

مقدمه:

در این پروژه ما از داده‌های مربوط به تعداد زیادی بازی استفاده می‌کنیم. که این داده‌ها شامل اطلاعاتی است مانند:

- امتیاز کلی هر بازی
- امتیاز کسب شده از جانب خریداران هر بازی
- میزان فروشی هر بازی
- متوسط زمان لازم برای به پایان رساندن هر بازی
- سال تولید هر بازی
- ژانر هر بازی
- و ...

حال هدف ما این است که به چندین مساله پاسخ دهیم تا طراحان بازی به کمک آنها بتوانند بازی‌های محبوب تر یا پرفروش‌تری تولید کنند. این مسائل عبارتند از :

مساله ۰: ابتدا باید سنجیده شود، بهترین داده برای استنتاج میزان محبوبیت (امتیاز) یک بازی کدام است، در نتیجه مساله صفرم برابر است با: آیا نظر اولیه کاربران، برای استنتاج داده معتبری است ؟

مساله ۱: آیا ژانرهای متفاوت محبوبیت (امتیاز) متفاوتی دارند؟ اگر بله، محبوب ترین ژانرها کدامند؟

مساله ۲: آیا در سال های اخیر، فروش بازی های کامپیوتری افزایش داشته است؟

مساله ۳: آیا بین میزان محبوبیت (امتیاز) و میزان فروش رابطه خطی وجود دارد؟ اگر بله، یک پیشبینی برای میزان فروش بر حسب محبوبیت ارائه می‌دهیم.

مساله ۴: آیا مدت زمان لازم برای به پایان رساندن یک بازی روی محبوبیت (امتیاز) آن تاثیر گذار است؟

منابع:

داده‌ها (data set) اصالتاً از سایت کاگل (kaggle) گرفته و در مسابقه هفتگی (tidytuesday) استفاده شده است و ما از آنجا به داده‌ها دسترسی پیدا کردیم.

- اطلاعات مربوط به امتیاز دهی و فروش هر بازی:

<https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2022/2022-01-20/ratings.csv>

- اطلاعات مربوط به طراحی هر بازی (متصل به جدول بالا از طریق آیدی هر بازی):

<https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2022/2022-01-20/details.csv>

نویسندگان:

- اشکان زرخواه ۶۱۰۳۹۹۱۹۶
- دلارام فرتوت ۶۱۰۳۹۹۱۵۳