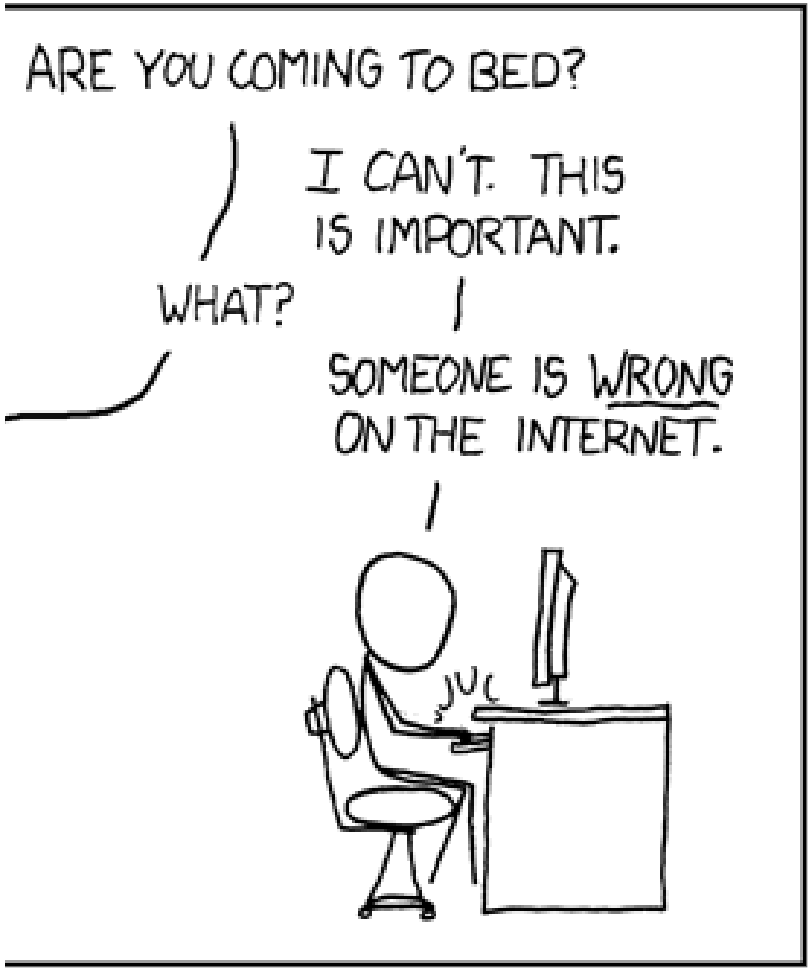
Being Bayesian

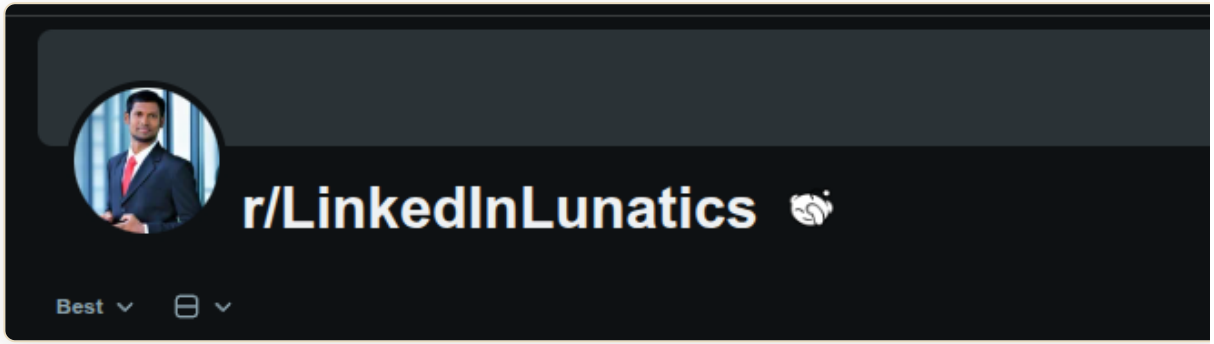
Stress-Testing India's Loudest Stories with Open Data and FOSS

Jaidev Deshpande · beingbayesian.in

IndiaFOSS 2026

How my sabbatical started





Quiz Time!



A loud opinion online, no data attached. It rests on one person's experience. **How much of India does that cover?**

	GROUP	SHARE OF ADULTS	SHARE OF THE ONLINE CROWD
Gender	Men	49.8%	56.8%
	Women	50.2%	**43.2%**
Age	Under 30	**34.2%**	49.8%
	30 and over	65.8%	50.2%
Education	Higher education	**14.3%**	**23.4%**
	Below higher	85.7%	76.6%
Region	Urban	**31.1%**	**38.7%**
	Rural	68.9%	61.3%

	GROUP	SHARE OF ADULTS	SHARE OF THE ONLINE CROWD
Income	Richest fifth	20%	26.7%
	Other four fifths	80%	73.3%

Shading compares the two groups within each pair, column by column · Share of adults: PLFS 2023–24, survey-weighted microdata (income fifths are 20% by definition) · Share of the online crowd: Bayes' rule on PLFS shares and Data For India internet-use rates, 2023

If they come with no evidence, the average influencer is speaking for no more than **1 in 12** Indians.

