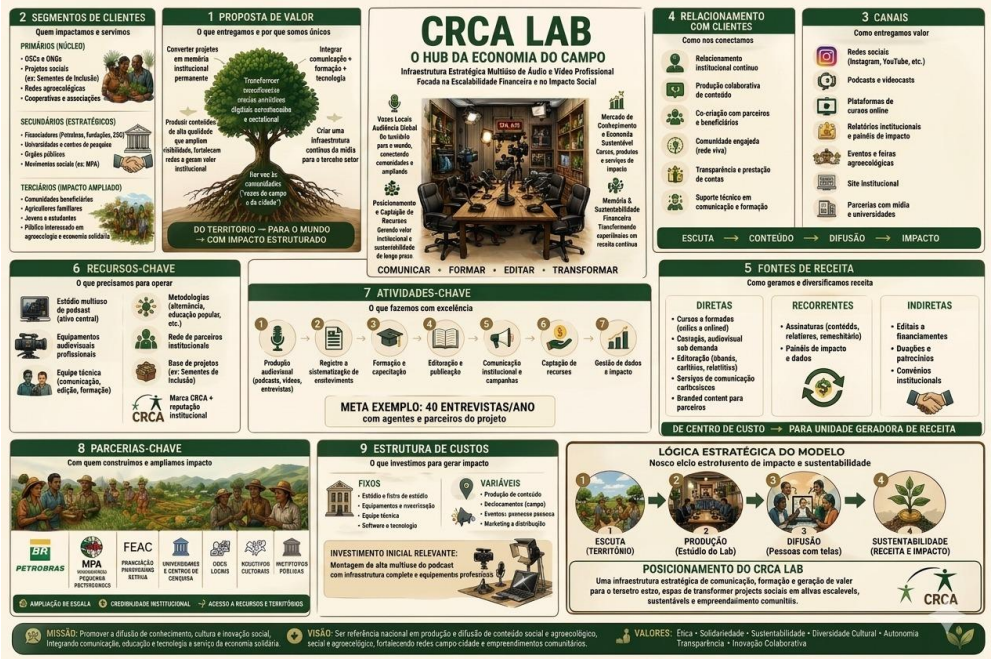
CRCA LAB

Modelo de Projeto Aplicado em Marketing Analytics e Inteligência Artificial

Projeto de "Spin-Off" | **impacto social estruturado**

Aprofundamento em Business Analytics, RStudio e governança de dados



CRCA Lab: modelo orientado à decisão para priorização de conteúdos, públicos e sustentabilidade financeira de um hub de comunicação, formação e impacto social

Curso: MBA em Marketing Analytics e Inteligência Artificial
Projeto aplicado: CRCA Lab — do território → para o social → com impacto estruturado.
Aluno(a): Juliano Vieira Gonçalves < Auxiliar de Projetos | CRCA – Centro de Referência em Cooperativismo e Associativismo >

Professor(a): [preencher]
Instituição: ECA-USP
Data: [preencher]

Comentado [MC1]: Paleta sugerida:
- Verde escuro institucional: #0F3B24
- Verde médio: #2F6B3F
- Verde claro/agroecológico: #8AB833
- Creme/fundo editorial: #F4EAD7
- Dourado/ênfase: #C9A85A
- Cinza técnico: #D9D9D9

Elementos gráficos recomendados:
- Ícones de escuta, produção, difusão, sustentabilidade, dados e impacto.
- Linhas ou setas representando fluxo analítico.
- Imagens de território, agroecologia, estúdio, redes comunitárias e painéis de dados.
- Boxes de destaque para decisões, indicadores e achados do Google Colab.

Sumário

1. Introdução

- *O CRCA Lab como spin-off institucional voltado à comunicação, formação, editoração e tecnologia social. Como priorizar conteúdos, públicos, canais e fontes de receita com base em dados.*

2. Objetivos

- Um modelo híbrido em Python/Google Colab e R/RStudio para apoiar decisões estratégicas do CRCA Lab.

3. Caracterização do CRCA Lab

- *Apresentação do CRCA Lab como uma iniciativa de inteligência institucional vinculada à comunicação, formação, editoração, tecnologia social e análise de dados. O item deve contextualizar sua origem, sua relação com o CRCA, seus públicos prioritários, suas frentes de atuação e seu papel na transformação de experiências territoriais em produtos de conhecimento, comunicação e sustentabilidade.*

4. Canvas analítico do projeto

- *Apresentação da lógica do Canvas analítico do CRCA Lab, conectando conteúdos, públicos, canais, custos, alcance, engajamento, impacto social e potencial de receita. O item introduz o score_prioridade como indicador central para apoiar decisões estratégicas.*

5. Base de dados

- *Base inicial inspirada em estudos práticos de analytics. Cada linha representa um conteúdo, produto ou entrega do CRCA Lab. Cada coluna representa uma variável estratégica, como formato, público, canal, custo, alcance, engajamento, impacto social e potencial de receita.*

6. Tratamento e organização dos dados em Python e R

- Estruturas fundamentais em Python e R. Organização inicial dos dados do CRCA Lab com Pandas e tidyverse. Classificação de conteúdos, públicos e canais, validação das variáveis e preparação da base para análise.

7. Construção de indicadores

- *Alcance, engajamento, impacto social, potencial de receita, indicadores derivados e score_prioridade. Criação de métricas comparativas para avaliar conteúdos, canais e formatos.*

8. Análises exploratórias

- *Exploração inicial da base com Python e Pandas. Verificação da dimensão do dataset, identificação de categorias estratégicas, estatísticas descritivas, comparação entre grupos, análise de relações entre variáveis e identificação de oportunidades.*

9. Visualizações no Google Colab e no RStudio

- Criação de gráficos, relatórios reproduzíveis e painéis interativos para comunicar padrões, rankings, comparações e oportunidades estratégicas identificadas na análise.

10. Uso de inteligência artificial

- *Comunicação, planejamento estratégico, gestão de dados simples, treinamento de habilidades, captação de recursos, monitoramento de impacto, governança, voluntariado, roteirização e apoio à tomada de decisão.* [1] *Apoio à formulação de perguntas analíticas, geração de códigos em Python, explicação de resultados, ajuste de visualizações, levantamento de hipóteses e interpretação estratégica dos dados do CRCA Lab.*

11. Resultados esperados

- *Ranking de conteúdos prioritários, canais mais promissores, formatos com maior potencial de receita, indicadores de impacto e identificação de oportunidades estratégicas para decisão editorial, institucional e financeira.*

12. Discussão e aprofundamento em Business Analytics

- Integra o modelo já desenvolvido em Python, Google Colab e RStudio aos conceitos de OKRs, Lean Analytics, BPR, maturidade analítica, Data Mart/Data Warehouse, modelos preditivos, vieses decisórios e inferência causal. O objetivo é demonstrar como o CRCA Lab pode transformar dados em rotinas de decisão, experimentação e aprendizagem institucional.

13. Conclusão

- *O CRCA Lab deixa de ser apenas uma estrutura de produção audiovisual e passa a operar como uma unidade de inteligência institucional, capaz de transformar escuta*

[1] ANEXO V – ChatGPT para reconfigurar OSCs...

territorial em conteúdo, conteúdo em impacto e impacto em sustentabilidade.

14. Próximos passos

- Tradução do modelo analítico para uma interface digital interna, com banco de dados, dashboard em Shiny e relatórios em Quarto, permitindo acesso inicial às equipes dos projetos, gestores e, em níveis definidos, parceiros e financiadores.

15. Apêndices

- Espaço destinado aos códigos utilizados no Google Colab e no RStudio, bases simuladas, scripts em R, prints de tabelas, protótipos de dashboard, fórmulas dos indicadores e materiais complementares produzidos durante o desenvolvimento do projeto.

16. Checklist de preenchimento final

- Lista de verificação para garantir que o projeto contenha todos os elementos necessários: problema, objetivos, caracterização do CRCA Lab, base de dados, tratamento em Python e R, indicadores, análises, visualizações, dashboard, uso de IA, resultados esperados, discussão, conclusão, próximos passos e referências.

17. Referências e documentos de apoio[2]

- Reunião das referências bibliográficas, apostilas do curso, materiais institucionais do CRCA Lab, documentos de apoio sobre análise de dados, inteligência artificial, impacto social, comunicação institucional e sustentabilidade de organizações da sociedade civil.

[2] ANEXO IV – (V02) Gemini_CRCA_Lab_Canvas...

Resumo Executivo

O presente projeto aplicado propõe a criação de um modelo analítico híbrido, utilizando Python no Google Colab e R no RStudio, para apoiar decisões estratégicas do CRCA Lab. O Colab atua como ambiente de prototipagem e aprendizagem em Python; o RStudio amplia a solução com tratamento estatístico, visualização, relatórios reproduzíveis e desenvolvimento de painéis web. O Lab é concebido como uma infraestrutura de comunicação, formação, editoração social e inovação, voltada à transformação de experiências territoriais em conteúdos compreensíveis, replicáveis e escaláveis.

A proposta do trabalho é demonstrar como dados sobre públicos, canais, formatos, custos, alcance, engajamento, impacto social e potencial de receita podem orientar a priorização de conteúdos, a escolha de canais e a sustentabilidade institucional do CRCA Lab.

Síntese da decisão analisada

[Preencher com a decisão principal: quais conteúdos produzir primeiro; quais canais priorizar; quais públicos atender na fase inicial; ou como combinar impacto e sustentabilidade financeira.] [Inserir aqui a principal decisão estratégica analisada: por exemplo, quais conteúdos produzir primeiro, quais canais priorizar ou quais públicos atender em uma primeira fase.]

Ideia do e-book antigo	Tradução atual para o CRCA Lab Analytics
ChatGPT como parceiro de OSCs →	IA como assistente estratégico do CRCA Lab
Prompts específicos para OSCs →	Modelos de classificação e recomendação
Comunicação mais eficiente →	Análise de canais, públicos e conteúdos
Gestão de dados simples	Organização em DataFrames no Pandas
Apoio à tomada de decisão	Score de prioridade e dashboards
Simulação de cenários	Funções em Python para testar hipóteses
Treinamento de habilidades	Geração de roteiros, entrevistas e respostas
Captação e sustentabilidade	Análise de potencial de receita por formato

Resultado esperado do modelo (Produto analítico final)

[Preencher com o resultado do Google Colab: ranking de prioridade, matriz de decisão, gráficos por canal, simulação de receita e recomendações executivas.] [Inserir aqui o resultado esperado: ranking de conteúdos, matriz de prioridade, gráfico de canais, simulação de receita ou painel de indicadores.]

A versão V06 aprofunda o componente de Business Analytics. Além do score de prioridade, o projeto passa a contemplar governança de dados, redesenho de processos, OKRs, MVP analítico, maturidade institucional, regressão múltipla e desenho de experimentos. A proposta não é construir um sistema sofisticado de uma única vez, mas criar uma trajetória verificável: organizar dados, produzir métricas acionáveis, testar hipóteses e somente depois ampliar modelos preditivos ou causais.

1. Introdução

1.1 Contextualização

O CRCA Lab nasce como proposta de infraestrutura estratégica para fortalecer a comunicação, a formação e a editoração social vinculadas à economia solidária, à agroecologia, ao cooperativismo e à inclusão socioprodutiva. Sua lógica parte da necessidade de transformar ações territoriais, muitas vezes dispersas ou efêmeras, em memória institucional, conteúdos formativos, publicações, campanhas, podcasts, videocasts e produtos digitais.

[Inserir aqui uma contextualização sobre o CRCA, o projeto Sementes de Inclusão, a sala multiuso de podcast e a importância da comunicação para projetos sociais.]

1.2 Justificativa

Projetos sociais produzem grande volume de conhecimento prático, relatos, metodologias, evidências e resultados. No entanto, parte significativa desse conhecimento se perde por falta de sistematização, difusão e análise de dados. O CRCA Lab surge como resposta a esse desafio, propondo uma estrutura capaz de organizar, produzir, distribuir e mensurar conteúdos de impacto.

No contexto do MBA em Marketing Analytics e Inteligência Artificial, o projeto permite aplicar recursos de Python para organizar dados, construir indicadores, analisar desempenho e apoiar decisões de marketing, comunicação e sustentabilidade.

1.3 Problema de decisão

Problema central

O CRCA Lab precisa decidir, com base em dados, quais conteúdos produzir primeiro, para quais públicos, em quais canais e com qual lógica de sustentabilidade: impacto institucional, captação, formação, prestação de contas ou geração de receita. **Como o CRCA Lab pode usar dados, Python e inteligência artificial para priorizar sua produção de conteúdos, medir impacto comunicacional e apoiar decisões de sustentabilidade institucional?**

Essa formulação conversa diretamente com o Canvas reorganizado, que define o Lab como uma infraestrutura de comunicação, formação, editoração e tecnologia social, voltada a transformar projetos sociais em conteúdos compreensíveis, replicáveis e escaláveis.

Perguntas orientadoras:

- Quais públicos devem ser priorizados pelo CRCA Lab em sua fase inicial?
- Quais formatos de conteúdo apresentam maior potencial de impacto, alcance e receita?
- Quais canais devem receber maior esforço de produção e distribuição?
- Como transformar dados de produção e comunicação em indicadores úteis para decisão?
- Como o Google Colab e o RStudio podem apoiar a construção de um modelo simples, reproduzível e acessível de priorização estratégica?

Em resumo, **quais conteúdos, canais e públicos devem ser priorizados pelo CRCA Lab para maximizar alcance, engajamento, impacto institucional e potencial de receita?**

2. Objetivos

2.1 Objetivo geral

Desenvolver um modelo analítico híbrido em Python e R, utilizando Google Colab e RStudio, para apoiar a tomada de decisão do CRCA Lab na priorização de conteúdos, canais, públicos e estratégias de sustentabilidade institucional.

