Data-Driven CRM Analytics:
Summer 2025 Campaign

Janhvi Goje

Our Goal: Maximize ROI Through Data-Driven Decisions

By analyzing the performance of the 2024 campaign and identifying key store characteristics that contribute to success, we aim to make informed decisions about which stores to target for maximum impact.

Therefore, we aim to leverage the power of data in transforming **raw data into actionable strategies**, answering the pivotal questions: **how can data drive better decisions?**

Data, Data and More Data!

We utilized two key datasets— *"CompanyX_Summer_Plan_24_activated_Stores"* and *"Store feat."*

The data we have has multiple columns, but all of these might not be of use of us.

Summer Precipitations IDX	Summer Temperature IDX	Nr of Campings	Nr of B&B	Latitude	Longitude	Total Population	Male Population	Female Population	Celibe Population	Married Population	Divorced Population	Popula
47	280	22	6	43.4548506	13.5962393	162377	79475	82902	65481	78139	3385	
36	274	22	3	42.1582256	14.6157881	94559	46473	48086	37165	47789	1461	
35	277	18	4	42.0626849	14.775524	102311	50229	52082	41002	51491	1617	
37	277	15	7	42.2659656	14.4341445	139226	67808	71418	54675	68401	2469	
32	280	15	0	41.94062	15.06898	54984	27036	27948	22389	26951	874	
46	242	6	0	44.506954	8.890496	89051	42758	46293	34187	40324	2843	
46	242	6	0	44.50657	8.8925	94316	45222	49094	36313	42595	3007	
37	277	7	3	42.232987	14.407132	88890	43399	45491	35003	43576	1622	
17	298	14	0	40.394896	16.800577	47495	23585	23910	19999	23202	526	
14	283	7	0	39.6282944	16.5133564	62747	30825	31922	28628	28972	875	
39	246	14	3	44.4040434	8.6773076	237741	111866	125875	87827	108662	8491	
18	259	6	0	39.3276554	16.2168357	165892	79895	85997	70191	78118	2986	
48	275	9	3	43.5576133	13.5103654	237118	114068	123050	94344	111881	5944	
63	290	13	1	45.2784198	11.9976768	155616	76736	78880	64241	75071	2900	
3	294	8	4	39.279295	9.021928	241844	115811	126033	108251	106269	5471	
5	298	6	0	39.16534	8.99546	42045	20633	21412	19000	18444	957	
48	275	9	3	43.5725585	13.5096235	219909	105790	114119	87222	103761	5583	
46	242	0	0	44.5690681	8.9488123	106058	51226	54832	41058	47937	3356	
46	242	0	0	44.570986	8.952866	125489	60352	65137	48378	56855	4016	
39	246	8	3	44.408278	8.683755	149890	70392	79498	54894	68962	5286	
17	298	14	3	40.372436	16.811808	39356	19623	19733	16607	19244	435	
38	238	14	3	44.3997824	8.6649888	342006	161215	180791	128251	154524	12384	
39	246	10	3	44.405538	8.678307	184274	86652	97622	67800	84670	6514	
48	275	23	4	43.52175	13.55876	218345	105570	112775	87144	103401	5303	
3	294	6	3	39.292166	9.0034569	215216	103663	111553	96332	95214	4724	
4	297	2	0	39.3104912	8.9724838	115073	56794	58279	51618	52471	2233	
4	297	0	1	39.292572	8.993641	154625	75373	79252	69505	69186	3254	

▶ Key Features: What Drives Campaign Success?

We utilized two key datasets— *"CompanyX_Summer_Plan_24_activated_Stores"* and *"Store feat."*

- **Summer Precipitation Index:** Weather impacts demand; this index helps pinpoint regions with higher seasonal potential.
- **Number of Campings:** The presence of nearby campgrounds signals increased foot traffic and also, a possible indication for tourism.
- **Distance from the Sea:** Proximity to coastal areas often correlates with unique consumer behavior during summer months.

