Dataset Generation Process

We built a mock dataset designed to reflect realistic demographics in Pakistan and the flavour/strength preferences for oral nicotine pouches that follow from them. The dataset is based on explicit assumptions.

Step 1. Build population and priors

* Each simulated respondent is assigned demographic attributes (region, metro/rural, gender, age band, occupation) according to target shares.
* From these attributes, we generate a prior probability distribution for both flavour category (Spice, Mint & Menthol, Fruit) and strength (Low, Medium, High).

Step 2. Sample choices

* From the category prior, we draw one flavour category.
* Within that category, three distinct flavours are drawn without replacement.
* Strength is drawn separately using the strength prior.

Priors and multipliers

Baseline category weights are set at Spice 0.30, Mint & Menthol 0.40, Fruit 0.30. These are then adjusted with multiplicative ‘bias multipliers’ based on demographics. After applying multipliers, probabilities are renormalized to sum to 1.

Examples of demographic-based adjustments:

Metro vs Rural

* Urban users: higher weight on café-style spice flavours (e.g. coffee/latte +25% within Spice), slightly higher Medium/High strength (+10%).
* Rural users: higher Mint & Menthol (×1.15) and simple fruit (citrus, grape).

Region

* North: small uplift for menthol (Mint & Menthol ×1.10).
* South: uplift for fruit (×1.10) and a modest boost to spice (×1.05).
* Central/SNB: close to baseline, with a tilt to mainstream Mint.

Gender

* Male: slightly more Medium/High strength (×1.10 each).
* Female: uplift on Fruit (×1.10) and Mint & Menthol (×1.05), tilt toward Low/Medium strength.

Age band

* 18-24: more experimental, small uplift for bold spice (×1.08) and novelty fruit (×1.05).
* 25-34: balanced, near baseline, slight tilt to Medium strength.
* 35+: stronger bias to Mint & Menthol (×1.10) and toward LoFw/Medium strength.

Application

1. Start with baseline weights.
2. Apply all relevant multipliers for metro/rural, region, gender, age.
3. Renormalize so probabilities sum to 1.
4. Repeat process for strength categories.
5. Within the chosen category, apply finer multipliers to specific flavours (e.g. for urban  creamy latte/coffee +20% and rural bright spearmint +10%), then sample three distinct flavours.

Sampling method

* Weighted random sampling with renormalization at each step.
* Flavours drawn without replacement to ensure flavours aren’t repeated
* Fixed random seed can be set for reproducibility.

This way the dataset mirrors plausible demographic differences in flavour and strength preferences while remaining reproducible and fully documented.