Increasing market transparency of residential air-to-water heat pumps in Europe by clustering the appliances

Ugne Potthoff[a], Felix Morlock[a], Felix Wortmann[b]

[a]D-MTEC, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
[b]ITEM-HSG, Dufourstrasse 40a, 9000 St. Gallen, Switzerland

Abstract

The transition to low-carbon heating requires not only efficient technologies but also transparent markets that enable informed policy and consumer decisions. Yet, users and installers face challenges in comparing air-to-water heat pumps (AWHPs) due to complex technical parameters and inconsistent model naming across countries and online platforms. This study aims at helping to navigate in choosing the heat pump by improving the understanding of how heat pumps are priced in Western Europe. Using the open-source *hplib* library, a dataset of 696 Keymark-certified AWHPs from 13 manufacturers was compiled, covering units up to 13 kW using low-GWP refrigerants R32 and R290. Retail prices were collected through automated web scraping of European online vendors and matched to certified technical parameters. Price modelling shows that the main price drivers are rated power, storage size and refrigerant type. Together with noise emission, standby power, reference COP (coefficient of performance), and tank configuration, these factors explain most of the observed price variation ($R^2 = 0.84$). The three dominant drivers substantially outweigh the influence of COP, which contributes only marginally to price formation. Unsupervised clustering identified five distinct product groups differentiated mainly by brand origin (EU vs. non-EU) and configuration (monoblock vs. split). European heat pumps show higher maximum supply temperatures and lower noise levels, indicating premium and suitability for retrofits, while non-EU heat pumps focusing on affordability. The efficiency remains similar between regional groups (SCOP = 4.70 EU vs. 4.76 non-EU). However, European models are priced higher (€1,362/kW vs. €1,023/kW). Overall, this study integrates model matching, price modeling, and clustering to contribute to a transparent, data-driven European AWHP market. The findings enhance product comparability and provide evidence to support policymakers and consumers in accelerating the heat-pump transition.

Keywords: Air-to-water heat pumps; Price drivers; Product naming; SCOP; Market transparency; European heating market; Web-scraping;

1. Introduction

Heat pumps (HPs) have become central to the decarbonization of Europe's building stock, providing an efficient alternative to fossil-fuel-based heating systems. Their market penetration has grown rapidly, with more than 24 million units now installed in Europe alone [1]. Air-to-water heat pumps (AWHPs) in particular play a crucial role in residential heating transitions due to their lower installation costs compared to ground- or water-source systems. Despite strong policy support and continuous technological improvements, adoption remains uneven across regions, and market transparency continues to be a significant barrier to wider deployment [2].

One of the central challenges in the European air-to-water heat pump (AWHP) market is the large variation in retail prices among technically comparable products. A survey by a German consumer advice center analyzing 160 installation offers revealed price differences of up to a factor of three for similar systems, creating significant uncertainty and distrust among potential adopters [3]. Such discrepancies hinder consumer confidence and slow down the market uptake of low-carbon heating technologies. Importantly, these price comparisons often conflate equipment costs with installation and site-specific expenses, which are rarely disclosed separately and vary strongly across regions and installers.

A key reason for these inconsistencies lies in the limited market transparency. Due to the technical complexity of heating systems and the dominance of business-to-business (B2B)-oriented sales channels, homeowners rarely receive standardized product information. In some cases, heat pumps are offered primarily as bundled solutions including installation, while the standalone hardware price is not explicitly visible to end users. Moreover, non-uniform product naming across countries, sales platforms, and even within manufacturer catalogues prevents direct price comparison. Two heat pumps with identical certified performance parameters may appear under different commercial names or configurations, obscuring comparability for consumers, installers, and analysts alike. This fragmentation fosters information asymmetry, where buyers cannot distinguish whether price differences reflect real performance or simply branding and configuration choices.

Beyond these data inconsistencies, there remains limited empirical understanding of what actually drives heat pump prices. Manufacturers often attribute higher prices to superior efficiency, commonly expressed by the Seasonal Coefficient of Performance (SCOP). However, empirical evidence supporting this relationship is scarce. Previous studies have primarily examined cost structures or technical performance in isolation. For example, Ban and Bungău [4] and Paiho et al. [5] identified strong cost differences between system types but did not link them to retail pricing. Bianco et al. [6] focused on primary energy and life-cycle efficiency, while Miara et al. [7] emphasized discrepancies between laboratory and field performance. Yet, the interaction between efficiency indicators, configuration type, and retail prices across manufacturers remains insufficiently quantified. Understanding this relationship is essential for designing transparent, data-driven policies and for enabling fair comparison in the rapidly expanding European heat pump market.

