JAVIER GARCÍA SELLER

Community of **Madrid, Spain**

TOTAL AVAILABILITY

EXPERIENCE

2023 – 2024:

1) **ETL** Software Architecture (**Python, SQL, Websockets**): **OPTIMIZED Software architecture** to easily make new WebScraping projects. Provides the **following advantages**:
   - **REDUCES the time** to start scraping
   - **IMPROVES project organization**, useful in company scenarios where you need to extract data from multiple sites such as Tiktok, Instagram, Twitter, LinkedIn

   Projets made:

   1. Tennis WebScraping**: Scrapes** tennis match data from 'flashscore.es', **transforms** the data into JSON format, **and loads** it into a MySQL database. This Project was made to practice **SQL** and Python library **pandas**.
   2. **Linkedin** WebScraping: **Extracts all available jobs** with their descriptions based on a job title and a location for further analysis using a **custom LLAMA3 model**
   3. Whois WebScraping: **Extracts contact info** from domain names registered. Used in Contacts Database.

2) **Contacts Database** (**SQL, React, Python**): **Website** created with React to get **contact information** given a **LinkedIn profile**, personal website, or email. **Multiple data sources** were collected, **transformed**, and loaded into **MySQL**. **FastAPI** is used for the backend, where Python handles querying MySQL in **parallel** to **REDUCE time response**.

2022 – 2023:

1) Tennis Match Prediction (**Python, Machine Learning, Deep Learning**): Public tennis data in CSV format was **loaded, processed, and transformed** using Python's **pandas** library. **Feature engineering** was performed to obtain multiple new variables, and an **AI model** was made to predict the match winner. The **probability** of correctly predicting the match winner **increased FROM 71% TO 77%, compared to the predictions made by bookmakers**

ACADEMIC DATA

- **(URJC)** Universidad Rey Juan Carlos 2020 – 2025
  - **6,69 SOFTWARE ENGINEER**

LANGUAGES

- Spanish Native
- **ENGLISH B2- C1**