

Ferramentas de Visualização de Dados para Business Analytics



Tópicos da Aula

01

Critérios para Seleção

Recursos obrigatórios, recomendados e fatores de avaliação críticos

02

Magic Quadrant Gartner 2025

Análise do posicionamento dos principais fornecedores no mercado

03

Análise Detalhada

Avaliação aprofundada de líderes, desafiadores e visionários

04

Ferramentas Complementares

Excel, Python, R e outras soluções relevantes

05

Tendências e Recomendações

Direções futuras e guias de seleção por cenário

A Importância da Visualização de Dados no Cenário Atual

A visualização de dados e o Business Analytics são fundamentais no cenário atual dos negócios digitais. As plataformas de Analytics e Business Intelligence (ABI) permitem às organizações modelar, analisar e visualizar dados complexos para apoiar a tomada de decisões informadas e criar valor estratégico sustentável.

Este material explora de forma abrangente os critérios essenciais para seleção de ferramentas, apresenta uma análise detalhada das principais opções disponíveis no mercado global, e oferece recomendações práticas baseadas em diferentes cenários organizacionais e necessidades específicas.



Insight de Kate Strachnyi: "Embora a ferramenta em si não seja tão importante quanto a capacidade de transmitir informações ao público, ela desempenha um papel fundamental ao ajudar as pessoas a formar visuais eficazes e impactantes."

Recursos Obrigatórios (Must-Have)

Visualização Avançada

- Suporte a visualizações customizadas e interativas
- Integração com bibliotecas de gráficos modernas
- Mapeamento geográfico robusto
- Recomendações inteligentes de tipos de visualização baseadas nos dados

Preparação de Dados

- Funcionalidades drag-and-drop intuitivas
- Combinação de dados de múltiplas fontes heterogêneas
- Criação de modelos analíticos complexos
- Gestão de medidas, pipelines, conjuntos e hierarquias



Recursos Obrigatórios (Must-Have)

Insights Automatizados

- Machine Learning integrado para descoberta automática
- Identificação de padrões e atributos importantes
- Previsões de séries temporais
- Detecção e análise de clusters

Consulta Natural (NLQ)

- Perguntas em linguagem natural digitada ou falada
- Interpretação contextual de consultas
- Sugestões inteligentes de análises
- Resposta automática com visualizações apropriadas



Conectividade e Integração de Dados


Conectividade de Fontes de Dados

A capacidade de conectar e ingerir dados de diversas plataformas é fundamental para o sucesso de qualquer solução de Business Analytics. As ferramentas modernas devem oferecer conectividade robusta tanto para ambientes on-premises quanto para infraestruturas de nuvem híbrida.

Fontes de Dados Essenciais:

- **Bancos Relacionais:** SQL Server, Oracle, MySQL, PostgreSQL
- **Data Warehouses:** Snowflake, Amazon Redshift, Google BigQuery
- **Armazenamento em Nuvem:** AWS S3, Azure Data Lake, Google Cloud Storage
- **APIs e Web Services:** REST, GraphQL, Salesforce, ServiceNow
- **Arquivos:** Excel, CSV, JSON, Parquet, Avro
- **Streaming:** Kafka, Event Hubs, Kinesis



 **Dica Técnica:** Priorize ferramentas que ofereçam conectores nativos para suas principais fontes de dados, reduzindo a complexidade de integração e melhorando a performance das consultas.

Recursos Recomendados (Nice-to-Have)

1

Camada de Métricas

Camada virtualizada que permite criar e definir métricas de negócio como código, garantindo consistência e governança across diferentes relatórios e dashboards. Esta funcionalidade é essencial para organizações que precisam de definições padronizadas de KPIs.

2

Analytics Embarcada

Integração seamless de capacidades analíticas diretamente em aplicações de negócio existentes, permitindo que usuários acessem insights sem sair do contexto de trabalho. Inclui APIs robustas e SDKs para desenvolvimento customizado.

3

Administração de Plataforma

Ferramentas abrangentes para rastreamento de estatísticas de uso, gerenciamento otimizado de custos, monitoramento proativo de performance e governança de dados. Essencial para ambientes empresariais complexos.

4

Geração de Linguagem Natural

Capacidade de produção automática de narrativas escritas ou faladas com controle granular sobre nível de detalhamento, tom comunicacional e audiência-alvo, transformando dados em histórias compreensíveis.

Fatores Críticos de Avaliação



Custo Total de Propriedade

Análise completa incluindo licenças, implementação, treinamento, manutenção e escalabilidade. Considere modelos por usuário, por capacidade ou híbridos, e custos ocultos como conectores premium ou recursos avançados.



Compatibilidade Mobile

Experiência responsiva e aplicações nativas para acesso mobile, considerando funcionalidades offline, sincronização de dados e interface otimizada para diferentes tamanhos de tela e contextos de uso.



Colaboração e Compartilhamento

Recursos robustos para publicação, compartilhamento seguro, colaboração em tempo real, controle de versões, comentários contextuais e workflows de aprovação para relatórios e dashboards.



Usabilidade e Adoção

Interface intuitiva que facilite a adoção por usuários com diferentes níveis técnicos. Avalie a curva de aprendizado, disponibilidade de templates e a capacidade de self-service analytics para usuários de negócio.



Ecossistema de Suporte

Qualidade do suporte técnico, documentação abrangente, comunidade ativa de usuários, disponibilidade de consultores especializados e recursos de treinamento e certificação profissional.



Escalabilidade e Performance

Capacidade de crescer com a organização, suportando volumes crescentes de dados, usuários concorrentes e complexidade analítica sem degradação significativa de performance ou experiência do usuário.

Magic Quadrant Gartner 2025: Visão Geral



Metodologia de Avaliação

O Magic Quadrant da Gartner de 2025 para Plataformas de Analytics e Business Intelligence posiciona 20 fornecedores principais em quatro quadrantes distintos, avaliando duas dimensões críticas:

Eixos de Avaliação:

- **Capacidade de Execução (Eixo Y):** Habilidade de entregar produtos e serviços, viabilidade financeira, responsividade ao mercado
- **Completeness da Visão (Eixo X):** Compreensão do mercado, estratégia de marketing, inovação, modelo de negócio

Esta análise considera fatores como penetração de mercado, crescimento de receita, satisfação do cliente, capacidade de execução técnica e visão estratégica para direções futuras do mercado ABI.

Posicionamento dos Fornecedores por Quadrante

Líderes

- **Google** (Looker)
- **Microsoft** (Power BI)
- **Oracle**
- **Qlik**
- **Salesforce** (Tableau)
- **ThoughtSpot**

Fornecedores com alta capacidade de execução e visão completa de mercado.

Desafiadores

- **Alibaba Cloud**
- **Amazon Web Services**
- **Domo**

Alta capacidade de execução, mas visão de mercado em desenvolvimento.

Visionários

- **IBM**
- **Pyramid Analytics**
- **SAP**
- **SAS**
- **Strategy**
- **Tellius**

Visão inovadora, mas capacidade de execução ainda em evolução.

Jogadores de Nicho

- **GoodData**
- **Incorta**
- **Sigma**
- **Sisense**
- **Zoho**

Focados em segmentos específicos ou mercados geográficos limitados.

Microsoft Power BI: Líder de Mercado

Posição: **Líder**

Pontos Fortes Estratégicos:

- **Dominância de Mercado:** Presença ubíqua facilita encontrar profissionais qualificados e recursos especializados no mercado
- **Stack Integrado D&A:** Microsoft Fabric oferece ecossistema completo (OneLake, Data Integration, Data Warehouse, Apache Spark)
- **IA Generativa:** Marca Copilot reconhecida mundialmente para impulsionar produtividade e democratizar analytics
- **Integração Office:** Sinergia natural com Excel, Teams, SharePoint e demais ferramentas Microsoft

Pontos de Atenção:

- **Aumentos de Preços:** Power BI Pro subiu de \$10 para \$14/mês; Premium de \$20 para \$24/mês
- **Limitação Azure:** Algumas funcionalidades avançadas restritas ao ecossistema Azure
- **Isolamento de Workload:** Desafios em ambientes multi-tenant complexos



Market Share: Power BI mantém posição dominante com crescimento consistente em adoção empresarial, especialmente em organizações já investidas no ecossistema Microsoft.