Previous research on heat pumps has focused mainly on efficiency metrics such as COP and SCOP or on life-cycle economics, while little work has linked certified technical data to real-world retail prices. To enable meaningful comparability across products and countries, this study focuses explicitly on the retail prices of heat pump hardware only, excluding installation and other customization-dependent costs, which are typically not transparently available on a large-scale. This study addresses that gap by combining HP Keymark data [8] with retail prices obtained through automated web scraping of German-led Western European online markets. Because identical models often appear under different commercial identifiers, a reference dictionary was developed to harmonize naming and create a unified dataset. Using this consolidated data, the study aims to quantify retail price variation across certified AWHPs, identify the technical and market-related factors that explain price differences, and detect structural product clusters based on combined technical and commercial attributes. This integrated approach provides new empirical insight into how performance characteristics, configuration types, and market practices shape AWHP pricing.

2. Methods

2.1. Data and scope

We constructed a dataset of residential air-to-water heat pumps (AWHPs) from the Heat Pump Library (hplib) based on Keymark certificates (snapshot: February 2025) [8,9]. Certificates provide harmonized technical specifications, including model identifiers, SCOP, rated heating capacity at low temperature, outdoor sound power level, and maximum water temperature. The study focuses on AWHPs with nominal capacity of ≤ 13 kW and using R32 or R290 refrigerants to represent the European residential segment [10,11]. Duplicate certificate entries and non-AWHP technologies were removed. The resulting technical base covers models from 13 manufacturers.

2.2. Retail price collection

Retail prices were collected from publicly accessible Western European e-commerce platforms using automated web scraping. Web scraping refers to the programmatic extraction of structured information from publicly available web pages using automated scripts, enabling systematic and reproducible data collection from online sources [12]. This approach is widely used in market analysis and energy economics to capture real-world price information that is otherwise fragmented across platforms.

The analysis focuses on online retailers operating primarily in Western Europe, with strongest coverage in Germany, and additional platforms serving the Netherlands, Italy, Switzerland, and the United Kingdom, where standalone heat-pump hardware prices are publicly disclosed. Scraped platforms and their associated markets are listed in Table 1. Data was collected exclusively from pages displaying equipment-only prices, excluding offers that include installation or additional site-specific services.

Table 1: Online retail platforms used for heat pump price collection and their primary markets.

Platform (domain)	Market	Platform (domain)	Market
breeze24.com	Germany	heima24.de	Germany
climamarket.eu	Italy / EU-wide	heizungsdiscount24.de	Germany
climamarket.it	Italy	idealo.de	Germany / EU-wide
climashop.nl	Netherlands	plumbnation.co.uk	United Kingdom
eibabo.ch	Switzerland	raleo.de	Germany
eibmarkt.com	Germany / EU-wide	selfio.de	Germany
theheatpumpwarehouse.co.uk	United Kingdom	vanwalraven.com	Netherlands / EU-wide

For each product listing, the scraper extracted the manufacturer name, model identifier, listed price, currency, product URL, and scraping date. Prices were converted to euros (€), exact duplicates were removed by retaining the most recent observation, and entries below €1,000 were excluded, as these typically correspond to accessories or non-comparable components rather than complete heat-pump units. This procedure ensures that the resulting dataset reflects comparable, market-visible hardware prices that consumers and analysts can independently verify.

2.3. Model configuration decoding

To enable like-for-like comparisons, we systematically annotated each Keymark-certified model with configuration attributes that affect price: monoblock vs. split, presence of a hydraulic module, and integrated domestic hot-water tank (recording tank capacity where available). Classification combined certificate metadata with manufacturer documentation and major reseller listings; when encoded hints were present (e.g., suffixes indicating “m/s” or tank size), they were resolved and verified against source pages. The resulting configuration labels were merged with the technical records, allowing products to be grouped not only by performance (e.g., power) but also by system architecture - an essential prerequisite for unbiased price comparisons and subsequent segmentation.