Each of these features was chosen for their predictive strength and alignment with seasonal consumer behavior, **letting the data lead, we uncover deeper insights to drive impactful decisions.**

Method used: Forward Feature Selection (Random Forests)

▶ Choosing the Best Predictive Model

MODELS	ACCURACY
Logistic Regression	88.8%
Decision Tree	89.4%
Random Forests	92.4%
XGBoost	90%

Therefore, we choose the Random Forests to predict the **"Outcome of Activation"** (incremental sales or no incremental sales)

▶ Calculating Expected Profit: Translating our Predictions into Numbers

$$E(Profit) = P(Incremental\ Sales) \cdot 1000 + P(No\ Incremental\ Sales) \cdot (-1500)$$

- The first term accounts for the probability of success multiplied (or propensity) by the profit per successful activation **(€1,000)**, while the second term factors in the probability of failure multiplied by the associated cost **(-€1,500)**.
- Assumption - "each successful store generates €2.5k in incremental sales, while running a campaign in a store costs €1.5k."

# ▶ Why not just say: +€1,000 for success and -€1,500 for failure?

E(Profit - Incremental Sales)=1000
E(Profit - No Incremental Sales)=−1500

- This approach assumes certainty in outcomes, which is unrealistic in a dynamic, data-driven environment.
- **Success is never guaranteed; even when the accuracy of the model is very high.**

Maximizing ROI: Insights from Propensity Thresholds

We seem to have reached maximum profit of around 163,700 Euro when we activate **55% of the the stores having a propensity of higher than 0.6.**
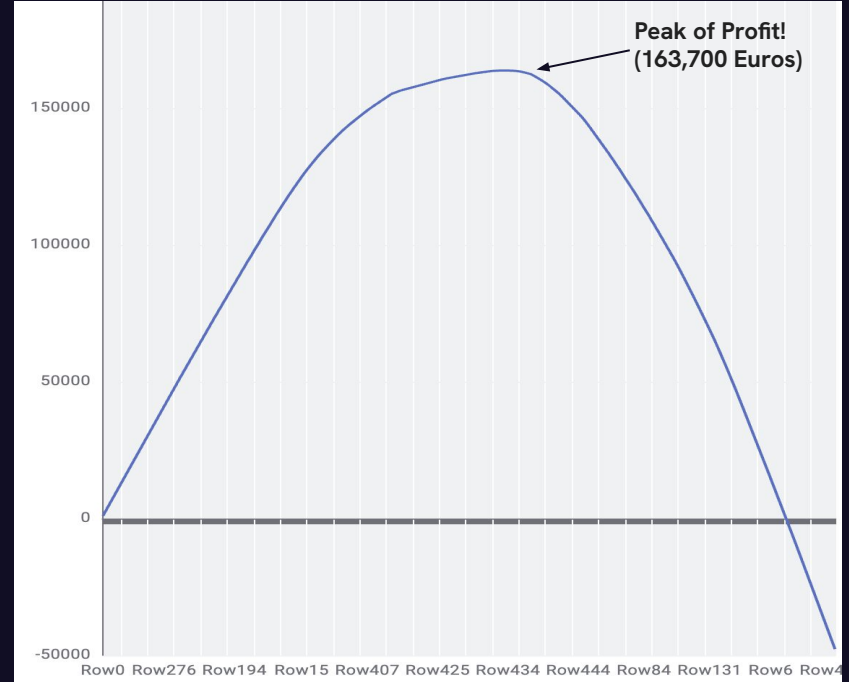
▶ Thinking beyond the Dataset...

🇮🇹 **Italy Cities Database**   Other Countries ▾

Below is a list of 307 prominent cities in Italy. Each row includes a city's latitude, longitude, region and other variables of interest. This is a subset of all 17,649 places in Italy (and only *some* of the fields) that you'll find in our World Cities Database. We're releasing this data subset for free under an MIT license. You're free to use the data below for personal or commercial applications. The data below can be downloaded in `.csv`, `.json`, and `.xlsx` formats.

Notable Cities: The capital of Italy is *Rome*, which is in the region of Lazio. The largest city in Italy is *Rome*, with an urban population of 2,748,109.

[Download Data in Spreadsheet (307 cities) ▾]   [Get all cities in Italy (17,649 cities)]   Looking for an Italy map?

