

ABENZOAR Manoah

Data Engineer

Disponible dès le 1er Juin 2024.

✉ manoah.abenzoar@gmail.com

🇫🇷 Français

📍 Paris

☎ +33 6 35 96 61 19

📅 23 ans

💻 Présentiel ou
Télétravail



EXPERIENCES PROFESSIONNELLES

Data Engineer - Stage

24s.com Maison de LVMH

Paris, FRANCE



10/2023 -
04/2024

- Réalisation et cadrage AGILE de cas d'usage métiers Retails (Supply Chain, CRM, achat, marketplace, merchandising):
 - Conception de dashboards Looker Studio
 - Automatisation de l'accès à la donnée
- Développement de pipelines ELT & orchestration du workflow
- Collecte de millions de données internes et externes via des API (Google Analytics 4 API, Google Sheet API,...)
- Maintenance & Sécurisation de la stack data pour assurer un accès opérationnel aux ressources à hauteur de 95%:
 - Versionnage Git & Revue de code
 - Gestion & Automatisation des ressources en Dev puis en Prod
- Développement d'une maîtrise de la Datawarehouse BigQuery pour contrôler les coûts:
 - Optimisation des requêtes SQL
 - Utilisation de tables (partitionnées, clusterisées, indexées) ou vues (matérialisées et standard) selon le besoin

Jira, GCP (IAM, GCS, Cloud Function, Schedulers, Event, Pub/Sub, Cloud Run), BigQuery, SQL, Python, Git, GitLab, Terraform, Bash, Looker, MacOS, Sequel Ace, Vs Code

Data Scientist - Stage

Capgemini Engineering

Strasbourg, FRANCE



03/2023 -
09/2023

- Réalisation et cadrage d'un projet de R&D sur le calcul de l'irradiation solaire à l'aide de données satellites
- Intégration de données via des API
- Gestion de milliards de données numériques et d'images satellitaires
- Conception de pipelines ETL Pyspark et Statsmodels Python traitant des séries temporelles, avec un gain de temps de 80%
- Création d'un modèle d'ensemble learning avec une précision de 83%, à l'aide de technique de Machine & Deep Learning

GCP, Colab, Python(Pandas, Plotly, Seaborn), PySpark(Dataframe, RDD), TensorFlow, SKLearn, NN, RF, LSTM, Ada & XG Boost

PROJETS PERSONNELS

Data Platform

Normandie, FRANCE



Depuis
04/2024

- Webscraping Python Selenium de données structurées et non structurées (images, messages)
- Création de pipelines de données & Orchestration GCP (2 niveaux : Events triggers & Schedulers; Pub/Sub & Appels HTTP)
- Ajout d'un 3è niveau d'Orchestration du workflow avec Airflow
- Maintien de l'intégrité de l'app avec des pipelines CI/CD GitLab avec Stages de test, build et deploy
- Conception d'un cluster Kubernetes pour l'app

Python (Selenium, Pandas, Flask), GCP (IAM, Cloud Storage, Cloud Function, Pub/Sub, Event, Scheduler, Composer), BigQuery, SQL, Airflow, Docker, Kubernetes (Minikube, kubectl), Conda, Git, GitLab, Bash, Linux, DigitalOcean, HTML CSS, JS

A venir : Ajout d'outils ETL (DBT, Talend ou SnapLogic) | Connexion de nouveaux environnements (Cloud Providers, Hadoop ou BDD)

Analyse Footballistique

Pau, FRANCE



10/2022 -
03/2023

- Webscraping de millions de données numériques et images
- Création & Optimisation de modèles identifiant le style de jeu des équipes et des joueurs, avec une précision de 83%
- Implémentation de pipelines ETL pour convertir des vidéos de matchs de football en un format structuré
- Conception d'outils de visualisations pour communiquer la donnée

Python(Selenium, Pandas, Numpy), NN, CNN, Yolo v5, OpenCV, Ubuntu, MplSoccer

FORMATION

2020-2023 **Diplôme d'Ingénieur / Master 2** CYTech
En Mathématiques et Informatique Spécialisé en
High Performance Big Data Analytics

2022 **Data Training** Hupi
Formation Professionnelle de Data :
SQL & NoSQL (MongoDB, PostgreSQL), Kafka, PowerBI, Spark & PySpark, NN

2022 **Master 1** Keio University
Semestre d'échange Informatique en anglais : ML,
DL, AR, Computer vision, Programmation parallèle,
Cuda, OpenMP, MPI, Python (Pandas, OpenCV), Yolo

LANGUES

Français
Langue Materelle

Anglais
Courant : semestre à l'étranger

Japonais
Débutant : en autodidacte

ATOUS

**Résolution de Problèmes
Communication**

- Expériences en milieux multiculturels
- Tutorat en Mathématique et Informatique

Sérieux & Rigueur

Organisation d'activités chez 24s

Curiosité

- Passionné par la Tec, la Data et l'Automatisation
- Attiré par le Continuous Learning