URW – Data Challenge

Group 3: *A. Arya, S. Bairagi, S. Capizzi, L. Passet, I. Sijelmassi, J. Xu*

January 2026

Our Reformulation

*“Using **real mall-level data**, propose a data-driven framework to **assess**, **simulate**, and **recommend optimal retail mixes** that **maximize the overall value of a shopping center** in terms of revenue, visitor engagement, efficiency, and future-readiness”*

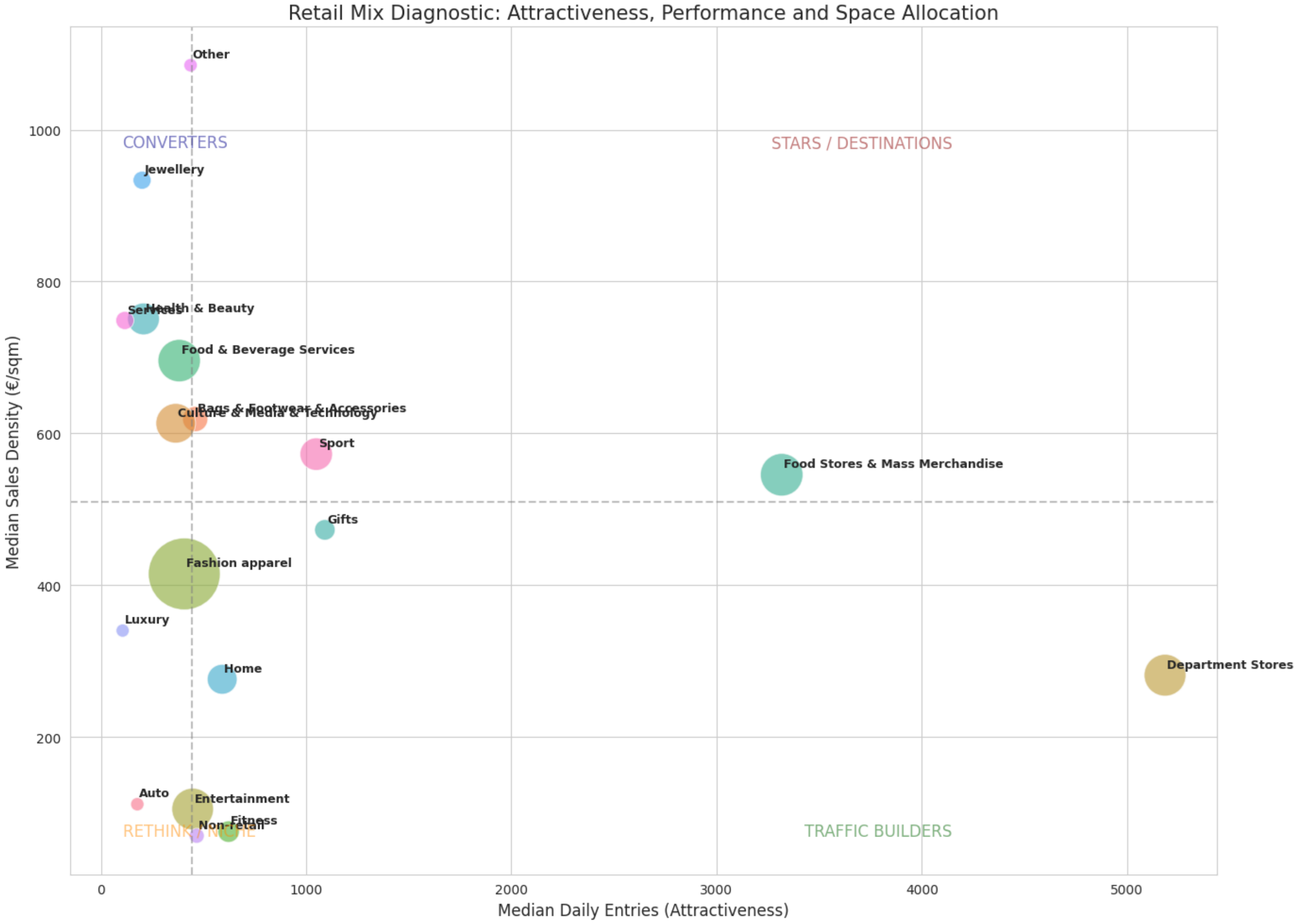
- Is there an **optimal retailer mix**?
- Is the assigned GLA sufficient to **maximize sales/m²**?
- Can **adjacent and nearby stores** act as a **driver of a store’s sales**? Is **store location** relevant?

