

АНАЛИЗ ОНЛАЙН-СЕРВИСА ЦИФРОВОГО РАСПРОСТРАНЕНИЯ КОМПЬЮТЕРНЫХ ИГР И ПРОГРАММ STEAM

ANALYSIS OF STEAM ONLINE SERVICE FOR DIGITAL DISTRIBUTION OF COMPUTER GAMES AND PROGRAMS

V. Gladun
A. Pantykhin
I. Malinin
S. Molodyakov

Summary. An analysis of Steam, an online service for digital distribution of computer games and programs, is presented. Stimulated by the rapid growth of digital content and the wide variety of game products, Steam has become an important source of data for researchers, developers, and investors in the game industry. Software for working with the service was developed. User preferences by genre, pricing trends, popularity of games, their ratings and reviews were analyzed. Sales dynamics, audience segmentation and evaluation of competitiveness of products on Steam platform are considered. Recommendations on how to apply the results in practice are presented.

Keywords: data analysis, online service, Steam, video games.

Гладун Владимир Вадимович

Санкт-Петербургский политехнический университет Петра Великого
vladimir.gldn@gmail.com

Пантюхин Андрей Максимович

Санкт-Петербургский политехнический университет Петра Великого
panandafog@gmail.com

Малинин Илья Игоревич

Санкт-Петербургский политехнический университет Петра Великого
malinin.ilja@gmail.com

Молодяков Сергей Александрович

Д.т.н., профессор, Санкт-Петербургский политехнический университет Петра Великого
molodyakov_sa@spbstu.ru

Аннотация. Представлен анализ онлайн-сервиса цифрового распространения компьютерных игр и программ Steam. Стимулированный быстрым ростом цифрового контента и широким разнообразием игровых продуктов, Steam стал важным источником данных для исследователей, разработчиков и инвесторов в игровой индустрии. Разработано программное обеспечение для работы с сервисом. Проведен анализ предпочтения пользователей по жанрам, тенденции ценообразования, популярность игр, их рейтинги и отзывы. Рассмотрены динамики продаж, сегментация аудитории и оценка конкурентоспособности продуктов на платформе Steam. Представлены рекомендации по применению результатов в практической деятельности.

Ключевые слова: анализ данных, онлайн-сервис, Steam, видеоигры.

Введение

Steam — это сервис цифровой дистрибуции видеоигр и магазин компании Valve. Магазин был запущен в сентябре 2003 года как клиент программного обеспечения для автоматического обновления игр компании Valve, а в конце 2005 года был расширен до распространения и продажи игр сторонних издателей. Steam предлагает различные функции, такие как управление цифровыми правами (DRM), организация матчей на игровых серверах, защита от мошенничества, социальные сети и сервисы потоковой трансляции игр. Сервис предоставляет пользователям автоматическое обновление игр, облачную синхронизацию сохраненных игр, а также функции сообщества, такие как обмен сообщениями с друзьями, внутриигровой чат и рынок сообщества.

Во многом успех Steam как платформы объясняется проведением частых распродаж, удобством единой цифровой библиотеки игр и вышеупомянутым перехо-

дом на цифровые игры вместо физических. В то время как другие платформы появляются и набирают обороты, лучшего ресурса для изучения игровой индустрии за последнее десятилетие, пожалуй, не найти. Учитывая это, если нам удастся создать набор данных на основе данных Steam, мы получим доступ к огромному количеству информации о почти 30 000 игр, выпущенных с 2003 года, когда Steam только появился.

Исследование магазина Steam помогает отсортировать и проанализировать данные этой платформы. Такой анализ может помочь игровым разработчикам и издателям лучше понять рынок и принимать более обоснованные решения о ценообразовании для своих игр.

Анализ данных

Сервис Steam предоставляет доступ к своим базам данных посредством наличия наборов интерфейсов программирования приложений (API), предоставляемых Valve Corporation, что в свою очередь позволяет разра-

ботчикам взаимодействовать с платформой Steam для разработки и интеграции различных функций.

