



Проекты игр с Кикстартера



Описание базы данных

- » База данных включает в себя обработанную информацию о проектах игр с 2009 по 2016 год. Датасет в формате таблицы сформирован на основе проектов, опубликованных на сайте www.kickstarter.com. Данные не содержат пропущенных значений
- » Данные были собраны с сайта www.kickstarter.com с помощью web-scraping методов и обработаны для преобразования в машиночитаемый формат. Каждая строка в таблице соответствует уникальному проекту (игре). Всего в датасете 8421 проект и для каждого проекта в таблице включено 37 характеристик (столбцов)



Поля таблицы 1



Финансовые переменные (указаны в долларах):

- goal - сколько проект планировал собрать средств
- pled – сколько средств проект собрал в итоге .
- min_pled_t, mean_pled_t, max_pled_t - информация о величине пожертвований от пользователей
- n_pled_t, step_pled_t - количество опций пожертвований и средняя разница между ними



Переменные, связанные с оформлением страницы проекта на сайте:

- text_am – количество слов в на странице
- n_img, n_vid – количество картинок и видео на странице
- mgr – суммарное количество пикселей у фотографий на странице (в миллионах)
- cont – средняя контрастность изображений на странице (EBImage, среднее среди стандартных отклонений по трем каналам)



Бинарные переменные, показывающие наличие тега в описании игры, символизирующие жанровую принадлежность (прокси для жанра игры):

- rpg, platformer, shooter, fighting, survival, horror, strategy, arcade, simulator, mmo, indie, action, quest, adventure,
- Уточнение: некоторые проекты имеют несколько тегов (до 10), а некоторые игры не имеют ни одного тега



Поля таблицы 2



Другие переменные

- URL – ссылка на страницу проекта на Kickstarter
- date – дата опубликования страницы проекта на Kickstarter
- cr_time – время со дня регистрации аккаунта автора проекта (прокси опыта автора)
- backed – количество проектов профинансированных автором проекта на текущий период
- created – количество проектов профинансированных автором проекта до опубликования этого проекта
- site – ссылка на сайт проекта (если у проекта есть сайт, если сайта нет, то значение пустое)
- is_site – бинарная переменная, показывающая наличие сайта у проекта
- curr – валюта в которой производится сбор средств
- period – продолжительность онлайн компании (сбора средств), в днях
- success – бинарная переменная, показывающая, собрала ли компания нужное количество средств (или больше) или нет



Дополнительные детали по базе



При прогнозировании успешности проекта стоит учитывать только информацию, доступную в момент опубликования проекта. Иначе прогнозирование теряет смысл.



Стоит отметить, что на сайте присутствуют компании, собиравшие средства в одиннадцати разных валютах. Все показатели, измеряющиеся в денежных единицах были конвертированы в доллары с использованием спотового обменного курса на дату запуска проекта. Аналогичный метод используется администрацией сайта Kickstarter для построения собственной статистики. Данные по ежедневному обменному курсу были предоставлены официальной базой данных Банка Англии через информативную платформу Quandl.



Примеры исследований



Сколько слов в описании проекта необходимо для максимальной успешности проекта?



Есть ли то количество видео, после которого публикация новых видеороликов не влияет на успешность проекта?



Проявление сезонности успешности проекта: влияет ли дата публикации проекта на его успешность?



Влияние тегов в описании на успешность тегов (влияние каждого тега по отдельности и/или влияние количества тегов)



Влияние успешности проекта на активность автора проекта в дальнейшем