Unpacking the Impact of High-Speed Rail on the Short-Term Rental Market

A Machine Learning Approach Using Airbnb Data in Italy

Fulponi J. I.[1], *Tartaglia M.*[2], *Nourbakhsh S.*[2]
[1]Freelance; [2]Ferrovie dello Stato Italiane

10th September 2025

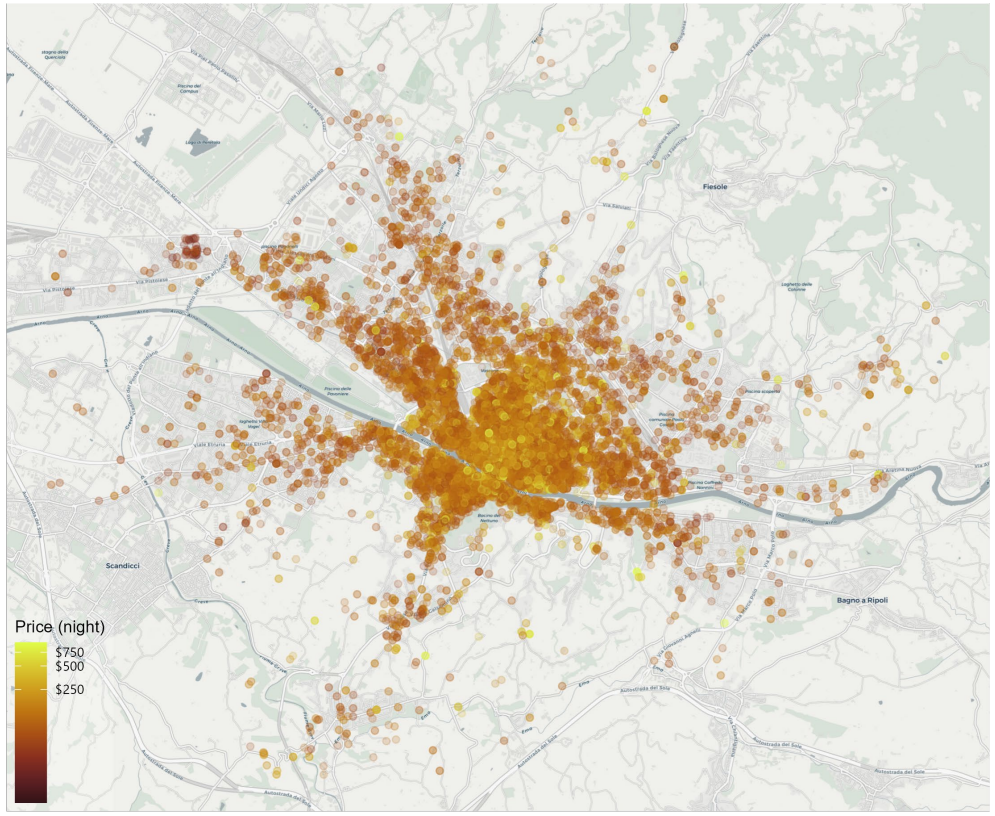
Gruppo FS
FS Research Centre
Il Centro Studi di Ferrovie dello Stato Italiane

Objectives

- Evaluate the spatial and socio-economic impacts of HSR infrastructure on Italian Real Estate Market, especially in the Short-Term Rental (STR) Market, through quantitative and geospatial methods.

Specific Objectives:

- Measure the effect of HSR on STR availability, prices, and intensity.
- Apply sentiment analysis on Airbnb reviews to assess perceived accessibility.
- Identify heterogeneous impacts across central, tourist, and peripheral areas.



Airbnb rentals in Florence (2010-2015)

Literature review and gaps

Main references:

- Baş & Yüzer (2017) – Spatial impacts of HSR in Turkey

  Investigates the Ankara–Konya HSR line's effects on population distribution, labor accessibility, travel time reduction, and urban sprawl. Combines a new accessibility model with socio-economic data and passenger interviews.
- Pagliara & Mauriello (2020) – HSR and tourist destination choice

  Analyzes how HSR influenced tourist behavior across 99 Italian provinces (2006–2016), using Geographically Weighted Poisson regression. Shows positive effects on destination choice and tourist flows.
- Wang et al. (2019) – HSR and spatial structure in China

  Uses difference-in-differences models to assess HSR's impact on urban spatial structure (centralization vs. dispersion) in Chinese prefecture-level cities, finding effects stronger in industry than population patterns.
- Gallo & La Rocca (2022) – Tourist attractiveness in Italy

  Studies 111 Italian provincial capitals (2018) and finds that availability of HSR services significantly increases tourist arrivals and demonstrates strong elasticity. HSR is the only significant accessibility variable.
- Tartaglia et al. (2024) – OMI-level HSR effects in Italy (basis for this project)

  This microscale study serves as the foundation for our project's spatial disaggregation — inspiring the move from OMI zones to listing-level effects.

