

Отзывы с Amazon Reviews разбиты на два типа файлов(два типа матриц) - User Reviews и Item Metadata. Приведем описание каждого из них по отдельности.

User Reviews:

Название поля	Тип данных	Объяснение
rating	float	Рейтинг от 1.0 до 5.0
title	str	Заголовок пользовательского отзыва
text	str	Основной текст пользовательского отзыва
images	list	Изображения продукта, которые выложил пользователь после его получения. Каждое изображение имеет свой размер(маленькое, среднее и большое), каждый размер представлен соответствующей ссылкой: small_image_url, medium_image_url, и large_image_url
asin	str	ID продукта
parent_asin	str	Родительский ID продукта. Замечание: продукты с различными цветами, стилями, размерами обычно относятся к одному и тому же parent ID. "asin" в предыдущих датасетах амазона и есть parent ID. Стоит использовать именно parent ID для поиска метаданных продукта.
user_id	str	ID пользователя
timestamp	str	Время публикации отзыва(в unix формате)
verified_purchase	bool	Подтверждение покупки товара пользователем
helpful_vote	int	Голоса полезности на отзыве

Item Metadata:

Название поля	Тип данных	Объяснение
main_category	str	Главная категория(т.е. домен) продукта
title	str	Название продукта
average_rating	float	Рейтинг продукта, показанный на странице продукта
rating_number	int	Количество оценок продукта
features	list	Буллет-список свойств продукта
description	list	Описание продукта
price	float	Цена продукта в долларах США(на момент сбора данных)
images	list	Изображения товара. Каждое изображение имеет разный размер (большой, увеличенный, увеличенный в размерах). В поле "вариант" отображается положение изображения.
videos	list	Видео продукта. Состоит из названия и ссылки.
store	str	Имя продукта в хранилище
categories	categories	Иерархические категории продукта
details	dict	Описание продукта, включающие в себя материалы, бренд, размер и прочее
parent_asin	str	parent ID продукта
bought_together	list	Рекомендуемые пакеты(дополнения) с веб-сайтов