

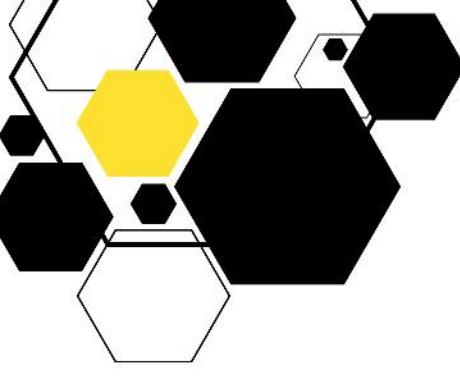


# PROJETO - 1

## Análise de Churn

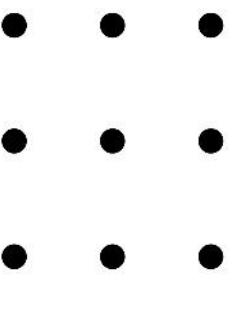
Prof. Rodrigo Vieira





# Descrição do Projeto

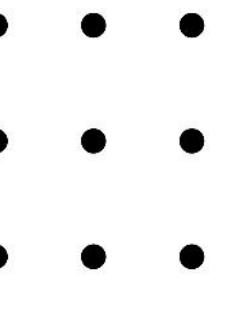
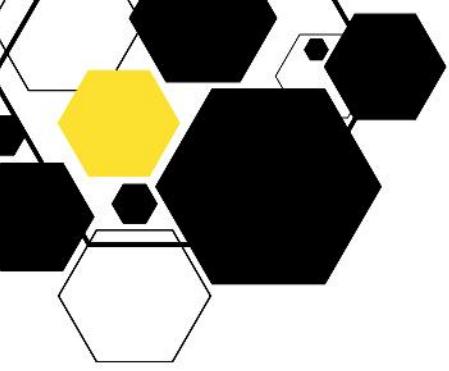
- **Previsão de Churn com Clusterização e Regressão Logística.**
- **Utilização do Iranian Churn Dataset do UCI Machine Learning Repository.**
- **Clusterização dos clientes para identificar padrões de comportamento.**
- **Regressão logística em cada cluster para prever o Churn.**

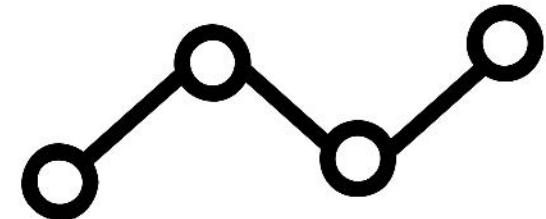
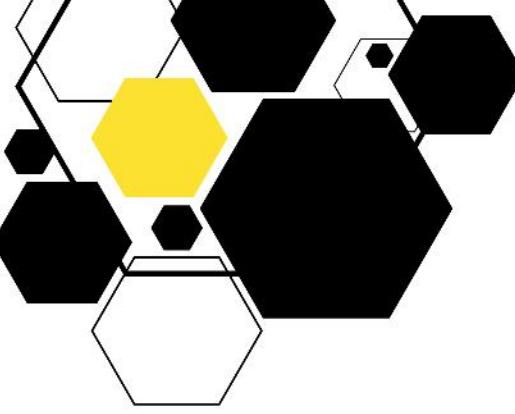




# Resultados Esperados

- Melhor compreensão dos diferentes perfis de clientes.
- Predições mais precisas ao considerar características específicas.
- Estratégias mais direcionadas para retenção de clientes.

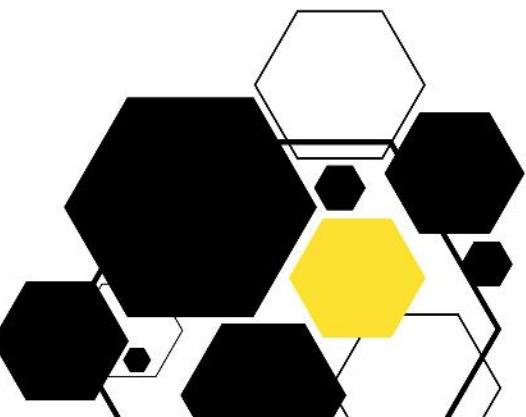


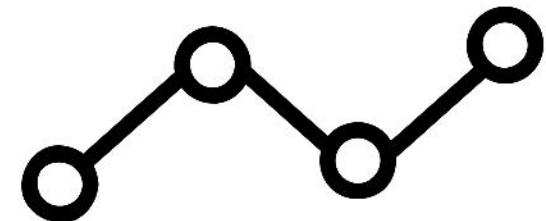
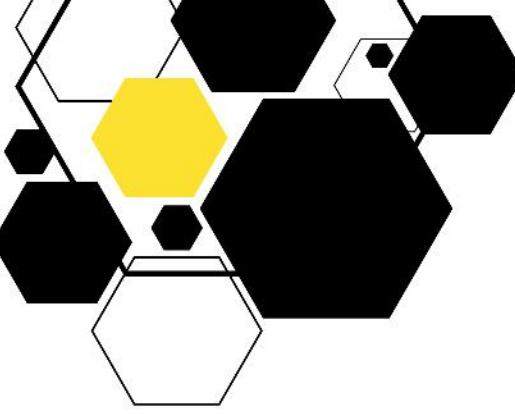


## O QUE É CHURN?

...  
...  
...  
...  
**Churn**, ou taxa de cancelamento, refere-se ao percentual de clientes que deixam de usar os serviços de uma empresa em um determinado período.

- Impacta diretamente a **receita e lucratividade** da empresa.
- Geralmente, **custa mais adquirir novos clientes** do que reter os atuais.
- Ajuda a entender **insatisfações e oportunidades de melhoria** nos serviços.



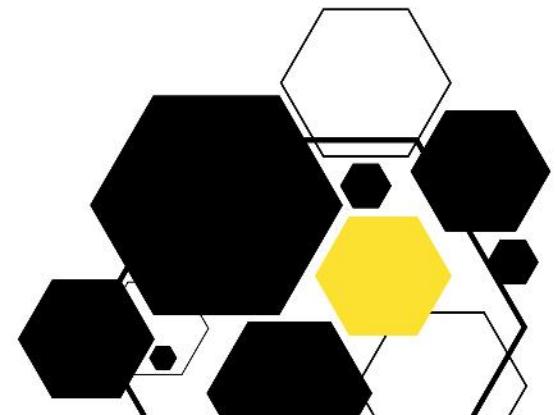


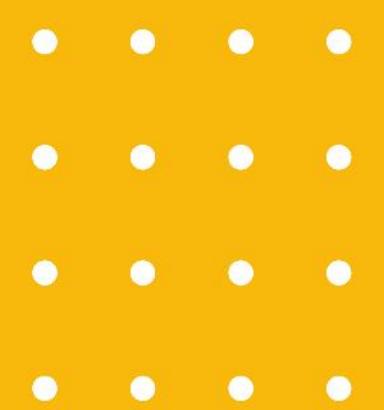
## Tipos de Churn

- **Churn Voluntário** - O cliente decide cancelar o serviço por insatisfação ou oferta melhor de um concorrente.
- **Churn Involuntário** - Ocorre por razões externas, como problemas de pagamento ou mudanças na situação do cliente.

## Como Prevenir o Churn?

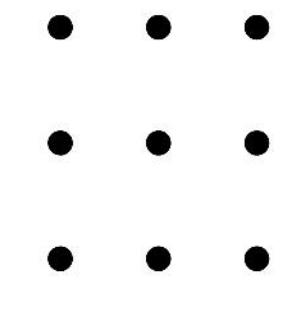
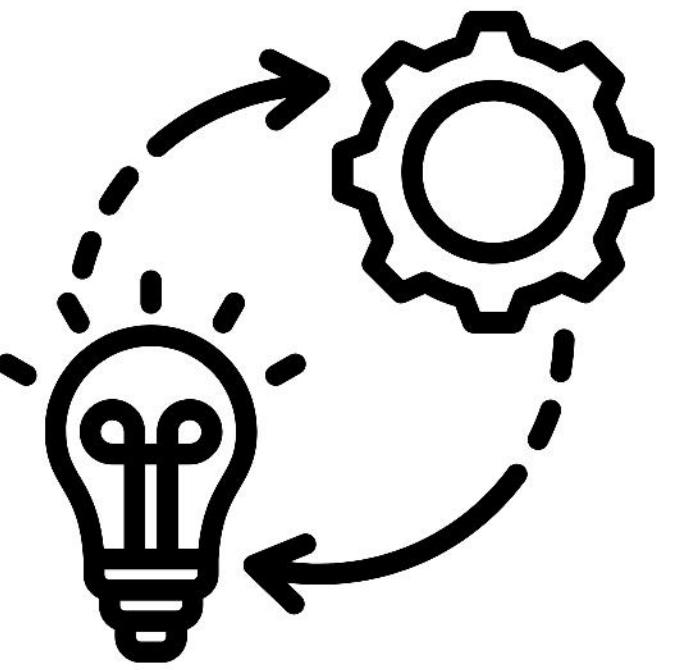
- Melhorando a experiência do cliente.
- Criando ofertas personalizadas.
- Monitorando sinais de insatisfação.