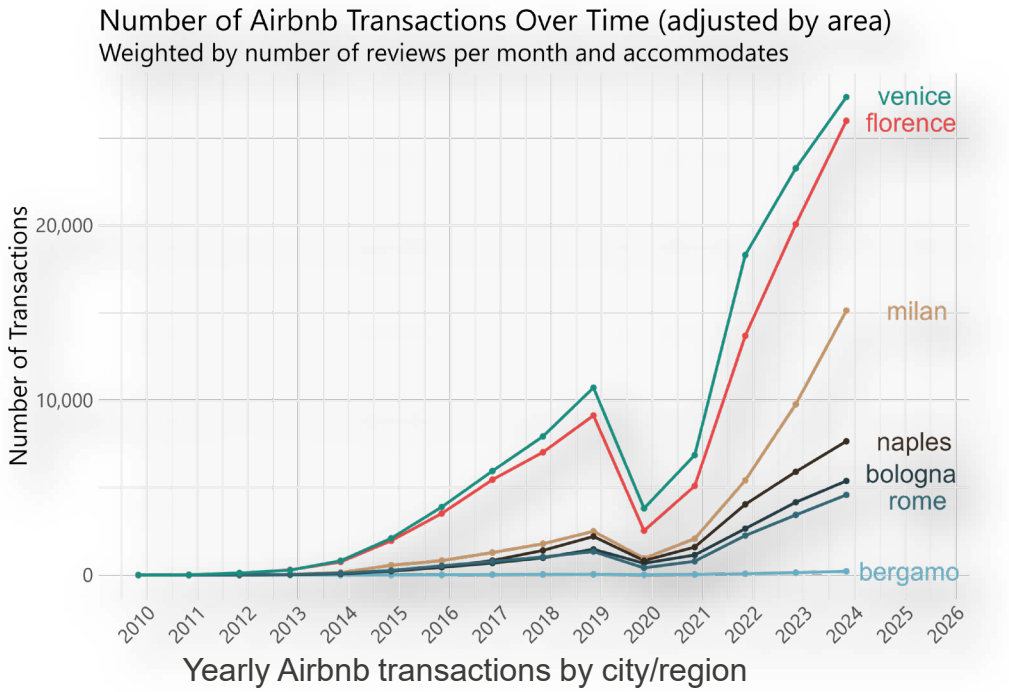
Gaps:

- Macro vs Micro: Prior studies often neglect intra-urban dynamics.
- Spatial models applied to provinces or communes but not to neighbourhoods.
- Real Estate Impacts: Limited use of georeferenced digital sources; mostly aggregated.

Data Sources: Real Estate

Airbnb database:

- Granular Listings Data and High Spatial Resolution
- Daily-level booking and availability data
- Nightly rates, average daily rates (ADR), and occupancy levels
- Review Scores & Sentiment
- Room Type Segmentation
- 2010-2025, selected cities



