JOÃO PEREIRA DOS SANTOS

Business Intelligence Engineer @ Amazon
EU Supply Chain Analytics

(+351)
joaopereiradsantos@gmail.com
joaopereiradsantos.com
London, United Kingdom

EXPERIENCE

Business Intelligence Engineer

Amazon EU

11/2021 - Ongoing London, United Kingdom
amazon.com

EU Supply Chain Analytics - Business Intelligence Administrator

- **On-Call**: Empower the users (+300 users) to produce and obtain data in the fastest, easiest and cheapest way, while maintaining and continuously improving the BI infrastructure. Ownership and management of 3 AWS accounts, including 3 Redshift clusters (+1 PB & +2k tables).
- **EXcel To REdshift Migration Engine - EXTREME**: An AWS architecture that automatically converts .xlsx files into Redshift local tables quickly and seamlessly (<1 min processing time). Centered around a micro-service orchestration workflow using AWS Step Functions while additionally inferring the most frugal table definition based on the present data. +20 active users and ~5 service calls per day within the first month of launch.
- **Redshift CloudWatch Dashboard**: Monitor cluster's health and performance at the node level. Proactively avoiding cluster slowness via automated trouble tickets and pager alerts. Viewed by a total of +100 distinct users ~75 times per week.
- **AWS Auto Tag**: Deployed via Lambda using CloudTrail consumed through CloudWatch, Auto Tag improves the ease of cost allocation and governance. As soon as a user creates a supported resource (such as EC2 instances, IAM users, RDS instances, S3 buckets, EMR clusters, VPC, etc), it automatically applies up to 3 tags: C (Resource Owner) + T (Resource Creation Datetime) + I (Resource Invoked by what other resource). +1k resources.
- **AWS Auto Scheduler**: Leverages AWS resource tags and AWS Lambda to automatically stop and restart EC2 and RDS instances across multiple AWS regions and accounts on a defined schedule (stored in DynamoDB deployed via CloudFormation). Reduces operational costs resulting in up to 70% cost savings on instances that are only necessary during regular business hours (weekly utilization reduced from 168 hours to 50 hours).
- **AWS SageMaker Auto Stop**: As the second largest cost in our AWS fleet, this solution auto shuts down both SageMaker Notebook and SageMaker Studio instances when they are idle for 1 hour using lifecycle configurations and the Jupyter API. 98% decrease in SageMaker costs YoY.

Business Intelligence Engineer

Amazon EU

11/2021 - Ongoing London, United Kingdom

EU Supply Chain Analytics - FC Launch & Transfer Analytics

- **Tech Project Manager & Mentor:** Manage a team of +5 BIE interns by applying Agile project management methodologies (Scrum & Kanban) on Asana. Enforce software development good practices (git for version control, pyenv for Python version management and poetry for Python packaging and dependency management). Data Engineer and Business Intelligence Engineer tech interviewer. Interviewed +10 intern candidates.
- **Data Pipelines**: Owner and provider of +30 tables sourced via batch ETL pipelines powered by AWS Redshift and Spark EMR (Amazon Cradle), orchestrated with Airflow (AWS MWAA) and Amazon BDT Maestro, tested using Great Expectations and internally monitored with Amazon Pyramid.
- **Pod Selection Algorithm - PSA**: Return a list of pods to transfer based on a set of user defined CLI inputs (argparse), business criteria (objective functions) and hard operational constrains with Integer Linear Programming (ILP) optimization using the FICO Xpress Solver. Compared to the previous production model, PSA provides ~10% transferred volume increase, ~15% units decrease and ~5% unique items decrease.
- **Pre-Launch Reactive TSO Max Capacity Alert & Automated Closure:** Automated control on pre-launch TSO assignments for new FC launches. Gets live TSO from RODEO REST API, updates a Chime room based on the predefined FC processing capacity threshold, destination FC and Critical Pull Time (CPT) and applies a Closure once 95% TSO capacity is breached.
- **EU SC FC Launch Performance Dashboard:** The EU SC FC Launch Performance Dashboard is a "One-Stop Shop" metrics compilation providing a user-friendly interface with visualizations of new warehouse's ramp up actual vs. week-1/locked S&OP plans. The QuickSight dashboard reduces the amount of planning time and serves as a centralized single source of truth for SC Analytics, GES Operations Engineering and FC Finance teams. Surpassing a total of 1k views since launch, by +60 distinct users and used weekly on average 30 times.