> *"What can be asserted without evidence can also be dismissed without evidence."*
>
> — Christopher Hitchens

INFLUENCER 1 OF 4

The one who's trying to sell you something



> *"Kirana stores are dying. Here's how they can bridge the gap."*

A PRODUCT, PITCHED AS A UBIQUITOUS, TRANSFORMATIVE SOLUTION · CASE 1: KIRANA STORES

Here's how they can bridge the gap:

→ **Tech Integration**
Adopting POS systems, digital payment options, and inventory management tools.

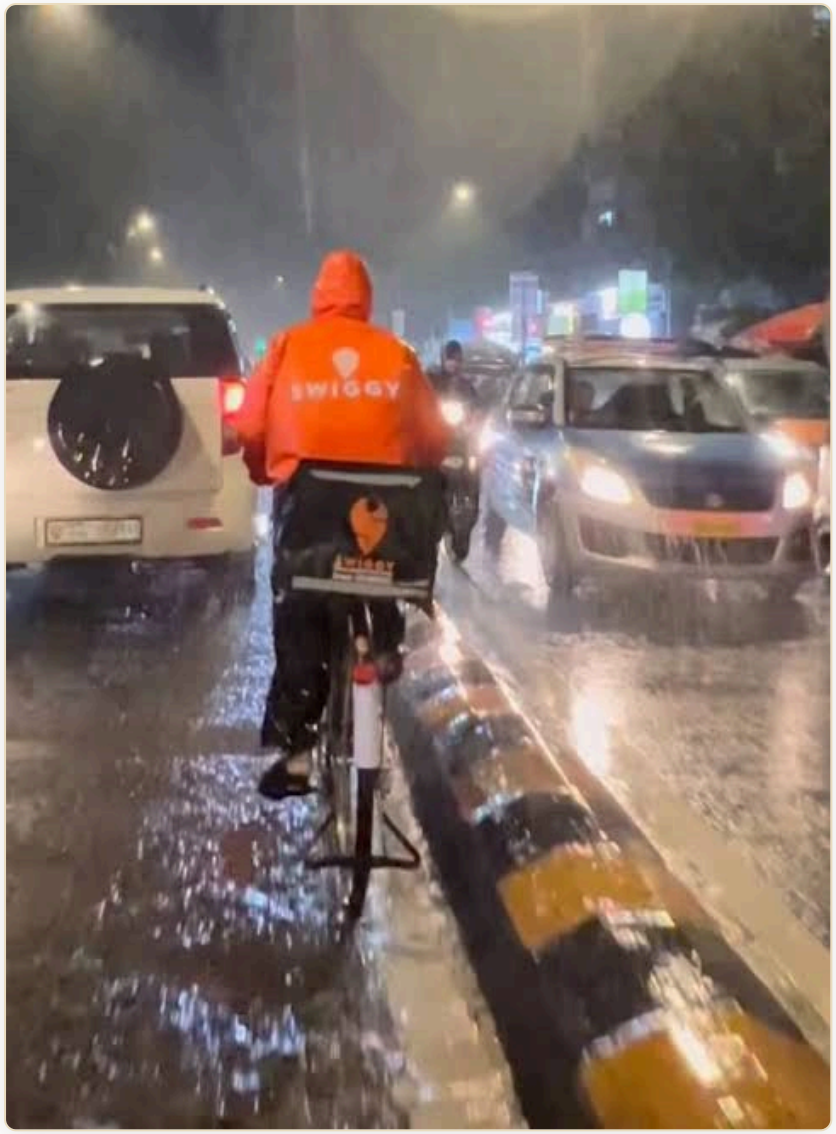
→ **Omni-Channel Presence**
Partnering with q-commerce apps to offer hyperlocal deliveries.

→ **Customer Experience**
Leveraging data analytics to provide personalized promotions and services.

→ **Training & Upskilling**
Learning how to optimize stock, pricing, and promotions through digital channels.