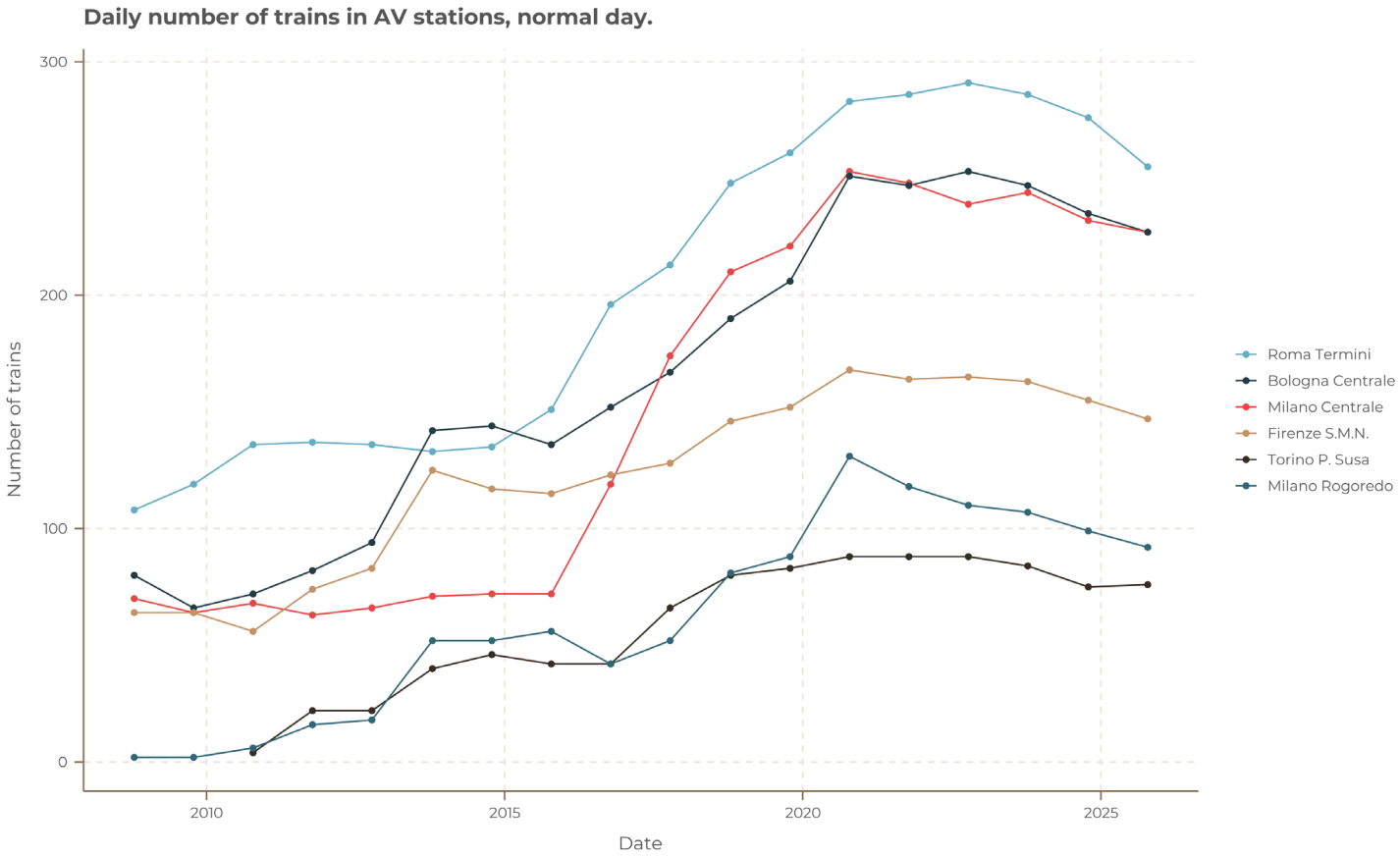
Yearly Airbnb transactions by city/region



