**UNIVERSIDADE PRESBITERIANA MACKENZIE**

**GABRIELA GERMANA DA SILVA**

**NATHALIA DE ALBUQUERQUE DO AMARAL**

**GRUPO – DATA 4 LIFE**

**PROJETO APLICADO II**

**LOJA ONLINE DE ROUPAS**

Sumário

[ETAPA 4](#_Toc136291959)

[GRUPO DE TRABALHO 4](#_Toc136291960)

[VISÃO GERAL 4](#_Toc136291961)

[OBJETIVO 4](#_Toc136291962)

[META 4](#_Toc136291963)

[BIBLIOTECA GITHUB 5](#_Toc136291964)

[BASE DE DADOS 5](#_Toc136291965)

[DEFINIÇÃO DA LINGUAGEM DE PROGRAMAÇÃO USADA NO PROJETO. 5](#_Toc136291966)

[BIBLIOTECAS PYTHON 5](#_Toc136291967)

[CRONOGRAMA 6](#_Toc136291968)

[ANÁLISE EXPLORATÓRIA DA BASE DE DADOS ESCOLHIDA. 7](#_Toc136291969)

[TRATAMENTO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA: 8](#_Toc136291970)

[Metodologia de Avaliação: 12](#_Toc136291971)

[ACURÁCIA 12](#_Toc136291972)

[ANOVA: 14](#_Toc136291973)

[Storytelling: 16](#_Toc136291974)

[Conclusão: 18](#_Toc136291975)

# ETAPA

Nesta etapa do projeto, você deve entregar um documento conforme os itens:

* Grupo de trabalho.
* Premissas do projeto: definição da organização escolhida, área de atuação e apresentação dos dados que serão utilizados (imagem ou texto).
* Objetivos e metas.
* Cronograma de atividades (Estimativa).

# GRUPO DE TRABALHO

Data 4 life (2 pessoas)

Gabriela Germana

Natália Albuquerque

# VISÃO GERAL

Neste projeto vamos trabalhar com uma loja online de roupa dos Estados Unidos e iremos extrair um conjunto de dados fictício sobre os dados de transação da loja online dos EUA. Onde iremos verificar o tipo de cliente, tipo de compras e trabalhar com os dados a fim de encontrarmos novas formas de melhor a situação de vendas da loja e aumento de faturamento.

# OBJETIVO

Segmentar os clientes (preditivo) – decidimos então segmentar os clientes em categorias de acordo com a frequência de compras, como por exemplo “Platinum”, “Gold” e “Silver”, de acordo com a análise e perfil de compras desses clientes. Ao observarmos o comportamento será possível a classificação e as melhores tomadas de decisões quanto aos próximos passa de vendas e capitação da loja.

# META

Aumentar as vendas em cada segmento (*Crosell - cross-selling*) – A meta então é aumentar as vendas, focando principalmente nos clientes que ainda não estão em uma categoria que indica uma fidelidade com a loja.

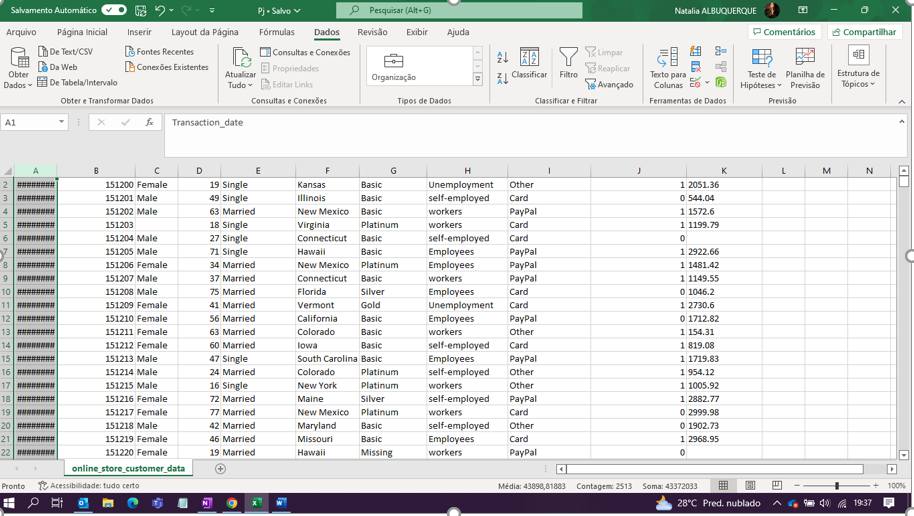
# BIBLIOTECA GITHUB

<https://github.com/gabigermana/projeto_aplicado_02>

# BASE DE DADOS

Foram extraídos dados no site “*Kaggle*”, seguem os links abaixo do site onde extraímos os dados e foto dos dados no Excel.

