

Описание выполнения тестового задания по вакансии менеджер-аналитик

1. Используемый датасет: В качестве основы для анализа был взят датасет "Google Play Store Apps" с платформы Kaggle (ссылка: <https://www.kaggle.com/datasets/lava18/google-play-store-apps/data>). Этот набор данных состоит из двух основных таблиц:
* googleplaystore.csv: содержит подробную информацию о приложениях в Google Play Store (категория, рейтинг, количество отзывов, размер, количество установок, тип (платное/бесплатное), цена, возрастной рейтинг, жанры, дата последнего обновления, текущая и минимальная версии Android).
* googleplaystore_user_reviews.csv: содержит отзывы пользователей, включая переведенный отзыв, его эмоциональную окраску (Sentiment), полярность и субъективность.

2. Анализ с применением Excel:

• Power Query:

Загрузка и объединение данных: Обе таблицы (googleplaystore.csv и googleplaystore_user_reviews.csv) были загружены в Power Query. Таблица googleplaystore_user_reviews.csv была предварительно обработана для удаления строк с некорректными данными или неполными отзывами.

Очистка и преобразование данных:

- В основной таблице приложений (googleplaystore.csv) были удалены дубликаты строк.
- Столбец "Installs" был очищен от символов '+' и ',' и преобразован в числовой тип данных.
- Столбец "Price" был очищен от символа '\$' и преобразован в числовой тип данных. Для бесплатных приложений цена была установлена как 0.
- Пропуски в данных (например, в столбце "Rating") были удалены
 - **Сводные таблицы:** На основе подготовленной и объединенной таблицы были созданы следующие сводные таблицы для выявления закономерностей:
 1. **Группировка приложений по ценовым диапазонам:** Исходные данные о ценах приложений были сгруппированы в определенные ценовые категории (например, 0, 0.99, 1-1.99, 2-2.99 и так далее). **Расчет количества установок для каждой ценовой категории:** Для каждой выделенной ценовой категории было суммировано общее количество установок приложений, попадающих в этот диапазон. **Визуализация распределения:** На основе полученных данных была построена гистограмма или столбчатая диаграмма, отображающая зависимость между ценовой категорией (по оси X, в логарифмическом масштабе) и средним количеством установок (по оси Y). Логарифмическая шкала по оси для наглядного представления данных, в которых значения цен могут сильно различаться.
 2. **Расчет среднего рейтинга по категориям:** Исходные данные о рейтингах приложений были сгруппированы по соответствующим категориям (например, EDUCATION, ART_AND_DESIGN, BOOKS_AND_REFERENCE и другие). Для каждой категории было рассчитано среднее значение рейтинга на основе оценок всех

приложений, входящих в эту категорию. **Визуализация среднего рейтинга:** Полученные средние рейтинги для каждой категории были визуализированы с использованием столбчатой диаграммы (гистограммы).

- **ВПР (VLOOKUP):** Хотя основное объединение данных могло быть выполнено с помощью Power Query (что является более мощным инструментом для таких задач), функция ВПР была использована для добавления столбца «Category» в таблицу Reviews

3. Выводы по результатам анализа:

Анализ распределения приложений по ценовым категориям показал, что бесплатные приложения занимают доминирующее положение на рынке, обеспечивая наибольшее количество установок. Среди платных приложений оптимальным ценовым диапазоном, демонстрирующим хорошее соотношение цены и количества установок, является 2.49-2.99 \$. При этом установлено, что цены выше 10 \$ приводят к существенному снижению числа установок, а цены, превышающие 30 \$, практически не пользуются спросом. Рекомендуется провести дополнительное A/B-тестирование цен в диапазонах 4.99 \$ и 6.99 \$, где наблюдается потенциально высокая конверсия, а также углубленно изучить приложения с аномально высокой ценой в 399.99 \$ и ценовые точки 9, 13.99, 24.99 \$ для выявления специфических факторов.

Анализ среднего рейтинга приложений по категориям выявил лидеров по оценкам: EVENTS, EDUCATION и ART_AND_DESIGN, средний рейтинг которых превышает 4.35. Категории DATING, TOOLS и MAPS_AND_NAVIGATION демонстрируют более низкие средние рейтинги (ниже 4.05). Общий средний рейтинг по рынку составляет 4.17. На основании полученных данных рекомендуется изучить успешные категории для перенимания лучших практик, сосредоточиться на улучшении приложений в категориях с низкими рейтингами и внимательно отслеживать оценки после выпуска обновлений.