<ins>We are defining few **mall-level KPIs**, e.g.:</ins>

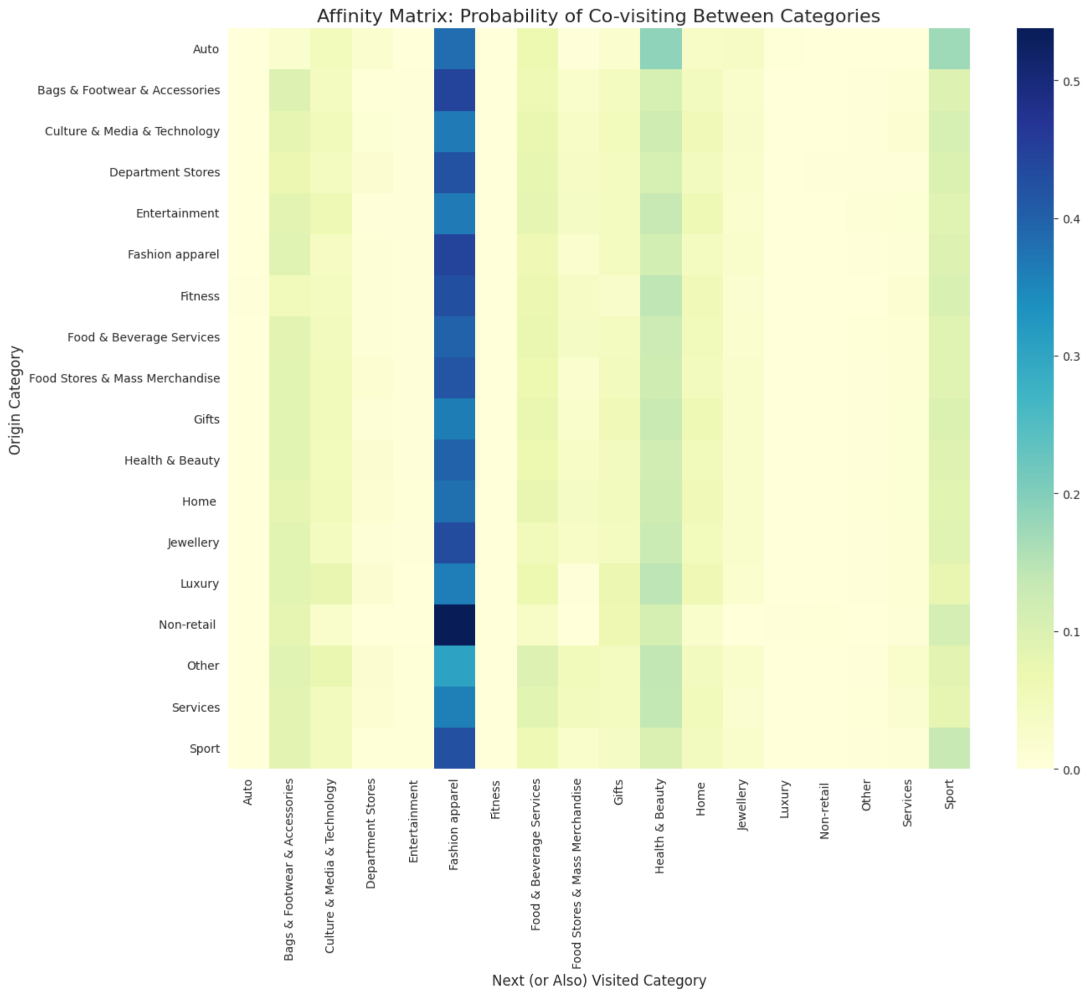
- **Revenue efficiency** (Σ sales / Σ GLA)
- **Network synergy** (cross-visits uplift)
- **Sustainability score** (weighted SRI)

