

Projeto Final

Contexto do Problema de Negócio

Parabéns! Você acaba de ser contratado como Cientista de Dados da empresa Fome Zero, e a sua principal tarefa nesse momento é ajudar o CEO Kleiton Guerra a identificar pontos-chaves da empresa, respondendo às perguntas que ele fizer utilizando dados!

A empresa Fome Zero é uma marketplace de restaurantes. Ou seja, seu core business é facilitar o encontro e negociações de clientes e restaurantes. Os restaurantes fazem o cadastro dentro da plataforma da Fome Zero, que disponibiliza informações como endereço, tipo de culinária servida, se possui reservas, se faz entregas e também uma nota de avaliação dos serviços e produtos do restaurante, dentre outras informações.

O Desafio

O CEO Guerra também foi recém contratado e precisa entender melhor o negócio para conseguir tomar as melhores decisões estratégicas e alavancar ainda mais a Fome Zero, e para isso, ele precisa que seja feita uma análise nos dados da empresa e que sejam gerados dashboards, a partir dessas análises, para responder às seguintes perguntas:

Geral

1. Quantos restaurantes únicos estão registrados?
2. Quantos países únicos estão registrados?
3. Quantas cidades únicas estão registradas?
4. Qual o total de avaliações feitas?
5. Qual o total de tipos de culinária registrados?

Pais

1. Qual o nome do país que possui mais cidades registradas?
2. Qual o nome do país que possui mais restaurantes registrados?
3. Qual o nome do país que possui mais restaurantes com o nível de preço igual a 4 registrados?
4. Qual o nome do país que possui a maior quantidade de tipos de culinária distintos?
5. Qual o nome do país que possui a maior quantidade de avaliações feitas?
6. Qual o nome do país que possui a maior quantidade de restaurantes que fazem entrega?
7. Qual o nome do país que possui a maior quantidade de restaurantes que aceitam reservas?
8. Qual o nome do país que possui, na média, a maior quantidade de avaliações registrada?
9. Qual o nome do país que possui, na média, a maior nota média registrada?
10. Qual o nome do país que possui, na média, a menor nota média registrada?
11. Qual a média de preço de um prato para dois por país?

Cidade

1. Qual o nome da cidade que possui mais restaurantes registrados?
2. Qual o nome da cidade que possui mais restaurantes com nota média acima de 4?
3. Qual o nome da cidade que possui mais restaurantes com nota média abaixo de 2.5?
4. Qual o nome da cidade que possui o maior valor médio de um prato para dois?

5. Qual o nome da cidade que possui a maior quantidade de tipos de culinária distintas?
6. Qual o nome da cidade que possui a maior quantidade de restaurantes que fazem reservas?
7. Qual o nome da cidade que possui a maior quantidade de restaurantes que fazem entregas?
8. Qual o nome da cidade que possui a maior quantidade de restaurantes que aceitam pedidos online?

Restaurantes

1. Qual o nome do restaurante que possui a maior quantidade de avaliações?
2. Qual o nome do restaurante com a maior nota média?
3. Qual o nome do restaurante que possui o maior valor de uma prato para duas pessoas?
4. Qual o nome do restaurante de tipo de culinária brasileira que possui a menor média de avaliação?
5. Qual o nome do restaurante de tipo de culinária brasileira, e que é do Brasil, que possui a maior média de avaliação?
6. Os restaurantes que aceitam pedido online são também, na média, os restaurantes que mais possuem avaliações registradas?
7. Os restaurantes que fazem reservas são também, na média, os restaurantes que possuem o maior valor médio de um prato para duas pessoas?
8. Os restaurantes do tipo de culinária japonesa dos Estados Unidos da América possuem um valor médio de prato para duas pessoas maior que as churrascarias americanas (BBQ)?

Tipos de Culinária

1. Dos restaurantes que possuem o tipo de culinária italiana, qual o nome do restaurante com a maior média de avaliação?
2. Dos restaurantes que possuem o tipo de culinária italiana, qual o nome do restaurante com a menor média de avaliação?
3. Dos restaurantes que possuem o tipo de culinária americana, qual o nome do restaurante com a maior média de avaliação?
4. Dos restaurantes que possuem o tipo de culinária americana, qual o nome do restaurante com a menor média de avaliação?
5. Dos restaurantes que possuem o tipo de culinária árabe, qual o nome do restaurante com a maior média de avaliação?
6. Dos restaurantes que possuem o tipo de culinária árabe, qual o nome do restaurante com a menor média de avaliação?
7. Dos restaurantes que possuem o tipo de culinária japonesa, qual o nome do restaurante com a maior média de avaliação?
8. Dos restaurantes que possuem o tipo de culinária japonesa, qual o nome do restaurante com a menor média de avaliação?
9. Dos restaurantes que possuem o tipo de culinária caseira, qual o nome do restaurante com a maior média de avaliação?
10. Dos restaurantes que possuem o tipo de culinária caseira, qual o nome do restaurante com a menor média de avaliação?
11. Qual o tipo de culinária que possui o maior valor médio de um prato para duas pessoas?
12. Qual o tipo de culinária que possui a maior nota média?
13. Qual o tipo de culinária que possui mais restaurantes que aceitam pedidos online e fazem entregas?

