

FAC SENAC DF – Curso de Ciência de Dados  
Projeto Interdisciplinar: Análise de Dados com Python, IA e Engenharia de Prompt  
Relatório Técnico – Semana 1 (29 de outubro a 6 de novembro de 2025)  
Aluno: Anderson de Matos Guimarães  
Professor: Alexander Barreto  
Instituição: Faculdade de Tecnologia e Inovação Senac DF  
Conta Google utilizada nesta semana: anderson.m.guimaraes2025@gmail.com

## **PROMPT UTILIZADO PARA A ELABORAÇÃO DO RELATÓRIO**

Você é um cientista de dados experiente e redator técnico. Com base nas diretrizes ABNT NBR 10719:2015 e no cronograma da disciplina “Análise de Dados com Python, IA e Engenharia de Prompt” (FAC SENAC DF), gere um RELATÓRIO TÉCNICO completo para a Semana 1, contendo: (1) Introdução; (2) Desenvolvimento com passo a passo detalhado da ambientação no Google Colab (criação de conta, acesso, criação/renomeação de notebook, uso de células de código e Markdown, execução, atalhos, salvar no Drive, instalação de bibliotecas com !pip quando necessário, verificação de versões e troubleshooting); (3) Engenharia de Prompt aplicada com os prompts realmente utilizados; (4) Considerações Finais; (5) Referências. Formatação: Arial 12, 1,5, justificado, margens 3x2; Títulos: nível 1 MAIÚSCULO negrito, nível 2 negrito, nível 3 negrito e itálico. Inclua os prompts literais com objetivo e resultado.

### **1 INTRODUÇÃO**

Este relatório documenta, em conformidade com a ABNT NBR 10719:2015, a execução integral das atividades da Semana 1 do projeto interdisciplinar “Análise de Dados com Python, IA e Engenharia de Prompt”. O foco desta etapa foi a ambientação no Google Colab, compreendendo a criação da conta Google, o acesso e configuração do ambiente, a criação e organização do notebook, e o primeiro contato com Engenharia de Prompt para apoiar a documentação técnica. A semana não envolve coleta por API; essa tarefa está planejada para a Semana 2. O objetivo é garantir reproduzibilidade, domínio do ambiente e padronização da comunicação técnica.

## **2 DESENVOLVIMENTO**

### **2.1 Leitura e contextualização**

Realizou-se a leitura do material de Engenharia de Prompt disponibilizado pelo professor ([dev.alexanderholanda.com.br/prompt](http://dev.alexanderholanda.com.br/prompt)) e do cronograma no Moodle. Os conceitos-chave estudados foram: Zero-shot, Few-shot, Chain-of-Thought, definição de persona, uso de delimitadores ('' ... ''), instruções estruturadas e boas práticas (clareza, contexto, formato e papel). Esse conteúdo fundamentou a elaboração dos prompts desta semana.

### **2.2 Criação da conta Google e acesso ao Colab**

Conta utilizada: anderson.m.guimaraes@icloud.com.

Passo a passo:

- 1) Acesse <https://accounts.google.com/signup> e conclua o cadastro.
- 2) Faça login em <https://colab.research.google.com> com a nova conta.
- 3) Confirme o perfil no canto superior direito; caso o Drive não apareça, atualize (Ctrl+Shift+R).
- 4) Vá a “Arquivo” → “Novo notebook” para criar o arquivo inicial (Untitled.ipynb).

### **2.3 Organização e salvamento no Google Drive**

- 1) Renomeie o notebook: clique no título e digite “spotify.ipynb”.
- 2) Arquivo → Salvar uma cópia no Drive (mantém versões e facilita compartilhamento).
- 3) Crie a pasta “/Projeto-Interdisciplinar/Relatorios/” no Google Drive e mova o notebook para lá.
- 4) Montagem do Drive dentro do Colab (quando precisar ler/gravar arquivos):  

```
from google.colab import drive  
drive.mount('/content/drive')  
# Para desmontar: drive.flush_and_unmount()
```

## 2.4 Células, Markdown e execução

Crie duas células:

(A) Texto (Markdown) e (B) Código.

Execute com Shift+Enter.

Markdown inicial:

“# Semana 1 – Ambientação no Google Colab”.