Salesforce Tableau: Excelência em Visualização

Posição: **Líder**

Forças Competitivas:

- **Insights Automatizados:** Einstein Discovery oferece ML avançado, Data Guide para exploração inteligente, Tableau Pulse para monitoramento proativo
- **Flexibilidade Multi-Cloud:** Deployment escalável em AWS, Azure, Google Cloud sem vendor lock-in
- **Comunidade Robusta:** Ecossistema vibrante com recursos extensivos de aprendizado e best practices
- **Visualização Avançada:** Capacidades superiores para storytelling visual complexo

Limitações Identificadas:

- **Interoperabilidade:** Catálogo analítico não integra com outras plataformas BI existentes
- **NLQ Limitado:** Consultas em linguagem natural carecem de sugestões preditivas em tempo real
- **Custo Premium:** Licenciamento pode ser proibitivo para organizações com orçamentos limitados
- **Complexidade Técnica:** Curva de aprendizado íngreme para usuários não-técnicos

Google Looker: Governança e Semântica

Posição: **Líder**

Diferenciação Principal

Excelência incomparável em governança analítica e administração de plataforma empresarial. Arquitetura fundamentalmente diferente baseada em modelos semânticos como código.

Arquitetura Robusta

Sistema aberto com camada semântica como fonte única de verdade (single source of truth). LookML permite definições reutilizáveis e versionamento de lógica de negócio, garantindo consistência organizacional.

Modelo Comercial

Precificação simplificada com capacidades de IA Generativa integradas nativamente. Sem custos adicionais para recursos básicos de GenAI, diferentemente de concorrentes premium.

Desafios Estratégicos

Preços premium podem ser barreira para grandes implementações. Frequentemente percebido como ferramenta para desenvolvedores, limitando adoção por usuários de negócio. Capacidades voltadas ao consumidor ainda em desenvolvimento.

ThoughtSpot: Search-Driven Analytics



Posição: **Líder**

Capacidades Distintivas:

- **SpotIQ:** Engine de insights automatizados que identifica padrões, anomalias e correlações sem intervenção manual
- **Spotter (NLQ):** Interface de busca em linguagem natural mais avançada do mercado, com compreensão contextual
- **Smart Apps:** Produtos de análise embarcada para integração seamless em aplicações existentes
- **Momentum GenAI:** Crescente interesse impulsionado por capacidades de IA generativa native

Pontos de Melhoria:

- **Catálogo Analítico:** Lacunas em recursos de catalogação e descoberta de conteúdo
- **Ecosistema Independente:** Fornecedor "puro" sem suite própria de aplicações empresariais
- **Diferenciação:** Desafios em mercado cada vez mais concorrido com recursos similares

Inovação Disruptiva: ThoughtSpot pioneering search-driven analytics está redefinindo como usuários interagem com dados, eliminando barreiras técnicas tradicionais.

Amazon Web Services QuickSight: Desafiador

Integração AWS Nativa

Integração perfeita e otimizada com todo o stack de Data & Analytics da AWS, incluindo Redshift, S3, Athena, Glue, SageMaker. Performance superior para workloads já na nuvem AWS.

Pricing Transparente

Modelo de precificação simples e competitivo: \$3/usuário/mês para readers, \$18/usuário/mês para authors. Sem custos ocultos ou taxas de conectores premium.

Arquitetura Serverless

Performance escalável com engine SPICE (Super-fast, Parallel, In-memory Calculation Engine) e capacidade de auto-scaling sem gerenciamento de infraestrutura.

Limitações Estratégicas

Deployment otimizado exclusivamente para AWS limita flexibilidade multi-cloud. Ausência de camada de métricas universal e suite de aplicações de negócio integradas.

Domo: Plataforma Cloud-Native

Posição: Desafiador

Pontos Fortes Operacionais:

- **Magic ETL:** Preparação visual de dados com interface drag-and-drop intuitiva
- **Conectividade Massiva:** Mais de 1.000 conectores pré-construídos para fontes diversificadas
- **Domo Everywhere:** Capacidades robustas de embedding em portais e aplicações corporativas
- **Agentes de IA:** Investimento estratégico em automação inteligente e assistentes virtuais

Desafios de Mercado:

- **Geografia Limitada:** Presença concentrada principalmente no mercado norte-americano
- **Pricing Complexo:** Modelo baseado em consumo pode ser difícil de prever e gerenciar
- **NLQ Básico:** Recursos de consulta em linguagem natural menos desenvolvidos que líderes

Domo posiciona-se como plataforma cloud-native com foco em democratização de dados através de interfaces amigáveis e conectividade extensiva. A estratégia de "data experience platform" visa eliminar silos organizacionais.



Tellius: Visionário em NLQ e AutoML

1

Kaiya Assistant

Assistente de IA conversacional que democratiza analytics avançados através de interfaces de linguagem natural sofisticadas e compreensão contextual profunda.

2

Vertical Strategy

Especialização estratégica em indústrias específicas como farmacêutica e CPG (Consumer Packaged Goods), oferecendo soluções pré-configuradas e KPIs específicos.

3

Customer-Centric

Foco intenso na experiência do usuário e facilidade de adoção, com interfaces intuitivas que reduzem significativamente a curva de aprendizado para usuários de negócio.

- ⊗ **Limitações Geográficas:** Estratégia de expansão ainda emergente, com maioria da receita concentrada no mercado americano. Momentum de mercado reduzido comparado aos líderes estabelecidos.



IBM Cognos Analytics: Visionário Empresarial

Posição: Visionário

Fortalezas Históricas:

- **Relatórios Empresariais:** Capacidades robustas e altamente customizáveis para relatórios complexos e compliance regulatório
- **IBM watsonx BI:** Investimentos significativos em IA com integração da plataforma Watson para analytics avançados
- **Vertical Expertise:** Soluções especializadas para indústrias regulamentadas como financeira, saúde e governo
- **Governança Avançada:** Recursos Enterprise-grade para segurança, auditoria e compliance

Gaps Competitivos:

- **Preparação de Dados:** Atraso em recursos de preparação habilitados por IA comparado aos líderes
- **Modern UX:** Interface menos intuitiva para usuários acostumados com soluções contemporâneas
- **Market Momentum:** Crescimento mais lento em novos clientes e mindshare de mercado

❏ **Legacy Advantage:** Para organizações com investimentos significativos em IBM, Cognos oferece continuidade e integração natural com o portfólio Watson AI e Cloud Pak for Data.



Sigma Computing: Jogador de Nicho Inovador

1

Interface Spreadsheet-Like

Diferenciação única com interface familiar de planilha que reduz drasticamente a barreira de entrada para usuários de Excel. Combina facilidade de uso com poder analítico de plataformas enterprise.

2

Alpha Query Engine

Performance e escalabilidade excepcionais através de engine proprietário otimizado para consultas complexas em grandes volumes de dados. Arquitetura cloud-native para elasticidade.

3

Embedded Analytics

Capacidades robustas de embedding que permitem integração seamless em aplicações customer-facing e portais internos com APIs flexíveis e SDKs abrangentes.

4

Limitações de Mercado

Ausência de ecossistema cloud empresarial próprio limita integração nativa. Concorrência acirrada em mercado maduro com players estabelecidos. Capacidades de NLQ e NLG ainda em desenvolvimento.

Microsoft Excel: A Ferramenta Universal



Relevância Contínua no Ecossistema Analytics

Embora não seja uma plataforma ABI dedicada, o Excel mantém relevância crítica no panorama de Business Analytics por várias razões estratégicas:

Vantagens Intrínsecas:

- **Universalidade:** Primeira ferramenta de visualização para milhões de profissionais globalmente
- **Familiaridade:** Base conhecida que facilita transição para ferramentas mais avançadas
- **Flexibilidade:** Liberdade total para análises ad-hoc e prototipagem rápida
- **Power Query/Pivot:** Capacidades nativas robustas para preparação e análise de dados

Integração Estratégica:

- **Power BI Desktop:** Transição natural para analytics enterprise
- **Office 365:** Integração seamless com Teams, SharePoint, OneDrive
- **Power Platform:** Gateway para Power Apps e Power Automate

Ferramentas de Programação: Python e R



Python (Plotly/Dash)

Vantagens: Máxima flexibilidade e customização, ecossistema rico (pandas, numpy, scikit-learn), custos reduzidos (open source), integração com ML pipelines, comunidade ativa e abundante documentação.

Requisitos: Conhecimento técnico significativo, necessidade de infraestrutura própria, desenvolvimento e manutenção customizados, curva de aprendizado íngreme para não-programadores.



R (Shiny/ggplot2)

Fortalezas: Excelência em análise estatística avançada, visualizações científicas superiores via ggplot2, Shiny para aplicações web interativas, forte presença em academia e pesquisa.