2.2 Objetivos específicos

- Organizar uma base de dados inicial sobre conteúdos, públicos, canais, formatos, custos, alcance e engajamento.
- Tratar e padronizar dados com recursos básicos de Python e Pandas.
- Criar indicadores de alcance, engajamento, eficiência de custo, impacto social e potencial de receita.
- Gerar rankings e visualizações para apoiar decisões estratégicas.

- Propor um modelo inicial de recomendação de formatos e prioridades editoriais.
- Relacionar os resultados analíticos ao Canvas estratégico do CRCA Lab.
- Demonstrar como o RStudio pode transformar a análise em relatórios reproduzíveis e em um dashboard web interno para as equipes dos diferentes projetos.

Decisão	Pergunta analítica	Saída em Python
Priorizar públicos	Quais segmentos têm maior potencial de impacto e receita?	ranking de clientes/personas
Priorizar conteúdos	Quais temas devem virar podcast, curso, relatório ou campanha?	matriz de prioridade
Avaliar canais	Instagram, YouTube, site, cursos ou eventos geram mais resultado?	gráficos por canal
Medir impacto	A produção do Lab aumenta alcance, memória institucional e prestação de contas?	indicadores e dashboards
Sustentabilidade	Quais fontes de receita podem reduzir dependência de editais?	simulação de receita

- Aplicar OKRs e Lean Analytics para transformar a implantação do CRCA Conecta em ciclos mensuráveis de execução e aprendizagem.
- Mapear processos, clientes internos, papéis e níveis de maturidade analítica do CRCA e de seus projetos.
- Testar modelos estatísticos em RStudio para estimar fatores associados a custos, atrasos, adoção interna, alcance e potencial de receita.
- Propor experimentos e métodos de avaliação capazes de distinguir associação, previsão e possível efeito causal.

Isso se encaixa bem na lógica estratégica do modelo: escuta → produção → difusão → sustentabilidade, já destacada na apresentação do CRCA Lab.[3]

3. Caracterização do CRCA Lab

3.1 Propósito institucional

O CRCA Lab é concebido como hub de comunicação, formação, editoração social e tecnologia aplicada ao impacto. Sua finalidade é transformar experiências territoriais em conteúdos permanentes, capazes de fortalecer a memória institucional, a prestação de contas, a formação de públicos e a sustentabilidade financeira de iniciativas sociais.

Elemento	Formulação para preencher
Missão	[Inserir missão do CRCA Lab]
Visão	[Inserir visão do CRCA Lab]
Valores	[Inserir valores institucionais: ética, solidariedade, sustentabilidade, diversidade cultural, autonomia, transparência e inovação colaborativa]

[3] ANEXO I – GPT_CRCA_Lab_Canvas_reorganizado

3.2 Proposta de valor

O CRCA Lab entrega valor ao converter projetos sociais em ativos de comunicação, formação e conhecimento. Seu diferencial está na integração entre escuta territorial, produção audiovisual, editoração, educação popular, dados e tecnologia social.

Proposta de valor resumida

Transformar projetos sociais em conteúdos compreensíveis, replicáveis e escaláveis, gerando memória institucional, impacto comunicacional e sustentabilidade.

3.3 Lógica estratégica do modelo

Etapa	Descrição	Produto de dados associado
Escuta	Captação de vozes, relatos, demandas, experiências e evidências do território.	Entrevistas, depoimentos, temas recorrentes, demandas.
Produção	Transformação da escuta em podcasts, vídeos, cursos, relatórios, publicações e campanhas.	Formato, custo, tempo de produção, status.
Difusão	Distribuição em redes sociais, site, plataformas EAD, eventos e parceiros.	Canal, alcance, engajamento, cliques.
Sustentabilidade	Geração de receita, captação, fortalecimento institucional e ampliação do impacto.	Potencial de receita, parceiros, conversões, contratos.

3.4 Segmentos de clientes e públicos

Segmento	Públicos	Papel no modelo
Primários	OSCs, ONGs, projetos sociais, redes agroecológicas, cooperativas e associações.	Núcleo de atendimento, produção de conteúdo e formação.
Secundários	Financiadores, universidades, gestores públicos e movimentos sociais.	B2B institucional, validação, captação, parcerias e escala.
Terciários	Comunidades beneficiárias, agricultores familiares, jovens, estudantes e público interessado em agroecologia, economia solidária e impacto social	Alcance social, difusão pública e legitimidade territorial.

3.5 Base conceitual do projeto

O CRCA Lab pode ser tratado como uma **infraestrutura estratégica de dados e comunicação**.

Nos documentos anexos, ele aparece como um hub que organiza:

1. **Escuta territorial**, com entrevistas, depoimentos e coleta de narrativas;
2. **Produção de conteúdo**, com podcast, videocast, cursos, publicações e relatórios;

3. **Difusão**, por redes sociais, plataformas, eventos, site e parceiros;
4. **Sustentabilidade**, com cursos, assinaturas, serviços, editoração, produção sob demanda e parcerias.

A sala multiuso de podcast também já possui objetivos mensuráveis, como produzir pelo menos **40 entrevistas anuais** com agentes e parceiros do projeto, além de criar acervo audiovisual sobre agroecologia, UTDs e cadeias produtivas fechadas.[4]

4. Canvas Analítico do Projeto

4.1 Da estratégia institucional ao modelo de dados

O Canvas do CRCA Lab será traduzido em variáveis analíticas. Cada dimensão estratégica do modelo se transforma em uma coluna, categoria ou indicador na base de dados.

Dimensão do Canvas	Tradução analítica no projeto	Variável sugerida (exemplo)
Segmentos de clientes	Públicos atendidos	publico
Proposta de valor	Tipo de entrega e benefício prometido	formato / tema
Canais	Meios de distribuição	canal
Relacionamento	Tipo de interação e suporte	nivel_relacionamento
Fontes de receita	Potencial econômico	potencial_receita
Recursos-chave	Capacidade produtiva	equipe / equipamento
Atividades-chave	Produção realizada	conteudo / categoria
Parcerias-chave	Apoio institucional	parceiro
Estrutura de custos	Investimento necessário	custo_estimado

4.2 Hipótese analítica

Hipótese do modelo

Conteúdos com maior combinação entre impacto social, prioridade institucional, alcance estimado, engajamento e potencial de receita devem ser priorizados na agenda inicial do CRCA Lab.

4.3 Modelo de decisão

O modelo de decisão será baseado em um score de prioridade, calculado a partir de pesos atribuídos a diferentes dimensões.

Indicador	Peso sugerido	Justificativa

[4] ANEXO II – Sala_PODCAST

Impacto social	35%	Centralidade da missão do CRCA Lab.
Prioridade institucional	30%	Alinhamento com projetos, parceiros e prestação de contas.
Potencial de receita	20%	Contribuição para sustentabilidade do Lab.
Taxa de engajamento	15%	Sinal de interesse e relevância para o público.

 # Fórmula conceitual:

Score de Prioridade = Impacto Social + Prioridade Institucional + Potencial de Receita + Engajamento + Eficiência de Custo

```
df['score_prioridade'] = (
    df['impacto_social'] * 0.35 +
    df['prioridade_institucional'] * 0.30 +
    df['potencial_receita'] * 0.20 +
    df['taxa_engajamento'] * 10 * 0.15
)
```

[Inserir justificativa dos pesos adotados.]

4.4 Base de Inteligência Artificial para OSCs



O CRCA Lab parte da compreensão de que a inteligência artificial pode apoiar OSCs em múltiplas dimensões: comunicação institucional, planejamento estratégico, captação de recursos, gestão de dados simples, monitoramento de impacto, formação de equipes e sistematização de conhecimento. Essa visão já estava presente em reflexão anterior sobre o uso do ChatGPT por organizações sociais, na qual a IA era compreendida como uma ferramenta complementar capaz de ampliar a capacidade técnica e comunicacional das equipes.

No presente projeto, essa reflexão é atualizada por meio do uso de Python e Google Colab. A IA generativa deixa de ser tratada apenas como recurso de produção textual e passa a integrar um modelo de decisão orientado por dados. Assim, conteúdos, públicos, canais, custos, engajamento, impacto social e potencial de receita podem ser organizados em bases estruturadas, analisados com bibliotecas como Pandas e visualizados por meio de gráficos e rankings.

Dessa forma, o CRCA Lab se posiciona como uma infraestrutura de inteligência institucional para OSCs, articulando escuta territorial, produção de conteúdo, análise de dados, recomendação estratégica e sustentabilidade.



5. Base de Dados

Estruturação de uma base inicial inspirada em estudos práticos de analytics. Cada linha representa um conteúdo, produto ou entrega do CRCA Lab, enquanto cada coluna representa uma variável relevante para análise, como formato, público, canal, custo, alcance, engajamento, impacto social e potencial de receita.

5.1 Origem dos dados

a) **Dados reais internos**: histórico de conteúdos, redes sociais, site, cursos, eventos, entrevistas e relatórios.

b) **Dados planejados**: metas do CRCA Lab, como entrevistas anuais, publicações, cursos e pessoas alcançadas.

c) **Dados simulados**: utilizados para demonstrar o funcionamento do modelo enquanto a base real ainda está em estruturação.

5.2 Estrutura inicial da base

Coluna	Descrição	Tipo esperado
conteudo	Nome ou descrição do conteúdo	texto
tema	Tema principal	texto
publico	Público prioritário	texto
canal	Canal de distribuição	texto
formato	Podcast, vídeo, curso, relatório etc.	texto
custo_estimado	Custo de produção	número
alcance_estimado	Pessoas alcançadas	número
engajamento_estimado	Interações esperadas	número
potencial_receita	Escala de 1 a 5	número
impacto_social	Escala de 1 a 5	número
prioridade_institucional	Escala de 1 a 5	número
data_planejada	Data prevista de publicação	data
status	Planejado, em produção, publicado	texto

5.3 Leitura da base no Colab

```
import pandas as pd


# Leitura de arquivo CSV
df = pd.read_csv('/content/base_crca_lab.csv')


# Visualização inicial
df.head()

</> Markdown

# Nesta primeira versão, a base foi criada de forma simulada no próprio notebook, com dados realistas sobre conteúdos, públicos, canais, formatos, custos, alcance, engajamento, impacto social e potencial de receita. Em uma etapa posterior, a estrutura poderá ser substituída por uma planilha CSV real, exportada das rotinas de produção do CRCA Lab.
```

5.4 Diagnóstico inicial

```
# Dimensões da base
print(df.shape)

# Informações gerais
df.info()

# Estatísticas descritivas
df.describe()

# Verificação de valores ausentes
df.isna().sum()
```

6. Tratamento e Organização dos Dados em Python e R

As anotações da aula foram incorporadas como uma trilha operacional: importar e visualizar a base, explorar dados, limpar campos, criar métricas, agregar resultados, construir rankings, gerar gráficos e finalizar com storytelling orientado à decisão. No RStudio, essa mesma trilha pode ser executada com o ecossistema tidyverse, mantendo scripts, dados, resultados e documentação em um projeto reproduzível.

No projeto aplicado do CRCA Lab, essa mesma lógica será adaptada para o campo da comunicação, formação e tecnologia social. Cada linha da base representará um conteúdo, produto, entrega ou ação estratégica do laboratório. As colunas, por sua vez, registrarão variáveis como formato, público prioritário, canal de distribuição, custo estimado, alcance esperado, engajamento, impacto social e potencial de receita.

Dessa forma, o projeto transforma elementos qualitativos do Canvas em uma base estruturada de dados, permitindo análises comparativas, cálculo de indicadores e geração de rankings de prioridade.

Passos da aula de analytics	Papel analítico	Adaptação ao CRCA Lab
1. Importar e visualizar a base	Data Collector / Data Explorer	Carregar CSV com conteúdos, públicos, canais e métricas do Lab.
2. Explorar os dados	Data Profiler	Verificar linhas, colunas, tipos, nulos e consistência.
3. Limpeza básica	Data Cleaner	Padronizar textos, datas, valores numéricos e categorias.
4. Criar novas colunas	Data Modeler	Criar taxa de engajamento, eficiência de custo e score de prioridade.
5. Análises simples	Data Analyst	Agrupar por canal, formato, público e tema.
6. Ranking	Performance Analyst	Listar conteúdos, canais e públicos prioritários.
7. Gráficos básicos	Data Visualizer	Comunicar resultados em barras e linhas.
8. Storytelling com dados	Data Storyteller	Construir narrativa executiva para decisão.