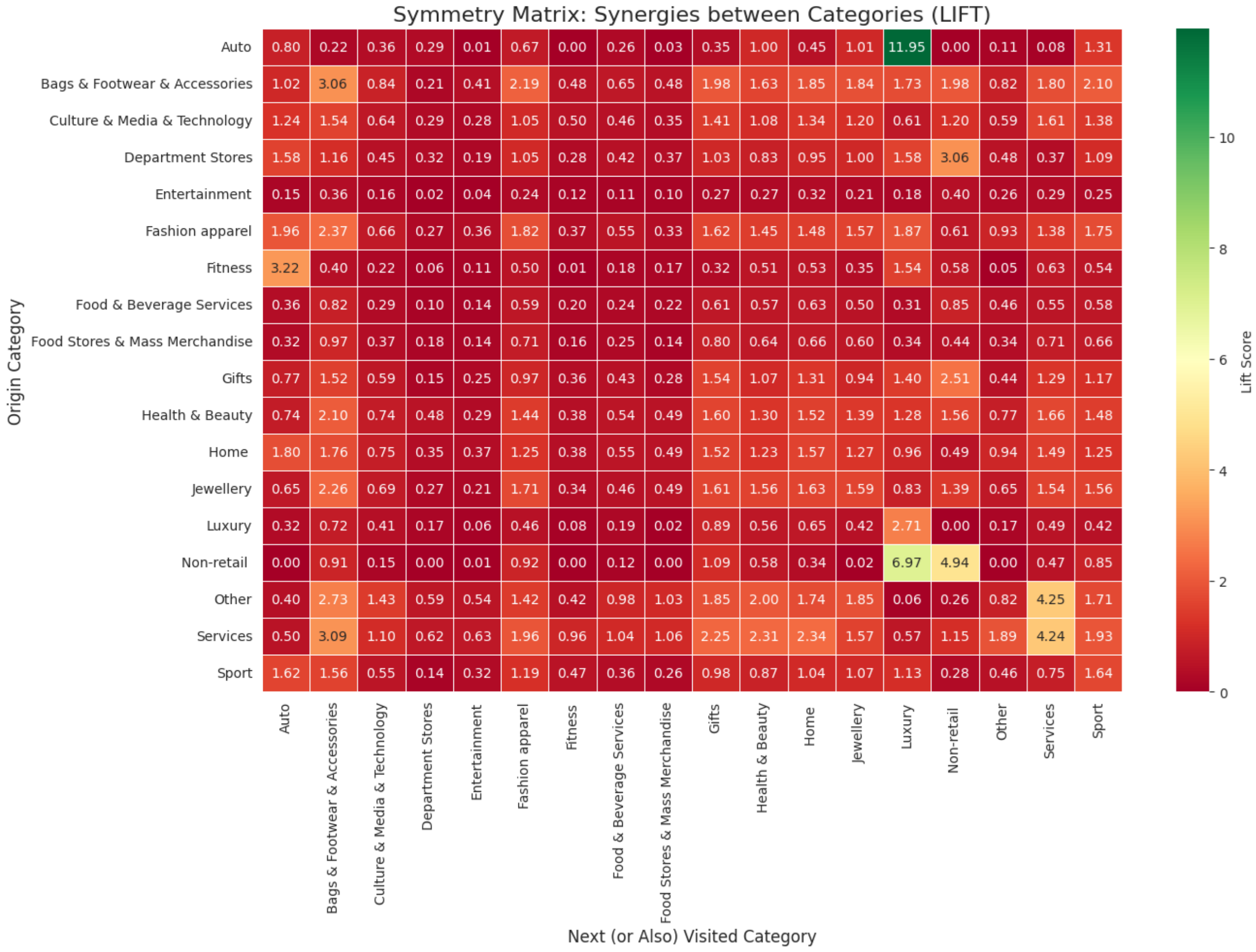
Westfield Euralille mall plan – level 0

EDA : Performance matrix (1/3)



