**Relatório atividade EBAC**

Para a primeira etapa selecionei a fonte de dados do Kaggle, referente ao comportamento dos consumidores em relação à compras feitas pelo E-comerce, disponível no link abaixo:

<https://www.kaggle.com/datasets/salahuddinahmedshuvo/ecommerce-consumer-behavior-analysis-data>

Através dos dados fornecidos, busquei entender um pouco melhor como está o comportamento de compra avaliando diversas diferenças, como idade, escolaridade, frequência de uso. Para entender um pouco melhor se esse tipo de comercio está mais amigável para toda a população.

Os dados utilizados foram dados estruturados, apenas em uma tabela. Inicialmente com 28 colunas, mas novas colunas foram adicionadas para análises especificas. Como dito inicialmente, dado foi retirado do kaggle, e fizemos todo tratamento através do PyCharm, utilizando python.

Segue o código abaixo:

import pandas as pd  
from sklearn.preprocessing import LabelEncoder  
  
  
pd.set\_option('display.width', None)  
pd.set\_option('display.max\_colwidth', None)  
  
df = pd.read\_csv(r'C:\Users\Felip\Downloads\Ecommerce\_Consumer\_Behavior\_Analysis\_Data.csv')  
print(df.head())  
  
print(df.shape)  
print(df.dtypes)  
print('Valores Nulos:\n', df.isnull().sum())  
  
df\_tratado = df.fillna('Unknown')  
print('Valores Nulos:\n', df\_tratado.isnull().sum())  
  
ensino = df\_tratado['Education\_Level'].unique()  
print(ensino)  
  
education\_ordem = {"Bachelor's": 1, "High School": 2, "Master's": 3}  
df\_tratado['Education\_Level\_Ord'] = df['Education\_Level'].map(education\_ordem)  
  
label\_encoder = LabelEncoder()  
df\_tratado['Gender\_cod'] = label\_encoder.fit\_transform(df['Gender'])  
df\_tratado['Marital\_Status\_cod'] = label\_encoder.fit\_transform(df['Marital\_Status'])  
  
df\_tratado['Purchase\_Amount'] = df\_tratado['Purchase\_Amount'].str.replace('$', '', regex=False)  
df\_tratado['Purchase\_Amount'] = df\_tratado['Purchase\_Amount'].astype(float)  
  
print('Correlação: \n',df\_tratado[['Purchase\_Amount', 'Age','Gender\_cod', 'Marital\_Status\_cod','Education\_Level\_Ord', 'Frequency\_of\_Purchase', 'Time\_Spent\_on\_Product\_Research(hours)', 'Customer\_Satisfaction', 'Return\_Rate']].corr())  
  
print(df\_tratado.dtypes)  
print(df\_tratado['Purchase\_Amount'].sum())  
print(df\_tratado['Purchase\_Amount'].max())  
print(df\_tratado['Purchase\_Amount'].min())  
print(df\_tratado['Purchase\_Amount'].mean())  
  
  
bins = [17, 23, 29, 35, 41, 47, 60]  
labels = ['18-23', '24-29', '30-35', '36-41', '42-47', '48-60']  
df\_tratado['Age\_range'] = pd.cut(df['Age'], bins=bins, labels=labels)  
  
  
df\_tratado.to\_csv('Df\_tratado.csv', index=False)

O arquivo será encaminhado junto a atividade.

No último passo foi feita a apresentação no Looker Studio, como solicitado, abaixo segue o link para acesso do relatório.

<https://lookerstudio.google.com/s/uVfQlcmiSCo>

Desenvolvimento:

Após a seleção de dados, se iniciou a preparação e o tratamentos deles, utilizando do pycharm. Após importação das bibliotecas necessárias (Pandas e sklearn), analisei as colunas presentes no dataframe e logo fiz a checagem se possuía valores nulos. Os valores nulos presentes eram provavelmente por falta de informação na coleta, como não atrapalharia nas análises, substitui os valores nulos para “desconhecidos”. Para fazer a correlação de algumas colunas foi necessário fazer a codificação de colunas com variáveis categóricas, para isso utilizei o LabelEncoder e outras técnicas, após isso foi feita a correlação de algumas colunas.

Com base nos resultados obtidos, decidi fazer um estudo para saber se no uso do ecommerce existe alguma preferência de gênero, idade ou escolaridade. Além disso, ter uma noção da importância do ecommerce, através dos valores envolvidos e da frequência do uso. Com essas repostas, conseguimos fornecer para as lojas varejistas, a informação se vale apena investir nessa modalidade. Para começar, peguei os valores monetários (US $) envolvidos (valor total, o máximo gasto por cliente, o mínimo gasto por cliente e a média gasta), para ter uma ideia da relevância desse estilo de comercio. Ainda no pycharm adicionei uma coluna de Faixa de idade, para facilitar a visualização gráfica futura. Salvei os dados alterados, para iniciar a apresentação via LookerStudio.

No Looker Studio coloquei as informações gerais dos valores monetários, que já havia retirado no pycharm e fiz mais 4 gráficos. O primeiro apresentando o valor gasto por gênero e escolaridade dos clientes, o segundo para mostrar o valor gasto por faixa de idade, o terceiro de tempo gasto nas compras por faixa de idade e o último que mostra a frequência de compras repetidas.

Analisando os gráficos e dados apresentados no LookerStudio, é possível fazer algumas reflexões. Utilizando o primeiro e segundo gráfico, conseguimos perceber que hoje em dia, a utilização do ecommerce não é mais restrita aos jovens, ou aqueles que possuem “mais afinidade” com o mundo digital. Não existe uma diferença significativa no valor gasto entre gênero, escolaridade ou idade das pessoas. A diferença por outros gêneros aconteceu mais por baixa quantidade de pessoas envolvidas neles. Para idade aconteceu algo parecido, pois consegue ver uma leve diferença da faixa de idade de 48-60 anos, mas isso aconteceu por que a pessoa mais velha envolvida nesse estudo, possui 50 anos. Logo essa última faixa de idade ficou restrita a pessoas de 48, 49 e 50 anos, uma das justificativas para o resultado estar um pouco fora do padrão das outras faixas.

Pelo Terceiro gráfico, conseguimos reforçar essa ideia. O tempo gasto nas compras de ecommerce é praticamente o mesmo para todas as faixas de idade, a única diferença mais significante, pode ser justificada pelo mesmo motivo anterior, de possuir menos pessoas avaliadas nessa última faixa de idade.

E o último gráfico mostra que o ecommerce é cada vez mais utilizado nos dias atuais, cada vez mais compras são repetidas o que mostra que as pessoas estão confiando nessa forma de comercio.

Conclusões:

Através dos insights, conseguimos perceber que Ecommerce está cada mais aceitável para a população de forma geral. Antigamente era comum esse método ser utilizados apenas por pessoas mais novas, com mais intimidade com o mundo digital. Mas podemos ver que através desses dados que essa diferença está cada vez menor, na maioria dos casos, até irrelevantes. Não importa o gênero, a idade ou escolaridade. Hoje em dia todos tem valores gastos bem próximos com Ecommerce.