6.1 Padronização de textos

```
import unicodedata

def remover_acentos(texto):
    texto = str(texto)
    texto = unicodedata.normalize('NFKD', texto)
    return ''.join([c for c in texto if not unicodedata.combining(c)])

colunas_texto = ['conteudo', 'tema', 'publico', 'canal', 'formato']

for coluna in colunas_texto:
    df[coluna] = df[coluna].astype(str).str.strip()
    df[coluna + '_normalizado'] = df[coluna].apply(remover_acentos).str.lower()
```

6.2 Conversão de valores numéricos e tratamento de datas

```
colunas_numericas = [
    'custo_estimado', 'alcance_estimado', 'engajamento_estimado',
    'potencial_receita', 'impacto_social', 'prioridade_institucional'
]

for coluna in colunas_numericas:
    df[coluna] = pd.to_numeric(df[coluna], errors='coerce').fillna(0)

df['data_planejada'] = pd.to_datetime(df['data_planejada'], errors='coerce')
df['mes_planejado'] = df['data_planejada'].dt.month
df['ano_planejado'] = df['data_planejada'].dt.year
```

6.3 Remoção de duplicidades

```
# Verifica linhas duplicadas
df.duplicated().sum()
```

```
# Remove duplicidades, se necessário
df = df.drop_duplicates()
```

6.4 Pensando o CRCA Lab como uma lista de receitas

</> Python

```
receitas_mensais = [2500, 3200, 2800, 4000, 4500, 5000, 5200, 5800, 6100, 6500, 7000, 7500]

# Cada número representa uma receita mensal aproximada.
# A lista funciona como uma sequência:

# Janeiro  -> índice 0 -> 2500
# Fevereiro -> índice 1 -> 3200
# Março -> índice 2 -> 2800
```

6.5 Código completo testado no Google Colab

</> Python

```
# Análise simples de previsibilidade de receitas do CRCA Lab

receitas_mensais = [2500, 3200, 2800, 4000, 4500, 5000, 5200, 5800, 6100, 6500, 7000, 7500]

# Receita total do ano
receita_total = sum(receitas_mensais)

# Receita média mensal
media_mensal = receita_total / len(receitas_mensais)

# Separação por semestre
```

```
primeiro_semestre = receitas_mensais[:6]
segundo_semestre = receitas_mensais[6:]

# Soma por semestre
total_primeiro_semestre = sum(primeiro_semestre)
total_segundo_semestre = sum(segundo_semestre)

# Diferença entre os semestres
diferenca = total_segundo_semestre - total_primeiro_semestre

# Crescimento percentual
crescimento_percentual = (diferenca / total_primeiro_semestre) * 100

# Exibição dos resultados
print("Análise de Receitas do CRCA Lab")
print("--------------------------------")
print(f"Receita total do ano: R$ {receita_total:.2f}")
print(f"Receita média mensal: R$ {media_mensal:.2f}")
print(f"Receita do primeiro semestre: R$ {total_primeiro_semestre:.2f}")
print(f"Receita do segundo semestre: R$ {total_segundo_semestre:.2f}")
print(f"Diferença entre os semestres: R$ {diferenca:.2f}")
print(f"Crescimento percentual: {crescimento_percentual:.2f}%")

File "/tmp/ipykernel_1318/3891158863.py", line 1
Análise de Receitas do CRCA Lab ^
```

SyntaxError: invalid syntax

[ “Por favor, explique este erro:]

The error SyntaxError: invalid syntax in cell vD5R0PVBEeEJ occurs because the cell contains plain text and not valid Python code. The Colab environment is attempting to execute this text as Python code, which causes the syntax error. It looks like this cell might be a copy of the output from the previous code cell.

Remova o conteúdo de texto da última célula
Okay, I will remove the text content from cell vD5R0PVBEeEJ.
The content of cell vD5R0PVBEeEJ has been successfully removed.”

Calcular a projeção de receita para o próximo ano com base na média
Sure, I can calculate the revenue projection for the next year based on the average monthly revenue.
The projection of the annual revenue based on the average monthly revenue is **R$ 60100.00**.

Gemini 2.5 Flash

Assim, o CRCA Lab poderia começar a responder perguntas como:

- **Qual fonte gera mais receita?**
- **Existe dependência excessiva de edital?**
- **A receita própria está crescendo?**
- **Qual é a média mensal de entrada?**
- **Quanto precisamos vender para sustentar o laboratório?**

Esse é o ponto em que Python deixa de ser apenas “linguagem de programação” e vira uma ferramenta de **gestão, planejamento e credibilidade institucional**, de fato. Com listas em Python, começamos a transformar registros financeiros simples em evidências de previsibilidade, crescimento e sustentabilidade para o CRCA Lab.

6.6 Cadastro de parceiros do CRCA Lab

</> Python

```
parceiros = {
    "IBiomas": "Parceiro institucional",
    "Petrobras": "Financiador / apoiador",
    "CRCA": "Organização responsável",
    "MOC 4.0": "Apoio em estratégia digital"
}

print(parceiros)
```

</> Python

```
parceiros = {
    "IBiomas": {
        "tipo": "Parceiro técnico",
        "area": "Agroecologia e território",
        "status": "ativo"
    },
    "Petrobras": {
        "tipo": "Financiador",
        "area": "Projeto socioambiental",
        "status": "ativo"
    },
    "MOC 4.0": {
        "tipo": "Consultoria estratégica",
        "area": "Comunicação, dados e tecnologia",
        "status": "em articulação"
    }
}
```

6.7 Controle de clientes do CRCA Lab

</> Python

```
clientes = {
    "OSC Parceira A": {
        "servico": "Diagnóstico de comunicação",
        "valor_previsto": 1500,
        "status": "proposta enviada"
    },
    "Projeto Social B": {
        "servico": "Produção de conteúdo e relatório de impacto",
        "valor_previsto": 2500,
        "status": "em negociação"
    },
    "Instituto C": {
        "servico": "Curso de formação em projetos sociais",
        "valor_previsto": 5000,
        "status": "reunião agendada"
    }
}
```

6.8 Controle de entregas do CRCA Lab

```
# Criando um dicionário para entregas
```

</> Python

```
entregas = {
    "Relatório de Impacto": {
        "cliente": "OSC Parceira A",
        "responsavel": "Juliano",
        "prazo": "30/06/2026",
        "status": "em andamento"
    },
    "Podcast Institucional": {
        "cliente": "Projeto Social B",
        "responsavel": "Equipe CRCA Lab",
        "prazo": "15/07/2026",
        "status": "planejado"
    },
    "Diagnóstico Digital": {
        "cliente": "Instituto C",
        "responsavel": "MOC 4.0",
        "prazo": "10/07/2026",
        "status": "não iniciado"
    }
}
```

6.9 “Minissistema” (painel simples) do CRCA Lab no Google Colab

</> Python

```
# Controle simples do CRCA Lab com dicionários

clientes = {
    "OSC Parceira A": {
        "servico": "Diagnóstico de comunicação",
        "valor_previsto": 1500,
        "status": "proposta enviada"
    },
    "Projeto Social B": {
        "servico": "Produção de conteúdo e relatório de impacto",
        "valor_previsto": 2500,
        "status": "em negociação"
    },
    "Instituto C": {
        "servico": "Curso de formação em projetos sociais",
        "valor_previsto": 5000,
        "status": "reunião agendada"
    }
}

entregas = {
    "Relatório de Impacto": {
        "cliente": "OSC Parceira A",
        "responsavel": "Juliano",
        "prazo": "30/06/2026",
        "status": "em andamento"
    },
    "Podcast Institucional": {
        "cliente": "Projeto Social B",
```

```
        "responsavel": "Equipe CRCA Lab",
        "prazo": "15/07/2026",
        "status": "planejado"
    },
    "Diagnóstico Digital": {
        "cliente": "Instituto C",
        "responsavel": "MOC 4.0",
        "prazo": "10/07/2026",
        "status": "não iniciado"
    }
}

# Cálculo da receita prevista

total_previsto = 0

for cliente in clientes:
    total_previsto = total_previsto + clientes[cliente]["valor_previsto"]

print("Painel Inicial do CRCA Lab")
print("--------------------------")
print(f"Total de clientes cadastrados: {len(clientes)}")
print(f"Total de entregas cadastradas: {len(entregas)}")
print(f"Receita prevista: R$ {total_previsto:.2f}")
```

6.10 Lista organizada de entregas do CRCA Lab

```
</> Python

print("Entregas do CRCA Lab")
print("--------------------")

for nome_entrega in entregas:
    cliente = entregas[nome_entrega]["cliente"]
    responsavel = entregas[nome_entrega]["responsavel"]
    prazo = entregas[nome_entrega]["prazo"]
    status = entregas[nome_entrega]["status"]

    print(f"Entrega: {nome_entrega}")
    print(f"Cliente: {cliente}")
    print(f"Responsável: {responsavel}")
    print(f"Prazo: {prazo}")
    print(f"Status: {status}")
    print("--------------------")

#Isso geraria uma saída parecida com:

                                                                    Entrega: Relatório de Impacto
                                                                         Cliente: OSC Parceira A
                                                                          Responsável: Juliano
                                                                             Prazo: 30/06/2026
                                                                           Status: em andamento
--------------------------------------------------------------------------------------------------

                                                                    Entrega: Podcast Institucional
                                                                         Cliente: Projeto Social B
                                                                    Responsável: Equipe CRCA Lab
                                                                             Prazo: 15/07/2026
                                                                           Status: planejado

--------------------------------------------------------------------------------------------------

# Com dicionários assim, o CRCA Lab pode começar a organizar uma base mínima de gestão:
```

```
| Área         | Pergunta que o Python pode responder            |
| ------------ | ----------------------------------------------- |
| Parceiros    | Quem são os parceiros ativos?                   |
| Clientes     | Quem está em negociação?                        |
| Receita      | Quanto temos de receita prevista?               |
| Entregas     | Quais entregas estão atrasadas ou em andamento? |
| Responsáveis | Quem está cuidando de cada frente?              |
| Projetos     | Qual cliente está ligado a qual serviço?        |
```

Com dicionários em Python, o CRCA Lab pode transformar informações dispersas sobre parceiros, clientes e entregas em uma estrutura organizada de gestão, permitindo consultas rápidas, cálculo de receitas previstas e acompanhamento das responsabilidades institucionais.

6.11 Estruturas do CRCA Lab para organização e decisão em Python

```
</> Markdown

# Antes de avançar para o tratamento de dados com bibliotecas como o Pandas, é importante compreender algumas estruturas fundamentais da linguagem Python: condicionais, laços de repetição e funções.

# No contexto do CRCA Lab, essas estruturas permitem transformar uma base inicial de conteúdos, públicos, canais, custos, alcance, engajamento e potencial de receita em uma ferramenta simples de apoio à decisão.

# As condicionais ajudam a classificar conteúdos conforme critérios definidos; os laços de repetição permitem percorrer vários registros automaticamente; e as funções organizam cálculos que podem ser reutilizados em diferentes partes da análise.

# Se o conteúdo tem alto impacto social e bom potencial de receita, ele deve ser priorizado.


</> Python

conteudo = "Podcast sobre agroecologia"
impacto_social = 9
potencial_receita = 8

if impacto_social >= 8 and potencial_receita >= 7:
    prioridade = "alta"
else:
    prioridade = "normal"

print(f"Conteúdo: {conteudo}")
print(f"Prioridade: {prioridade}")


</> Markdown

# Neste exemplo, o Python avalia dois critérios: impacto social e potencial de receita.
Caso ambos estejam em nível elevado, o conteúdo recebe prioridade alta. Esse tipo de regra ajuda o CRCA Lab a tomar decisões editoriais com base em dados, e não apenas em percepção subjetiva.

# Esse trecho conversa diretamente com o score_prioridade, mencionado no item 4 do sumário.


</> Python

conteudos = ["Podcast agroecologia", "Curso de projetos sociais", "Relatório de impacto"]
```

```
for item in conteudos:
    print(f"Analisando conteúdo: {item}")


</> Markdown

# O laço de repetição permite percorrer uma lista de conteúdos do CRCA Lab.

# Em vez de escrever uma linha de código para cada item, o Python executa a mesma ação automaticamente para todos os registros.



# Essa lógica prepara o terreno para quando o CRCA Lab tiver uma base maior com conteúdos, formatos, públicos, canais e indicadores.

</> Python

conteudos = {
    "Podcast agroecologia": 9,
    "Post institucional": 5,
    "Curso de projetos sociais": 8,
    "Vídeo curto para redes": 7
}

for nome, impacto in conteudos.items():
    if impacto >= 8:
        print(f"{nome} possui alto impacto social.")


</> Markdown

# Neste exemplo, combinamos repetição e decisão.

# O Python percorre todos os conteúdos cadastrados e exibe apenas aqueles com impacto social igual ou superior a 8.

# Essa combinação é útil para filtrar conteúdos prioritários dentro da estratégia do CRCA Lab.
```

No projeto aplicado do CRCA Lab, as estruturas fundamentais do Python — condicionais, laços de repetição e funções — são utilizadas como mecanismos de organização e decisão. Elas permitem classificar conteúdos, percorrer registros, aplicar critérios de priorização e reutilizar fórmulas de cálculo, preparando a base de dados para análises mais avançadas com Pandas e visualizações no Google Colab.

6.12 Lógica de analytics

```
</> Python

import pandas as pd

dados_crca_lab = {
    "conteudo": [
        "Podcast sobre agroecologia",
        "Curso de projetos sociais",
        "Relatório de impacto",
        "Vídeo curto institucional",
        "Ebook sobre captação de recursos"
    ],
    "formato": [
```

```
        "podcast",
        "curso",
        "relatório",
        "vídeo",
        "ebook"
    ],
    "publico": [
        "parceiros institucionais",
        "OSCs",
        "financiadores",
        "comunidade",
        "gestores de OSCs"
    ],
    "canal": [
        "YouTube",
        "Plataforma EAD",
        "Site institucional",
        "Instagram",
        "Landing page"
    ],
    "custo_estimado": [
        800,
        2500,
        1200,
        300,
        700
    ],
    "alcance_estimado": [
        3000,
        500,
        800,
        5000,
        1500
    ],
    "engajamento": [
        7,
        8,
        6,
        9,
        7
    ],
    "impacto_social": [
        9,
        8,
        10,
        7,
        8
    ],
    "potencial_receita": [
        6,
        9,
        7,
        5,
        8
    ]
}

df = pd.DataFrame(dados_crca_lab)
```

6.13 O papel do RStudio na solução analítica do CRCA Lab

O RStudio é o ambiente de desenvolvimento no qual a equipe pode organizar scripts em R, importar dados, validar a qualidade da base, realizar análises estatísticas, produzir visualizações e transformar resultados em produtos digitais.

Ele não substitui o banco de dados nem o sistema de acesso dos usuários: funciona como a camada analítica que conecta os registros dos projetos às decisões do CRCA Lab.

No modelo proposto, Python e R não competem. O Google Colab pode continuar sendo utilizado para prototipagem, aprendizagem e testes rápidos, enquanto o RStudio assume tarefas em que a rastreabilidade, a análise estatística, a documentação reproduzível e a publicação de painéis web sejam especialmente relevantes.

Capacidade no RStudio	Aplicação no CRCA Lab	Resultado esperado
Conexão e importação	Ler CSV, Excel, formulários e bancos relacionais.	Base integrada por projeto e período.
Tratamento com tidyverse	Padronizar públicos, territórios, atividades, custos e indicadores.	Dados consistentes e auditáveis.
Estatística e modelagem	Comparar grupos, testar relações e estimar tendências.	Hipóteses mais claras para decisão.
Visualização	Produzir gráficos, mapas e matrizes com ggplot2, Plotly ou Leaflet.	Leitura executiva dos resultados.
Relatórios com Quarto	Gerar Word, PDF, HTML e apresentações usando os mesmos dados e códigos.	Prestação de contas reproduzível.
Aplicações com Shiny	Criar painéis acessíveis pelo navegador para diferentes perfis internos.	Acesso às equipes dos projetos.

