Data Scientist Positions – Maya Breslauer

Description of positions sources:

The aggregated dataset is built from multiple sources:

- **LinkedIn** – Collected via the Serper API, focusing on Data Scientist positions in Israel.
- **Riskified** – Extracted via the company's public Greenhouse API.
- **Similarweb** – Retrieved from Greenhouse; required custom text parsing and regex-based section splitting due to less structured job descriptions.
- **Taboola** – Retrieved from Greenhouse; HTML structure required manual parsing to split descriptions into standardized sections.
- **Melio** – Retrieved from Greenhouse; presented in a relatively well-structured format.

In summary, all company career page data (Riskified, Similarweb, Taboola, Melio) was sourced from **Greenhouse**, while LinkedIn data came from the **Serper API**.

Dashboard:

The interactive dashboard was built with **Streamlit** to visualize and explore aggregated Data Scientist job postings in Israel. The data is merged from multiple sources, including LinkedIn and several company career pages. **Main Features:**

- **Interactive Filters** — Filter jobs by keyword, company, source, location (city), posting date, and whether the role is open.
- **Key Metrics** — Displays total filtered roles, number of open roles, unique companies, and unique sources.
- **Charts & Insights:**
    - **Roles by Posting Date** – Shows the distribution of new job postings over time.
    - **Top Companies** – Highlights companies with the highest number of open roles.
    - **Roles by Seniority (heuristic)** – Categorizes job titles into entry-level, mid-level, senior, lead, or managerial positions based on title keywords.
- **Job Table** — A searchable and scrollable table listing key details for each posting, with direct links to the original job description.
- **Download Option** — Export any filtered view as a CSV file for offline analysis.

Technical Details:

- The dashboard uses **Plotly** for interactive charts and **custom CSS** for a bright, slightly blurred background design.

- Data loading and processing are optimized with caching to ensure smooth performance.
- The UI is responsive, making it accessible from both desktop and mobile browsers.