Konstantinos Kompogiannopoulos

A Senior Data Scientist/Analytics Engineer/Data Engineer building Data Infrastracture at Zettablock. Ex Polygon Labs.

Email: **kkombos@gmail.com**
Location: **Greater London Area, United Kingdom**
Website: **Blockchain Data Analytics**

Linkedin, **Twitter**, **Github**

WORK EXPERIENCE

ZettaBlock
San Francisco, California, United States

Senior Data Scientist/Analytics Engineer/Data Engineer
Mar 2023 to Present

- In my role as a Data Scientist/Analytics Engineer/Data Engineer, I Implemented product-centric schemas in the Postgres database instance to enhance data exploration for clients, both in historical **ETL** and **real-time** streaming pipelines across multiple blockchains.
- Developed real-time streaming pipelines for **Bitcoin**, **Ethereum**, and other **EVM** blockchains by leveraging Alchemy and Quicknodes APIs, utilizing **Python** and **asyncio** to minimize latency.
- Led data architecture projects and analyzed user behaviour, establishing metrics to measure the impact of client-side and server-side changes on the Zettablock **SQL** playground.
- Designed **AWS Datalake batch ingestion** and transformation pipeline to server client-side **GraphQL** APIs with decoded and transformed blockchain data.
- Conducted thorough data decoding and implemented logic for both historical and snapshot **ETL** pipelines, ensuring consistent data accessibility and enriching user experience on the Zettablock platform. Built Abstraction tables using **dbt**, **Spark** and **Iceberg**.

Polygon Labs
New York, New York, United States

Data Scientist/Analytics Engineer
April 2022 to April 2023

- I my role as a Data Scientist/Analytics Engineer, I developed and maintained blockchain and web3-specific metrics, which enabled the data science team at Polygon to track product performance, user engagement, and revenue growth., I developed and maintained blockchain and web3-specific metrics, which enabled the data science team at Polygon to track product performance, user engagement, and revenue growth.
- I standardized the methodology for creating metrics on various data types, including Network, **DeFi**, **NFT**, and Gaming, by establishing proper nomenclature, parametrized and reusable aggregation logic, and respective definitions, and documentation for creating metrics in dbt from existing table columns.
- I implemented best practices for data ingestion, **data standardization**, and **quality control**, which reduced data processing time and ensured **data consistency** across all data sources.

SKILLS

Technical

Core Experience: `Python`, `dbt`, `Airflow`, `Postgres`, `Spark`, `SQL`, `Redshift`, `Flask`, `FastAPI`

Analytics Experience: `Looker`, `Google Analytics`

Other Experience: `Go`, `Docker`, `Kubernetes`

Cloud: `GCP`, `AWS`

Most used packages: `Pytorch`, `asyncio`, `Tensorflow`, `Huggingface`, `Pandas`

Project Management

`Kanban Methodology`, `Scrum`, `Clickup`

Languages
English: Fluent, C2 Profficiency
Greek: Mother tongue, C2

EDUCATION

University of Surrey
Pittsburgh, Pennsylvania
Data Science MSc
2019-2020
Specialization: Deep Learning and Reinforcement Learning.

University of Nottingham
Nottingham, UK
Theoretical and Mathematical Physics BSc
2017 - 2019
Specialization: Mathematical approaches and physics simulations through programming.

- I worked with **cross-functional** teams to integrate on-chain and off-chain data sources to provide comprehensive insights for partners using **Looker** dashboards.
- I developed internal and external dashboards and visualizations to help organizations and partners monitor the performance of their dApps, gaming platforms, and NFT collections on the Polygon network. I also refactored all **SQL** queries in use with ETL pipelines in a one table per file format. I updated and refactored all **Airflow** DAGs associated with their respective querying jobs.
- In addition to my work in Data Science, I also implemented **dbt** & **Spark**, set up a **Google Cloud data proc** cluster configured it with Spark to receive **PySpark** jobs from dbt and enabled the computer power necessary for daily Polygon wallet balances and large-scale machine-learning projects.
- I integrated the **dbt** data pipeline orchestration framework into our existing workflow and refactored all computationally intensive **Python** jobs into **PySpark** syntax. I also created documentation for Data Engineering on guidelines for using **Spark**. Furthermore, I deployed and developed various virtual machine instances, distributed computing clusters, **ETL** pipelines, and automated scheduling frameworks (**Airflow**) using the **Google Cloud platform**.
- I created an automated pipeline that aggregates and decodes blockchain data into easy-to-understand derived dimension and fact table metrics that immediately translate into actionable insights.

LeasePlan   *Slough, England, United Kingdom*

Data Scientist
November 2021 to April 2022

- In my role, I developed **Credit Risk Scoring** models for private individuals, collaborating with **Finance** and **Credit Risk Analysts** to identify and utilize both traditional and non-traditional data sources to enhance feature selection for the model. I led the feature discovery, analytics, and model development phases, and successfully delivered the initial version to shareholders.
- Additionally, I contributed to the Leaseplan Digital Organization by creating text-**extraction pipelines**, using **ML**, to find inconsistencies between different PDF contract documents.

Toyota   *Epsom, England, United Kingdom*

Junior Data Scientist
June 2020 to October 2021

- In my role as a Data Scientist and Analyst, I have led machine learning projects involving vehicle **telematics crash data**, specifically focusing on crash driver liability and accident severity prediction within the crashmatics group. I developed a comprehensive machine learning deployment workflow, which encompassed data preprocessing, feature engineering, **model training**, and API deployment. I refined the business problem and data into pre-processing steps and feature engineering in a **CRISP-DM**-type lifecycle. My experience includes utilizing **AWS** for

University of Leicester
Leicester, UK
Physics BSc
2016 - 2017

data storage and deployment, Python and Bash scripting for model development and deployment, and **MLFlow** for development tracking and **Docker** for deployment.

- In the realm of Data Engineering, I implemented and maintained an **AWS**-based technology stack, gaining substantial experience in cloud computing and storage. I deployed **ML** models on **AWS**, refining data preprocessing and model deployment practices. I utilized **MLFlow** to log, load, register, and deploy machine learning models, improving the efficiency and traceability of model development and deployment workflows. I also employed **Docker** for development tracking and deployment, allowing for consistent and reproducible environments.

- Furthermore, I led the development of **ETL** pipelines and automated scheduling frameworks on **AWS**, improving data processing efficiency and consistency. I developed and maintained machine learning feature extraction and **model pipelines** for predicting accident **severity** and driver **liability**, contributing to the refinement of business problems and pre-processing steps.