city	lat	lng	country	iso2	admin_name	capital	population	population_proper
Rome	41.8933	12.4828	Italy	IT	Lazio	primary	2748109	2748109
Milan	45.4669	9.1900	Italy	IT	Lombardy	admin	1354196	1354196
Naples	40.8333	14.2500	Italy	IT	Campania	admin	913462	913462
Turin	45.0792	7.6761	Italy	IT	Piedmont	admin	841600	841600
Palermo	38.1111	13.3517	Italy	IT	Sicilia	admin	630167	630167
Genoa	44.4111	8.9328	Italy	IT	Liguria	admin	558745	558745
Bologna	44.4939	11.3428	Italy	IT	Emilia-Romagna	admin	387971	387971
Florence	43.7714	11.2542	Italy	IT	Tuscany	admin	360930	360930
Bari	41.1253	16.8667	Italy	IT	Puglia	admin	316015	316015
Catania	37.5000	15.0903	Italy	IT	Sicilia	minor	298762	298762
Verona	45.4386	10.9928	Italy	IT	Veneto	minor	255588	255588
Venice	45.4375	12.3358	Italy	IT	Veneto	admin	250369	250369
Messina	38.1936	15.5542	Italy	IT	Sicilia	minor	218786	218786
Padova	45.4167	11.8667	Italy	IT	Veneto	minor	206496	206496
Cosenza	39.3000	16.2500	Italy	IT	Calabria	minor	200257	63760
Trieste	45.6503	13.7703	Italy	IT	Friuli Venezia Giuli	admin	198417	198417
Parma	44.8015	10.3280	Italy	IT	Emilia-Romagna	minor	196764	196764
Brescia	45.5417	10.2167	Italy	IT	Lombardy	minor	196446	196446
Prato	43.8800	11.0983	Italy	IT	Tuscany	minor	195736	195736
Taranto	40.4711	17.2431	Italy	IT	Puglia	minor	188098	188098

Traditionally, Istat develops classifications of Italian municipalities that are based on geo-morphological or urban settlement characteristics, measured for statistical purposes only.

Municipalities are therefore assigned a series of attributes, corresponding to the following physical and/or anthropological characteristics:

- Coastal nature
- Altitude zone
- Altitude of the main town centre (msl)
- Land area (kmq)
- Degree of urbanization ←
- Coastal areas

to which is added information relating to the size in terms of surface area and population (legal and resident).

The description of the possible values that each characteristic can assume are explained in the legend (classification metadata).

Warning : *To access the new version of the reports on the topics of "Dimension" and "Characteristics of the territory", please refer to the* Situas *Portal (section "Home/Links/Most consulted", or the "Territorial units" section to search for the date of interest)*

⬇ Statistical classifications – year 2024 (zip) – published on 8 July 2024

▶ Integrating Deg. of Urbanization

- Now, **XGBoost** is our best model with accuracy **94%**!
- Calculating 'Expected Profit' similarly, the ROI changes as follows…
- If we limit our campaign to customers with a propensity score higher than **0.6** (which will be around **60%** of the total population), we will maximize our profit of around **246,000 Euro**.

New Peak of Profit! (246,000 Euros)
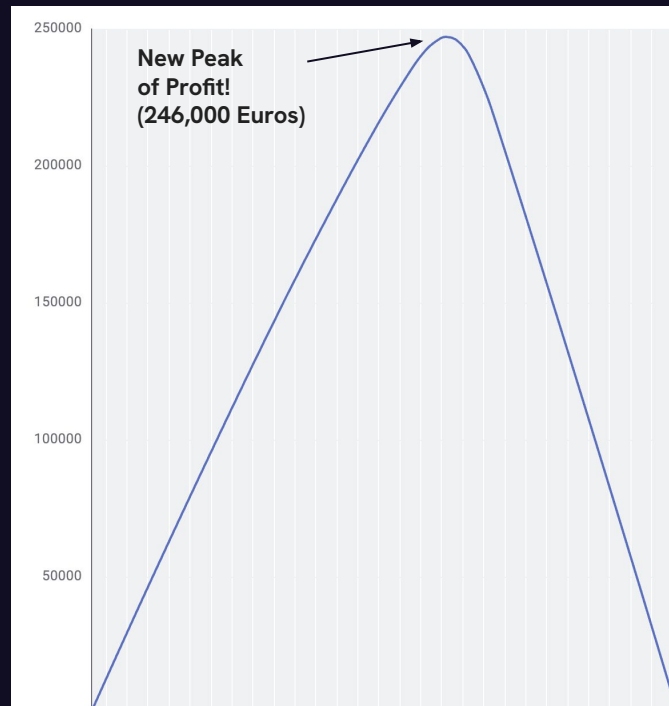
'Size of the Prize' Comparison: From €15K to €246K

2024	Prediction for 2025	Prediction for 2025 (Using Deg. of Urbanization)
	Accuracy - **92.4%**	Accuracy - **94%**
15,000 Euros (for 170 stores)	**163,700 Euro** (for 267 stores)	**246,000 Euro** (for 294 stores)