INFLUENCER 2 OF 4

The jingoist



> *“Leave our founders alone. It's not exploitation, it's public service!”*

CRITICISM OF A BUSINESS, RECAST AS AN ATTACK ON THE NATION · CASE 2: GIG WORK

INFLUENCER 3 OF 4

The pointless contrarian



> *"Work-life balance is a western concept. I work 140 hours every week."*

CONTRARIAN FOR ITS OWN SAKE · CASE 3: THE 90-HOUR WEEK

INFLUENCER 4 OF 4

The bully

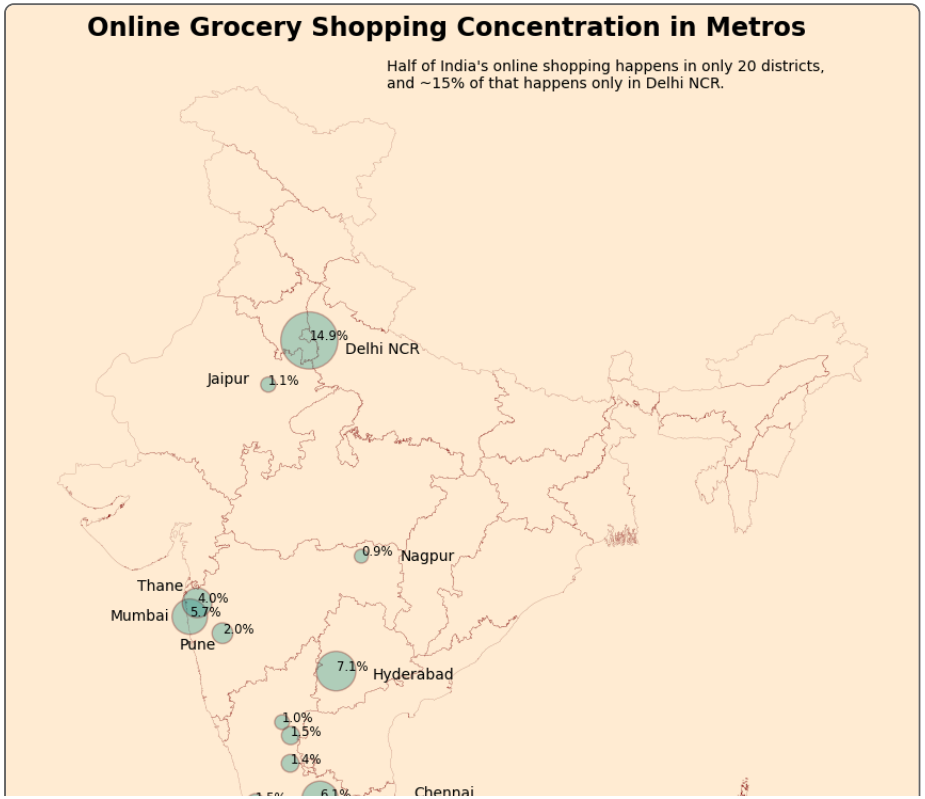
> *"I am heavily taxed. Income tax is the only kind of tax there is. It funds freebies, which makes people lazier. I get nothing."*

PUNCHING DOWN, AT THE PEOPLE WITH THE LEAST · CASE 4: WELFARE



CASE 1 · RETAIL

Quick commerce is killing kirana stores

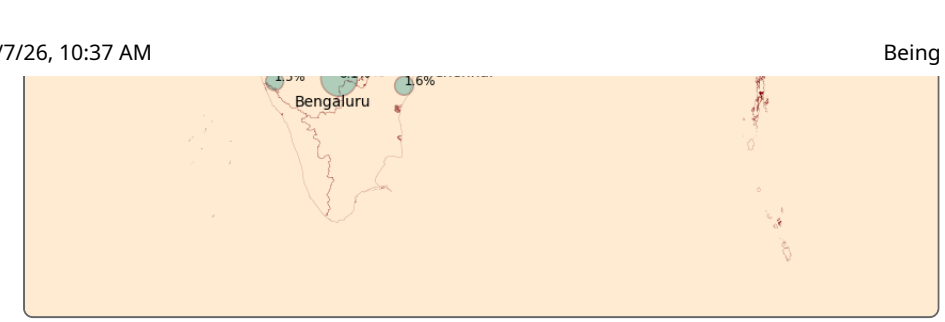


> *"About 2,00,000 kirana stores have shut down in the last year because of predatory e-commerce practices."*
>
> AICPDF (TRADERS' FEDERATION), WIDELY RESHARED AS BREAKING, SHOCKING DATA

GUT CHECK **200,000 sounds huge. Huge compared to what?**

INDIA HAS ABOUT 13 MILLION KIRANA STORES

200,000 closures against ~13M kirana stores	**1.5%** national closure rate
... even limited to metro India alone	**5%**



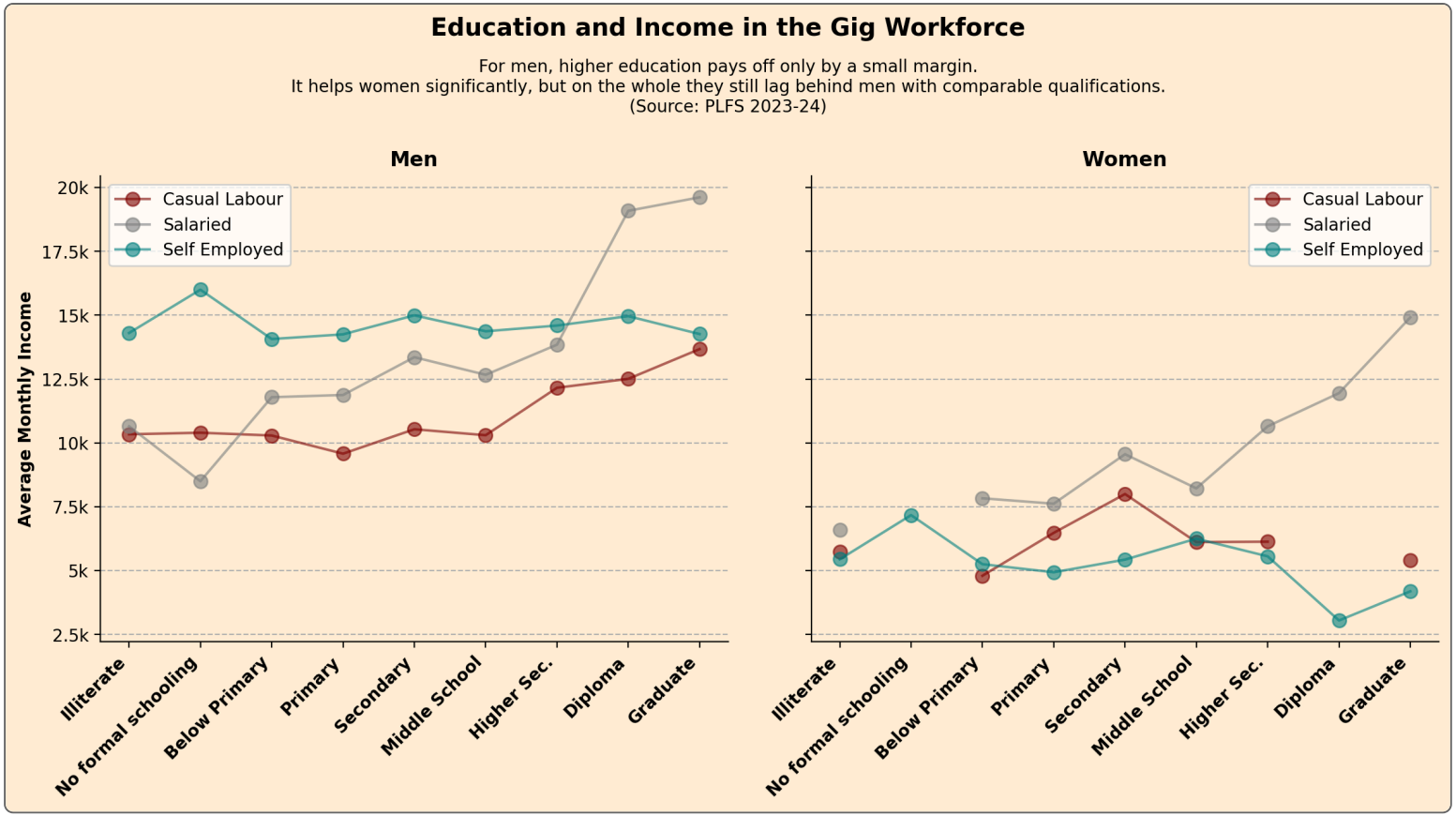
Households that shop online for groceries at all	**7%** — 93% stay offline
… even inside urban India specifically	**83%** still don't shop online
Historical base rate: Amazon/Flipkart-era, COVID lockdown surge	displaced only **1–2%**, then **~0.5%**

