Our Roles, Aspirations, and News in the Data Space

By Daniel Solomon, Guilherme Soares, Guney Kuran & Madalina Samoila

Where we see ourselves...

Data Scientists

Guilherme Soares

1. Complete Sparta's as a Data Engineer. Learn Python, SQL, etc...

2. Enter role with Sparta's Client to gain real experience in Data Engineering

3. Finish contract with Sparta Global.

4. After 5/10 years: Be working in the US for a renowned tech company (Microsoft, Apple, etc...) as an AI Dev or other area of ML.

Madalina Samoila

1. Completing Sparta course as a data engineer. Proficient in sql, python & database management.

2. Bridging the gap between technical expertise & real world impact through working with Sparta client.

4. Further qualifications/ world projects i.e Microsoft Fabric to build a strong portfolio.

5. Experienced & strategic data leader.

Guney Kuran

Bridging the skill gap with Sparta Global to kickstart my career in Data Engineering

Data Engineering placement with a clients organisation

Established or experienced role with more responsibilities within the organisation

Daniel Solomon

Complete training with Sparta Global

Kickstart my career with a client, as a Junior Data Engineer

Finish Placement with Sparta Global

See what further progression opportunities arise from there

Data Engineers

Aspiring Data Scientists

What skills needed?

- **Statistical & analytical skills** - descriptive, inferential statistics, hypothesis testing, probability theory
- **Programming languages** - Python, R, SQL, Jupyter Notebook
- **Data visualisation** - Python libraries (matplotlib,seaborn, plotly), dashboard (Power BI, Streamlit), storytelling with data
- **Machine Learning** - Supervised & Unsupervised Learning
- **Data Cleaning & Transformation**
- **Data Management & Cloud Basics** - Big Data Tools, Hadoop
- **Soft skills** - problem solving, analytical thinking, communication & storytelling
- **NoSQL & SQL** - MongoDB & MySQL

Guilherme Soares:
What skills do I currently have?
- **Python, Java**, R, SQL
- Machine Learning
- **Database Management + SQL** (MySQL, Access)
- **Statistical and Analytical skills** (from R)

Madalina Samoila
What skills do I currently have?
- **SQL, Python,** R
- **Statistical skills** (regression analysis, hypothesis testing)
- **Data Visualisation** (Streamlit, Power BI)
- **Data Analysis** (MySQL)

Aspiring Data Engineers

What skills needed?

- **ETL Data Pipeline** - creating pipelines to automate data workflows and ensure accurate and clean data is delivered
- **SQL** - extract, transform, and load (ETL) data from various sources
- **NoSQL** - e.g. MongoDB
- **Python** - ETL pipelines, data transformation scripts, API's
- **Cloud** - e.g. Amazon Web Services
- **Machine Learning**
- **GitHub**
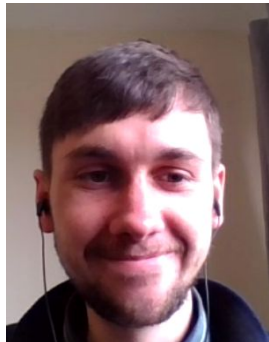
Guney Kuran

What skills do I currently have?

- **Python** - including a PCAP certification
- **PostgreSQL** - Databases, Queries
- **Tableau** - Data Connection, Preparation, Visualisation and Analysis.
- **GitHub**

Daniel Solomon

What skills do I currently have?

- **ETL Data Pipeline**
- **PostgreSQL** - Databases, Queries
- **Python** - Functional Programming, Object Oriented Programming (OOP), Pandas, Test Driven Development (TDD)
- **AWS Cloud** - S3, Lambda, Step Function
- **GitHub** - Version Control, Branches

Current Data News

Space-Based Data Centers

Axiom Space (Space infrastructure developer) are planning to launch two orbital data centre nodes into low earth orbit by the end of 2025.

Nodes will operate independently of terrestrial infrastructure which will offer real-time data processing, cybersecurity and AI capabilities for satellites and spacecrafts

Previous testing such as the 2022 AWS Snowcone on the International Space Station which leads to future plans for a commercial space station by 2027.

Legal Battles over Data: Legal problems faced over using copyrighted Data for Training Models. News, media, science, and health sectors are opting out of handing their data for AI datasets. As less data enters the datasets AIs use to train from, this may impact model performance, prompting companies to make more efficient models.

- Some AI are already being scrutinised for the way they use training data…
- Meta is also mishandling data…
- More efficient models that use less data help here (DeepSeek)

Laws are taking some time to catch up with the evolution of AI and new regulations might be added soon to ensure that training data is sourced ethically.

Current Data News Cont.

- **Data Privacy and Regulation (DeepSeek)**
  - DeepSeek, a Chinese AI company, is suspected as falling very short of GDPR regulations.
  - Suspected of storing EU user's data in China.
  - Blocked in some Italian app stores.
  - South Korea banned new downloads of the app.
  - Australia banned it on government devices.

AI Center Data Challenges (Sustainability)

Data experts are now focusing more on **AI skills** to get the most out of data.

i.e: **Amazon**, **Netflix**, **Spotify** looks at customer data to recommend products the user might like, making user experience faster and more personalised.

A lot of energy is being used, globally:

- Data centers used about **1.5% of electricity** in 2024 roughly **415 terawatt-hours** (TWh)
- Estimates suggesting it could drive data center demand to **945 TWh by 2030.** For context, that's Japan's annual electricity use today.

So data centers need more power, which worries some environmental experts.

Companies looking to save energy **Amazon**, **Microsoft**, **Google**, and **Meta.**

Amazon(largest corporate buyer of renewable energy) **- aims for 100% by 2025**

Google achieved 67% carbon-free energy in its data centers 24/7 in recent years, up from 61% in 2019, through solar, wind, and other renewables.

Microsoft - sourcing **wind, solar, and hydropower** aims for **100% by 2025**.

Meta - Achieves 100% renewable energy for data center operations through PPAs

THE

END!