

# DESENVOLVIMENTO DE UM SISTEMA DE MENSAGERIA

*Autores:*

*Bruno Alexandre Amaral Silva*

*David Emanuel Gentil Silva*

*Gustavo Moraes*

*Lucas Duarte Roballo*

*Maria Ribeiro Machado Pires*

*Walife Oliveira Rodrigues*

**BECOME**X

# CONTEXTUALIZAÇÃO

## SETOR DE ATUAÇÃO:

- *Análise de dados públicos*
- *Aplicações: compliance, vendas B2B, controle institucional*

## PROBLEMA IDENTIFICADO

- Dados públicos fragmentados, desatualizados e difíceis de utilizar
- Informações importantes (como telefone e e-mail) ausentes

### **Por que esse problema?**

- Processo manual: demorado, suscetível a erros
- Necessidade de dados confiáveis para tomada de decisão

## INDICADORES DE SUCESSO

- Extração automatizada mensal
- Enriquecimento de +90% dos CNPJs
- Geração de CSV estruturado
- Envio automatizado via e-mail

# RESUMO EXECUTIVO DOS RESULTADOS

*SISTEMA FUNCIONAL DE AUTOMAÇÃO QUE COLETA, FILTRA E ENRIQUECE DADOS DE EMPRESAS BRASILEIRAS, COM ENTREGA AUTOMATIZADA VIA E-MAIL.*

## O que o sistema faz?

- Filtra empresas por CNAE (3 tipos específicos)
- Consulta API ReceitaWS por CNPJ
- Enriquecimento dos dados com nome, e-mail, situação, etc.
- Geração e envio automático de CSV com os resultados

## Resultados

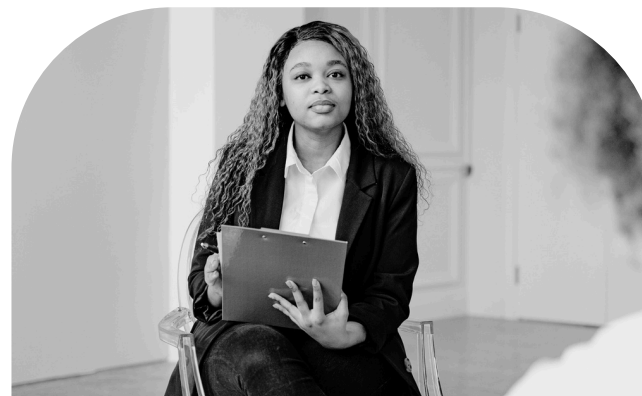
- Enriquecimento de +90% dos CNPJs;
- Arquivo CSV final padronizado;
- Envio automático via e-mail com Z-API.

## Principais ações

- Construção do pipeline com Python e APIs;
- Filtragem por CNAE;
- Automatização do envio final.

# BUSINESS UNDERSTANDING

*Agilizar e automatizar a coleta e análise de dados públicos sobre empresas para uso por equipes de compliance, vendas e controle.*



## Objetivo de negócio

Melhorar acesso a dados públicos relevantes para diferentes setores.

## Objetivos da Mineração de Dados

- Aplicar filtros por CNAE
- Obter dados como e-mail, telefone e situação cadastral
- Enriquecer +90% dos CNPJs
- Gerar arquivo CSV padronizado e funcional

## Avaliação da Situação

- API ReceitaWS tem limites e requer monitoramento
- Z-API requer autenticação e plano ativo
- Dados públicos podem ter erros de encoding e inconsistências
- Projeto executado com ferramentas gratuitas, viável mesmo com baixo orçamento

# PLANO DO PROJETO

O PLANO DO PROJETO FOI DIVIDIDO EM 5 FASES TÉCNICAS PRINCIPAIS:

DOWNLOD  
DOS  
DADOS

EXTRAÇÃO  
DO CSV

FILTRAGEM  
DOS DADOS

CONSULTA E  
ENRIQUECI-  
MENTO

EXPORTA-  
ÇÃO  
E ENVIO

# DATA UNDERSTANDING

## FONTE DE DADOS

- Portal da transparência - Favorecidos PJ (CSV mensal em ZIP)
- ReceitaWS API - enriquecimento por CNPJ

## ESTRUTURA DOS DADOS ORIGINAIS

- ~200.000 linhas por mês
- Principais colunas: CNPJ, COD\_CNAE, Razão Social, UF, Município, Valor Transferido, Data do Pagamento

