



Desafio | Estágio em Ciência de dados

2021

Instruções – Desafios Keycash

Olá! Estamos felizes que tenha decidido fazer nosso desafio! ;)

Neste documento, você encontrará:

- Um desafio de **banco de dados** com duas atividades.
 - Um desafio de **modelagem**.
- Um desafio bônus(opcional) de **disponibilização de modelos**.

Ao finalizar, você deve enviar um e-mail com os resultados e também o seu CV para: vempira@keycash.io

Não se esqueça: se tiver outros projetos/materiais no **Kaggle**, **Github**, **Blogs/Medium**, etc... Pode nos enviar os links também!!

Boa sorte!



Desafio 1 – Manipulando bases de dados (SQL)

Dados são o CORE do nosso negócio e todos os dias precisamos manipulá-los de forma eficiente e segura.

A forma mais comum de manipular dados é através de Queries SQL.



Como um Cientista de Dados na Keycash, você deve ser capaz de navegar pelos nossos bancos de dados com conforto.

Vamos ao seu **primeiro desafio!**



Desafio 1 – Manipulando bases de dados (SQL)

Atividade A – Os passos devem ser feitos no formato de uma Query em SQL, na qual você irá nos enviar (em forma de texto ou script SQL)

Caso queira utilizar um software open-source (ex: MySQL) disponibilizamos um script chamado Desafio1_imports.sql **apenas para importação dos dados**. Mas não se preocupe! Não é obrigatório para a resolução do exercício ;)

Seu **desafio** 1 será composto de **duas** atividades:

Na **primeira**, você deve utilizar os dados da tabela **Desafio1_imoveis.csv** para:

1. Calcular o valor do m^2 (**Preço/área**) por imóvel;
2. Calcular a **média** do **valor do m^2** para cada bairro;
3. **Ordenar** o resultado de forma **decrescente** e **filtrando** apenas os bairros com IDs iguais a **1 e 3**;
4. A tabela resultante deve se chamar **valor_medio**;



Desafio 1 – Manipulando bases de dados (SQL)

Atividade B – Os passos devem ser feitos no formato de uma Query em SQL, na qual você irá nos enviar (em forma de texto ou script SQL)

Caso queira utilizar um software open-source (ex: MySQL) disponibilizamos um script chamado Desafio1_imports.sql **apenas para importação dos dados**. Mas não se preocupe! Não é obrigatório para a resolução do exercício ;)

Na **segunda** atividade, você deve utilizar os dados da tabela **valor_medio** criada anteriormente e também os dados da tabela **Desafio1_nome_bairros.csv**. A ideia aqui é trazer as informações com o nome do bairro (NOME_BAIRRO) da tabela Desafio1_nome_bairros para a tabela valor_medio, de acordo com a correspondência do ID_BAIRRO.



Desafio 2 - Precificando um imóvel

Precificar um imóvel não é uma tarefa simples...

Felizmente, hoje temos técnicas avançadas, skill's, ferramentas e o mais importante: os dados.



Como um Cientista de Dados na Keycash, você deve ser capaz de transformar dados em Insights de negócio.

Pois bem, precificar um imóvel será seu **segundo desafio!**



O Desafio 2

Seu **desafio** será criar um **modelo de precificação** com as informações da base de dados anexa (**Desafio2_input.csv**) a este documento. Não se esqueça de dar uma olhadinha no arquivo **"Descrição dados dados.txt"**!



Regras – Desafio 2

01

Utilize **qualquer técnica** (pode ser uma sua se quiser) para precificar os imóveis.



02

Escreva seu racional de **forma didática**. **Disponibilize** o código para avaliarmos! Se quiser pode enviar em formato Jupyter.



Desafio Bônus – Quer nos mostrar outras habilidades? (opcional)

O desafio a seguir refere-se à uma forma de disponibilização (deploy) de modelos.

03

Uma forma de disponibilização de modelo é através de uma aplicação Web. Por exemplo, você pode fazer uma API utilizando alguma “biblioteca pronta” para te ajudar!

Dica: a API pode ser disponibilizada localmente (localhost), basta você nos enviar o código.



Dicas para os desafios



Não seja herói...
Existe muita
informação na
internet, utilize
sua criatividade
para alavancar no
que for necessário.



Fique a vontade para
entrar em contato e
tirar dúvidas sobre
o problema,
compartilhar as
dúvidas costuma ser
sempre bom.



Não se preocupe demais
com os resultados das
métricas do modelo. O
importante para nós é
a forma como você
abordou o problema.

keyc^sh

Para tempos de mudanças