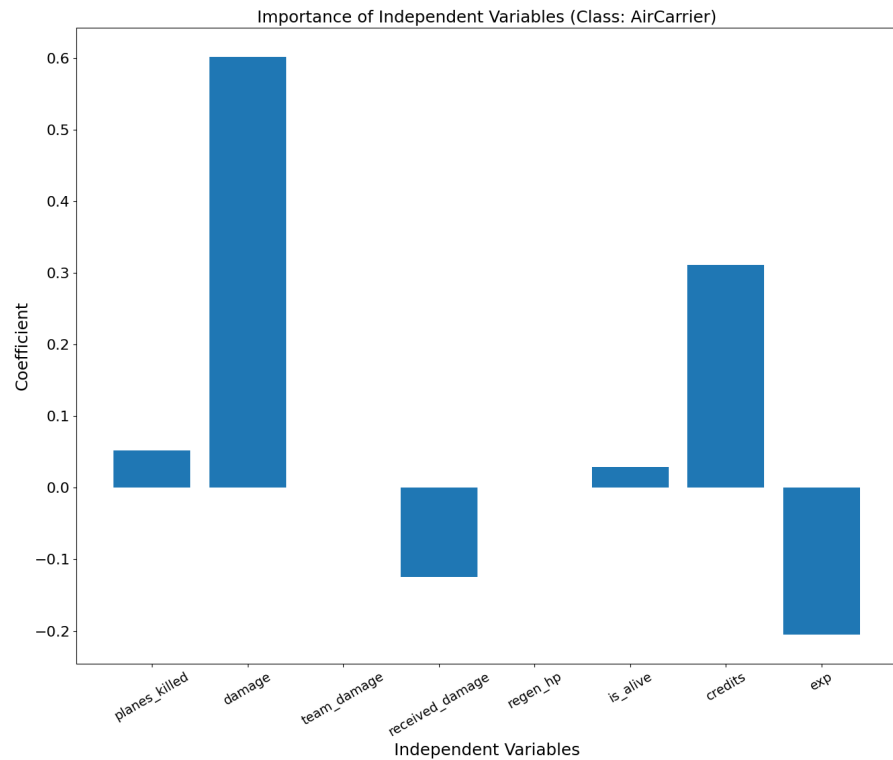


Задание 1.2. Линейная регрессия.

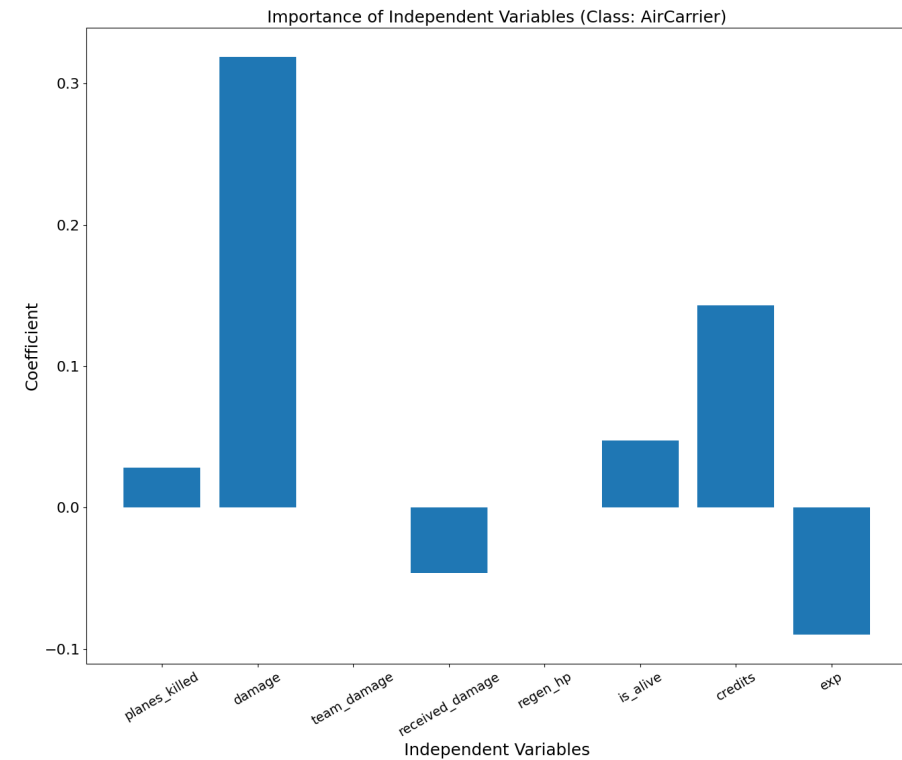
- **Важность рассмотрения** данного аспекта: Построение линейной регрессии для оценки взаимосвязи между показателями эффективности и другими факторами может помочь идентифицировать факторы, которые влияют на эффективность кораблей. Это может быть полезной информацией для принятия решений и оптимизации стратегии в игре.
- **Вывод:**
 1. Количество уничтоженных кораблей сильнее всего зависит от количество нанесенного урона (прямая зависимость) и количество полученного урона (обратная зависимость).
 2. Интересная ситуация по регрессии полученных кредитов и опыта. Если смотреть на общую гистограмму: наибольший вклад в получение опыта вносит нанесенный урон и выживаемость. А для кредитов – нанесенный и полученный урон. Однако, если смотреть на гистограмму без ботов, появляется аномальный эффект по выживаемости: она резко снизила свою «важность» для полученного опыта а для кредитов и вовсе ушла в обратную зависимость. Это может происходить по разным причинам – от несовершенства моделирования до ошибок в данных. Требуется дальнейший углубленный анализ.

