CORONATHON

*Equipe 28*

*Participantes: Allan Rogge, Érico Encarnação, Ivan Ormenesse, Jonas Lima, Maurício Trujillo, Rodrigo Figueira*

*PROJETO: PROPOSTA DE API PÚBLICA BASEADA EM MODELAGEM ESTATÍSTICA PARA MELHORIA DE ASSERTIVIDADE DO SINE*

## Resumo do Projeto

Nosso produto final será uma API pública pela qual o SINE e, eventualmente, outros stakeholders possam acessar a propensão, descrita como uma probabilidade de um trabalhador ser contratado para uma determinada vaga. Nossa missão é reduzindo o atrito e agilizar o processo de contratação para trabalhadores e empregadores, melhorando a visualização de vagas no front-end e facilitando encontrar pessoas e posições, inclusive aquelas portadoras de deficiência.

## Etapa de Imersão

Iniciamos o entendimento do problema discutindo detalhadamente os principais problemas apresentados pelo SINE e, a partir da explicação das bases apresentadas nas primeiras etapas do evento, escolhemos alguns problemas para resolver.

Documentamos todo o processo por meio do Excalidraw gerando tanto uma imagem em PNG como um arquivo JSON que detalha não apenas o processo de Imersão, mas também a ideação, prototipação e a solução proposta.