Программное обеспечение, разработанное для проведения исследований, состоит из нескольких модулей: GetSteamData, DataCleaning, DataExploration. Модули отвечают за соответствующие функции:

1. **GetSteamData:** загрузка данных из Steam DB по Steam API [1], запись полученных данных в формате csv-файлов.
2. **DataCleaning:** создание копии csv-файлов, оптимизация файлов посредством удаления ненужной информации для последующего анализа и запись в один новый csv-файл [2].
3. **DataExploration:** Построение диаграмм на основе данных, полученных после обработки модулем DataCleaning.

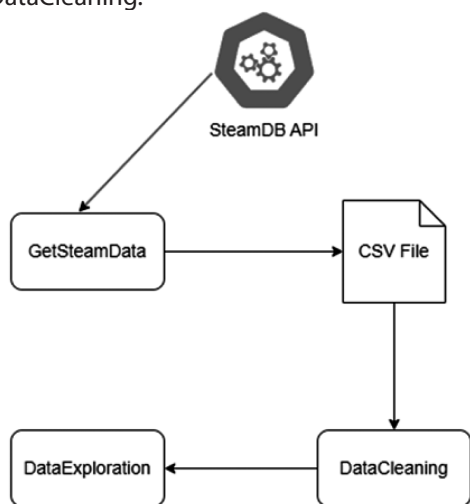


Рис. 1. Архитектура проекта

Проект реализован на языке Python [3] с использованием библиотек: Numpy, Pandas, Seaborn, Matplotlib. Numpy, Pandas используются для обработки и анализа данных. Seaborn, Matplotlib выбраны для реализации визуализации данных.

Данные, полученные посредством Steam API, можно представить в виде таблицы, имеющей 39 столбцов и 20245 строк. Для анализа использовались данные, полученные из SteamDB API за 16 лет (с 2006 по 2022 год). Каждая строчка представляет собой набор открытых данных по отдельной игре. Выделим нужные нам для исследования параметры: название игры, id steam-приложения, необходимый возраст, информация о поддерживаемых платформах, в каких категориях находится игра, жанры игры, игровые достижения, дата выхода, цена игры, поддерживаемые языки, разработчики игры, издатели игры, название игры, количество положительных и отрицательных отзывов количество владельцев игр [4].

Для того чтобы начать работать с данными, нам нужно провести оптимизацию полученных данных. Оптимизация посредством удаления ненужных столбцов полученных данных позволяет упростить процесс загрузки и обработки данных программой. Выберем несколько столбиков, нужных нам для исследований: жанры, владельцы, цена, год и рейтинги. Используя seaborn, мы можем легко просматривать и сравнивать оценку плотности ядра (Kernel Density Estimation) и диаграммы рассеяния, описывающие данные, наблюдая любые закономерности и потенциальные взаимосвязи в данных. Мы применим логарифмическое преобразование к столб-



Рис. 2. Количество релизов по годам в разбивке по количеству владельцев

цам «Владельцы» и «Общий рейтинг», чтобы сравнить их в более удобном масштабе [5].

Исследование зависимости количества релизов в разбивке по количеству владельцев

Проведен анализ данных по параметрам: количество владельцев игры, количество релизов в год. Была построена диаграмма (рис. 2), демонстрирующая информацию о количестве релизов видеоигр и пользователей, обладающих играми. Черным и серым цветами обозначены категории пользователей, приобретших тот или иной релиз видеоигр.

По полученной диаграмме видно, что многие игры в наборе данных имеют от 0 до 20 000 владельцев.

Исследование корреляции нескольких параметров к типу распространения платных/бесплатных игр

Второе, проведенное исследование, позволило построить несколько взаимосвязанных диаграмм (рис. 3)

Для расчета рейтинга игр мы использовали формулу команды SteamDB (1), Формулы расчета общего числа отзывов (2) и оценка по отзывам (3) также представлены ниже:

Пусть P — рейтинг, OO — оценка по отзывам, BO — всего отзывов, $ПО$ — положительные отзывы, $НО$ — негативные отзывы.

$$P = OO - (OO - 0.5) * 2^{-\log_{10}(BO)} \quad \#(1)$$

$$BO = ПО + НО \quad \#(2)$$

$$OO = \frac{ПО}{BO} \quad \#(3)$$

Очевидно, что при 0 отзывах мы на 100 % не уверены в том, каким должен быть рейтинг. Давайте немного растянем это и скажем, что мы на 100 % не уверены даже в 1 отзыве, тогда мы можем применить предыдущую мысль. При 10 отзывах мы должны быть в 2 раза менее неопределенными, то есть 50 % неопределенности. При 100 отзывах — 25 %. При 1000 обзорах — 12,5 % и так далее.