2.4. Cross-source model matching

Market listings frequently use variant model strings that differ from Keymark identifiers (regional codes, formatting, added descriptors). We therefore built a reference dictionary to reconcile vendor/model strings to their corresponding Keymark entries, as shown in Table 2. Matching applied token normalization (lower-casing, removal of special characters and regional suffixes) and rule-based reconciliation; ambiguous or low-confidence pairs were manually reviewed. This iterative procedure - refining rules whenever unmatched items were detected - produced a unified dataset that reliably links certified specifications with observed retail prices, forming the basis for further data analysis.

Table 2: Exemplary name matches between scraped vendor data and official hplib entries. For each manufacturer, a selected row shows how the scraped Model/Type column was matched with the hplib Titel column.

Manufacturer	Model (Keymark database)	Title (Keymark database)	Model (Scraped data)
Bosch	CS5800i/6800iAW 10/12 OR	CS5800iAW 10 ORM-T	BOSCH Luftwärmepumpen-Paket BOPA CS788 AW 10 OR-T, CS5800iAW 12 M, 7739623429
Daikin	ALTHERMA 4 H ECH2O 08–10 kW 500L (3ph)	EPSK08AW1 / EPSX(B)10P50A	Daikin Altherma 4 H ECH2O Set EPSK-08AW1 + EPSX-10P50A Hydrobox/Speicher 500 L WiFi R290 Heizstab 9 kW
Johnson Hitachi	Yutaki M 5.0 HP R32 (mono)	RASM-5VR1E – with cooling kit	Hitachi Yutaki M RASM-5VR1E Monoblock Wärmepumpe zum Heizen / Kühlen / Brauchwasserbereitung 12,0 kW

... full table available in the supplementary GitHub repository.

2.5. Binarized region

To account for potential regional patterns while maintaining model interpretability, the origin of each manufacturer was binarized into two categories - European (EU) and non-European (non-EU) - based on the location of the brand's headquarters. This binary representation reduces model complexity, avoids sparsity across many country-level categories, and enables robust comparison between broad manufacturing regions.

2.6. Price modelling

To quantify how technical attributes influence heat pump prices, several regression and machine-learning models were trained and tested, and the best-performing model was selected based on out-of-sample accuracy. Testing multiple model types is necessary because price formation may involve both approximately linear relationships, which are easy to interpret, and more complex non-linear patterns, which simpler models may fail to capture.

We therefore compared parametric linear models (ordinary least squares and Ridge regression), which provide transparent and easily interpretable estimates, with non-parametric ensemble methods (Random Forest and Gradient Boosting), which can represent non-linear relationships and interactions between technical features without requiring explicit functional assumptions [13]. While linear models assume constant marginal effects, ensemble methods allow effects to vary across operating ranges and feature combinations, which may better reflect heterogeneous market behavior.

Rather than normalizing prices by capacity or pre-grouping heat pumps into fixed size classes, rated power was included directly as a model feature alongside all other technical attributes. This multivariate formulation allows the contribution of parameters such as SCOP, acoustic performance, or configuration to be estimated after accounting for capacity, without restricting the analysis to predefined capacity bins. Given the wide capacity range covered by the dataset (approximately 3–13 kW), sufficient overlap exists to disentangle capacity effects from other features while retaining the full information content of the data.

All models were evaluated using five-fold cross-validation on an 80/20 train–test split. Model performance was assessed using standard metrics: the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE). Based on this comparative evaluation, Gradient Boosting was selected as the final modelling approach. Unlike linear models, Gradient Boosting allows for non-linear and interaction effects, providing a flexible representation of price formation across capacity ranges.

To retain interpretability despite model complexity, feature contributions were analyzed using SHapley Additive exPlanations (SHAP), a widely used, model-agnostic framework for quantifying how individual features influence predicted prices in observed data [14]. This approach enables the identification of key price drivers while preserving transparency in a non-linear modelling context.

2.7. Market segmentation (clustering)

To identify structural patterns within the European air-to-water heat pump market, we applied unsupervised clustering to group products with similar technical and commercial characteristics. Clustering is a data-driven method that identifies recurring product archetypes without imposing predefined categories, making it well suited for exploratory market analysis [13]. In this study, Gaussian Mixture Model (GMM) clustering was used, as it allows clusters to differ in size, shape, and overlap, which better reflects heterogeneous technology markets than distance-based methods such as k-means [13].