Considerações: Performance limitada para grandes datasets, menor adoção empresarial comparado ao Python, recursos de produção menos robustos que plataformas comerciais.

Cenários Ideais: Organizações com equipes técnicas fortes, necessidades analíticas altamente customizadas, orçamentos limitados para licenças comerciais, ou requisitos específicos não atendidos por soluções comerciais.

Impacto Transformador da IA Generativa

Revolução em Analytics: Como a GenAI Está Redefinindo o Mercado

01

Análises Conversacionais

Interfaces de linguagem natural que permitem fazer perguntas complexas aos dados como se conversando com um analista especializado. Democratização total do acesso a insights avançados.

02

Geração Automática de Conteúdo

Criação automatizada de relatórios, dashboards, narrativas explicativas e apresentações executivas com base em padrões identificados nos dados e preferências organizacionais.

03

Assistentes Inteligentes

Agentes de IA que atuam como consultores analíticos 24/7, oferecendo recomendações proativas, alertas contextuais e sugestões de análises baseadas em comportamento histórico.

04

Democratização Analítica

Eliminação de barreiras técnicas tradicionais, permitindo que qualquer usuário de negócio realize análises sofisticadas sem conhecimento de SQL, Python ou ferramentas especializadas.



Mudanças no Panorama Competitivo

Forças Disruptivas Emergentes

Novos Entrantes:

- **Fornecedores DBMS:** Snowflake, Databricks, BigQuery expandindo para analytics
- **Aplicações AI-Native:** Startups especializadas em NLQ e automated insights
- **Hyperscalers:** AWS, Azure, GCP intensificando investimentos em BI nativo

Preocupações do Mercado:

- **Vendor Lock-in:** Crescente dependência de ecossistemas específicos de nuvem
- **Transparência de Preços:** Demanda por modelos mais simples e previsíveis
- **Interoperabilidade:** Necessidade de integração entre multiple platforms

Tendências de Consolidação

- **Aquisições Estratégicas:** Grandes players adquirindo inovadores de nicho
- **Partnerships:** Alianças para complementar capacidades tecnológicas
- **Platform Plays:** Movimento em direção a plataformas integradas end-to-end
- **Specialized Solutions:** Crescimento de soluções verticais específicas



Recomendações por Tipo de Organização

Organizações Empresariais (1000+ usuários)

- **Microsoft Power BI:** Para organizações já investidas no ecossistema Microsoft 365/Azure
- **Tableau:** Para necessidades avançadas de visualização e storytelling complexo
- **Looker:** Para governança robusta e modelos semânticos como código
- **Oracle/SAP:** Para integração com ERP existente e compliance rigoroso

Casos de Uso Específicos

- **ThoughtSpot:** Para democratização através de search-driven analytics
- **Qlik:** Para análise associativa e exploração interativa de dados
- **Sigma:** Para usuários intensivos de planilhas que resistem à mudança
- **Domo:** Para conectividade massiva e embedding em aplicações

Organizações com Orçamento Limitado

- **Power BI:** Melhor custo-benefício para recursos enterprise (\$14/mês/usuário)
- **Zoho Analytics:** Opção de baixo custo para PMEs (\$4.50/mês/usuário)
- **AWS QuickSight:** Para workloads já na AWS (\$3/mês/reader)
- **Open Source (Python/R):** Para organizações com capacidade técnica interna

Conclusões e Próximos Passos

A escolha da ferramenta de visualização de dados adequada deve ser baseada em uma avaliação holística que considere não apenas capacidades técnicas, mas também fatores organizacionais, culturais e estratégicos de longo prazo.

1

Não Existe Solução Única

Cada organização possui necessidades específicas que determinam a ferramenta ideal. O Magic Quadrant da Gartner fornece direcionamento valioso, mas a decisão final deve sempre considerar contexto específico e objetivos estratégicos únicos.

2

IA Generativa é Game-Changer

A integração de capacidades de IA Generativa está democratizando analytics avançados e redefinindo expectativas de usuário. Organizações devem priorizar fornecedores com roadmaps claros de GenAI.

3

Foco na Adoção Usuário

Como enfatizado por Kate Strachnyi, o mais importante é a capacidade de transmitir informações e insights impactantes ao público, garantindo que a mensagem desejada seja compreendida e acionável, independentemente da ferramenta utilizada.


Criando uma análise exploratória de dados com o ChatGPT e Google Colab



Escolha e baixe um dataset

Você pode recorrer ao repositório de dataset como o UCI por exemplo:

<https://archive.ics.uci.edu/dataset/222/bank+marketing>



Bank Marketing

Donated on 2/13/2012

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

Dataset Characteristics	Subject Area	Associated Tasks
Multivariate	Business	Classification
Feature Type	# Instances	# Features
Categorical, Integer	45211	16

Dataset Information

Additional Information

The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed. ...

SHOW MORE

Has Missing Values?

No

Introductory Paper

[A data-driven approach to predict the success of bank telemarketing](#)

By Sérgio Moro, P. Cortez, P. Rita. 2014

Published in Decision Support Systems

DOWNLOAD (999.8 KB)

IMPORT IN PYTHON

CITE

9 citations

502754 views

Creators

- S. Moro
- P. Rita
- P. Cortez

DOI

10.24432/C5K306

License

This dataset is licensed under a [Creative Commons Attribution 4.0 International](#) (CC BY 4.0) license.

This allows for the sharing and adaptation of the datasets for any purpose, provided that the appropriate credit is given.

Entre no chat GPT e pergunte

ChatGPT ▾

Eu gostaria de fazer uma análise de exploratória de dados com o chat GPT pode explicar como fazer isso ?



Claro! 🚀

Análise Exploratória de Dados (ou **EDA – Exploratory Data Analysis**) é o primeiro passo em qualquer projeto de ciência de dados. A ideia é **entender a estrutura, padrões e possíveis problemas nos dados** antes de aplicar modelos.

Você pode fazer essa análise com o ChatGPT de duas formas:

1. **Conversando sobre seus dados** (colando amostras ou descrevendo) → eu ajudo a interpretar.
 2. **Rodando código em Python** (ex.: com Pandas, Matplotlib, Seaborn, etc.) → eu escrevo os scripts e você executa localmente.
-

Entre no chat GPT e pergunte

Passos típicos de uma EDA

1. Entender o dataset

- Quais são as colunas?
- Qual o tipo de cada variável (numérica, categórica, temporal, texto)?
- Quantas linhas existem?

👉 Em Python:

python

 Copiar código

```
import pandas as pd

df = pd.read_csv("seus_dados.csv") # ou Excel etc.
print(df.shape)
print(df.info())
print(df.head())
```

Entre no chat GPT e pergunte

2. Estatísticas descritivas

- Médias, medianas, desvio padrão
- Valores máximos e mínimos
- Distribuição de cada variável

👉 Em Python:

python

 Copiar código

```
print(df.describe(include="all"))
```

Entre no chat GPT e pergunte

3. Qualidade dos dados

- Há valores nulos?
- Há duplicatas?
- Algum outlier?

👉 Em Python:

python

 Copiar código

```
print(df.isnull().sum())  
print(df.duplicated().sum())
```

Entre no chat GPT e pergunte

4. Visualizações iniciais

- Histogramas → ver distribuição
- Boxplots → outliers
- Scatterplots → relação entre duas variáveis
- Heatmap de correlação → dependência entre variáveis numéricas

👉 Exemplo:

python

 Copiar código

```
import matplotlib.pyplot as plt
import seaborn as sns

# Histograma
df["coluna_numerica"].hist(bins=30)
plt.show()

# Boxplot
sns.boxplot(x=df["coluna_numerica"])
plt.show()

# Heatmap
sns.heatmap(df.corr(), annot=True, cmap="coolwarm")
plt.show()
```


Entre no chat GPT e pergunte

5. Exploração de variáveis categóricas

- Frequência de cada categoria
- Relação com variáveis numéricas

👉 Exemplo:

python

 Copiar código

```
print(df["coluna_categorica"].value_counts())  
sns.boxplot(x="coluna_categorica", y="coluna_numerica", data=df)  
plt.show()
```

6. Insights iniciais

- Quais variáveis parecem mais importantes?
- Existem padrões, tendências ou anomalias?
- O dataset precisa de limpeza antes de seguir?

👉 Com o ChatGPT, você pode:

- Me mostrar um **trecho dos dados** (umas 20 linhas já ajuda).
- Pedir para eu **gerar o código da análise**.
- E depois você executa no seu ambiente (Jupyter Notebook, VSCode, Colab, etc.).