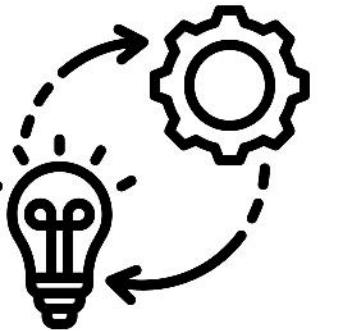
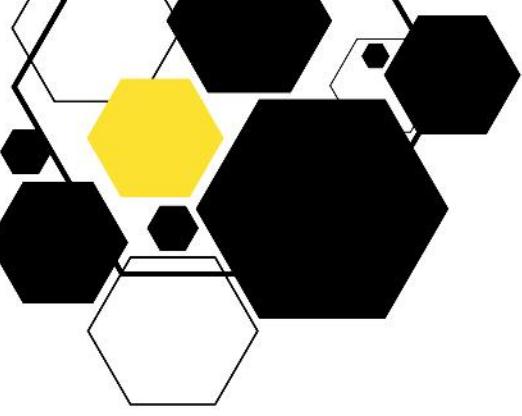




# Considerações Sobre as Técnicas

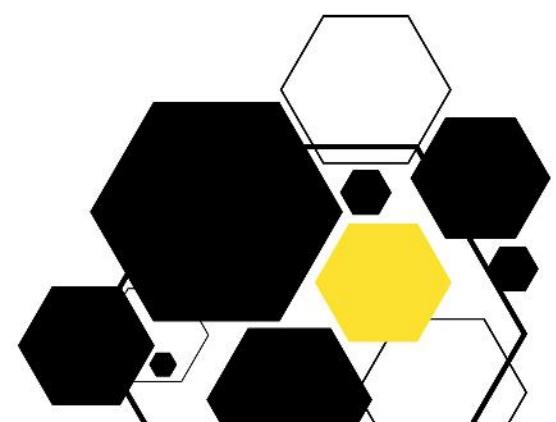
- CLUSTERIZAÇÃO.
- REGRESSÃO LOGÍSTICA.

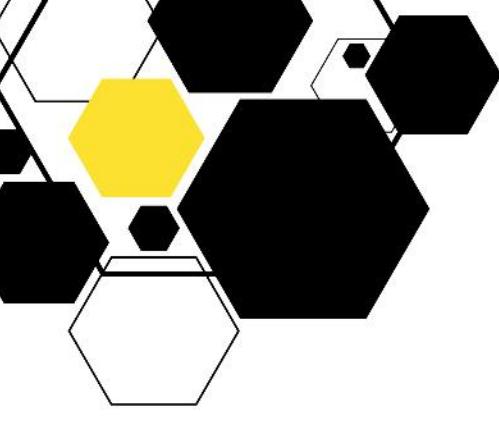
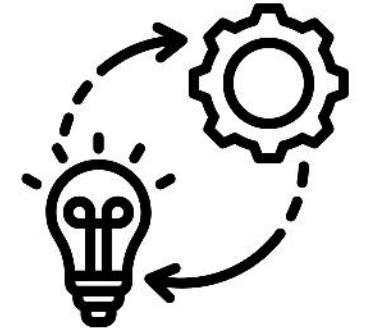




## Clusterização

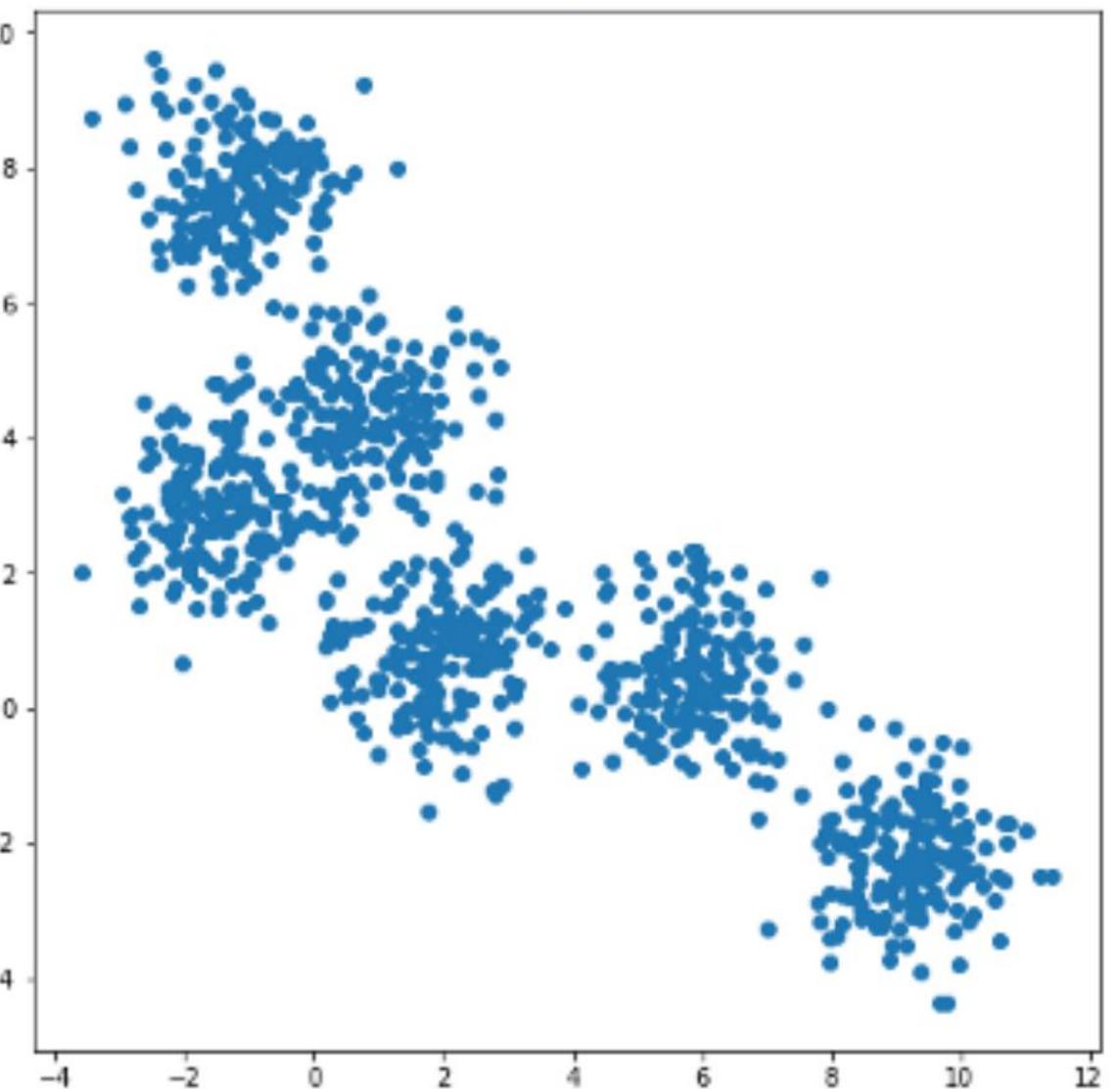
- Técnica de **aprendizado não supervisionado** para segmentação de dados.
- Neste caso o objetivo é **agrupar clientes com perfis semelhantes**.
- Mede a **similaridade entre os dados** usando distâncias (ex: Euclidiana).
- Algoritmos comuns: **K-Means**, DBSCAN, Hierárquico.



