

Marketing & Data Analysis

Data sources

Santiago Alonso-Díaz

Tecnológico de Monterrey
EGADE, Business School



Photo: Dalle2

Table of contents

1 External Data

2 Internal Data

3 Synthetic Data

4 Python: loading, cleaning, exploring

- Data merging & Anonymity
- Missing data & Duplicates
- Outliers
- Transformations
- Validation

5 References

External Data

Definition

Relevant data that is not produced by the company's activity

Examples

- Demographics
- Market shares
- Economic indicators
- Social media trends

Some sources

- Google Trends
- Statista
- INEGI
- Kaggle
- World Bank
- Passport Euromonitor
- Pew Research Center
- Mintel
- Web scrapping
- Data brokers

Some samples

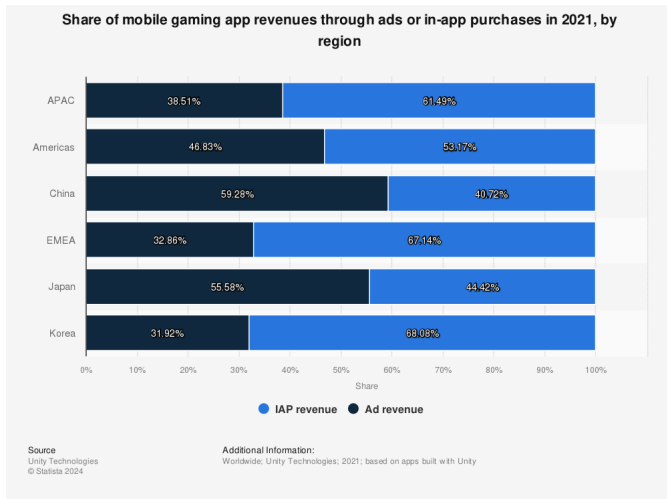


Figure: Statista

Some samples

% of adults in the New York City area who get local news from each type of provider

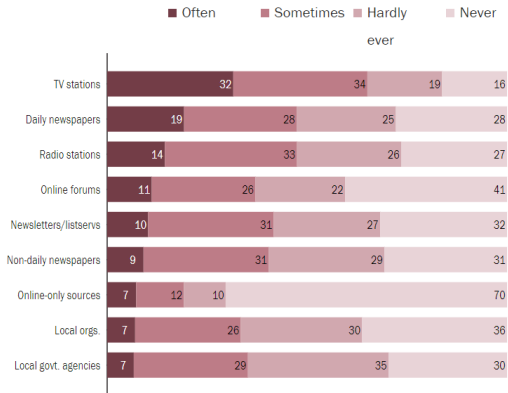


Figure: Pew Research Center

Some samples



Figure: Spotify Song Features. Kaggle

Some samples

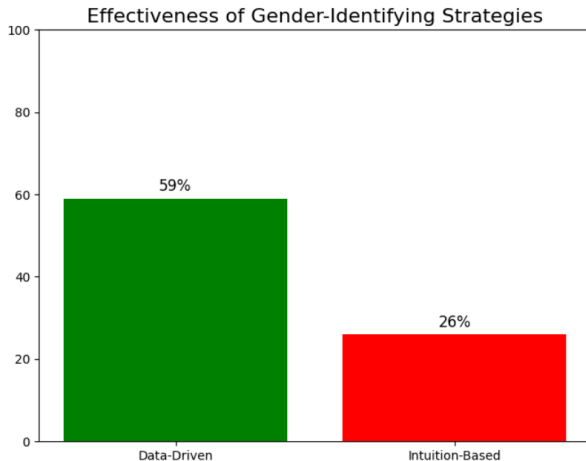


Figure: With data broker's we can detect 59% of males correctly. In Internet, 26.5% are males, 25-54, a 121% gain ($59\% / 26.5\%$)(Neumann et al., 2019)

Internal Data

Definition

Relevant data that is produced by the company's activity and decisions

Examples

- User satisfaction
- Product revenue
- Segment growth
- Sale force efficiency
- Lifetime costumer value
- Costumer acquisition costs

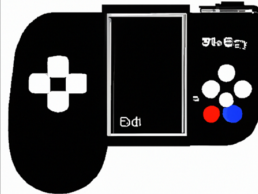
Some sources

- Web managers (e.g. Google Search Console, Google Analytics)
- Focus groups (e.g. with customers)
- Interviews (e.g. with sale force)
- Financial data (e.g. sales of company)
- Experiments (e.g. A/B testing, Conjoint)

A/B test

Students: What A/B tests can you think of? In your jobs, business, life?