Учитывая, что у игры 100 отзывов, из которых 90 % положительные, мы на 25 % не уверены, что эти 90 % — правильный рейтинг. Мы на 75 % уверены, что это правильный рейтинг. Другими словами, 75 % итогового рейтинга — это 90 %, а остальные 25 % — это средний рейтинг 50 %, что также хорошо согласуется с нашим вторым правилом. Таким образом, мы получаем итоговый рейтинг $75\% * 90\% + 25\% * 50\% = 80\%$. Это выглядит хорошо, и эти правила можно перевести в формулу, которая дает нам скорректированный рейтинг с учетом количества отзывов и «реального рейтинга».

На основе этой визуализации мы можем сделать несколько первых наблюдений:

- Платные игры в среднем имеют более высокий рейтинг, чем бесплатные.
- Рейтинги бесплатных игр распределены по более широкому диапазону.
- Похоже, что владельцев платных игр может быть больше, чем бесплатных, но это неясно.
- Платные игры получают чуть больше оценок, чем бесплатные.
- Чем больше владельцев игры, тем больше у нее оценок.
- Оценки игр с большим количеством владельцев в среднем выше или, по крайней мере, распределены в меньшем и большем диапазоне.
- Если посмотреть на год выпуска, то можно заметить, что в последние годы наблюдается сдвиг в сторону более бесплатных игр.
- Более поздние игры стоят дороже, но обратите внимание, что все цены — это текущие цены, а не цены на момент выхода.

Для оценки корреляции рейтинга от количества владельцев игры и типа распространения, была построена дополнительная диаграмма (рис. 4).

Разбор распределения рейтинговых баллов игр с 20000 владельцами показал частоту и значения рейтинга оценок (рис. 5). Здесь следует обратить внимание на два важных момента. Во-первых, средний рейтинг платных игр значительно ниже, если учитывать все данные, а не только игры с 20000 и более владельцами. Во-вторых, средние оценки платных игр действительно ниже средних оценок бесплатных игр при учете всех данных, тогда как для игр с 20000 и более владельцами оценки платных игр заметно выше.

По диаграмме (рис. 5) видно, что большинство оценок обладают значениями рейтинга в 80 %. Негативные отзывы сильно реже положительных. Возможно, это связано с несколькими факторами:

1. Качество контента: В Steam часто предлагают игры, которые прошли тщательный отбор, что снижает вероятность того, что игра окажется совсем плохой. Разработчики проходят через процесс модерации, и, как правило, только качественные игры попадают в магазин.
2. Отзывы от игроков: Steam предоставляет платформу для пользовательских отзывов, что помогает другим игрокам принимать более обоснованные решения о покупке. Положительные отзывы могут привлечь больше внимания, чем отрицательные, особенно если они содержат конструктивную критику.
3. Взаимодействие с сообществом: Многие игры на Steam активно взаимодействуют с сообще-