Задание 1.2. Линейная регрессия. Класс AirCarrier.

Общая регрессия

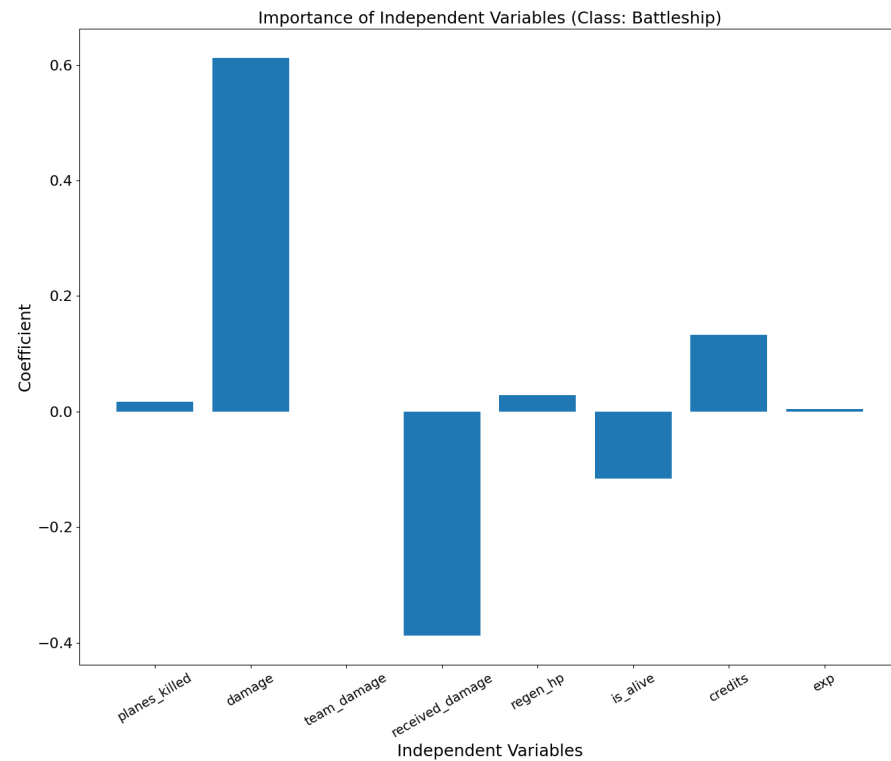


Регрессия без ботов

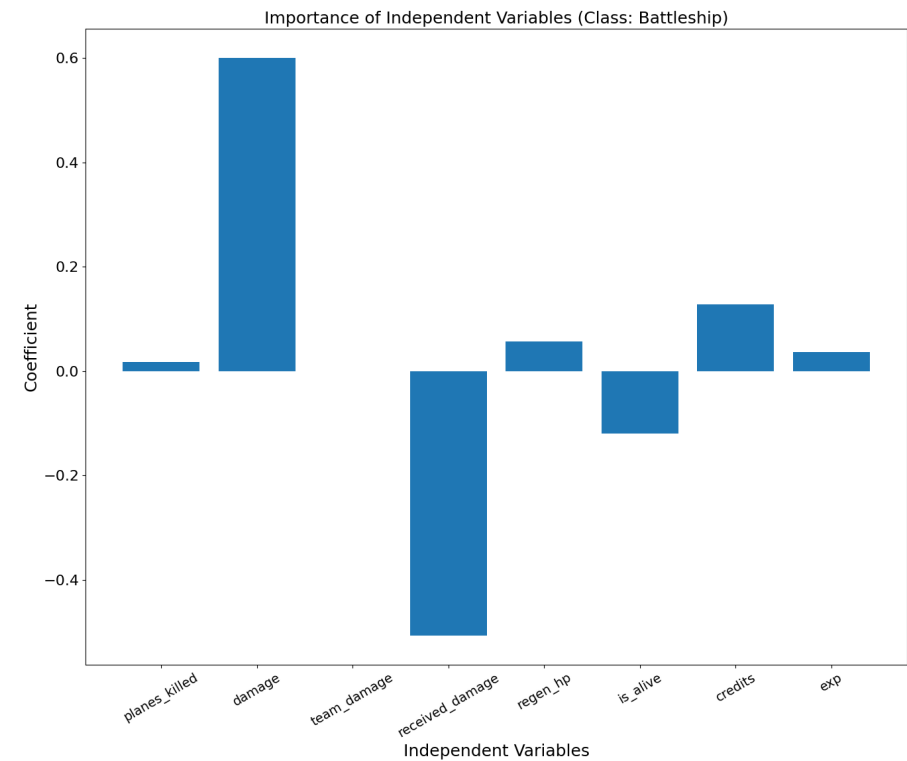


Задание 1.2. Линейная регрессия. Класс Battleship.