<https://www.kaggle.com/datasets/mountboy/online-store-customer-data?resource=download>



**Imagem do banco de dados extraído do site Kaggle**

# DEFINIÇÃO DA LINGUAGEM DE PROGRAMAÇÃO USADA NO PROJETO.

A linguagem de programação utilizada será em *Python*.

# BIBLIOTECAS *PYTHON*

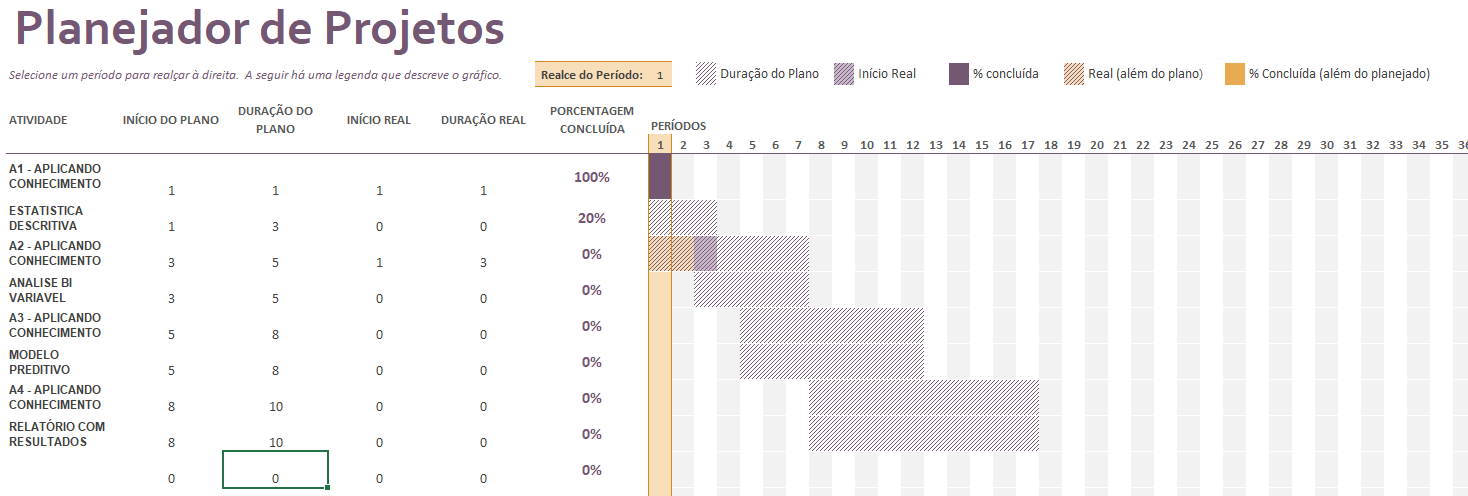
*NumPy*

*Pandas*

*Matplotlib.pyplot*

*Seaborn*

# CRONOGRAMA



OBS: O Cronograma das atividades ainda será ajustado de acordo com a quantidade de semanas que temos disponível para completá-las.

# ANÁLISE EXPLORATÓRIA DA BASE DE DADOS ESCOLHIDA.

A análise exploratória escolhida será a análise ABC, onde iremos classificar categorias de acordo com sua importância para conseguirmos analisar os dados e tomarmos decisões com base nas amostragens.

Iremos usar o valor total da loja, número de vendas, valor de vendas por categoria e assim modelar nossos dados de acordo com as tomadas de decisões.

Para analisar o desempenho das vendas da nossa loja de roupas, iremos utilizar técnicas de análise de dados, em particular, modelos de classificação para identificar os produtos que mais e menos vendem. Os dados disponíveis incluem informações sobre o tipo de roupa, idade dos clientes, gênero, estado etc. Utilizaremos essas informações para classificar os produtos em diferentes categorias, a fim de identificar padrões e determinar quais produtos têm maior ou menor demanda. Para complementar a análise, podemos utilizar a técnica de *clustering* para agrupar os produtos em diferentes categorias com base em suas características comuns, a fim de identificar quais grupos de produtos são mais populares ou menos populares entre os clientes. O objetivo é identificar quais produtos têm maior ou menor demanda, a fim de otimizar o estoque e as vendas da loja.

