



# Machine Learning aplicado para análise de comportamento de consumo e previsão de evasão de cliente

Letícia Junqueira Inglez de Souza

Profª Daniele Aparecida Cicillini Pimenta

# Apresentação



Administradora, pós graduada em Gestão e Estratégia em Varejo e cursando MBA em Data Science & Analytics.

Possui 20 anos de experiência na área de varejo. Desenvolveu trabalhos em Business Intelligence, Gestão de Franquias, Gestão Comercial, Gestão de Marketing e Ciência de Dados, atuando em empresas como Schering, Bayer, O Boticário, Humboldt, Liz Lingerie e Caixa Seguradora.

Em 2021 realizou uma transição de carreira e atualmente está dedicada a área de Ciência de Dados realizando modelagens de Machine Learning e Inteligência Artificial no Grupo Via – banQi.

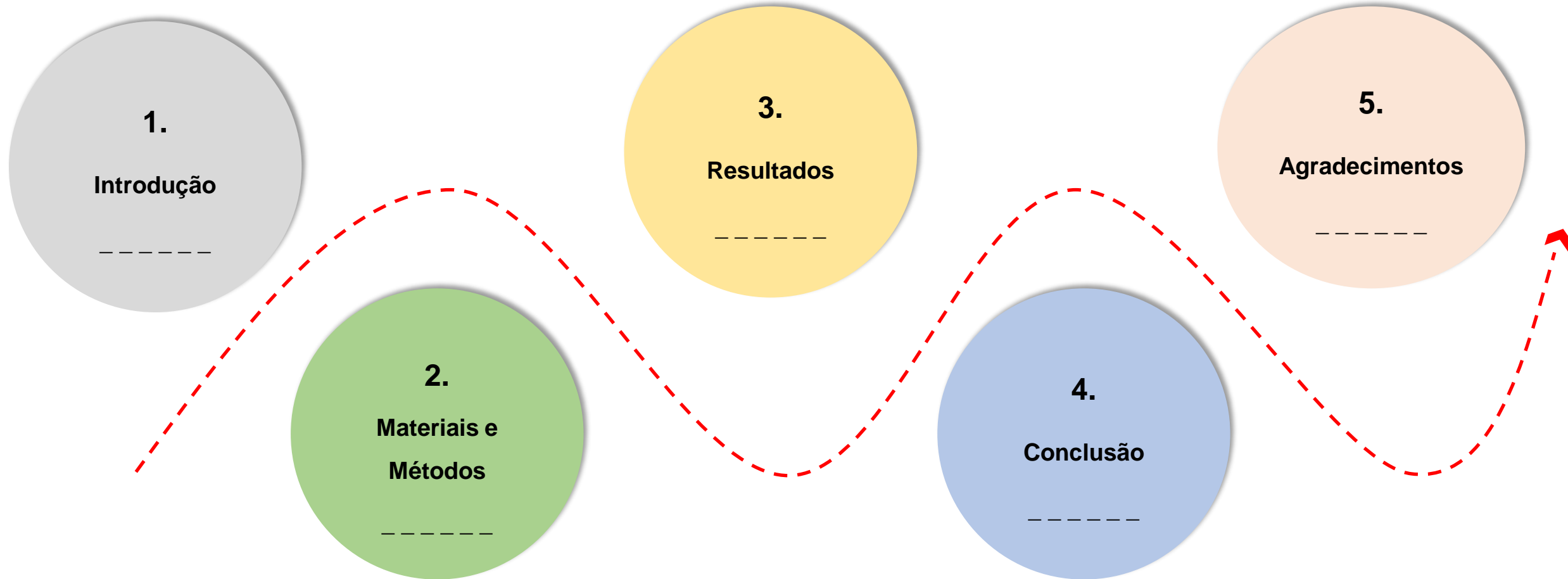
Apaixonada por comportamento de consumo, deseja por meio da análise de dados somada às suas experiências profissionais e pessoais, proporcionar soluções criativas para o mercado.

Acredita que a educação é a solução para promover um mundo inclusivo, diverso, equilibrado, responsável e justo.