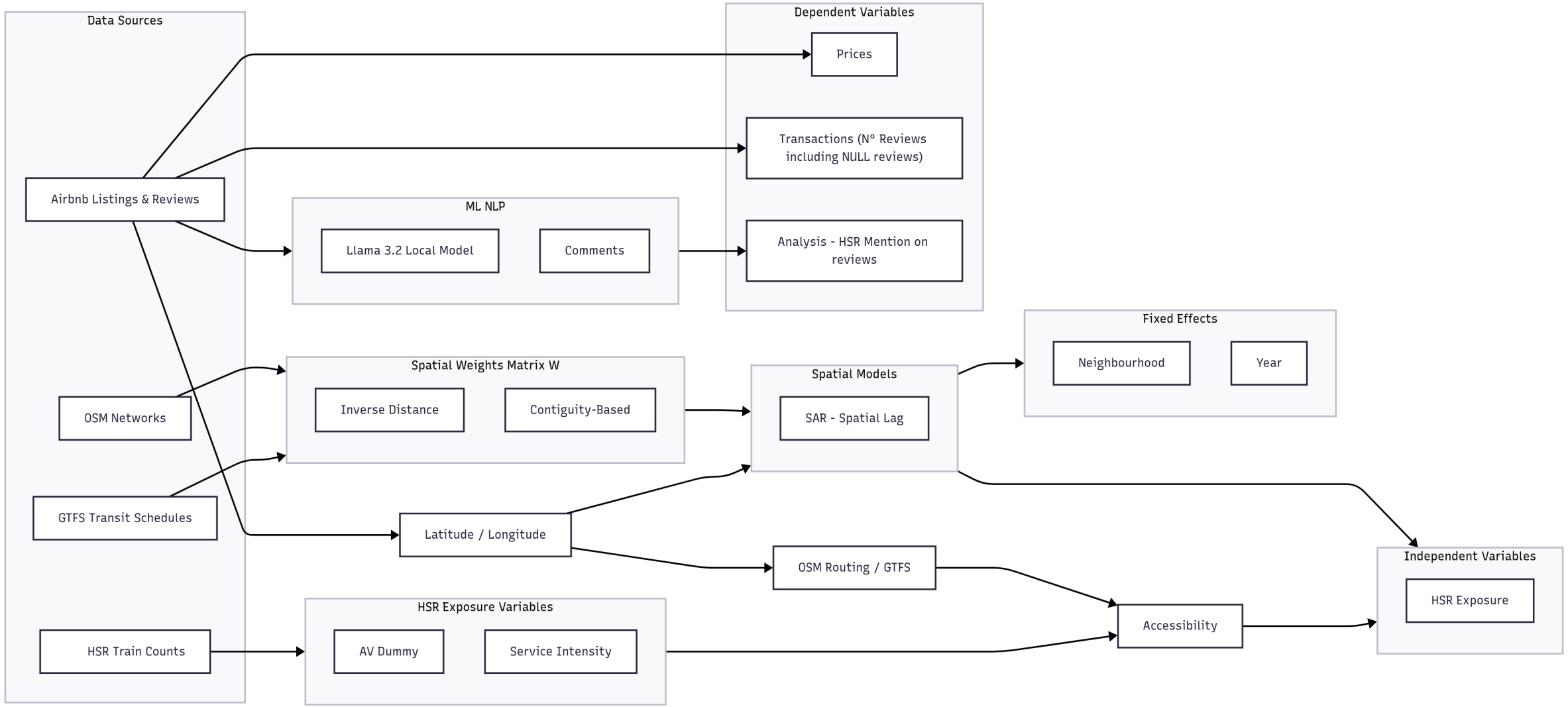
Analysed cities and regions

Data Sources: Mobility and Land Use

- OSM:
  - Street networks & station locations
  - Points of interest & land use
  - Basis for spatial accessibility metrics
- GTFS:
  - Public transit timetables & routes
  - Multimodal accessibility computation
  - Integration with station-level catchments
- Train Timetable:
  - HSR train counts per station (2008–present)
  - Separate data for weekdays & holidays
  - Temporal trends in service frequency



Methodology: Schema



Methodology: Formula

- The **econometric model is implemented within a machine learning logic**: using *fixed effects* as high-dimensional embeddings to absorb latent heterogeneity, and a *Continued Difference-in-Differences* setup as an engineered interaction feature to identify causal effects.
- Fixed Effects: Control for unobserved, time-invariant heterogeneity (e.g. intrinsic listing quality, persistent city traits).
- Continuous DiD: Compares treated vs. control units before and after a shock (here: HSR expansion) but with continued variables

$$\text{HSRIntensity}_{i,t} = \text{Trains}_{s,t} \cdot \exp\left(-\lambda_{i,s,t}\right)$$

Where:

- $\text{Trains}_{s,t}$: number of HSR departures at station s in year t
- $\lambda_{i,s,t}$: accessibility impedance from listing i to station s, defined as:

$$\lambda_{i,s,t} = \frac{d_{i,s}}{\exp(-\gamma g_{i,t})}$$

- $d_{i,s}$: straight-line distance (km) from listing i to station s
- $g_{i,t}$: generalized PT cost (in minutes), where:
  - $g_{i,t} = \text{walk}_{i\to\text{stop}} + \text{wait}_{i,t} + \text{inveh}_{i,t}$
  - $\text{walk}_{i\to\text{stop}}$: walking time to nearest PT stop
  - $\text{wait}_{i,t}$: expected wait time (from GTFS headways)
  - $\text{inveh}_{i,t}$: in-vehicle time to HSR station
- $\exp(-\gamma g_{i,t})$: accessibility weight → higher PT quality reduces effective distance penalty