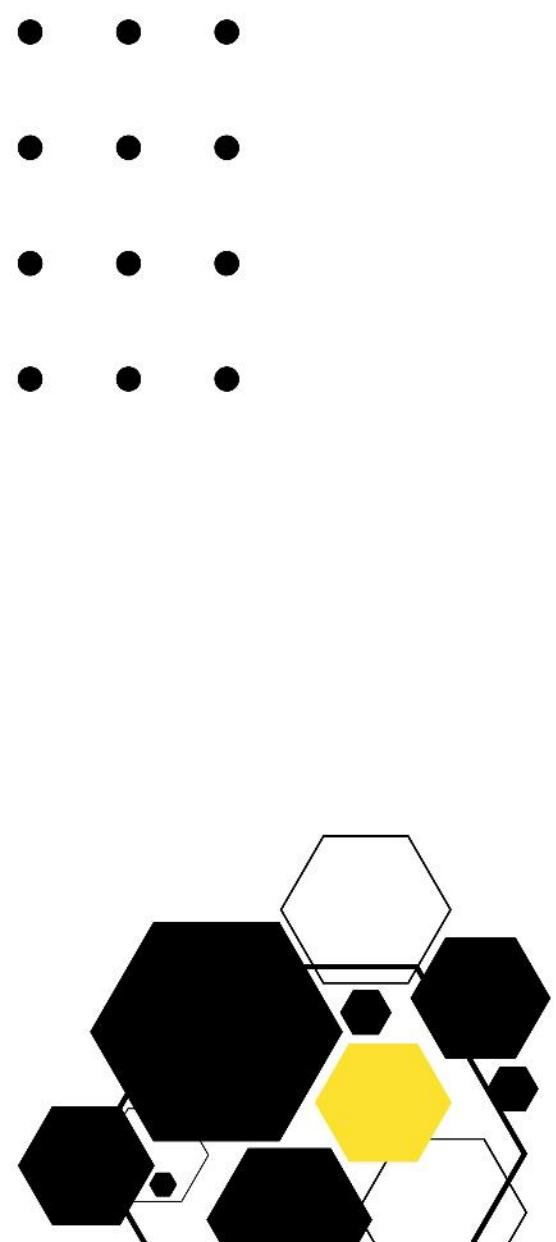
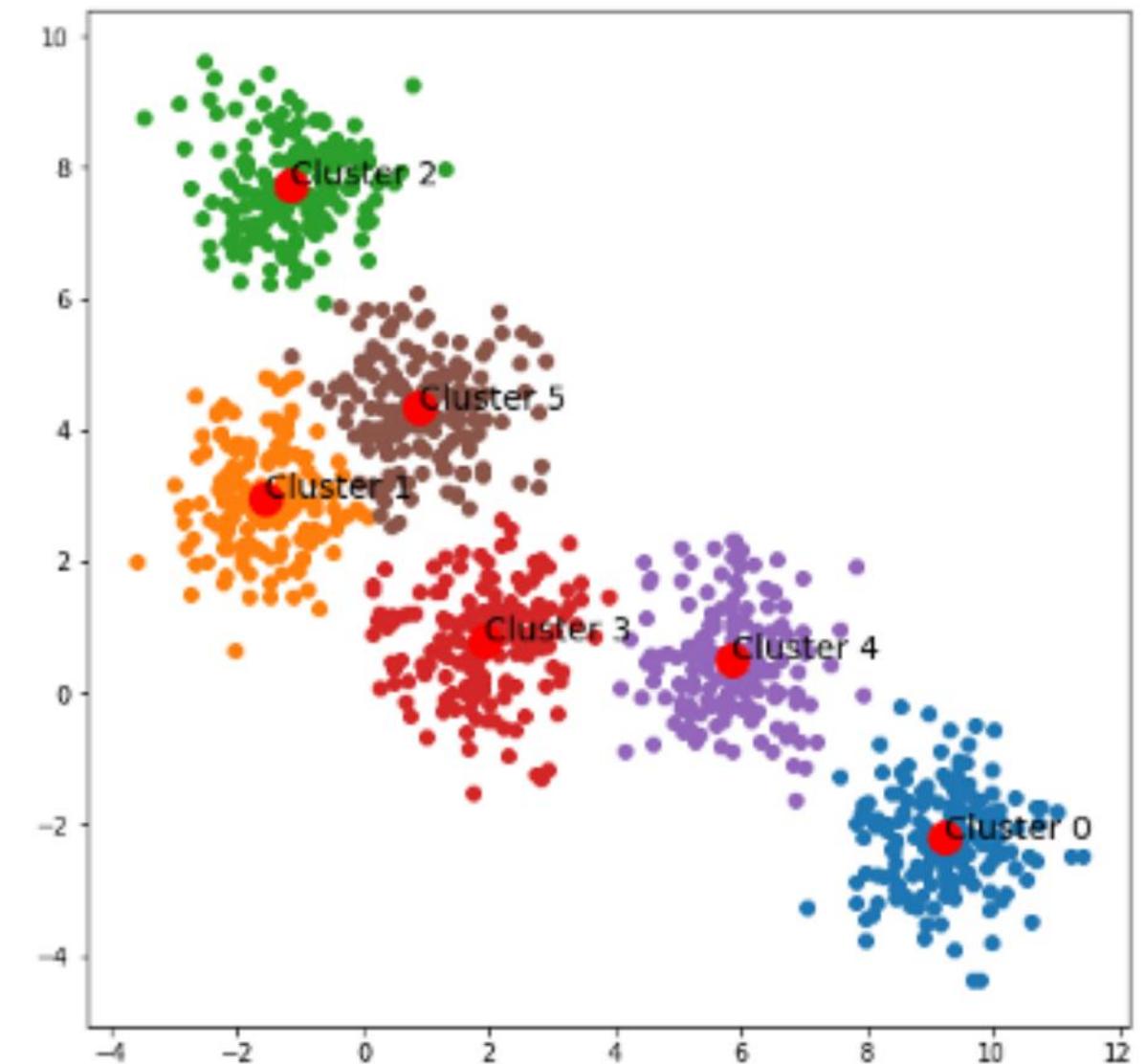


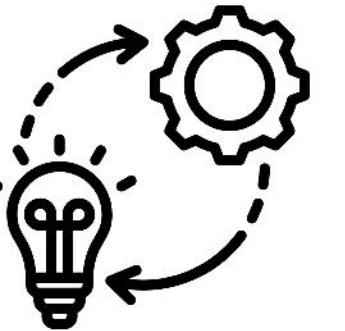
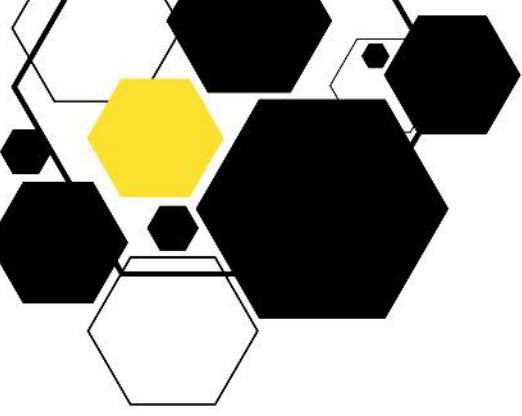
# Clusterização

Input



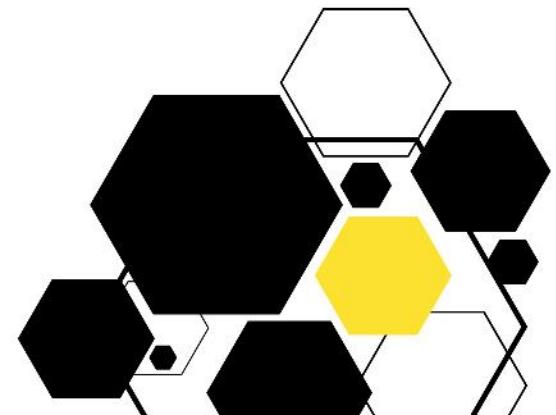
Output

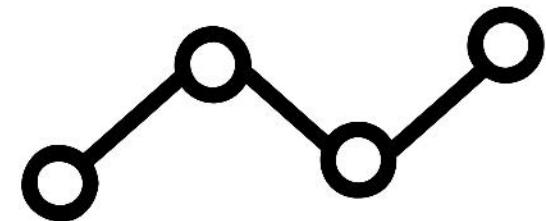
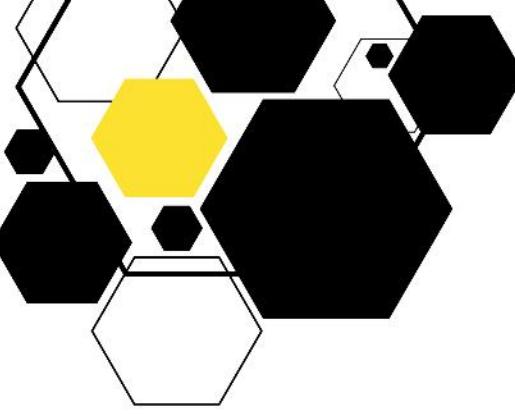




## Clusterização

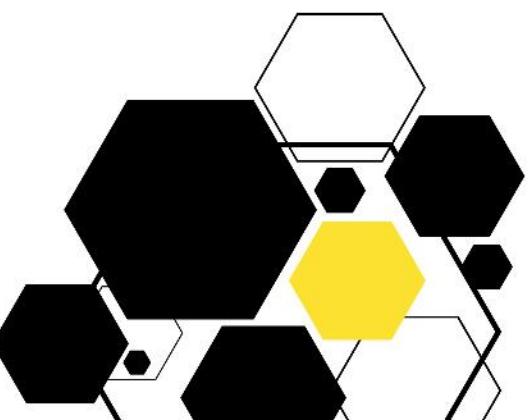
- • •
- Avaliação por Coesão e Separação
  - **Inércia** (Within-Cluster Sum of Squares - WCSS).
  - Mede a compactação dos clusters.
  - Quanto menor, melhor.
- Número Ideal de Clusters
  - **Método do Cotovelo** (Elbow Method).
  - Usa a inércia para encontrar um ponto ótimo.

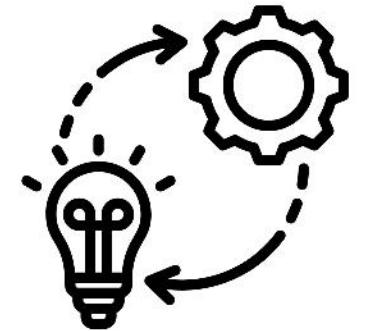
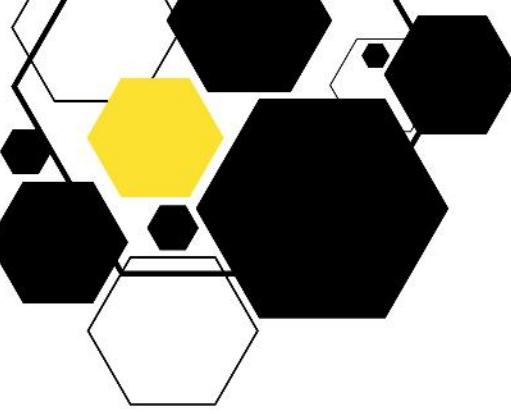




## 📌 Objetivo da Clusterização

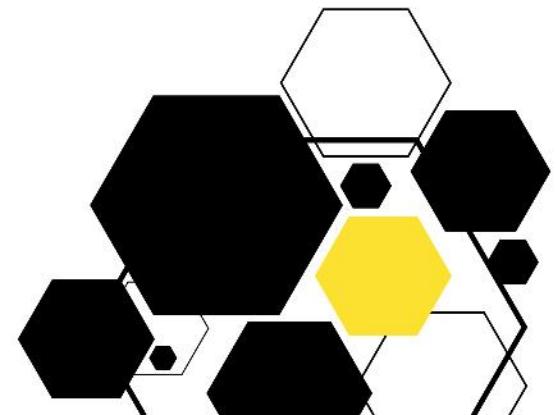
- • •
- • •
- • •
- • •
  - Agrupar clientes com características similares para entender padrões de comportamento.
  - Como interpretar os clusters?
    - Cada cluster representa um grupo de clientes com perfis distintos.
    - Os valores são normalizados (positivos = acima da média, negativos = abaixo da média).