Agora um prompt mais elaborado

5. Exploração de variáveis categóricas

- Frequência de cada categoria
- Relação com variáveis numéricas

👉 Exemplo:

python

 Copiar código

```
print(df["coluna_categorica"].value_counts())  
sns.boxplot(x="coluna_categorica", y="coluna_numerica", data=df)  
plt.show()
```

6. Insights iniciais

- Quais variáveis parecem mais importantes?
- Existem padrões, tendências ou anomalias?
- O dataset precisa de limpeza antes de seguir?

👉 Com o ChatGPT, você pode:

- Me mostrar um **trecho dos dados** (umas 20 linhas já ajuda).
- Pedir para eu **gerar o código da análise**.
- E depois você executa no seu ambiente (Jupyter Notebook, VSCode, Colab, etc.).

Agora um prompt mais elaborado

Você é um analista de dados experiente com extensos conhecimentos em python e estatística. Sua tarefa é realizar uma análise exploratória de dados bastante detalhada referente aos dados amostrados em anexo que referem-se a campanhas de marketing direto de uma instituição bancária portuguesa.

As campanhas de marketing baseavam-se em chamadas telefónicas. Muitas vezes, era necessário mais do que um contacto com o mesmo cliente para verificar se o produto (depósito a prazo bancário) seria ('sim') ou não ('não') subscrito.

Eu quero que você gere um código para que eu execute no Google Colab, esse código deve facilitar o upload dos dados e a geração de um relatório em pdf para baixar.

A análise deve envolver todas as variáveis, mostrar distribuições das mesmas, suas correlações e quando houverem correlações fortes mostrar as distribuições bivariadas

Faça comentários e traga insights sobre estes dados.

Passe-me instruções sobre as etapas a realizar

A resposta

Plano de Análise Detalhada

1. Carregamento e inspeção inicial

- Dimensão do dataset
- Tipos de variáveis
- Valores ausentes ou categorias "desconhecidas"

2. Análise univariada

- Distribuição das variáveis numéricas (`age`, `balance`, `duration`, `campaign`, `pdays`).
- Frequência das variáveis categóricas (`job`, `marital`, `education`, `housing`, `loan`, `month`, `outcome`).
- Variável alvo (`accept`).

3. Análise bivariada

- Relação entre a variável alvo (`accept`) e cada variável explicativa.
- Gráficos de barras para categóricas x target.
- Boxplots e histogramas comparativos para numéricas x target.

4. Correlação entre variáveis numéricas

- Heatmap de correlação.
- Scatterplots apenas para correlações fortes.

5. Insights específicos sobre o dataset

- Fatores mais relacionados à aceitação da campanha.
- Possíveis problemas de balanceamento da variável alvo.
- O impacto da duração da chamada (`duration`), do histórico (`previous`, `pdays`) e da situação financeira (`balance`).


6. Exportação de relatório em PDF

- Contendo gráficos, tabelas e comentários.

Código Colab-ready (completo e detalhado)

python

 Copiar código


```
# =====  
#  EDA Detalhada - Campanhas de Marketing Bancário  
# =====  
  
# 1. Instalação de pacotes  
!pip install pandas matplotlib seaborn reportlab  
  
# 2. Importação das bibliotecas  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from reportlab.platypus import SimpleDocTemplate, Paragraph, Spacer, Image  
from reportlab.lib.styles import getSampleStyleSheet  
from reportlab.lib.pagesizes import A4  
from reportlab.lib.units import inch  
import os
```

.1.

A resposta

Código Colab-ready (completo e detalhado)

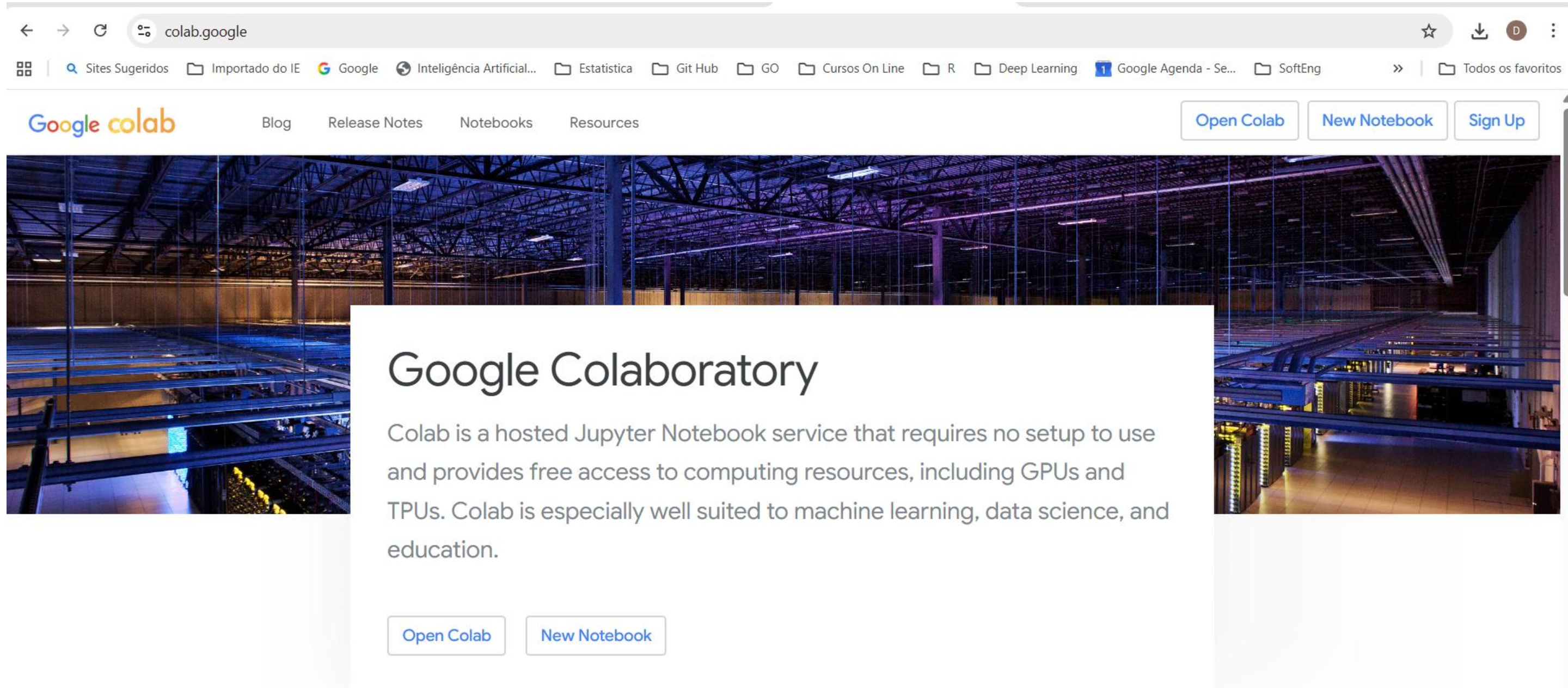
python

 Copiar código

```
# =====  
#  EDA Detalhada - Campanhas de Marketing Bancário  
# =====  
  
# 1. Instalação de pacotes  
!pip install pandas matplotlib seaborn reportlab  
  
# 2. Importação das bibliotecas  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from reportlab.platypus import SimpleDocTemplate, Paragraph, Spacer, Image  
from reportlab.lib.styles import getSampleStyleSheet  
from reportlab.lib.pagesizes import A4  
from reportlab.lib.units import inch  
import os
```

Vá ao Google Colab

<https://colab.google/>



The image is a screenshot of the Google Colab website. At the top, there is a browser address bar showing 'colab.google'. Below it is a navigation bar with the Google Colab logo on the left and three buttons: 'Open Colab', 'New Notebook', and 'Sign Up' on the right. The main content area features a large background image of a server room with blue lighting. Overlaid on this image is a white text box containing the heading 'Google Colaboratory' and a paragraph describing the service. At the bottom of this text box are two buttons: 'Open Colab' and 'New Notebook'.