Общая регрессия

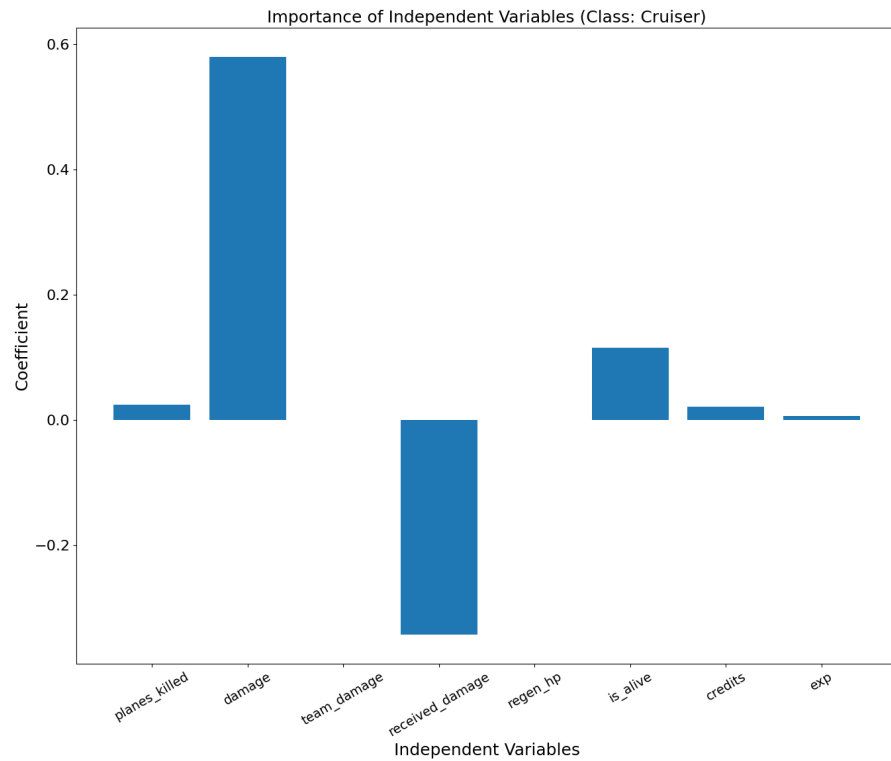


Регрессия без ботов

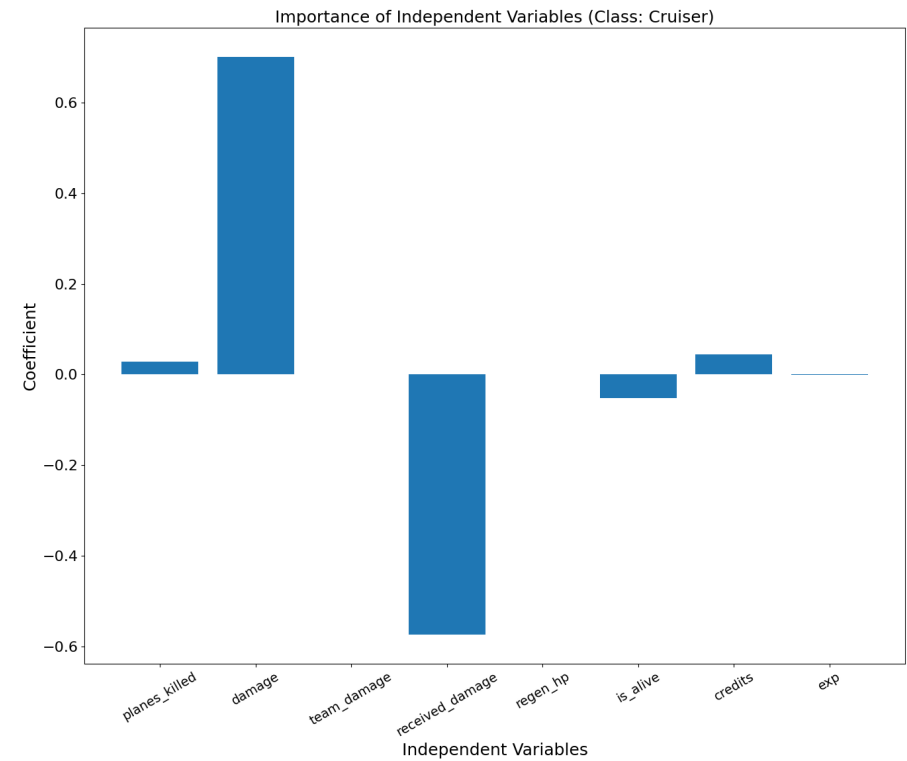


Задание 1.2. Линейная регрессия. Класс Cruiser.