EDA : Synergies analysis (2/3)





EDA : Mall profiling (3/3)

Mall profiling

- **Vélizy 2 :** *"The Destination Mall- Focused on Entertainment & Family Experience.“*
- **Euralille :** *“The Urban Hub- Focused on Convenience & High Capture.“*
- **Parly 2 :** *“The lifestyle Mall- Focused on High Spending & Selective Curation."*

Classification of the Malls:

	mall_id	avg_entries	avg_density	profile
9	19	121301.427637	824.480898	Destination
1	10	143987.784361	43.424466	Destination
16	30	177186.819510	256.036955	Destination
3	13	152494.214153	947.383438	Destination
15	28	178008.890892	114.653846	Destination
14	26	190416.344904	588.678297	Destination
12	24	96030.268208	423.579419	Destination
11	23	74386.524171	77.566136	Destination
8	18	155526.289902	772.110287	Destination
18	33	127454.459474	289.243674	Destination
10	22	137112.905357	936.845697	Destination
0	7	114247.325616	1266.467023	Lifestyle
5	15	119650.682945	2734.159774	Lifestyle
7	17	153045.679627	1141.014345	Lifestyle
6	16	200308.703880	1280.802697	Urban Hub
13	25	191271.092234	647.643144	Urban Hub
4	14	297531.723461	1341.529378	Urban Hub
2	12	205479.168967	1723.573234	Urban Hub
17	32	298868.252687	2202.313853	Urban Hub
19	36	275624.711402	1045.154360	Urban Hub

UNIBAIL-RODAMCO-WESTFIELD

Feature Engineering

We created rolling statistics to decrease the **seasonality noise** within the data and to observe the **short term trends**

Temporal Rolling Features

- Rolling statistics (7 / 14 / 28 days): level, volatility, and stability
- Short-term trends and growth rates (Δ, %Δ)
- Strict use of lagged information to prevent temporal leakage

Cross-visit Structure

- Modeled store-to-store interactions using cross-visit data
- Features capture:
    - Network centrality and traffic intensity
    - Quality of neighboring stores (sales and sustainability)
    - Concentration along shopping paths (top-K neighbors)

Relative Performance

- Store performance normalized against:
    - Same-day mall environment
    - Same-category peer stores
    - Store's own recent history

- Enables separation of absolute demand from contextual opportunity

The Model : Comparison & Selection

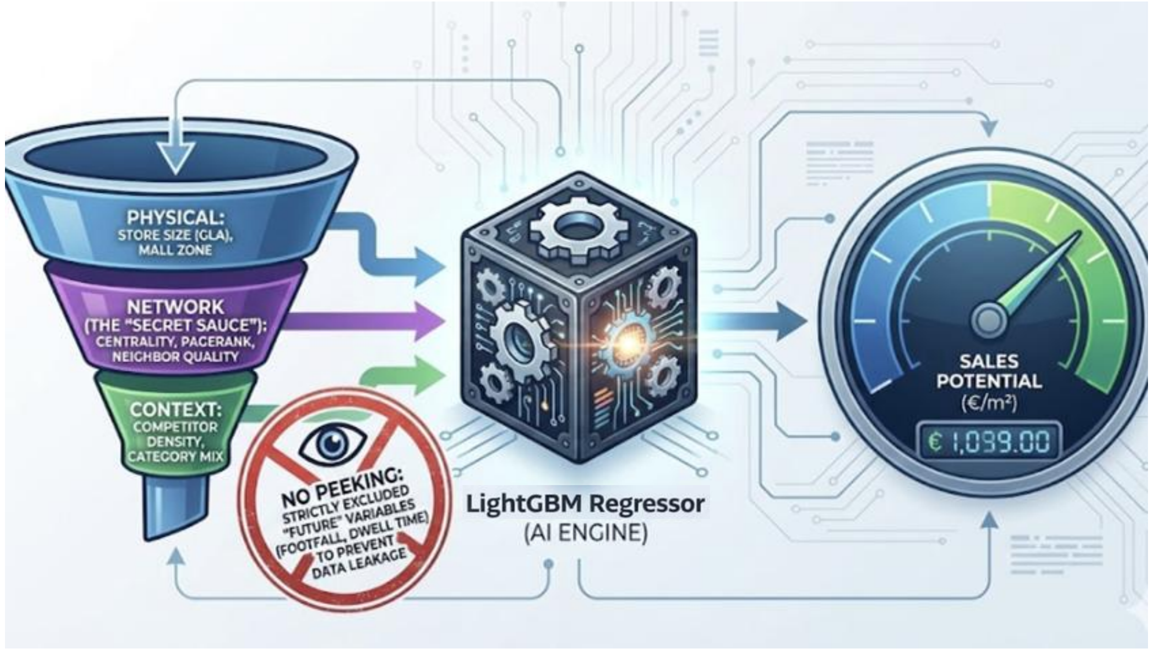
- Boosting models consistently outperform linear and tree baselines
- LightGBM achieves strong accuracy with stable generalization
- Selected model balances **performance, robustness, and interpretability**

