

Задание по EDA: Анализ инвестиционной стратегии в игровую индустрию

Задача:

Вы — аналитики инвестиционного фонда, который планирует вложить 10 миллионов долларов в игровую индустрию. Ваша задача — определить, как распорядиться с этими деньгами, в какие игры / игру их инвестировать (имеется в виду портрет новой игры, а не инвестиции в какую-то текущую игру)

Для этого вам нужно провести исследовательский анализ данных (EDA), найти закономерности в успешных играх и дать обоснованные рекомендации

Описание колонок датасета

Основные идентификационные данные

- **appid** — Уникальный идентификатор игры
- **name** — Название игры.
- **release_date** — Дата выхода игры в формате YYYY-MM-DD.

Информация о разработчике и издателе

- **developer** — Название студии-разработчика (например, Valve, Ubisoft).
- **publisher** — Название издателя игры (например, Electronic Arts, Bethesda). Может совпадать с разработчиком.

Языковые и возрастные ограничения

- **english** — Бинарный признак (1 — игра поддерживает английский язык, 0 — нет).
- **required_age** — Возрастное ограничение игры (например, 18+).

Оценки и отзывы пользователей

- **positive_ratings** — Количество положительных отзывов пользователей в Steam.
- **negative_ratings** — Количество отрицательных отзывов пользователей.
- **achievements** — Количество достижений, доступных в игре.

Игровые механики

- **multiplayer** — Бинарный признак (1 — игра поддерживает многопользовательский режим, 0 — нет).
- **singleplayer** — Бинарный признак (1 — игра поддерживает одиночный режим, 0 — нет).

Игровые жанры и теги

- **Жанры с большой буквы**
- **Пользовательские теги с маленькой буквы:** (сколько раз и какой тег проставляли пользователи)

Технические параметры

- **windows** – Бинарный признак (1 – игра доступна на Windows, 0 – нет).
- **mac** – Бинарный признак (1 – игра доступна на MacOS, 0 – нет).
- **linux** – Бинарный признак (1 – игра доступна на Linux, 0 – нет).

Финансовые показатели

- **price** – Цена игры в долларах на момент выгрузки данных.
- **owners_num** – Количество владельцев игры (сколько людей купили или получили игру).
- **budget** – Оценка бюджета разработки игры.

Игровая активность

- **average_playtime** – Среднее время, проведенное в игре пользователями (в минутах).
- **median_playtime** – Медианное время, проведенное в игре пользователями (в минутах).

Требования к анализу:

- Переформулировка задачи **из описания в плоскость метрик** (какую игру вы ищете)
- **Не менее 5 инсайтов** о том, какая игра наиболее перспективна.
- **Выводы должны быть видны из графиков и прописаны в виде комментариев** (барплоты, scatter-плоты, heatmap и т.д.).
- **Как минимум 1 новая метрика**, введенная в ходе анализа
- **Не менее 1 вывода на примере графика Box plot** (из распределений, которые вы увидите на нем)
- Все выводы должны иметь отношение к итоговой цели анализа, графики и метрики посчитанные без связи с основной целью не засчитываются

Формат сдачи работы

- **Файл с расчетами**

- **ИЛИ Python file** с чистым кодом и комментариями. Все графики должны иметь подписи осей, самих графиков (штраф 2 балла)
 - **ИЛИ Orange file** с комментариями (там есть оставлять поле для комментария – не найдете – пишите)
- **Презентация** в формате: графики и ключевые выводы из них со связью с бизнес задачами. При необходимости вынесете на слайды дополнительную информацию. По презентации должна прослеживаться логика, как удалось прийти к выводам, которые вынесены на финальный слайд

Финальный слайд – список ключевых выводов и рекомендаций