

Modelo preditivo de engajamento de postagens no Instagram para o mercado de papelarias

Parte 3



11/01/2022



Pós-graduação Análise de Dados, Data Mining e Inteligência Artificial

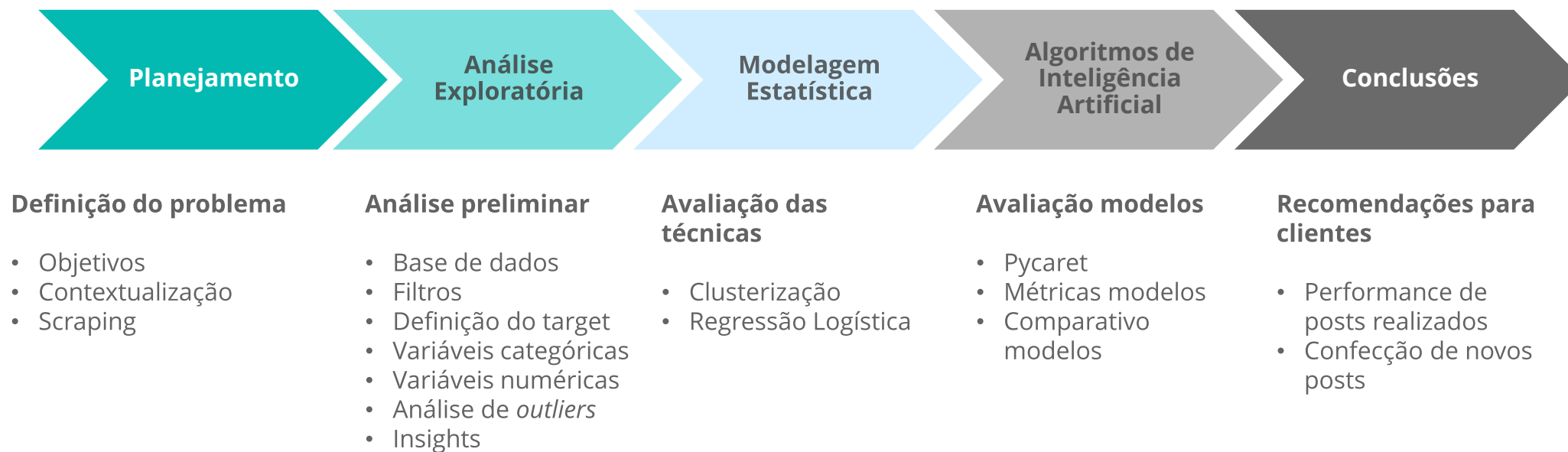


Nome do Aluno:
Paula Gonçalves Meneses

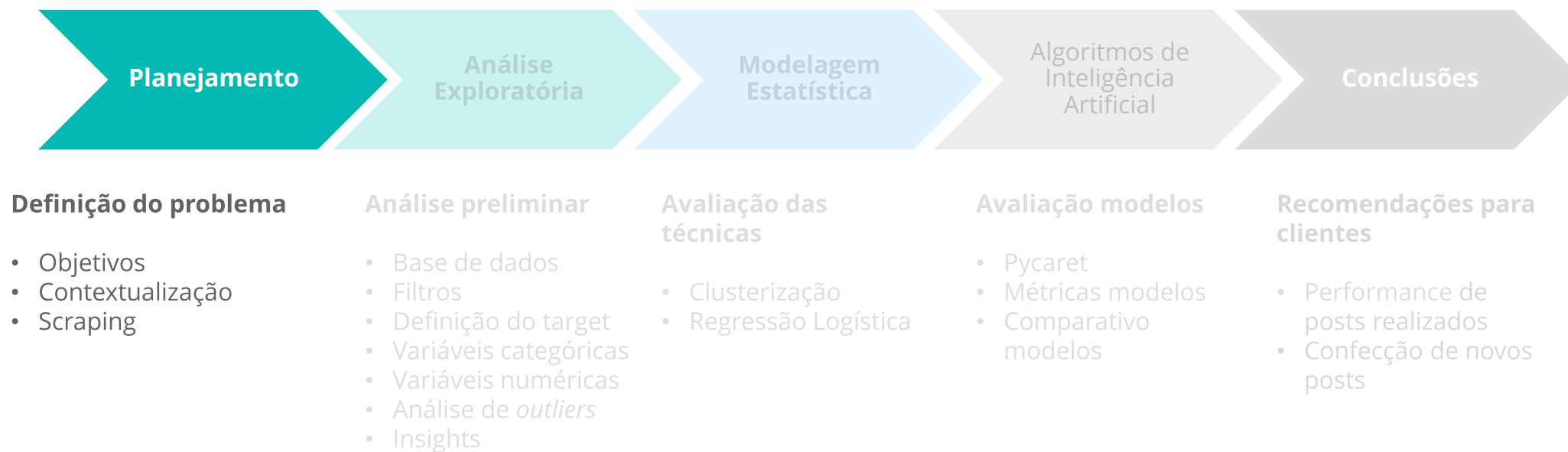
Coordenadores:
Profª Drª Alessandra de Ávila Montini
Profª Dr. Adolpho Walter Pimazoni Canton



Agenda



Agenda



1.1. Objetivo do Trabalho

O presente trabalho tem como objetivo **predizer o engajamento de uma postagem no Instagram**, realizada por um perfil de papelaria.

O projeto apresenta **duas principais fases**: a primeira constitui na **elaboração de uma base de dados** bem estruturada, através de **técnicas de scraping**. Na segunda fase será abordada a **modelagem** que, utilizando dos dados coletados das postagens de diferentes perfis de papelarias brasileiras no Instagram, identificará quais são as **características mais relevantes da postagem para explicar seu engajamento com o público**.

Por fim, espera-se que a implementação deste modelo auxilie na **estratégia de elaboração e publicação de postagens**, tornando-as cada vez mais assertivas para **conversão em vendas**.





1.2. Contextualização do Problema

No último ano foi possível observar um **aumento significativo** na presença de empresas do setor de papelaria no Instagram, com intuito de realizar **vendas online**. Este movimento foi marcado não só por novas empresas que já nasceram na modalidade online, mas também por lojas físicas que passaram a **investir na digitalização**.

Mas como se **destacar** em um contexto de tanta concorrência? Como **otimizar o tempo** na elaboração de postagens, ganhar relevância pelo público e **aumentar a conversão em vendas**?

Foi neste contexto que o presente trabalho foi pensado. É conhecido pelo mercado a importância de valorizar, por exemplo, o **dia da semana e horário** das postagens e o próprio Instagram oferece uma ferramenta que apresenta esses dados. Mas quais são as **outras features** que podem contribuir de forma relevante para o **engajamento** de uma postagem?

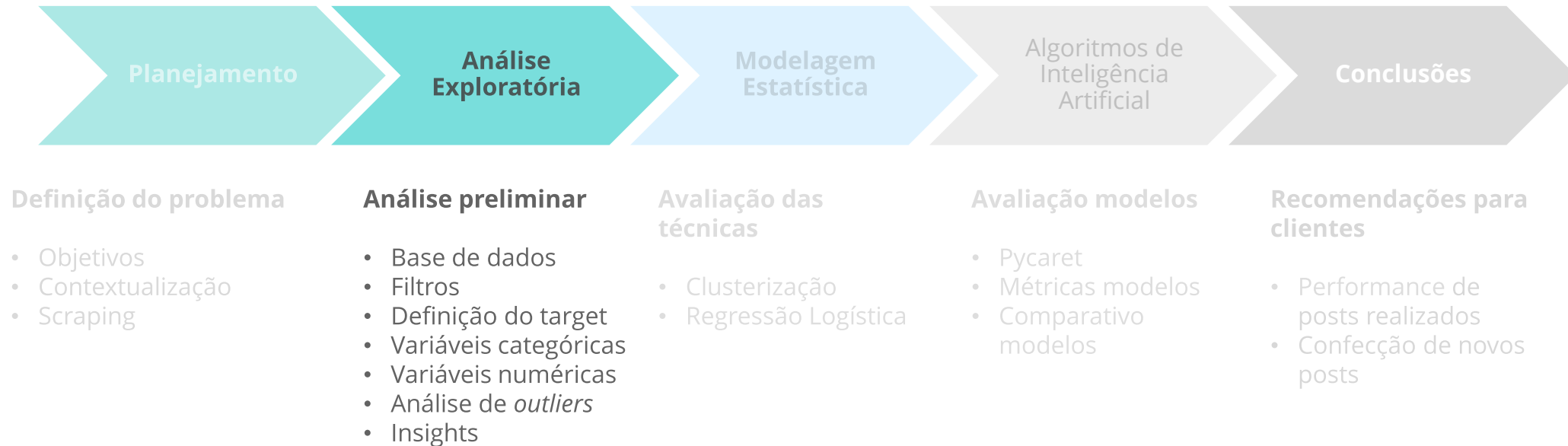
Por isso, o intuito do projeto é **potencializar** o mercado de papelarias online, através de análises mais robustas, e dando visibilidade para as características que mais influenciam no **aumento de seu engajamento**.

1.3. Bases de Dados | Scraping

Como não havia uma base de dados consolidada com as informações necessárias para o projeto, essa primeira fase foi um pouco mais extensa. Algumas estratégias foram adotadas para realização do scraping:

- I) Criação de um código utilizando a biblioteca Selenium para raspar dados do Instagram através de requisições pelo xpath.
Esta estratégia foi interessante para validar a viabilidade de conseguir informações, porém funcionou apenas para pequena escala. Ao executar o script para a quantidade de perfis necessários, o Instagram começou a limitar o acesso e se tornou inviável seguir por este caminho, mesmo inserindo *sleeps* no código rodando apenas alguns perfis.
- II) Ao pesquisar melhor a experiência de outros trabalhos, surgiu-se a possibilidade de utilizar uma API não oficial do Instagram, chamada **igramscraper**. Foi realizada a adaptação do código, inserção de *sleeps* e, apesar de haver certa limitação, o cenário foi mais favorável para adquirir os dados, até mesmo pensando na quantidade e qualidade das features inclusas. A base original foi então estruturada a partir deste caminho.

Agenda



2.1. Filtros

9



Base Original

- Extraída pela API igramscraper.
- **99** diferentes perfis de papelarias analisadas.
- Média de **28 posts** por perfil.
- Total de **2.835** posts (linhas) e **26** features.

Visão Geral

- Cada linha da base é referente à um post do instagram e as features estão relacionadas às características dessa publicação relacionadas à número de curtidas, perfil da da papelaria, tipo de post, entre outras.
- Algumas features resultantes da API vieram faltantes ou incompletas e foram avaliadas para dar continuidade ao estudo.

Legenda

- ✗ Features eliminadas por informação ou dados inconsistentes
- ✓ Features de input para ABT