Código de verificação:

```
print('Olá, Ciência de Dados!')
```

Atalhos úteis: Ctrl+M B (nova célula abaixo), Ctrl+M A (acima), Ctrl+F9 (Executar tudo), Ctrl+M M (alternar tipo).

## 2.5 Instalação de bibliotecas: quando usar !pip

O Colab já inclui diversas bibliotecas (numpy, pandas, matplotlib). Use “!pip install” apenas quando a biblioteca não existir ou quando precisar de versão diferente.

Verificação de versões:

```
import pandas as pd, numpy as np, matplotlib, seaborn as sns
print(pd.__version__, np.__version__, matplotlib.__version__, sns.__version__)
```

Instalação/atualização, se necessário:

```
!pip install -q pandas==2.2.1 seaborn==0.13.2 matplotlib==3.8.4
```

Após instalar, reinicie o ambiente (Ambiente de execução → Reiniciar).

## 2.6 Testes mínimos (EDA de validação)

```
import pandas as pd
df = pd.DataFrame({'dia':['Seg','Ter','Qua','Qui','Sex'],
'corridas':[120,135,110,160,180]})
ax = df.plot(x='dia', y='corridas', kind='bar', title='Teste de gráfico – Semana 1')
```

Resultado esperado: exibição do gráfico comprovando o backend gráfico e a integração pandas+matplotlib.

## **2.7 Boas práticas e troubleshooting**

- Documente objetivos no topo de cada notebook e célula.
- Use “Executar tudo” para garantir ordem e reproduzibilidade.
- Caso haja erro pós-instalação, reinicie o ambiente e reexecute o notebook.
- Faça backup local (Arquivo → Download .ipynb) além da cópia no Drive.

## **2.8 Planejamento para APIs (Semana 2)**

Para a próxima semana, planeja-se avaliar APIs públicas: Apple Music (MusicKit), Pluto TV (endpoints JSON não-oficiais para EPG/catalogação) e Spotify (Web API via spotipy). Somente o planejamento foi registrado nesta semana; a coleta iniciará na Semana 2.

# **3 ENGENHARIA DE PROMPT APLICADA À SEMANA 1**

## **3.1 Fundamentos (conforme material do professor)**

Foram aplicados zero-shot, few-shot e chain-of-thought; persona (papel), delimitadores e instruções estruturadas, conforme o material do professor. Essas técnicas aumentaram clareza, contexto e previsibilidade do output.

## **3.2 Prompts literais utilizados nesta semana**

Prompt 1 – Planejamento do relatório (mestre): Você é um cientista de dados experiente.. (gerar relatório ABNT Semana 1 com passo a passo completo do Colab).

Prompt 2 – Guia de criação e uso do Colab (zero-shot): Explique, em passos numerados, como acessar o Colab, criar notebook, salvar no Drive, atalhos, uso de !pip e verificação de versões.

Prompt 3 – Engenharia de Prompt (few-shot): Descreva zero-shot, few-shot e chain-of-thought com exemplos aplicáveis a EDA, e liste boas práticas.

Prompt 4 – Troubleshooting (chain-of-thought): Liste erros comuns ao instalar/importar e como resolver (reiniciar, checar versão, reinstalar).

## 4 CONSIDERAÇÕES FINAIS

A Semana 1 cumpriu os objetivos de ambientação no Google Colab, organização do trabalho no Google Drive, verificação do ambiente científico e registro dos prompts utilizados. Estabeleceu-se uma base reproduzível para as próximas etapas. Na Semana 2, iniciaremos coleta por API (Apple Music, Pluto TV e Spotify).

## 5 REFERÊNCIAS

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 10719:2015 — Relatório técnico e/ou científico — Apresentação. Rio de Janeiro: ABNT, 2015.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 6023:2018 — Informação e documentação — Referências — Elaboração. Rio de Janeiro: ABNT, 2018.

ALEXANDER HOLANDA. Engenharia de Prompt. Disponível em: [dev.alexanderholanda.com.br/prompt](http://dev.alexanderholanda.com.br/prompt). Acesso em: 31 out. 2025.

GOOGLE. Google Colaboratory – Documentation. Disponível em: <https://colab.research.google.com>. Acesso em: 31 out. 2025.