## ENRIQUECIMENTO

- API ReceitaWS retornou: Nome, E-mail, Telefone, Situação Cadastral, Tipo, Abertura



## DESAFIOS

*Arquivos ausentes e com codificação inconsistente*

*API com limite de requisições*

*Encoding latin1 e separador ‘;’*

*Dados nulos ou duplicados*



## SOLUÇÕES

*Logs de verificação e alertas*

*Delay automático entre requisições*

*Tratamento com pandas (read\_csv com encoding e delimiter)*

*Salvamento incremental para evitar perdas*

# DATA PREPARATION: SELEÇÃO DE DADOS

## OBJETIVOS

Transformar dados brutos em informações úteis e alinhadas ao objetivo do projeto.

### SELEÇÃO DE COLUNAS

- Mantidos: CNPJ, COD\_CNAE, UF, Município, Valor Transferido
- Adicionados: Nome, E-mail, Telefone, Situação Cadastral (via ReceitaWS)
- Removidos: Campos redundantes ou irrelevantes

### SELEÇÃO DE REGISTROS

- Foco em 3 códigos CNAE definidos
- Remoção de CNPJs inválidos ou ausentes
- Linhas com repasses diferentes mantidas (transações distintas)

## RESULTADOS

Base direcionada, limpa e relevante ao negócio, com menor ruído e alta precisão para análise.



# DATA PREPARATION: LIMPEZA E SELEÇÃO



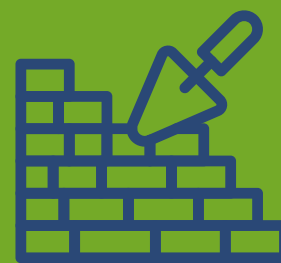
## LIMPEZA DE DADOS

- Conversão de tipos: de texto para numérico (ex.: VALOR\_TRANSFERIDO)
- Padronização: datas no formato ISO (YYYY-MM-DD)
- Normalização: campos em maiúsculas e sem espaços extras
- Remoção de caracteres especiais (ex.: em e-mails e nomes)



## TRATAMENTO DE NULOS

- CNPJs inválidos ou vazios removidos
- Demais campos: preenchidos com "N/D" ou string vazia



## CONSTRUÇÃO DOS DADOS

- Novos campos adicionados:
  - Nome empresarial, e-mail, telefone
  - Situação cadastral, tipo, data de abertura
- Enriquecimento linha a linha com salvamento incremental



## RESULTADOS

Base final sem nulos críticos, com formatação padronizada e dados enriquecidos para análise confiável.



# DATA PREPARATION: INTEGRAÇÃO E FORMATAÇÃO



## INTEGRAÇÃO DE FONTES

- Fonte A: Dados públicos (Portal da Transparência)
- Fonte B: ReceitaWS (dados enriquecidos por CNPJ)
- Técnica: junção via chave primária CNPJ
- Desempenho: ~10 CNPJs/min com delay de 1,5s



## FORMATAÇÃO FINAL

- Exportação para CSV com codificação utf-8-sig (compatível com Excel)
- Campos organizados em ordem lógica:
  - CNPJ
  - COD\_CNAE
  - Nome, E-mail, Telefone
  - Situação, Abertura
  - Valor, UF, Município, Data de Pagamento
- Numéricos com ponto decimal
- Textos com aspas duplas



## RESULTADOS

Arquivo limpo, compatível, padronizado e pronto para consumo por sistemas e usuários finais.

# MODELING: ESTRATÉGIA E TÉCNICAS



OBJETIVO	Executar um pipeline funcional para enriquecer dados com base em regras de negócio, sem uso de modelo estatístico ou preditivo.
ESTRATÉGIA DE MODELAGEM	<ul style="list-style-type: none"><li>• Filtragem booleana com pandas (seleção por CNAE)</li><li>• Iteração controlada com <code>df.iterrows()</code></li><li>• Consulta à API ReceitaWS linha a linha</li><li>• Enriquecimento condicional (só sobrescreve se o campo estiver ausente)</li></ul>
REGRAS E PREMISSAS	<ul style="list-style-type: none"><li>• CNPJs válidos = 14 dígitos (validados com <code>zfill</code>)</li><li>• Dados da ReceitaWS considerados confiáveis</li><li>• Respeito aos limites de requisições da API</li></ul>
TESTES REALIZADOS	<ul style="list-style-type: none"><li>• Casos com CNPJs válidos, inválidos, duplicados e já consultados</li><li>• Verificação manual de dados consultados no site da Receita Federal</li><li>• Arquivos validados em Excel, Google Sheets e Notepad++</li></ul>