O CEO também pediu que fosse gerado um dashboard que permitisse que ele visualizasse as principais informações das perguntas que ele fez. O CEO precisa dessas informações o mais rápido possível, uma vez que ele também é novo na empresa e irá utilizá-las para entender melhor a empresa Fome Zero para conseguir tomar decisões mais assertivas.

Seu trabalho é utilizar os dados que a empresa Fome Zero possui e responder as perguntas feitas do CEO e criar o dashboard solicitado.

Como é o seu primeiro trabalho, o CEO indicou um Cientista de Dados pleno para te ajudar. Ele observou os dados e te ajudou criando duas funções:

1. Para colocar o nome dos países com base no código de cada país
2. Criar a categoria do tipo de comida com base no range de valores.
3. Criar o nome das cores com base nos códigos de cores
4. Para renomear as colunas do DataFrame

Preenchimento do nome dos países

```
COUNTRIES = {
    1: "India",
    14: "Australia",
    30: "Brazil",
    37: "Canada",
    94: "Indonesia",
    148: "New Zeland",
    162: "Philippines",
    166: "Qatar",
    184: "Singapore",
    189: "South Africa",
    191: "Sri Lanka",
    208: "Turkey",
    214: "United Arab Emirates",
    215: "England",
    216: "United States of America",
}

def country_name(country_id):
    return COUNTRIES[country_id]
```

Criação do Tipo de Categoria de Comida

```
def create_price_tye(price_range):
```

```

if price_range == 1:
    return "cheap"
elif price_range == 2:
    return "normal"
elif price_range == 3:
    return "expensive"
else:
    return "gourmet"

```

Criação do nome das Cores

```

COLORS = {
    "3F7E00": "darkgreen",
    "5BA829": "green",
    "9ACD32": "lightgreen",
    "CDD614": "orange",
    "FFBA00": "red",
    "CBCBC8": "darkred",
    "FF7800": "darkred",
}

def color_name(color_code):
    return COLORS[color_code]

```

Renomear as colunas do DataFrame

```

def rename_columns(dataframe):
    df = dataframe.copy()
    title = lambda x: inflection.titleize(x)
    snakecase = lambda x: inflection.underscore(x)
    spaces = lambda x: x.replace(" ", "")
    cols_old = list(df.columns)
    cols_old = list(map(title, cols_old))
    cols_old = list(map(spaces, cols_old))
    cols_new = list(map(snakecase, cols_old))
    df.columns = cols_new

    return df

```

Além disso, o Cientista de Dados também sugeriu categorizar, inicialmente, todos os restaurantes somente por um tipo de culinária. Para isso, ele te entregou o seguinte código:

```

df["cuisines"] = df.loc[:, "cuisines"].apply(lambda x: x.split(",")[0])

```

Utilizando as funções criadas, você deverá ter um DataFrame pronto para ser analisado

O CEO viu um Dash interno da empresa, e pediu pra que você monte um novo dashboard tendo como base esse dashboard interno. O link para acessar o dashboard interno é:

<https://pedroferraresi-ftc-analise-python-01--main-page-ozp460.streamlit.app/>

Os Dados

O conjunto de dados que representam o contexto está disponível na plataforma do Kaggle. O link para acesso aos dados :

<https://www.kaggle.com/datasets/akashram/zomato-restaurants-autoupdated-dataset?resource=download&select=zomato.csv>

Baixe e utilize somente o arquivo .csv.

Como Solucionar Esse Desafio

Aqui vão algumas dicas para você começar a resolver esse problema:

- **Utilize os conhecimentos adquiridos no curso:** Reviste as suas anotações e reveja algumas aulas se necessário.
- **Tenha calma e não tenha medo:** Dê um passo de cada vez e quebre as suas tarefas em tarefas menores e mais simples. Tudo que você precisa para resolver esse desafio foi ensinado no curso.
- **Responda as perguntas antes de Codificar:** Como você faria, se não tivesse que programar, para responder as perguntas feitas? Pensando antes em como responder, facilitará no momento de codificar o algoritmo para achar a resposta, porque isso faz parte do planejamento da solução!
- **Tenha Paciência e Resiliência:** Criar soluções e, principalmente, pensar e planejá-las leva tempo. Assuma uma postura resiliente e não desista! Afinal, você quer se tornar um Cientista de Dados, ganhar ótimos salários e trabalhar em ótimas empresas, certo?
- **A Comunidade:** Caso você tenha tentado várias estratégias e não tenha chegado a uma solução ou não tenha avançado, peça ajuda dentro da

Comunidade de DS! Estamos todos para nos ajudar nessa jornada que é a Ciência de Dados.