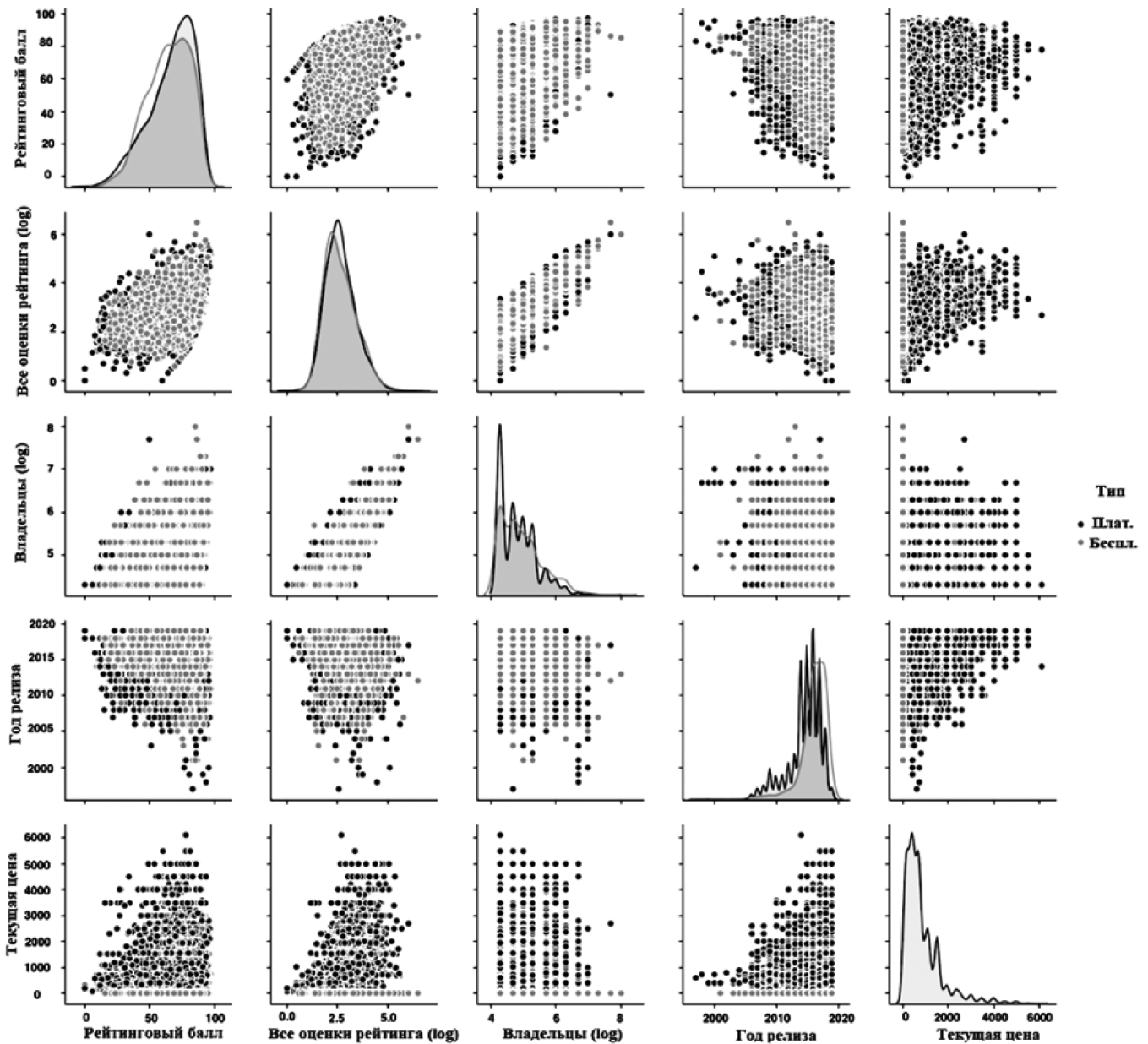


Рис. 3. Визуализация корреляции нескольких параметров к типу распространения платных/бесплатных игр

ством игроков, внимательно слушая их отзывы и внося изменения в игру в соответствии с пожеланиями. Это может привести к тому, что даже если игра имеет недостатки, активная поддержка разработчиков может улучшить общее впечатление.

- Продвижение хороших игр: Steam может активно продвигать игры, которые получили много положительных отзывов, делая их более видимыми для потенциальных покупателей. Это может повысить общую долю положительных отзывов на платформе.

В целом, сочетание этих факторов может привести к тому, что на Steam в среднем больше положительных отзывов, чем отрицательных.

Исследование корреляции жанра игры от количества владельцев и процента популярности. По данным, представленным на диаграмме (рис. 6), можно сразу отметить лидирующие позиции жанров: «Инди» по проценту популярности и «Массовый мультиплеер» по среднему количеству владельцев на игру.

