File structure

SoniDash/
├── data/
│ └── SoniDash_KPI.csv
├── analysis/
│ └── SoniDash_Preprocessing.ipynb ← Google Colab notebook
├── web/
│ ├── index.html
│ ├── script.js
│ └── style.css
└── README.md

The project asks the question:

> "Can analysts detect anomalies and compare cross-variants better using sound?"

The code generates 2 files:

1. sonidash_out/luxury_cosmetics_cleaned.csv
    - Full processed dataset
    - Contains original columns, derived metrics (`total_visitors`, `revenue_est`, `conversion_est`)
    - Normalized columns (*_norm, *_z), per-brand normalizations (*_norm_by_brand)
    - Contains all audio mappings (pitch, volume, instrument)
    - This is the **complete analysis version** — great for debugging, visual analysis, or deeper exploration in Colab.
2. sonidash_out/SoniDash_KPI.csv
    - **lightweight dashboard-ready file** — stripped down to only what your **Tone.js web app** or **interactive dashboard** needs.

Column	Purpose
`brand`, `region`, `event_type`	grouping/filtering
`units_sold`, `sell_through_pct`, `revenue_est`	key KPIs
`value_norm`, `value_roc_norm`	normalized values for mapping
`pitch_hz`, `volume`, `instrument`	audio mapping parameters for Tone.js

- Have to load this csv file in the front end file (JS) to:
    - Draw charts with Chart.js or Vega-Lite
    - Play tones via Tone.js (using `pitch_hz` and `volume`)

KPIs in Luxury Cosmetics data

KPI in your data	Meaning	Analogous generic KPI
Units Sold	How many products were sold	Same as "Units Sold"
Sell Through %	% of stock sold (sales efficiency)	Like "Conversion Rate"
Avg. Daily Footfall	Avg. number of visitors per day	Like "Website Traffic"
Revenue Estimate (you derived this)	Sales × Price	Like "Revenue"
Conversion Estimate (you derived this)	Units Sold ÷ Total Visitors	Like "Conversion Rate"

KPIs and their mapped variables:

1. `sell_through_pct → pitch`
2. `units_sold → loudness`
3. `region → instrument`

- **Pitch** = how high the note is (data value)
- **Loudness** = how strong the note is (magnitude/change)
- **Instrument** = what kind of sound it is (which variant or region)

DERIVED KPIs (for graphs and charts) Formulae

1. Total visitors: avg_daily_footfall × lease_length_days
2. Revenue_est: units_sold × price_usd
3. Conversion_est: units_sold / total_visitors

Method	Range	Meaning	Used for
Min-max	0 – 1	Rescales proportionally	Sound mapping (pitch/volume)
Z-score	~−3 – +3	Distance from average	Anomaly or pattern detection