

Marcos Pereira

Data Scientist



Profile

My journey in data started with BI where I had the opportunity to implement the entire process that involves a BI project from requirements acquisition, through ETL and dashboard construction, specializing later in Data Science in which I have been working for more than 3 years in this area.

Education and Certifications

Post Degree in Data Science | UNI7

Degree in Technologist Analysis and System Development | University Estacio de Sá

Online courses:

- Microsoft PowerBI for Data Science(72h) – DSA (Data Science Academy)
- Big Data Analytics with Microsoft Azure Machine Learning(108h) - DSA
- Data Engineering with Hadoop and Spark(44h) - DSA
- Business Analytics(64h) - DSA
- Machine Learning(120h) - DSA
- Real-Time Analytics with Python and Spark(72h) - DSA
- Data Visualization and Dashboard Design(54) - DSA

Certifications:

DSCP Certification - Data Scientist Certified Professional (Data Science Academy)

Key Skills

Technical Skills:

- Python
- R/Rstudio/Markdown
- SQL
- Spark
- Hadoop
- Hive
- NoSQL (MongoDB, Redis and Neo4J)
- Git
- Docker
- PowerBI
- Flask
- Airflow

Soft Skills:

- Team Player
- Critical Thinking

• Pró-active

Language Skills:

- Portuguese (mother tongue)
- English (B2 level)
- Spanish (A2 Level)



Experience Highlights

Company: Capgemini Engineering

Duration: 11 months

Role: Data Scientist

Responsibility:

Project 1:

- Exploratory analysis between meteorological data and equipment defects to check correlations and predict failures. **Technologies:** Python, sklearn, SQL

Project 2:

- Anomaly detection in time series. **Technologies:** Python, sklearn, pycaret, SQL, Kafka

Project 3:

- Put into production and propose improvements in cluster models for fraud detection. **Technologies:** Python, sklearn, oracle, PowerBI, Streamlit.

Company: Instituto Atlântico

Duration: 2 years

Role: Data Scientist

Responsibility:

Project 1:

- Development Machine learning models can offer suggestions for product pricing. **Technologies:** Python, Sklearn, Flask, Docker.

Project 2

- Helped architect a corporate Data Lake to allow access to fraud data between teams and created a machine learning model to perform data deduplication. **Technologies:** Python, Kafka, Docker, Greenplum, Teradata, Sklearn, Power BI.

Project 3

- Development of Machine learning models to fraud classification by priority order based on the potential impact. **Technologies:** Python, Sklearn, Salesforce Einstein.

Project 4

- Development of a tool to assist in the execution of functional and usability tests using Cognitive Computing techniques in desktop applications and web pages. **Technologies:** Python, Selenium, docker, Django

Company: Atlanta Tecnologia

Duration: 7 years

Role: Developer/Business Intelligence

Responsibility: As Business Intelligence I was responsible for implementing the entire structure of BI project. Identifying use cases, creating pipelines for ETL and performing data analysis to create dashboards. **Technologies:** Power BI, Pentaho Data Integration, Postgres, Python

As a developer responsible for developing and maintaining solutions in Java and development applications for Android (mobile).

Technologies: Java, Postgres, MongoDB, Android Studio