HCES 2023–24

CASE 2 · LABOUR

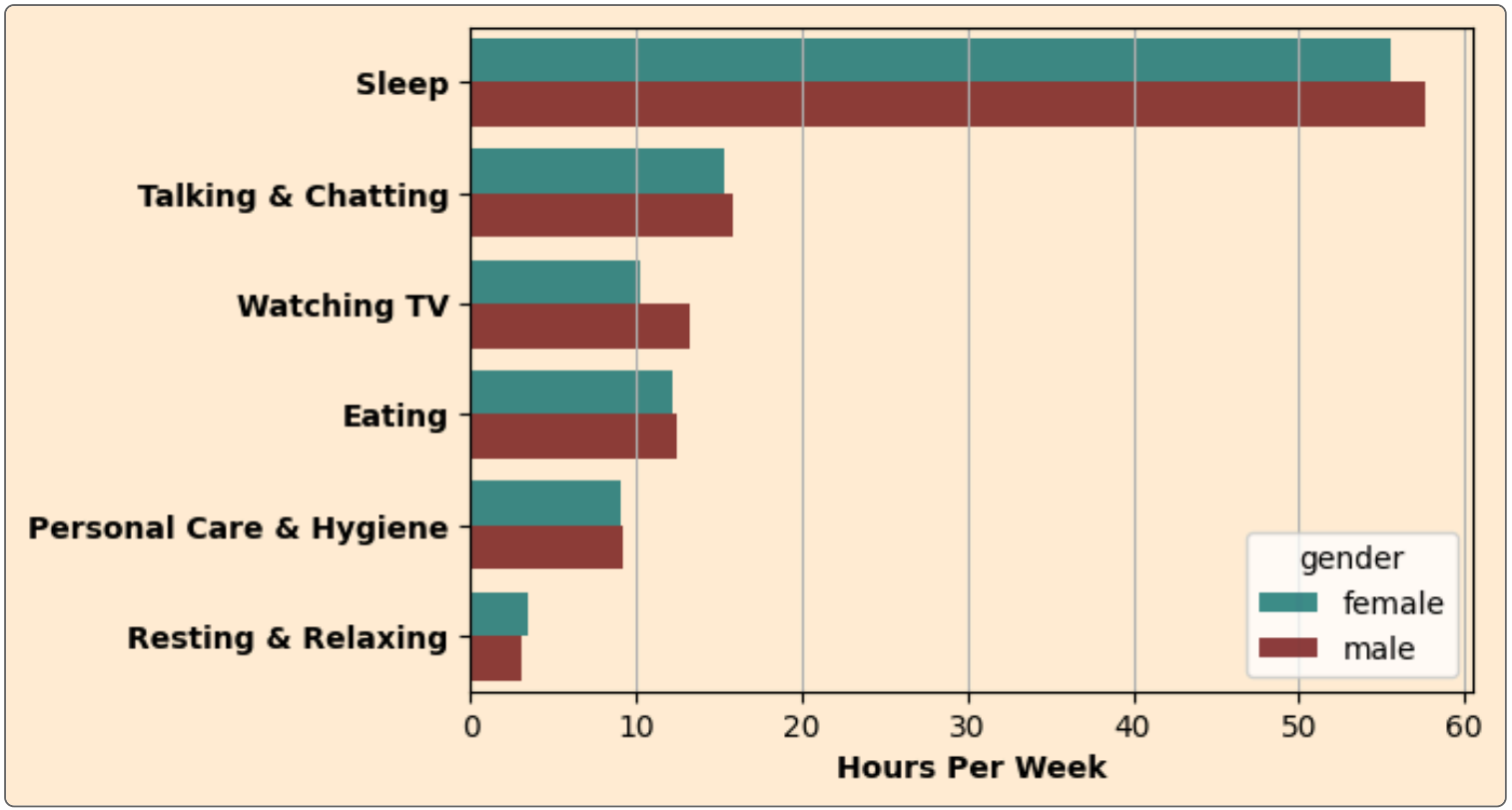
Gig work liberates India's poor and uneducated

