14/9/2023

Alexsandre Aguiar

CODERHOUSE

**Segmentação de clientes para estratégia de vendas**

Projeto Final

Sumário

[1- Descrição do Caso de Negócio 2](#_Toc147332314)

[2- Tabela de Versionamento 2](#_Toc147332315)

[3- Objetivo do Modelo 2](#_Toc147332316)

[4- Descrição dos Dados 2](#_Toc147332317)

[5- Resultados Encontrados pela EDA 3](#_Toc147332318)

[a. Análise Univariada 3](#_Toc147332319)

[b. Análise Bivariada 4](#_Toc147332320)

[6- Algoritmo Escolhido 5](#_Toc147332321)

[7- Métrica de Desempenho 5](#_Toc147332322)

[8- Iterações de Otimização 5](#_Toc147332323)

[9- Métricas Finais 6](#_Toc147332324)

[10- Linhas de Ações Futuras 6](#_Toc147332325)

[11- Conclusão 7](#_Toc147332326)

# Descrição do Caso de Negócio

A empresa Buiatchaka sofre com a falta de insights e compreensão sobre as características, necessidades e comportamentos específicos dos seus clientes. Como a sua base de clientes não possui segmentação em grupos distintos, a empresa não consegue direcionar suas estratégias de venda e ações de marketing de maneira eficiente e personalizada.

# Tabela de Versionamento

|  |  |  |
| --- | --- | --- |
| Versão | Autor | Data |
| V 1.0 | Alexsandre Aguiar | 14/09/2023 |

# Objetivo do Modelo

Segmentar os clientes em grupos com características semelhantes, com o objetivo de melhorar as estratégias de marketing, aumentar o engajamento dos clientes, impulsionar os resultados do negócio, além de trabalhar com a previsão da localidade de novos clientes para elaborar e executar ações de venda direcionadas geograficamente.

# Descrição dos Dados

Base de dados de clientes da empresa Buiatchaka, possuindo as seguintes colunas: first\_name, last\_name, title, gender, email, city, country, country\_code, latitude, longitude, phone, street\_address, street\_name, street\_number, street\_suffix, time\_zone, company\_name, department, job\_title, language, university, linkedin\_skill, ip\_address, age, salary\_year.

A base foi coletada através do arquivo Clientes\_atualizado.csv.

O Conjunto de dados possui 51 mil registros. Desse conjunto, não há linhas duplicadas, 35 registros são nulos e ele têm um total de 25 características, das quais apenas 5 são numéricas e o restante são características categóricas;

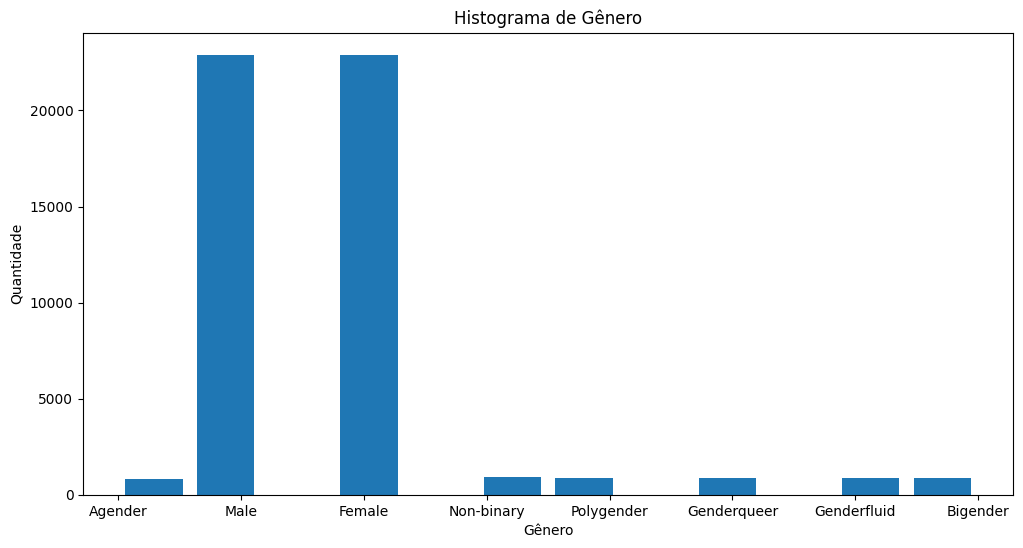
# Resultados Encontrados pela EDA

# Análise Univariada

Gráfico

Descrição gerada automaticamente

Como podemos observar, o departamento está igualmente distribuído.