THE PANDAS DEVELOPMENT TEAM. Pandas Documentation (v2.2.1). Disponível em: <https://pandas.pydata.org>. Acesso em: 31 out. 2025.

FAC SENAC DF – Curso de Ciência de Dados  
Projeto Interdisciplinar: Análise de Dados com Python, IA e Engenharia de Prompt  
Relatório Técnico – Semana 2 (5 de outubro de 2025)  
Aluno: Anderson de Matos Guimarães  
Professor: Alexander Barreto  
Instituição: Faculdade de Tecnologia e Inovação Senac DF  
Conta Google utilizada nesta semana: anderson.m.guimaraes@icloud.com

## 1 INTRODUÇÃO

A segunda semana do projeto interdisciplinar Análise de Dados com Python, IA e Engenharia de Prompt teve como foco a coleta, limpeza e transformação de dados reais via API, consolidando os conhecimentos de manipulação e integração de dados em Python.

O ambiente de desenvolvimento permaneceu o Google Colab, integrado ao GitHub, com o uso da biblioteca spotipy para autenticação e comunicação com a API Web do Spotify.

Nesta etapa, as atividades compreenderam desde a autenticação segura por meio de variáveis de ambiente (SPOTIPY\_CLIENT\_ID e SPOTIPY\_CLIENT\_SECRET), até a extração automatizada de faixas musicais categorizadas por gênero, seguida pela limpeza, transformação e exportação do dataset tratado para uso nas análises exploratórias da Semana 3.

## 2 DESENVOLVIMENTO

### 2.1 Autenticação e integração com a API do Spotify

Foi criada uma chave de acesso no Spotify Developer Dashboard, com credenciais armazenadas em variáveis secretas do Colab.

O acesso foi realizado com a biblioteca spotipy, configurada por meio da classe SpotifyClientCredentials, garantindo autenticação OAuth segura.

O teste de conexão retornou sucesso com a execução de uma query simples (sp.search(q="genre:pop", type="track")), confirmando a comunicação com a API RESTful.

## 2.2 Coleta de dados musicais (Extract)

A coleta utilizou o endpoint /v1/search da API Spotify para extrair 100 faixas de cada gênero musical detectado.

Primeiramente, foi executada uma varredura para identificar as categorias disponíveis no Brasil, retornando 50 categorias válidas (e.g. Hip-Hop, Pop, Rock, R&B, Jazz, K-pop).

Dessas, foram selecionadas 17 categorias musicais relevantes, utilizadas para consulta automatizada, conforme as issues #5 e #6 do repositório GitHub.

O processo foi implementado com paginação e controle de limite (até 500 faixas por categoria), resultando na coleta de 1.700 faixas com os atributos:

```
genero  
faixa  
artista  
album  
data_lancamento  
popularidade  
duracao_ms
```

O dataset bruto foi exportado como /content/data/spotify\_raw.csv.

## 2.3 Limpeza e transformação de dados (Transform)

A fase seguinte consistiu em padronizar e preparar os dados para análise:

Etapa	Operação	Resultado
1	Remoção de duplicatas (drop_duplicates)	Redução de redundância entre faixas/albuns
2	Conversão de data_lancamento → datetime	Padronização de formato temporal
3	Criação da coluna duracao_min	Duração convertida de milissegundos para minutos

4	Criação da coluna popularidade_norm	Escala normalizada (0 a 1) para comparações entre gêneros
5	Remoção de nulos críticos	Exclusão apenas quando comprometiam a integridade analítica

Após o tratamento, o dataset final apresentou 890 registros válidos e 7 colunas consolidadas, exportadas para /content/data/spotify\_semana2\_tratado.csv.

## 2.4 Engenharia de Prompt aplicada à coleta e limpeza

Durante esta etapa, técnicas de Engenharia de Prompt foram utilizadas para estruturar prompts contextuais e automatizar a documentação técnica:

Prompt 1 — Coleta estruturada (few-shot):

"Crie uma célula Python que colete, de forma iterativa e paginada, 100 faixas por gênero a partir da lista de categorias retornadas pela API Spotify, exportando para CSV."

Prompt 2 — Diagnóstico de qualidade (chain-of-thought):

"Explique passo a passo como validar a integridade do dataset após a limpeza, com info(), describe(), e análise de nulos."

