

Sthefany Spina

Data Engineer | Data Analyst | Python Developer | Data Intern

✉️ sthefany@outlook.com | ☎ +55 19 994337044 | 🔖 <https://sthefanys-tech.vercel.app/>
⌚ <https://github.com/sthefany@outlook.com> | 💬 <https://www.linkedin.com/in/sthefany-spina-02bb11202/>

Resume

Professional transitioning to the Data Engineering field, focused on building scalable pipelines, processing and validating data, and supporting data-driven decision making.

I have knowledge of Python, analysis and visualization tools such as Excel, Power BI, Tableau and Looker Studio, databases such as MySQL and PostgreSQL, and strong communication skills.

Skills

Languages: Native: Portuguese | Advanced: English | Intermediate: Spanish | Basic: Italian

Programming Languages: Python, R, Html, CSS, Javascript

Frameworks: Node.js, Pandas, Pyspark, Matplotlib, Seaborn, Numpy, Flask, Django;

Database: MySQL, PostgreSQL, MongoDB, Redis;

Machine Learning: scikit-learn, TensorFlow, Keras, PyTorch

Technologies: Excel, Power BI, Tableau, Looker Studio, Apache Airflow, Databricks, Apache Spark, Apache Hadoop, Apache Kafka, BigQuery, Redshift, Snowflake;

Cloud: AWS, Azure and Google Cloud Platform;

Developer Tools: VsCode, Docker, Git, Github, Google Colab;

Soft Skills: Strong analytical skills, problem-solving, and attention to detail; Interest in learning new tools and technologies; Comfortable working independently and as part of a team.

Education

Postgraduate Degree in Data Analytics

Aug 2025 – Jul 2026

North Paraná University (UNOPAR)

Bachelor's degree in Computer Engineering

Jan 2025 – Dec 2028

Unifatec University Center

Postgraduate Degree in Data Engineering and Machine Learning

Jan/2024 – Dec 2025

North Paraná University (UNOPAR)

Bachelor's degree in Systems Analysis and Development

Jan 2023 – Dec 2024

Unifatec University Center

Experience

Grande Hotel São Pedro – Receptionist

Jul 2022 – at the moment

- Recorded and maintained guest, reservation, and billing data with high accuracy, ensuring the integrity, quality, and consistency of the information.

- Structured and followed complex operational processes, developing a useful systemic view for automating and orchestrating data flows.
- Resolved inconsistencies in records and system failures, applying analytical thinking and troubleshooting skills.

Licenses and Certifications

Microsoft Certified - Azure Data Fundamentals	Aug 2025 – Aug 2026
Databricks – Databricks Fundamentals	Mar 2025 – Mar 2026
Google – Google Analytics	
IBM – Data Analyst Professional Certificate	
IBM – Data Engineering Professional Certificate	

Projects

Data Pipeline with Monitoring Dashboard

Built an automated pipeline to collect data from CSV/API sources, handle inconsistencies (names, dates, formats), and generate a standardized dataset. Cleaned data is stored in a relational database and connected to a BI tool for visualization.

Stack: Python, Pandas, Numpy, PostgreSQL, ETL, Power BI.

Link: <https://github.com/sthefanyspina/DataEngineeringProjects/tree/main/pipeline>

Data Ingestion and Normalization Pipeline with PySpark

Processed large CSV/JSON files using PySpark, performing joins, cleaning, normalization (dates, names), duplicate removal, and null handling. Produced a final Parquet dataset with quality logs and rejected records. Includes exploratory analysis with aggregations and format performance comparison.

Stack: PySpark, ETL, Data Quality, Parquet

Link: https://github.com/sthefanyspina/DataEngineeringProjects/tree/main/pyspark%20projects/spark_pipeline

E-commerce Sales Analysis and Forecasting

Developed an interactive dashboard to identify best-selling products, top revenue categories, and seasonal trends. Included a demand forecasting model (ARIMA/Prophet) for upcoming months.

Stack: Python (Pandas, Prophet, NumPy, Matplotlib / Plotly), Power BI

Link: https://github.com/sthefanyspina/DataAnalysisProjects/tree/main/Dashboards/sales_dashboard_powerbi