6.14 Arquitetura proposta: CRCA Conecta + RStudio

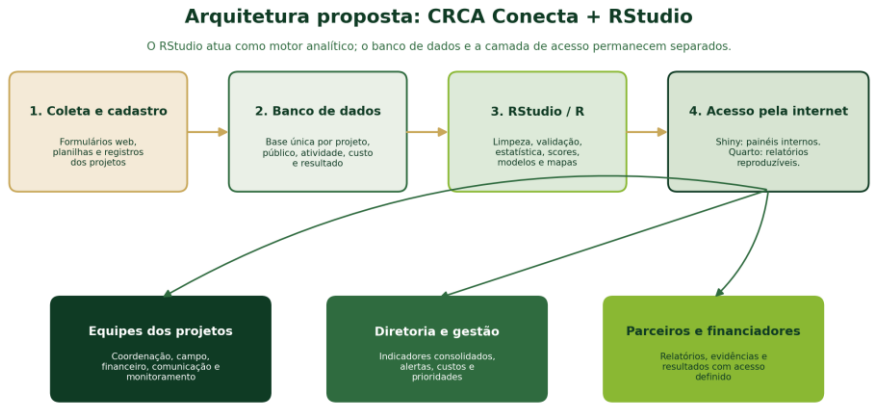


Figura – Fluxo proposto para coleta, análise e disponibilização dos dados do CRCA Lab.

Na primeira fase, o acesso será restrito às equipes dos diferentes projetos do CRCA. Cada perfil visualizará apenas os módulos e projetos necessários ao seu trabalho. A abertura para beneficiários poderá ocorrer posteriormente, depois da validação dos campos, das permissões, da qualidade dos dados e dos procedimentos de proteção de informações pessoais.

A arquitetura separa quatro responsabilidades: coleta, armazenamento, análise e acesso. Essa separação reduz a dependência de uma única ferramenta e permite que o CRCA comece com uma base simples e evolua, posteriormente, para um banco de dados institucional.

6.15 Exemplo de preparação dos indicadores em R

O script abaixo reproduz a lógica do score de prioridade e acrescenta uma etapa de normalização. A normalização é recomendada porque impacto, receita, engajamento e prioridade institucional podem ser medidos em escalas diferentes.

```
library(tidyverse)
library(scales)

ranking_r <- read_csv("base_crca_lab.csv") |>
  mutate(
    taxa_engajamento = engajamento_estimado / alcance_estimado,
    eficiencia_custo = alcance_estimado / custo_estimado
  ) |>
  mutate(
    impacto_n = rescale(impacto_social),
    prioridade_n = rescale(prioridade_institucional),
    receita_n = rescale(potencial_receita),
    engajamento_n = rescale(taxa_engajamento)
  ) |>
  mutate(
    score_prioridade =
      impacto_n * 0.35 +
      prioridade_n * 0.30 +
      receita_n * 0.20 +
      engajamento_n * 0.15
  ) |>
  arrange(desc(score_prioridade))
```

A saída desse processamento será uma tabela ordenada de conteúdos, públicos, canais ou projetos. Os pesos permanecem transparentes e podem ser alterados em reuniões de validação com a Diretoria e as coordenações.

7. Construção dos Indicadores

A etapa de indicadores transforma a base de dados em variáveis úteis para a tomada de decisão. O objetivo não é apenas descrever a produção, mas orientar escolhas estratégicas. Seguem os indicadores principais do CRCA Lab.:

Indicador	Fórmula / lógica	Uso na decisão
Taxa de engajamento	engajamento_estimado / alcance_estimado	Avaliar relevância do conteúdo para o público.
Eficiência de custo	alcance_estimado / custo_estimado	Comparar retorno comunicacional por real investido.
Score de prioridade	média ponderada de impacto, prioridade, receita e engajamento	Definir o que entra primeiro na agenda de produção.
Potencial de receita	escala 1 a 5	Avaliar contribuição para sustentabilidade.
Impacto social	escala 1 a 5	Manter coerência com missão e território.

```
# Taxa de engajamento
df['taxa_engajamento'] = df['engajamento_estimado'] / df['alcance_estimado']

# Eficiência de custo
df['eficiencia_custo'] = df['alcance_estimado'] / df['custo_estimado']

# Tratamento de divisões inválidas
df = df.replace([float('inf'), -float('inf')], 0)
df = df.fillna(0)
```

7.1 Score de prioridade

</> Python

```
df['score_prioridade'] = (
df['impacto_social'] * 0.35 +
df['prioridade_institucional'] * 0.30 +
df['potencial_receita'] * 0.20 +
df['taxa_engajamento'] * 10 * 0.15
)

ranking = df.sort_values('score_prioridade', ascending=False)
ranking.head(10)
```

7.2 Interpretação do score

[Inserir aqui a interpretação dos resultados obtidos no Colab. Explicar quais conteúdos aparecem no topo do ranking e por quê.]

7.3 Métricas recomendadas para o projeto

 # Proposta de divisão dos indicadores em quatro grupos:

7.3.1 Indicadores de produção

- Número de podcasts gravados;
- Número de vídeos curtos editados;
- Número de publicações técnicas;
- Número de cursos ou módulos produzidos;
- Número de entrevistas com agentes, técnicos, parceiros e produtores.

7.3.2 Indicadores de alcance

- Visualizações;
- Impressões;
- Cliques;
- Visitantes no site;
- Pessoas alcançadas nas redes.

7.3.3 Indicadores de engajamento

- Comentários;
- Compartilhamentos;
- Taxa de retenção em vídeos;
- Inscrições em cursos;
- Downloads de materiais;
- Taxa de engajamento.

7.3.4 Indicadores de impacto social

- Impacto social estimado
- Impacto por real investido
- Número de públicos vulneráveis alcançados
- Número de territórios contemplados
- Número de conteúdos vinculados a projetos sociais estratégicos

7.3.5 Indicadores de sustentabilidade

- Receita por serviço;
- Receita por curso;
- Receita por assinatura;
- Captação por edital;
- Custo por conteúdo;
- Retorno institucional por canal.

Esses indicadores dialogam com a própria minuta institucional do CRCA Lab, que já propõe metas como **40 podcasts/ano, 10 publicações/ano, 300 pessoas capacitadas, 15 OSCs atendidas e 50 mil pessoas alcançadas nas redes**.[5]

7.4 Função para calcular as prioridades do CRCA Lab

 # "Fórmula" para calcular o *score*:

</> Python

```
def calcular_score_prioridade(impacto_social, potencial_receita, baixo_custo):
    score = (impacto_social * 0.4) + (potencial_receita * 0.4) + (baixo_custo * 0.2)
    return score
```

</> Python

```
score = calcular_score_prioridade(9, 8, 7)

print(f"Score de prioridade: {score:.2f}")
```

</> Markdown

```
# A função `calcular_score_prioridade` organiza a fórmula de cálculo do indicador.

# Ela recebe três informações: impacto social, potencial de receita e baixo custo.

# A partir desses critérios, retorna uma pontuação final que pode apoiar a priorização dos conteúdos do CRCA Lab.[6]
```

8. Análises Exploratórias | CRCA Lab

Comparação entre canais, formatos e públicos. Uso de estatísticas descritivas para compreender custos, alcance, engajamento, impacto social e potencial de receita. Construção de rankings de conteúdos prioritários e identificação de oportunidades estratégicas. Um vídeo curto no Instagram tem alcance acima da média do canal, baixo custo e bom impacto social? Então, ele pode ser considerado uma oportunidade estratégica. Ou, um curso online tem alto potencial de receita, mas custo elevado e alcance baixo? Então talvez precise de revisão de estratégia comercial antes de ser priorizado.

[5] ANEXO III – CRCA_Lab_Estrutura_Estatuto...

[6] Essa parte é muito importante, porque o material explica que funções permitem encapsular uma lógica e reutilizá-la várias vezes com diferentes conjuntos de dados, aumentando produtividade e reduzindo erros.

</> Markdown

```
# No projeto aplicado do CRCA Lab, essa lógica será utilizada para comparar conteúdos, produtos e entregas com a média de seus canais, formatos ou públicos.

# Ela recebe três informações: impacto social, potencial de receita e baixo custo.

# Dessa forma, torna-se possível identificar quais conteúdos apresentam desempenho acima da média, quais possuem alto potencial com baixo custo e quais devem ser priorizados em razão de sua combinação entre alcance, engajamento, impacto social e potencial de receita. A análise deixa de ser apenas descritiva e passa a cumprir uma função estratégica: orientar decisões editoriais, institucionais e financeiras.
```

8.1 Identificação de categorias estratégicas

- Canais
- Formatos
- Públicos
- Territórios
- Linhas editoriais
- Fontes de receita

Exemplos:

</> Python

```
df["canal"].unique()
```

Ou:

```
df["formato"].unique()
```

</> Markdown

```
# O CRCA Lab pode identificar categorias estratégicas únicas, como canais de distribuição, formatos de conteúdo e públicos prioritários. Essa etapa permite compreender a diversidade da estratégia e verificar se há concentração excessiva em determinados canais ou públicos.
```

8.2 Conteúdos prioritários

</> Python

```
ranking[['conteudo', 'publico', 'canal', 'formato', 'score_prioridade']].head(10)
```

[Inserir interpretação dos conteúdos mais promissores.]

```
</> Markdown

# A etapa de análise exploratória tem como objetivo compreender o comportamento inicial da base de dados do CRCA Lab. Essa fase utiliza recursos do Pandas para observar médias, valores mínimos e máximos, distribuição dos indicadores e diferenças entre conteúdos, canais e formatos.
```

Essa leitura inicial permite identificar quais ações apresentam maior alcance, quais formatos possuem melhor relação entre custo e impacto, quais conteúdos têm maior potencial de receita e quais canais merecem prioridade na estratégia institucional.

8.3 Dimensão da base de dados

- Quantos conteúdos, produtos ou entregas foram cadastrados?
- Quantos critérios de análise estamos usando?

```
</> Python

df.shape
```

```
</> Markdown

# A primeira etapa da análise exploratória consiste em verificar a dimensão da base de dados.

# No projeto do CRCA Lab, essa operação permite identificar quantos conteúdos, produtos ou entregas foram cadastrados e quantas variáveis estão disponíveis para análise. Essa leitura inicial ajuda a compreender a escala da base e a planejar as próximas etapas do tratamento dos dados.
```

8.4 Alcance por canal

```
</> Python

alcance_por_canal = df.groupby('canal')['alcance_estimado'].sum().sort_values(ascending=False)
alcance_por_canal
```

[Inserir análise dos canais com maior potencial de alcance.]

```
</> Markdown

# A função `calcular_score_prioridade` organiza a fórmula central do modelo analítico. Ela recebe quatro variáveis: impacto social, prioridade institucional, potencial de receita e taxa de engajamento. A partir dos pesos definidos no
```

projeto, a função retorna uma pontuação final que permite comparar conteúdos, canais e formatos de maneira mais objetiva.

Essa lógica evita que a priorização da agenda do CRCA Lab dependa apenas de percepção subjetiva. O score não substitui a análise institucional, mas oferece uma camada inicial de evidência para orientar decisões editoriais, financeiras e estratégicas.

8.4.1 Análise e distribuição dos conteúdos por canal

Podemos identificar canais com maior concentração de conteúdos da seguinte forma no Python:

</> Python

```
df["canal"].value_counts()
```

</> Markdown

 # Esse comando mostra quantos conteúdos estão associados a cada canal. A análise permite identificar se o CRCA Lab está concentrando sua estratégia em poucos canais ou se há diversidade suficiente para alcançar diferentes públicos.

8.4.2 Comparação entre canais e alcance médio por canal

</> Python

```
df.groupby("canal")["alcance_estimado"].mean()
```

</> Markdown

 # Esse comando calcula o alcance médio estimado por canal. Com isso, o CRCA Lab pode comparar, por exemplo, se Instagram, YouTube, site institucional ou plataforma EAD apresentam maior potencial de distribuição.

8.5 Engajamento por formato

</> Python

```
engajamento_por_formato = df.groupby('formato')['engajamento_estimado'].sum().sort_values(ascending=False)
engajamento_por_formato
```

[Inserir análise dos possíveis formatos com maior engajamento.]

8.6 Estatísticas descritivas por Python

- Custo médio
- Alcance médio
- Engajamento médio
- Impacto social (médio)
- Potencial de receita médio
- Maior custo
- Maior alcance
- Maior impacto

</> Python

```
df.describe()
```

Ou:

```
df["custo_estimado"].mean()

df["alcance_estimado"].max()

df["impacto_social"].mean()

df.groupby("formato")["potencial_receita"].mean()
```

</> Markdown

```
# As estatísticas descritivas permitem compreender o comportamento geral da base do CRCA Lab. Por meio de médias, valores mínimos e máximos, torna-se possível observar o custo médio das ações, o alcance estimado, o nível de engajamento, o impacto social e o potencial de receita dos conteúdos cadastrados.
```

8.7 Potencial de receita por público

</> Python

```
receita_por_publico = df.groupby('publico')['potencial_receita'].mean().sort_values(ascending=False)
receita_por_publico
```

[Inserir análise dos públicos com maior potencial de sustentabilidade financeira.]

8.8 Indicadores derivados

No CRCA Lab, podemos criar indicadores derivados equivalentes. Por exemplo:

a) Impacto por real investido

- Quais conteúdos geram mais impacto social em relação ao custo?

</> Python

```
df["impacto_por_real"] = df["impacto_social"] / df["custo_estimado"]
```

b) Alcance por real investido

- Quais ações têm melhor relação entre custo e alcance?