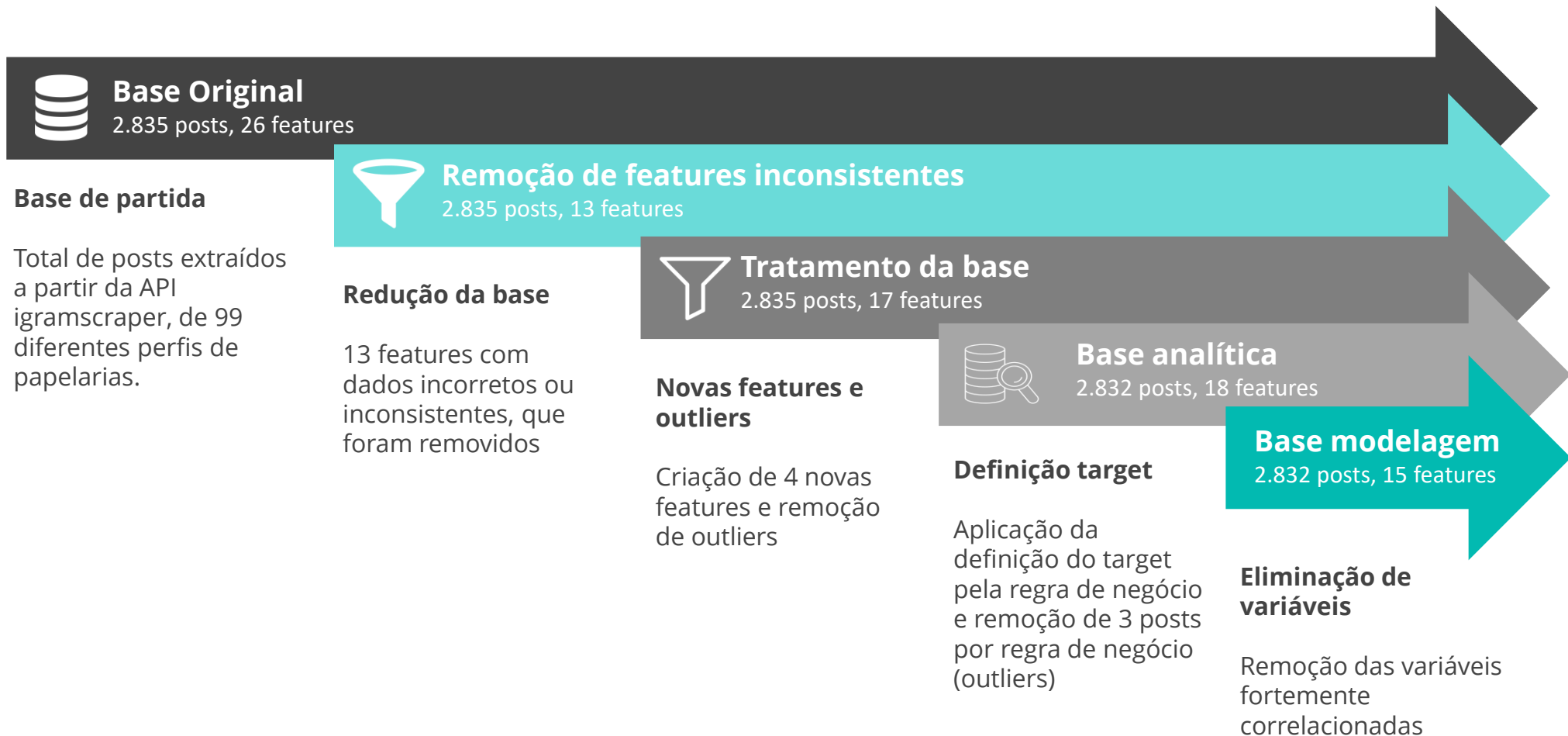
Lista de variáveis Base Original

- ✓ **account_name**: nome do perfil da papelaria.
- ✓ **full_name**: descrição do nome da papelaria.
- ✗ **biography**: descrição do perfil da pepelaria referente ao post.
- ✓ **is_business_account**: identificação se aquele perfil é business ou não.
- ✗ **is_joined_recently**: se o perfil é recente no instagram ou não.
- ✗ **is_verified**: se a conta é verificada ou não.
- ✓ **business_category_name**: categoria que aquele perfil se posiciona.
- ✗ **business_email**: e-mail do perfil.
- ✗ **business_phone_number**: telefone para entrar em contato com o perfil.
- ✗ **connected_fb_page**: se a conta é conectada ao facebook.
- ✓ **followed_by_count**: número de seguidores do perfil.
- ✓ **follows_count**: número de perfis que a papelaria segue.
- ✗ **has_channel**: se a papelaria tem algum canal.
- ✓ **media_count**: quantidade de publicações do perfil.
- ✓ **caption**: descrição do post.
- ✓ **created_time**: dia e horário que o post foi publicado.
- ✓ **comments_count**: quantidade de comentários do post.
- ✗ **comments_next_page**: não ficou claro o que seria esta feature
- ✗ **has_more_comments**: não ficou claro o que seria esta feature
- ✗ **image_high_resolution_url**: url da imagem do post.
- ✗ **is_ad**: se o post é patrocinado.
- ✓ **likes_count**: quantidade de likes do post
- ✓ **link**: link do post
- ✗ **location_name**: localização do post.
- ✗ **video_views**: total de visualização dos vídeos.
- ✓ **type**: tipo de post (vídeo/foto/carrossel).



2.1. Filtros

10



2.2. Definição do target



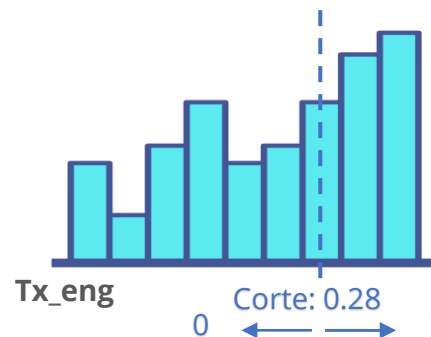
- Para classificar a relevância do post, primeiramente é necessário calcular a taxa de engajamento.
- Por definição da área de negócio, a taxa de engajamento é descrita pela fórmula:

$$\frac{qtd \text{ likes do post}}{total \text{ seguidores do perfil}}$$

- Foi realizada uma pesquisa para entender referências de engajamento de estudos realizados previamente para setores semelhantes ao de papelarias. A partir de um estudo recente, realizado em agosto de 2021, viu-se uma referência de **2,8% como uma taxa média de engajamento para perfis de “Shopping & Retail” no Instagram**. Esse foi o valor base adotado no presente estudo para definir os posts com engajamento ou não (política estabelecida).
- A partir do cálculo e análise do engajamento por post, foi possível identificar que **aproximadamente 25% da base apresenta a taxa de engajamento superior à 2,8%**.

count	2832.000000
mean	0.023179
std	0.034646
min	0.000000
25%	0.004341
50%	0.013083
75%	0.029601
max	0.868234
Name: eng_rate, dtype: float64	

- Por fim, foi criada uma nova feature “target” para definir, por post, aquele que apresentava engajamento ou não.



A definição de uma postagem engajada será classificada em 0 ou 1, baseada no corte, sendo:

- 0: postagens com baixo engajamento**
- 1: postagens com alto engajamento**

2.3. Base Analítica (ABT)

A base analítica apresenta **2 variáveis chave**, **15 variáveis explicativas** e a **variável resposta (target)**.



Variáveis chave

- link
- account_name



Variáveis categóricas

- is_business_account
- business_category_name
- type
- weekday
- hour
- time_range



Variáveis numéricas

- followed_by_count
- follows_count
- media_count
- comments_count
- likes_count
- eng_rate
- caption_count
- caption_tag
- caption_hash



Variável Resposta

Target:
1 = postagens com alto engajamento
0 = postagens com baixo engajamento



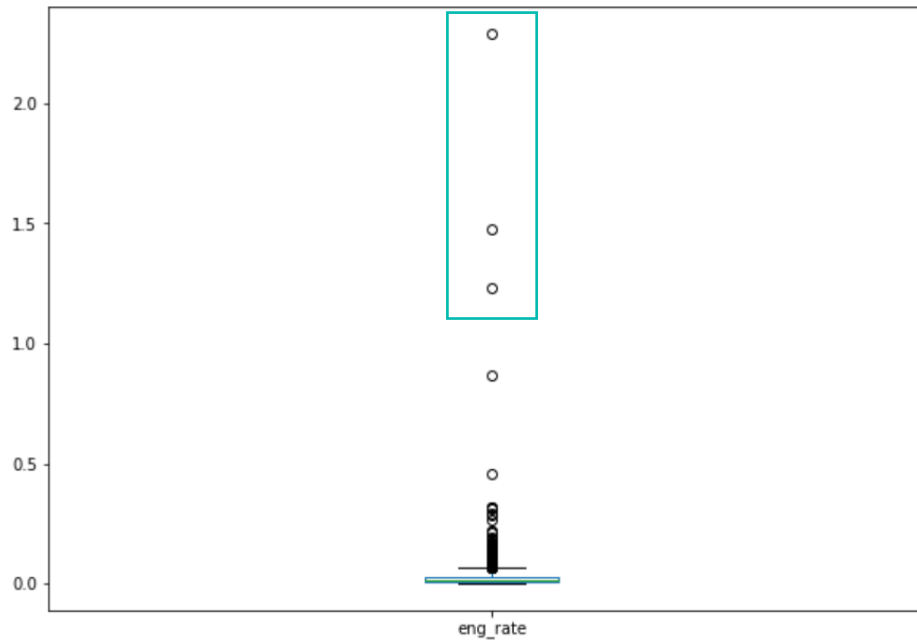


2.3. Base Analítica (ABT) | Descrição das variáveis

Variável	Descrição
link	Link do post (url)
account_name	Perfil da papelaria no instagram
is_business_account	Yes – conta business; No – não é conta business
business_category_name	Tipo de categoria definida pelo perfil
followed_by_count	Número de seguidores
follows_count	Número de perfis que a papelaria segue
media_count	Quantidade de posts publicados
comments_count	Quantidade de comentários daquele post
likes_count	Quantidade de likes do post
type	Tipo de post: image – foto; sidecar – carrossel (mais de uma foto); video – reel ou video
eng_rate	Taxa de engajamento
weekday	Dia da semana do post: mon, tue, wed, thu, fri, sat e sun.
hour	Hora do dia do post
time_range	Momento do dia do post: manha, tarde, noite e madrugada.
caption_count	Quantidade de caracteres da descrição do post.
caption_tag	Quantidade de marcações (@) na descrição do post.
caption_hash	Quantidade de hashtags na descrição do post.
target	0 – post com baixo engajamento; 1- post com alto engajamento

2.4. Base Analítica (ABT) | Outliers

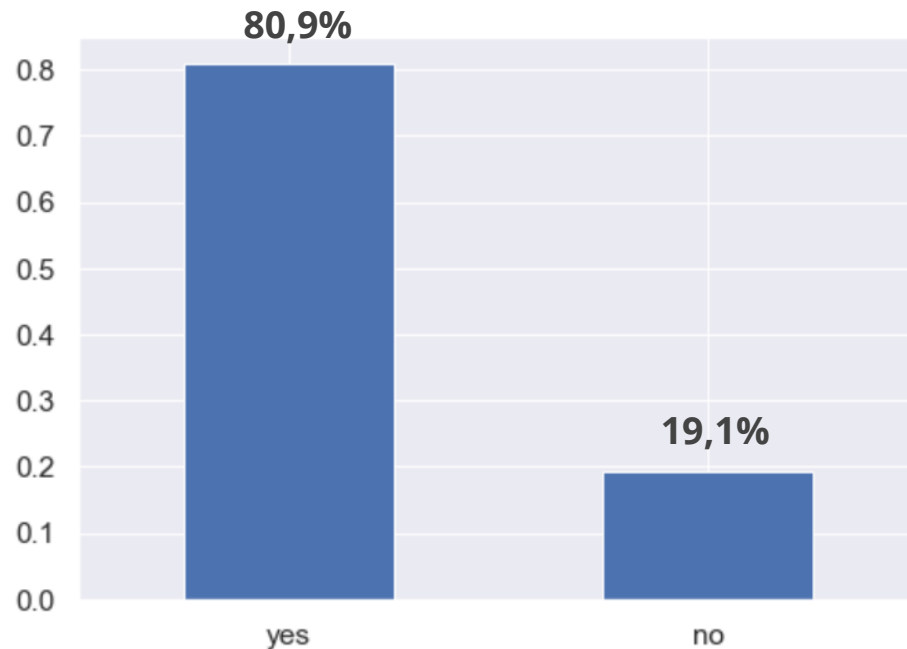
- Engajamento **superior a 1** representa que o post teve mais curtidas do que o número total de seguidores daquele perfil.
- Taxas superiores a 1 podem indicar algum **comportamento ilegal** como compra de likes, etc, portanto, por definição de regra de negócio, tomou-se a iniciativa de **remover** posts com taxa de engajamento superior a 1.
- A quantidade de posts removidos, como é possível verificar no boxplot abaixo, foi pequena. Apenas **3 posts** foram removidos da base.



Boxplot da taxa de engajamento da base

2.5. Análise Exploratória | Variáveis Categóricas

Variável: is_business_account



Posts dos perfis que são conta business apresentam engajamento superior (30%) à média de 26,5%.

	target	0	1	All	eng
is_business_account					
no		486	56	542	0.103321
yes		1594	696	2290	0.303930
All		2080	752	2832	0.265537

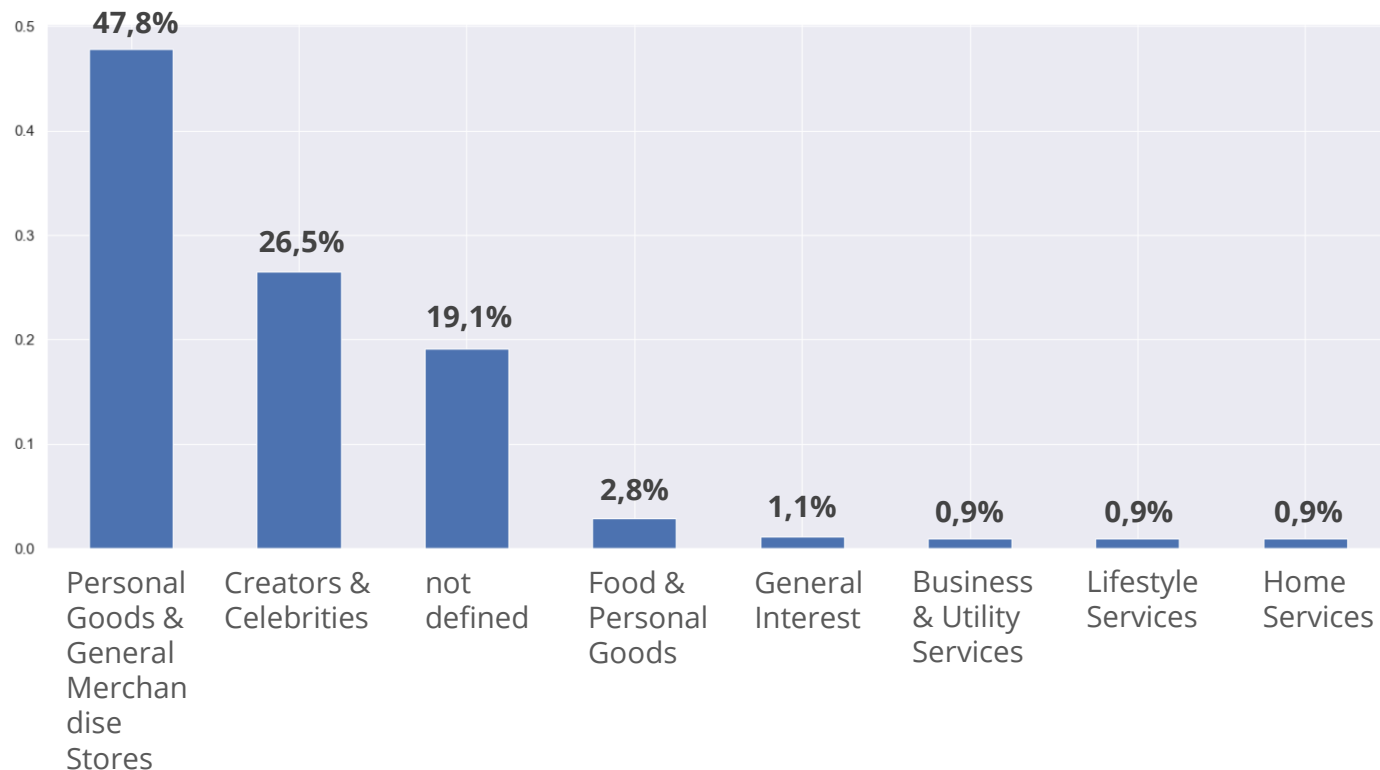
- Variável pode ser relevante para o modelo.



