Andhra Trade License Extractor: Java + Selenium for Scalable Lead Generation Using Hash Map data structure

Unlocking business potential through automated data extraction.

Why this project

- Introduction: The Challenge of Manual Data Collection

- Our Solution: Project Highlights and Technology

- Why Java? Core Advantages for Data Mining

- Key Differentiator: Multi-System Parallel Scraping

- Tangible Outcomes and Impact

The Challenge: Inefficient Data Collection

Manual data collection from trade license portals is a time-consuming and inefficient process, often leading to:

- High labour costs
- Delayed lead generation
- Inaccurate or outdated information

The Need: Active Customer Data for Tele-Sales

Sales and marketing teams require large volumes of **active customer data**, particularly mobile numbers, to:

- Improve tele-sales outreach effectiveness
- Enhance lead qualification
- Drive targeted marketing campaigns

Our Solution: Project Highlights

Automated Extraction

Utilises Selenium WebDriver with generic XPath for intelligent traversal of dynamic web elements.

Active User Filtering

Only "Active" users are filtered and stored, ensuring high-quality leads.

Duplicate Prevention

HashMap, with License Number as key, avoids duplicate entries through efficient hashing.

CSV Export

Mined data is organised in-memory via ArrayLists and exported to readily usable CSV files.

Why Java for This Solution?

Platform Independence

Java ensures the solution runs seamlessly across various operating systems without modifications.

Speed & Scalability

Leverages Java's performance for rapid data processing and scalability to handle large datasets.
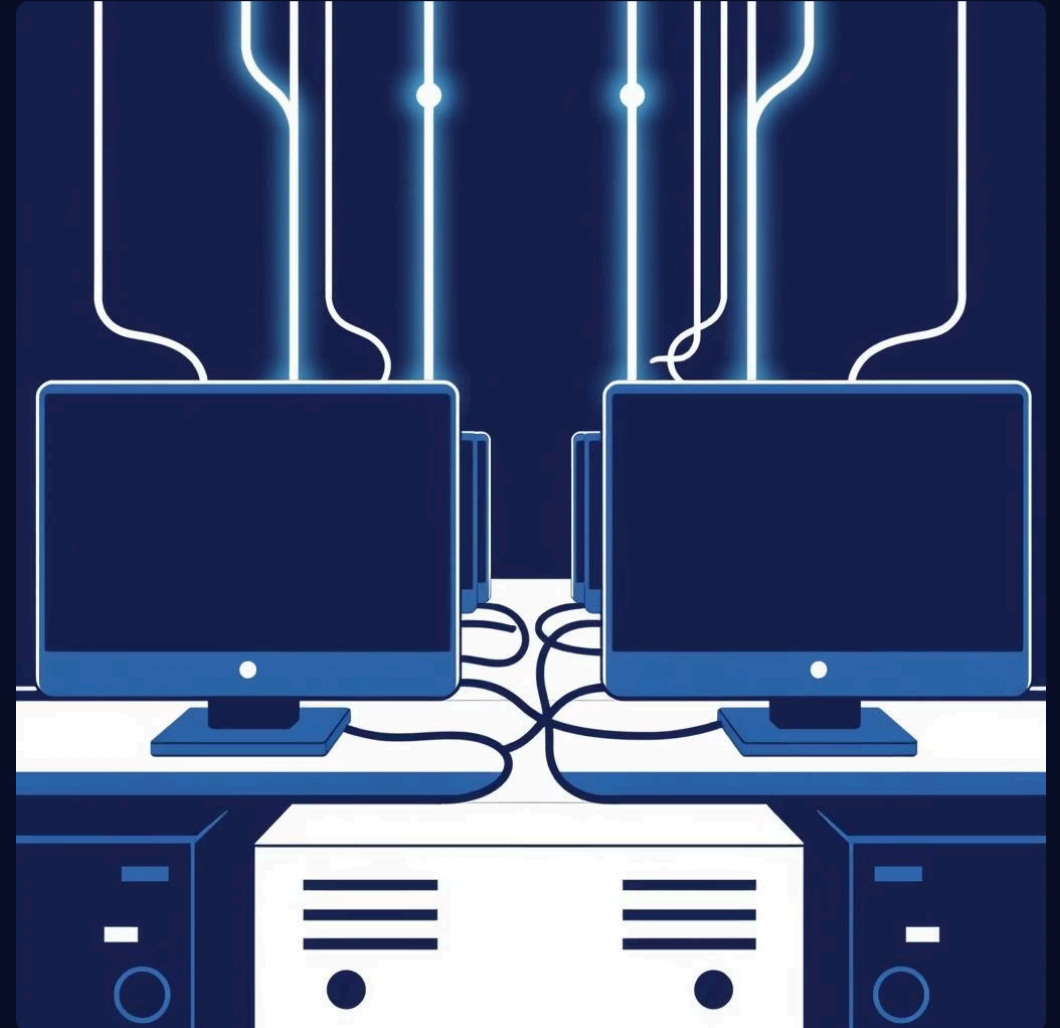
Easy Deployment

Executable via JAR files and BAT scripts, eliminating the need for an IDE on deployment systems.

Best Feature: Multi-System Parallel Scraping

The extractor supports **simultaneous operation across multiple systems**, enabling:

- Massive-scale data mining

- Rapid extraction of records (e.g., 20k records per system)

- Significant reduction in overall data acquisition time

Achieving Scale: 50,000+ Active Licenses

Our objective is to build a scalable, automated web scraping system to extract and organise over **50,000+ active trade license entries** from the Andhra Pradesh Trade License portal.

Tangible Outcomes & Business Impact

50K+

Verified Records

Automated pipeline for mining a high volume of accurate customer records.

20%+

Tele-Sales Boost

Provides precise, deduplicated leads, directly enhancing tele-sales performance.

50%+

Manpower Reduction

Significantly reduces the time and human resources required for lead generation.

Next Steps

- Demonstration of the system in action.

- Discussion on deployment and integration within your existing infrastructure.

- Customisation options to fit specific data requirements.

Let's unlock the potential of automated lead generation for your business.

Data Source Webpage Image

Trade License

Note: The information provided online is updated, and no physical visit is required.

Search | Reset | Close

Application Number	TL Number	Trade Owner	Mobile Number	Status	PT Assessment No.	Arrears Demand	Current Demand	Total Collection	Actions
06883-2024-AP	TL/00008/2016	Shaik Mahaboob Basha	8688862484	Cancelled	N/A	10125	500	3000	Select from Below ⌄
01832-2023-NZ	TL/00012/2016	N P Mahboob Khan	9440248207	Active	1001016065	6900	300	6600	Select from Below ⌄
08090-2025-PM	TL/00017/2016	D Subramanyam	9705895397	Active	1001003679	6700	300	7000	Select from Below ⌄
00491-2016-DK	TL/00032/2016	Masthan Sab	9666479635	Active	N/A	9100	1000	1500	Select from Below ⌄
10697-2025-DC	TL/00033/2016	S Mahaboob Basha	9490596931	Cancelled	N/A	14650	1500	2500	Select from Below ⌄
00524-2016-TH	TL/00043/2016	V Padmanabhamaiah	9949727092	Active	N/A	3100	200	0	Select from Below ⌄
00607-2016-KY	TL/00058/2016	S Mahaboob Basha	9704051853	Active	N/A	10750	1000	625	Select from Below ⌄
00657-2016-LI	TL/00067/2016	J Babu	9885967667	Active	N/A	10250	1000	750	Select from Below ⌄

Showing 1 to 10 of 450 entries   Show 10 entries   Previous 1 2 3 4 5 ... 45 Next