</> Python

```
df["alcance_por_real"] = df["alcance_estimado"] / df["custo_estimado"]
```

c) Receita potencial por real investido

- Quais produtos parecem mais sustentáveis financeiramente em relação ao custo?

</> Python

```
df["receita_por_real"] = df["potencial_receita"] / df["custo_estimado"]
```

d) Média de impacto por real e por formato

- ~~Quais produtos parecem mais sustentáveis financeiramente em relação ao custo?~~

</> Python

```
media_por_formato = df.groupby("formato")["impacto_por_real"].transform("mean")
```

e) Classificar conteúdos acima ou abaixo da média do formato

- ~~Quais produtos parecem mais sustentáveis financeiramente em relação ao custo?~~

</> Python

```
df["classificacao_oportunidade"] = df["diferenca_media_formato"].apply(
                lambda x: "acima da média" if x > 0 else "abaixo da média"
                                                                        )

df[["conteudo", "formato", "impacto_por_real", "diferenca_media_formato", "classificacao_oportunidade"]]
```

</> Markdown

```
# Primeiro criamos um indicador de impacto por real. Depois calculamos a média desse indicador por formato. Em seguida
```

comparamos cada conteúdo com a média do seu grupo. E, por fim, classificamos se ele está acima ou abaixo da média, de acordo com os cálculos feitos pela IA.

</> Markdown

```
# Ranking por impacto e por real investido pode ser traduzido assim:
```

</> Python

```
df["alcance_por_real"] = df["alcance_estimado"] / df["custo_estimado"]
```

O projeto do CRCA Lab propõe a criação de indicadores derivados. Esses indicadores permitem comparar conteúdos e produtos diferentes de forma mais justa, considerando sua relação entre custo, impacto, alcance e potencial de receita. o projeto aplicado do CRCA Lab propõe uma análise por canais, formatos, públicos, custos, impacto social e potencial de receita, permitindo transformar dados institucionais em critérios objetivos de decisão.

Comentado [MC2]: Rever parágrafo



Entre os indicadores possíveis, destacam-se o impacto por real investido, o alcance por real investido e o potencial de receita por real investido. Esses cálculos ajudam a identificar quais ações apresentam melhor relação entre esforço, custo e retorno estratégico.

8.9 Comparação entre canais, formatos e públicos[7]

Para obter dados sobre o impacto social médio por formato:

</> Python

```
df.groupby("formato")["impacto_social"].mean()
```

</> Markdown

```
# Comparação entre canais, formatos, públicos e tipos de entrega. Análise de distribuição, médias, concentração, diferenciais estratégicos e relações entre variáveis, como custo, alcance, engajamento, impacto social e potencial de receita. Construção de rankings de conteúdos prioritários e identificação de oportunidades estratégicas.
```

[7] Comparação entre canais, formatos, públicos e tipos de entrega. Análise de distribuição, médias, concentração, diferenciais estratégicos e relações entre variáveis, como custo, alcance, engajamento, impacto social e potencial de receita. Quando houver dados territoriais disponíveis, poderão ser testadas correlações entre indicadores socioeconômicos e resultados do CRCA Lab.

```
# A análise socioeconômica apresentada amplia a compreensão da análise exploratória ao demonstrar que os resultados observados em uma base de dados podem ser influenciados por variáveis externas de contexto.


# No projeto aplicado do CRCA Lab, essa lógica pode ser adaptada para compreender se características territoriais, sociais ou institucionais dos públicos atendidos influenciam indicadores como alcance, engajamento, impacto social e potencial de receita. Assim, a análise deixa de considerar apenas os conteúdos isoladamente e passa a observar também o contexto em que eles circulam.
```

</> Python

```
df[["territorio", "indice_vulnerabilidade", "alcance_estimado", "impacto_social"]]
```

- Territórios com maior vulnerabilidade social geram maior impacto social percebido?
- Conteúdos voltados a públicos institucionais geram maior potencial de receita do que conteúdos voltados diretamente à comunidade?

</> Python

```
correlacao = df["indice_vulnerabilidade"].corr(df["impacto_social"])

print(f"Correlação entre vulnerabilidade e impacto social: {correlacao:.2f}")
```

</> Markdown

```
# Esse código mede se existe relação entre o índice de vulnerabilidade do território e o impacto social estimado do conteúdo. O resultado varia entre -1 e 1. Quanto mais próximo de 1, maior a tendência de as duas variáveis crescerem juntas. Quanto mais próximo de -1, maior a tendência de uma crescer enquanto a outra diminui. Valores próximos de 0 indicam pouca ou nenhuma relação linear.
```

8.10 Levantamento de hipóteses para priorização

Essa abordagem permite identificar padrões relevantes, como canais com maior alcance médio, formatos com maior potencial de sustentabilidade, conteúdos com melhor relação entre custo e impacto e públicos mais adequados para cada tipo de entrega. Com isso, a análise deixa de ser apenas descritiva e passa a apoiar decisões editoriais, institucionais e financeiras.

Podemos até mesmo criar um diferencial estratégico no CRCA Lab através de uma variável equivalente:

</> Python

```
tem_produto_digital
```

Ou:

```
tem_potencial_comercial
```

Exemplo:

```
df.groupby("tem_potencial_comercial")["potencial_receita"].mean()
```

</> Markdown

```
# Essa análise compara o potencial médio de receita entre conteúdos com e sem potencial comercial. Ela ajuda a verificar se os produtos mais voltados à venda, como cursos, ebooks e consultorias, apresentam maior capacidade de sustentabilidade financeira.
```

9. Visualizações no Google Colab e no RStudio

Criação de gráficos e painéis a partir das bases do CRCA Lab, permitindo visualizar rankings de conteúdos, comparação entre canais, custos estimados, potencial de receita e impacto social. No Google Colab, a proposta permanece voltada ao protótipo em Python. No RStudio, os mesmos dados podem gerar análises estatísticas, gráficos com ggplot2, relatórios em Quarto e aplicações web em Shiny. As visualizações apoiarão a comunicação dos resultados para equipes, gestores, parceiros e possíveis investidores.

O Colab pode ser estruturado como um notebook com este roteiro:

Notebook: CRCA_Lab_Analytics_MVP.ipynb

Bloco 1 — Introdução do projeto

</> Python

```
# Projeto: CRCA Lab Analytics
# Objetivo: criar um modelo simples de apoio à decisão para priorizar conteúdos, públicos, canais e fontes de receita do CRCA Lab.
```

Bloco 2 — Importação das bibliotecas

```
# Usa exatamente as funcionalidades do resumo do curso: pandas, numpy, matplotlib, tratamento de texto, datas, agrupamentos e gráficos.
```

```
</> Python

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from pathlib import Path
import unicodedata
```

Bloco 3 — Criação da base de dados

```
# Como ainda não existe uma base histórica completa, a criação de uma base simulada, mas realista como essa, com
dados do CRCA Lab, é possível.

</> Python

dados = [
    {
        "conteudo": "Podcast com agricultor familiar",
        "tema": "Agroecologia",
        "publico": "Agricultores familiares",
        "canal": "YouTube",
        "formato": "Podcast",
        "custo_estimado": 850,
        "alcance_estimado": 2500,
        "engajamento_estimado": 180,
        "potencial_receita": 2,
        "impacto_social": 5,
        "prioridade_institucional": 5
    },
    {
        "conteudo": "Curso online sobre economia solidária",
        "tema": "Formação",
        "publico": "OSCs e cooperativas",
        "canal": "Plataforma EAD",
        "formato": "Curso",
        "custo_estimado": 3200,
        "alcance_estimado": 900,
        "engajamento_estimado": 220,
```

```
        "potencial_receita": 5,
        "impacto_social": 4,
        "prioridade_institucional": 5
    },
    {
        "conteudo": "Relatório de impacto para financiadores",
        "tema": "Prestação de contas",
        "publico": "Financiadores",
        "canal": "Site institucional",
        "formato": "Relatório",
        "custo_estimado": 1800,
        "alcance_estimado": 400,
        "engajamento_estimado": 80,
        "potencial_receita": 4,
        "impacto_social": 4,
        "prioridade_institucional": 5
    },
    {
        "conteudo": "Série de vídeos sobre cozinhas comunitárias",
        "tema": "Segurança alimentar",
        "publico": "Comunidades beneficiárias",
        "canal": "Instagram",
        "formato": "Vídeo curto",
        "custo_estimado": 1200,
        "alcance_estimado": 5000,
        "engajamento_estimado": 420,
        "potencial_receita": 2,
        "impacto_social": 5,
        "prioridade_institucional": 4
    }
]

df = pd.DataFrame(dados)
```

As visualizações devem ser simples, legíveis e orientadas à decisão. A recomendação é utilizar gráficos de barras para rankings e comparações por canal, formato ou público.

Bloco 4 — Tratamento e padronização dos dados

```
# Domínio das funções do curso: strings, tipos, nulos, valores numéricos e normalização
```

</> Python

```
def remover_acentos(texto):
    texto = str(texto)
    texto = unicodedata.normalize('NFKD', texto)
    return ''.join([c for c in texto if not unicodedata.combining(c)])

df["tema_normalizado"] = df["tema"].apply(remover_acentos).str.lower().str.strip()
df["publico_normalizado"] = df["publico"].apply(remover_acentos).str.lower().str.strip()

colunas_numericas = [
    "custo_estimado",
    "alcance_estimado",
    "engajamento_estimado",
    "potencial_receita",
    "impacto_social",
    "prioridade_institucional"
]

for coluna in colunas_numericas:
    df[coluna] = pd.to_numeric(df[coluna], errors="coerce").fillna(0)

df.info()
```

Bloco 5 — Criação de indicadores

```
# Parte orientada à tomada de decisão em que o CRCA Lab possa usar uma pontuação ponderada para decidir o que
produzir primeiro.


</> Python

df["taxa_engajamento"] = df["engajamento_estimado"] / df["alcance_estimado"]

df["eficiencia_custo"] = df["alcance_estimado"] / df["custo_estimado"]

df["score_prioridade"] = (
    df["impacto_social"] * 0.35 +
    df["prioridade_institucional"] * 0.30 +
    df["potencial_receita"] * 0.20 +
    df["taxa_engajamento"] * 10 * 0.15
)

df.sort_values("score_prioridade", ascending=False)
```

No projeto aplicado do CRCA Lab, essa lógica será adaptada para transformar conteúdos, canais, públicos, custos e indicadores de impacto em uma base analítica capaz de orientar decisões editoriais, institucionais e financeiras.

9.1 Gráfico de ranking de prioridades dos conteúdos[8]

[8] O score_prioridade responde **"o que parece mais importante?"** A recomendacao responde **"o que fazer com isso?"**. O ranking responde **"qual vem primeiro?"**. E os indicadores_gerais respondem **"como resumir isso em cards para o website?"**

```
</> Python

import matplotlib.pyplot as plt

ranking_grafico = df.sort_values('score_prioridade', ascending=True)

plt.figure(figsize=(10, 6))
plt.barh(ranking_grafico['conteudo'], ranking_grafico['score_prioridade'])
plt.title('Ranking de Prioridade de Conteúdos - CRCA Lab')
plt.xlabel('Score de Prioridade')
plt.ylabel('Conteúdo')
plt.tight_layout()
plt.show()
```

9.2 Gráfico de alcance estimado por canal

```
</> Python

plt.figure(figsize=(8, 5))
alcance_por_canal.plot(kind='bar')
plt.title('Alcance Estimado por Canal')
plt.xlabel('Canal')
plt.ylabel('Alcance Estimado')
plt.tight_layout()
plt.show()
```

9.3 Gráfico de potencial de receita por formato

```
</> Python

receita_por_formato = df.groupby('formato')['potencial_receita'].mean().sort_values(ascending=False)

plt.figure(figsize=(8, 5))
receita_por_formato.plot(kind='bar')
plt.title('Potencial Médio de Receita por Formato')
plt.xlabel('Formato')
plt.ylabel('Potencial de Receita')
plt.tight_layout()
plt.show()
```

9.4 Leitura dos padrões

- Quantidade de conteúdos por formato
- Quantidade de conteúdos por canal
- Ranking de impacto social
- Ranking de potencial de receita
- Comparação entre custo e alcance
- Score de prioridade

</> Python[9]

```
df["formato"].value_counts().plot(kind="bar", title="Distribuição de conteúdos por formato")

Ou:

df.groupby("canal")["alcance_estimado"].mean().plot(kind="bar", title="Alcance médio por canal")
```

</> Markdown

```
# As visualizações foram criadas para facilitar a leitura dos padrões identificados na análise exploratória do CRCA Lab. Os gráficos representam a distribuição dos conteúdos por formato, a comparação entre canais, os rankings de impacto social, o potencial de receita, o custo estimado, a eficiência de custo e o score_prioridade. Essa leitura permite comparar canais, formatos, públicos e tipos de entrega, identificando relações entre custo, alcance, engajamento, impacto social e potencial de receita. Com isso, o modelo apoia a construção de rankings de conteúdos prioritários e a identificação de oportunidades estratégicas para decisões editoriais, institucionais e financeiras.
```

Comentado [MC3]: Juntar os dois parágrafos em somente 1

[9] O CRCA Lab não está usando Python apenas como exercício técnico, mas como ferramenta de inteligência institucional.

9.5 Dashboard web interno com Shiny

O Shiny permite transformar códigos e análises em uma aplicação web acessível pelo navegador. Para o CRCA Lab, essa aplicação poderá funcionar como a interface analítica do CRCA Conecta, sem exigir que coordenadores, técnicos de campo ou equipes administrativas utilizem diretamente o RStudio.

Perfil de acesso	Visão inicial no dashboard	Possíveis ações
Diretoria	Portfólio consolidado, gastos, metas, alertas e comparações.	Filtrar projetos e acompanhar prioridades.
Coordenação de projeto	Metas, atividades, público, cronograma, evidências e orçamento do projeto.	Validar registros e gerar relatórios.
Equipe de campo	Formulários, atividades atribuídas e pendências de evidências.	Cadastrar e corrigir informações autorizadas.
Financeiro/administrativo	Despesas, categorias, documentos e execução por projeto.	Conferir registros e exportar bases.
Comunicação	Conteúdos, canais, alcance, engajamento e banco de imagens.	Priorizar pautas e selecionar evidências.
Monitoramento e avaliação	Indicadores, qualidade dos dados, séries históricas e resultados.	Revisar métricas e produzir análises.