2.5. Análise Exploratória | Variáveis Categóricas

16

Variável: `business_category_name`



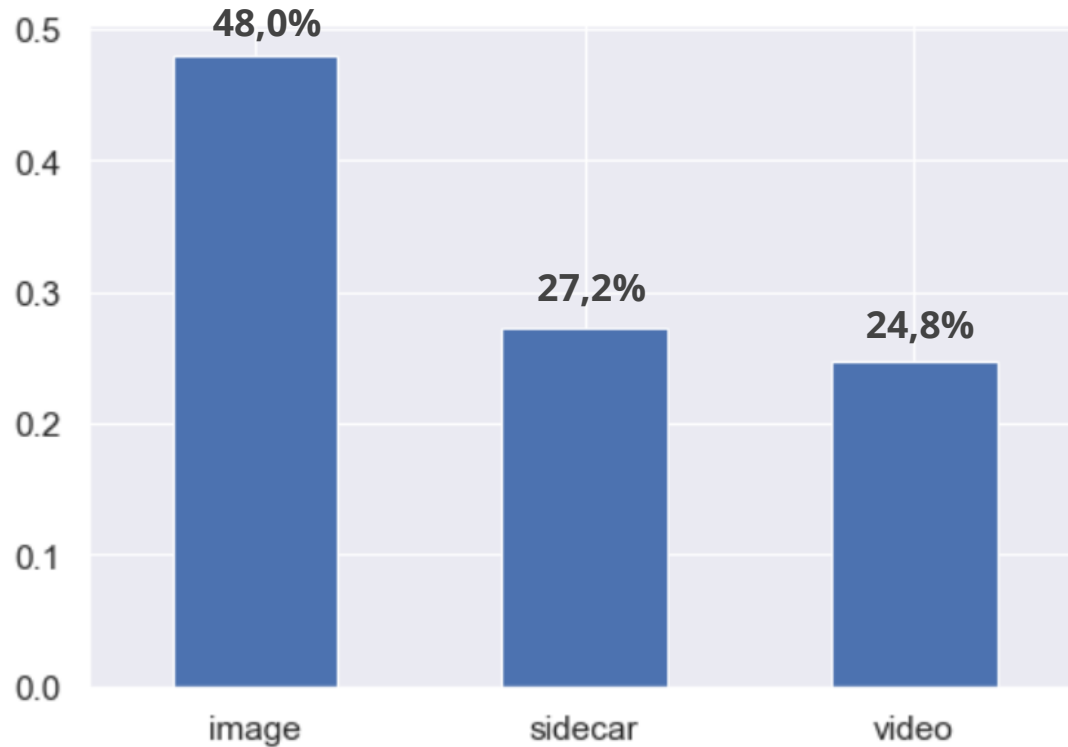
Algumas categorias parecem se destacar mais em termos de engajamento, como por exemplo “Business & Utility Services” e “Creators & Celebrities”

	target	0	1	All	eng
business_category_name					
Business & Utility Services	15	10	25	0.400000	
Creators & Celebrities	476	275	751	0.366178	
Food & Personal Goods	75	5	80	0.062500	
General Interest	28	2	30	0.066667	
Home Services	25	0	25	0.000000	
Lifestyle Services	25	0	25	0.000000	
Personal Goods & General Merchandise Stores	950	404	1354	0.298375	
not defined	486	56	542	0.103321	
All	2080	752	2832	0.265537	

Apesar da taxa de engajamento para “Business & Utility Services” ser mais alta, esses posts são do mesmo perfil de papelaria. As categorias “Creators e Celebrities” e “Personal Goods & General Merchandise Stores” chamam atenção pelo engajamento acima da média.

2.5. Análise Exploratória | Variáveis Categóricas

Variável: type



Posts do tipo carrossel (**sidecar**), aqueles que apresentam mais de uma foto, apresentam **engajamento superior**, comparado aos outros dois.

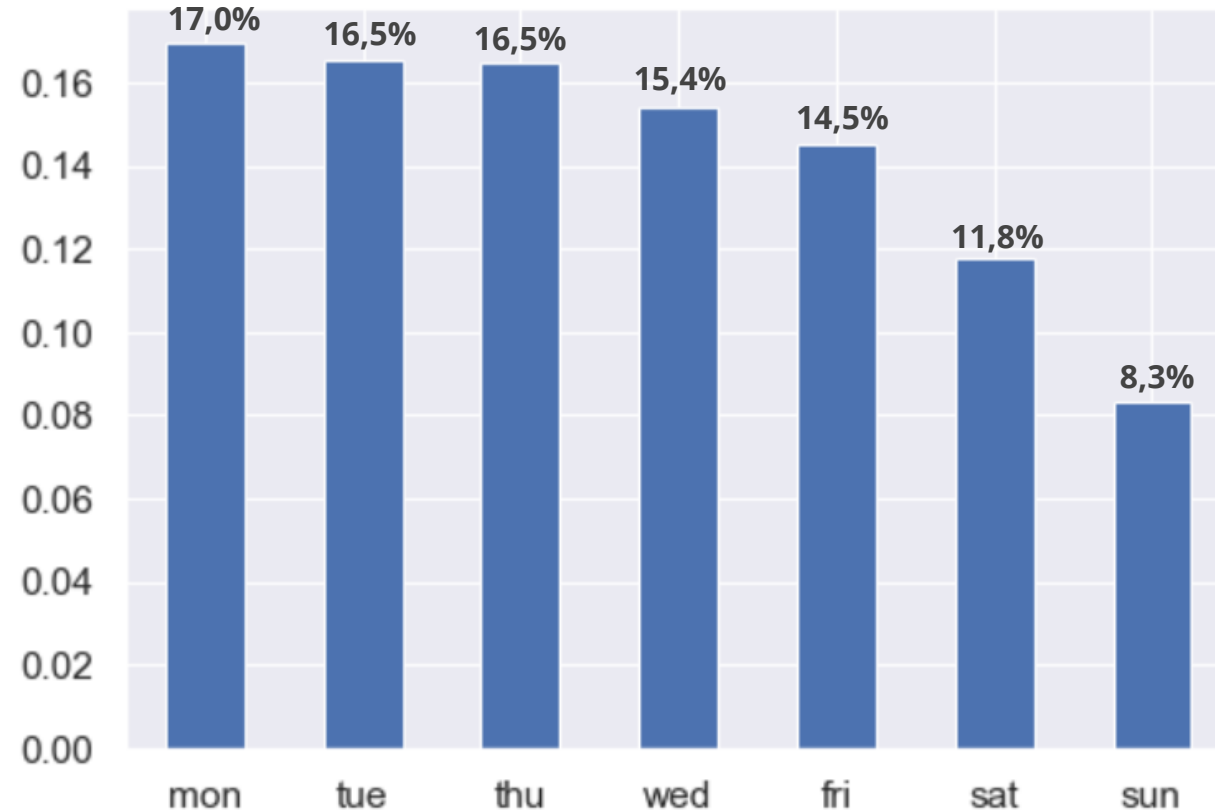
target	0	1	All	eng
type				
image	1048	311	1359	0.228845
sidecar	512	259	771	0.335927
video	520	182	702	0.259259
All	2080	752	2832	0.265537

- Variável pode ser relevante para o modelo.



2.5. Análise Exploratória | Variáveis Categóricas

Variável: weekday



Apesar de ter maior volume de posts, a **segunda** não é o dia com maior taxa de engajamento. **Quinta** se destaca no engajamento e **domingo** é um dia que chama atenção, aparecendo em terceiro lugar, e com menor volume de postagens.

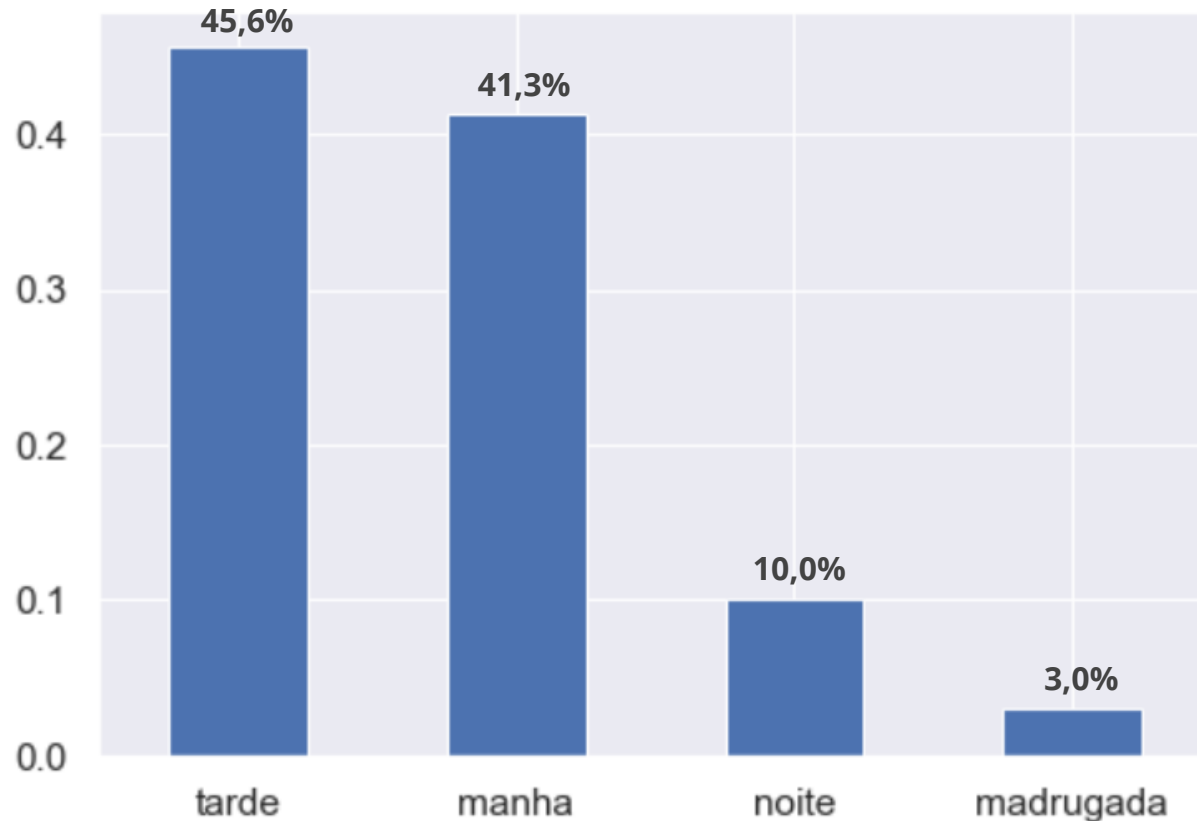
target	0	1	All	eng
weekday				
thu	327	139	466	0.298283
mon	342	138	480	0.287500
All	2080	752	2832	0.265537
sun	174	62	236	0.262712
fri	306	105	411	0.255474
wed	327	110	437	0.251716
tue	352	116	468	0.247863
sat	252	82	334	0.245509

- Variável parece ser relevante para o modelo.



2.5. Análise Exploratória | Variáveis Categóricas

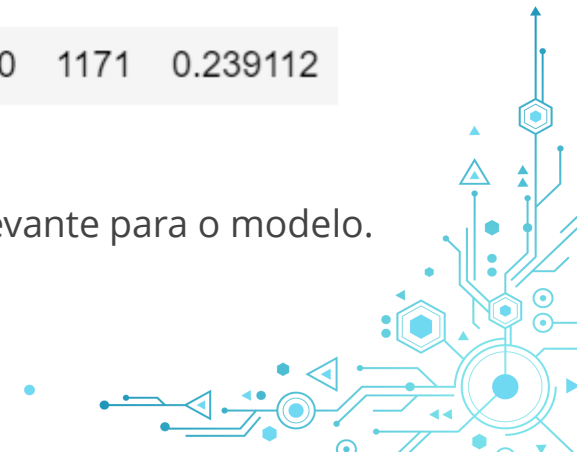
Variável: time_range



Posts no período da tarde (12h às 18h) se destacam no engajamento. Período da manhã foi o que apresentou menor taxa.