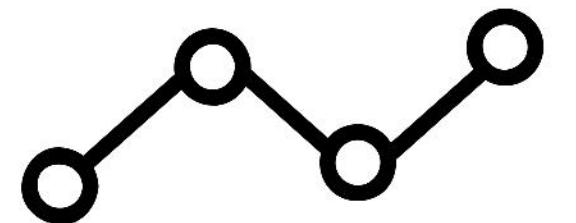
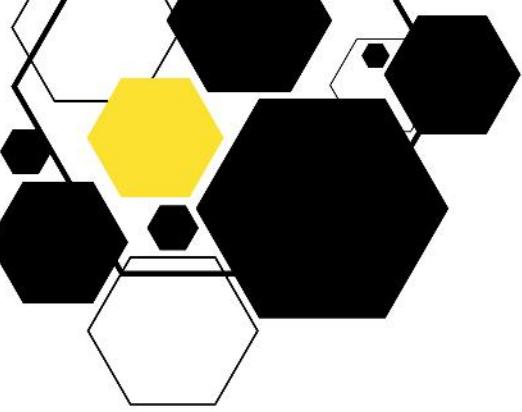




## Regressão Logística

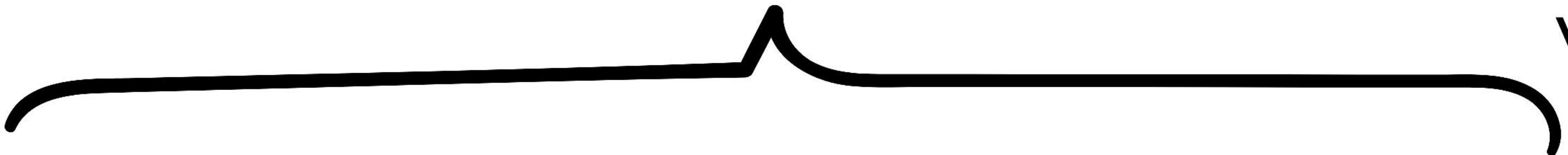
- Técnica de **aprendizado supervisionado** que busca prever a probabilidade de um evento ocorrer.
- No caso, prever o **Churn** (1 = cliente cancela, 0 = cliente permanece).
- Usa uma função logística (**sigmoide**) para transformar variáveis de entrada em uma probabilidade.
- Modelo é **treinado dentro de cada cluster** para prever Churn com mais precisão.



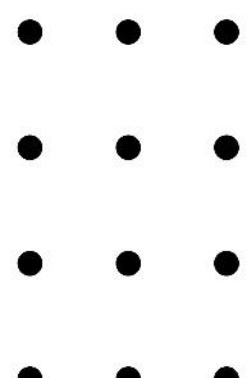


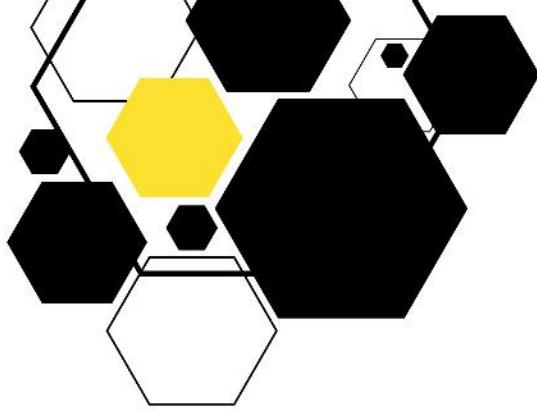
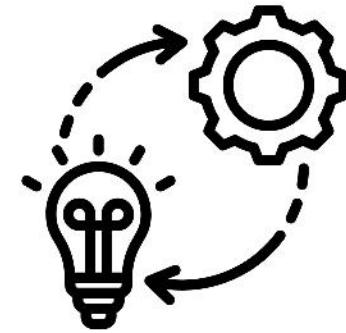
## Variáveis Preditoras

## Variável Alvo



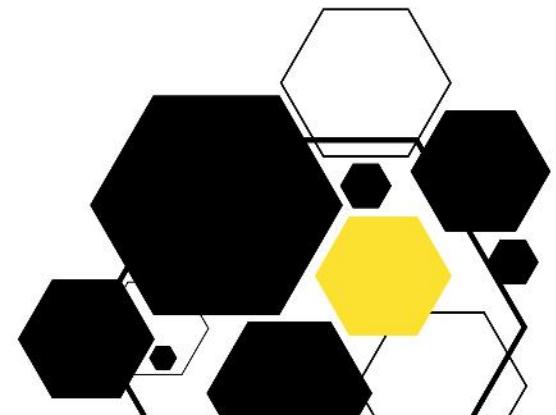
Chamada_Falha	Reclamac... ↓	Tempo_Assina...	Valor_Cobrado	Segundos_Uso	Frequencia_Uso	Frequencia_SMS	Numeros_Disti...	Faixa_Etaria	Plano_Tarifario	Status	Idade	Valor_Cliente	Churn
Number (integer)	Number (double)	Number (integer)											
8	0	Reclamacoes	0	4370	71	5	17	3	1	1	30	197.64	0
0	0	39	0	318	5	7	4	2	1	2	25	46.035	0
10	0	37	0	2453	60	359	24	3	1	1	30	1,536.52	0
10	0	38	0	4198	66	1	35	1	1	1	15	240.02	0
3	0	38	0	2393	58	2	33	1	1	1	15	145.805	0
11	0	38	1	3775	82	32	28	3	1	1	30	282.28	0
4	0	38	0	2360	39	285	18	3	1	1	30	1,235.96	0
13	0	37	2	9115	121	144	43	3	1	1	30	945.44	0
7	0	38	0	13773	169	0	44	3	1	1	30	557.68	0

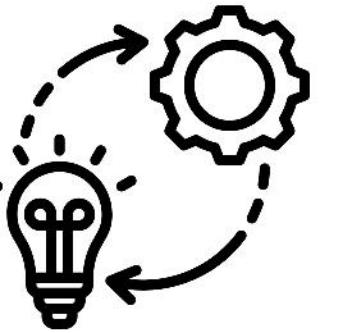




## Regressão Logística

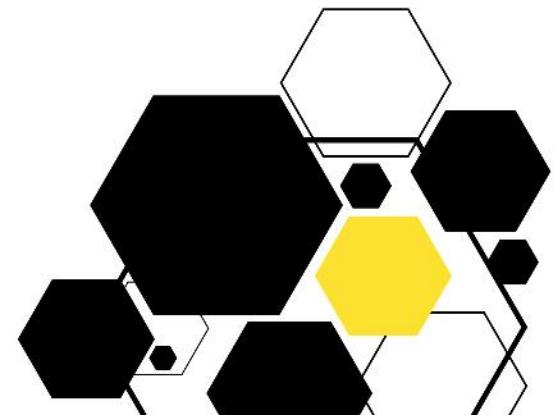
- . . .
- Métricas de Avaliação
  - **Acurácia**: Percentual de previsões corretas.
  - **Precisão** (Precision): % de clientes previstos como Churn que realmente saíram.
  - **Revocação** (Recall): % de clientes que saíram e foram corretamente identificados.
  - **F1-Score**: Média harmônica entre precisão e revocação (bom para dados desbalanceados).
  - **AUC-ROC**: Mede a capacidade do modelo em distinguir classes (curva entre taxa de verdadeiros positivos e falsos positivos).