Business Intelligence Intern

Amazon EU

02/2021 - 08/2021 Luxembourg, Luxembourg

EU Supply Chain Analytics - Inbound Cross-Dock (EU IXD)

- **Why Spread is Biased & How to Overcome it: Spread Bias:** A complementary metric to FC Spread (how many FCs on average an ASIN is sent to). Mathematically, it's a weekly weighted average share of total volume cross-docked at each FC per ASIN. The more biased the volume is towards one particular FC, the more the spread bias will tend to 1. Less Spread Bias leads to a more uniform FC level placement which increases unique inventory and reduces the risk of TRB (constraint in outbound capacity).
- **ITS 2% Rule & Impact on Placement:** Understanding of hard constrains in optimization models (SCOT heuristic approach to reduce latency of the request easing the algorithm decision time by removing the possibility of case break) and impact on placement and financial outcomes (spread, item selection, period 1/ period 2 AR share and misplacement volume). Estimated 10% of total IXD volume to fall under this rule, leading to misplacement and increased CBF (~$35M annualized benefits based on increase in sortation eligibility for UIS by ~1.5MMu/week; Improving UIS utilization saving 22cents/unit by sorting via UIS vs Manual sortation; Reduced long distance fulfillment)
- **Tote Utilization Dashboard:** Tote monitoring for the IXD Sr. Ops Managers & Area Managers. Aimed at knowledge sharing and improved users tote filling best practices in order to increase truck fill rate and tote optimization.
- **Centralized Fluid Loading Dashboard:** Power BI dashboard for the overall Productivity (fluid loading share, volume, fill rate and labor) Sustainability (saved number of trucks, CO2 emissions, plastic waste) and Savings (transportation, productivity gain, unloading cost) metrics.
- **IXD Optimization Model EDA:** Collaboration with the EU S&OP & SC Data Science Team on an exploratory statistical analysis regarding ITS bias in actual/ ideal product level (PL) and group level (GP) assignments for a more accurate heuristic ITS simulation algorithm.
- **S&OP Arc Bin Level Forecast:** Arc bin level forecast with uni-variate multi forecasting time series using Exponential Moving Average (EMA), Auto-ARIMA and FBProphet models.

EXPERIENCE

Data Analyst Intern

Amkor Technology Portugal

09/2020 - 01/2021 Mindelo, Portugal

amkor.com

Physical Failure Analysis & Reliability Request Management Process Optimization

- **Requests Management Workflow Automation & Optimization:** Applied DMAIC methodology (Define, Measure, Analyze, Improve, and Control) as a data-driven improvement cycle to clearly articulate the business problem, goal, potential resources, project scope, and high-level project timeline. Aimed at improving the Request Scheduling Efficiency by over 48% YoY.

Process Engineer Intern

Amkor Technology Portugal

02/2020 - 07/2020 Mindelo, Portugal

repositorio-aberto.up.pt/handle/10216/132835

Master Thesis Project: Qualitative and Quantitative Statistical Analysis in Copper Electroplating Baths

- **Modeling of Electroplating Chemical Components and Total Organic Contamination Concentrations:** Ease the decision-making process associated with the control of the main chemical components concentrations and total organic contamination (TOC).

Summer Research Intern

Continental ITA

08/2019 - 09/2019 Lousado, Portugal

continental.com

Anti-foaming agent study for process optimization in DIP content tire textile reinforcement.