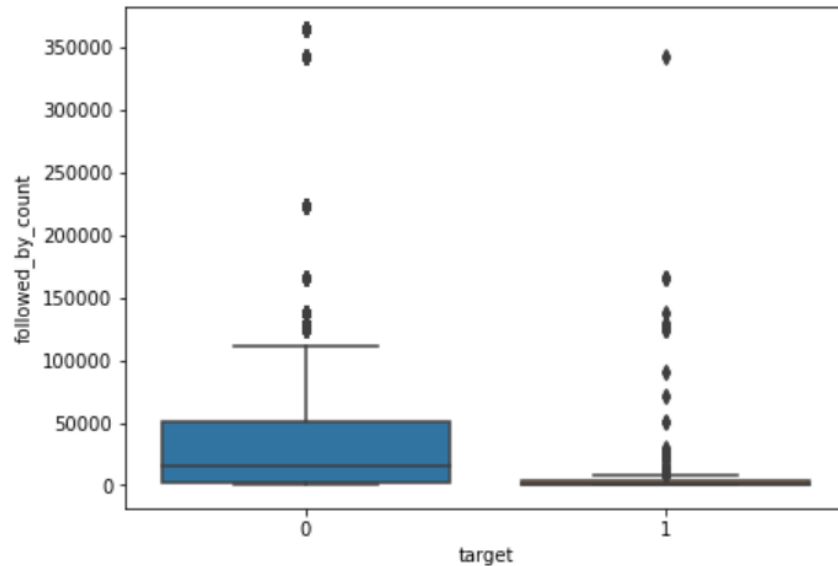
target	0	1	All	eng
time_range				
tarde	920	372	1292	0.287926
noite	206	78	284	0.274648
All	2080	752	2832	0.265537
madrugada	63	22	85	0.258824
manha	891	280	1171	0.239112

- Variável parece ser relevante para o modelo.



2.6. Análise Exploratória | Variáveis Numéricas

Variável: followed_by_count



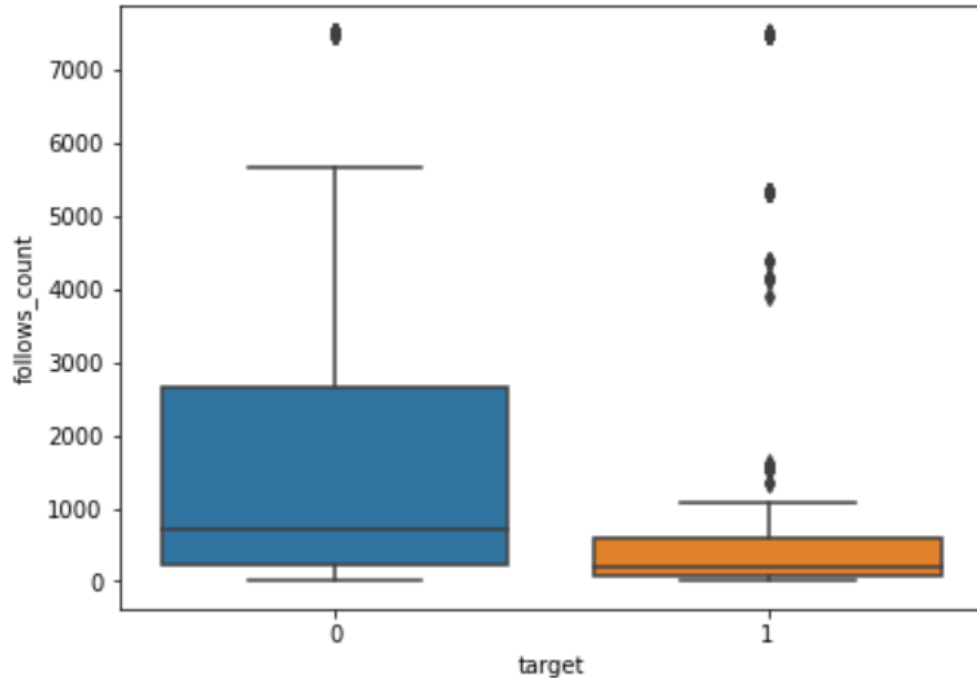
- Não necessariamente aqueles perfis com mais seguidores são os que apresentam maior quantidade de posts com engajamento.
- O desvio está bem alto, mostrando que isoladamente, essa variável não contribui tanto para explicar o engajamento.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	40803.787500	2303.25	15125	50916	78	364602	70128.392425
1	752	8153.865691	362.00	684	3444	78	342715	29532.048231



2.6. Análise Exploratória | Variáveis Numéricas

Variável: follows_count



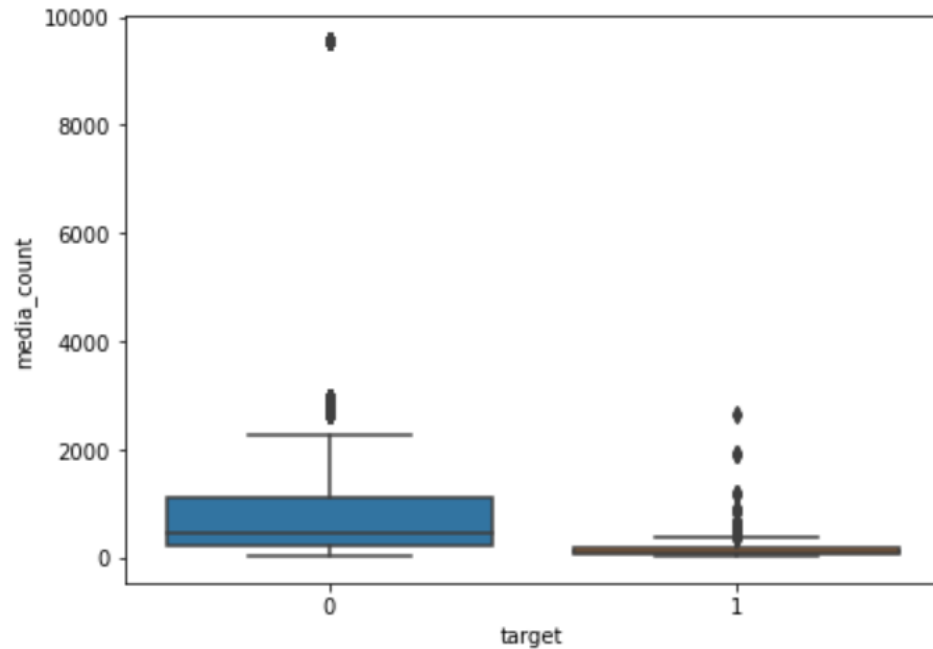
- O volume de perfis que a papelaria segue não parece influenciar no engajamento.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	1569.773558	235	722	2664	1	7500	1784.296810
1	752	730.811170	81	193	576	1	7500	1581.509898



2.6. Análise Exploratória | Variáveis Numéricas

Variável: media_count



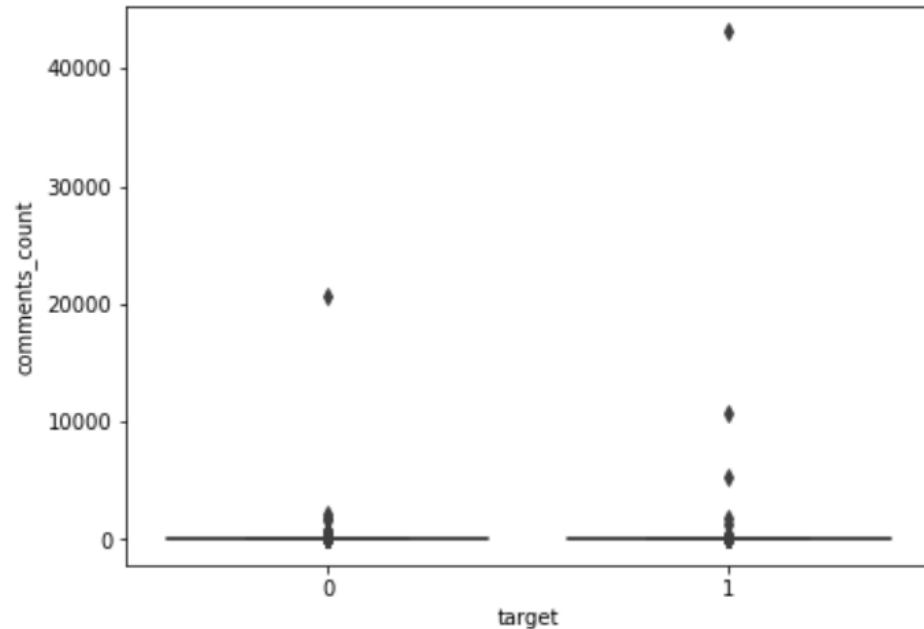
- O volume de postagens que a papelaria faz não parece influenciar de forma significativa no engajamento.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	888.301923	204	435	1094	19	9547	1409.279750
1	752	176.642287	59	99	188	12	2662	266.717061



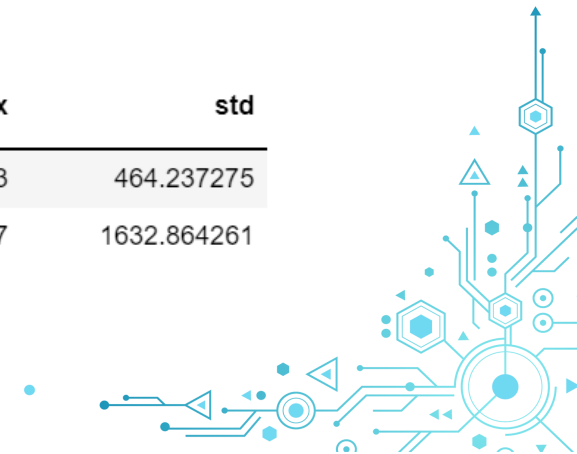
2.6. Análise Exploratória | Variáveis Numéricas

Variável: comments_count



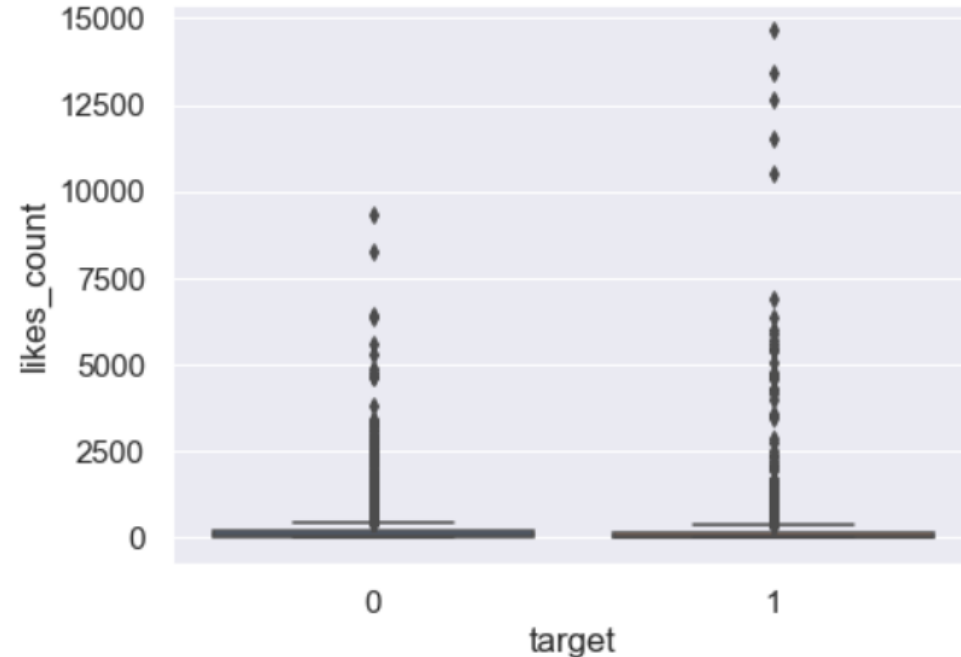
- 75% dos posts tem até 10 comentários.
- O desvio acaba sendo alto, possivelmente devido aos posts do tipo “sorteio”, que aumentam de forma significativa este volume.

target	size	mean	p25	median	p75	p90	amin	amax	std
0	2080	40.724038	0	2	9	57.2	0	20593	464.237275
1	752	93.156915	0	2	8	22.0	0	43127	1632.864261



2.6. Análise Exploratória | Variáveis Numéricas

Variável: likes_count



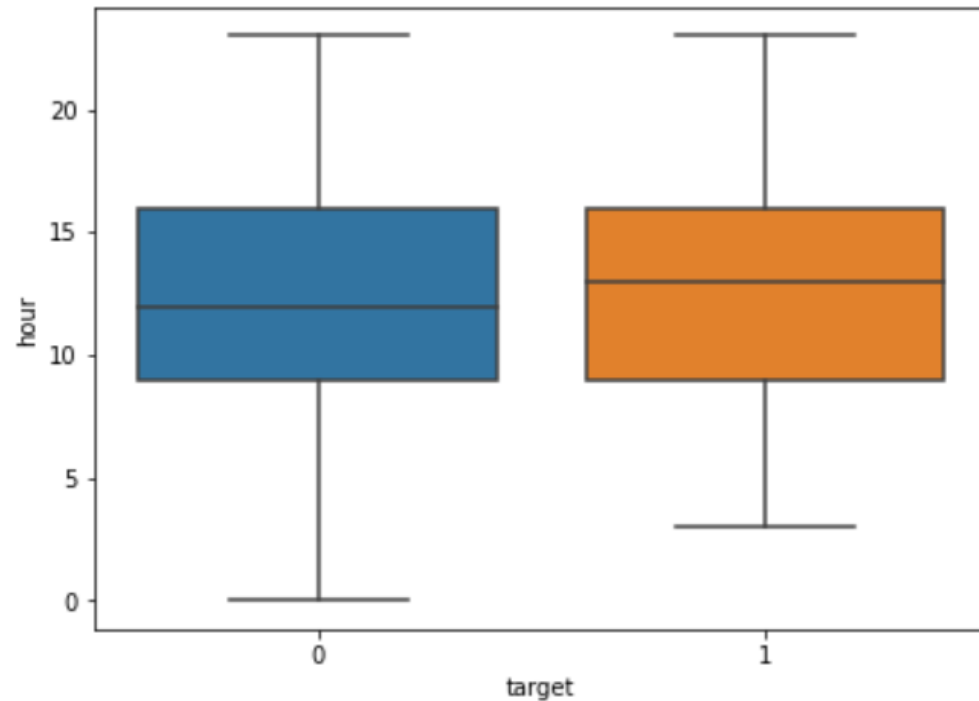
- Isoladamente, a quantidade de likes não implica necessariamente em engajamento.
- Por mais que a média dos posts engajados seja maior nos posts engajados, a mediana e o P75 são menores.
- Variável foi combinada à quantidade de seguidores do perfil para calcular o target.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	237.485577	19	58	191.00	0	9339	603.240084
1	752	395.660904	17	33	152.25	3	14648	1344.189370



2.6. Análise Exploratória | Variáveis Numéricas

Variável: hour



- Em relação ao horário da postagem, o engajamento não parece ser tão sensível à hora específica, apesar da leve tendência dos posts publicados mais tarde terem uma média e mediana um pouco mais elevados.
- Interessante observar que **90%** dos posts são realizados **até 18h**.