<https://www.linkedin.com/in/leticiainglez/>

# Estrutura da apresentação:



# Introdução:



**1ª Revolução Industrial**  
(séc. XVIII)

Mecanização,  
introdução da  
**máquina a vapor** e  
do carvão



**2ª Revolução Industrial**  
(séc. XIX)

Produção em  
massa, **linha de montagem**, com  
base em petróleo e  
eletricidade



**3ª Revolução Industrial**  
(séc. XX)

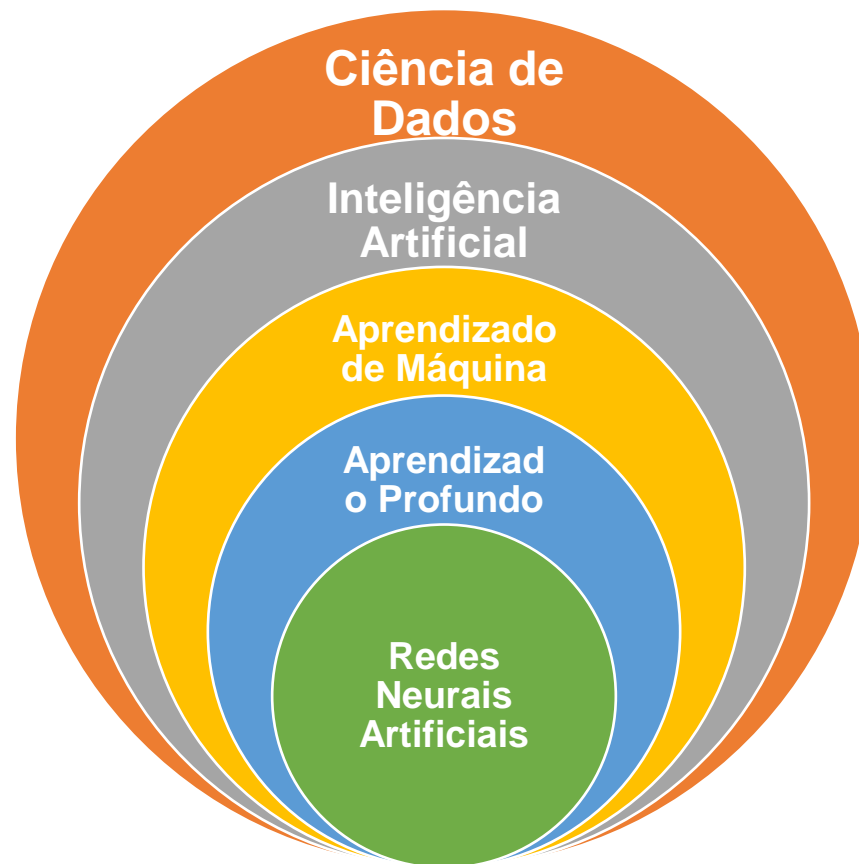
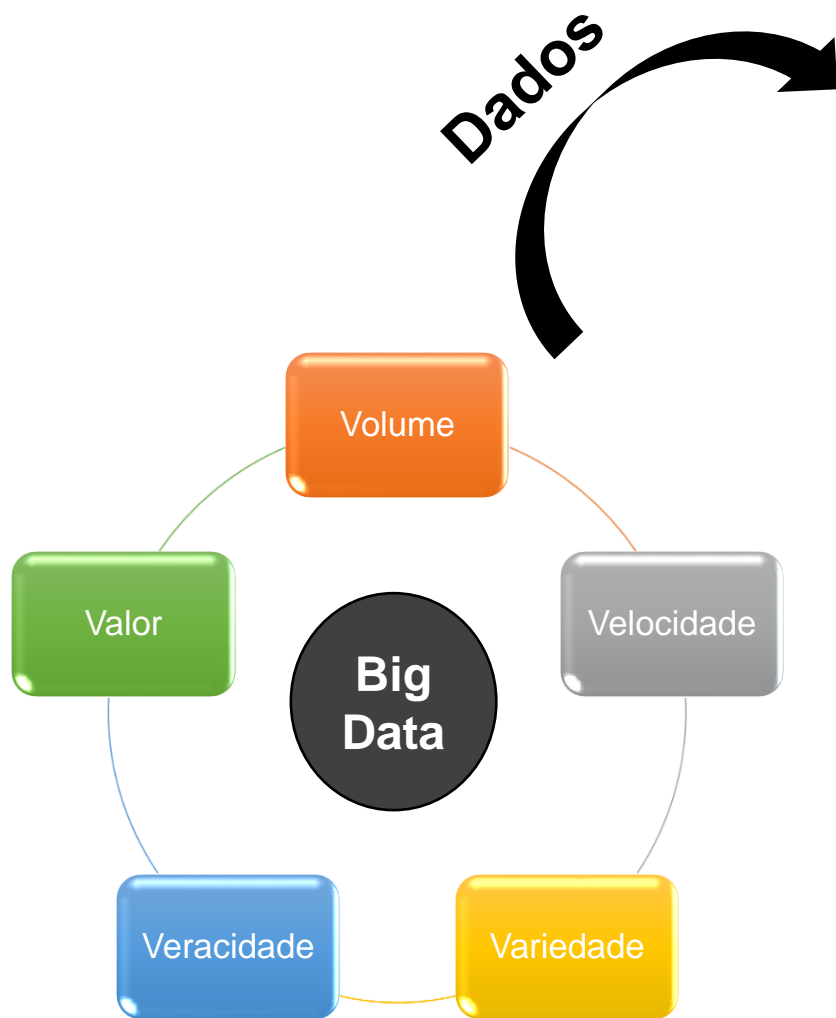
**Produção automatizada**,  
utilizando  
computadores,  
eletrônico e TI