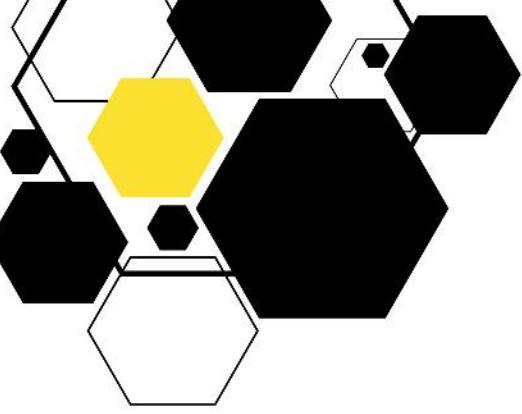




## Regressão Logística

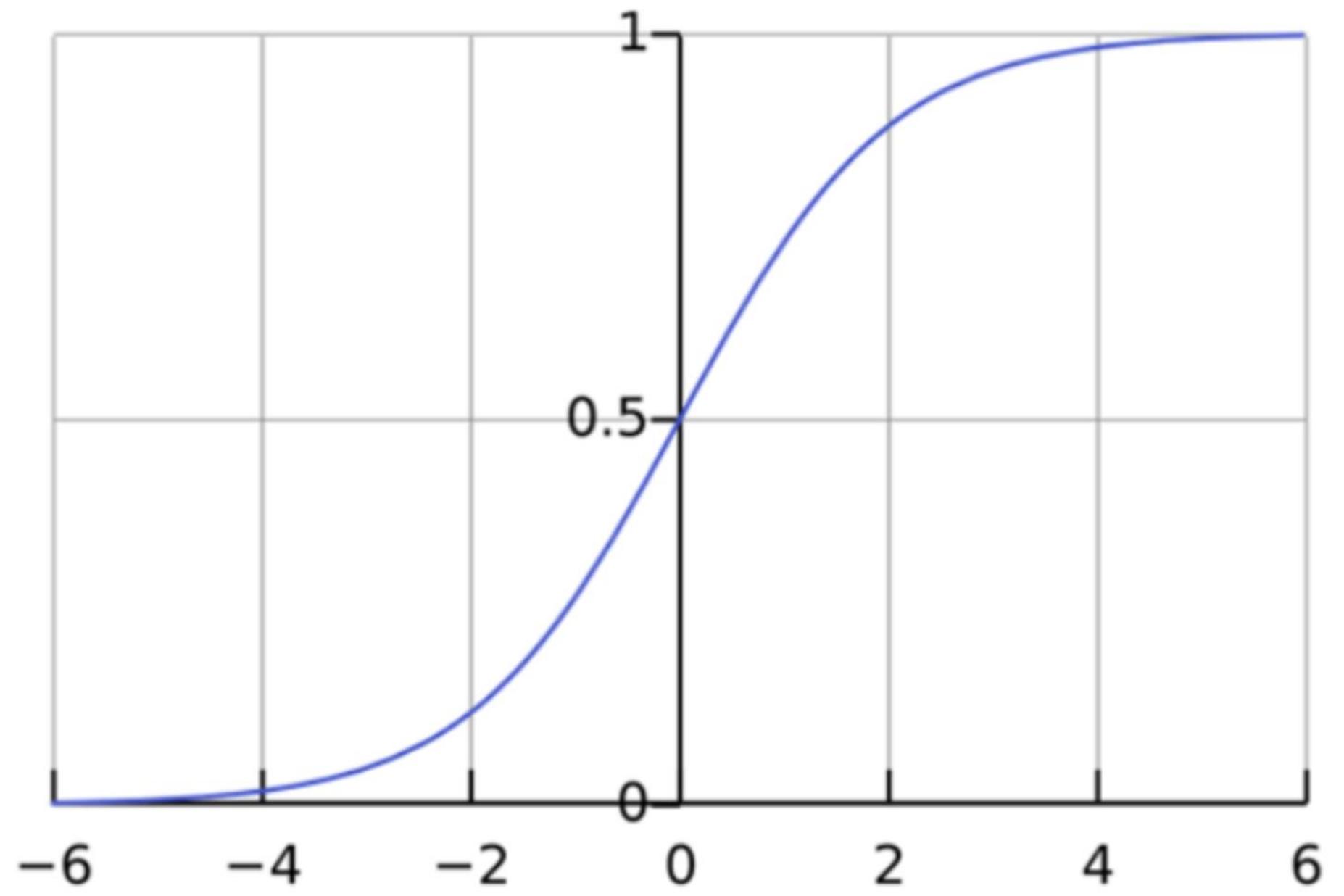
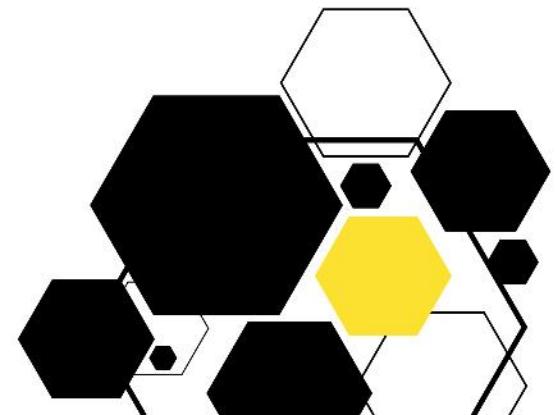
- . . .
- Interpretação das Probabilidades
  - Valores acima de 0.5 geralmente indicam Churn, mas o threshold pode ser ajustado conforme necessidade.
- . . .





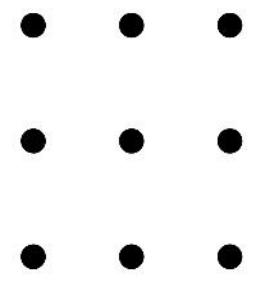
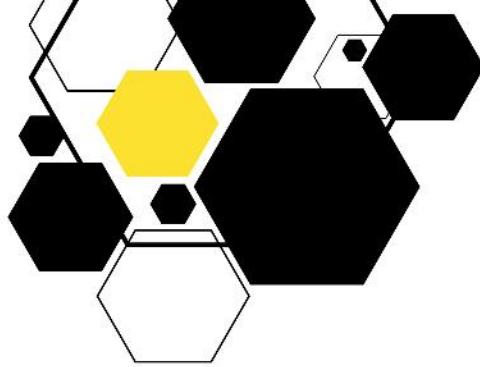
## Regressão Logística

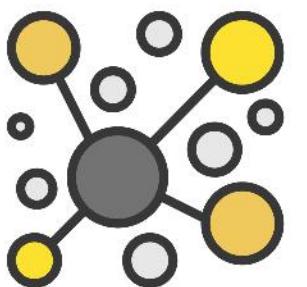
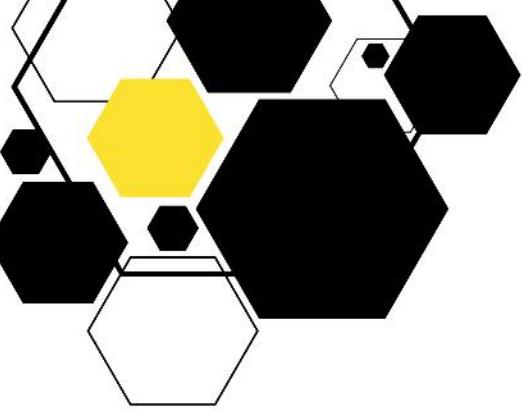
• • •  
• • •  
• • •  
• • •





**BORA POR A MÃO  
NA MASSA!?!**

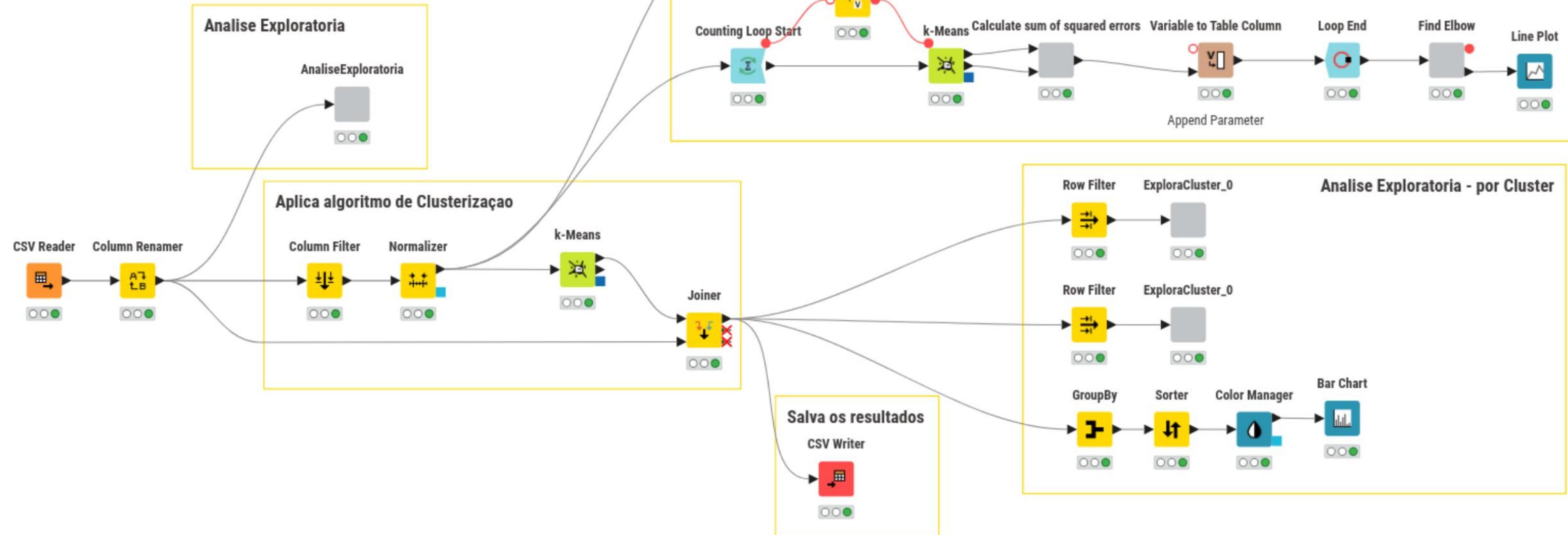


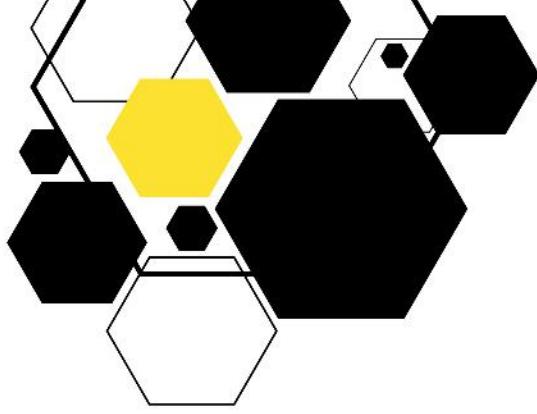
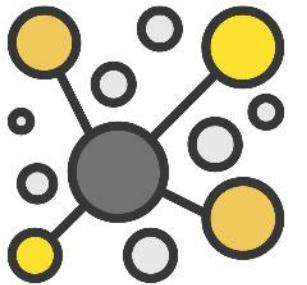


# Clusterização

## Projeto - Iranian Telecon Churn

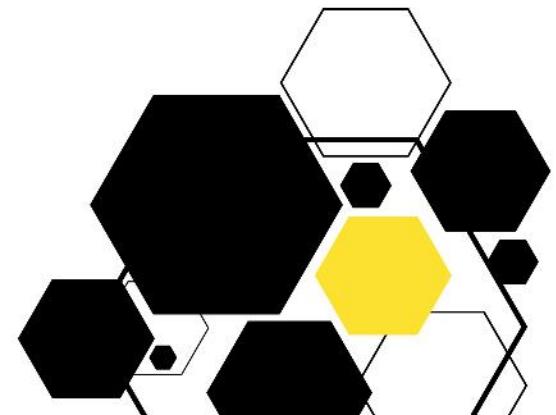
Este workflow demonstra como aplicar o K-means para segmentar dados de clientes de uma empresa de Telecon.  
Posteriormente a segmentação sera usada para treinar modelos de regressao para prever o churn em cada segmento.



