

RISHABH KUMAR

✉ resumetozero3@gmail.com

📄 [rishabh-kumar-719a0a258](https://www.linkedin.com/in/rishabh-kumar-719a0a258)

🌐 [resumetozero](https://resumetozero.com)

☎ +91-9167993467

Education

Rajiv Gandhi Institute of Petroleum Technology (An Institute of National Importance)

Nov 2022 - Present

Bachelor of Technology in Computer Science Engineering

Bengaluru, Karnataka, India

Summary

Data Engineering and Analytics Enthusiast with hands-on experience in designing scalable ETL pipelines, workflow automation, and delivering data-driven insights for business impact.

Experience

Sambin Technologies Pvt. Ltd.

Jun 2024 - Aug 2024

Data Engineer Intern (Remote) — Selenium, Bs4, multithreading, requests, pandas

Visakhapatnam, India

- Scalable web scraping pipelines to extract structured data from multiple online platforms.
- Performed data cleaning, transformation, and validation to ensure data accuracy and consistency.
- Boosted pipeline performance by 40% through implementation of multithreading and automation scripts.

Projects

Football Data Engineering Pipeline | *Apache Airflow, Python, Azure, SQL, Tableau*

- Developed an end-to-end ETL pipeline extracting football data from Wikipedia using BeautifulSoup.
- Standardized and transformed datasets; automated workflows with Apache Airflow, achieving 99% reliability.
- Deployed data to Azure Data Lake via ABFS protocol and visualized insights through Tableau dashboards.

LLM-Analyser Webscraping | *Python, Playwright, Streamlit, Ollama LLM, Tor, obfs4, Async IO*

- Built an async web scraping pipeline with Playwright for 50+ dynamic sites, processing 10,000+ pages and handling network restrictions via Tor and obfs4 bridges.
- Applied Ollama LLMs for NLP tasks including summarization, entity extraction, and semantic analysis on 1M+ words.
- Developed a Streamlit dashboard for interactive visualization and optimized API workflows, reducing runtime by 40% while ensuring data accuracy.

Ola Bike Ride Analysis | *Python, Pandas, Matplotlib, Scikit-learn*

- Analyzed ~10K ride records to uncover demand trends, pricing dynamics, and customer behavior.
- Engineered features (datetime splits, holiday flags, outlier handling) to enhance model performance.
- Built and compared ML models (Linear Regression, Random Forest, XGBoost), achieving ~85% forecast accuracy.

Technical Skills

Programming Languages: Python, SQL(PostgreSQL, MySQL), C/C++

Backend & APIs: FastAPI, Flask, SupaBase, Rest API, SQLAlchemy, Postman

Data Engineering: Airflow, PySpark, Docker, AWS(S3, EC2), Dataricks

Data Analysis: Pandas, NumPy, Matplotlib, scikit-learn, Power-Bi/Tableau

Web Scraping: BeautifulSoup, Selenium, playwright

Version Control & CLI Tools: Git, Bash, Custom CLI

Platforms & IDEs: Jupyter Notebook, VS Code, Sublime Text, Google colab

Operating Systems: Linux (Ubuntu), Windows

Achievements

• **LevelSuperMind Hackathon 2024 — Finalist**

Jan 2025

Extracted user behavior data from Google, YouTube, Reddit, and Quora using Selenium for ad creation analysis.

Analyzed competitor ad strategies to identify high-performing hooks and CTAs, leveraging Flask for data processing.

• **Smart India Hackathon 2023 — Winner**

Dec 2023

Built a Flask-based price comparison tool for GeM by scraping 2L+ listings (Amazon, Flipkart, eBay) using BeautifulSoup; processed and compared data for government cost optimization.

• **StrataScratch** – Completed all free SQL analytics questions

• **HackerRank** – Certified in Python & SQL with 5-Star Rating (Advanced Level)

Positions of Responsibility

Design Head, IEEE Student Branch RGIPT

Nov 2023 - Feb 2025

Design Head, URJOTSAV '24 – Technical Fest, RGIPT

Nov 2024

Secretary, KALTARANG '25 – Cultural Fest, RGIPT

Apr 2025

Cultural Council Executive, Music Club RGIPT