O controle de autenticação e de permissões deve ser implementado na camada de hospedagem e no banco de dados. Portanto, a proposta não é simplesmente "publicar um dashboard", mas disponibilizar uma aplicação interna com perfis, registro de acesso, regras de visualização e procedimentos compatíveis com a LGPD.

9.6 Estrutura mínima do protótipo em Shiny

```
library(shiny)
library(tidyverse)

ui <- fluidPage(
  titlePanel("CRCA Conecta – Painel Interno"),
  sidebarLayout(
    sidebarPanel(
      selectInput("projeto", "Projeto", choices = unique(base$projeto))
    ),
    mainPanel(
      textOutput("total_publico"),
      plotOutput("grafico_gastos"),
      dataTableOutput("ranking_prioridades")
    )
  )
)

server <- function(input, output, session) {
  dados_filtrados <- reactive(
    base |> filter(projeto == input$projeto)
  )
  # Indicadores e visualizações usam o projeto selecionado.
}

shinyApp(ui, server)
```

Esse código representa apenas a estrutura do MVP. O protótipo acadêmico pode utilizar dados simulados; a implantação institucional exigirá validação dos campos, autenticação, testes com usuários, política de backup e definição das responsabilidades por atualização e qualidade da base.

10. Uso de Inteligência Artificial no Projeto

A inteligência artificial (IA) pode ser apresentada como camada de apoio à classificação, síntese e recomendação. Para fins do projeto aplicado, começamos com um modelo heurístico simples, transparente e explicável, antes de avançar

para técnicas mais complexas. Por ser muito utilizada na categorização de temas, a IA facilmente soube identificar as palavras-chave. Pretendemos resumir as 40 primeiras entrevistas de forma semelhante com a IA, além de analisar as demandas recorrentes desse laboratório de audiovisual, esperamos dessa tecnologia recomendações para possíveis formatos inovadores.

10.1 IA como apoio à classificação de conteúdos

Primeiro, como **classificação de conteúdo**: transformar entrevistas, depoimentos e relatórios em categorias como agroecologia, economia solidária, formação, captação, segurança alimentar e prestação de contas.

- Classificar temas de entrevistas e depoimentos.
- Identificar palavras-chave recorrentes.
- Apoiar a criação de resumos executivos.
- Sugerir formatos de conteúdo a partir do público e do objetivo.
- Transformar dados e relatos em narrativas de impacto.

10.2 Modelo inicial de recomendação

Segundo, como **apoio à decisão editorial**: sugerir quais temas devem virar podcast, artigo, curso, relatório ou campanha.

</> Python

```
def recomendar_formato(row):
    if row['impacto_social'] >= 5 and row['alcance_estimado'] >= 3000:
        return 'Campanha pública'
    elif row['potencial_receita'] >= 5:
        return 'Curso ou produto educacional'
    elif row['publico'] == 'Financiadores':
        return 'Relatório de impacto'
    else:
        return 'Podcast ou videocast'

df['recomendacao_formato'] = df.apply(recomendar_formato, axis=1)

df[['conteudo', 'publico', 'formato', 'recomendacao_formato', 'score_prioridade']]
```

Terceiro, como **análise de texto**: usar resumos, palavras-chave, sentimento, recorrência de temas e identificação de demandas territoriais. Não é uma IA complexa. Podemos chamar de modelo heurístico de recomendação, uma primeira etapa antes de usar modelos mais avançados de linguagem.

10.3 Possíveis usos futuros de IA generativa

- Resumir entrevistas de campo.
- Gerar títulos e chamadas para redes sociais.
- Classificar depoimentos por tema.
- Identificar narrativas de impacto.

- Sugerir roteiros de podcast.
- Apoiar relatórios de prestação de contas.
- Transformar dados em textos executivos para financiadores.

10.4 Exploração da base do CRCA Lab

</> Python

```
# Verificar tamanho da base:
print("Dimensão da base:")
print(df.shape)

# Verificar formatos únicos:
print("Formatos cadastrados:")
print(df["formato"].unique())

# Verificar canais únicos:
print("Canais cadastrados:")
print(df["canal"].unique())

# Estatísticas descritivas:
print("Resumo estatístico:")
display(df.describe())

# Indicadores derivados[10]:
df["impacto_por_real"] = df["impacto_social"] / df["custo_estimado"]
df["alcance_por_real"] = df["alcance_estimado"] / df["custo_estimado"]
df["receita_por_real"] = df["potencial_receita"] / df["custo_estimado"]

# Visualizar tabela atualizada:
df
```

[10] Esse código cria um indicador chamado **impacto_por_real**, que mede quanto impacto social cada conteúdo gera em relação ao seu custo estimado. Em seguida, organiza os conteúdos em ordem decrescente, mostrando quais entregas oferecem melhor relação entre custo e impacto.

A exploração inicial de dados permite que o CRCA Lab compreenda a estrutura de sua base, identifique categorias estratégicas, calcule estatísticas descritivas e crie indicadores derivados capazes de orientar decisões sobre conteúdos, canais, públicos e sustentabilidade financeira. O CRCA Lab demonstra, enfim, como uma base real pode ser organizada em linhas e colunas, carregada no Google Colab com Pandas e explorada por meio de estatísticas iniciais, seleção de variáveis e apoio de inteligência artificial.

11. Resultados Esperados

Espera-se que o modelo permita:

- Identificar conteúdos prioritários para produção.
- Relacionar públicos a formatos mais adequados.
- Avaliar canais com maior potencial de alcance e engajamento.
- Apoiar a sustentabilidade financeira do CRCA Lab.
- Transformar o Canvas estratégico em um painel analítico de decisão.
- Demonstrar a aplicação prática de Python em um caso real de marketing analytics e impacto social.

Campo para resultados

[Inserir aqui os resultados obtidos após rodar o notebook no Google Colab: tabelas, gráficos, rankings e principais descobertas.]

</> Markdown

```
# Como resultado esperado, o projeto deverá gerar rankings de oportunidades para o CRCA Lab, indicando conteúdos, formatos e canais com maior potencial estratégico. Esses rankings serão construídos a partir de critérios objetivos, como custo estimado, alcance, engajamento, impacto social e potencial de receita.

# A proposta é reduzir a dispersão de decisões e apoiar a priorização de ações com maior capacidade de gerar impacto e sustentabilidade. O modelo analítico do CRCA Lab buscará filtrar múltiplas possibilidades de conteúdo e entrega para indicar aquelas mais relevantes para a estratégia institucional.
```

11.1 Filtros múltiplos para priorização de oportunidades (ranking de conteúdo)

- Quais conteúdos têm alto impacto social, bom potencial de receita, custo abaixo da média e bom engajamento?

</> Python

```
oportunidades = df[
    (df["impacto_social"] >= 8) &
    (df["potencial_receita"] >= 7) &
    (df["custo_estimado"] <= df["custo_estimado"].mean()) &
    (df["engajamento"] >= 7)
]

oportunidades
```

</> Python

```
ranking_oportunidades = df.sort_values(
    by="score_prioridade",
    ascending=False
)

ranking_oportunidades[["conteudo", "formato", "canal", "score_prioridade"]]
```

- Disponibilizar um protótipo de dashboard interno em Shiny, com filtros por projeto, público, período, território e tipo de indicador.
- Definir um OKR de implantação e uma Métrica Mais Importante do Momento para cada fase do MVP.
- Produzir um mapa de processo atual e futuro, com clientes internos, responsabilidades e regras de governança.
- Estimar, quando houver dados suficientes, modelos de custo, adoção e potencial comercial no RStudio.
- Distinguir claramente resultados descritivos, preditivos e causais nas recomendações executivas.
- Posicionar a spin-off como serviço de implantação de inteligência institucional para OSCs.

12. Discussão

Os resultados devem ser analisados não apenas do ponto de vista quantitativo, mas também estratégico. Um conteúdo com menor alcance pode ser prioritário caso dialogue com financiadores, fortaleça a prestação de contas ou contribua para a memória institucional. Da mesma forma, conteúdos de grande alcance podem ter menor potencial de receita direta, mas alto valor para reputação, mobilização e impacto público.

[Inserir aqui análise crítica sobre limites do modelo, qualidade dos dados, necessidade de base histórica, riscos de enviesamento e próximos passos.]

A inclusão do RStudio fortalece a dimensão metodológica do projeto porque permite manter código, dados e resultados em um fluxo reproduzível. Entretanto, a qualidade da análise continuará dependente da consistência dos registros, da definição dos indicadores e da validação humana. O dashboard não elimina a interpretação institucional e não deve transformar estimativas de impacto em afirmações causais.

12.1 Evolução conceitual do CRCA Lab

A principal contribuição da análise não está apenas na execução dos códigos em Python, mas na capacidade de formular perguntas relevantes, interpretar resultados e transformar dados em decisões. A experiência metodológica apresentada demonstra que problemas complexos devem ser divididos em etapas menores: formular perguntas claras, resolver cada análise individualmente, validar os resultados e interpretá-los com foco em decisão.

No CRCA Lab, essa lógica é especialmente importante, pois permite transformar um conjunto amplo de possibilidades — conteúdos, cursos, relatórios, podcasts, produtos digitais, canais e públicos — em uma matriz de decisão orientada por dados. Dessa forma, o laboratório passa a operar não apenas como espaço de produção, mas como uma unidade de inteligência institucional.

</> Python

```
capacidades_ia = [
    {
        "capacidade": "Redação institucional",
```

```
        "area_osc": "Comunicação",
        "uso_crca_lab": "Produção de textos, posts, releases e relatórios",
        "nivel_impacto": 4,
        "nivel_complexidade": 2,
        "prioridade": 5
    },
    {
        "capacidade": "Análise de dados simples",
        "area_osc": "Monitoramento e Avaliação",
        "uso_crca_lab": "Organização de indicadores de alcance, engajamento e impacto",
        "nivel_impacto": 5,
        "nivel_complexidade": 3,
        "prioridade": 5
    },
    {
        "capacidade": "Simulação de entrevistas",
        "area_osc": "Gestão de Pessoas e Voluntariado",
        "uso_crca_lab": "Treinamento de equipes, voluntários e facilitadores",
        "nivel_impacto": 3,
        "nivel_complexidade": 2,
        "prioridade": 3
    },
    {
        "capacidade": "Captação de recursos",
        "area_osc": "Sustentabilidade",
        "uso_crca_lab": "Apoio à escrita de projetos, propostas e campanhas",
        "nivel_impacto": 5,
        "nivel_complexidade": 4,
        "prioridade": 5
    },
    {
        "capacidade": "Classificação de conteúdos",
        "area_osc": "Editoração Social",
        "uso_crca_lab": "Organizar podcasts, cursos e publicações por tema e público",
        "nivel_impacto": 5,
        "nivel_complexidade": 3,
        "prioridade": 5
    }
]
```

</> Markdown

```
# Depois, isso vira um DataFrame :
```

</> Python

```
import pandas as pd

df_ia = pd.DataFrame(capacidades_ia)
df_ia
```

</> Markdown

```
# E podemos criar um indicador simples:
```

</> Python

```
df_ia["score_aplicacao"] = (
    df_ia["nivel_impacto"] * 0.5 +
    df_ia["prioridade"] * 0.3 -
    df_ia["nivel_complexidade"] * 0.2
)

df_ia.sort_values("score_aplicacao", ascending=False)
```

Esse código traduz uma pergunta estratégica: **quais capacidades de IA devem ser implantadas primeiro no CRCA Lab, considerando impacto, prioridade e complexidade?** A intenção aqui é criar uma metodologia para transformar IA generativa em inteligência institucional para OSCs, usando Python como camada de organização, mensuração e decisão. **OSCs precisam transformar essa conversa com IA em método, dados, indicadores e decisão**.

12.2 Matriz de capacidades generativas para OSCs/ONGs (Non-profit)

Dimensão	Aplicação no CRCA Lab	Tradução em Python
Comunicação	Posts, releases, roteiros, campanhas	classificação por canal e formato
Formação	Cursos, oficinas, materiais didáticos	análise de temas e públicos
Gestão de dados	Indicadores, tabelas, relatórios	Pandas, groupby, gráficos
Captação	Editais, propostas, relatórios de impacto	score de potencial de receita
Governança	atas, estatutos, prestação de contas	organização documental e indicadores

Se conteúdos de determinado público têm mais engajamento, isso não prova automaticamente que o público causa o engajamento. Pode haver outros fatores envolvidos, como canal usado, qualidade do conteúdo, tema, linguagem, momento de publicação ou investimento em divulgação.

</> Markdown

```
# Um cuidado metodológico importante é reconhecer que correlação não significa causalidade. Caso a análise identifique relação entre um determinado público, canal ou território e melhores resultados de engajamento, isso não significa necessariamente que uma variável cause diretamente a outra. A interpretação deve considerar fatores externos, como linguagem utilizada, investimento em divulgação, qualidade da produção, momento de publicação e aderência temática.
```

12.3 Aprofundamento em Business Analytics aplicado ao CRCA Lab

Recomendação: aprofundar, mas em camadas. Os conteúdos da disciplina são diretamente aplicáveis ao CRCA Lab porque o desafio atual não é apenas produzir gráficos. O desafio é criar uma rotina em que cada análise esteja ligada a uma pergunta de gestão, a um responsável, a uma decisão e a um ciclo de revisão. O avanço deve ocorrer da base para o topo: primeiro processos e governança; depois análise descritiva e diagnóstica; em seguida previsão; e somente com histórico suficiente, avaliação causal e prescrição.