Google colab

Blog Release Notes Notebooks Resources

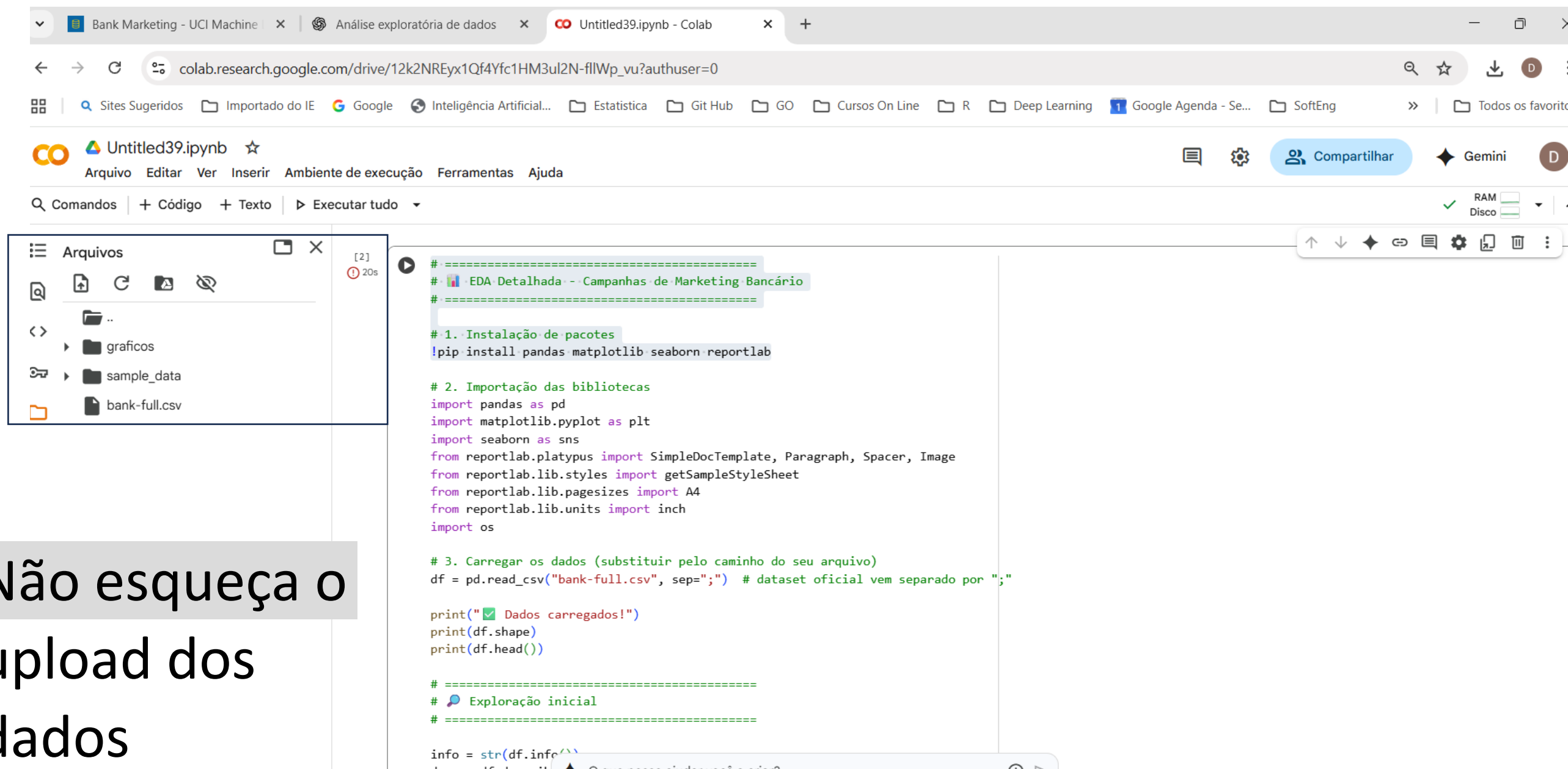
Open Colab New Notebook Sign Up

Google Colaboratory

Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education.

Open Colab New Notebook

Vá ao Google Colab crie um novo notebook e copie e cole o Código Gerado pelo GPT



The screenshot displays the Google Colab web interface. The browser's address bar shows the URL `colab.research.google.com/drive/12k2NREyx1Qf4Yfc1HM3ul2N-flIWp_vu?authuser=0`. The notebook is titled "Untitled39.ipynb". The left sidebar, labeled "Arquivos", shows a file explorer with a folder named "graficos" and a file named "bank-full.csv". The main area contains Python code for data analysis, including package installation, library imports, and data loading. The code is as follows:

```
# =====  
# EDA Detalhada - Campanhas de Marketing Bancário  
# =====  
  
# 1. Instalação de pacotes  
!pip install pandas matplotlib seaborn reportlab  
  
# 2. Importação das bibliotecas  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
from reportlab.platypus import SimpleDocTemplate, Paragraph, Spacer, Image  
from reportlab.lib.styles import getSampleStyleSheet  
from reportlab.lib.pagesizes import A4  
from reportlab.lib.units import inch  
import os  
  
# 3. Carregar os dados (substituir pelo caminho do seu arquivo)  
df = pd.read_csv("bank-full.csv", sep=";") # dataset oficial vem separado por ";"  
  
print("✅ Dados carregados!")  
print(df.shape)  
print(df.head())  
  
# =====  
# Exploração inicial  
# =====  
  
info = str(df.info())
```

Não esqueça o upload dos dados

Vá ao Google Colab crie um novo notebook e copie e cole o Código Gerado pelo GPT

```
-----  
ValueError                                Traceback (most recent call last)  
/tmp/ipython-input-3037652580.py in <cell line: 0>()  
    76  
    77 plt.figure(figsize=(12,8))  
--> 78 sns.heatmap(df.corr(), annot=True, cmap="coolwarm", center=0)  
    79 plt.title("Matriz de Correlação")  
    80 plt.tight_layout()  
  
----- 3 frames -----  
/usr/local/lib/python3.12/dist-packages/pandas/core/internals/managers.py in _interleave(self, dtype, na_value)  
    1751         else:  
    1752             arr = blk.get_values(dtype)  
-> 1753             result[rl.indexer] = arr  
    1754             itemmask[rl.indexer] = 1  
    1755  
  
ValueError: could not convert string to float: 'management'  
  
<Figure size 1200x800 with 0 Axes>  
-----
```


Volte ao GPT

Executei o código e encontrei o seguinte erro:

ValueError Traceback (most recent call last)

/tmp/ipython-input-3037652580.py in <cell line: 0>()

76

77 plt.figure(figsize=(12,8))

---> 78 sns.heatmap(df.corr(), annot=True, cmap="coolwarm",
center=0)

79 plt.title("Matriz de Correlação")

80 plt.tight_layout()

3 frames

/usr/local/lib/python3.12/dist-

packages/pandas/core/internals/managers.py in _interleave(self, dtype,
na_value)

1751 else:

1752 arr = blk.get_values(dtype)

-> 1753 result[rl.indexer] = arr

1754 itemmask[rl.indexer] = 1

1755

ValueError: could not convert string to float: 'management'

<Figure size 1200x800 with 0 Axes>

Volte ao GPT

Esse erro aconteceu porque o `df.corr()` só funciona com variáveis **numéricas** e o seu dataset contém várias variáveis categóricas (como `job`, `marital`, `education`, etc.).

O Pandas tenta calcular a correlação entre todas as colunas, mas não consegue converter strings em números.

👉 Temos duas soluções:

🔧 Ajuste no código da correlação

Substituímos:

python

 Copiar código

```
sns.heatmap(df.corr(), annot=True, cmap="coolwarm", center=0)
```

por:

python

 Copiar código

```
# Seleciona apenas variáveis numéricas
numeric_df = df.select_dtypes(include=['int64', 'float64'])

plt.figure(figsize=(12,8))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", center=0)
plt.title("Matriz de Correlação (variáveis numéricas)")
plt.tight_layout()
plt.show()
```

Volte ao GPT

Esse erro aconteceu porque o `df.corr()` só funciona com variáveis **numéricas** e o seu dataset contém várias variáveis categóricas (como `job`, `marital`, `education`, etc.).

O Pandas tenta calcular a correlação entre todas as colunas, mas não consegue converter strings em números.

