**Sprint 6 - Projeto**

Parabéns! Você concluiu oficialmente a primeira parte do programa na plataforma interativa. Agora é hora de reunir tudo o que você aprendeu até agora em seu primeiro projeto integrado, um estudo de caso analítico da vida real.

Este projeto vai abranger todas as habilidades que você desenvolveu até este momento:









Quando terminar o projeto, envie seu trabalho para ser avaliado na revisão. Você vai receber feedback dentro de 24 horas. Use o feedback para fazer alterações e, em seguida, envie a nova versão de volta ao revisor do projeto.

Você talvez receba mais feedback sobre a nova versão. Isso é completamente normal. Não é incomum passar por vários ciclos de feedback e revisão.

Seu projeto será considerado concluído assim que o revisor do projeto o aprovar.

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 ganhar experiência trabalhando com dados. 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

* Substitua os nomes das colunas (transforme tudo em minúsculos).
* Converta os dados para os tipos necessários.
* Descreva as colunas onde os tipos de dados foram alterados e explique o motivo.
* Se necessário, decida como lidar com valores ausentes: 1. Explique sua abordagem ao preencher valores ausentes ou ao deixá-los em branco. 2. Na sua opinião, por que os valores estão ausentes? Dê possíveis razões. 3. Preste atenção à abreviação TBD (que significa "to be determined", ou "a ser determinado"). 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 a cada ano. 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 nos dados para cada ano. Encontre as plataformas que costumavam ser populares, mas que 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 fazer isso, olhe para suas respostas das perguntas anteriores. Os dados devem permitir que você construa 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 das plataformas populares (você escolhe). Construa um gráfico de dispersão e calcule a correlação entre avaliações e vendas.
* Tendo suas conclusões em mente, compare as vendas dos mesmos jogos 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 por conta própria o valor do limiar *alfa*.

Explique:

— Como você formuoua as hipóteses alternativas e nulas.

— Qual o nível de significância que você escolheu para testar as hipóteses, e por quê.

Etapa 6. Escreva uma conclusão geral

Formato: conclua 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 dos 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 milhões de USD)

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

—*User\_Score* (Pontuação dos usuários) (máximo de 10)

—*Classificação* (ESRB)

Os dados de 2016 podem estar incompletos.

Como o meu projeto será avaliado?

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

Veja o que os revisores vão analisar ao avaliar seu projeto:

* Como você descreve os problemas que identifica nos dados?
* Como você prepara um conjunto de dados para análise?
* Como você constrói gráficos de distribuição e como você os explica?
* Como você calcula o desvio padrão e a variância?
* Você formula hipóteses alternativas e nulas?
* Que métodos você aplica ao testá-los?
* Você explica os resultados dos testes de hipóteses?
* Você segue a estrutura do projeto e mantém seu código limpo e compreensível?
* Quais conclusões você tira?
* Você deixou comentários claros e relevantes a cada etapa?

Tudo o que você precisa para concluir esse projeto está nas folhas de conclusões e resumos de capítulos anteriores. Boa sorte!