**Descrição do projeto**

Você trabalha para a loja online Ice, que vende videogames no mundo todo. As avaliações de usuários e especialistas, gêneros, plataformas (por exemplo, Xbox ou PlayStation) e dados históricos sobre vendas de jogos estão disponíveis em fontes abertas. Você precisa identificar padrões que determinam se um jogo tem sucesso ou não. Isso vai permitir que você identifique possíveis sucessos e planeje campanhas publicitárias.

Os dados disponibilizados remontam a 2016. Vamos imaginar que estamos em dezembro de 2016 e você está planejando uma campanha para 2017.

(O importante é ter experiência trabalhando com dados. Realmente não importa se você está prevendo as vendas de 2017 com base nos dados de 2016 ou as vendas de 2027 com base nos dados de 2026.)

O conjunto de dados contém uma coluna de "rating" (classificação) que armazena a classificação ESRB de cada jogo. O Entertainment Software Rating Board avalia o conteúdo de um jogo e atribui uma classificação etária, como Teen (Adolescente) ou Mature (Adulto).

**Instruções para concluir o projeto**

Etapa 1. Abra o arquivo de dados e estude as informações gerais

Caminho do arquivo:

*/datasets/games.csv* . [Download dataset](https://practicum-content.s3.us-west-1.amazonaws.com/datasets/games.csv)

Etapa 2. Prepare os dados

* Substituir os nomes das colunas (transformar tudo em minúsculos).
* Converta os dados para os tipos necessários.
* Descreva as colunas onde os tipos de dados foram alterados e por quê.
* Se necessário, decida como lidar com valores ausentes:
  1. Explique por que você preencheu os valores ausentes da forma que você fez ou por que você decidiu deixá-los em branco.
  2. Por que acha que os valores estão ausentes? Dê possíveis razões.
  3. Preste atenção à abreviação TBD (a ser determinada). Especifique como pretende lidar com esses casos.
* Calcule o total de vendas (a soma das vendas em todas as regiões) para cada jogo e coloque esses valores em uma coluna separada.

Etapa 3. Analise os dados

* Veja quantos jogos foram lançados em anos diferentes Os dados de cada período são significativos?
* Veja como as vendas variaram de plataforma para plataforma. Escolha as plataformas com as maiores vendas totais e construa uma distribuição com base em dados para cada ano. Encontre as plataformas que costumavam ser populares, mas agora não têm vendas. Quanto tempo leva para as novas plataformas aparecerem e as antigas desaparecerem?
* Determine para qual período você deve pegar dados. Para fazê-lo, olhe para suas respostas para as perguntas anteriores. Os dados te deveriam permitir construir um modelo para 2017.
* Trabalhar apenas com os dados que você decidiu que são relevantes. Desconsidere os dados de anos anteriores.
* Quais plataformas estão liderando em vendas? Quais estão crescendo ou diminuindo? Selecione várias plataformas potencialmente lucrativas.
* Construa um diagrama de caixa para as vendas globais de todos os jogos, divididos por plataforma. As diferenças nas vendas são significativas? E quanto às vendas médias em várias plataformas? Descreva suas descobertas.
* Veja como as avaliações de usuários e profissionais afetam as vendas de uma plataforma popular (você escolhe). Construa um gráfico de dispersão e calcule a correlação entre revisões e vendas. Tire conclusões.
* Com essas conclusões em mente, compare as vendas em outras plataformas.
* Dê uma olhada na distribuição geral de jogos por gênero. O que podemos dizer sobre os gêneros mais lucrativos? Você pode generalizar sobre gêneros com vendas altas e baixas?

Etapa 4. Crie um perfil de usuário para cada região

Para cada região (AN, UE, JP), determine:

* As cinco plataformas principais. Descreva as variações das suas quotas de mercado de região para região.
* Os cinco principais gêneros. Explique a diferença.
* As classificações do ESRB afetam as vendas em regiões individuais?

Etapa 5. Teste as seguintes hipóteses:

— As classificações médias dos usuários das plataformas Xbox One e PC são as mesmas.

— As classificações médias de usuários para os gêneros Action (ação) e Sports (esportes) são diferentes.

Defina o valor do limiar *alfa* você mesmo.

Explique:

— Como você formula as hipóteses alternativas e nulas.

— Quais critérios você usou para testar as hipóteses, e por quê

Etapa 6. Escreva uma conclusão geral

Formato: Complete a tarefa no Jupyter Notebook. Insira o código nas células *code* e textos de explicação nas células *markdown*. Aplique a formatação e adicione títulos.

**Descrição de dados**

—*Name* (nome)

—*Platform* (plataforma)

—*Year\_of\_Release* (Ano de lançamento)

—*Genre* (gênero)

—*NA\_sales* (vendas norte-americanas em milhões de USD)

—*EU\_sales* (vendas na Europa em milhões de USD)

—*JP\_sales* (vendas no Japão em milhões de USD)

—*Other\_sales* (vendas em outros países em em milhões de USD)

—*Critic\_Score* (Pontuação crítica) (máximo de 100)

—*User\_Score* (Pontuação do usuário) (máximo de 10)

—*Classificação* (ESRB)

Os dados de 2016 podem estar incompletos.

Como o meu projeto será avaliado?

Leia atentamente esses critérios de avaliação do projeto antes de começar a trabalhar.