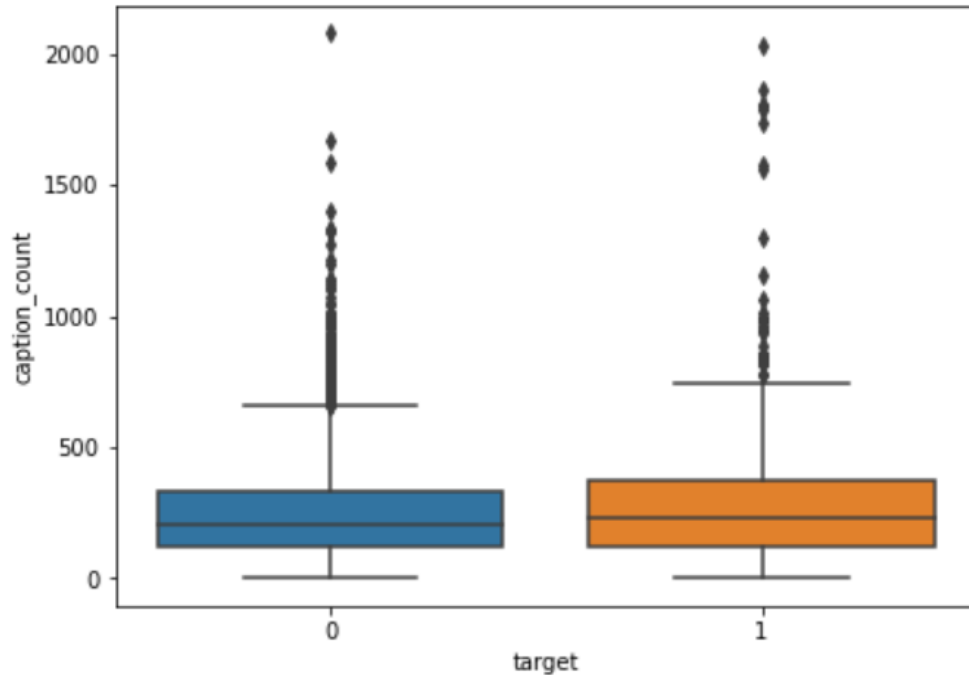
target	size	mean	p25	median	p75	p90	amin	amax	std
0	2080	12.214904	9	12	16	17	0	23	4.11177
1	752	12.501330	9	13	16	18	3	23	4.09052



2.6. Análise Exploratória | Variáveis Numéricas

26

Variável: caption_count



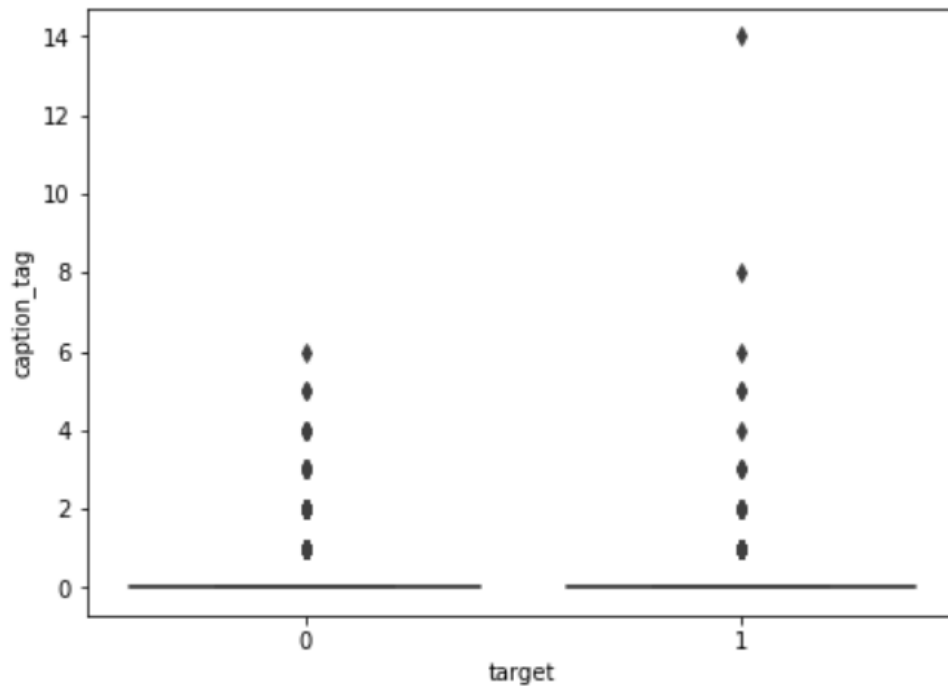
- Posts com maior número de caracteres apresentam leve tendência de impactar no maior engajamento do post.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	256.946154	119.00	208.5	335	0	2080	207.121052
1	752	284.513298	122.75	229.5	376	0	2030	247.108390



2.6. Análise Exploratória | Variáveis Numéricas

Variável: caption_tag



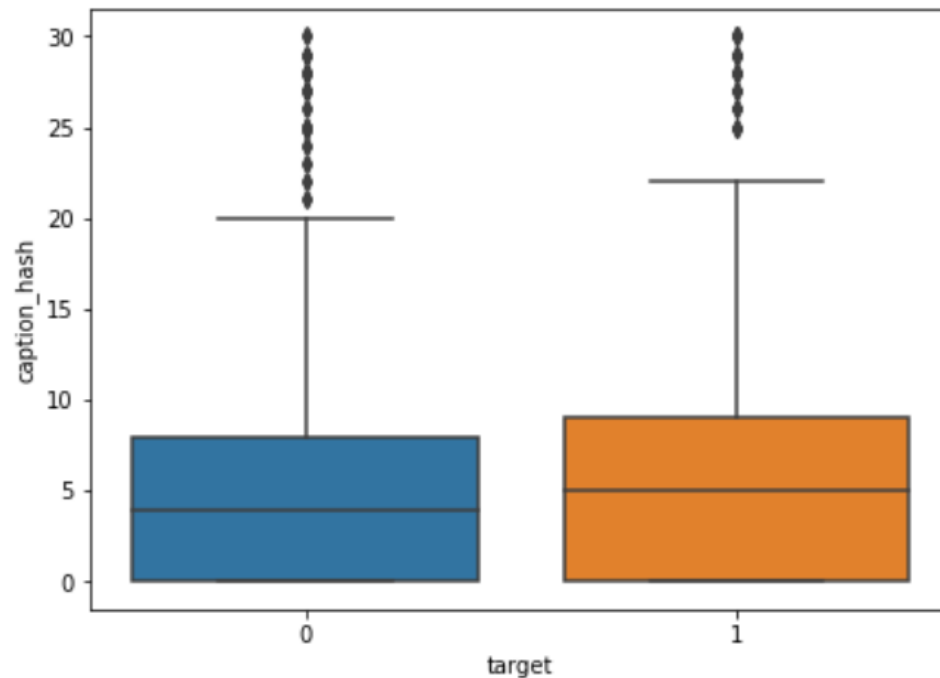
- De forma geral, os perfis fazem poucas marcações nas descrições dos posts.
- Apesar disso, os posts mais engajados apresentaram maior número de marcações médio.

targe	size	mean	p25	median	p75	amin	amax	std
0	2080	0.177885	0	0	0	0	6	0.525530
1	752	0.206117	0	0	0	0	14	0.811905



2.6. Análise Exploratória | Variáveis Numéricas

Variável: caption_hash



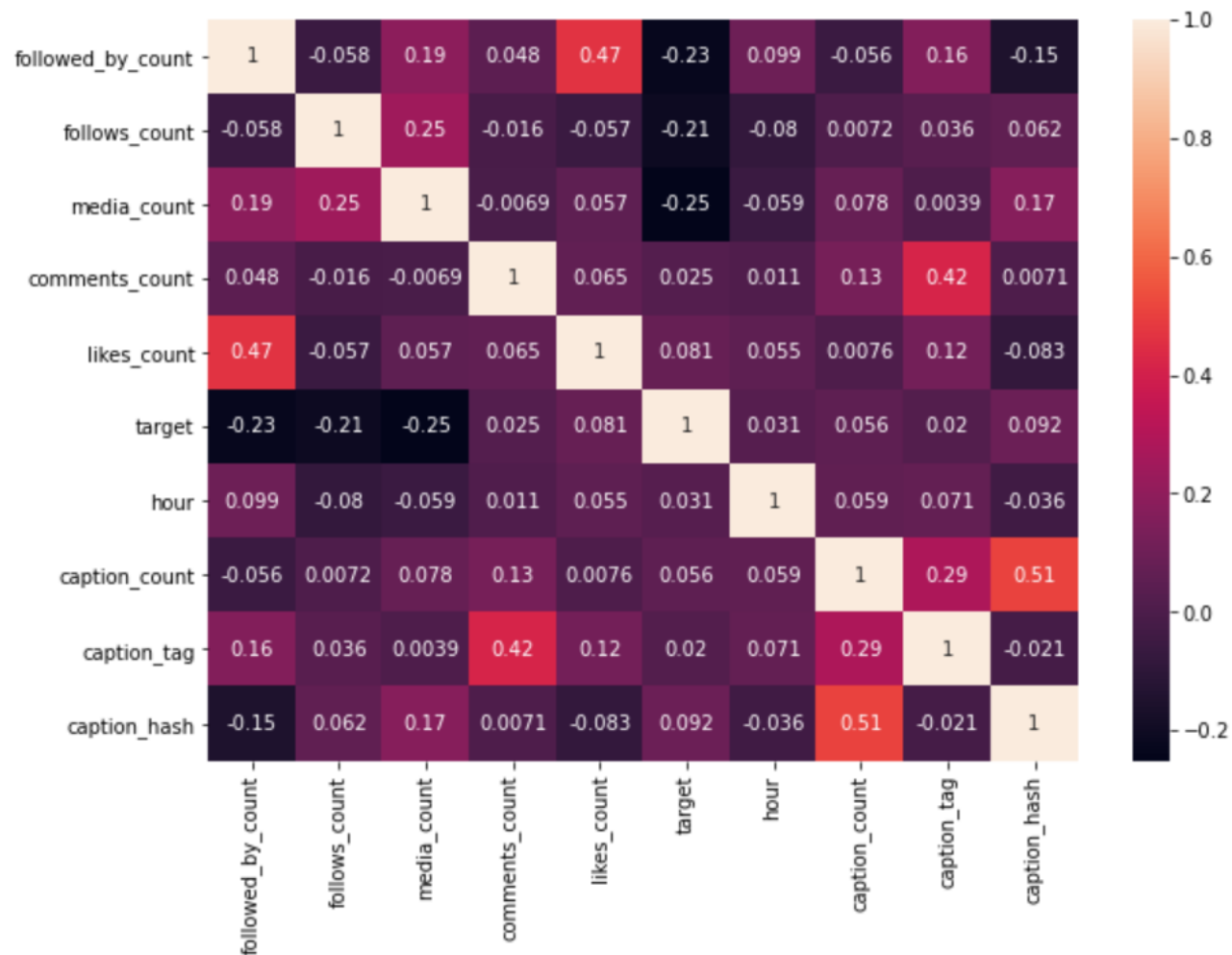
- A quantidade de hashtags **parece influenciar** no engajamento dos posts.
- Importante destacar que o Instagram limita o número máximo a 30 hashtags.

target	size	mean	p25	median	p75	amin	amax	std
0	2080	4.884615	0	4	8	0	30	5.423672
1	752	6.067819	0	5	9	0	30	6.294539



