

# Data Science Workflow Canvas\*

Start here. The sections below are ordered intentionally to make you state your goals first, followed by steps to achieve those goals. You're allowed to switch orders of these steps!

Conceptualized by Jasmine Vasandani using notes from General Assembly's Data Science Immersive. Format inspired by Business Model Canvas.

<b>Title:</b> A identificação de clientes em uma Instituição Financeira, com histórico de reclamações na Ouvidoria, suscetíveis a registrar reclamações no Banco Central do Brasil			
<div><b>1</b></div> <div><b>Problem Statement</b></div> <div>What problem are you trying to solve? What larger issues do the problem address?</div> <p>Os registros de Reclamação no Banco Central são fonte de extrema relevância Estratégica para as Instituições Financeiras brasileiras.</p> <p>Como identificar perfis de clientes mais suscetíveis a apresentar suas demandas ao Bacen, de forma proativa mantendo a satisfação e o relacionamento com os clientes em patamares "saudáveis"?</p>	<div><b>2</b></div> <div><b>Outcomes/Predictions</b></div> <div>What prediction(s) are you trying to make? Identify applicable predictor (X) and/or target (y) variables.</div> <p>Variáveis explicativas: dados armazenados pela própria Instituição (dados cadastrais, bem como dados das reclamações registradas pelos clientes, na Ouvidoria da Instituição).</p> <p>Variável de interesse(alvo): suscetibilidade de registro de demanda junto ao Banco Central do Brasil. Registra (1), ou Não Registra (0).</p>	<div><b>3</b></div> <div><b>Data Acquisition</b></div> <div>Where are you sourcing your data from? Is there enough data? Can you work with it?</div> <p>Bancos de Dados mantidos pela Instituição, contendo bases diversas, cada uma contendo dados específicos, de acordo com a Diretoria responsável.</p> <p>1) Base Ouvidoria (dados referentes a histórico de reclamações de clientes);</p> <p>2) Base cadastral (dados cadastrais dos clientes da Instituição).</p>	<div><b>6</b></div> <div><b>Data Preparation</b></div> <div>What do you need to do to your data in order to run your model and achieve your outcomes?</div> <p>Label Encoding das classes utilizadas;</p> <p>Tratamento de Nulls ;</p> <p>Ajuste nos tipos de dados, frente à natureza dos dados registrados (números armazenados como texto, por exemplo);</p> <p>Criação de variáveis de interesse, com base nas informações à disposição;</p> <p>Balanceamento de classes.</p>
<div><b>4</b></div> <div><b>Modeling</b></div> <div>What models are appropriate to use given your outcomes?</div> <p>Algoritmo de aprendizado supervisionado, já que as Bases de Dados utilizadas são todas estruturadas e mantidas em formato de tabelas, com a variável explicativa já com resultado conhecido.</p> <p>Utilizarei 5 algoritmos diferentes e escolherei o de melhor performance em explicar o problema definido: LogisticRegression(), LinearSVC(), KNeighborsClassifier(), RandomForestClassifier() e GaussianNB()</p>	<div><b>5</b></div> <div><b>Model Evaluation</b></div> <div>How can you evaluate your model's performance?</div> <p>Métricas de avaliação dos algoritmos:</p> <p>Recall</p> <p>Fbeta-Score</p> <p>(Precisão, apenas como elemento presente no cálculo do Fbeta-Score, já que o problema apresentado exige foco maior nos Falsos Negativos).</p>		

## ✓ Activation

When you finish filling out the canvas above, now you can begin implementing your data science workflow in roughly this order.