> *"Workers who once faced barriers like lack of education or technical skills now earn ₹15,000 to ₹45,000 or even more."*
>
> THE PLATFORM-ECONOMY NARRATIVE, ECHOING NITI AAYOG'S OWN OPTIMISTIC FRAMING


Education and Income in the Gig Workforce
For men, higher education pays off only by a small margin.
It helps women significantly, but on the whole they still lag behind men with comparable qualifications.
(Source: PLFS 2023-24)
Men
Women
Average Monthly Income
20k
17.5k
15k
12.5k
10k
7.5k
5k
2.5k
Casual Labour
Salaried
Self Employed
Illiterate
No formal schooling
Below Primary
Primary
Secondary
Middle School
Higher Sec.
Diploma
Graduate

CASE 3 · WORK & CARE

The 90-hr Workweek



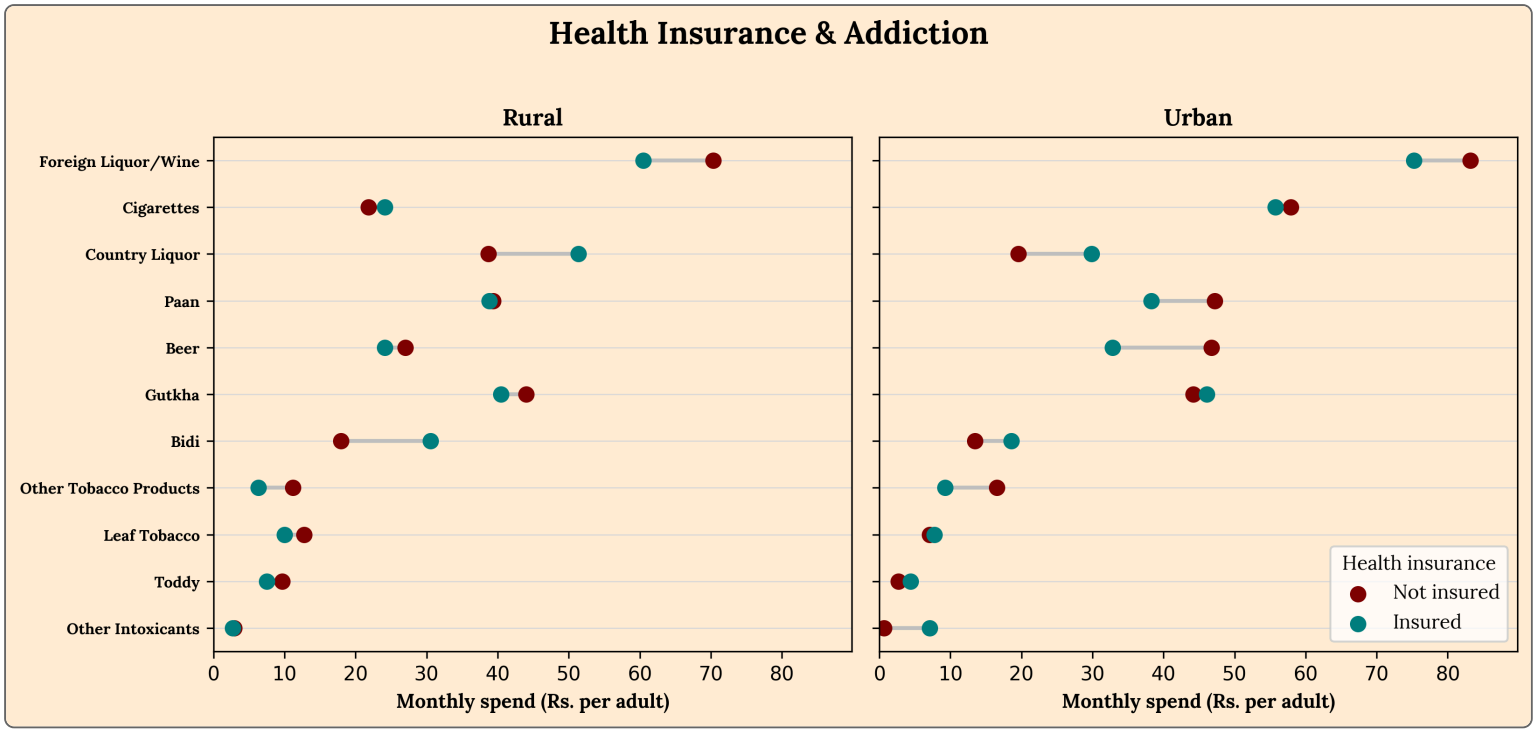
Best case: both genders cap chores at 1 hr/day

`women's ceiling rises to only` **`~67 hrs/wk`**

To reach 90, you'd have to cut	~1 hr sleep/night, all TV (10–12 hrs/wk), most talk & text time (15 hrs/wk)
Deficit still remaining after those cuts	women **22 hrs** · men **34 hrs**