Общая регрессия

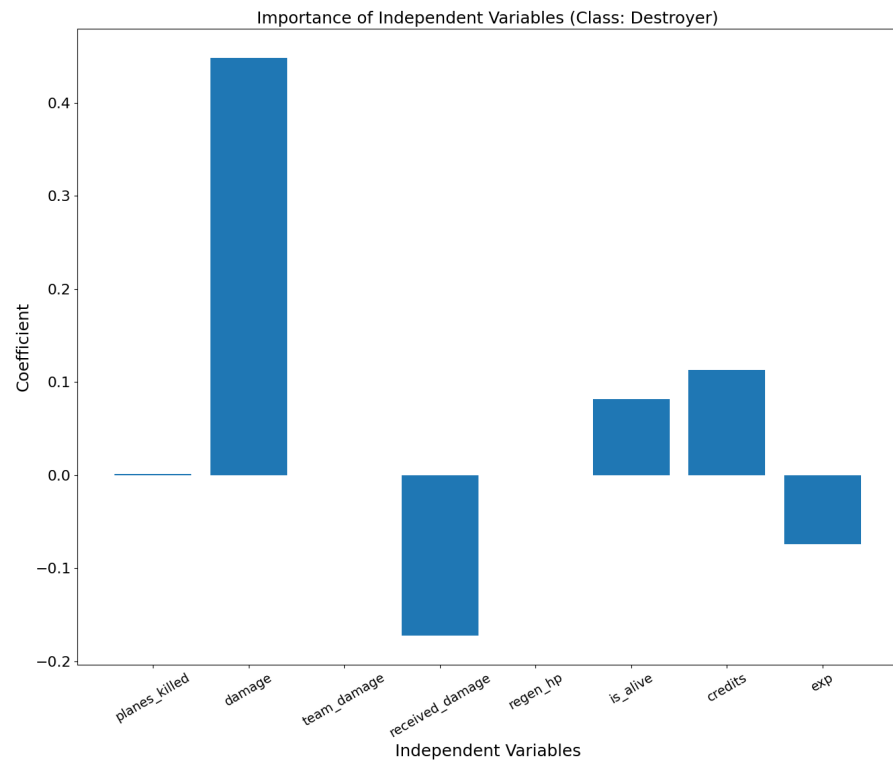


Регрессия без ботов

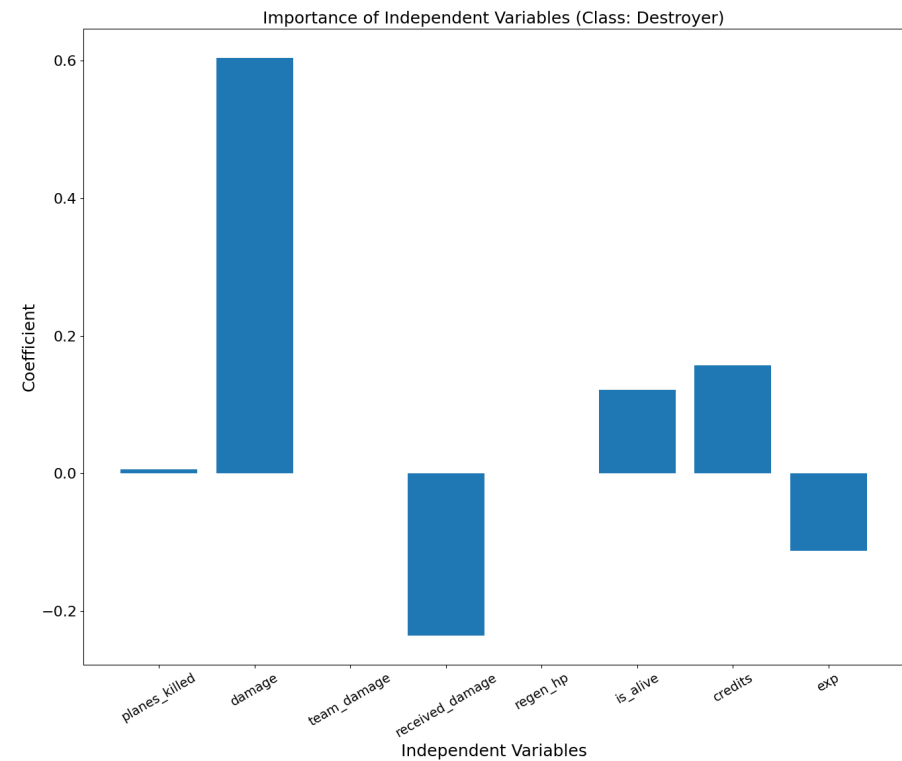


Задание 1.2. Линейная регрессия. Класс Destroyer.