- Incremental Sales = **€1,000**
- No Incremental Sales = **-€1,500**

▶ 'Size of the Prize' Comparison

Venus Campaign Predictions (Profit per Store)

88.24 Euro/store — 2024

613.11 Euro/store — Prediction for 2025

836.73 Euro/store — Prediction for 2025 (Using Deg. of Urbanization)
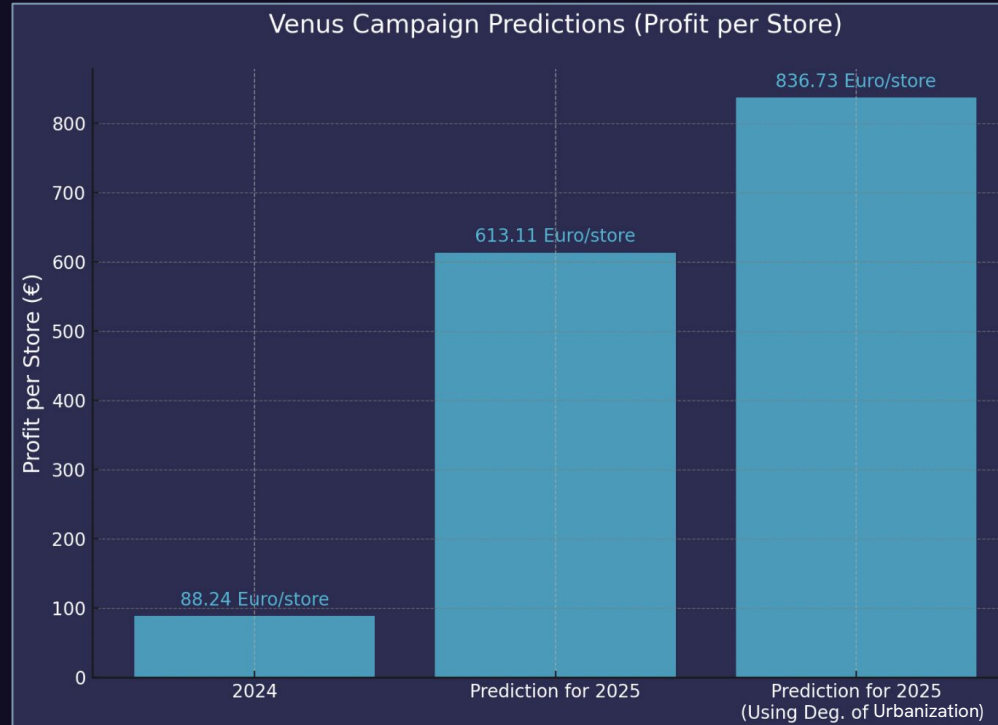
We export the predictions of XGBoost as our 'Potential Stores' to look at their key characteristics in PowerBI.