CASE 4 · WELFARE

Welfare makes people lazy — and wasteful



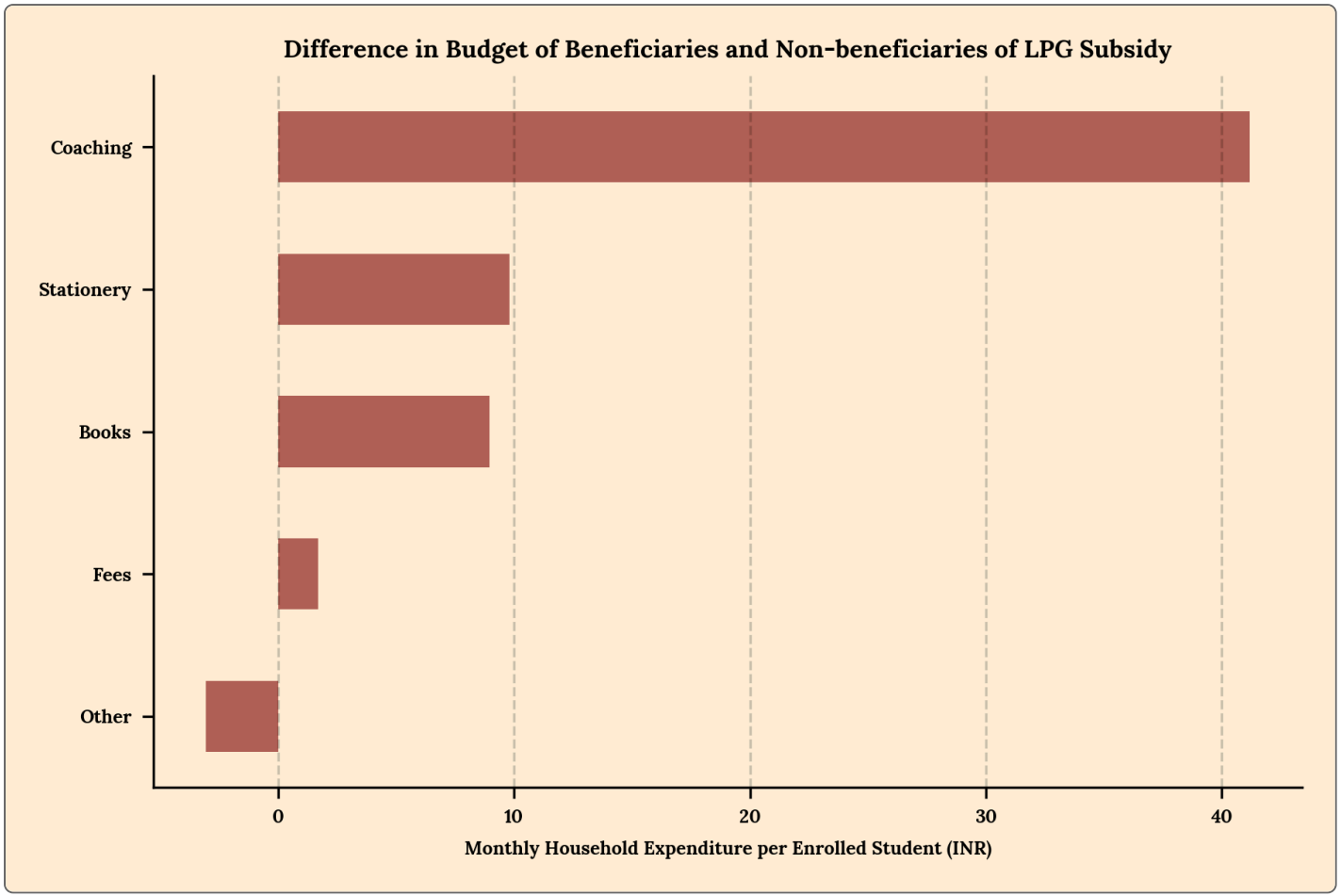
Urban PMJAY/scheme beneficiaries vs. non-beneficiaries	spend **₹2/month more** on gutkha
Rural beneficiaries vs. non-beneficiaries	spend **₹4/month less**