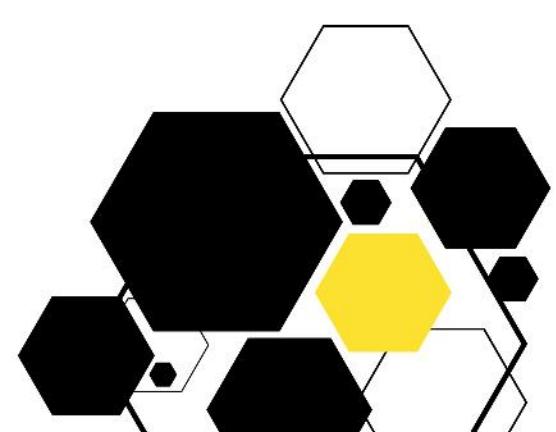
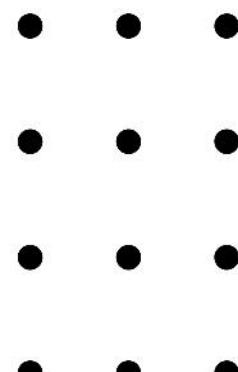
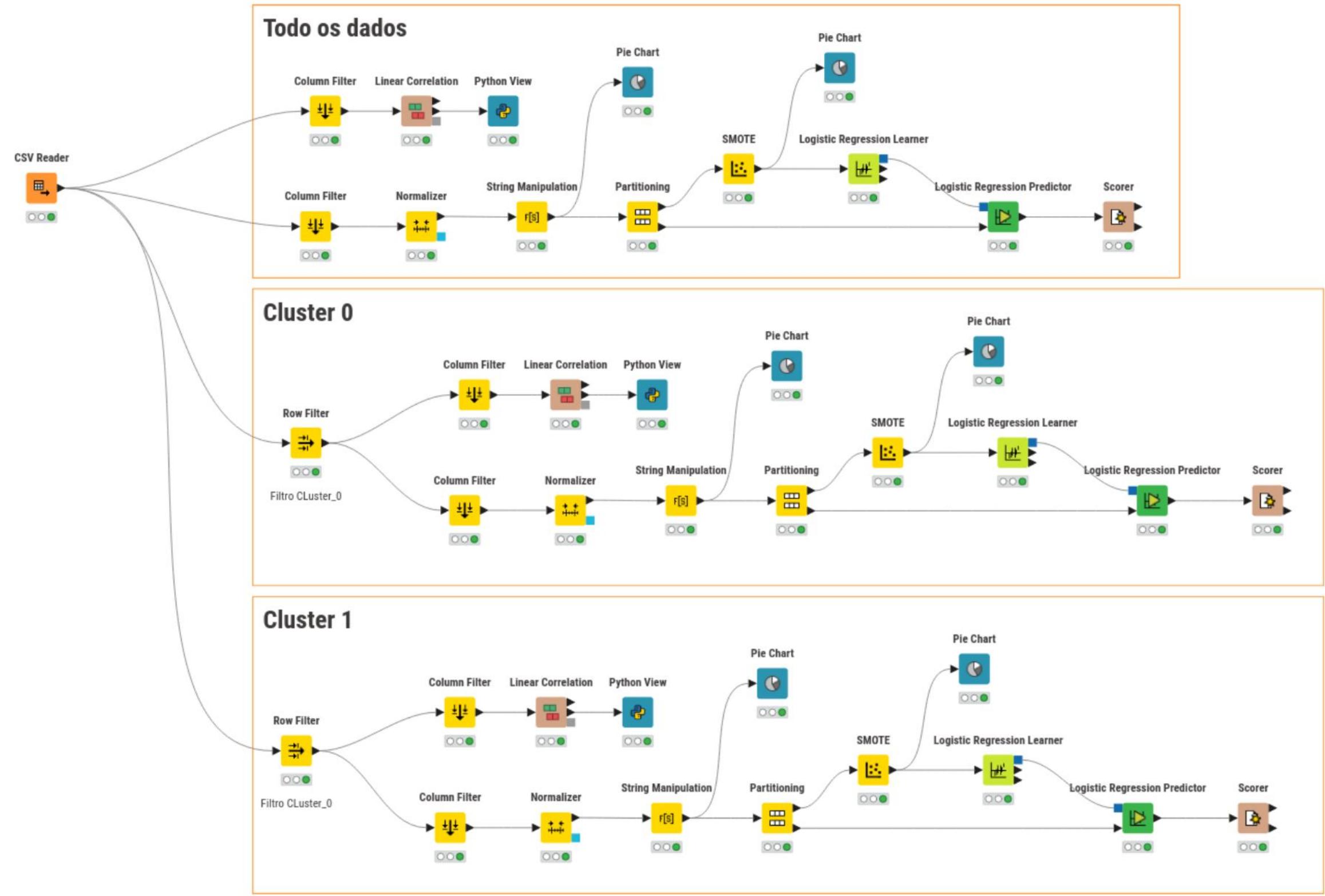
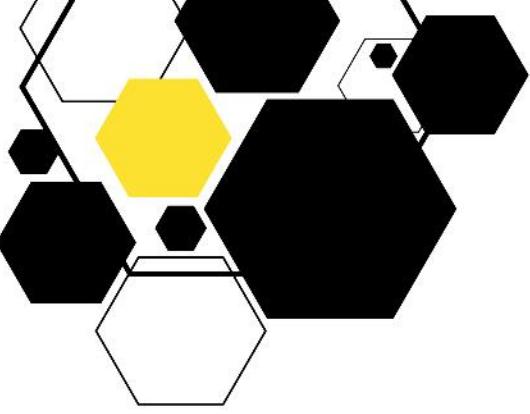
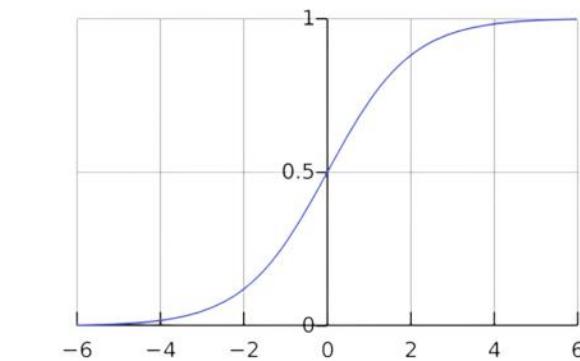


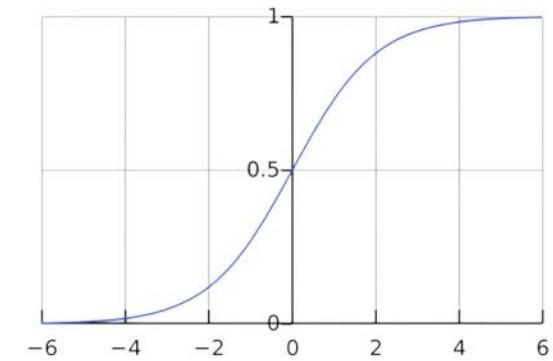
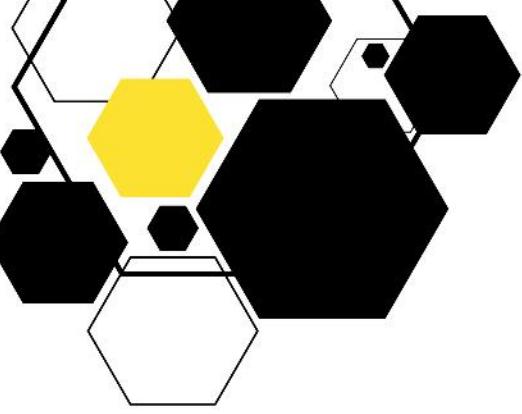
## Resumo sobre os Clusters Identificados

- Cluster 0 – Clientes Fiéis e de Alto Valor
  - • •  Alto uso do serviço (frequência de chamadas, SMS e tempo de uso)
  - • •  Contratos mais longos e maior faturamento
  - • •  Menos reclamações → provável maior satisfação
  - • •  Idade e valor do cliente acima da média
  - ⚠ Baixo risco de churn
  
- Cluster 1 – Clientes de Baixo Engajamento e Maior Risco de Churn
  - ⚠ Menor uso dos serviços (menos chamadas e SMS)
  - ⚠ Assinaturas mais curtas e menor faturamento
  - ⚠ Mais propensos a reclamar → possível insatisfação
  - ⚠ Idade abaixo da média
  - ⚠ Maior risco de churn