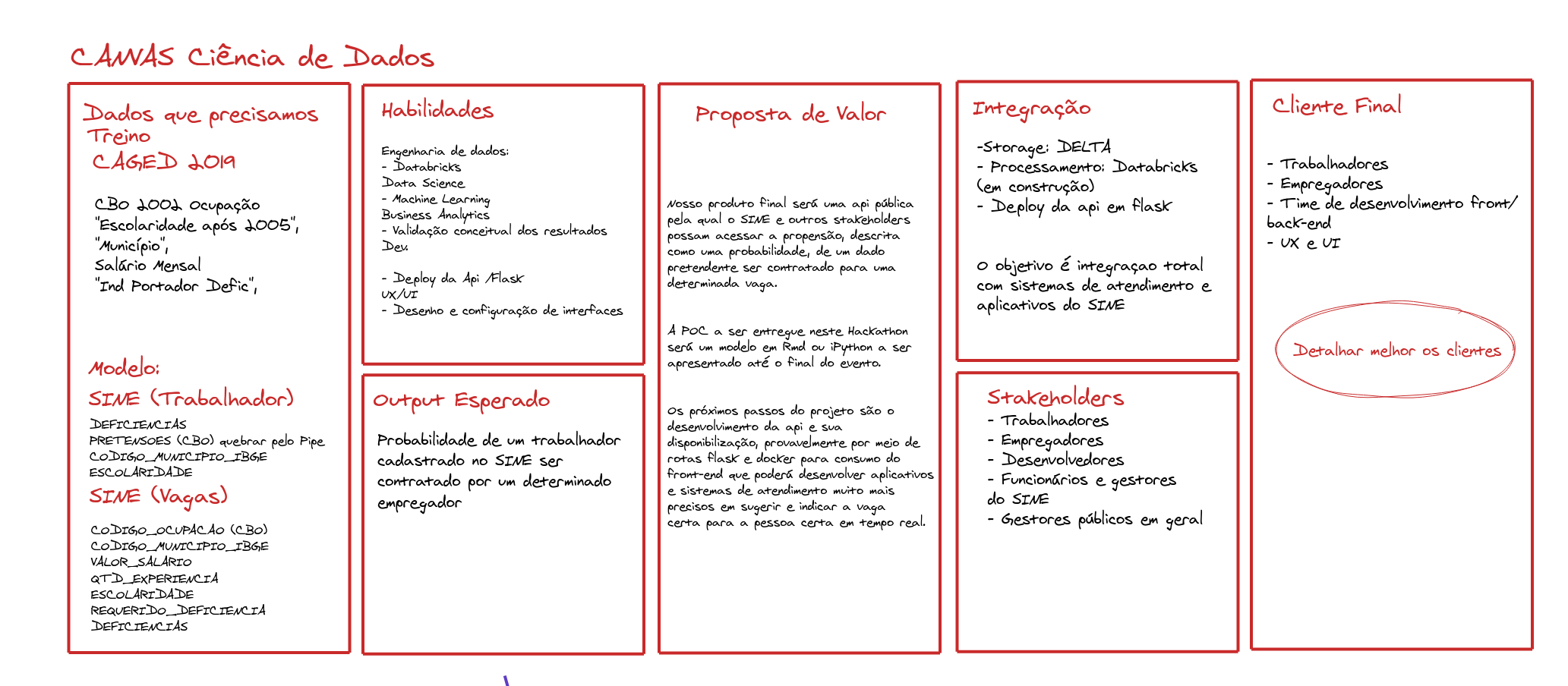
Uma imagem contendo texto, mapa

Descrição gerada automaticamente

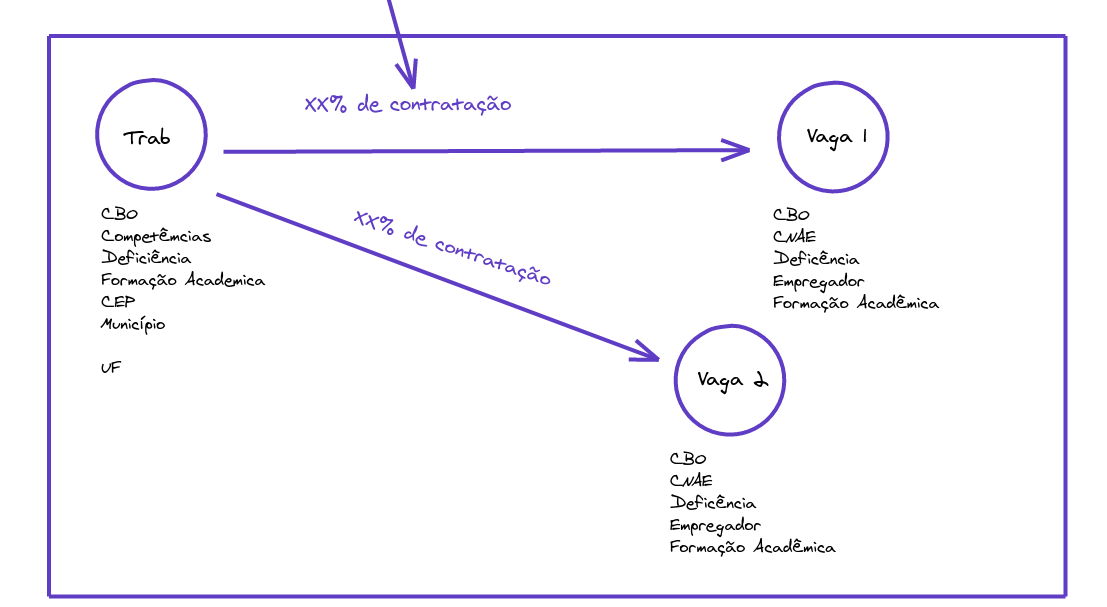
Além disso, fizemos algumas análises descritivas com o propósito de levantar e validar algumas hipóteses que se mostraram valiosas para a etapa de ideação:

## Etapa de Ideação

*CANVAS:*



*Funcionamento do Modelo:*

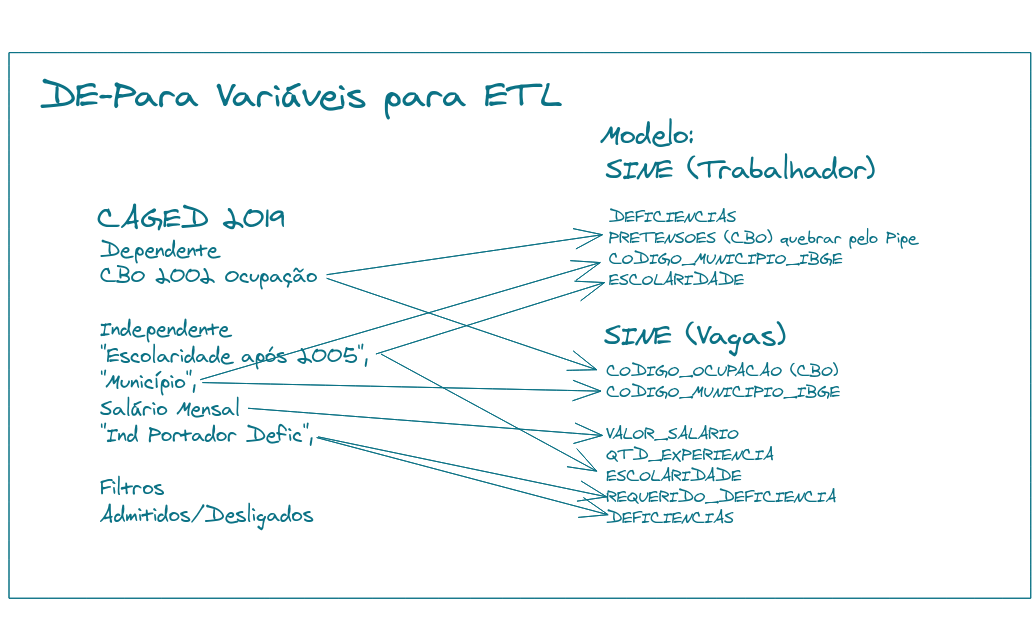


*O algoritmos será construído com base nas informações da base do CAGED e será treinado com um grande volume de informações para que, com base nas informações também disponíveis na base do SINE possam alimentar o modelo em tempo real e responder a probablidade de contração para uma dada relação colaborado-vaga.*

*O modelo será treinado com base nos dados do Sudeste da base da CAGED, validado com base nos dados do Sul e aplicado aos dados disponibilizados pelo SINE.*

*Detalhamento do ETL*

*Para isto avaliamos os dicionários de dados disponíveis em cada base e planejamos como o ETL deveria ser realizado para assegurar que haja informações suficientes para a estimação de um MVP do modelo.*



*Primeiramente, tentamos uma abordagem tradicional com bancos relacionais do Azure, no entanto a performance foi instatisfatória e pivotamos a solução para o Databricks.*

*Desenho final do protótiopo em desenvolvimento*

*Armazenamento de dados Delta com apoio de solução gratuita do Databricks.*

*MVP do modelo baseado em regressões logísticas*

*Apresentação em Rmd dos primeiros resultados*