Entire spread across every group in the chart

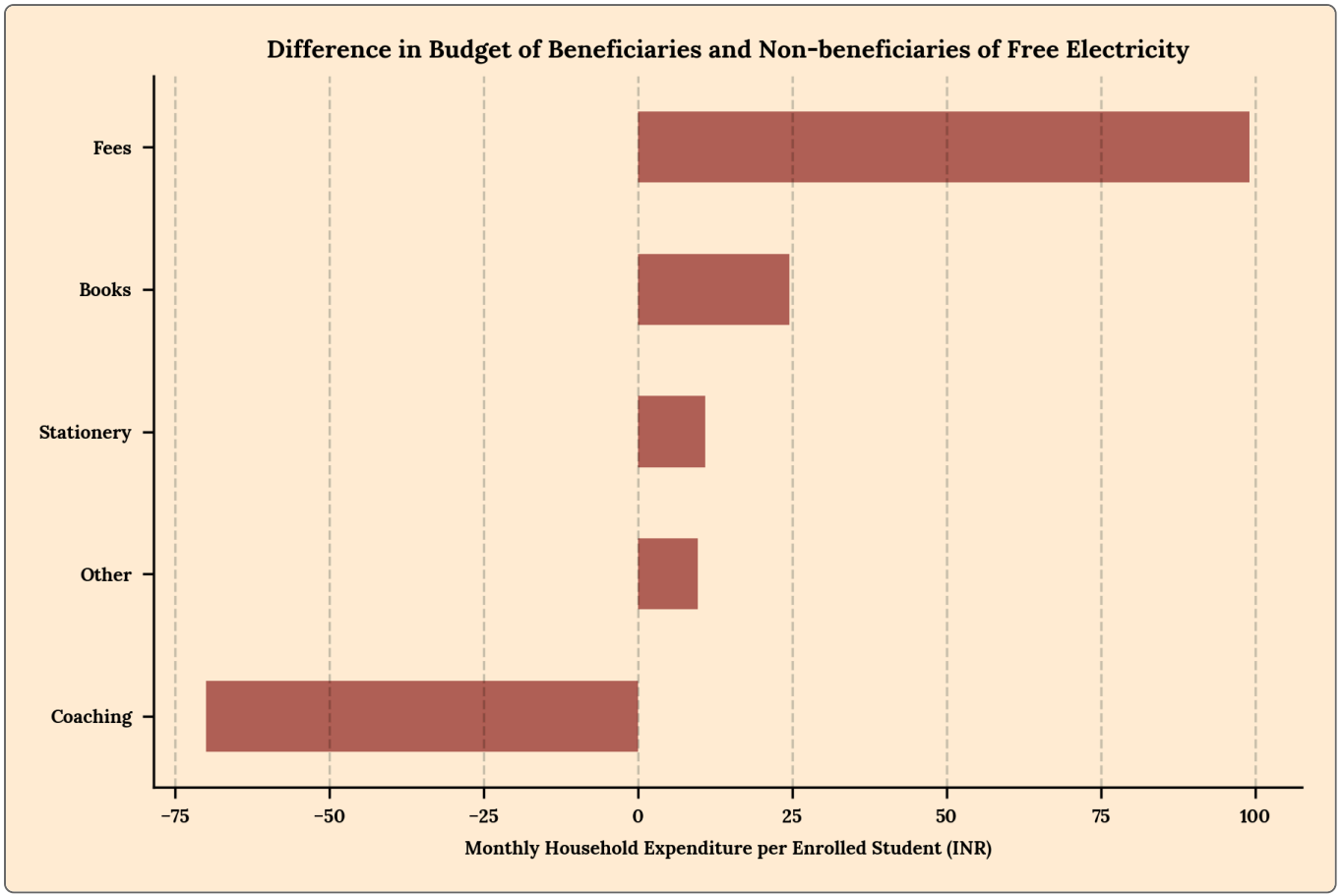
±₹15

HCES, propensity-score matched on household size, caste, religion & region — not a raw comparison

But welfare's real effects hide where headlines don't look



LPG-subsidy households spend nearly **₹60 more** on education per student — and two-thirds of that goes straight into private coaching.



LPG Subsidy → Education	+₹59 total — **+₹41 to coaching** (published)

Health Insurance → Education	−₹44 total, but **+₹54 to coaching** — hidden by fees falling −₹74 (never published)
Free Electricity → Education	+₹74 total, but **−₹70 from coaching**, +₹99 to fees — the opposite direction entirely

Elements of the Analysis

1 · The Data

2 · Survey Design & Interpretation

3 · The Maths & Stats

4 · LLMs

ELEMENT 1 · THE DATA

Looking for Data, and Inspiration



microdata.gov.in

HCES · PLFS · Time Use Survey

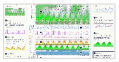
Data{Meet}

datameet.org

Data For India

dataforindia.com

Diagram Chasing

Visual stories  and tools 

that make public data worth looking at

diagramchasing.fun

BharatViz

Mapping India

bharatviz.org

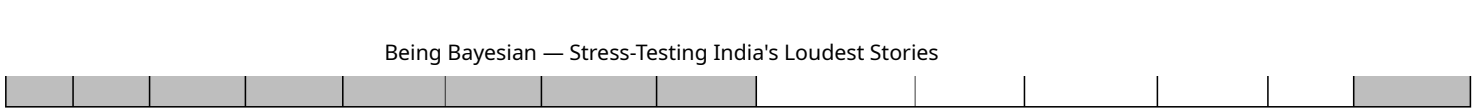
ELEMENT 2 · SURVEY DESIGN & INTERPRETATION

Q5.1 What is the quantity and value of the ***cereals and cereal substitutes*** consumed by the household during last 30 days?

Item	item code	quantity and value of the item consumed during last 30 days				
		consumption out of home produce		total consumption (incl. consumption out of home produce)		source#
		quantity* (0.000)	value (Rs.)	quantity* (0.000)	value (Rs.)	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
rice-free$	061					1
wheat/atta- free$	062					1
coarse grains&- free$ (jowar, bajra, maize, barley, ragi, small millets)	070					1
rice – PDS						
rice – other sources						
chira						
muri						
other rice products (khoi/lawa,etc.)						
wheat/ atta – PDS						
wheat/ atta – other s						
maida						
suji, rawa						
vermicelli (sewai)						
other wheat produc						
coarse grains-PDS& (jowar, bajra, barley, ragi, small r						
coarse grai sources&(jowar, maize, barley, rag millets & its produc						
other cereals& proc						
cereal: sub-total						
cereal subs tapioca, etc.						