# TRATAMENTO DA BASE DE DADOS E ANÁLISE EXPLORATÓRIA:

Tabela

Descrição gerada automaticamente

Imagem 1 – Base de dados

Interface gráfica do usuário, Aplicativo, Word

Descrição gerada automaticamente

Imagem 2 – Base de dados

Gráfico

Descrição gerada automaticamente

Imagem 3 - Podemos identificar que a idade dos clientes está entre 30 e 60 anos.

Interface gráfica do usuário, Aplicativo

Descrição gerada automaticamente

Imagem 4 - Assim como 53,98% dos clientes são mulheres e 44,90 são homens.

Texto

Descrição gerada automaticamente

Imagem 5 - 58,6% dos clientes são casados e 41,4% são solteiros.

Uma imagem contendo Texto

Descrição gerada automaticamente

Imagem 6 - O tipo de cliente que estão cadastrados na loja dentro de cada perfil.

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

Imagem 7 - Os status de cliente que estão empregados, desempregados etc.

Uma imagem contendo Tabela

Descrição gerada automaticamente

Imagem 8 - Amostra de segmento por tipo de ocupação dos clientes.

Tela de computador

Descrição gerada automaticamente

Imagem 9 - Amostragem de clientes por estados, conforme base de dados.

Calendário

Descrição gerada automaticamente com confiança baixa

Imagem 10 - Amostra do tipo de pagamento realizado pelos clientes.

# Metodologia de Avaliação:

Para avaliar o desempenho de vendas da nossa loja de roupas, iremos adotar medidas de desempenho.

Acurácia, precisão, revogação e F1, obtidas por meio da análise dos dados de vendas utilizando ferramentas estatísticas disponíveis no Python. Também iremos usar a tabela ANOVA para nos ajudar em nosso objetivo é identificar os produtos que têm maior e menor demanda entre nossos clientes e o grau de confiança nas nossas análises, medindo a quantidade de vendas realizadas corretamente. Para escolher os melhores produtos a serem comercializados, utilizaremos a matriz de confusão para identificar quais produtos vendem mais ou menos. Procuraremos por produtos com a maior quantidade de vendas corretas (Verdadeiro Positivo - TP), aqueles que têm maior demanda entre nossos clientes e que trazem mais lucro para a loja.

# ACURÁCIA

O que é Acurácia?

O termo "acurácia" tem sua origem na palavra "*accuracy*" em inglês. No Brasil, a Física e a Matemática começaram a utilizar o conceito para descrever a proximidade entre um resultado experimental e seu valor real, sendo que uma maior acurácia indica maior autenticidade do resultado.

Na tecnologia, a acurácia refere-se à proximidade entre um resultado experimental e um valor de referência real, determinando o grau de exatidão. Em outras palavras, ela é uma medida que quantifica o nível de erro entre um valor desejado ou ideal e um valor real ou medido.

Níveis de Acurácia:

Os níveis de acurácia não são um conceito binário de "sim" ou "não". Eles são classificados em diferentes níveis percentuais para determinar se a acurácia de um modelo é satisfatória. É um processo complexo que auxilia na interpretação, especialmente para avaliar a efetividade de áreas, setores, processos ou produtos de uma empresa.

Vamos utilizar um exemplo para entender melhor como esses níveis funcionam. Imagine um teste de laboratório para detectar uma doença, em que a acurácia é medida com base na quantidade de vezes em que ele acertou se o paciente possui ou não a enfermidade.

Dependendo da doença em questão, a acurácia pode estar em torno de 80%. Isso pode parecer interessante em um aspecto, mas ter um valor relativamente baixo se comparado a outros métodos e processos que podem atingir uma acurácia de 99,99%. Ao lidar com assuntos que exigem interpretação precisa, é essencial separar os percentuais em diferentes níveis.

Por esse motivo, foram estabelecidos níveis para classificar cada sistema ou processo. São eles:

1. Baixo:

- Percentuais entre 0% e 30% são considerados como nível baixo de acurácia.

- Nesse nível, há pouca ou nenhuma confiança de que os resultados obtidos se aproximem da realidade.

- Em termos de combate a fraudes, um nível baixo de acurácia representa um risco relativamente alto, o que não é favorável para uma solução.