Четверку самых популярных по частоте создания новых игр замыкают жанры: «Экшен», «Приключенческие», «Казуальные». «Массовый мультиплеер» несмотря на свою популярность среди пользователей, не обладает высокой частотой пополнения новыми играми. Возможно, это обуславливается сложностью разработки игр этого жанра.

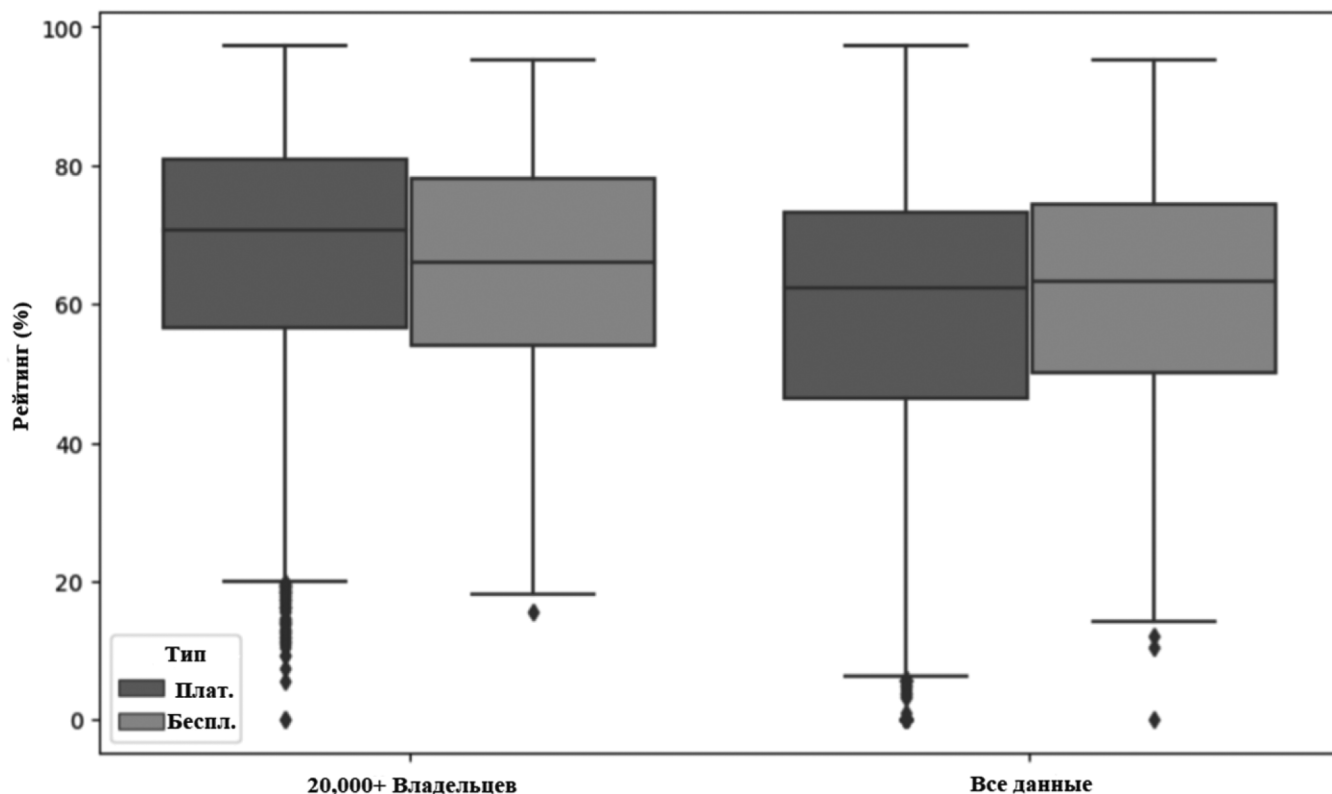


Рис. 4. Визуализация рейтинга платных/бесплатных игр
Распределение рейтинговых баллов платных игр с 20 000 владельцами