Общая регрессия



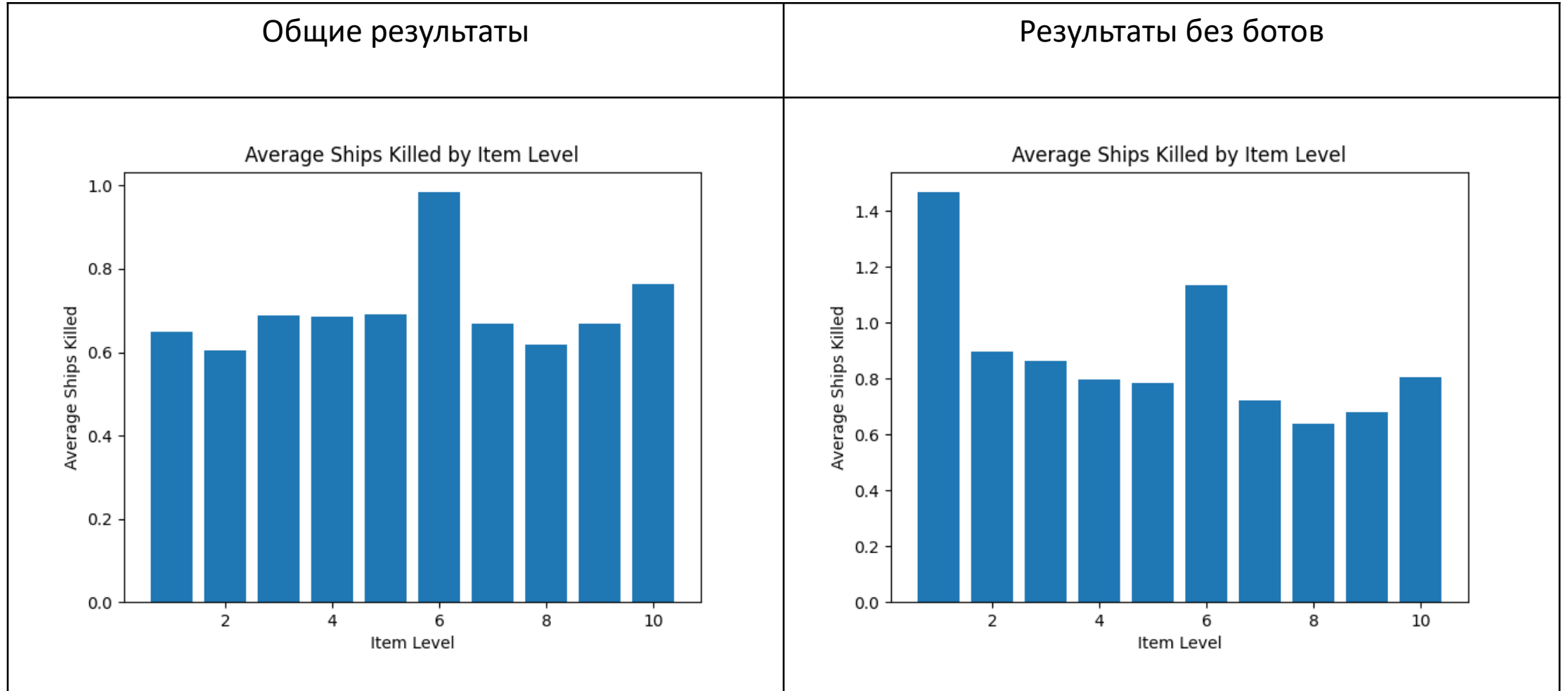
Регрессия без ботов



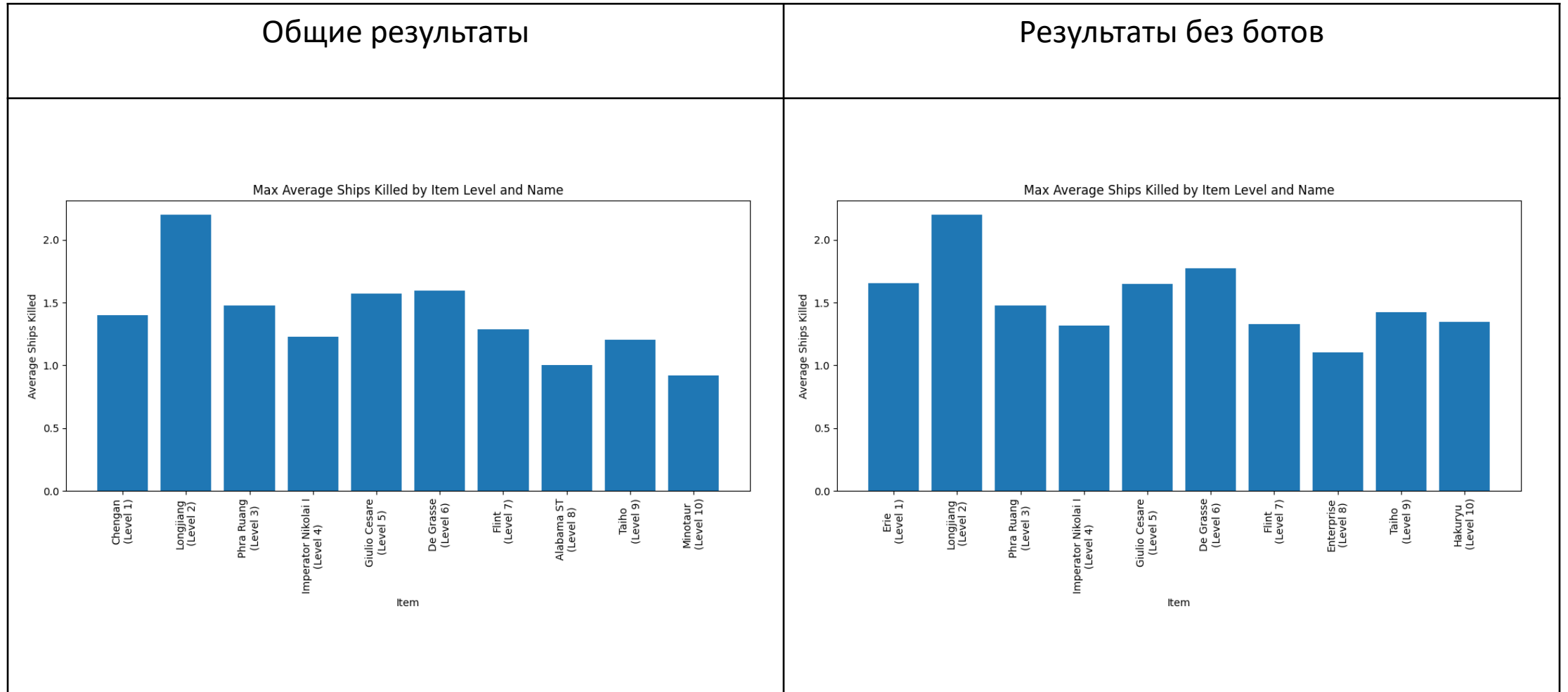
Задание 1.2. Линейная регрессия.

- **Важность рассмотрения** данного аспекта: Построение линейной регрессии для каждого класса позволяет увидеть разницу во взаимосвязи между показателями эффективности в разных классах кораблей. Это может быть полезной информацией для принятия управления балансом в игре и оптимизации стратегии.
- **Вывод:**
 1. Связь количество уничтоженных кораблей с другими характеристика эффективности отличается от класса к классу.
 2. Кроме того, во всех регрессиях выше нанесенный урон по союзникам никак не влияет на количество уничтоженных кораблей.