- Na prevenção de crimes virtuais, esse nível deve ser avaliado com cautela, pois uma baixa acurácia pode resultar em seleção inadequada de consumidores, classificando como fraudes pessoas que realmente desejam fazer negócios ou permitindo a presença de fraudadores como consumidores legítimos.

2. Médio:

- Percentuais entre 30% e 70% são considerados como nível médio de acurácia.

- Esse nível indica um risco moderado de não apresentar resultados aceitáveis em relação aos valores de referência.

- Embora não seja o nível mais seguro de acurácia, alguns modelos de negócio que toleram um grau mais elevado de risco podem considerá-lo aceitável.

- Na prevenção de fraudes, os sistemas que oferecem um nível médio de acurácia também devem ser utilizados com cautela. Quando o objetivo é detectar tentativas de captura de dados ou invasões, os percentuais precisam ser mais altos.

3. Alto:

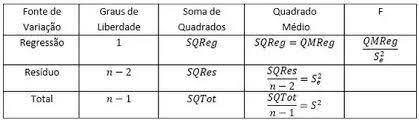
- Percentuais entre 70% e 100% são considerados como nível alto de acurácia.

- Nesse caso, os resultados são altamente precisos, com baixo risco de erro, proporcionando maior segurança para a tomada de decisões.

- Quando a acurácia ultrapassa os 90%, os resultados são considerados comprovados, pois a variação em relação ao valor real de referência é mínima.

- Quando a decisão não pode ter margem para erro e a tolerância ao risco é baixa, trabalhar com níveis elevados de acurácia é essencial.

# ANOVA:



(imagem internet)

Uma definição rápida de ANOVA é que é uma técnica de análise estatística na qual conjuntos de dados são comparados e medidos para determinar sua significância.

Para uma definição mais detalhada, é necessário percorrer o processo do teste e aplicação dos resultados. Este teste mede efetivamente a significância da interação entre variáveis, analisando a variância, começam criando uma hipótese nula (Ho), que afirma que não há diferença significativa entre as variáveis medidas, se o teste apresentar resultados estatisticamente significativos, então o testador pode rejeitar a hipótese nula e aceitar a hipótese alternativa (H1), afirmando que a interação entre as variáveis é significativa.

A estatística F é o coeficiente da ANOVA que indica se os resultados são significativos ou não. Um valor F próximo a 1 denota pouca ou nenhuma diferença nos valores, o que significa que não há uma variação significativa entre os grupos. Existem muitos sistemas de software estatístico que podem executar os testes de ANOVA para você, ou alguns cálculos estatísticos inteligentes podem ser feitos para descobrir isso. A fórmula para encontrar a estatística F é pegar o erro quadrático médio do conjunto de dados e dividi-lo pela soma média de quadrados do conjunto de dados.

É recomendado o uso de um sistema de software, pois tudo isso pode ser feito em questão de segundos.

# Storytelling:

Empresa “Joaninha Moda”: Segmentando Clientes para o Crescimento Sustentável e Lucrativo

Introdução:

A Empresa “Joaninha Moda” é uma organização dedicada a oferecer produtos e serviços excepcionais aos seus clientes. Com o objetivo de aprimorar seu desempenho e aperfeiçoar suas estratégias de vendas, a empresa embarcou em uma jornada emocionante de segmentação de clientes. Através de uma abordagem preditiva, a Empresa “Joaninha Moda “decidiu classificar seus clientes em categorias distintas com base na frequência de compras e no perfil de compras. Essa segmentação permitiu uma melhor compreensão dos clientes, possibilitando a tomada de decisões mais assertivas em relação às vendas futuras e à expansão do negócio.

Desenvolvimento:

1. Compreensão do desafio:

- A Empresa “Joaninha Moda “reconheceu a importância de entender profundamente seu público-alvo e personalizar suas estratégias de vendas.

- O desafio era identificar padrões e comportamentos de compra entre os clientes para segmentá-los de forma eficaz.

2. Coleta e análise de dados:

- A Empresa “Joaninha Moda “implementou um sistema robusto de coleta de dados, capturando informações relevantes sobre as transações e histórico de compras dos clientes.

- Utilizando técnicas de análise preditiva, a empresa avaliou os dados coletados para identificar padrões de comportamento de compra e tendências.

3. Segmentação de clientes:

- Com base nas análises, a Empresa “Joaninha Moda “definiu três categorias principais de clientes: Platinum, Gold e Silver.