2.6. Análise Exploratória | Variáveis Numéricas

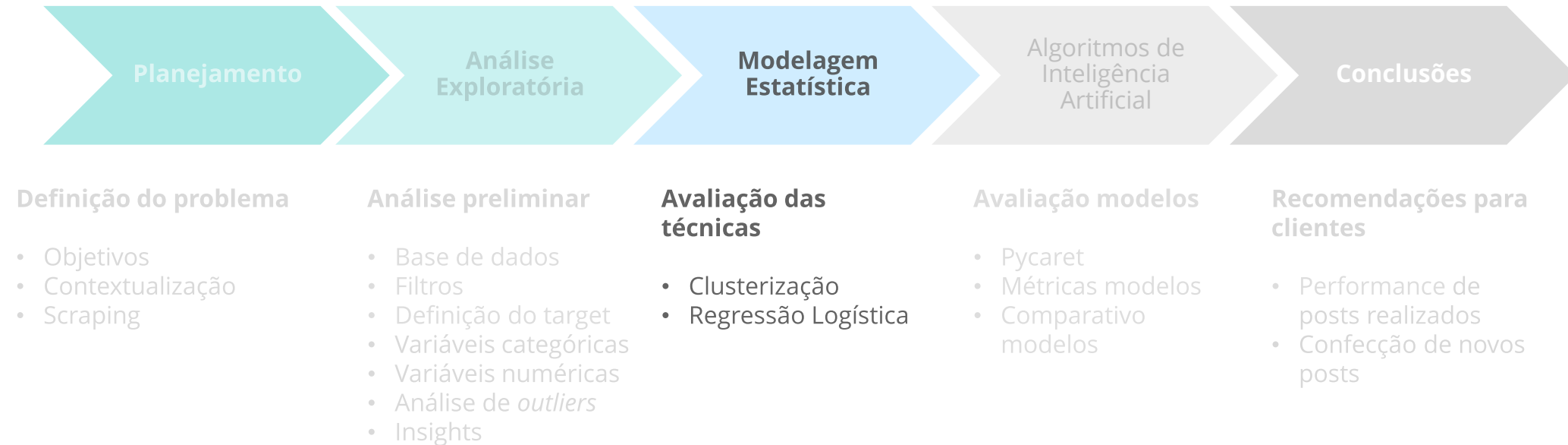
Análise de correlação entre variáveis numéricas



Pontos relevantes

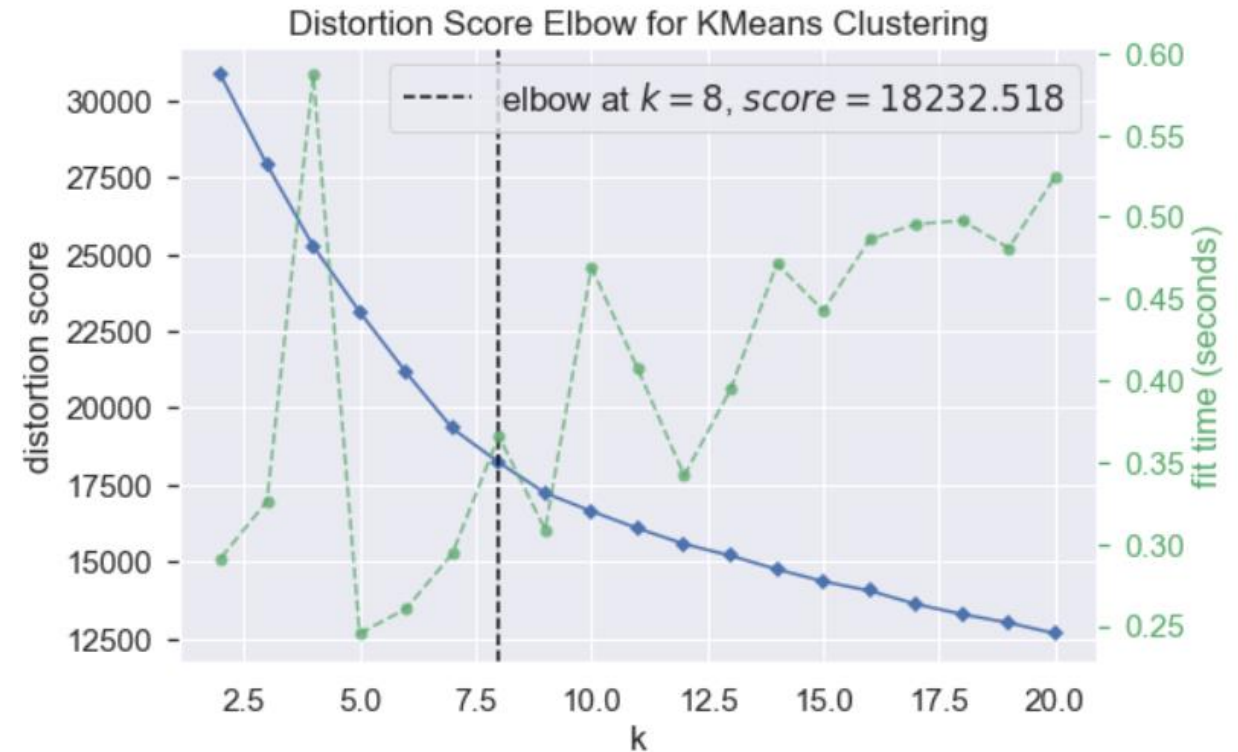
- Foi aplicada a técnica de StandardScaler nas variáveis antes de plotar a correlação, de forma a trazer resultados mais comparáveis.
- Isoladamente, variáveis como quantidade de seguidores do perfil, quantidade de perfis que a papelaria segue e total de posts não influenciam diretamente no engajamento. Por isso a importância do cálculo da taxa de engajamento para definir o target.
- Interessante a correlação entre quantidade de marcações e comentários. É possível inferir que, se a quantidade de comentários leva a um engajamento do post, então inserir marcações no post pode levar a mais comentários e, consequentemente, mais engajamento.
- Quantidade de hashtags é a variável que tem mais correlação com o target.

Agenda

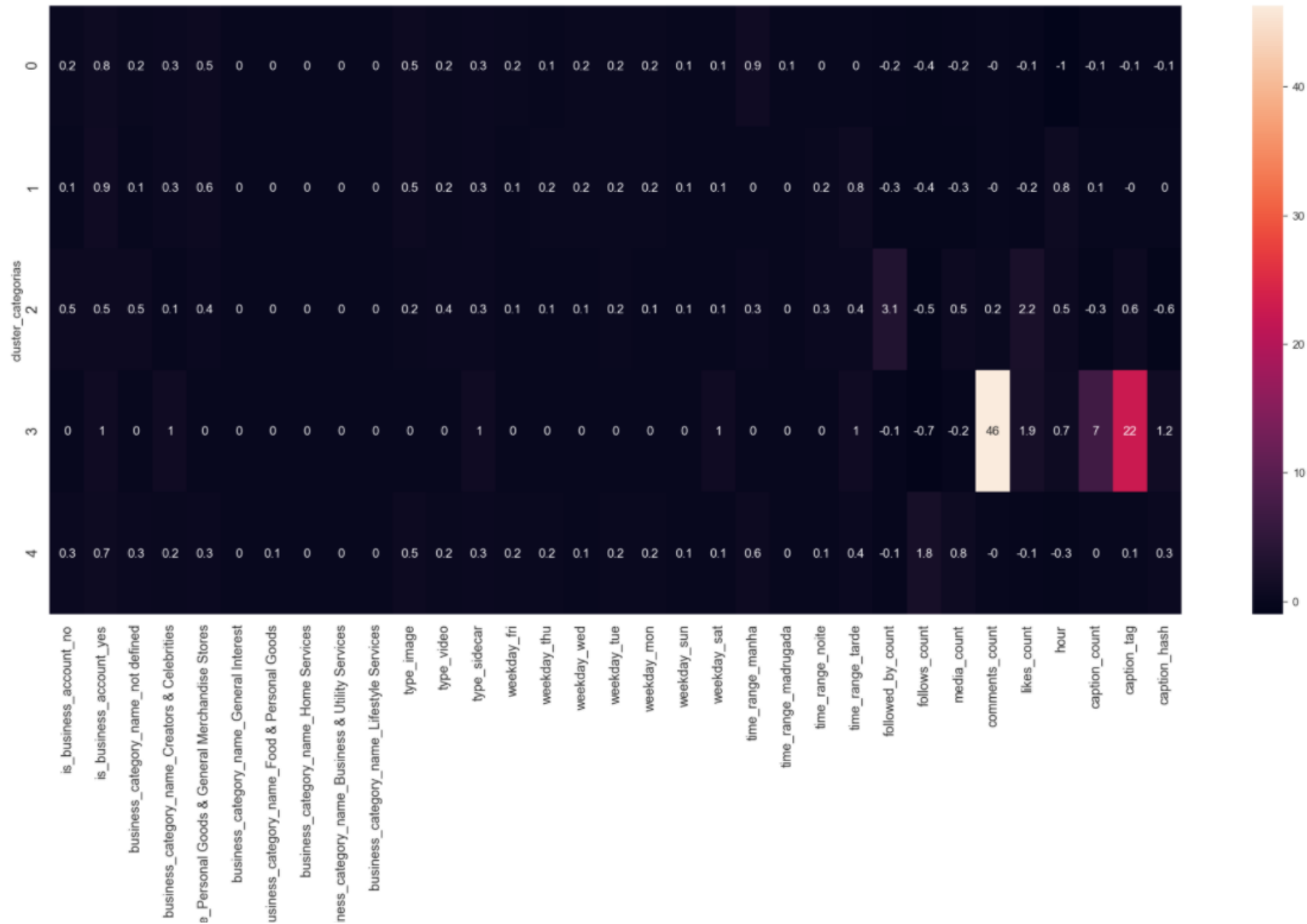


3.1. Modelagem estatística | Clusterização

- Na base de input para a clusterização foi aplicada a técnica de **one hot encoder** para variáveis categóricas e o **standard scaler** para as variáveis numéricas.
- Como observado no gráfico ao lado, o método aplicado sugere a quebra da base em **8 diferentes clusters**.
- Apesar da sugestão, para simplificar um pouco a análise e caracterização de grupos, tomou-se a decisão de agregar os posts em **5 diferentes clusters**.

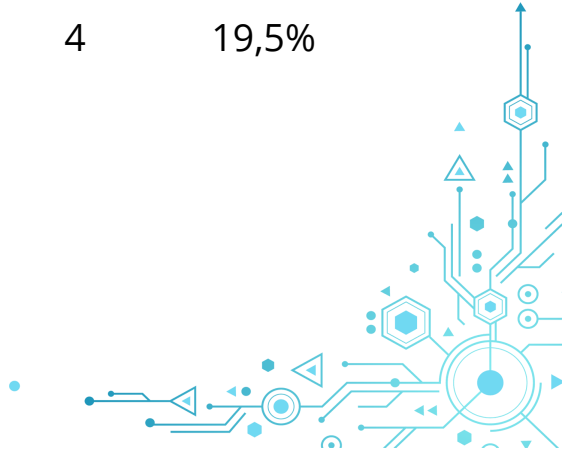


3.1. Modelagem estatística | Clusterização



- A matriz apresenta de forma geral as características mais presentes em cada um dos clusters.
- Em termos de volume de posts, os clusters estão distribuídos da seguinte forma:

Cluster	% posts
0	31,2%
1	43,0%
2	6,3%
3	0,0003%
4	19,5%



3.1. Modelagem estatística | Descrição dos Clusters

Cluster 0



883 posts, **84%** conta business (acima da média, 81%). Tipo de post segue a mesma tendência da base, apresentando mais **images (50%)**.



Perfis diversos em número de seguidores



93% de posts **manhã** e **7%** **madrugada**



30% de posts **com engajamento (target 1)**

Cluster 1



1.216 posts, **88%** conta business. Tipo de post segue a mesma tendência da base, apresentando mais **images (48%)** e um pouco a mais de post do tipo **sidecar (+0,4%)**.



Perfis diversos em número de seguidores



83% de posts **tarde** e **17%** **noite**



33% de posts **com engajamento (target 1)**

Cluster 2



177 posts, **45%** conta business. **45%** posts tipo **vídeo**, **32%** carrossel e **23%** foto. Perfil diferencia da média da base.



Perfis com **maior** número de seguidores



42% de posts **tarde** e **30%** **noite**



15% de posts **com engajamento (target 1)**

Cluster 3



1 post, do perfil carrossel, postado no sábado, à tarde.



Post foi analisado no detalhe e viu-se que era um caso de **sorteio**. O número de comentários (**65% maior que o número de seguidores**) e demais métricas destoaram bastante do padrão.

Cluster 4



555 posts, **71%** conta business. **53%** posts tipo **image**, **26%** carrossel e **21%** vídeo. Perfil diferencia da média da base.