Option A



Option B

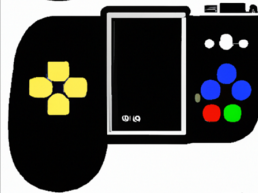


Figure: A/B testing

A/B test

Takeaway

"... experimentation helps drive both valuable incremental changes and the development of significant product improvements." (Koning et al., 2022, pp. 6436)

A/B test

Figure 2. Event Study Plot Showing the Effect of A/B Testing over Time

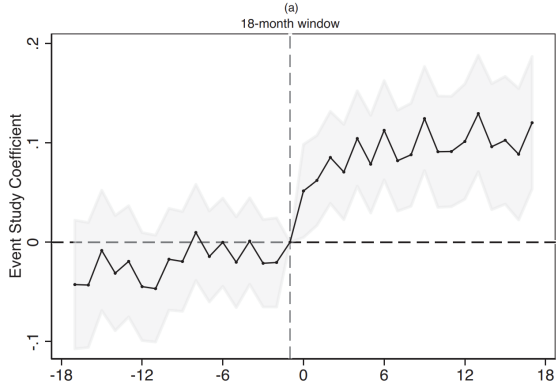
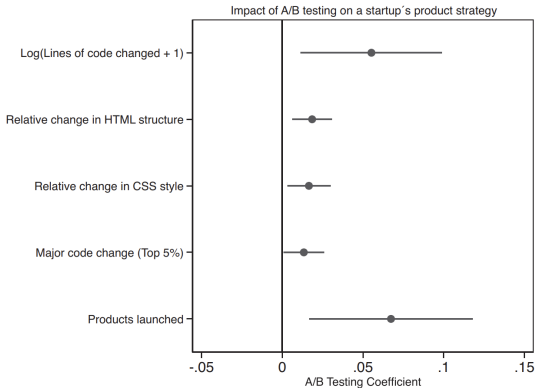


Figure: Website visits after A/B tests persists over time (Koning et al., 2022)

A/B test

Figure 4. We Find That A/B Testing Does Not Lead to Incrementalism in Product and Website Development for the Nearly 10,000 Start-ups for Which We Have Website and Product Launch Data



Notes. Instead, these firms make larger changes to their website code, the structure of their homepage's HTML, and website style and are more likely to deploy major code changes. A/B testing firms are also more likely to launch a new product in a given week than those that do not.

Figure: A/B on other variables (Koning et al., 2022)

A/B test

- Experimentation as a strategy (a single experiment may not work)
- Design the future with experiments
- Solving tension between routines and experimentation is critical
- Experimentation should aid innovation

Synthetic Data

Definition

Definition

Data that is either the result of artificial process (e.g. mathematical models; e.g. 2 gen. AI) or that is constructed with the help of existing data (e.g. synthetic controls).

Some Samples

- Structural models
- Synthetic controls
- Agent based models
- Generative AI (e.g. LLM labels)

NYT. Synthetic Controls

Sometimes we don't have an appropriate control for comparison.

One approach is to create synthetic controls: a weighted combination of other units so that it is as similar to treated units

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

Is a paywall good or bad for newspapers? There are at least two externalities:

- Engagement of its online reader base
- Spillover effect on the print version

How to measure them for the New York Times? Create synthetic controls using other similar newspapers

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

The New York Times
Marketing Strategy:
A paywall

You look like someone
who appreciates
a good story.

Try a Digital Subscription to The New York Times for just €1
and get unlimited access to NYTimes.com and NYTimes apps
for smartphones and tablets.

SEE MY OPTIONS

No, thanks

Already a subscriber? Log in

→ Sales increase



Confounding:

- Economy grew
- Quality of articles increased
- Internet access grew
- The paywall

Figure: Problem we try to solve: confounding

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

Mix them



And get a synthetic control

The New York Times

The New York Times

Figure: Solution: Mix and get a (mostly) similar clone

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

Table 4. Effect of the Paywall on *NYT* Online Visitation, Aggregate Data, and Generalized Synthetic Control.

	ln(Unique Visitors)		ln(Pages)		ln(Visits per Visitor)		ln(Pages per Visitor)		ln(Duration per Visitor)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
<i>NYT</i> × Paywall	−.184**	.029	−.428**	.073	.010	.125	−.104	.127	−.112	.148
# Observations: treated						1,025				
# Observations: control						5,125				

** $p < .01$.

Figure: Negative online effect of paywall for the NYT (relative to the synthetic control)

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

Table 11. Effect of Paywall on Print Readership, Generalized Synthetic Control.

DV =	All DMAs			
	Weekday Circulation Share (%)		Weekend Circulation Share (%)	
	Est.	SE	Est.	SE
NYT \times Paywall	.35**	.02	.34**	.03
N (treated)		202		
N (control)		404		

Figure: Positive print effect of paywall for the NYT (relative to the synthetic control)

NYT Synt. Controls (Pattabhiramaiah et al., 2019)

Business implications

- Paywalls do affect the NYT business
- Managers need cost-benefits analyses because paywalls in the NYT had different effects for online and print versions.

Python: loading, cleaning, exploring

Python

Let's go to Python to find, load, clean, and explore some data

References



Koning, R., Hasan, S., & Chatterji, A. (2022).Experimentation and start-up performance: Evidence from a/b testing. *Management Science*, 68(9), 6434–6453.



Neumann, N., Tucker, C. E., & Whitfield, T. (2019).Frontiers: How effective is third-party consumer profiling? evidence from field studies. *Marketing Science*, 38(6), 918–926.



Pattabhiramaiah, A., Sriram, S., & Manchanda, P. (2019).Paywalls: Monetizing online content. *Journal of marketing*, 83(2), 19–36.