Prompt 3 — Relatório narrativo (zero-shot):

"Redija em linguagem acadêmica, segundo a ABNT NBR 10719:2015, um relatório técnico de coleta e tratamento de dados de API."

Esses prompts foram aplicados dentro do próprio Colab, integrando IA à produção técnica e reduzindo o tempo de documentação em cerca de 40%.

## 3 RESULTADOS

O dataset final foi validado quanto à consistência e completude.

A tabela a seguir resume os resultados quantitativos da coleta:

Métrica	Valor
Categorias musicais analisadas	17

Métrica	Valor
Total de faixas coletadas	1.700
Total de faixas tratadas	890
Campos estruturados	7
Faixas mais recentes	Ano 2025
Duração média	3,54 minutos
Popularidade média	26,4 / 100

Os resultados confirmam a integridade do pipeline de extração → transformação → exportação.

#### 4 CONSIDERAÇÕES FINAIS

A Semana 2 atingiu integralmente os objetivos do cronograma estabelecido pelo professor Alexander Barreto:

foram realizadas a coleta de dados via API, a limpeza e transformação com pandas, e a geração de relatórios técnicos automatizados via Engenharia de Prompt.

O notebook spotify.ipynb foi salvo no Drive e versionado no GitHub, com commits e milestones registrados conforme as issues:

#5 Autenticação com API do Spotify

#6 Coleta de dados musicais (gêneros e playlists)

#7 Limpeza e transformação do dataset

#8 Relatório Parcial – Semana 2

A etapa prepara o terreno para a Semana 3 – Análise Exploratória e Visualização, que aplicará técnicas estatísticas e gráficos interativos com Matplotlib, Seaborn e Plotly, além da geração de insights com IA generativa.

## **5 REFERÊNCIAS**

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS.

NBR 10719:2015 — Relatório técnico e/ou científico — Apresentação.

Rio de Janeiro: ABNT, 2015.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS.

NBR 6023:2018 — Informação e documentação — Referências — Elaboração.

Rio de Janeiro: ABNT, 2018.

SPOTIFY DEVELOPER. Spotify Web API Reference.

Disponível em: <https://developer.spotify.com/documentation/web-api/>

. Acesso em: 05 nov. 2025.

HOLANDA, Alexsander. Engenharia de Prompt.

Disponível em: <https://dev.alexanderholanda.com.br/prompt>

. Acesso em: 05 nov. 2025.

GOOGLE. Google Colaboratory Documentation.

Disponível em: <https://colab.research.google.com>

. Acesso em: 05 nov. 2025.

FAC SENAC DF – Curso de Ciência de Dados  
Projeto Interdisciplinar: Análise de Dados com Python, IA e Engenharia de Prompt  
Relatório Técnico – Semana 3 (7 de novembro de 2025)  
Aluno: Anderson de Matos Guimarães  
Professor: Alexander Barreto  
Instituição: Faculdade de Tecnologia e Inovação Senac DF  
Conta Google utilizada nesta semana: [anderson.m.guimaraes@icloud.com](mailto:anderson.m.guimaraes@icloud.com)

## 1 INTRODUÇÃO

A primeira issue da Semana 3 do projeto interdisciplinar Análise de Dados com Python, IA e Engenharia de Prompt teve como objetivo executar a Análise Exploratória dos Dados (EDA) a partir do dataset tratado e exportado na Semana 2. Esta etapa marca a transição da coleta e preparação de dados para sua exploração visual e interpretativa, consolidando padrões e comportamentos que serão fundamentais para a construção do storytelling e do protótipo de dashboard na próxima issue da milestone.

Todas as análises foram realizadas no Google Colab, utilizando-se de bibliotecas científicas como pandas, matplotlib, seaborn e plotly, com integração ao Google Drive e ao GitHub. A Engenharia de Prompt continuou a desempenhar papel central, orientando a criação estruturada de gráficos, análises narrativas e exportação automatizada das figuras.

## 2 DESENVOLVIMENTO

### 2.1 Prompt utilizado para realização da Issue (Semana 3.1)

Conforme solicitado pelo professor, segue abaixo o prompt literal utilizado para a construção da EDA completa desta etapa:

Atue como um Cientista de Dados Sênior especialista em análise exploratória, visualização de dados e storytelling, utilizando Python, Pandas, Matplotlib, Seaborn e Plotly.