Perfis com **maior** número de seguidores



56% de posts **manhã** e **36%** **tarde**



10% de posts **com engajamento (target 1)**

- A análise dos cluster possibilita o levantamento de hipóteses a serem testadas de forma mais aprofundada nos modelos.
- O período de publicação do post (time_range) parece influenciar no engajamento, sendo os períodos da tarde e manhã os mais relevante.
- A conta ser business parece trazer algum tipo de relevância também ao engajamento.
- Campanha de sorteio surgiu como uma oportunidade de engajamento, como observado pelo cluster 3.
- Distribuição de tipos de posts da base - **images: 48,0%, sidecar: 27,2% e vídeo: 24,8%**
- Distribuição de momento do dia dos posts da base - **manhã: 41,3%, tarde: 45,6%, noite: 10,0% e madrugada: 3,0%.**

Cluster	% posts
0	31,2%
1	43,0%
2	6,3%
3	0,0003%
4	19,5%

3.2. Modelagem estatística | Regressão Logística

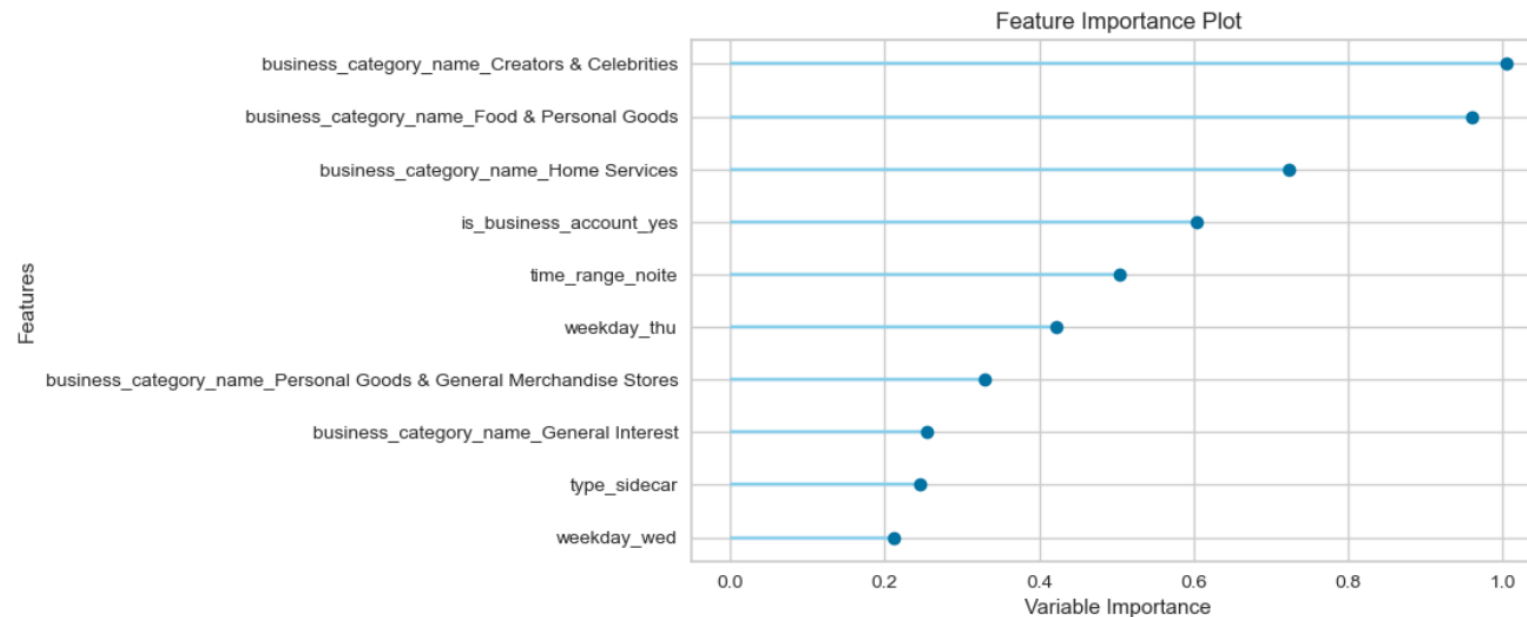
- A partir da mesma base de dados transformada para a clusterização, foi implementado um modelo de regressão logística com o objetivo inicial de avaliar qualidade e importância das variáveis presentes.
- O modelo teve um AUC interessante, de **81,3%**.
- É possível observar que a **categoria dos perfis** parece impactar no engajamento do post, assim como se a **conta é business** e o **período do dia** em que o post foi publicado.

Matriz de confusão

	Predicted class	
	0	1
True class	0	1
	543	77
1	129	80

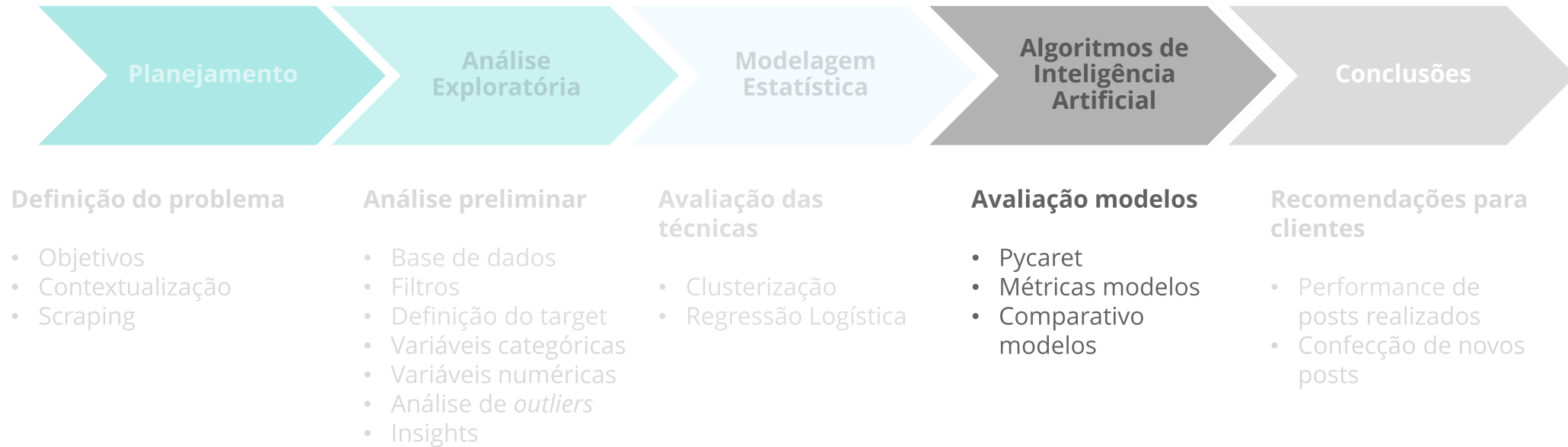
Avaliação do modelo

Métricas	Resultado
Acurácia	0.751
AUC	0.813
Recall	0.461
Precision	0.597
F1	0.483



Relevância das variáveis

Agenda



4.1. Técnicas de Inteligência Artificial | Resultado PyCaret

- Utilização da **biblioteca PyCaret** para avaliação de diferentes **modelos de classificação**, com o objetivo de identificar aquele com **melhor performance** para o problema em questão.
- Para o problema de **predição de engajamento de posts** foi identificado que a **melhor métrica para avaliação dos modelos seria o AUC**, uma vez que se trata de uma base desbalanceada (~25% posts com engajamento).
- Na tabela abaixo temos o output dos **3 melhores modelos, ordenados pelo AUC**, e suas principais métricas de avaliação.

Modelos	Acurácia	AUC	Recall	Precision	F1
Light Gradient Boosting Machine (LGBM)	0.8476	0.9042	0.7054	0.7241	0.7136
Gradient Boosting Classifier (GBC)	0.8315	0.9038	0.7630	0.6651	0.7103
Extreme Gradient Boosting (XGBoost)	0.8471	0.9033	0.6998	0.7250	0.7116

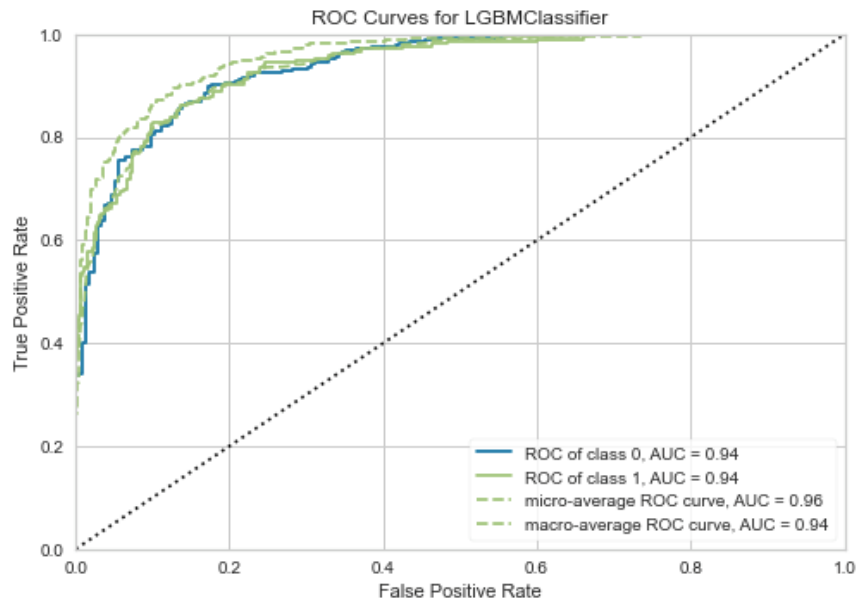


4.2. Técnicas de Inteligência Artificial | Outputs LGBM

37

AUC: 0.9042

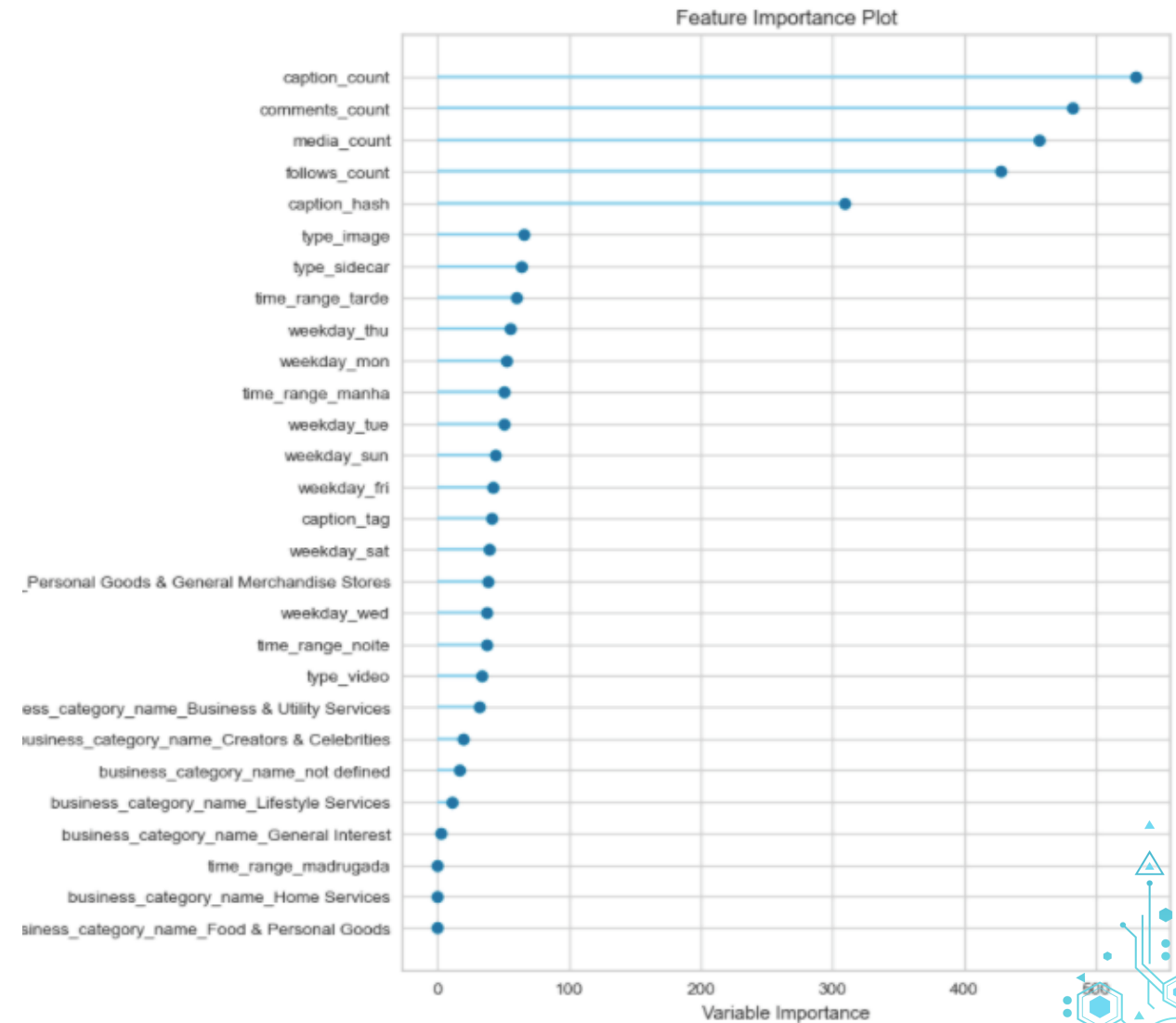
Gráfico da curva ROC deixa evidente que o balanceamento do **oversample foi bem aplicado**, a partir do comportamento semelhante das curvas para as duas classes do modelo.



True class	0	1
0	580	54
1	47	169

Predicted class

Bom resultado do modelo em termos de acerto das classes.