Roteiro Sugerido para Resolução

Um roteiro que pode ser usado como resolução do desafio, seria:

1. Colete os dados:
 - a. Os dados estão disponíveis no link neste post
2. Entenda os dados que você está trabalhando:
 - a. O que cada coluna representa?
 - b. Existem colunas que podem ser removidas do DataFrame, por possuir somente 1 valor em todas as linhas?
2. Faça uma limpeza nos dados:
 - a. Verifique se existem dados duplicados
 - b. Entenda as variáveis disponíveis na base de dados fazendo uma tabela de estatística descritiva.
 - c. Verifique se há dados faltantes e remova-os
3. Explore os dados e responda as perguntas:
 - a. Comece respondendo as perguntas sem usar programação, para planejar a sua solução. Ou seja, responda primeiro como você faria para responder a pergunta. Por exemplo: Calcularia a média dos valores dos pratos para duas pessoas agrupados por país.
 - b. Com o planejamento feito, implemente os passos/ações utilizando a linguagem de programação.
 - c. Se necessário, utilize gráficos para consolidar e validar a sua resposta.
 - d. Caso uma pergunta específica se mostre falsa, encontre a situação ou fato que a torne falsa. Por exemplo: "Os restaurantes do tipo de culinária japonesa dos Estados Unidos da América possuem um valor médio de prato para duas pessoas maior que as churrascarias americanas?". Caso essa pergunta seja falsa, demonstre com os dados que ela é falsa.
 - e. **Caso uma pergunta tenha dois registros como resposta, selecione sempre o registro com o valor da coluna "Restaurant**

ID" menor, assim iremos pegar o restaurante que está há mais tempo cadastrado na base de dados.

4. Deixe suas respostas disponíveis:
 - a. Utilizando o Streamlit, pense e planeje uma solução para deixar os seus Insights e as respostas disponíveis para o investidor possa acessá-las

O Ferramental da Solução

Como ideia desse projeto é colocar em prática e revisar todas as técnicas e conhecimento adquirido dentro do curso de FTC - Analisando Dados com Python, recomendo a utilização das ferramentas usadas durante o curso, como o Jupyter Lab, para prototipar e te auxiliar no desenvolvimento da solução que será criada e o framework Streamlit para possibilitar a criação do dashboard.

Aproveite para treinar e revisar todo o conteúdo visto, melhorando assim a sua velocidade de manipulação de dados, raciocínio analítico e resolução de problemas! Tenho certeza de que com esses projetos e tempo, você escreverá códigos e ferramentas cada vez mais robustas e ficará cada vez mais fluente e seguro nas técnicas e ferramentas usadas dentro do universo de Ciência de Dados.

Vá em frente!

Como sempre conversamos dentro da Comunidade, não existe caminho fácil e de curto prazo em nenhuma profissão, e ciência de dados não é exceção. Através de estudo e desenvolvimento de projeto, você irá aumentar a sua capacidade de resolver problemas e entregar resultados.

E os seus projetos de portfólio servem para esse fim: Demonstrar as suas capacidades de resolver problemas. Ou seja, demonstram que você é tão capaz quanto um cientista de dados que já atua profissionalmente nas empresas.

Conclusão

Você recebeu um desafio de Ciência de Dados próximo dos desafios reais das empresas e você irá utilizar todo o conhecimento adquirido no curso FTC - Analisando Dados com Python para resolvê-lo.

Os problemas nas empresas chegam em forma de perguntas abertas, desestruturadas e sem nenhuma dica sobre como resolver, então, utilizar os conhecimentos adquiridos no curso serão fundamentais para o seu sucesso na sua jornada como um profissional de Ciência de Dados.

É papel do Cientista de Dados entender a causa raiz, planejar o desenvolvimento e criar a melhor solução para o problema de negócio.

Aproveite esse desafio para colocar em prática tudo que você aprendeu ao longo do curso! E caso tenha alguma dificuldade, utilize o poder da Comunidade DS: Publique ela dentro do canal do curso no Discord para discutir com a galera as suas dúvidas!