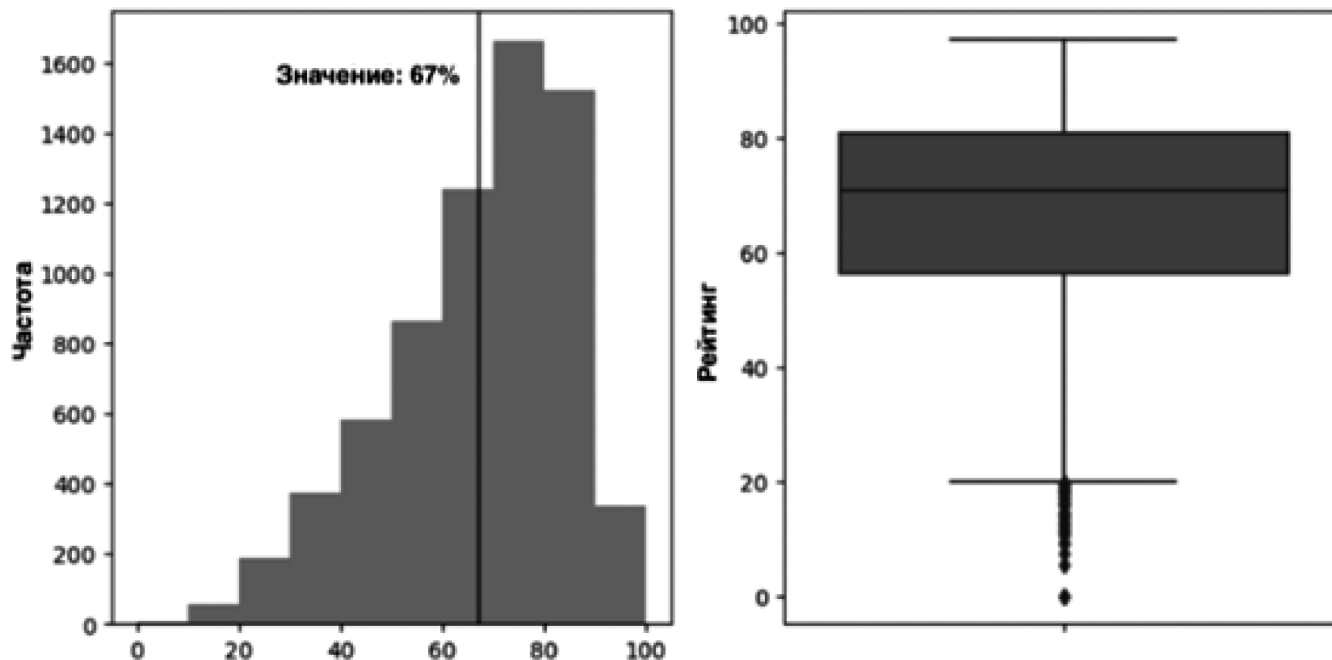


Рис. 5. Диаграмма распределения рейтинговых баллов платных игр с 20,000 владельцами

Исследование зависимости стоимости игры от жанра

По диаграмме видно частоту попадания цен на то или иное значение. Цена большинства игр находится в пре-

делах от 0 до 1000 рублей, мультиплееры игры вообще бьют рекорды по низкой цене за игру, возможно, разработчики зарабатывают не на продаже самой игры, а на событиях внутри игры.