# MODELING: AVALIAÇÃO E ARQUITETURA



## ARQUITETURA DO SCRIPT PRINCIPAL

- main.py orquestra o pipeline completo
- Módulos auxiliares:
- download\_data.py (coleta)
- process\_data.py (tratamento e enriquecimento)
- send\_data.py (envio via e-mail)

## PARÂMETROS DEFINIDOS

- CNAEs alvo: [5250803, 5231102, 5250804]
- Delay entre requisições: 1,5 segundos
- Salvamento incremental: ativado
- Execução inicial limitada a 10 registros (para testes)

## AVALIAÇÃO TÉCNICA

- Tempo médio por CNPJ: 1,4 segundos
- Enriquecimento completo em +90% dos casos
- Arquivo CSV gerado com estrutura compatível
- Envio via e-mail bem-sucedido

## RESULTADOS

***Pipeline confiável, eficiente e pronto para escalabilidade. Estrutura modular facilita manutenção e reuso.***

# EVALUATION

## OBJETIVOS ATINGIDOS

- Coleta automatizada com base em data
- Filtragem por CNAE aplicada com sucesso
- +90% dos CNPJs enriquecidos com dados externos
- Arquivo CSV padronizado gerado
- Envio automatizado via e-mail com Z-API

## DESTAQUES POSITIVOS

- Simplicidade: execução via script único
- Robustez: recuperação automática em caso de falhas
- Rapidez: 100 empresas processadas em <3 minutos

## LIMITAÇÕES OBSERVADAS

- ReceitaWS tem limite de requisições
- Nem todos os CNPJs possuem e-mail ou telefone
- Impossibilidade de paralelização por restrição da API



# DEPLOYMENT

## **IMPLANTAÇÃO ATUAL**

- Script main.py executável localmente
- Exportação de CSV automatizada
- Envio via API Z-API para e-mail

## **PRÓXIMAS AÇÕES (CURTO PRAZO)**

- Agendamento automático com cron
- Interface web simples para configurar CNAEs e destinatário

## **EVOLUÇÕES FUTURAS (MÉDIO PRAZO)**

- Substituição do CSV por banco de dados SQL
- Relatórios históricos e controle de múltiplos usuários
- Integração com APIs pagas para dados mais completos
- Aplicação em outros segmentos (ex: fornecedores)

## **ENTREGÁVEIS SUGERIDOS**

- Manual de uso do sistema
- Plano de controle e monitoramento
- Dashboard simples (futuro)

```
m "../assets/icons/arrow.svg";  
n "../assets/icons/bolt.svg";  
on } from "../assets/icons/right-arrow.svg";
```

```
ef } from "react";  
nsition-group";
```

```
"eslintConfig": {  
  "extends": [  
    "react-app"  
  ]  
},
```

# ONE-PAGER

## IMPACTOS TÉCNICOS E DE NEGOCIO

- +90% dos CNPJs enriquecidos com dados externos
- Redução de horas de trabalho manual para minutos
- Diminuição de erros por digitação e busca manual



EFICIÊNCIA

## GANHOS OPERACIONAIS

- Tempo médio economizado por mês: 20h+
- Redução de falhas: -90%
- Maior confiabilidade no uso de dados para compliance e vendas



CUSTOS

## VALOR ENTREGUE

- ✓ Dados confiáveis
- ✓ Processo escalável
- ✓ Solução leve, de baixo custo e alto impacto



QUALIDADE

# OBRIGADA!

DÚVIDAS? FALE COM A GENTE!

1BRUNOALEXANDRE@GMAIL.COM  
DAVID.GENTIL85@GMAIL.COM  
GUSTAVO.MORAIS3119@GMAIL.COM  
RIBEIROMARIAMP@GMAIL.COM  
WALIFE.OR@GMAIL.COM

*“AUTOMAÇÃO EFICIENTE É AQUELA QUE ENTREGA VALOR COM SIMPLICIDADE.”*

*ESTE PROJETO FOI DESENVOLVIDO COMO PARTE DA FORMAÇÃO EM DADOS DA DNC.*