**4ª Revolução Industrial**  
(Hoje)

**Produção**  
inteligente,  
incorporada com a  
**internet das coisas**  
(IoT) e **Big Data**

# Introdução:



Processamento  
e Interpretação

- ✓ Compreensão comportamentos
- ✓ Soluções personalizadas
- ✓ Competitividade
- ✓ Satisfação cliente



# Introdução:

Gestão relacionamento  
e satisfação cliente



**“Churn” / Evasão**

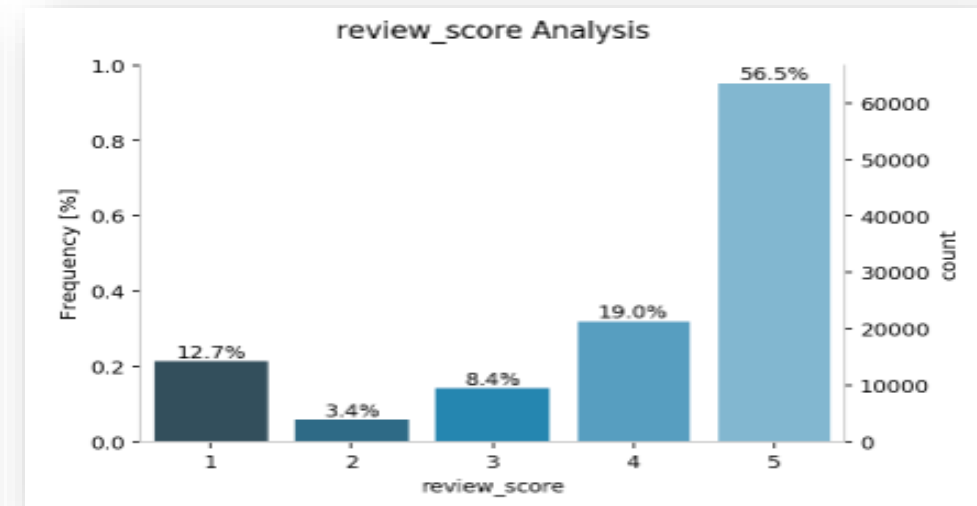
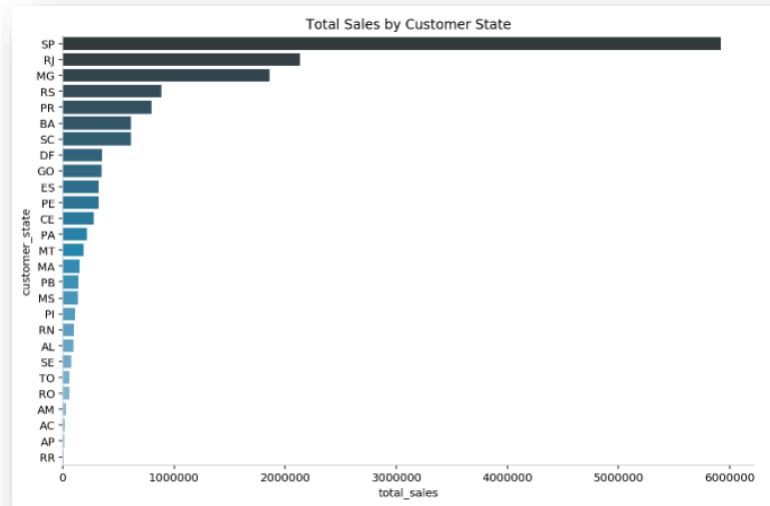
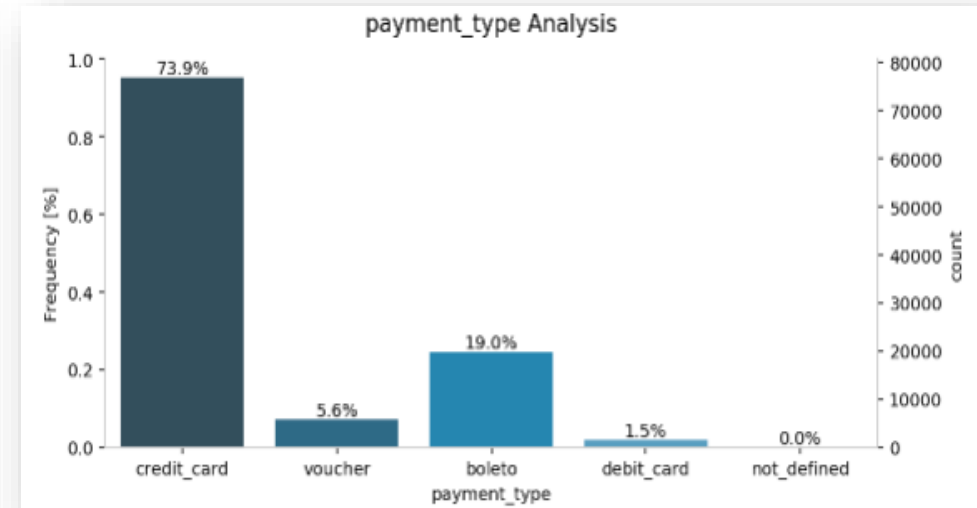
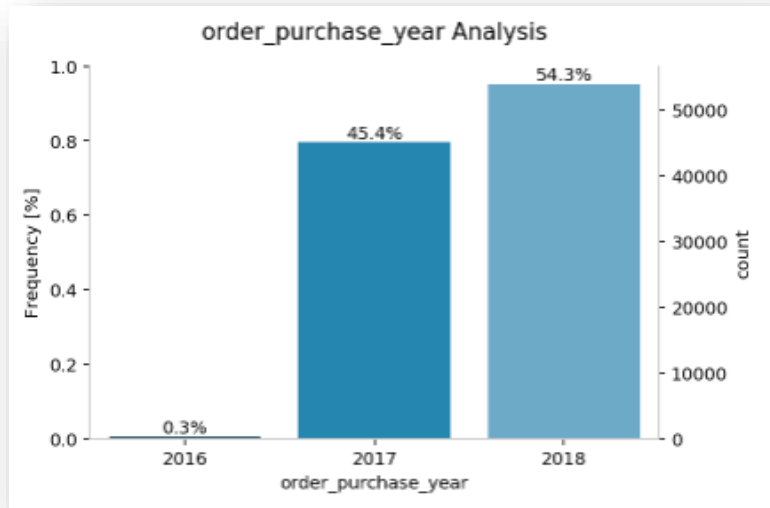
# Materiais e Métodos:

## Considerações sobre dos Dados

<b>Metodologia</b>	CRISP - processo de execução do projeto de Ciência de Dados de forma cíclica
<b>Fonte de Dados</b>	Dados públicos. Pedidos realizados na loja de comércio eletrônico (“Olist Store”) plataforma Kaggle
<b>Período da análise</b>	Entre os anos 2016 e 2018
<b>Ferramentas</b>	Jupyter Notebook e Linguagem Python
<b>Bibliotecas Python</b>	Pandas, Numpy, Seaborn, Matplotlib, Datetime, Lifetimes, Yellowbrick, Sklearn
<b>Bases de Dados</b>	<ul style="list-style-type: none"><li>○ Pagamentos dos pedidos (olist_order_payments_dataset)</li><li>○ Pedidos (olist_orders_dataset)</li><li>○ Clientes (olist_customers_dataset)</li><li>○ Avaliações feitas pelos clientes (olist_order_reviews_dataset)</li><li>○ Itens dos pedidos (olist_order_items_dataset)</li><li>○ Produtos pedidos (olist_products_dataset)</li></ul>

# Materiais e Métodos:

Análise  
exploratória  
dos dados  
após merge  
das bases





# Materiais e Métodos:

“Churn” (abandono) está diretamente associado ao tempo em que o indivíduo permanece como cliente de uma organização, portanto está diretamente relacionado ao conceito de “Customer Lifetime Value” (valor do ciclo de vida do cliente) ou da análise RFM (Recência, Frequência, Monetária)

## Recência



**Quão recente  
foi a última  
compra do  
cliente**

## Frequência



**Com que  
frequência o  
cliente  
compra**

## Valor Monetário



**Quanto o cliente  
costuma gastar**

$CLV = \text{número esperado de transações} * \text{receita por transação} * \text{margem}$

Onde o primeiro elemento refere-se ao modelo BG/NBD, o segundo elemento ao modelo Gamma-Gamma e a margem é definida pelo negócio