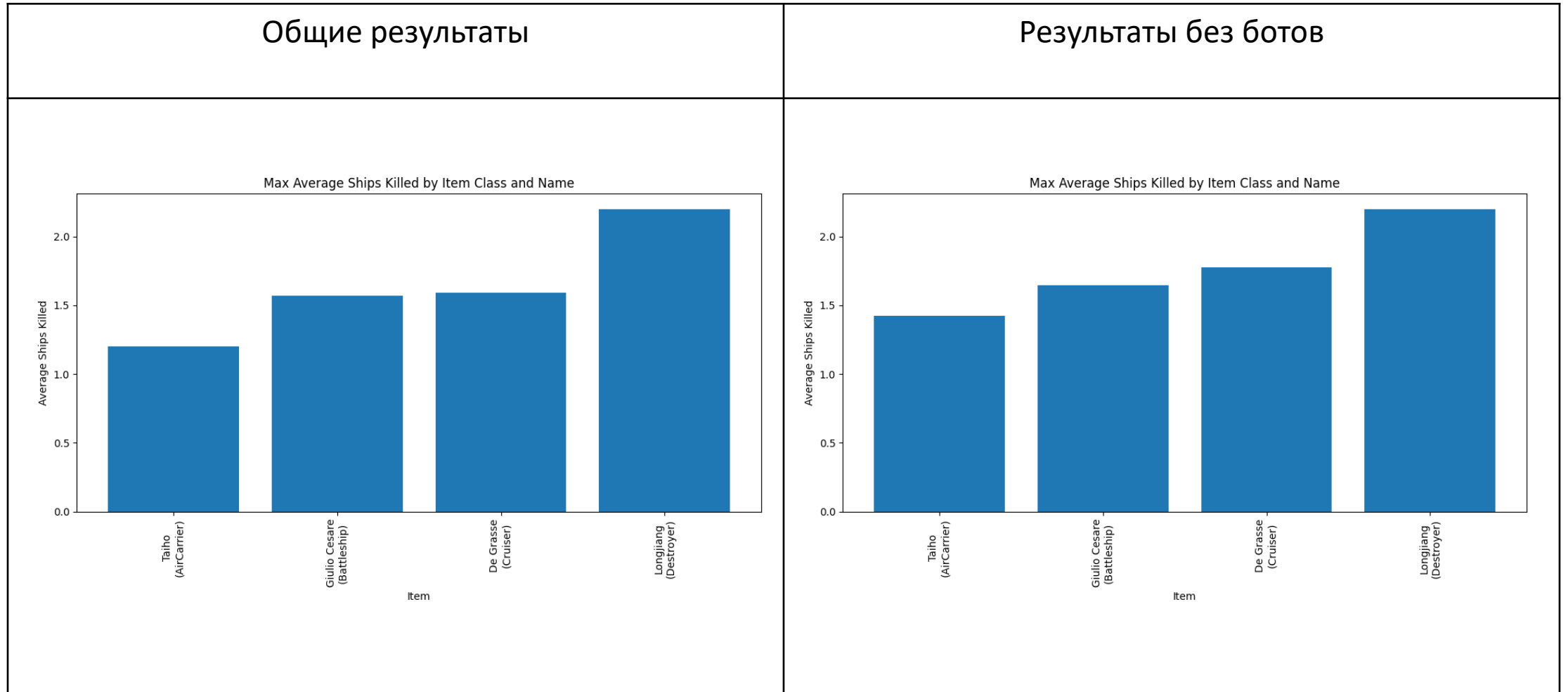
Задание 1.2. «Средние» показатели. Среднее количество уничтоженных кораблей по уровню техники



Задание 1.2. «Средние» показатели. Максимальные средние количества уничтоженных кораблей по уровню техники и названию корабля



Задание 1.2. «Средние» показатели. Среднее количество уничтоженных кораблей по классу корабля и названию корабля.



Задание 1.2. «Средние» показатели .

- **Важность рассмотрения** данного аспекта: расчет средних показателей, таких как среднее количество уничтоженных кораблей по различным категориям (уровню техники, названию корабля, классу и т.д.), позволяет получить обобщенную информацию о характеристиках кораблей.
- **Вывод:**
 1. По общему графику наибольшее среднее количество уничтоженных кораблей наблюдается на 6 уровне техники. Для не ботов – это 1 уровень.
 2. Самый эффективный корабль по уничтожению кораблей, как на своем уровне, так и в сравнении с другими уровнями - корабль “Longjiang” 2 уровня.
 3. Самый эффективный класс по уничтожению кораблей – “Destroyer” и самый лучший представитель этого класса снова “Longjiang”.