[6] particulars of activities for each household member of age 6 years and above from 4:00 AM on the day before the date of interview to 4:00 AM on the day of the interview

srl. no. (as in item. 1, bl. 5)	age (years) (as in item 2, bl. 5)	number of activities performed in the time slot	srl. no. of activities	record in 24- hour clock format		whether performed multiple activity in the time slot (*yes-1 no-2*)	**if 1 in col. 6**, whether simul-taneous activity (*yes-1, no-2*)	description of the activity (at most 3 activities for each 30 minutes time slot)	whether a major activity or minor activity (*major- 1 / minor - 2*)	3-digit activity code of TUS classification of activities	where the activity was performed? (code)	unpaid/ paid status of the activity (code)	**if any of 07 to 18 in col. 12,** enterprise type (code)
				from 4:00 AM on the day before the date of interview	to 4:00 AM on the day of the interview								
(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1	60	1	1	4:00	6:00	2		Night sleep	1	911	1	01	
1	60	1	2	6:00	7:00	2		Personal hygiene	1	931	1	01	
1	60	3	3	7:00	7:30	1	1	Drinking tea	1	922	1	01	
1	60	3	4	7:00	7:30	1		Reading newspaper	2	841	1	01	
1	60	3	5	7:00	7:30	1		Talking with household members	2	711	1	01	
1	60	1	6	7:30	8:30	2		Watching morning news in TV	1	842	1	01	
1	60	1	7	8:30	9:30	2		Teaching own children	1	413	1	02	
1	60	3	8	9:30	10:00	1	1	Eating meal	1	921	1	01	
1	60	3	9	9:30	10:00	1		Dressing	2	931	1	01	
1	60	3	10	9:30	10:00	1		Talking with household members	2	711	1	01	



ELEMENT 3 · THE MATHS & STATS

It all comes down to comparing aggregates of separate groups. Then look for statistical and numerical significance.

WEIGHTED AGGREGATE OF A GROUP

$$\bar{y}_g = \frac{\sum_{i \in g} w_i \, y_i}{\sum_{i \in g} w_i}$$

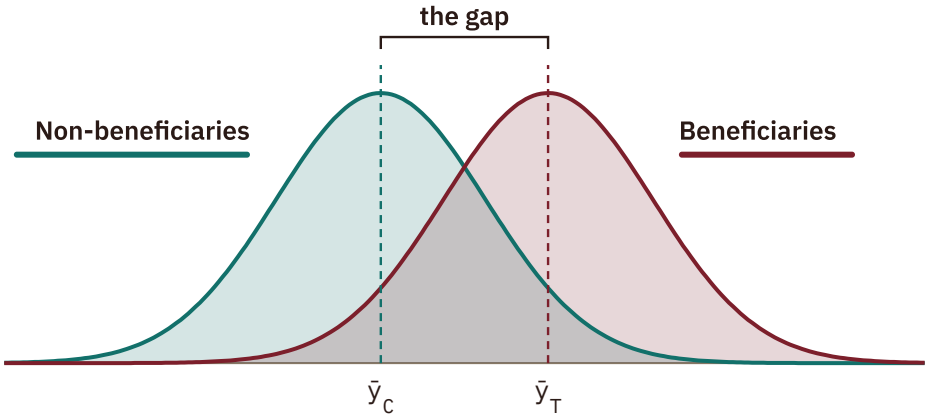
w_i = the survey multiplier: how many households row i stands for

ORDINARY LEAST SQUARES

$$y_i = \beta_0 + \beta_1 T_i + \gamma^\top x_i + \varepsilon_i$$

β_1 = the gap between the groups, holding everything in x_i fixed

TWO-SIDED T-TEST



Is the gap between the means big, next to the spread within each group?

ELEMENT 4 · LLMS

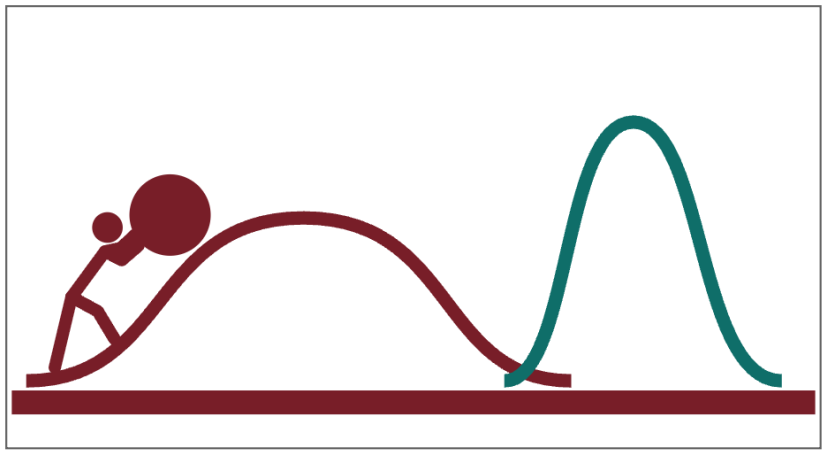
LLMs

- Survey documentation runs to **hundreds of pages**
- Almost everything is **coded**
- Questionnaires are **not linear**
- LLMs ***will*** hallucinate

PRO-TIP **Write unit tests**

A Confession

> *"The amount of energy needed to refute bullshit is an order of magnitude bigger than that needed to produce it."*
>
> — Alberto Brandolini, the Bullshit Asymmetry Principle



"One must imagine Sisyphus happy."

— Albert Camus, *The Myth of Sisyphus*

beingbayesian.in



Speaker notes