4.3. Técnicas de Inteligência Artificial | Comparativo com modelo LR inicial

Modelo de Regressão Logística (LR) inicial

Matriz de confusão		
True class	0	1
0	543	77
1	129	80
Predicted class		

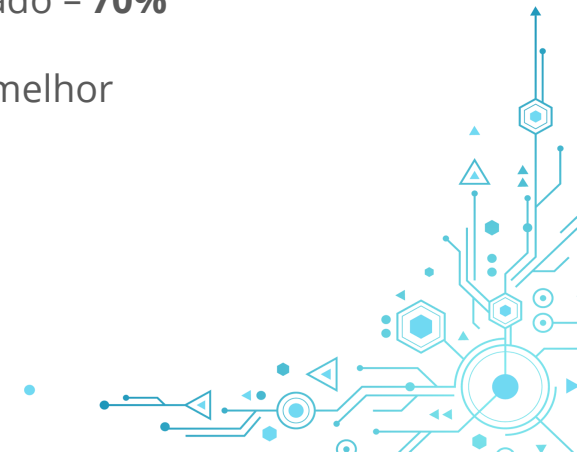
Modelo	AUC
LR Inicial	0.813
LGBM	0.904

Modelo LGBM

Matriz de confusão		
True class	0	1
0	580	54
1	47	169
Predicted class		

Após análise aprofundada do problema e realização de diferentes testes, foram identificadas **oportunidades de melhoria no modelos**, não consideradas previamente, e aplicadas no **PyCaret**. São elas:

- No modelo de LR já foi realizada previamente a **separação da base em treino e teste**, enquanto no PyCaret foi passada a **base completa** para que a ferramenta tomasse a decisão da segregação (o default foi aplicado – **70% treino e 30% teste**).
- Realização do **balanceamento** da base (**oversample**) com método **adasyn**. Este método apresentou melhor performance comparado ao default **SMOTE**.
- Inicialmente não havia sido aplicada a **normalização** das variáveis.



4.4. Técnicas de Inteligência Artificial | Modelo Original x Modelo Tunado

- A partir da seleção do **modelo LGBM**, que apresentou melhor performance, aplicou-se o **tuning**, otimizando pelo **AUC**, com o intuito de verificar a possibilidade de aprimorar ainda mais sua qualidade.
- Abaixo temos o comparativo **do LGBM Original x LGBM Tunado**, onde foi possível verificar melhoria de performance do modelo.

Modelo LGBM Original

		Matriz de confusão	
		0	1
True class	0	580	54
	1	47	169
		Predicted class	

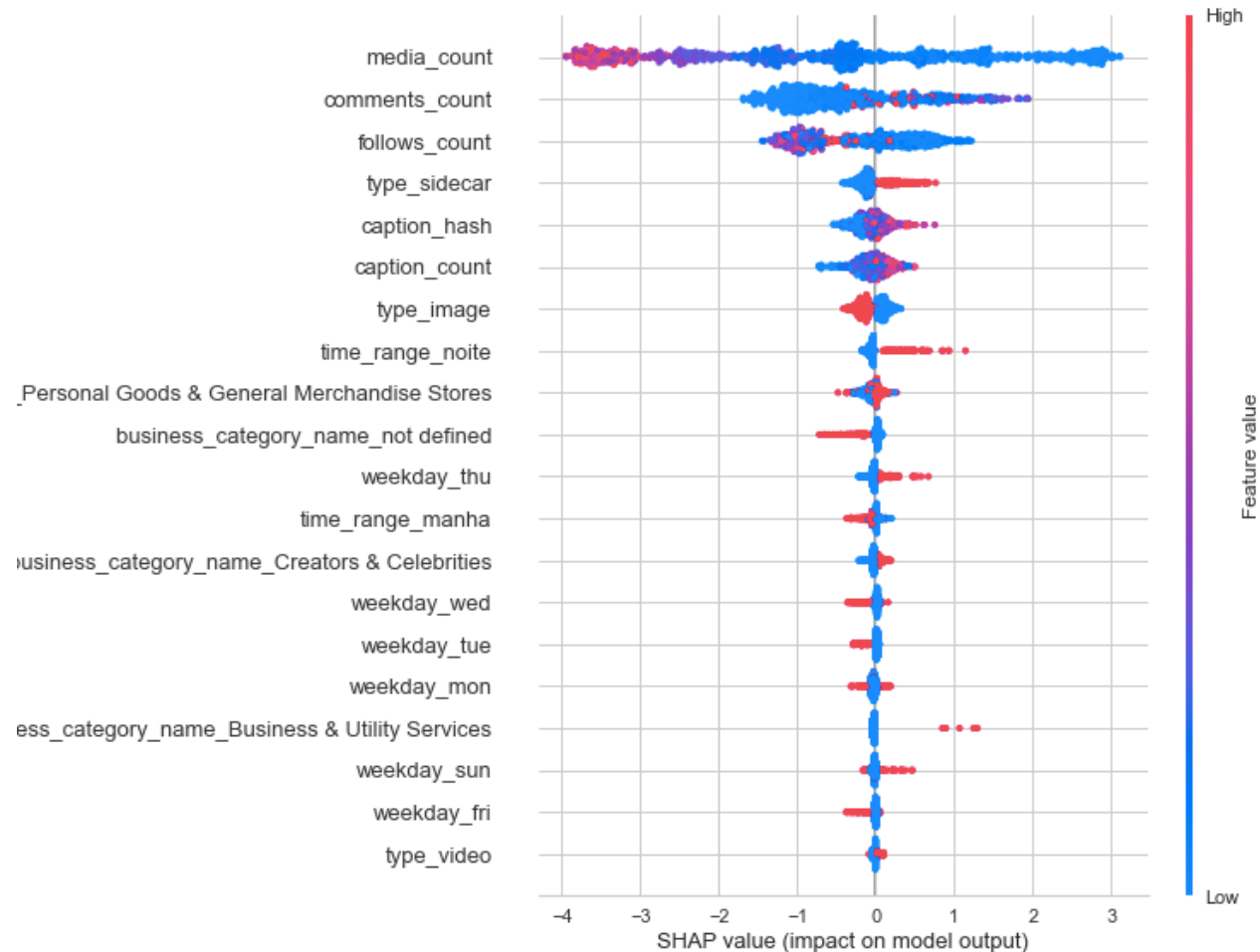
Modelo	AUC
LGBM Original	0.904
LGBM Tunado	0.909

Modelo LGBM Tunado

		Matriz de confusão	
		0	1
True class	0	582	52
	1	46	170
		Predicted class	



Gráfico Shap Value Modelo LGMB Tunado

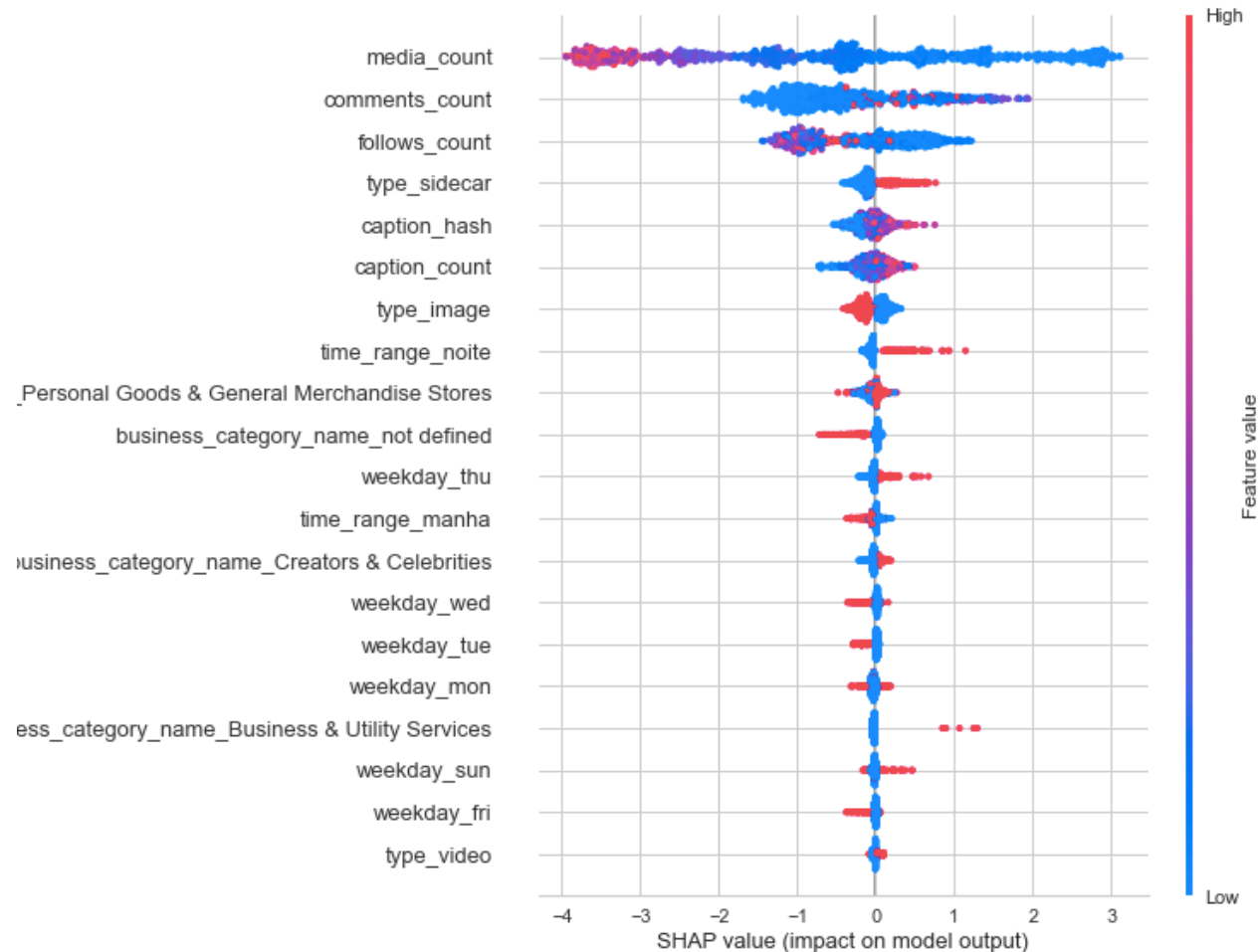


Principais inferências a partir do resultado do modelo:

- **media_count:** feature que melhor discrimina posts engajados ou não. Resultado apresenta que quanto maior o número de posts publicados pelo perfil, menor o engajamento, ou seja, o volume maior de posts realizados não impacta de forma relevante.
- **comments_count:** não temos tantos posts com grandes quantidades de comentários, mas é possível observar que quanto maior o volume de comentários, tende-se a um maior engajamento.
- **follows_count:** quantidade de perfis seguidos pela papelaria não impacta de forma relevante no engajamento.
- **type:** sidecar (carrossel de fotos) é o tipo de post que gera maior engajamento, seguido de vídeos e imagens, que, como observado, não traz engajamento significativo.



Gráfico Shap Value Modelo LGMB Tunado

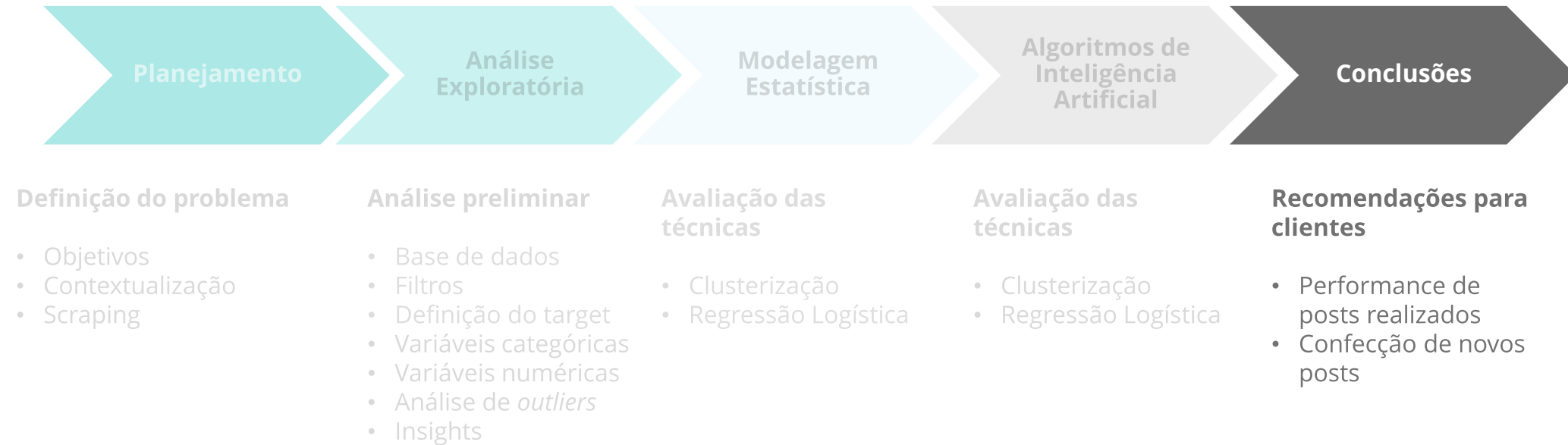


Principais inferências a partir do resultado do modelo:

- **time_range**: período da noite chama a atenção em termos de engajamento e relevância. Já o período da manhã apareceu como não tão recomendado.
- **business_category**: as categorias mais recomendadas para classificação do perfil são Goods & General Merchandise Stores e Creators & Celebrities.
- **weekday**: quinta-feira aparece como o dia de melhor performance e domingo, apesar do baixo volume de posts, aparece como boa oportunidade de teste para gerar engajamento.



Agenda



5. Conclusões

O presente trabalho pode ser aplicado em duas principais abordagens na indústria:



Performance dos posts realizados

Análise e avaliação retroativa dos posts já realizados, identificando principais oportunidades de melhorias.



Confecção de novos posts

A partir da identificação das features que mais impactam no engajamento de um post, torna-se possível a confecção de publicações que tenham maior probabilidade de sucesso, bem como realização de testes A/B mais direcionados.



Obrigada!



Modelo preditivo de engajamento de postagens no Instagram para o mercado de papelarias

Parte 3



11/01/2022