Fixed-effects Panel Specification

We estimate the effect of HSR accessibility at the **listing** × **year** level with the following specification:

$$Y_{i,t} = \beta \cdot \text{HSRIntensity}_{i,t} + \mu_{c,t} + \varepsilon_{i,t}$$

Where:

- $Y_{i,t}$: outcome (e.g., log price, log location review score, share of station mentions)
- $\mu_{c,t}$: city × year fixed effects (absorbs macro and local shocks)
- Standard errors clustered at the station level

Dynamic Lagged Effect Specification

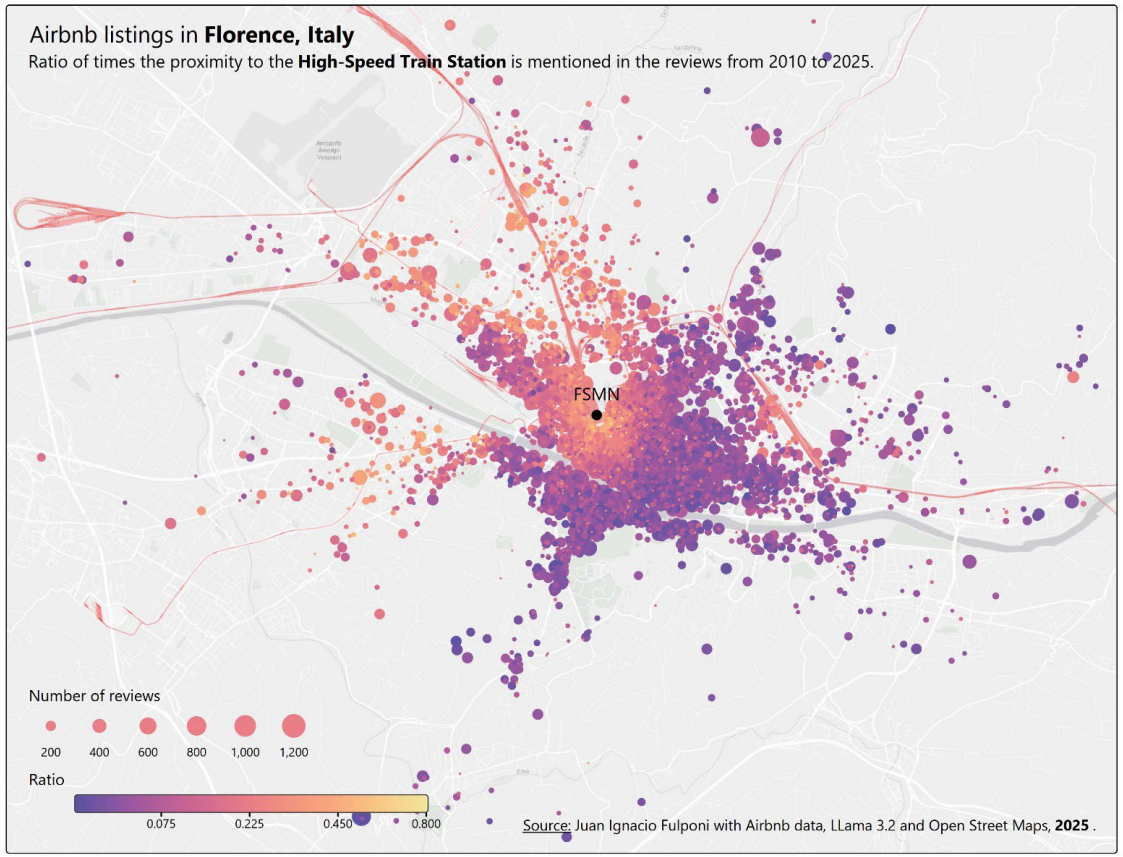
To assess delayed responses to infrastructure improvements:

$$Y_{i,t} = \beta_0 \cdot \text{HSRIntensity}_{i,t} + \beta_1 \cdot \text{HSRIntensity}_{i,t-1} + \beta_2 \cdot \text{HSRIntensity}_{i,t-2} + \mu_{c,t} + \varepsilon_{i,t}$$

- Tests whether demand adjusts **immediately** or with a **lag**
- $\beta_1 > 0$: suggests **delayed uptake** of improved HSR accessibility

Definition of independent variable ‘Mention on reviews’

- Local Llama 3.2 model with customised prompts to check if the reviews mention the HSR Station and its proximity.
- Model indifferent to the language of the review, not need to regex and look for misspellings.
- Effect visible not only on trains but also on others transit networks like trams with heavy centrality on HSR.