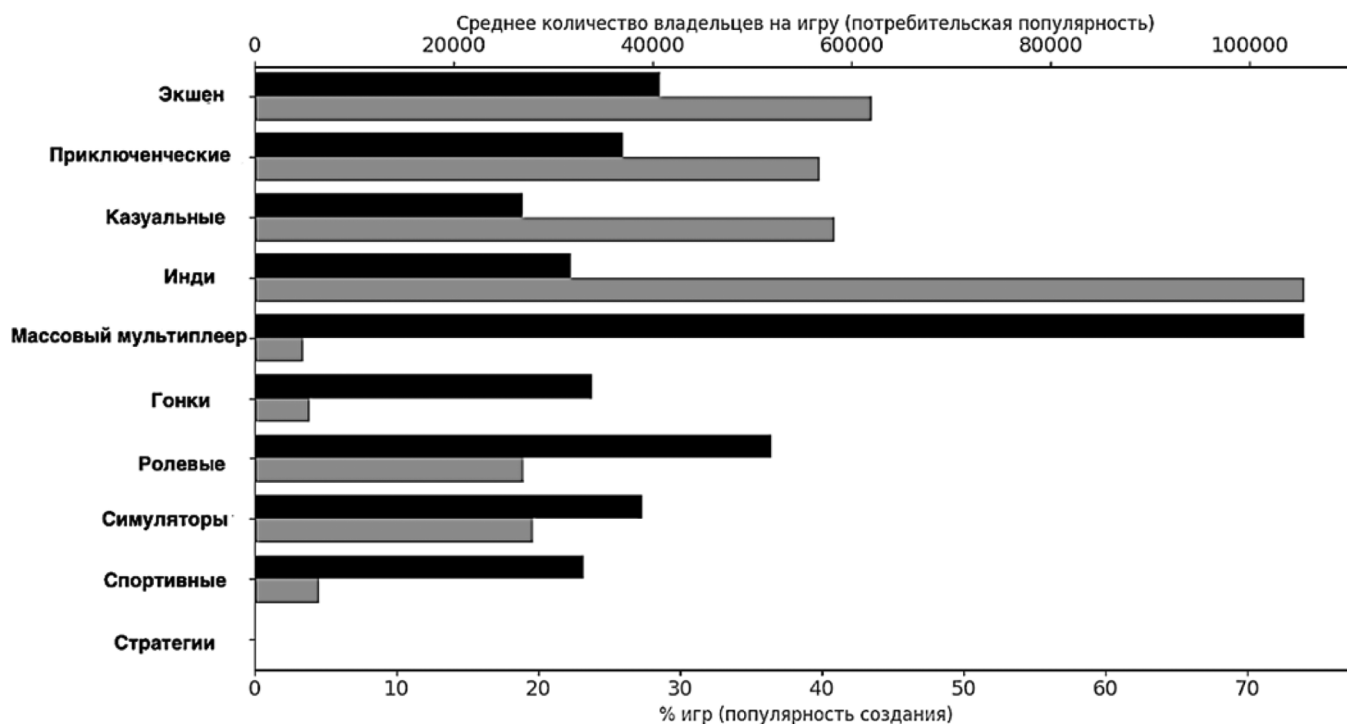


Рис. 6. Диаграмма корреляции жанра игры от количества владельцев и процента популярности

