Astronomical Stats Lookup

Pipeline Reference & Methodology Summary

Version 3.0 · Constellation Guide Data Pipeline

1. Purpose

The Stats Lookup script populates an astronomical observation-planning spreadsheet with the physical and observational parameters needed to plan a night at the eyepiece. It takes a spreadsheet in which each target is identified only by its **catalog ID** and **object type**, researches the missing data from authoritative astronomical catalogs, and writes back a fully populated, consistently formatted sheet. It is one numbered stage in a larger, reusable pipeline for producing constellation guide spreadsheets, and is designed to be re-run on any constellation dataset with minimal setup.

The guiding design principle is **data integrity over convenience**: the script never guesses, averages, or silently discards. Where a value cannot be confirmed, it is flagged rather than fabricated, and any value already present in the source sheet is preserved rather than overwritten.

2. What the Script Does

Working from the target list, the script performs the following operations in order:

- **Reads** the source spreadsheet, recording each object's ID, type, and which cells already contain values.
- **Inserts** an Equipment column (if not already present) immediately after the Type column, preserving all original formatting.
- **Researches** each object using live web sources appropriate to its type (see Section 5).
- **Fills blank cells only** — any value already in the sheet is left untouched, protecting hand-verified or previously researched data.
- **Calculates** surface brightness and the recommended equipment class from the researched figures (see Section 4).
- **Flags** unconfirmed, conflicting, or tentative values by highlighting the cell yellow and attaching an explanatory comment naming the issue and the source to consult.
- **Validates** the output file's internal structure, then saves it under a fixed output name ready for the next pipeline stage.

3. Expected Inputs

The script expects an Excel (.xlsx) file with a header row and one row per target. At minimum the **ID** (column A) and **Type** (column B) columns must be populated; every other column may be blank or partly filled. The full 17-column layout is:

Col	Header	Role
A	ID	Catalog identifier — the lookup key (source)
B	Type	Object type code — governs which fields apply (source)
C	Equipment	Binoculars / Telescope class — calculated
D	Magnitude	V-band magnitude, or carbon-star maximum — filled

Col	Header	Role
E	Size	Angular size in arcminutes — filled
F	Surface Brightness	mag/arcsec² — calculated
G / H	mBortle / mApt	Reserved for later pipeline stages (left blank)
I / J	Name / Components	Aliases and component notes (source)
K	Magnitude 2	Secondary magnitude, or carbon-star minimum — filled
L	Separation	Double-star separation in arcseconds — filled
M	Position Angle	Double-star position angle in degrees — filled
N	Period	Orbital or pulsation period in years — filled
O / P / Q	ColorDS / M Class / References	Descriptive metadata (source)

Object type codes recognised include GC (globular cluster), OC (open cluster), PN (planetary nebula), DN (dark nebula), CS (carbon star), DS (double star), MS-X (multiple star with X components), AS (asterism), and Star (single star). Multiple-star rows may carry a component-pair suffix in the ID — for example `67 Ophiuchi (AB)` — so that each pair receives its own separation, position angle, and magnitudes.

4. Values Calculated

4.1 Surface Brightness

Surface brightness expresses how bright an extended object appears per unit area, in **magnitudes per square arcsecond (mag/arcsec²)**. It is the key predictor of whether a diffuse object will stand out against the sky background. It is computed from the integrated magnitude and the angular dimensions:

$$SB = m + 2.5 \cdot \log_{10}(\pi \cdot (a''/2) \cdot (b''/2))$$

where `m` is the integrated magnitude, and `a″` and `b″` are the major and minor axes converted to arcseconds (arcminutes × 60). The full diameter is passed to the formula, which halves it internally; passing a radius by mistake would yield a value roughly 1.5 magnitudes too faint. Open clusters and asterisms, which resolve into individual stars, are recorded as `N/A`; dark and stellar objects are left blank.

4.2 Equipment Class

Each target is assigned a recommended equipment class — Binoculars (B), Telescope (T), or both (B, T) — using two simple visibility thresholds:

- **Binoculars** if the double-star separation exceeds 15 arcseconds, or the object's size exceeds 15 arcminutes.
- **Telescope** if the object's size is under 60 arcminutes (1°), or the object is stellar (no meaningful size).

Objects satisfying both criteria are marked `B, T`; very large objects (≥ 1°) become binocular-only; single stars are always `B, T`. This column is always recalculated from the current size and separation, even where those were supplied in the source file.

5. Website Resources Utilised

The script draws each value from the recognised authoritative catalog for that object class. Sources are chosen for provenance and accuracy, and a value is ideally corroborated across two independent sources before it is treated as confirmed.

Object Type	Primary Source(s)	Data Retrieved
Globular clusters	Harris 1996 (2010 ed.), McMaster; SEDS MWGC	Integrated magnitude, half-light radius, apparent diameter
Open clusters / asterisms	SEDS, WEBDA, Wikipedia, Collinder/Melotte	Magnitude, angular size
Planetary nebulae	Strasbourg-ESO PN Catalogue, Kaler, SIMBAD	Magnitude, main-shell size
Dark nebulae	Barnard 1927 Atlas, LDN (Lynds 1962), SIMBAD	Angular extent
Double / multiple stars	Washington Double Star (WDS) / Stelle Doppie	Separation, position angle, period, component mags
Carbon / variable stars	AAVSO VSX, GCVS	Maximum and minimum magnitude, period

A key discipline embedded in source selection is the use of the correct physical quantity. For globular clusters, for example, the Harris catalog publishes a half-light radius rather than a visual diameter, and separately tabulated tidal radii can be several times larger; the script deliberately uses the photometric/apparent diameter for size and never the tidal radius, which would misrepresent the object's telescopic appearance.

6. Data Integrity & Source Logic

6.1 How Source Data Is Selected and Applied

The script follows a strict set of rules for turning catalog data into spreadsheet values:

- **Fill-blanks-only.** A value already present in the source sheet is always preserved. Only empty cells are populated, protecting the user's prior work.
- **Type-appropriate fields.** Each object type has a defined set of applicable fields; the script writes only those. Carbon stars, for instance, use the two magnitude columns as a maximum/minimum brightness range rather than as two separate components.
- **Correct units, enforced.** Size is always in arcminutes, separation always in arcseconds, and periods in years — with conversions applied and sanity-checked so that, for example, a pulsation period quoted in days is not mistaken for years.
- **Preferred over disputed values.** Where authoritative sources disagree — such as competing orbital-period solutions for a binary — both values are retained and the discrepancy is noted rather than one being silently chosen.

6.2 How Data Is Validated

Integrity is enforced at three levels:

- **Formula verification.** The surface-brightness routine is checked against a set of known benchmark objects before any value is written; if a benchmark fails, the run halts.
- **Confirmation flagging.** Every value that cannot be confirmed from an authoritative source — or a pre-filled value found to conflict with the catalogs — is highlighted yellow with a comment stating the specific concern and where to verify it. Unconfirmed data is never presented as though it were solid.
- **File-structure validation.** Before delivery the saved workbook is inspected to confirm its cells are in strict column order, guarding against the file corruption that raw manipulation can introduce and ensuring the spreadsheet opens cleanly.

Together these rules make the pipeline reproducible and trustworthy: the same source file yields the same output, every calculated value is traceable to a formula and a catalog, and every uncertain value is visibly marked for follow-up rather than hidden.

This summary describes the Stats Lookup stage of the constellation guide data pipeline. Object-type rules, source priorities, and pre-verified reference data are maintained in the accompanying prompt script and Python implementation.