Example of Airbnb listings by HSR presence on reviews (Florence)

Results

- Dependent variable: **Log Price**
  - Coefficient of HSR intensity ≈ negative, small magnitude.
  - More HSR trains do not systematically increase Airbnb prices.
  - Suggests increased supply/competition around stations offsets demand growth.
  - *Price effects are limited or even slightly negative* → accessibility benefits are not directly capitalized in listing prices.

- Dependent variable: **Log Reviews** (number of transactions)
  - *Strong, positive and significant coefficient*.
  - Example: +100 additional trains ≈ +20% more reviews on average.
  - Robust across different λ (distance decay specifications).
  - HSR expansion drives higher demand for listings near stations.
  - Accessibility gains translate primarily into higher occupancy/usage, not higher prices.

- Dependent variable: **Share of Mentions on Reviews**
  - *Positive, weaker but significant effect*.
  - Guests explicitly mention station proximity more often as train frequencies increase.
  - Effect on STR demand stronger in large cities (Milan, Rome, Florence, Naples, Venice).

$$Y_{i,t} = \beta \cdot \text{HSRIntensity}_{i,t} + \mu_{c,t} + \varepsilon_{i,t}$$

	Price (log)	Reviews (log)	Mentions (%)
hsr_intensity_scaled	-0.490*	2.830*	0.393*
	(0.168)	(1.020)	(0.163)
Num.Obs.	472954	517159	517159
R2	0.185	0.251	0.214
RMSE	0.58	0.93	0.15
Std.Errors	by: station_id	by: station_id	by: station_id
FE: city_id	X	X	X
FE: year	X	X	X

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Continuous Differences in Differences with Fixed Effects

Spatiotemporal approach

$$Y_{i,t} = \beta_0 \cdot \text{HSRIntensity}_{i,t} + \beta_1 \cdot \text{HSRIntensity}_{i,t-1} + \beta_2 \cdot \text{HSRIntensity}_{i,t-2} + \mu_{c,t} + \varepsilon_{i,t}$$

- Current intensity (t): negative short-run effect (–0.027), marginally significant.
- Lag 1 (t–1): positive and significant (+0.057) → strongest STR demand response appears one year after train frequency increases.
- Lag 2 (t–2): negative again (–0.028), smaller magnitude.
- Guests respond with a delayed increase in STR demand after HSR service growth, suggesting adjustment time in travel behaviour and awareness.
- Effects fade or reverse after two years, indicating short- to medium-term impacts dominate.
- Fixed-effects: City × Year → controls for local shocks.

Dependent Variable:	ln_reviews
Model:	(1)
Variables	
hsr_intensity	-0.0269*
	(0.0148)
L1_int	0.0566**
	(0.0229)
L2_int	-0.0279*
	(0.0163)
Fixed-effects	
city_id	Yes
year	Yes
Fit statistics	
Observations	2,773
R^2	0.43353
Within R^2	0.00289

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Temporally lagged model

Conclusions and future developments

- ✓ HSR intensity boosts Airbnb STR demand near stations.
- ✓ No evidence of systematic price premium: accessibility benefits absorbed in quantity, not price.
- ✓ Station proximity is increasingly valued by guests (text mentions).

Future developments:

- ➢ Explore distributional effects: which neighbourhoods benefit most?
- ➢ Incorporate short-term rental regulations and housing market outcomes.
- ➢ Expand to cross-country analysis of HSR & platform economies. Compare market dynamics between STR and formal rental market (Long Term and Transitory rentals with contract according to the law).

Gruppo FS

The Mobility Leader

Thanks for your attention

#FSResearchCentre

https://www.fsitaliane.it/content/fsitaliane/en/fs-research-centre.html

https://www.linkedin.com/in/mario-tartaglia-81979527/

https://www.linkedin.com/in/jfulponi/