👉 Temos duas soluções:

🔧 Ajuste no código da correlação

Substituímos:

python

 Copiar código

```
sns.heatmap(df.corr(), annot=True, cmap="coolwarm", center=0)
```

por:

python

 Copiar código

```
# Seleciona apenas variáveis numéricas
numeric_df = df.select_dtypes(include=['int64', 'float64'])

plt.figure(figsize=(12,8))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", center=0)
plt.title("Matriz de Correlação (variáveis numéricas)")
plt.tight_layout()
plt.show()
```

Faça a Mudança Sugerida

```
# =====  
# 📊 Correlações  
# =====  
  
plt.figure(figsize=(12,8))  
#sns.heatmap(df.corr(), annot=True, cmap="coolwarm", center=0)  
# Seleciona apenas variáveis numéricas  
numeric_df = df.select_dtypes(include=['int64', 'float64'])  
  
plt.figure(figsize=(12,8))  
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", center=0)  
plt.title("Matriz de Correlação (variáveis numéricas)")  
plt.tight_layout()  
plt.savefig("graficos/correlacao.png")  
plt.close()
```


Relatório pronto para download depois de 20 s

Comandos | + Código | + Texto | ▶ Executar tudo

Arquivos

gráficos

sample_data

EDA_Marketing_Completo.pdf

bank-full.csv

[3] ✓ 21s ▶

```
if df[col].dtype in ["int64", "float64"]:
    sns.boxplot(x=df[target], y=df[col])
    plt.title(f"{col} vs {target}")
else:
    sns.countplot(x=df[col], hue=df[target])
    plt.title(f"{col} vs {target}")
    plt.xticks(rotation=45)
plt.tight_layout()
plt.savefig(f"graficos/{col}_vs_target.png")
plt.close()

# =====
# 📊 Correlações
# =====

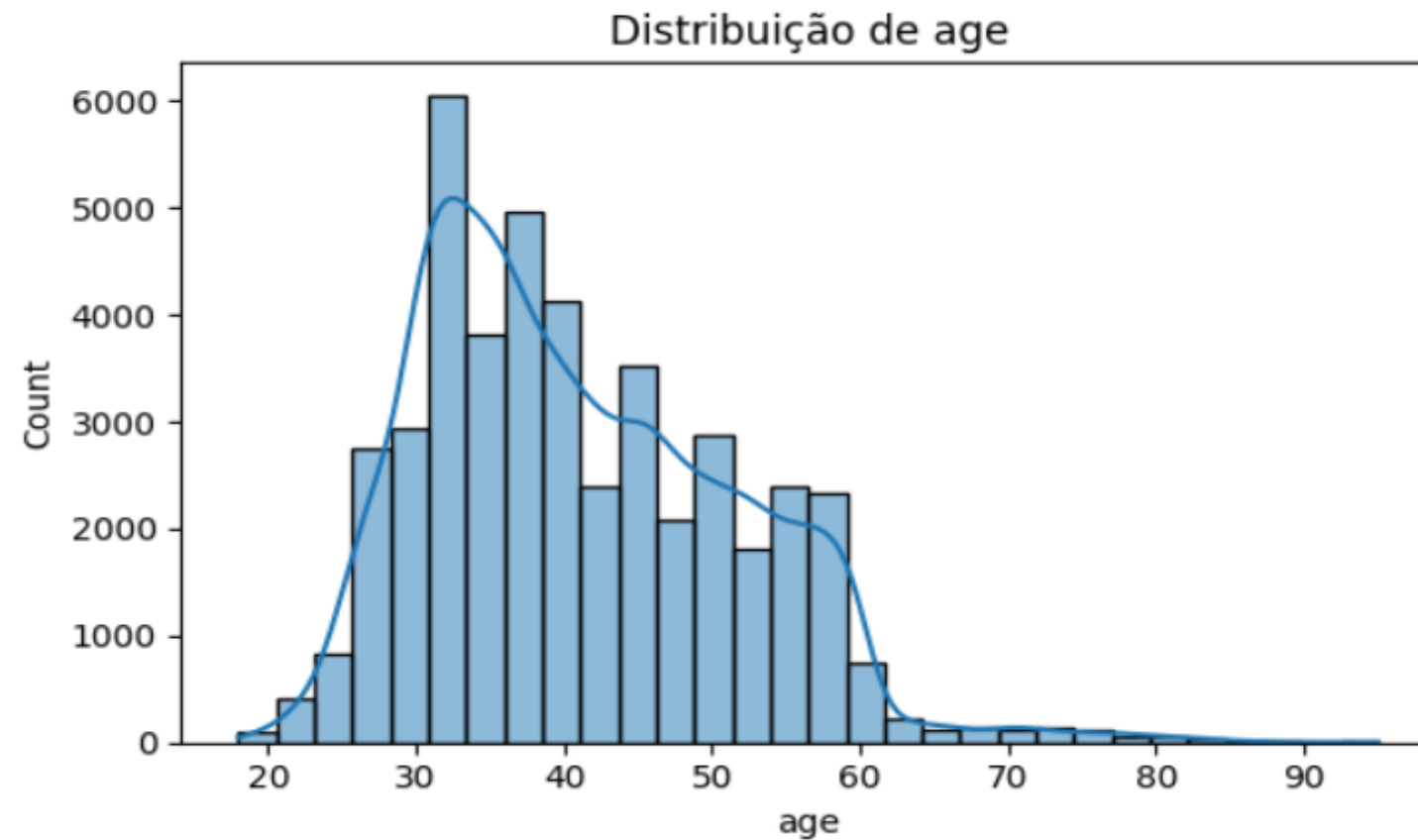
plt.figure(figsize=(12,8))
#sns.heatmap(df.corr(), annot=True, cmap="coolwarm", center=0)
# Seleciona apenas variáveis numéricas
numeric_df = df.select_dtypes(include=['int64', 'float64'])

plt.figure(figsize=(12,8))
sns.heatmap(numeric_df.corr(), annot=True, cmap="coolwarm", center=0)
plt.title("Matriz de Correlação (variáveis numéricas)")
plt.tight_layout()
plt.savefig("graficos/correlacao.png")
plt.close()

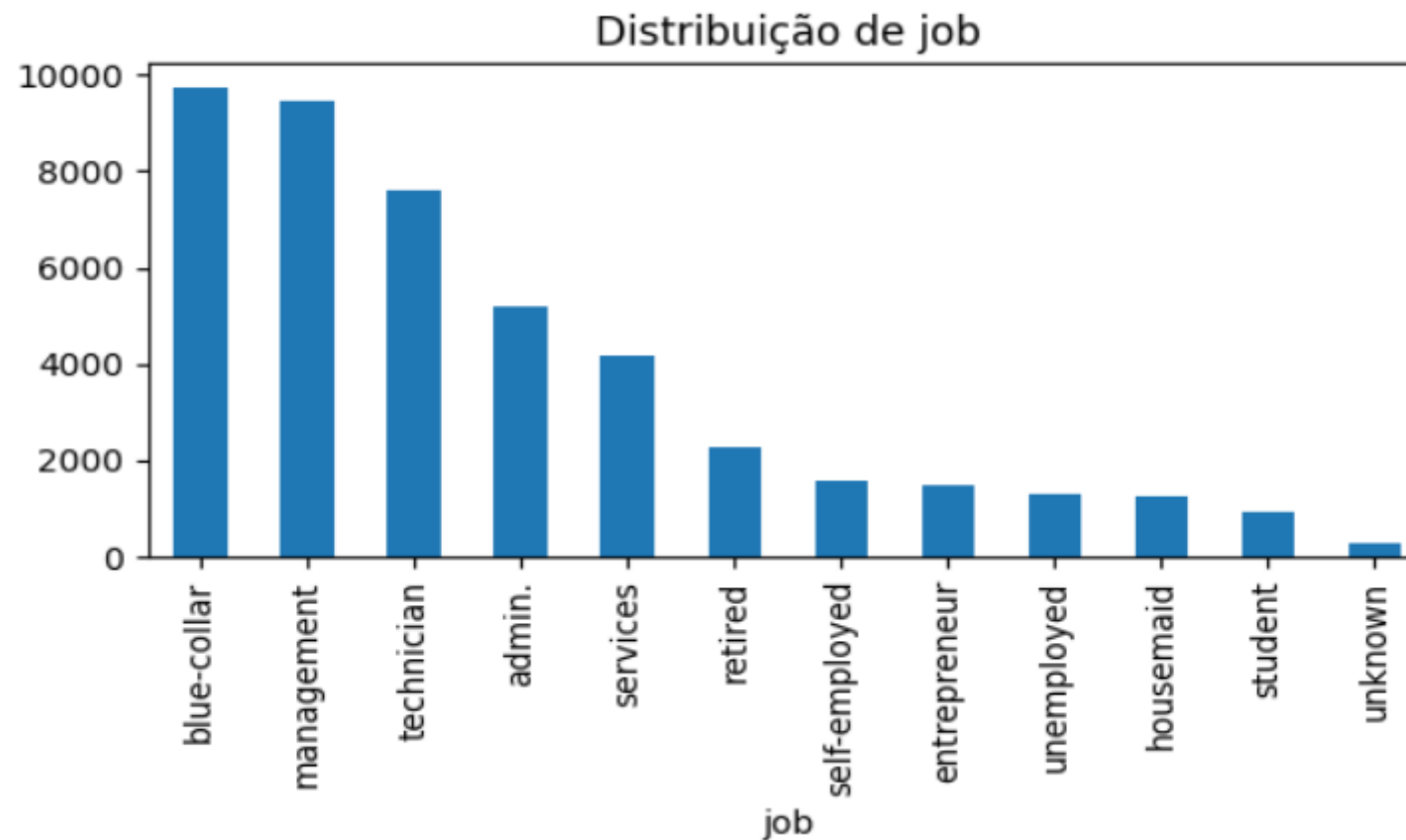
# =====
```

