Mariana Costa Almeida | Engenheira de Dados Sênior | Plataformas de Dados e Analytics

Rua dos Pinheiros, 450, Apto 82, São Paulo, SP, Brasil, 05422-000
+55 11 98765-4321 • mariana.costa.almeida@example.com
https://marianaalmeida.dev • @mariana-costa-almeida • @mariana-costa-data

Informações Básicas

- Engenheira de dados com mais de 6 anos de experiência na construção de plataformas de dados escaláveis, confiáveis e orientadas a impacto de negócio
- Especialista em pipelines batch e streaming, modelagem analítica, qualidade de dados e governança em ambientes de nuvem
- Experiência com Apache Spark, Airflow, dbt, Kafka, AWS, Databricks e ferramentas modernas de lakehouse
- Atuação colaborativa com times de produto, ciência de dados, BI e engenharia para traduzir necessidades analíticas em soluções robustas
- Forte compromisso com boas práticas de engenharia, observabilidade, automação e mentoria técnica

Formação

Universidade de São Paulo **https://www.usp.br**
Bacharelado, Ciência da Computação, Nota: 8.7/10 *mar. de 2014–dez. de 2018*

- Formação sólida em fundamentos de computação, bancos de dados e engenharia de software
- Desenvolvimento de projetos acadêmicos com foco em processamento e análise de grandes volumes de dados
- Participação em grupos de estudo de ciência de dados e computação distribuída

Cursos: Estruturas de Dados, Algoritmos e Complexidade, Bancos de Dados, Engenharia de Software, Sistemas Distribuídos, Estatística Aplicada, Aprendizado de Máquina, Computação em Nuvem

Insper **https://www.insper.edu.br**
Mestrado, Ciência de Dados, Nota: 9.1/10 *mar. de 2019–dez. de 2020*

- Especialização em arquitetura de dados, modelagem analítica e pipelines escaláveis
- Projeto de conclusão sobre qualidade de dados em plataformas de analytics em tempo real
- Desenvolvimento de soluções com foco em impacto de negócio e boas práticas de engenharia

Cursos: Engenharia de Dados, Estatística para Ciência de Dados, Aprendizado de Máquina, Visualização de Dados, Governança de Dados

Experiência Profissional

Nubank **https://nubank.com.br**
Engenheira de Dados Sênior *fev. de 2022–Atual*

- Liderança técnica na construção e evolução de pipelines de dados batch e streaming para produtos de crédito e risco
- Modelagem de dados analíticos em lakehouse com Apache Iceberg, Spark e dbt, apoiando times de analytics e ciência de dados
- Implementação de contratos de dados, testes automatizados e monitoramento de qualidade com foco em confiabilidade
- Otimização de custos e performance em AWS, reduzindo tempo de processamento e consumo de recursos
- Mentoria de engenheiras e engenheiros de dados júnior e colaboração com produto, engenharia e negócio

Palavras-chave: Apache Spark, Apache Airflow, dbt, Kafka, AWS, Data Quality, Lakehouse

iFood **https://www.ifood.com.br**
Engenheira de Dados *jan. de 2020–jan. de 2022*

- Desenvolvimento de pipelines ETL e ELT para ingestão de dados de pedidos, entregadores e restaurantes
- Construção de modelos dimensionais e tabelas analíticas para dashboards de operações e growth
- Automação de orquestração com Airflow e padronização de transformações com SQL e Python
- Atuação em iniciativas de observabilidade de dados, catálogo e linhagem para maior confiabilidade
- Parceria com times de BI e produto para traduzir necessidades analíticas em soluções escaláveis

Palavras-chave: ETL, ELT, SQL, Python, Airflow, Data Warehouse

Itaú Unibanco **https://www.itau.com.br**

Analista de Dados Júnior *mar. de 2019–dez. de 2019*

- Manutenção e evolução de rotinas de carga de dados para relatórios regulatórios e gerenciais
- Criação de consultas SQL e scripts Python para tratamento, validação e consolidação de dados
- Apoio à documentação de processos, regras de negócio e dicionários de dados
- Colaboração com áreas de risco e finanças para garantir consistência das informações

Palavras-chave: SQL, Python, Dados Financeiros, Relatórios, Qualidade de Dados

Idiomas

Português: Nativo ou Bilíngue **Palavras-chave**: Comunicação técnica

Inglês: Proficiência Profissional Fluente **Palavras-chave**: TOEFL iBT 108, Inglês para negócios

Espanhol: Proficiência Profissional Limitada **Palavras-chave**: Leitura técnica

Competências

Engenharia de Dados: Especialista **Palavras-chave**: Apache Spark, Apache Airflow, dbt, Apache Kafka, Databricks, Apache Iceberg, Modelagem de Dados, Data Quality

Nuvem e Infraestrutura: Avançado **Palavras-chave**: AWS, Terraform, Docker, Kubernetes, Amazon S3, AWS Glue, Snowflake

Programação e Analytics: Avançado **Palavras-chave**: Python, SQL, Scala, PySpark, Git, Power BI, Looker

Prêmios

Prêmio de Excelência em Engenharia de Dados

Nubank *nov. de 2023*

Reconhecimento por liderança técnica na modernização da plataforma de dados de risco, com ganhos de confiabilidade e redução de custos operacionais.

Certificados

AWS Certified Data Analytics - Specialty **https://aws.amazon.com/certification/certified-data-analytics-specialty/**

Amazon Web Services *mar. de 2023*

Google Cloud Professional Data Engineer **https://cloud.google.com/learn/certification/data-engineer**

Google Cloud *ago. de 2022*

Publicações

Blog de Engenharia do Nubank **https://medium.com/@mariana.costa.almeida/plataforma-de-dados-confiav**

Construindo uma Plataforma de Dados Confiável em Escala *ago. de 20*

- Artigo sobre práticas de contratos de dados, testes automatizados e observabilidade em pipelines críticos
- Discussão de arquitetura lakehouse, governança e métricas de confiabilidade para times de engenharia

Referências

Prof. Ricardo Menezes **+55 11 91234-5678**

Professor de Ciência de Dados *ricardo.menezes@insper.edu.br*

Mariana combina rigor técnico, visão de negócio e capacidade de liderança. Seu trabalho em plataformas de dados trouxe ganhos concretos de confiabilidade e eficiência para equipes multidisciplinares.

Projetos

Pipeline em Tempo Real para Detecção de Fraudes **https://github.com/mariana-costa-data/fraud-st**

Plataforma de streaming para ingestão, enriquecimento e scoring de eventos transacionais *mai. de*

- Arquitetura com Kafka, Spark Structured Streaming e armazenamento em lakehouse para análise de eventos em near real time
- Implementação de regras de qualidade, alertas e monitoramento de latência e throughput
- Integração com modelos de machine learning para apoio à decisão em risco e prevenção a fraudes

Palavras-chave: Kafka, Spark Streaming, Lakehouse, Fraude, Data Quality

Interesses

Corrida: Corrida de rua, Meia maratona, Trilhas

Comunidade: Mentoria, Palestras, Open Source

Música: Violão, MPB, Jazz

Voluntariado

Data Girls Brasil **https://www.datagirlsbrasil.org.br**

Mentora Voluntária *mar. de 2021–Atual*

- Mentoria de mulheres em transição de carreira para dados, com foco em SQL, Python e projetos práticos
- Participação em eventos e rodas de conversa sobre carreira em engenharia de dados
- Apoio na revisão de currículos e preparação para entrevistas técnicas