12.3.1 Conteúdos prioritários e aplicação prática

Conteúdo da disciplina	Aplicação no CRCA Lab	Produto ou evidência
OKRs e times	Traduzir a implantação do CRCA Conecta em objetivos comuns e resultados-chave construídos com as equipes dos projetos.	OKRs trimestrais, check-ins e backlog priorizado.
Lean Analytics e MVP	Testar uma versão interna do sistema em dois projetos antes de expandir para todos os públicos.	MVP, hipótese, métrica principal e decisão de manter, alterar ou interromper.
Métricas acionáveis	Substituir volume de arquivos e postagens por indicadores que orientem ação: completude, prazo, custo, uso e resultado.	Dicionário de métricas e painel gerencial.
Clientes internos	Definir quais decisões cada perfil precisa tomar: Diretoria, coordenação, campo, financeiro, comunicação e monitoramento.	Matriz usuário-decisão-indicador.
BPR	Redesenhar o fluxo atual de dados, evidências, relatórios e prestação de contas para reduzir retrabalho e versões divergentes.	Mapa do processo atual e processo futuro.
Maturidade analítica	Avaliar se o CRCA está no estágio de planilhas isoladas, relatórios padronizados, alertas, previsão ou recomendação.	Diagnóstico e roteiro de evolução.
Data Mart / Data Warehouse	Consolidar gradualmente projetos, atividades, públicos, gastos, evidências e resultados em uma fonte comum.	Base institucional governada e histórico longitudinal.
Correlação e regressão	Estimar quais variáveis estão associadas a custos, atrasos, alcance, engajamento e receita.	Modelos em RStudio com coeficientes e intervalos de confiança.
Inferência causal	Avaliar se uma mudança de processo realmente produziu resultado, em vez de apenas acompanhar uma associação.	Teste A/B, implantação escalonada ou diferenças em diferenças.
Vieses decisórios	Reduzir decisões por confirmação, ancoragem, disponibilidade, excesso de confiança e manutenção do status quo.	Registro de decisão, hipóteses prévias e revisão pós-resultado.

12.3.2 Modelo analítico-operacional: do problema à aprendizagem

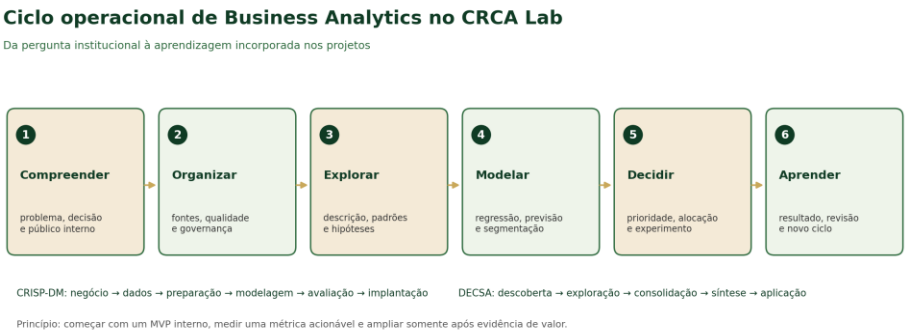


Figura 2 – Ciclo de Business Analytics proposto para o CRCA Lab, integrando CRISP-DM, DECSA e aprendizagem institucional.

Como usar o ciclo: cada demanda começa com uma decisão explícita — por exemplo, reduzir o tempo de fechamento de relatórios ou identificar quais serviços possuem maior potencial comercial. Somente depois são definidas fontes, variáveis, tratamento, modelo, critério de avaliação e forma de implantação. O resultado deve voltar ao processo por meio de check-ins, atualização do backlog e registro do aprendizado.

12.3.3 Pressão–Estado–Resposta–Suporte no contexto do CRCA Lab

Dimensão	Leitura para o CRCA Lab	Resposta proposta
Pressão	Projetos simultâneos, exigências de financiadores, prazos de prestação de contas e necessidade de demonstrar impacto.	Priorizar problemas decisórios com maior risco institucional ou potencial de geração de valor.
Estado	Dados dispersos entre planilhas, mensagens, arquivos e memórias individuais; indicadores heterogêneos entre projetos.	Diagnóstico das fontes, responsáveis, periodicidade, qualidade e nível de acesso.
Resposta	Criação do CRCA Conecta como base interna e do RStudio como ambiente de tratamento, modelagem e validação.	MVP em dois projetos, com formulários padronizados, rotina de atualização e painel.
Suporte	Governança, capacitação, documentação, patrocínio da Diretoria e regras de proteção de dados.	Comitê analítico leve, dicionário de dados, perfis de acesso e reuniões mensais de decisão.

12.3.4 OKRs para a implantação interna do CRCA Conecta

Objetivo trimestral: transformar os dados de dois projetos-piloto em uma fonte confiável e utilizada pela equipe para decidir, acompanhar e prestar contas.

- KR1 — alcançar pelo menos 90% de completude nos campos obrigatórios das atividades registradas.
- KR2 — reduzir em 40% o tempo médio entre a realização de uma ação e a validação de suas evidências.
- KR3 — garantir que 80% dos usuários convidados acessem o sistema e realizem ao menos uma atualização mensal.
- KR4 — produzir três decisões documentadas com base no painel, indicando dado observado, decisão tomada e resultado posterior.
- KR5 — entregar um relatório reproduzível em R/Quarto para cada projeto-piloto ao final do trimestre.

Regra de governança: o backlog do sistema deve derivar dos OKRs. Funcionalidades que não alterem um resultado-chave, não reduzam risco ou não gerem aprendizagem verificável devem aguardar.

12.3.5 Lean Analytics: MVP e Métrica Mais Importante do Momento

Fase	Hipótese principal	Métrica mais importante	Critério de decisão
1. Organização	As equipes conseguem registrar dados com um padrão comum.	Taxa de registros completos e válidos.	Se a completude permanecer baixa, simplificar campos e treinar usuários.
2. Adoção interna	A plataforma entra na rotina e reduz controles paralelos.	Retenção mensal de usuários internos ativos.	Se o uso cair, revisar utilidade, experiência e responsabilidades.
3. Eficiência	O novo fluxo reduz tempo e retrabalho.	Tempo entre ação, evidência validada e relatório.	Se não houver redução, redesenhar o processo antes de adicionar tecnologia.
4. Inteligência	Os dados geram decisões melhores.	Percentual de decisões gerenciais apoiadas por evidências registradas.	Se o painel não orientar decisões, rever indicadores e público interno.

Fase	Hipótese principal	Métrica mais importante	Critério de decisão
5. Spin-off	Outras OSCs pagariam pela implantação e acompanhamento.	Conversão de diagnóstico em piloto pago e margem de contribuição.	Se não houver demanda paga, revisar proposta de valor ou segmento.

12.3.6 Funil de adoção e métricas digitais adaptadas ao uso interno

Métrica digital	Tradução para o CRCA Conecta	Fórmula inicial
Aquisição	Pessoas das equipes convidadas para usar o sistema.	usuários convidados
Ativação	Usuários que acessam e concluem o primeiro registro válido.	ativados ÷ convidados
Engajamento	Usuários que consultam ou atualizam dados no período.	usuários ativos mensais
Retenção	Usuários que permanecem ativos nos meses seguintes.	ativos no mês t e t+1 ÷ ativos no mês t
Churn interno	Usuários que deixam de usar o sistema e retornam a controles paralelos.	usuários inativos ÷ usuários ativos anteriores
Reativação	Usuários que voltam após treinamento, melhoria ou nova rotina.	reativados ÷ inativos contatados
CAC comercial	Custo para conquistar uma OSC ou projeto como cliente da spin-off.	despesas comerciais ÷ novos clientes
Payback	Tempo necessário para recuperar o custo de aquisição e implantação.	CAC ÷ margem mensal por cliente

Cuidado com métricas de vaidade: quantidade de acessos, arquivos ou gráficos não comprova valor. O indicador precisa revelar mudança de comportamento, redução de custo, melhoria de qualidade, decisão tomada ou resultado social e financeiro.

12.3.7 Reengenharia do processo de dados e evidências

Etapa	Situação atual provável	Processo futuro proposto
Coleta	Formulários diferentes, mensagens e anotações locais.	Formulário padronizado por tipo de atividade, com identificador único.
Validação	Conferência tardia, dependente de memória e busca manual.	Regras automáticas, campos obrigatórios e responsável por validação.
Evidências	Fotos e documentos sem vínculo inequívoco com metas e atividades.	Repositório com metadados, projeto, meta, data, território e autorização.
Financeiro	Gastos agregados e dificuldade para relacionar custo a entrega.	Centro de custos, horas, deslocamento, terceiros e custo por produto.
Análise	Planilhas isoladas e indicadores calculados de formas distintas.	Scripts versionados em RStudio e dicionário único de indicadores.
Decisão	Reuniões baseadas em relatos e urgências do momento.	Painel, pauta analítica, decisão registrada e responsável pelo próximo passo.
Prestação de contas	Consolidação manual próxima ao prazo final.	Relatório reproduzível atualizado ao longo da execução.

12.3.8 Data Mart antes de Data Warehouse

Decisão recomendada: não começar por um Data Warehouse corporativo completo. Para a realidade atual, é mais seguro implantar um Data Mart do CRCA Lab — uma base temática e governada com os dados essenciais de projetos, atividades, públicos, custos, conteúdos, evidências e resultados. O Data Warehouse institucional pode ser uma etapa posterior, quando houver fontes estáveis, histórico e capacidade de manutenção.

Camada	Exemplos de conteúdo	Responsabilidade
Fatos	Atividades, gastos, entregas, acessos, interações, receitas e resultados.	Registros mensuráveis e datados.

Camada	Exemplos de conteúdo	Responsabilidade
Dimensões	Projeto, território, público, equipe, canal, formato, financiador e período.	Categorias comuns para comparação.
Metadados	Origem, responsável, data de atualização, regra de cálculo e nível de sensibilidade.	Rastreabilidade e governança.
Produtos analíticos	Painel, score, previsão, alerta, relatório e recomendação.	Uso decisório das informações.

12.3.9 Modelos preditivos no RStudio

Pergunta 1 — gastos: quais variáveis estão associadas ao custo total de uma entrega ou atividade? O modelo deve comparar efeitos individuais e coletivos, reproduzindo a lógica trabalhada pelo professor. A regressão simples mostra a associação isolada; a regressão múltipla estima cada efeito mantendo as demais variáveis constantes.

```
# Exemplo conceitual: dados simulados ou reais previamente anonimizados
modelo_custos <- lm(
  custo_total ~ horas_equipe + deslocamento_km + participantes +
    numero_revisoes + tipo_entrega + municipio,
  data = dados_crca
)

summary(modelo_custos)
broom::tidy(modelo_custos, conf.int = TRUE)
performance::check_model(modelo_custos)
```

Interpretação esperada: o coeficiente de cada variável indica a variação média esperada no custo quando aquela variável muda, mantendo as demais constantes. O R^2 mostra quanto da variação dos custos é explicada pelo conjunto de variáveis, mas não comprova causalidade.

Pergunta 2 — adoção interna: quais fatores aumentam a probabilidade de uma pessoa usar o CRCA Conecta de forma recorrente? Como a variável de resultado é binária — usuário ativo ou inativo — uma regressão logística pode ser mais adequada.

```
modelo_adocao <- glm(
  usuario_ativo ~ treinamento + tempo_resposta_sistema +
    clareza_responsabilidade + utilidade_percebida + projeto,
  data = usuarios_crca,
  family = binomial()
)

broom::tidy(modelo_adocao, exponentiate = TRUE, conf.int = TRUE)
```

Pergunta 3 — potencial comercial: quais perfis de organização, necessidades e tipos de entrega apresentam maior probabilidade de contratar um piloto ou gerar receita recorrente? A resposta depende de uma base comercial histórica. Antes disso, o score de potencial deve ser tratado como hipótese, não como previsão validada.

12.3.10 Associação, previsão e causalidade: três perguntas diferentes

Nível	Pergunta	Método possível	Limite
Descritivo	O que aconteceu e como os dados estão distribuídos?	Tabelas, médias, histogramas e painéis.	Não explica fatores nem prevê.
Diagnóstico	Quais variáveis se relacionam com o resultado?	Segmentação, correlação e regressão.	Associação não é causalidade.
Preditivo	Qual resultado é provável em novos casos?	Regressão, classificação e validação fora da amostra.	Prever bem não significa explicar causa.
Causal	O que mudaria se o CRCA adotasse uma intervenção?	Teste A/B, rollout escalonado, diferenças em diferenças ou pareamento.	Exige desenho, grupo comparável e histórico.

Nível	Pergunta	Método possível	Limite
Prescritivo	Qual ação deve ser priorizada sob restrições?	Regras, otimização e simulação de cenários.	Depende da qualidade dos modelos e dos critérios éticos.

12.3.11 Experimentos possíveis no CRCA Lab

Hipótese	Intervenção	Indicador principal	Desenho inicial
Formulário padronizado reduz retrabalho.	Um projeto usa novo formulário e outro mantém o processo anterior por período controlado.	Tempo de validação e número de correções.	Piloto comparativo ou implantação escalonada.
Treinamento aumenta adoção do sistema.	Grupo recebe onboarding guiado; grupo de comparação recebe apenas manual.	Ativação e retenção em 30 dias.	Teste A/B quando operacional e eticamente possível.
Lembretes melhoram atualização.	Mensagens em frequências diferentes para usuários com tarefas pendentes.	Percentual de registros entregues no prazo.	Experimento de comunicação.
Relatório visual melhora compreensão.	Versão resumida com gráficos versus relatório textual padrão.	Acerto em perguntas, tempo de leitura e decisão.	Teste com usuários internos.
Novo fluxo reduz prazo de prestação de contas.	Comparar períodos antes e depois, com projeto semelhante não afetado.	Dias de fechamento e pendências.	Diferenças em diferenças, se as condições forem comparáveis.

Cuidado ético e metodológico: experimentos não podem retirar direitos, dificultar acesso a serviços ou expor beneficiários. Na primeira fase, os testes devem se concentrar em processos internos, comunicação, formulários e experiência das equipes.