Relatório pronto para download depois de 20 s

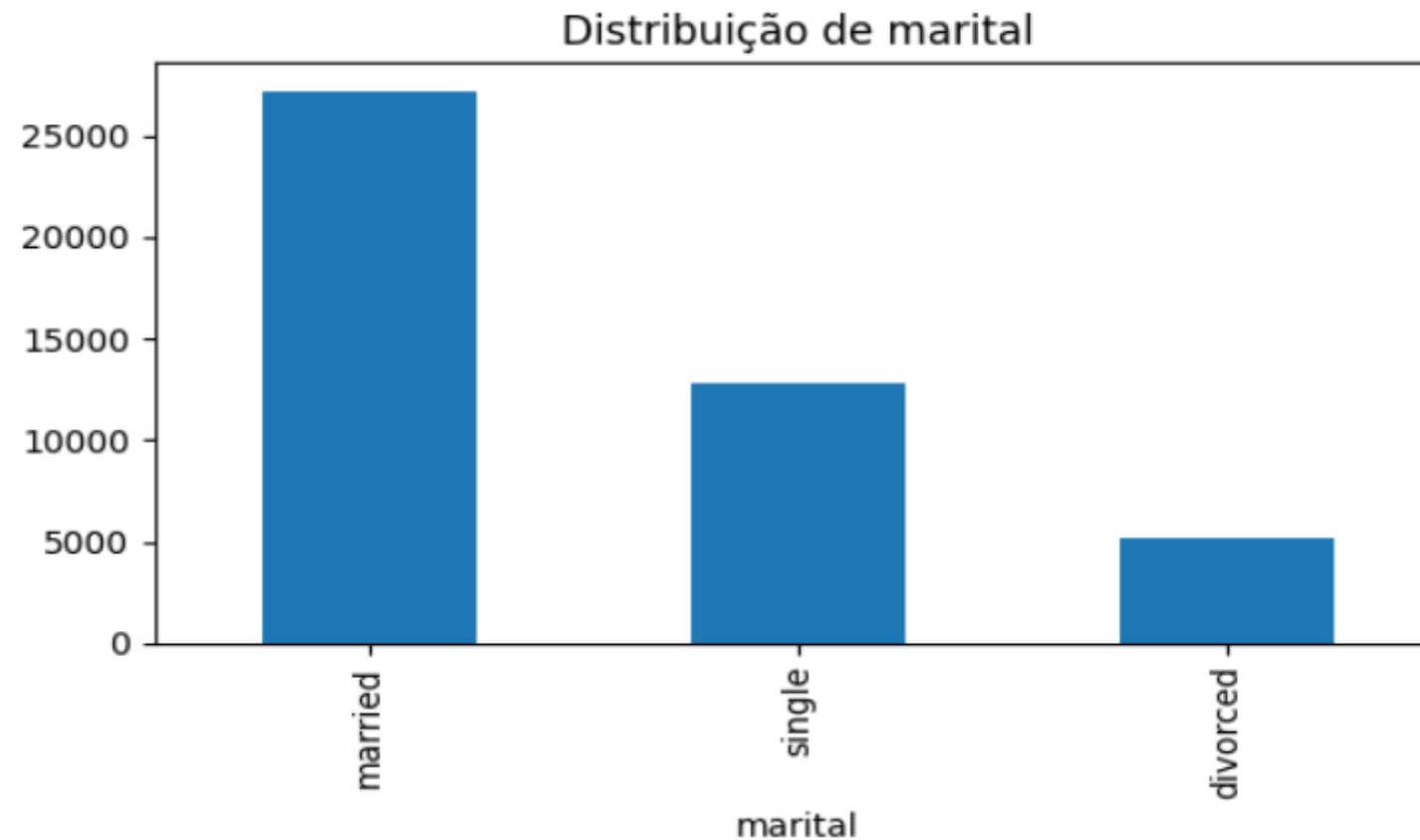
■ Distribuições das variáveis:



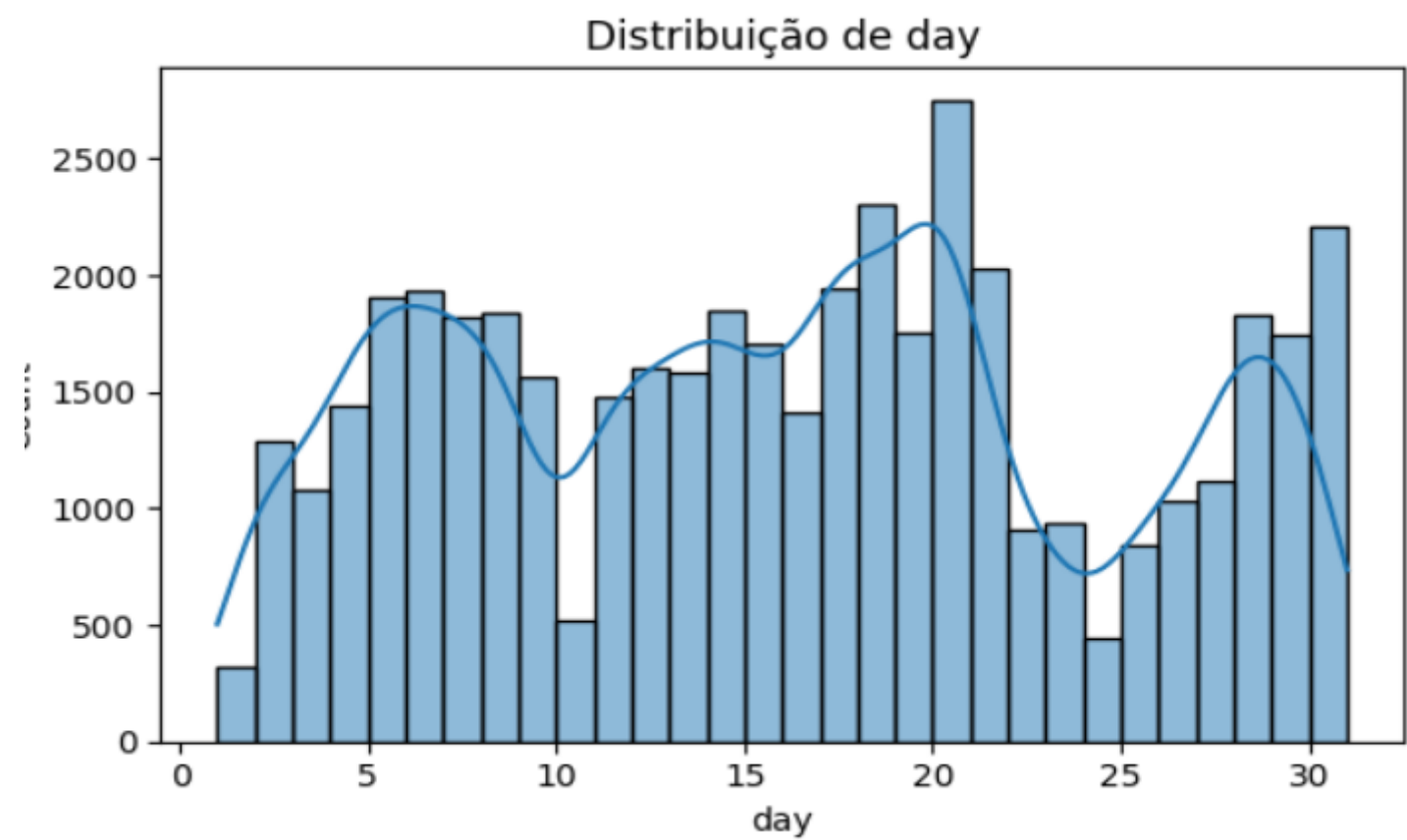
Relatório pronto para download depois de 20 s