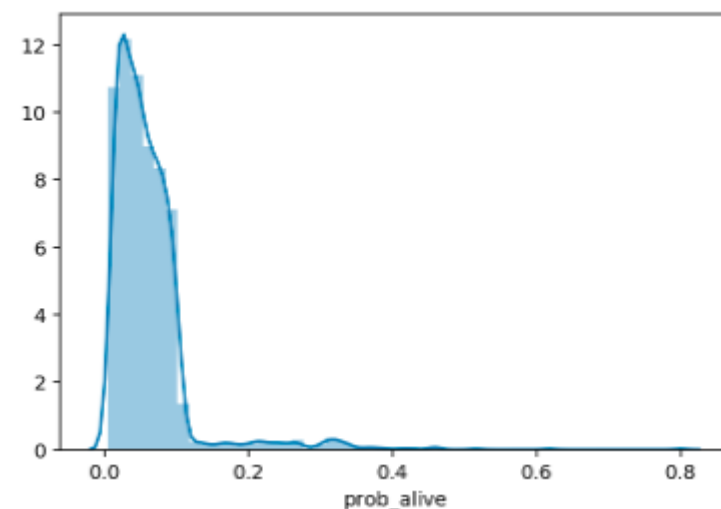
# Materiais e Métodos:

**Modelo Beta Geométrica / Binomial Negativa (BG/NBD):** modela a distribuição dos comportamentos de compra de cada cliente e prevê o número esperado de transações para cada cliente

**Submodelo Gamma-Gamma:** modela a distribuição de lucro médio esperado e prevê o lucro médio esperado para cada cliente

Biblioteca Python: Lifetimes

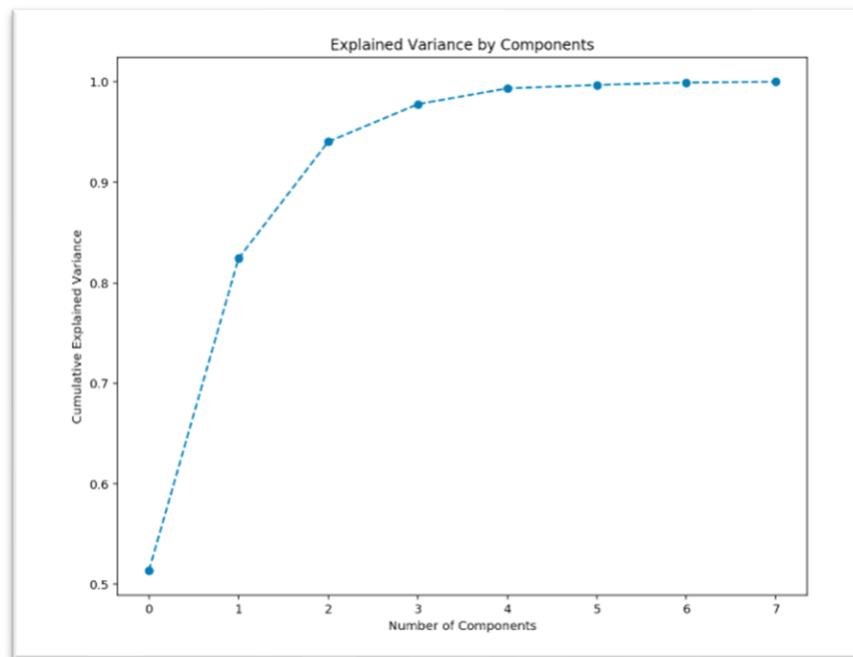
	customer_unique_id	frequency	recency	T	monetary_value	prob_alive	churn	purchase_next_6_month	LTV_next_6_month
0	004288347e5e88a27ded2bb23747066c	1.0	171.0	403.0	87.90	0.053200	churned	0.017419	8.255746
1	004b45ec5c64187465168251cd1c9c2f	1.0	267.0	367.0	27.00	0.082272	churned	0.028814	8.015718
2	0058f300f57d7b93c477a131a59b36c3	1.0	31.0	196.0	72.58	0.033793	churned	0.017722	5.146212
3	00a39521eb40f7012db50455bf083460	1.0	11.0	103.0	11.55	0.039748	churned	0.028697	3.651495
4	011575986092c30523ecb71ff10cb473	1.0	60.0	198.0	63.90	0.046973	churned	0.024491	6.718611



Definido que clientes abaixo de 10% de chance de estar vivo já é considerado churn/evasão