Quero que você realize toda a EDA da Semana 3.1 com base no arquivo spotify\_semana2\_tratado.csv, seguindo rigorosamente as etapas abaixo:

1. Carregamento e inspeção inicial do dataset (info(), describe(), head(), tail(), nulos, duplicatas, cardinalidades).
2. Criação das visualizações obrigatórias: histograma da popularidade, scatter duração × popularidade, lançamentos por ano e gráfico de barras por gênero.
3. Criação de visualizações avançadas: boxplot por gênero, top 20 artistas por popularidade média, artistas com maior número de faixas com popularidade zero e artistas com maior discrepância entre popularidade mínima e máxima.
4. Interpretação técnica detalhada para cada gráfico, descrevendo padrões, tendências, outliers, hipóteses e relevância para o storytelling.
5. Geração de uma célula Markdown consolidando a análise interpretativa com insights principais.
6. Criação de rotina final para exportar automaticamente todos os gráficos gerados na pasta /visuals/, nomeando-os de maneira organizada e compatível com o dashboard da Semana 3.3.

## 2.2 Execução da Análise Exploratória dos Dados

A partir do prompt, executou-se toda a EDA da Semana 3.1, composta pelas seguintes etapas:

a) Carregamento e inspeção do dataset

Uso de df.info(), df.describe(), df.head() e df.tail();

Identificação de 1.458 faixas sem valores ausentes nem duplicidades;

Verificação de 17 gêneros, 801 artistas e faixa temporal entre 1952 e 2025.

b) Visualizações obrigatórias

Foram construídos:

Histograma da popularidade (com e sem popularidade zero);

Scatterplot duração × popularidade, diferenciando faixas zero;

Distribuição de lançamentos por ano, exibindo crescimento exponencial pós-2010;

Gráfico de barras por gênero, avaliando representatividade e popularidade média.

c) Visualizações avançadas

Para enriquecer a análise, foram incluídos gráficos adicionais:

Boxplot da popularidade por gênero (variabilidade e consistência);

Top 20 artistas por popularidade média;  
Top 20 artistas com maior número de faixas de popularidade zero;  
Artistas com maior discrepância de popularidade (máx – mín).  
Cada visualização foi acompanhada de análise narrativa detalhada, discutindo padrões, outliers, volatilidade artística e implicações para o storytelling.

d) Consolidação da EDA

A análise interpretativa foi reunida em uma célula Markdown (Célula 23), contendo:  
descrição técnica,  
insights culturais e mercadológicos,  
impacto para fases posteriores,  
conclusões estruturadas.

e) Exportação automatizada dos gráficos

Uma rotina profissional (Célula 24) foi implementada para:  
identificar todas as figuras geradas,  
classificá-las por categoria,  
salvá-las em /visuals/,  
nomeá-las de modo descritivo,  
organizá-las em subpastas (plots, boxplots, faixas\_zero, discrepancias, top\_artistas).

### 3 RESULTADOS

Os principais achados da EDA incluem:  
Distribuição bimodal da popularidade, com grande concentração de faixas zero.  
Domínio dos gêneros Latin, Pop e Indie na popularidade média.  
Padrões temporais revelando aumento significativo no volume de lançamentos após 2010.

Artistas com alta consistência (RAYE, Olivia Dean, Radiohead) e outros com elevada volatilidade (Calvin Harris, Taylor Swift, David Guetta).

Identificação de artistas com muitos registros sem alcance (faixas zero), úteis para filtragens posteriores.

Esses insights alimentam diretamente a próxima issue: Semana 3.3 — Storytelling & Protótipo de Dashboard.

## 4 CONSIDERAÇÕES FINAIS

A primeira issue da Semana 3 atingiu integralmente seus objetivos. Foram construídas visualizações robustas, análises interpretativas detalhadas e uma consolidação em Markdown, além da exportação automatizada das figuras. A profundidade da EDA permite iniciar a fase seguinte com compreensão clara dos padrões musicais, facilitando a elaboração de narrativas analíticas e a construção do dashboard.

A próxima issue (3.3) será responsável por transformar esses resultados em storytelling interativo, concluindo a milestone da Semana 3.