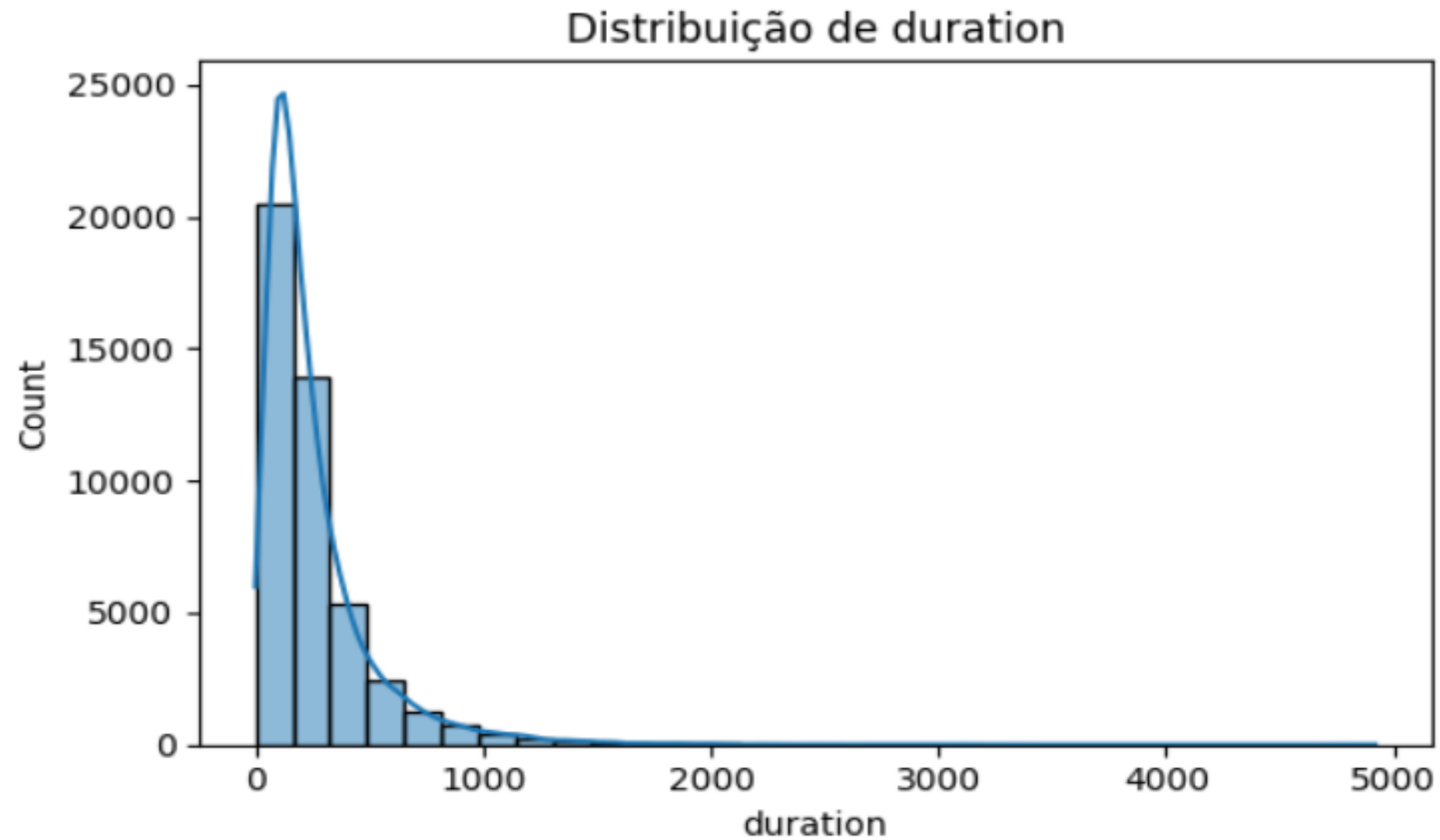
Relatório pronto para download depois de 20 s



Relatório pronto para download depois de 20 s

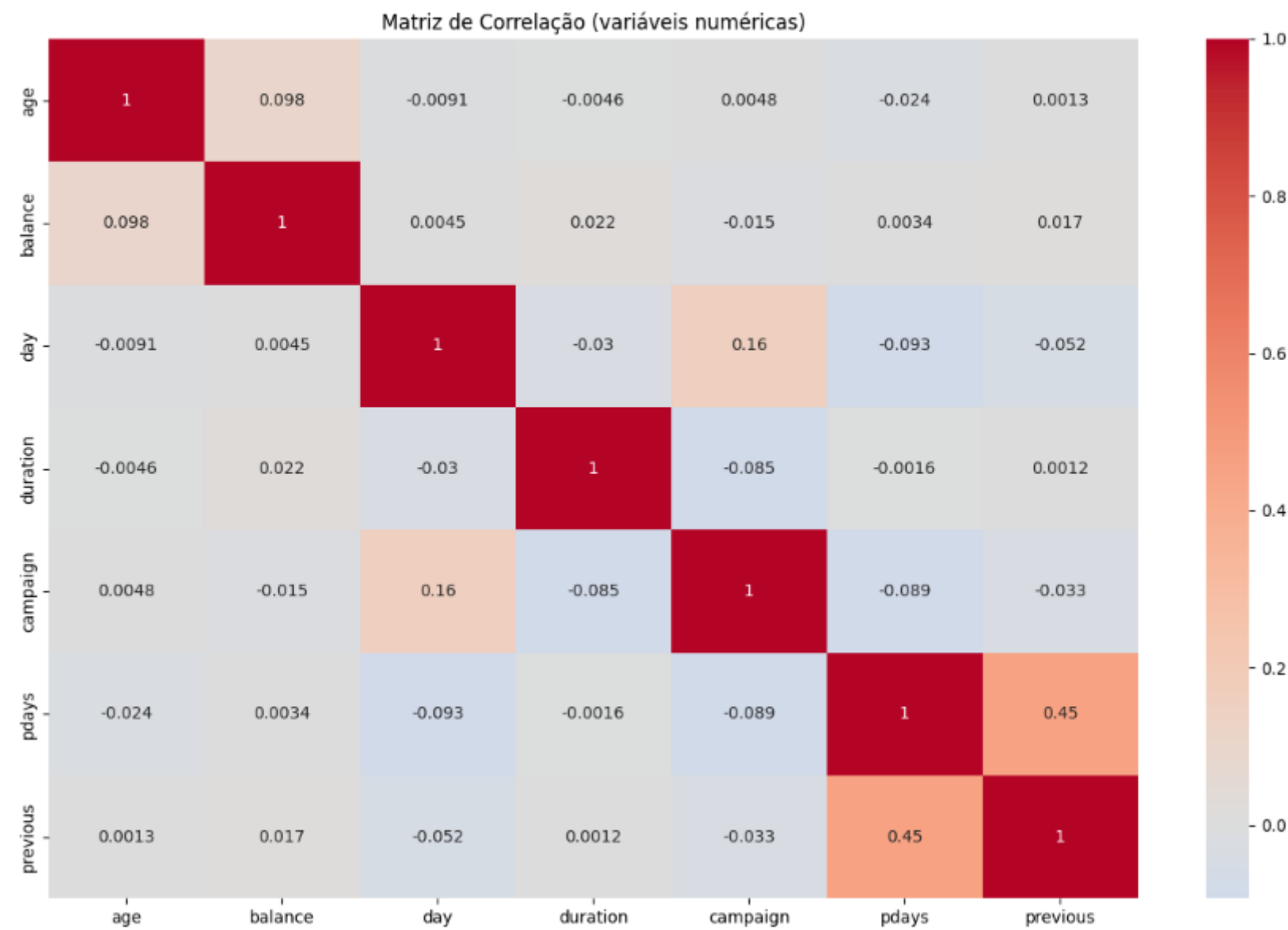


Relatório pronto para download depois de 20 s



Relatório pronto para download depois de 20 s

■ Correlação entre variáveis numéricas:



Relatório pronto para download depois de 20 s

■ Insights iniciais:

- A variável alvo está desbalanceada: muito mais 'não' do que 'sim'.
- A duração da chamada ('duration') é o fator mais fortemente associado ao sucesso da campanha.
- Clientes com histórico de contato prévio ('pdays' e 'previous') apresentam maior probabilidade de adesão.
- O saldo bancário ('balance') tem distribuição assimétrica: muitos clientes com saldos baixos ou negativos, poucos com saldos muito altos.
- Variáveis categóricas como profissão ('job'), estado civil ('marital') e nível de educação também impactam a probabilidade de aceitar o produto.
- O número de contatos na campanha atual ('campaign') tende a ser maior entre clientes que não aceitam.