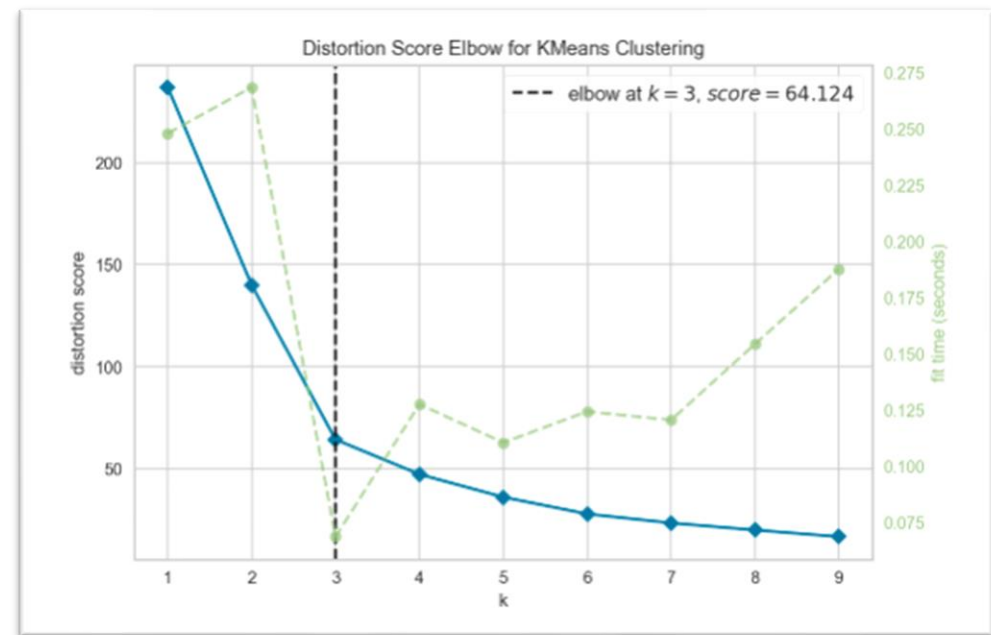
# Materiais e Métodos:

Adicionada a análise RFM, aplicou-se o modelo **PCA** – **Análise dos Componentes Principais** para a redução de dimensionalidade da base de dados



Método variance ratio para identificação quantidade de componentes principais que explicam a variância

Após o PCA aplicou-se o modelo de clusterização ou **agrupamento K-means** com o objetivo de reconhecer os padrões dos clientes quanto ao abandono do negócio e segmentá-los, de formar a propor estratégias de comunicação e/ou novos produtos/serviços dirigida a eles.