EDUCATION

MEng in Chemical Engineering - Processes and Product

Faculdade de Engenharia da Universidade do Porto

09/2015 - 07/2020 Porto, Portugal

Sciences et Technologies: Mathématiques, Informatique, Sciences de la Matière et de l'Ingénieur (MISMI)

Université de Bordeaux

02/2015 - 06/2015 Bordeaux, France

Classes Préparatoires aux Grandes Écoles, Mathématiques Supérieures: Physique, Chimie et Sciences de l'Ingénieur (PCSI)

Lycée Fabert

09/2014 - 01/2015 Metz, France

Baccalauréat Scientifique (BS)

Lycée Français International de Porto

09/2011 - 07/2014 Porto, Portugal

LANGUAGES

Portuguese Native

French Proficient

English Proficient

Spanish Advanced

AWARDS

Winner of the Hack@Home III by Tech@Catolica A Second Wave Analysis of COVID-19

12/2020

linkedin.com/company/tech-at-catolica/

Forecasting one week of COVID-19 cumulative confirmed cases with Gompertz, ARIMA, FBProphet, XGBoost and Neural Network models. Mean Absolute Percentage Error (MAPE) of 0.7%

Winner of the 2018 Academic Games of Engineering by AEFEUP

02/2018 - 04/2018

aefeup.pt/age/

Designing a continuous hydrothermal liquefaction (HTL) biomass processing unit for Paralab, SA

PROJECTS

Tennis Betting Bot

08/2021 joaopereiradsantos.com/project/tbb/

TBB is both a machine learning model (Neural Network) that outputs match-based win probabilities and an automated web scraping bot that scrapes *oddsportal.com* future tennis matches in order to find the best bookmaker odd. The bot preprocesses both outputs and if the bet is considered valuable by over *n* %, an automated email is sent with information regarding the tournament, players, predicted odd and the maximal bookmaker odd.

Are You Too Old For Data Science?

01/2021 kaggle.com/code/joaopereiradsantos/are-you-too-old-for-data-science

Kaggle's annual Machine Learning and Data Science Survey competition. The challenge objective: tell a data story about a subset of the data science community represented in this survey, through a combination of both narrative text and data exploration.

CERTIFICATIONS & MOOC

AWS Certified Cloud Practitioner | Amazon Web Services

AWS Cloud and basic architectural principles; Value proposition; Key services; Basic security and compliance; Billing, account management, and pricing models; Basic/core characteristics of deploying and operating in the AWS Cloud. (Exam Score: 89.3%)

Interactive Python Dashboards with Plotly and Dash | Udemy

Bar Charts, Line Charts, Scatter Plots, Heat Maps, etc; Create Layouts with Plotly's Dash library; Dash for interactive components with Plotly; Connect multiple inputs and outputs with a dashboard; Live interactive graphs with clicks and hover overs; Live updating data for streaming information; Security with App Authorization; Deploying to the internet with services like Heroku. (9.5 Hours)

The Data Science Course 2020: Complete Data Science Bootcamp 365Carrers | Udemy

Statistical analysis; NumPy; Pandas; Matplotlib; Seaborn; Tableau; Machine Learning with stats models and scikit-learn; Under-fitting, over-fitting, training, validation, k-fold cross validation, testing, hyper-parameters tuning; Deep learning with TensorFlow. (29 Hours)

Data Analyst with SQL Server | DataCamp

Database Objects: Create and alter tables, views, modify constraints, DML triggers; Work With Data: Queries, sub-queries, data types, aggregate queries, manage XML data; Modify Data: Stored procedures; Combine datasets, Work with functions; Troubleshoot And Optimize, Implement error handling. (45 Hours)

Process Mining Expert | Celonis

Data Engineer Training: Technically connect and maintain Celonis processes; Analyst Training: Translate data into actionable business insights. Create target oriented analyses to transform your business; Business User Training: Simplify your daily work. Get intelligent recommendations to execute the right actions. (30 Hours)

FIND ME ONLINE

Portfolio Website
joaopereiradsantos.com

LinkedIn
linkedin.com/in/joaopereiradsantos

GitHub
github.com/joaopereiradsantos