Model	RMSE (log sales)	R2
LightGBM	0.530858	0.827951
CatBoost	0.560312	0.808330
Random Forest	0.600866	0.779581
XGBoost	0.817927	0.591565
Elastic Net	0.959935	0.437429

The Model : Intrinsic Location Value

Quantifying how valuable a location is because of ***where it is and who surrounds it***

- We train a **gradient boosting model (LightGBM)** to predict **sales density (€/m²)** at store level
- The model uses **these families of features**:
  - **Structural**: GLA, category mix, positioning within the mall
  - **Network**: cross-visit centrality, neighbor quality, adjacency effects
- Contextual: mall profile and category-level dynamics
- The **GroupKFold by mall** ensures leakage-free evaluation and generalization across assets
- Output is a **"fair value" estimate of the location**, independent of brand or tenant execution



The Model : Recommendations

Simulating **who should occupy a location**, not by changing the location itself.

```
--- Optimizing Store: 1314339 (Mall: 25) ---
Current: Cards & Gadgets (Gifts) | Sales: 3.62
RECOMMENDATION: Replace with 'Imaterial services' (Services)
Predicted Uplift: +€869.43 / m²
```

	New_SubCat	New_Category	Optimized_Potential_Sales	Uplift
3	Imaterial services	Services	873.047485	869.425330
2	Computer Products & Electronics games	Culture & Media & Technology	792.371033	788.748877
4	Electronics and household appliances	Culture & Media & Technology	791.391785	787.769629

For each store, we compare: **Current performance** (observed sales density) and**Location potential** (model-predicted fair value)

Stores with **large negative residuals** are flagged as **underperforming assets**
For each flagged store, we **simulate alternative tenant categories**:

- Replace category features while keeping location & network fixed
- Re-run the model to estimate **counterfactual sales density**

The recommended tenant is the one that **maximizes predicted uplift** under realistic constraints

The Model : Streamlit App

Bridging **advanced modelling** and **real-world leasing decisions**

The app translates the model into a **decision-support interface** for asset managers

Top KPIs at mall level:

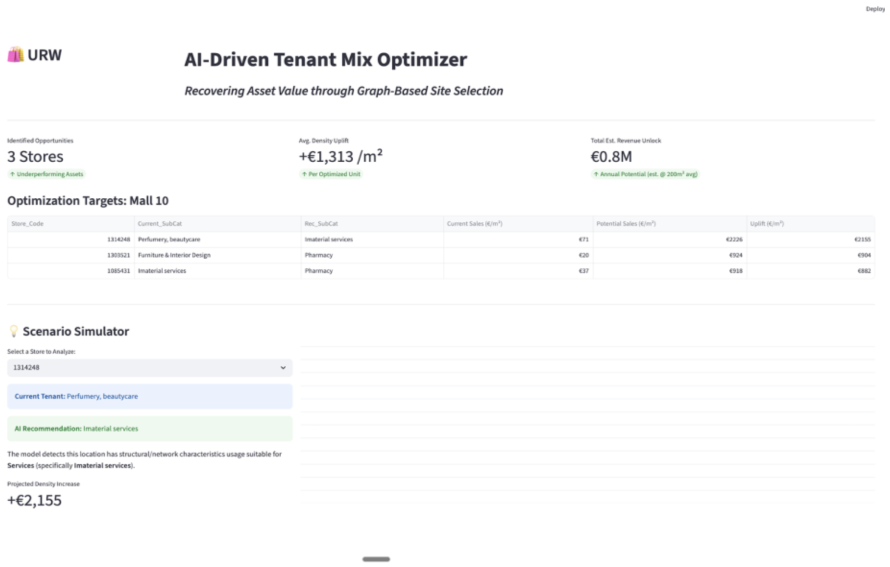
- Number of underperforming stores identified
- Average sales density uplift per store

Total estimated revenue unlock

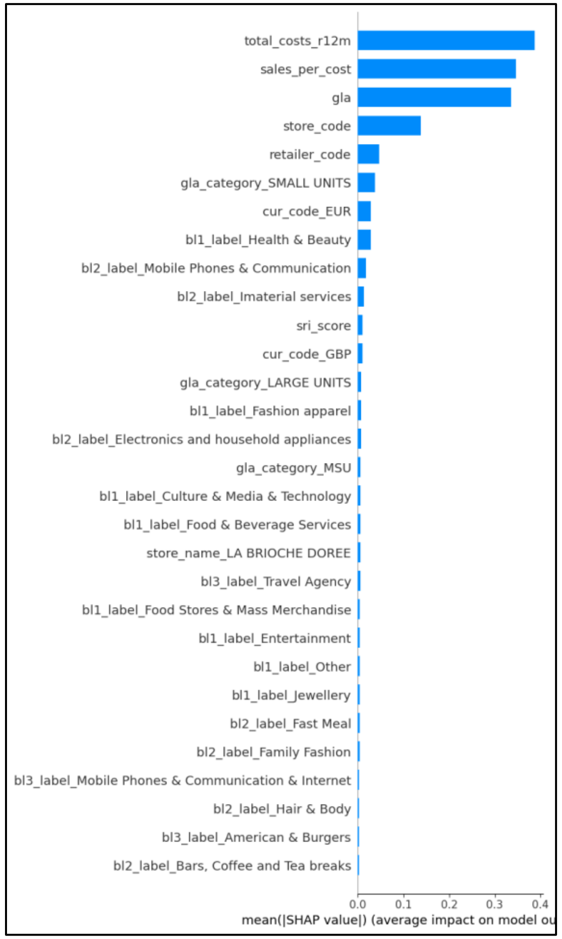
Store-level scenario simulator:

- Compare *Current Performance* vs *Location Potential* vs *AI-Optimized Tenant*
- Visualize expected uplift in €/m²