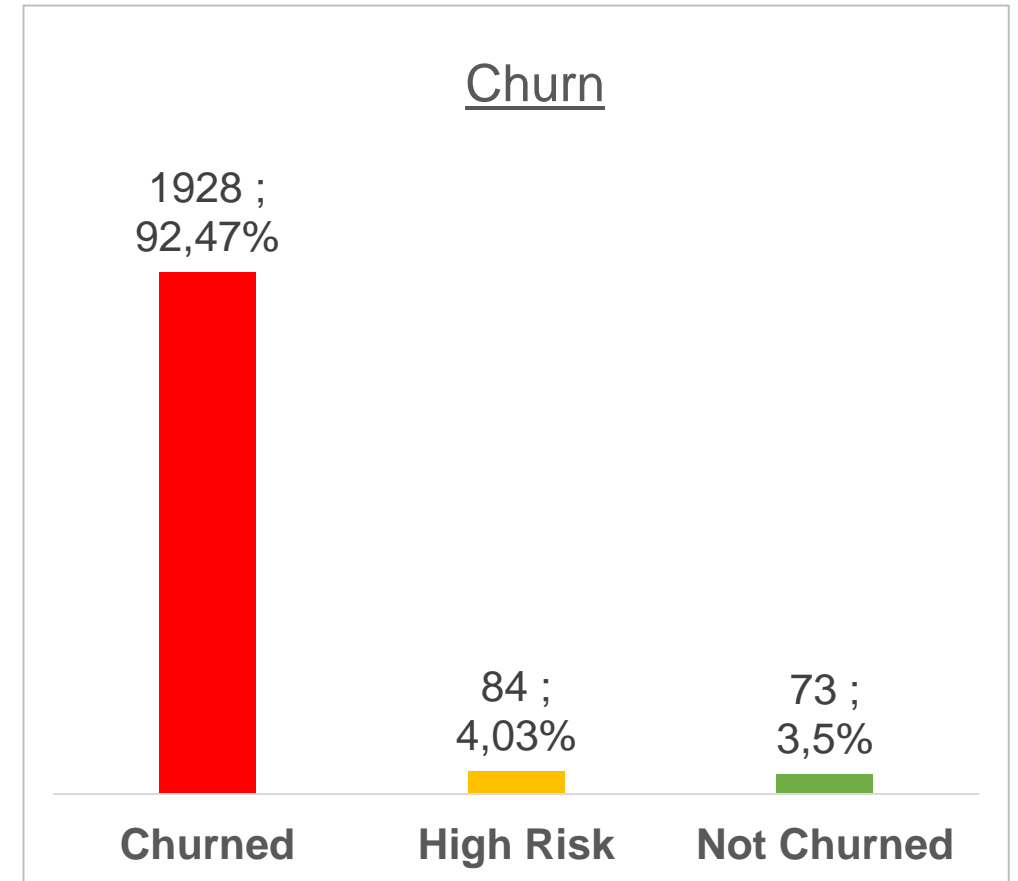
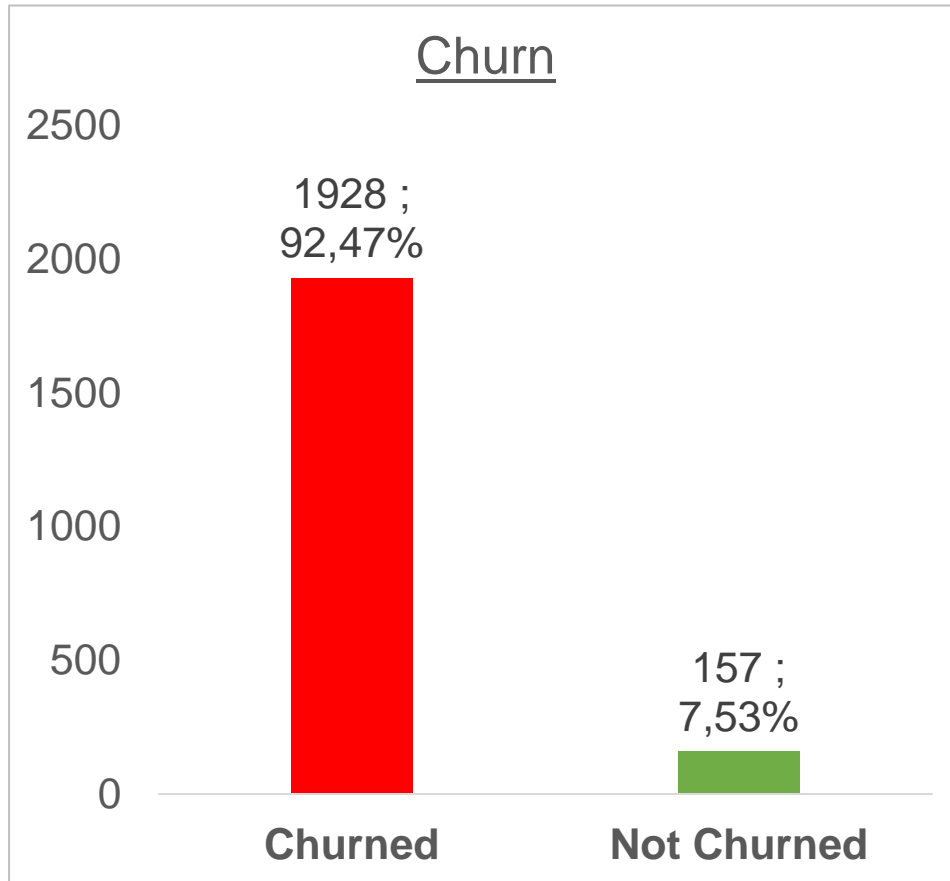
Método Cotovelo para identificação quantidade ideal de clusters

# Resultados:

	frequency	recency	T	monetary_value	prob_alive	purchase_next_6_month	LTV_next_6_month
count	2085	2085	2085	2085	2085	2085	2085
mean	1.107	122.805	319.305	113.264	0.061	0.035	13.294
std	0.516	120.263	146.804	140.686	0.058	0.087	27.771
min	1.000	1.000	10.000	3.850	0.005	0.001	0.555
25%	1.000	25.000	202.000	40.900	0.028	0.011	4.138
50%	1.000	80.000	321.000	78.000	0.049	0.020	7.718
75%	1.000	188.000	438.000	135.910	0.077	0.034	13.116
max	15.000	633.000	700.000	1999.990	0.800	3.045	646.779