- Os clientes Platinum eram aqueles com maior frequência de compras e um histórico consistente de compras de alto valor.

- Os clientes Gold representavam aqueles com uma frequência de compras moderada e um histórico de compras de valor médio.

- Os clientes Silver eram aqueles com menor frequência de compras, mas ainda com potencial de crescimento.

4. Tomada de decisões e estratégias personalizadas:

- Com a segmentação em vigor, a Empresa “Joaninha Moda “pôde direcionar suas estratégias de vendas de maneira mais eficiente.

- Para os clientes Platinum, foram oferecidos programas de fidelidade exclusivos, brindes e benefícios especiais para incentivar a lealdade e aumentar o valor das compras.

- Os clientes Gold receberam ofertas personalizadas com base em seus históricos de compras anteriores, aproveitando seus interesses e preferências específicos.

- Para os clientes Silver, foram criadas campanhas de marketing direcionadas, incentivando-os a aumentar sua frequência de compras por meio de descontos exclusivos e promoções especiais.

5. Monitoramento e ajustes contínuos:

- A Empresa “Joaninha Moda “acompanhou de perto o desempenho de cada segmento de clientes por meio de métricas específicas.

- Com base nas análises contínuas, a empresa ajustou suas estratégias conforme necessário, aprimorando constantemente a experiência do cliente e maximizando o retorno sobre o investimento.

# Conclusão:

A segmentação de clientes realizada pela Empresa "Joaninha Moda" foi um passo fundamental para o crescimento sustentável e lucrativo do negócio. Ao compreender profundamente seu público-alvo e personalizar suas estratégias de vendas, a empresa obteve uma vantagem competitiva significativa. A coleta e análise de dados desempenharam um papel essencial nesse processo, permitindo que a empresa identificasse padrões de comportamento de compra e tendências entre os clientes. Essa compreensão detalhada levou à criação de três categorias principais de clientes: Platinum, Gold e Silver.

A segmentação permitiu que a Empresa "Joaninha Moda" direcionasse suas estratégias de vendas de forma mais eficiente, adaptando-as às necessidades e preferências específicas de cada grupo de clientes. Os clientes Platinum receberam benefícios exclusivos para incentivar a fidelidade e aumentar o valor das compras. Os clientes Gold foram abordados com ofertas personalizadas com base em seus históricos de compras, enquanto os clientes Silver foram alvo de campanhas de marketing direcionadas para estimular o aumento da frequência de compras.

Acompanhando de perto o desempenho de cada segmento de clientes e realizando ajustes contínuos, a Empresa "Joaninha Moda" conseguiu melhorar constantemente a experiência do cliente e maximizar o retorno sobre o investimento. A segmentação de clientes não apenas impulsionou o crescimento sustentável do negócio, mas também fortaleceu a relação com os clientes existentes e atraiu novos clientes. A abordagem personalizada e direcionada permitiu que a empresa cultivasse relacionamentos duradouros, aumentando a satisfação do cliente e a fidelidade à marca.

Em um mercado cada vez mais competitivo, a segmentação de clientes é uma estratégia essencial para qualquer empresa que busque crescimento e sucesso a longo prazo. A Empresa "Joaninha Moda" demonstrou como essa abordagem pode impulsionar os resultados, tornando-se um exemplo inspirador para outras organizações em busca de estratégias eficazes de vendas e crescimento sustentável e lucrativo.

Referencias Bibliográficas

<https://docs.python.org/pt-br/3/> (Acessado em todas as etapas do trabalho para consultas de funções)

<https://blogbr.clear.sale/conheca-e-saiba-como-aplicar-a-acuracia> (Acessado 24/05/2023 às 23:46)

<https://blog.minitab.com/pt/entendendo-analise-de-variancia-anova-e-o-teste-f> (Acessado em 25/05/2023 às 02:46)

<https://www.fm2s.com.br/blog/curva-abc-o-que-e-para-que-serve> (Acessado 24/05/2023 às 03:18)

<https://www.financestrategists.com/wealth-management/fundamental-vs-technicalanalysis/anova/?gclid=Cj0KCQjwmtGjBhDhARIsAEqfDEd_Muu0ujIlKqG6atY99Wd0ivUqG2w7-uD1JQIEa5u1OrtEkk3l-7kaAg9cEALw_wcB> (Acessado em 25/05/2023 às 22:14)