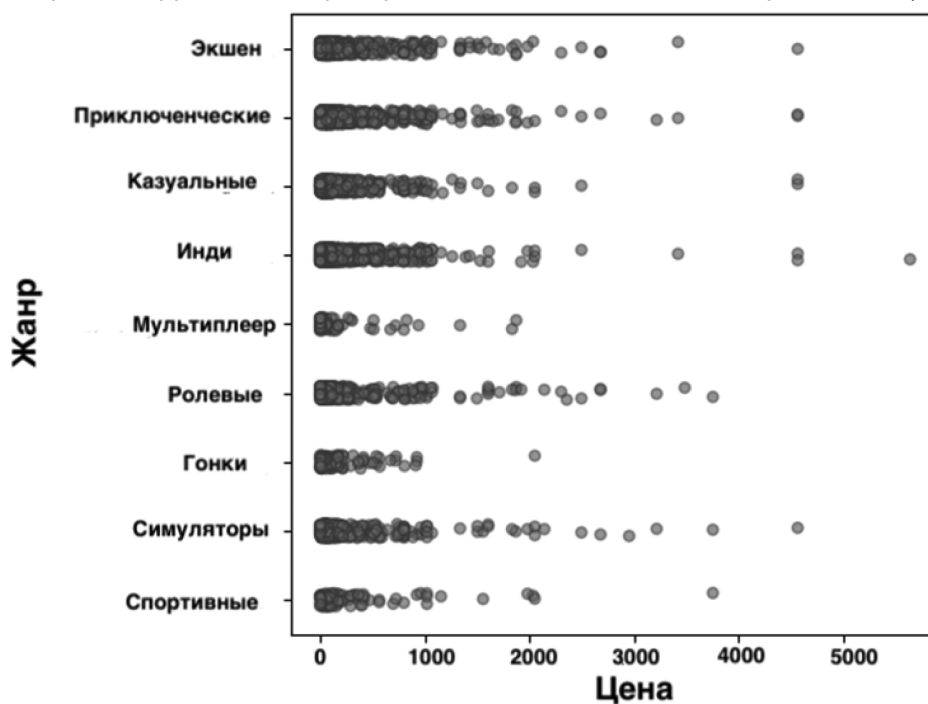


Рис. 7. Диаграмма зависимости стоимости игры от жанра

Заключение

Данные, полученные в ходе анализа магазина Steam, обладают высокой востребованностью и имеют широкое применение. Они могут применяться для решения самых разных задач:

- Понимание рынка: Исследование предпочтений потребителей, ценовой политики и популярности

различных жанров игр на Steam помогает понять текущие тенденции и направления развития игровой индустрии.

- Принятие решений о разработке игр: Разработчики могут использовать данные о продажах игр, их жанрах и ценах для принятия решений о том, какие игры разрабатывать, на каких платформах выпускать и как ценить свои продукты.

- Оптимизация маркетинга и продвижения: Изучение успешных игр и их характеристик на Steam помогает понять, какие стратегии маркетинга и продвижения могут быть наиболее эффективными для привлечения внимания к новым играм.
- Повышение конкурентоспособности: Понимание того, какие игры наиболее успешны на рынке Steam, может помочь разработчикам создавать более конкурентоспособные продукты и лучше позиционировать их на рынке.
- Определение ценовой стратегии: Анализ цен на игры в зависимости от их жанра и других характеристик помогает разработчикам опреде-

лить оптимальную ценовую стратегию для своих продуктов.

- Инвестиционные решения: Инвесторы и издатели могут использовать данные о продажах и популярности игр на Steam для принятия решений о финансировании проектов и предсказания потенциальной доходности инвестиций.

В целом, анализ магазина Steam позволяет получить ценную информацию о состоянии и динамике игрового рынка, что помогает разработчикам, издателям и инвесторам принимать более обоснованные решения в своей деятельности.

ЛИТЕРАТУРА

1. This third-party website gives you better insight into the Steam platform and everything in its database. URL: <https://steamdb.info/>
2. Implementation and Analysis of Algorithms for Pitch Estimation in Musical Fragments / N.V. Voinov, D.A. Ivanov, T.V. Leontieva, S.A. Molodyakov // Proceedings of 2021 24th International Conference on Soft Computing and Measurements, SCM 2021 : 24, St. Petersburg, 26–28 мая 2021 года. — St. Petersburg, 2021. — P. 113–116. — DOI 10.1109/SCM52931.2021.9507134. — EDN CEIXTQ.
3. Python for data analysis / Mrs Deepa R., Dr. Ravikumar G.K., Ms Kavitha H.M., Mrs Divya B. M. // IJRAR — International Journal of Research and Analytical Reviews (IJRAR), E-ISSN 2348-1269, P— ISSN 2349-5138, Volume.7, Issue 1, Page No pp.740–743, March 2020
4. Data Analysis with the SolarSoft System / S. L. Freeland, B.N. Handy // Solar Physics 182, October 1998. — P. 497–500. — DOI 10.1023/A: 1005038224881
5. Big data processing system for analysis of GitHub events / N. Voinov, K. Rodriguez Garzon, I. Nikiforov, P. Drobintsev // Proceedings of 2019 22nd International Conference on Soft Computing and Measurements, SCM 2019 : 22, St. Petersburg, 23–25 мая 2019 года. — St. Petersburg, 2019. — P. 187–190. — DOI 10.1109/SCM.2019.8903782.

© Гладун Владимир Вадимович (vladimir.gldn@gmail.com); Пантюхин Андрей Максимович (panandafog@gmail.com);
Малинин Илья Игоревич (malinin.ilja@gmail.com); Молодяков Сергей Александрович (molodyakov_sa@spbstu.ru)
Журнал «Современная наука: актуальные проблемы теории и практики»