



# Overview do Dataset

O trabalho que será desenvolvido entre Prepi e DNC visa trabalhar em um **pipeline** de ciência de dados com o objetivo de gerar **Lead Scoring** preciso para a Prepi, **reduzindo o CAC** e encontrando o **perfil ideal de Lead**, que tenha **maior probabilidade** de tornar-se **cliente**.

## Base de dados

A base de dados é resultante da união de 3 diferentes palhinhas, são elas:

- Lista de Leads
- Lista de Clientes e Usuários
- Custo da Campanhas de Anúncios Pagos

Podem existir linhas duplicadas, dados inconsistentes e valores nulos (vazios), dado a natureza dos dados, de informações do mundo real, que são alimentados por toda uma arquitetura de serviços. Por isso, salientamos a importância do cuidado durante a fase de análise e pré-processamento dos dados, focando na eliminação desses ruídos.

## Meta Dados

### Leads

Campo	Descrição	Campo	Descrição
Data	Data em que o Lead se cadastrou	UTM Medium	Canal de contato com a campanha
Hora	Hora em que o Lead se cadastrou	UTM campaign	Identificador (Nome ou título) da campanha
WeekNum	Semana do ano do cadastro do Lead	UTM content	Assunto da Campanha

Campo	Descrição	Campo	Descrição
loja_name	Nome da Loja	landing page version	Versão da página institucional da campanha
Nome	Nome do responsável pela Loja	referrer	Página de origem em que acessou a campanha da Prepi e abriu o anúncio da campanha.
instagram_lead	Instagram da Loja	city	Cidade onde foi realizado o cadastro
Email	Email para cadastro	region	Estado onde foi realizado o cadastro
Whatsapp_lead	Whatsapp para cadastro	became_client	Flag para caso o Lead se tornou cliente
fluxo	Fluxo de contato com anúncios da Prepi. TD representa tráfego direto (primeiro contato diretamente com oferta de plano pago) e PLG representa anúncios para cadastro gratuito.	became_user	Flag para caso o Lead se tornou usuário da Prepi. Todo Cliente é um usuário, automaticamente.
Tamanho do negocio	Média de pedidos mensal online que a Loja tem no momento. Respondida pelo próprio usuário	idade	Idade do responsável pela Loja
UTM source	Origem da campanha		

## Instagram

Campo	Descrição	Campo	Descrição
insta_followers	Quantidade de Seguidores.	Average ReelsLike	Média de curtidas nos Reels

Campo	Descrição	Campo	Descrição
instagram_photos	Quantidade de postagens no Feed.	Average ReelsComment	Média de comentários nos Reels
FotoPerfil	URL com a foto do perfil	Engajamento Reel	Média geral de engajamento nos Reels
Bio	Texto da Biografia do Perfil.	Alcance	Proporção entre a quantidade de 'plays' dos Reels para a quantidade de seguidores. Ou seja, a quantidade de vezes que cada seguidor deu 'play' nos Reels. <b>Uma média.</b>
Bio Link	Link presente na Biografia do Perfil.	Reels View Through	Proporção de pessoas que deram 'play' nos Reels e assistiram de fato o conteúdo.
Frequencia Posts	Média de postagens no feed por dia do perfil em questão.	StoriesCount	Quantidade Stories ativos no momento de cadastro.
Postou recente	Se realizou alguma postagem no Feed nos últimos 7 dias (considerando a data de cadastro como Lead da Prepi).	Has Highlight	Se possui conteúdo em destaque em algumas áreas do Instagram.
Postou muito recente	Se realizou alguma postagem no Feed nos últimos 3 dias (considerando a data de cadastro como Lead da Prepi).	UserTag Count	Quantidade de marcações daquele perfil em postagens de outros perfis.
Criacao estimada	Estimativa de criação do perfil no Instagram.	Is new to instagram	Criou a conta no Instagram recentemente.
OverAll Engajamento	Média de engajamento dos seguidores nas postagens.	Is Bussiness	Se o perfil é de negócios.

Campo	Descrição	Campo	Descrição
Average Comments	Média de comentários dos seguidores nas postagens.	Shoppable Post Count	Quantidade de postagens que possui as ferramentas de Shopping ativas.
Average Likes	Média de curtidas dos seguidores nas postagens.	Public Whatsapp	Número de Whatsapp público no perfil.
Reels Count	Quantidade de medias do tipo Reels	Public Email	Email público no perfil.
Average ReelPlay	Média 'plays' nos Reels	Public Zip	CEP público no perfil.
Average ReelsView	Média de visualizações nos Reels	Category Name	Categoria de negócio do Perfil

## Clientes e Usuários

Campo	Descrição	Campo	Descrição
status	Status da assinatura na Prepi.	older_indamplencia_date	Inadimplência mais antiga, ou seja, primeira fatura em débito.
instagram	Instagram cadastrado durante a criação da Loja virtual.	lead at	Data em que se cadastrou e tornou-se Lead
Whatsapp	Whatsapp cadastrado durante a criação da Loja virtual.	cidade da loja	Cidade da Loja
plano	Nome do Plano escolhido.	estado da loja	Estado da Loja
Payment	Método de pagamento escolhido.	P1	Resposta da primeira pergunta no primeiro acesso ao aplicativo Prepi.
Itv	Valor total da assinatura do plano.	P2	Resposta da segunda pergunta no primeiro acesso ao aplicativo Prepi.

Campo	Descrição	Campo	Descrição
MRR	Valor mensal do plano. LTV / periodicidade	P3	Resposta da terceira pergunta no primeiro acesso ao aplicativo Prepi.
periodicidade	Período de tempo, em meses, da assinatura.	P4	Resposta da quarta pergunta no primeiro acesso ao aplicativo Prepi.
data_trial	Data em que tornou-se testador da Prepi.	P5	Resposta da quinta pergunta no primeiro acesso ao aplicativo Prepi.
data_cliente	Data em que, de fato, tornou-se cliente da Prepi. Normalmente 7 dias após ser testador ou no mesmo momento.	Diagnostico	Com base nos dados e nas respostas, durante o primeiro acesso, é feito um diagnóstico sobre o estado atual da Loja.
data_churn	Data prevista para churn da Prepi. Normalmente, é a data em que a assinatura atual se encerra, sendo necessário a renovação.	Diagnostic Date	Data que o diagnóstico foi feito.
valor_indimplencia	Soma de toda a inadimplência	cost	Custo Prepi para adquirir o Lead/Cliente/Usuário

Nesse primeiro momento, estamos enviando um subconjunto dos dados, devido a carga de trabalho para unificar as diversas bases que temos, principalmente aquelas que possuem os custos das campanhas Prepi ao longo do tempo. Porém, salientamos que a base de dados final terá os mesmos campos do sample enviado. Ao finalizar todo o processo de unificação das bases de dados, enviaremos o **dataset** completo para degustação de todos vocês. Como unificamos várias planilhas, é possível que existam diversos campos nulos, devido a não correspondência.

Ex: existem Clientes que não estavam na base de Leads, então, os campos que originalmente estavam lá, devem aparecer como nulos para esse determinado cliente. Campos como UTM, e informações de Instagram não devem está na linha de clientes que não estão presentes por Email.