# Regressão Logística

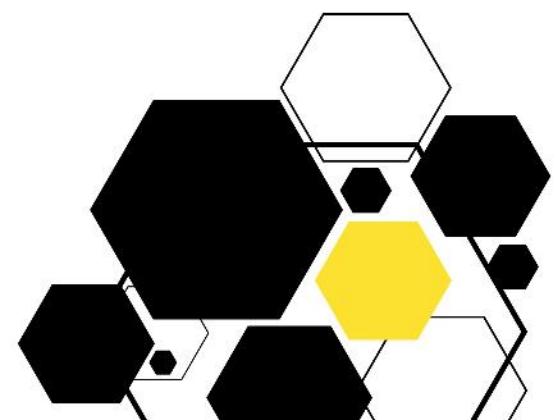


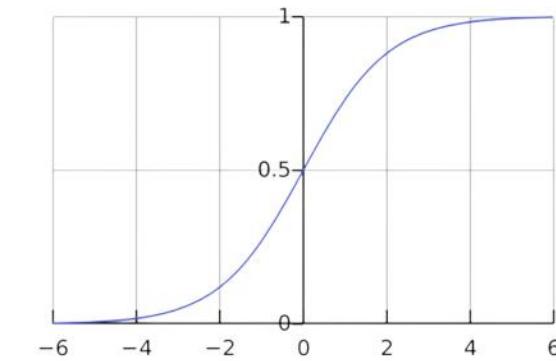
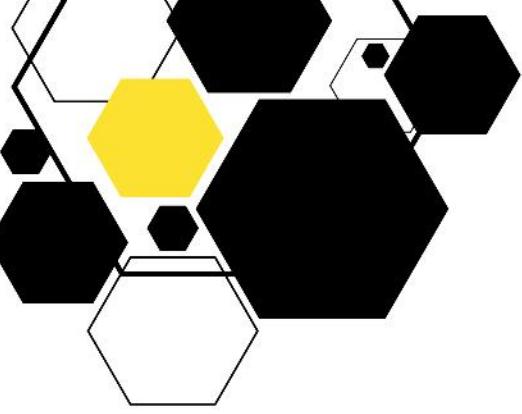


## 📌 Resumo sobre o Modelo Preditivo

### ✓ Para a base como um todo (sem segmentação):

- • •
- • •
- • •
- • •
- Minimizar reclamações deve ser a prioridade número 1 para reduzir churn.
- Incentivar maior gasto e engajamento melhora a retenção, pois clientes que investem mais tendem a permanecer.
- Oferecer planos personalizados para clientes mais antigos e de alto valor, já que esses fatores reduzem o churn.

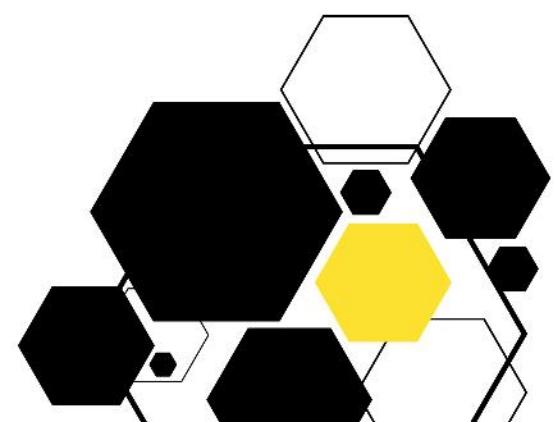


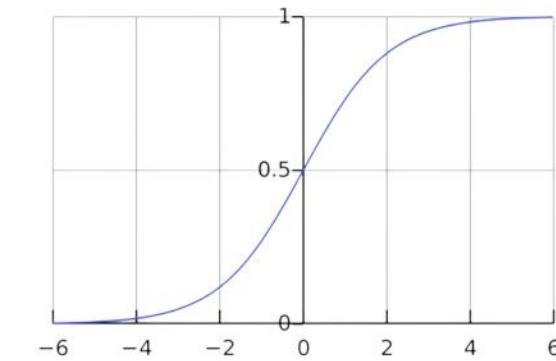
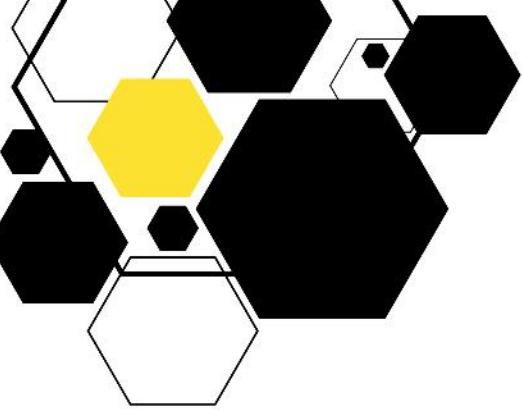


## 📌 Resumo sobre o Modelo Preditivo

### ✓ Para Cluster 0:

- • •
- • •
- • •
- • •
- Foco em clientes com alta retenção natural → reforçar benefícios para quem gasta mais e tem longa assinatura.
- Melhorar atendimento ao cliente para evitar churn por reclamações.
- Explorar planos diferenciados para clientes mais experientes, pois idade e diversidade de contatos influenciam a retenção.



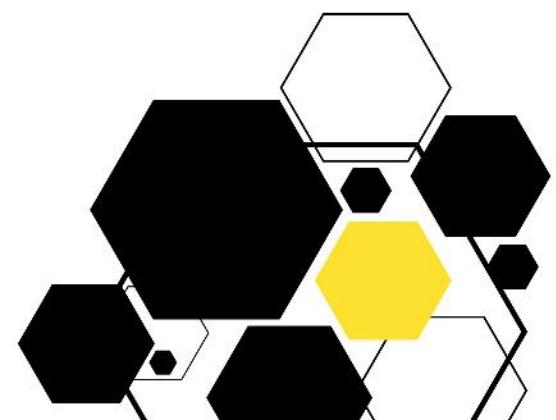


## 📌 Resumo sobre o Modelo Preditivo

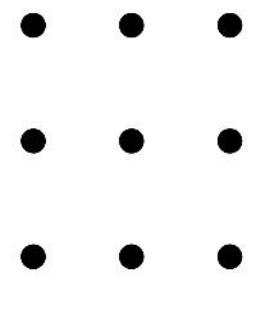
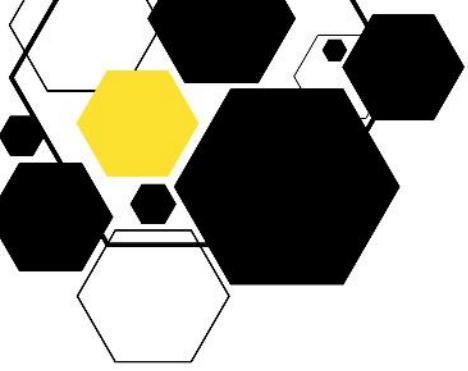
### ✓ Para Cluster 1:

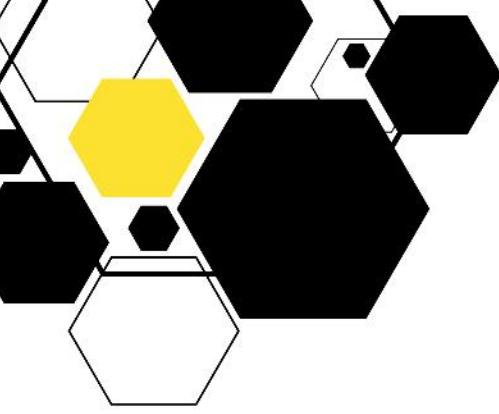
- Identificar e resolver problemas de serviço (**Call Failures** e **Complaints**) rapidamente.
- Analisar o impacto de planos tarifários – pode ser necessário reformular ofertas para esse grupo.
- Avaliar o motivo pelo qual uso intenso de minutos está correlacionado ao churn, pois isso não ocorre nos outros segmentos.

• • •  
• • •  
• • •  
• • •



**PRA FECHAR...**

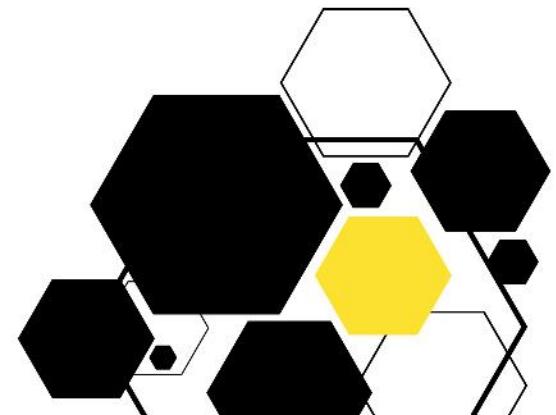


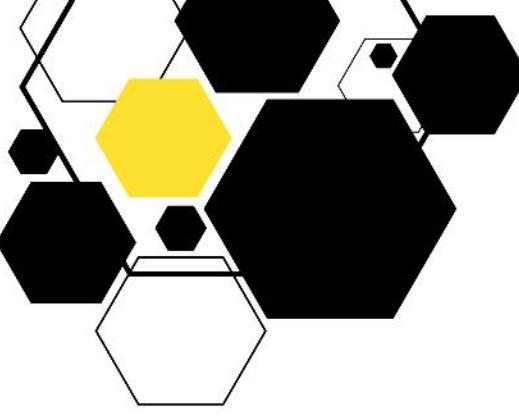


## 🏁 Considerações Finais

Os resultados forneceram insights valiosos sobre os fatores que influenciam o **Churn**, tanto de forma geral quanto segmentado por clusters.

• • •  
• • •  
• • •  
• • •



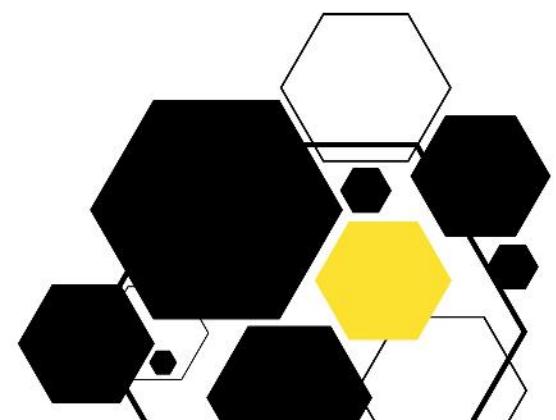


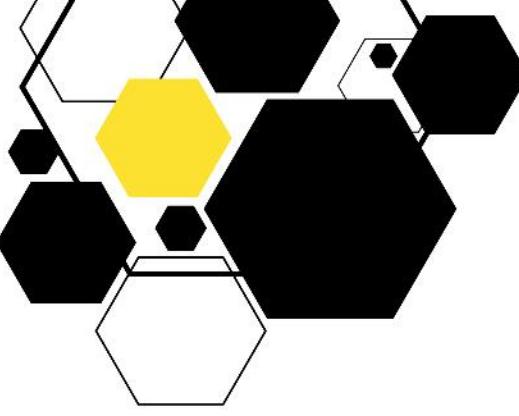
## 🏁 Considerações Finais

• • •  
• • •  
• • •  
• • •  
**🎯 Principais fatores:** Tempo de assinatura, valor gasto e frequência de uso impactam significativamente a retenção de clientes.