12.3.12 Vieses decisórios e contramedidas

Viés	Risco no CRCA Lab	Contramedida
Confirmação	Selecionar apenas dados que sustentem uma proposta ou leitura prévia.	Registrar hipótese antes da análise e apresentar evidências favoráveis e contrárias.
Ancoragem	Manter preços, metas ou expectativas com base em experiências antigas.	Usar faixas, benchmarks internos e revisão periódica dos parâmetros.
Disponibilidade	Dar peso excessivo ao caso mais recente ou memorável.	Comparar séries históricas e distribuição completa, não apenas exemplos.
Excesso de confiança	Tratar experiência da equipe ou resultado do modelo como certeza.	Intervalos de confiança, cenários, revisão por pares e registro de incerteza.
Status quo	Rejeitar novo processo porque a planilha atual é conhecida.	Pilotos curtos, demonstração de ganho e transição gradual.
Enquadramento	Mudar a decisão conforme a forma positiva ou negativa da apresentação.	Apresentar resultados em formatos equivalentes e explicitar critérios.

12.3.13 Roteiro de maturidade analítica do CRCA Lab

Nível	Capacidade	Entrega de referência	Critério para avançar
1. Fundacional	Inventário de dados, dicionário, responsáveis e qualidade.	Base mínima dos projetos-piloto.	Campos estáveis e rotina de atualização.
2. Descritivo	Painéis e relatórios padronizados.	Indicadores de produção, prazo, custo e público.	Usuários consultam e compreendem o painel.

Nível	Capacidade	Entrega de referência	Critério para avançar
3. Diagnóstico	Comparações, segmentações e análise de fatores.	Custo por entrega, gargalos e grupos prioritários.	Decisões documentadas com evidência.
4. Preditivo	Modelos de risco, custo, atraso, adoção ou potencial comercial.	Previsões validadas em dados não usados no ajuste.	Erro conhecido e utilidade operacional.
5. Causal/prescritivo	Experimentos, avaliação de efeito e recomendação de ação.	Evidência de intervenção e regras de priorização.	Governança, histórico e revisão ética.

12.3.14 Impacto na proposta comercial da spin-off

A oferta deixa de ser apenas "produção de comunicação" e passa a ser uma implantação de inteligência institucional. O CRCA Lab pode vender um ciclo completo: diagnóstico, redesenho do fluxo, base de dados, painel, análise em RStudio, capacitação, acompanhamento e avaliação de resultados.

Módulo da oferta	Entregas	Valor percebido pelo cliente
1. Diagnóstico analítico	Entrevistas, mapa de decisões, inventário de dados e maturidade.	Clareza sobre problemas, prioridades e riscos.
2. Estruturação de dados	Dicionário, formulários, Data Mart e governança.	Fonte comum e redução de informações dispersas.
3. Painel e relatórios	Dashboard interno, cards, alertas e Quarto.	Acompanhamento e prestação de contas mais rápida.
4. Business Analytics	Segmentação, regressão, previsão e cenários.	Decisões apoiadas por evidências e incerteza explicitada.
5. Experimentação	MVP, teste de processo e avaliação de efeito.	Aprendizagem antes de investir em escala.
6. Formação e sustentação	Treinamento, check-ins, documentação e revisão de OKRs.	Apropriação pela equipe e continuidade após a implantação.

Delimitação importante: a V06 não afirma que o CRCA Lab já possui base histórica suficiente para modelos preditivos robustos. Ela apresenta uma arquitetura de aplicação e um plano de maturidade. Resultados estatísticos deverão ser inseridos apenas depois da coleta, limpeza, validação e análise de dados reais ou de bases simuladas claramente identificadas.

13. Conclusão

Com o aprofundamento de Business Analytics, o projeto passa a demonstrar não apenas quais ferramentas podem ser utilizadas, mas como a organização deve estruturar perguntas, processos, métricas, modelos e ciclos de aprendizagem. O RStudio assume o papel de ambiente estatístico e reprodutível; o CRCA Conecta organiza o acesso e a coleta; e a governança conecta pessoas, dados e decisões.

O projeto demonstra que o CRCA Lab pode ser estruturado como uma unidade de inteligência institucional, capaz de transformar escuta territorial em conteúdo, conteúdo em dados, dados em decisão e decisão em sustentabilidade. O uso combinado de Python no Google Colab e R no RStudio permite organizar informações, criar indicadores, testar hipóteses e visualizar prioridades. Com Shiny e Quarto, os resultados podem ser convertidos em painéis internos e relatórios reproduzíveis, aproximando a análise das equipes que executam e coordenam os diferentes projetos.

Ao final, o CRCA Lab deixa de ser apenas uma sala de mídia ou um núcleo de produção audiovisual, assumindo o papel de infraestrutura estratégica para comunicação, formação, editoração social, prestação de contas e geração de valor para projetos sociais, comunidades e parceiros institucionais.

14. Próximos Passos

O CRCA Lab pode investigar relações entre contexto territorial, perfil dos públicos, formatos de conteúdo, engajamento, impacto social e potencial de receita, sempre reconhecendo que correlação não significa causalidade.

1. Validar a estrutura da base de dados.
2. Preencher a planilha inicial com dados reais ou simulados.
3. Rodar o notebook no Google Colab e o projeto analítico no RStudio.
4. Gerar gráficos e rankings.
5. Interpretar os resultados.
6. Revisar o texto final do projeto aplicado.
7. Preparar uma apresentação executiva com os principais achados.

```
</> Markdown
```

8. Criar um projeto RStudio com pastas padronizadas para dados, scripts, relatórios e aplicação.
9. Desenvolver um dashboard Shiny com acesso inicial às equipes dos diferentes projetos.
10. Testar autenticação, perfis de acesso, qualidade dos dados e experiência dos usuários internos.
11. Gerar um relatório reproduzível em Quarto para apresentação acadêmica e institucional.

Como próximo passo, os resultados obtidos no Google Colab e no RStudio poderão alimentar uma interface digital interna. O dashboard Shiny servirá para exploração e acompanhamento, enquanto o Quarto permitirá gerar relatórios consistentes em HTML, PDF ou Word. A primeira fase deve priorizar as equipes dos projetos; a abertura a beneficiários deverá ocorrer apenas após a validação dos dados, das permissões e das rotinas de proteção.

12. Definir a Métrica Mais Importante do Momento do MVP e os respectivos critérios de continuidade.
13. Mapear o processo atual de coleta, validação, evidências, análise e prestação de contas.
14. Construir o Data Mart inicial com dois projetos e histórico mensal.
15. Estimar um primeiro modelo de custos em RStudio, comparando regressões simples e múltiplas.
16. Desenhar um experimento interno de baixo risco para avaliar adoção ou redução de retrabalho.
17. Aplicar o diagnóstico de maturidade analítica e revisar o roadmap a cada trimestre.

15. Apêndices

Apêndice A - Estrutura sugerida do notebook Google Colab

Bloco	Conteúdo	Observações (#)
1	Apresentação do projeto em Markdown.	
2	Importação de bibliotecas: Pandas, NumPy, Matplotlib, Pathlib e Unicodedata.	-
3	Leitura da base CSV.	`# Carregamento de arquivo CSV.`
4	Diagnóstico inicial (uso de): **`head(), shape, info(), describe(), isna(), duplicated().`**	-
5	Tratamento dos dados: padronização de textos, números (conversão numérica), datas, remoção de duplicidades e preenchimento de dados nulos.	-
6	Construção de indicadores: taxa de engajamento, eficiência de custo, potencial de receita e score de prioridade.	-
7	Agregações (agrupamentos): **`groupby(), agg(), sort_values() e value_counts() .`**	-
8	Visualizações: gráficos de barras e linhas com Matplotlib.	-
9	Modelo de recomendação de formatos.	
10	Exportação dos resultados em CSV: **`ranking.to_csv('ranking_prioridades_crca_lab.csv', index=False)`**	

Apêndice B - Exemplo de base simulada

Conteúdo	Tema	Público	Canal	Formato
Podcast com agricultor familiar	Agroecologia	Agricultores familiares	YouTube	Podcast
Curso online sobre economia solidária	Formação	OSCs e cooperativas	Plataforma EAD	Curso
Relatório de impacto para financiadores	Prestação de contas	Financiadores	Site institucional	Relatório
Série sobre cozinhas comunitárias	Segurança alimentar	Comunidades beneficiárias	Instagram	Vídeo curto

Apêndice C - Estrutura sugerida do projeto RStudio

A organização abaixo separa dados originais, dados tratados, funções, relatórios, aplicação web e resultados exportados. Essa estrutura ajuda a equipe a localizar arquivos, repetir análises e reduzir alterações manuais.

```
CRCA_Lab_Analytics/
├── CRCA_Lab_Analytics.Rproj
├── data/
│   ├── raw/
│   └── processed/
├── R/
│   ├── importar_dados.R
│   ├── tratar_dados.R
│   └── calcular_indicadores.R
├── app/
│   └── app.R
├── reports/
│   └── relatorio_executivo.qmd
├── output/
│   ├── ranking_prioridades.csv
│   └── graficos/
└── README.md
```

Produto acadêmico sugerido: um arquivo .Rproj, um script principal em R, uma base simulada ou anonimizada, um relatório Quarto e um protótipo Shiny. O conjunto demonstra a passagem da coleta de dados para a análise, da análise para a visualização e da visualização para a decisão.

Artefato	Função no projeto aplicado
CRCA_Lab_Analytics.Rproj	Mantém o ambiente de trabalho e os caminhos do projeto.
base_crca_lab.csv	Reúne os registros usados na demonstração.
calcular_indicadores.R	Executa tratamento, normalização, métricas e ranking.
relatorio_executivo.qmd	Gera documento reproduzível com texto, tabelas e gráficos.
app.R	Disponibiliza filtros e indicadores em uma aplicação Shiny.
ranking_prioridades.csv	Exporta o resultado final para anexar ao trabalho.

Apêndice D - Plano mínimo de análise em RStudio

Script	Finalidade	Saída
01_importacao_qualidade.R	Importar bases, validar tipos, nulos, duplicidades e chaves.	Relatório de qualidade e base tratada.
02_descritiva_segmentos.R	Descrever custos, prazos, públicos, territórios, canais e formatos.	Tabelas, histogramas e comparações.
03_modelo_custos.R	Comparar regressões simples e múltiplas para custo total.	Coeficientes, R^2, resíduos e diagnóstico.
04_modelo_adocao.R	Estimar fatores associados ao uso recorrente do sistema.	Odds ratios, probabilidades e grupos de risco.
05_experimentos.R	Avaliar testes ou rollout escalonado de mudanças internas.	Efeito estimado, intervalo e limitações.
06_relatorio_quarto.qmd	Consolidar método, resultados, gráficos e recomendações.	HTML, Word ou PDF reproduzível.

16. Checklist de preenchimento final

☐ Formular um OKR trimestral e uma Métrica Mais Importante do Momento.
☐ Identificar clientes internos, decisões, periodicidade e responsáveis.
☐ Mapear processo atual e processo futuro de dados e evidências.
☐ Classificar a maturidade analítica atual e o nível desejado.
☐ Explicar a diferença entre correlação, regressão, previsão e causalidade.
☐ Avaliar R^2, resíduos, incerteza e qualidade dos dados dos modelos.
☐ Documentar vieses, limitações, hipóteses e decisões tomadas.
☐ Diferenciar claramente dados reais, planejados e simulados.
☐ Inserir nome do curso, aluno e professor.

☐ Inserir resumo executivo final.

☐ Validar problema de decisão.

☐ Preencher missão, visão e valores.

☐ Consolidar base de dados.

☐ Rodar notebook no Google Colab e projeto no RStudio.

☐ Inserir gráficos gerados.

☐ Interpretar resultados.

☐ Revisar conclusão.

☐ Inserir referências.

☐ Preparar versão em PDF ou apresentação final.

☐ Validar o script em R e o cálculo do score normalizado.

☐ Inserir print ou link do protótipo Shiny.

☐ Inserir arquitetura CRCA Conecta + RStudio.

☐ Registrar limitações de acesso, qualidade dos dados e LGPD.

17. Referências e documentos de apoio

- Apostila da Aula 02 — Business Analytics, OKRs, Times e Lean Analytics.
- Apostila da Aula 03 — Estudo de caso, cultura organizacional, CRISP-DM, DECSA e vieses decisórios.
- Apostila da Aula 04 — Modelos preditivos, correlação, regressão e uso do R.

- Apostila da Aula 05 — Regressão múltipla, R^2, causalidade, contrafactuais e experimentos.
- Apostila da Aula 06 — Manipulação de dados no RStudio e formulação das perguntas analíticas.
- Apostila da Aula 07 — Modelagem preditiva aplicada e interpretação de modelos simples e completos.
- Apostila da Aula 08 — BPR, maturidade analítica, Data Warehouse, governança e integração entre pessoas, processos e tecnologia.
- Canvas reorganizado do CRCA Lab.
- Apresentação executiva CRCA Lab: Impacto Social Estruturado.
- Documento de estruturação da sala multiuso de podcast.
- Plano de estruturação institucional do CRCA Lab.
- Resumo geral das funcionalidades utilizadas em Python.
- Notebook de aula sobre organização de dados em Python.
- Base exemplo vendas.csv utilizada para exercitar leitura, limpeza, métricas, agregações, ranking e gráficos.

- POSIT. RStudio IDE: ambiente integrado para R e Python. Disponível em: https://posit.co/products/open-source/rstudio. Acesso em: 12 jul. 2026.
- POSIT. Shiny: framework para criação de aplicações web interativas com R e Python. Disponível em: https://shiny.posit.co. Acesso em: 12 jul. 2026.
- QUARTO. Sistema de publicação técnica reproduzível. Disponível em: https://quarto.org. Acesso em: 12 jul. 2026.
- POSIT. RStudio Server: acesso ao ambiente RStudio por navegador. Disponível em: https://posit.co/products/open-source/rstudio-server. Acesso em: 12 jul. 2026.