**Scan Unlimited Script: Auto Clean, Copy & Prep**

📌 **Lesson Focus:**  
A **script that transforms raw Scan Unlimited exports** into a clean, lightweight file — ready for Keepa analysis and eventual merging.

We’ll eliminate extra columns, trim unnecessary rows, and extract just the ASINs we need — all with a **single, reusable script.**

**🔁 Script Workflow: What It Does**

The script named SU Cleaned will:

**✅ Auto-detect the SU File**

It finds a CSV file in the folder that includes "ScanUnlimited" in the name.

**🧹 Clean the File**

* Deletes all **unneeded columns** (there are 36+ columns in some cases — we only keep what we use)
* Rename Source header for the source links if its named something else
* Deletes all rows with **junk ROI** (e.g., -99% to -1%)

**📋 Copy ASINs to Clipboard**

* After trimming, copies all ASINs to clipboard to upload to Keepa
* ✅ Console prints:  
  "Copied 9,434 ASINs to clipboard"

This is what we’ll use for Keepa Product Viewer (KPF).

**💾 Outputs a Clean File**

* Final export file:  
  📄 SU Cleaned.xlsx
* This will serve as the Scan Unlimited half of the sourcing list — to be merged later with Keepa

**👨‍💻 Human Workflow: What You’ll Do**

To run this system properly:

1. ✅ Download the Scan Unlimited data **after your preset is applied**
2. ✅ Paste the exported file into the **same folder** as the leads list script
3. ✅ Run the script SU Cleaned.py
4. ✅ Copy the ASINs from clipboard into **Keepa Product Viewer**
5. ✅ Download the Keepa results and place them into the script folder
6. ✅ Proceed to the next script phase (Keepa Clean → Format)

**🧠 Design Phase – Building a Dynamic, Future-Proof Script**

Before we write the logic, we need to think long-term.

Our goal isn’t just to clean one file — it’s to build a script that works across all future Scan Unlimited exports, no matter the site or variation.

**🧩 The Problem: Custom Columns**

Every site we work with may have different **custom columns**:

* Some exports will have 0
* Others might include 1, 2, or even 5 extra fields

If we try to hardcode the script to **only keep specific columns**, we risk accidentally deleting useful custom fields unless we predefine every variation — which becomes impossible to scale as we add more sites.

**🔧 The Smarter Approach**

Instead of whitelisting what to keep, we’ll take the opposite route:

✅ We’ll identify all the **default Scan Unlimited columns**  
✅ The script will delete those if they appear  
✅ What’s left behind = **our core + any custom columns**

This ensures:

* Custom data always survives
* New columns from future sites are preserved automatically
* The script remains flexible even if **Scan Unlimited changes the column order or layout**

💡 **Side Tip:**  
This approach gives us *guard rails without rigidity* — perfect for scaling across dozens of lead list sources.

**Column Strategy – What Gets Deleted?**

Scan Unlimited exports over **36 default columns**, most of which we don’t need for Keepa merging. To future-proof our script and minimize clutter, we’ll proactively clean up anything unnecessary.

**🔍 How We Decided What to Delete**

This is a **one-time setup** step to future-proof our logic:

1. ✅ Enabled **all columns** inside Scan Unlimited
2. ✅ Downloaded the file
3. ✅ Transposed the headers to inspect every column name side-by-side

From there, we created **two categories**:

* Columns to **keep**
* Columns to **delete**

**See reference file Header Review in Reference Files**

We’ll copy/paste the “delete” list into our script and tell ChatGPT:  
👉 *“If you see these columns in the export, delete them.”*

💡 **Important:**  
⚠️ Do **not** include any **custom columns** in the delete list.  
Example: Source, Tags — These are **site-specific fields** we want to preserve.

**🧹 Columns to Delete:**

* UPC
* Sales Rank
* Title
* Variation Parent
* Price
* Size
* Color
* Estimated Sales
* Offers
* Average90
* Average180
* Package Quantity
* Review Count
* Review Stars
* Category
* Package Height
* Package Length
* Package Width
* Package Weight
* Amazon Last Seen
* FBA Fee
* Referral Fee
* Closing Fee
* Fixed Additional Cost
* Shipping Cost
* Percent Additional Cost
* Margin
* Brand
* Variations
* Average30

**🧠 Why Remove Columns Like Sales Rank, Title, or Selling Price?**

We’ll be getting those values from the Keepa export.

If needed, we can always reference the raw SU export — but for the purposes of our **scripted output**, we keep only what’s essential.

We’ll also be recalculating **Profit and ROI** using Keepa values. That means:

* We keep Profit/ROI for now for what did calculate correctly
  + We’ll use this to eliminate unprofitable inventory from Scan Unlimited and limit the amount of data that goes to Keepa

**📄 File to Open:**

Follow prompt chain P1 SU Clean Up in Tony Folder

Run the script when done:   
📂 Phase 1 - SU Clean Up.py

This is the first step in your automated sourcing list pipeline.

**✅ Summary:**

You’ll have a fully working script that:

* Auto-cleans your SU export
* Deletes junk ROI rows
* Extracts ASINs and preps them for Keepa upload
* Creates a repeatable, no-guesswork starting point for sourcing