Typical Traits of Potential Stores: Geographical and Demographic
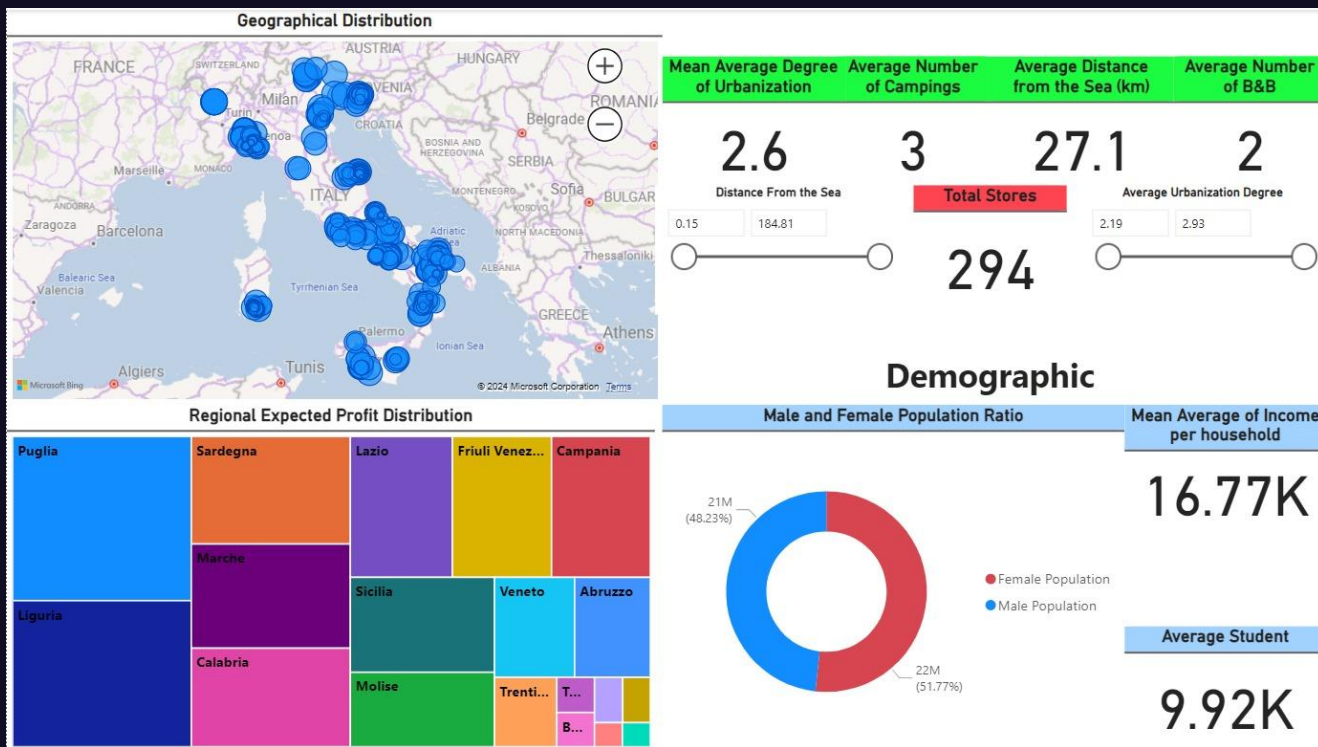
▶ AI-Driven Personalization: Geo-Localized Communication Strategies for Potential Stores

Based on the selected stores, expected to produce the best results, we used an LLM *(Llama 3.1)* to provide insights and advice for organizing the campaign, according to the specific characteristics of each store.

- **Role:** Act as the head of marketing in Company X, now focused on working on this summer's marketing campaign for Company X's Product in Italy. The campaign will focus on selected stores, which have been proven to give successful results in the past.

- **Task:** You are responsible of generating personalized communication contents on selected stores for this summer's campaign in Italy. You will be given specific information about each store, and the location it's in, and based on that information, you will generate the content for the campaign. Your answer should provide valuable insights and ideas for marketing experts to implement in the campaign.

- **Expected Features:** Your response should include a title, a message, the target audience, key messages, and any other information that can be useful like visuals, social media, in-store promotions, etc.

- **Data:** Store ID, Location, Population, Male/Female Ratio, Celibe/Married/Divorced Ratio, % of Students, % of Graduated Students, % of Unoccupied, and Population Age Distribution.
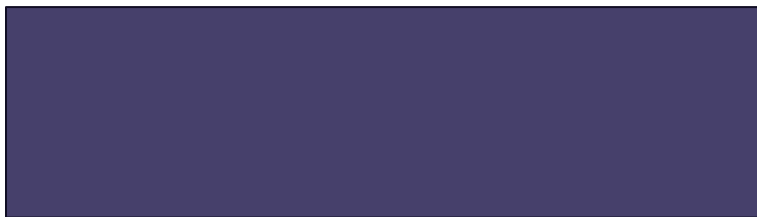
The Recommendations

Summer Campaign for Store ID 50231 in Porto Potenza Picena, Marche

Message: ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮

Target Audience: Women aged 20-49 living in Porto Potenza Picena and surrounding areas, with a focus on young professionals, mothers, and women who value convenience, comfort, and style.

Key Messages:

Visuals:

Social Media:
To reach our target audience, we will create engaging social media content for Facebook, Instagram, and Pinterest. The content will include:

In-Store Promotions:

Additional Ideas:

Key Takeaways

- **Summer Precipitation Index, Number of Campings and Distance from the Sea** were recognised as the most important features **(using Forward Feature Selection Node and Random Forest as the model)**
- Limiting activation to high-propensity stores, in particular propensity scores **>0.6 yields €246,000 profit** and hence maximizes ROI.
- Using GenAI, we **tailor feature-specific marketing and communication strategies for our potential stores**, making each campaign the most compatible with its demographic characteristics.

Thank You!