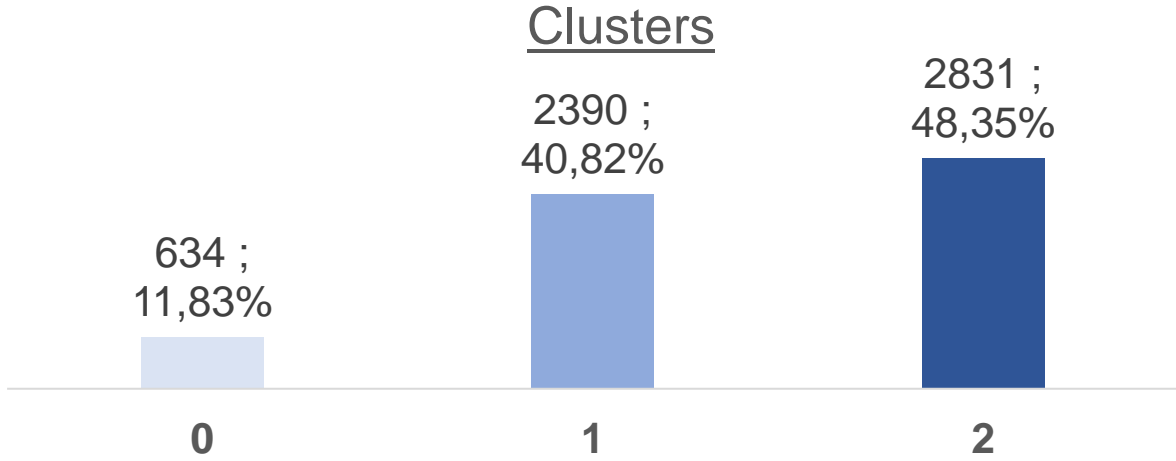
- ☐ O melhor cliente comprou 15 vezes durante o tempo de vida de 700 dias e sua compra mais recente foi há 67 dias ou no 633° (sexcentésimo trigésimo terceiro) dia.
- ☐ Probabilidade de manter o relacionamento com a empresa é de 80%, dentro dos próximos 6 meses a previsão é que ele compre 3 vezes e retorne um valor de \$646,78

# Resultados:



Considerado linha de corte para classificação: prob\_alive <10% (churned) ; >10% e <20% (high risk churned) ; > 20% (not churned)

# Resultados:



	frequency	recency	T	monetary_value	purchase_next_6_month	LTV_next_6_month
Cluster_KmPCA						
0	2.611987	293.235016	382.498423	106.431263	0.274540	83.810413
1	1.086192	164.719247	444.444351	104.379096	0.017087	8.979826
2	1.060049	52.358884	203.394560	110.239000	0.029533	9.204991

Cluster_KmPCA	churn	
0	high risk	260
	not churned	374
1	churned	2390
2	churned	2814
	high risk	17

**Grupo 0** – representam **11% do volume** dos pedidos dos clientes. Apresentam **alta frequência e recência de compras**, apresentam um **valor monetário médio**, são clientes de non-“churn” (não desistentes) e alto risco de “churn”, possuem um volume de compras médio para os próximos seis meses e um altíssimo valor de LTV.

**Grupo 1** – representam **41% do volume** dos pedidos dos clientes. Apresentam **média frequência e recência de compras**, apresentam um **valor monetário baixo**, são clientes desistentes (“churn”), possuem um valor de compras baixo para os próximos seis meses e um baixo valor de LTV

**Grupo 2** – representam **48% do volume** dos pedidos dos clientes. Apresentam **baixa frequência e recência de compras**, apresentam um **valor monetário alto**, são clientes de “churn” e (desistentes) e alto risco de “churn”, possuem um valor de compras alto para os próximos seis meses e um médio valor de LTV.

# Conclusão:

Pode-se supor que os clientes que desistiram já estão perdidos, porém ainda existem os **clientes** com **alto risco de “churn”**, mas que ainda não desistiram, o que significa que há **oportunidades** de **melhoria** em relação à **retenção**.

Com foco nesses clientes (presentes nos clusters 0 e 2), é possível **realizar ações de Marketing** como: comunicações personalizadas, promoções, programas de fidelização para tratamentos direcionados de modo a conquistá-los e evitar o “churn”.



# Agradecimentos:



À professora  
**Daniele** Aparecida  
Cicillini Pimenta,  
pela orientação e  
oportunidade de  
crescimento  
acadêmico.



Ao professor Luiz  
**Paulo Fávero**, pela  
coordenação do  
MBA USP/Esalq  
em Data Science &  
Analytics



À **todos os professores** que  
ministraram as aulas do curso e  
compartilharam seus  
conhecimentos

À **todos os profissionais** que  
colaboraram com a excelente  
organização e suporte do curso  
MBA USP/Esalq em Data  
Science & Analytics



Ao **Paulo Cotta**,  
**colega de trabalho**  
que acreditou na  
minha transformação e  
me forneceu o suporte  
para os novos desafios  
no desenvolvimento da  
carreira na área de  
Ciência de Dados.



À minha  
**companheira Lis**  
**de Oliveira** pelo  
incentivo, suporte,  
lições de vida,  
paciência e  
carinho.