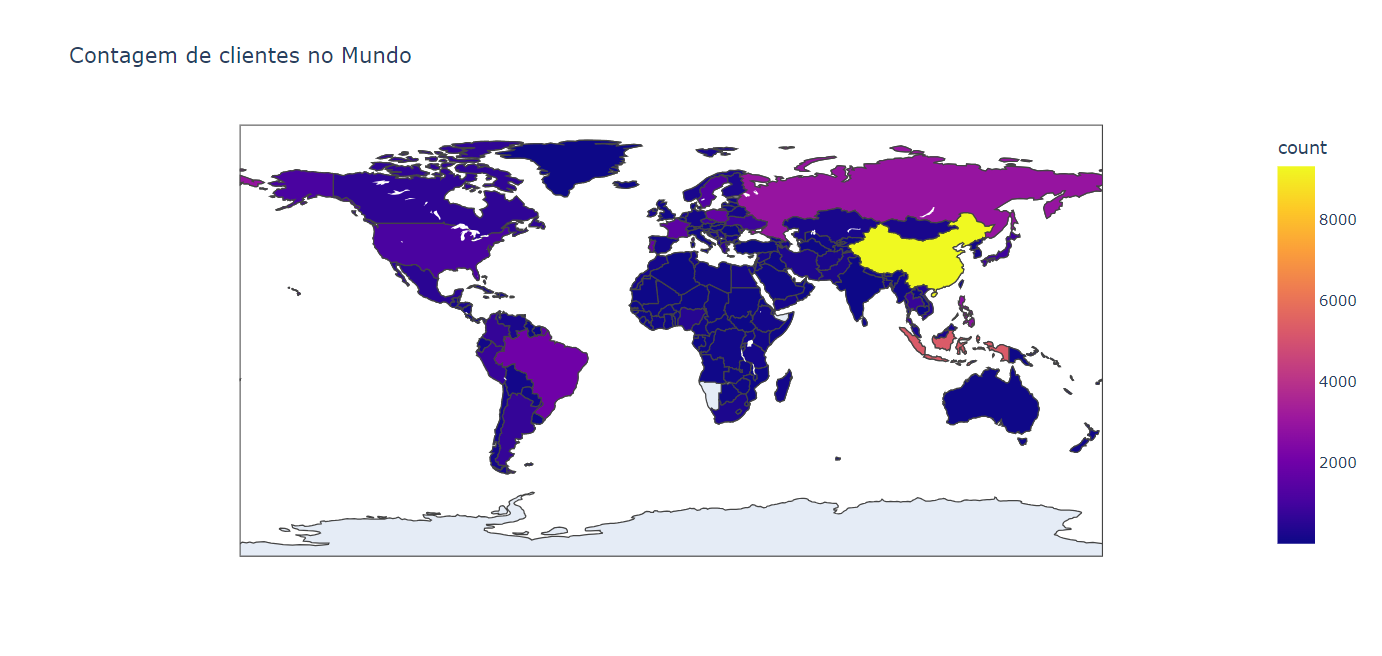
Como podemos observar, há um número igual de clientes masculinos e femininos no conjunto de dados e eles dominam a característica de gênero.

# Análise Bivariada

Gráfico, Gráfico de barras

Descrição gerada automaticamente

Executivo Júnior e Analista de Programação são as profissões dominantes. Outras profissões importantes somente possuem clientes do sexo feminino ou masculino.



O gráfico acima mostra claramente a distribuição de clientes em todo o mundo. A maior parte dos clientes é da China como já visto no gráfico anterior. A área azul indica um número muito menor de clientes daquele local, como Austrália, Índia, África, Groenlândia.

# Algoritmo Escolhido

Com base nas variáveis selecionadas, meu intuito é prever o país dos meus novos clientes para que se possa concentrar a comercialização e as ações de marketing dos meus produtos baseados nessa característica geográfica.

Baseado no ponto citado e para um melhor retorno, a idéia inicial é usar o modelo KNN, utilizando o número de Vizinhos (n\_neighbors), para determinar o número de vizinhos a serem considerados ao fazer a previsão.

# Métrica de Desempenho

O modelo escolhido foi o algoritmo de classificação K-Nearest Neighbors (KNN), para previsão do país de origem de novos clientes com base em características como idade (Age), salário anual (Salary\_Year) e gênero (gender).

O conjunto de dados foi dividido em duas partes: um conjunto de treinamento (X\_treino, y\_treino) usado para treinar o modelo e um conjunto de teste (X\_teste, y\_teste) usado para avaliar o desempenho do modelo.

A variável categórica 'gender' é codificada numericamente usando o LabelEncoder para que possa ser usada como entrada no modelo KNN.

Um modelo KNN com um número de vizinhos (k) igual a 10 é criado e treinado com os dados de treinamento (X\_treino, y\_treino) usando o método fit.

A precisão do modelo no conjunto de teste é calculada usando o método score, que compara as previsões do modelo (com base em X\_teste) com as verdadeiras classes (y\_teste). A precisão é então impressa na tela.

# Iterações de Otimização

Com o intuito de otimizar o código e melhorar o desempenho do modelo (KNN), realizei o ajuste dos hiperparâmetros.

Utilizei a validação cruzada para ajustar o hiperparâmetro k (número de vizinhos) para encontrar o valor que maximizasse o desempenho do modelo, com o auxílio da classe GridSearchCV do scikit-learn.

Com esse ajuste aumentei a taxa de precisão do modelo, que está longe do ideal, mas que foi a possível a partir dos conhecimentos adquiridos.



# Métricas Finais

Ao utilizar o KNN para criar um modelo de classificação com base nas características de idade, salário e gênero dos clientes, a distância entre pontos é considerada para classificar os novos clientes com base nos vizinhos mais próximos no conjunto de treinamento. Com isso notamos que, por conta das características informadas e por utilizar os 10 vizinhos mais próximos, dos 10 clientes novos informados, é previsto que 60% venham da China, 30% da Indonésia e 10% da Suécia.

Texto

Descrição gerada automaticamente

Tabela

Descrição gerada automaticamente com confiança média

# Linhas de Ações Futuras

Para evolução do modelo, seria de grande valia também utilizar o histórico de compras anteriores dos clientes.

Com os dados do comportamento passado, o modelo poderia também agrupar os clientes com base em seu gosto ou necessidade, o que aumentaria e muito a precisão do modelo.

# Conclusão

Com base na análise do conjunto de dados e dos modelos utilizados, a empresa pode focar na comercialização de seus produtos com base em características geográficas.

Pode focar no mercado asiático, principalmente China e Indonésia, o que aumentaria seu público-alvo, aumentando as chances de vendas.