Enables **transparent, explainable, and scalable decision-making**



Feature Importance



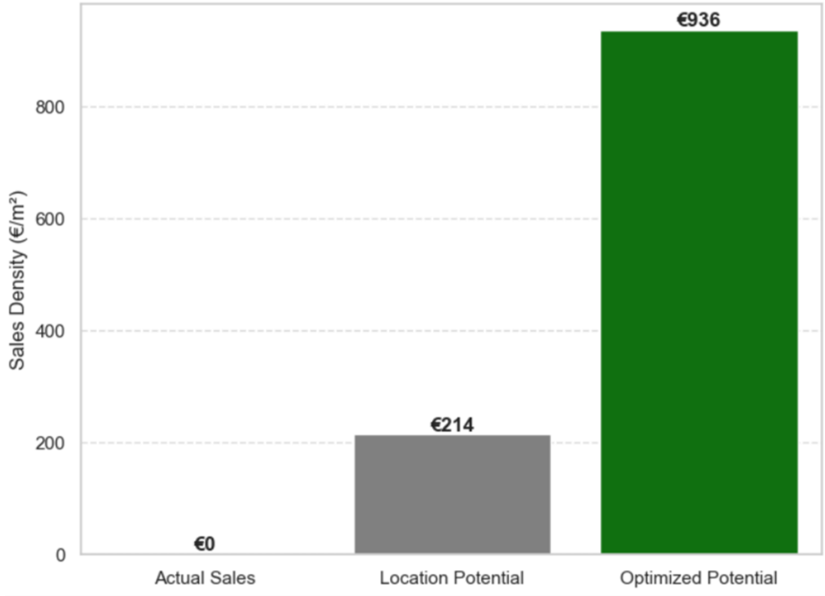
- Network and spatial factors are **top drivers of sales performance**
- Effects are **directional and heterogeneous** across stores
- Enables **store-level diagnosis and targeted tenant mix decisions**

Business Consequences

The "**Tenant Recommender System**"

- Built a script that simulates replacing tenants with **50+ alternative categories**
- Identifies **"Value Gaps"**: Locations where *Potential > Actual Performance*
- **Optimizes for Synergy**: Recommends categories that benefit from the specific neighbours of that unit

Optimized potential



<u>**Business Applications**</u>

URW "Leasing Relations" could use the tool both when **concluding a lease agreement**, to assess the **effectiveness of changing a tenant**, and in **broader portfolio-wide re-planning scenarios.**

Forward looking – Positioning

*We used **Euclidean distances as a proxy** for the **actual distances** within the mall.*

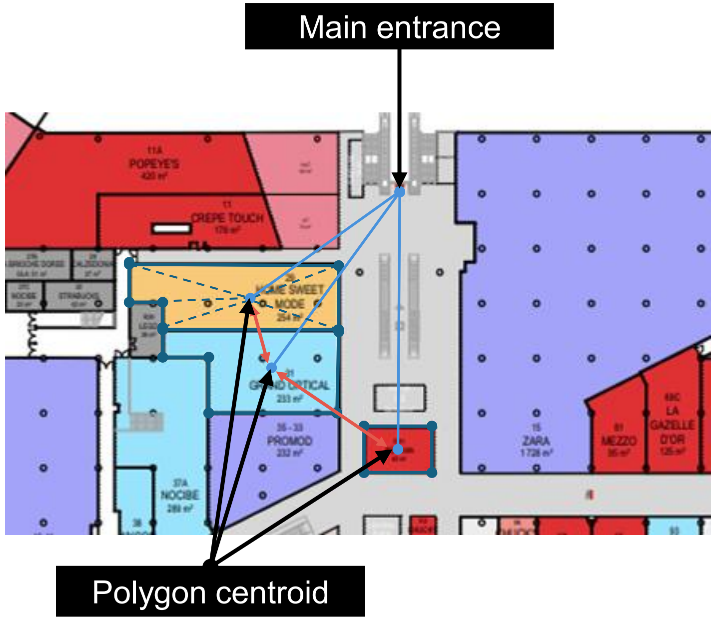
- **How to take positioning in consideration?**

To answer, we had to represent the shops using **polygons**.

We projected the maps provided to us onto a **coordinate plane**, thereby defining the **shop vertices as points**.

i.e. **POLYGON**((0,0);(0,1);(1,1);(1,0);(0,0))

At this point, we identified **the centroid of the resulting polygons**.
So, we discovered **Euclidean distances between the shops' centroid and the main entrances***
Also**, we calculated the distance between each possible pair of stores**



Executive Summary

*Focusing on the **intrinsic quality of the location itself,** we created a **tenant recommendation system**.*

-

1. We **contextualise the mall type** and **spot the location below potential**

2. Simulate **50+ alternative tenant categories**

3. Share the analysis with the "**Leasing Relations**" team