• • •  
**🎯 Segmentação por K-Means:** Identifica padrões distintos de comportamento, mostrando que diferentes perfis de clientes possuem motivações variadas para cancelar o serviço.

• • •  
**🎯 Regressão Logística:** Indicou que os fatores determinantes do **Churn** variam entre os grupos, reforçando a necessidade de estratégias personalizadas.





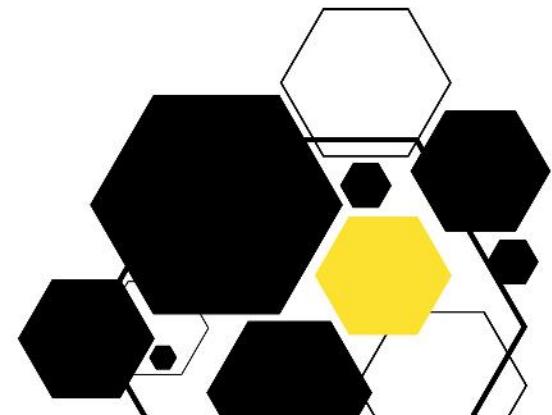
## 🏁 Considerações Finais

• • •  
• • •  
• • •  
• • •  
• • •  
• • •  
**🎯 Principais fatores:** Tempo de assinatura, valor gasto e frequência de uso impactam significativamente a retenção de clientes.

• • •  
**🎯 Recomendações:**

- Adoção de ações direcionadas a cada perfil de cliente para aumentar a retenção.
- Aprimoramento do atendimento e ajuste de planos tarifários conforme o comportamento dos usuários.

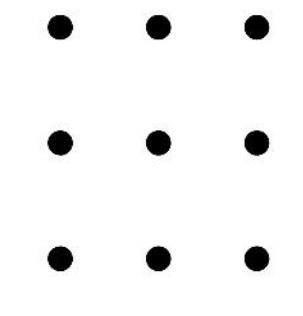
• • •  
**🎯 Análises Futuras:** Explorar modelos mais avançados para prever **Churn** com maior precisão.

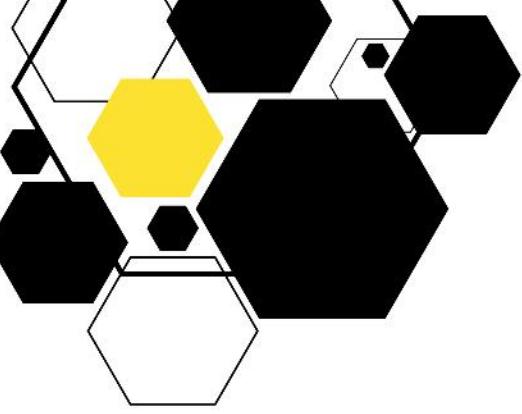




# Esse foi o primeiro projeto...

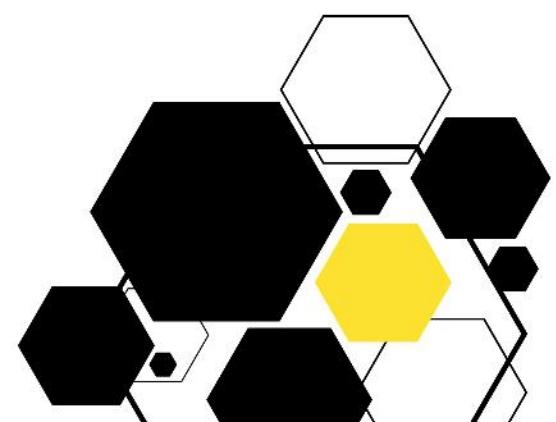
- Queremos apresentar 10 laboratórios...
- + 4 Projetos como esse...

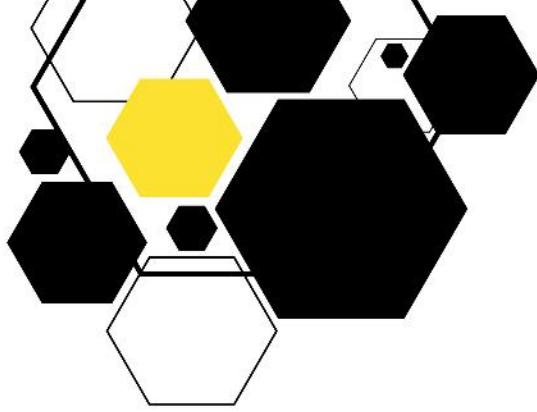
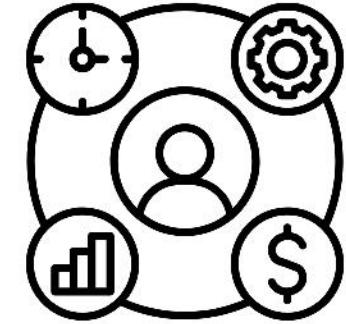




# Laboratórios...

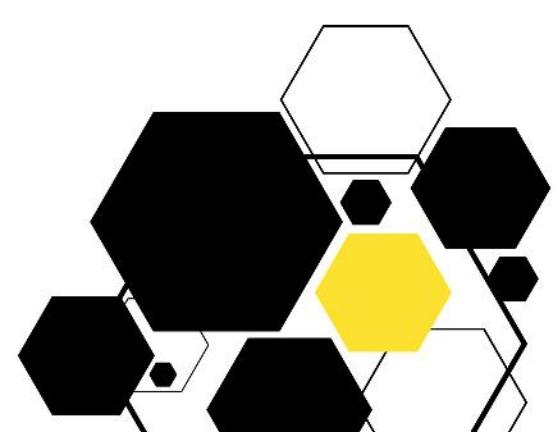
Laboratório	Técnica	Fonte de Dados
Lab_01a	Regressão Linear	Advertising.csv
Lab_01b	Regressão Logística	Social Network Ads
Lab_02a	Classificação (Naive Bayes)	Steel Plates Faults
Lab_02b	Árvore de Decisão	Loan Prediction Dataset
Lab_03	Random Forest	Absenteeism at Work
Lab_04	K-means	Consumer Services Mediated Complaints
Lab_05	PCA	Epileptic Seizure Recognition
Lab_06	Séries Temporais	RSXFS - Federal Reserve
Lab_07	Sistemas de Recomendação	Goodbooks-10k
Lab_08	SVM	Automobile Dataset
Lab_09	Gradiente	Adult Income Dataset
Lab_10	Redes Neurais Artificiais	CO2 Data





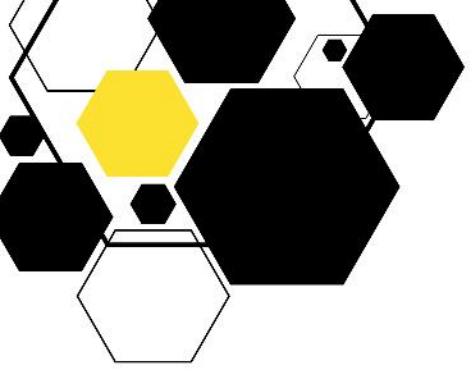
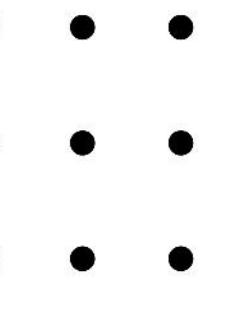
# Projetos...

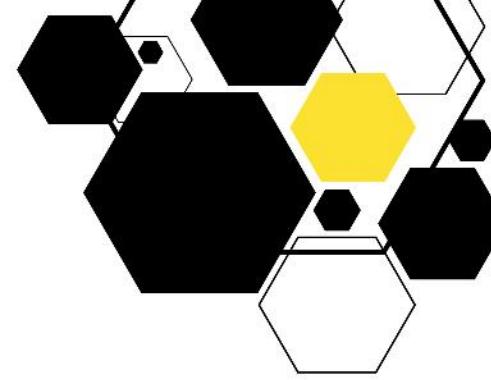
#	Problema	Tec1	Tec2	Fonte de Dados	Data
1	Diagnóstico de doenças baseado em exames	PCA (Redução de Dimensões)	Classificação (SVM)	UCI - Breast Cancer Wisconsin Dataset	22/03
2	Detecção de anomalias em sensores industriais	Métodos Ensemble (XGBoost, Isolation Forest)	Clusterização (DBSCAN)	UCI - Condition Monitoring of Hydraulic Systems	22/03
3	Diagnóstico de falha em peças automotivas	Redes Neurais Convolucionais (CNN)	Classificação (Random Forest)	Kaggle - MVTec Anomaly Detection Dataset	29/03
4	Otimização do consumo de energia em fábricas	Séries Temporais (Prophet)	Aprendizado por Reforço (DQN)	Kaggle - Steel Industry Energy Consumption	29/03





# Bora começar !?!





DATA  
[KN]ience

Muito  
**Obrigado!!!**

Até a proxima aula.