Clustering was performed using nine standardized features that jointly describe heat pump characteristics (Table 3). To ensure consistent interpretation, several variables were expressed as inverse indicators (e.g. affordability, value, and quietness), such that higher values correspond to more favorable product attributes.

Table 3: Feature description used for clustering.

Feature	Description
SCOP	Seasonal Coefficient of Performance
Quietness	Inverse of outdoor noise level (dBA)
Value	Inverse of price per kW
Max temp	Maximum water heating temperature (°C)

Feature	Description
Power	Rated power at low temperature (kW)
Affordability	Inverse of total price
Storage size	Size of installed water storage (L)
Region	Binarized origin of the manufacturer brand
Config	Monoblock / Split

The optimal number of clusters was determined by minimizing the Bayesian Information Criterion (BIC), with the Akaike Information Criterion (AIC) serving as a consistency check [13]. Based on this evaluation, five clusters were selected. Each product was assigned to one of the five Gaussian mixture clusters, and summary statistics (mean, median, and percentile values) were computed for key technical and price-related variables. To derive interpretable market archetypes, a rule-based labeling system was applied using percentile thresholds across multiple attributes, summarized in Table 4. Clusters were named according to their dominant characteristics - meaning that if a feature showed systematically similar values within a cluster, it served as the defining label, as summarized in the table below. Up to three dominant traits were combined to form concise and interpretable market labels, such as EU-Monoblock-Retrofit or non-EU-Split.

Table 4: Trait assignment criteria based on percentile thresholds. Traits are used to label each cluster with interpretable market personas.

Trait	Selection Criterion
Retrofit	Cluster average max. water temperature > 65 °C (suitable for older radiator systems)
Quiet	Average noise level below 20th percentile (low acoustic output)
EU	>80% of cluster items have manufacturers with EU origin
Non-EU	>80% of cluster items have manufacturers with non-EU origin
Monoblock	>80% of cluster items are monoblock heat pumps
Split	>80% of cluster items have manufacturers with EU origin

2.8. Reproducibility

All data preprocessing, model matching, and analysis scripts are fully reproducible. The codebase, matched dataset, and scripts to regenerate all figures and tables are maintained in repository at: https://github.com/UgneP/Heat-pump-market-transparency.

3. Results

3.1. Dataset overview

The final dataset comprised 696 certified air-to-water heat pumps (AWHPs) from 13 manufacturers, combining technical specifications from the Keymark certification database with retail prices collected via automated web scraping. Valid retail prices were successfully matched for 26.5% of all certified models, as shown in Table 5. Unmatched cases primarily resulted from: (1) inconsistent or unidentifiable product names and technical specifications, (2) incomplete system configurations where models are sold as separate components and a comparable bundle couldn't be created, and (3) units that are not yet available or no longer on the market.

Table 5: Overview of selected HPs matched with a price. The table shows the number of HPs from the filtered hplib database, the number of scraped items associated with the corresponding manufacturer, the number of matches, and the matched percentage.

Manufacturer	# Database	# Scraped	# Matched	% Matched
Alpha Innotec	27	257	22	85.19 %
Bosch	83	732	57	68.67 %
Buderus	84	1128	72	86.71 %
Daikin	875	1025	102	11.66 %
Johnson Hitachi	139	49	22	15.83 %
LG	80	195	58	72.50 %
Mitsubishi	610	480	123	20.16 %
Panasonic	137	215	48	35.04 %
Samsung	148	90	38	25.68 %
Toshiba	162	39	22	13.58 %
Vaillant	89	635	38	42.70 %
Viessmann	171	1015	71	41.52 %
WOLF	23	657	23	100.00 %
Total	2628	6517	696	26.48 %

Across all models, the average retail price was €8,898 ± 2,843 (range: €2,725–17,499), corresponding to a mean price of €1,162 ± 421 per kW. The mean SCOP was 4.72 (3.84-5.45), and the average rated power was 8.25 kW. Outdoor noise levels averaged 54.6 dBA, and the mean maximum water temperature was 59.4 °C (Figure 1).

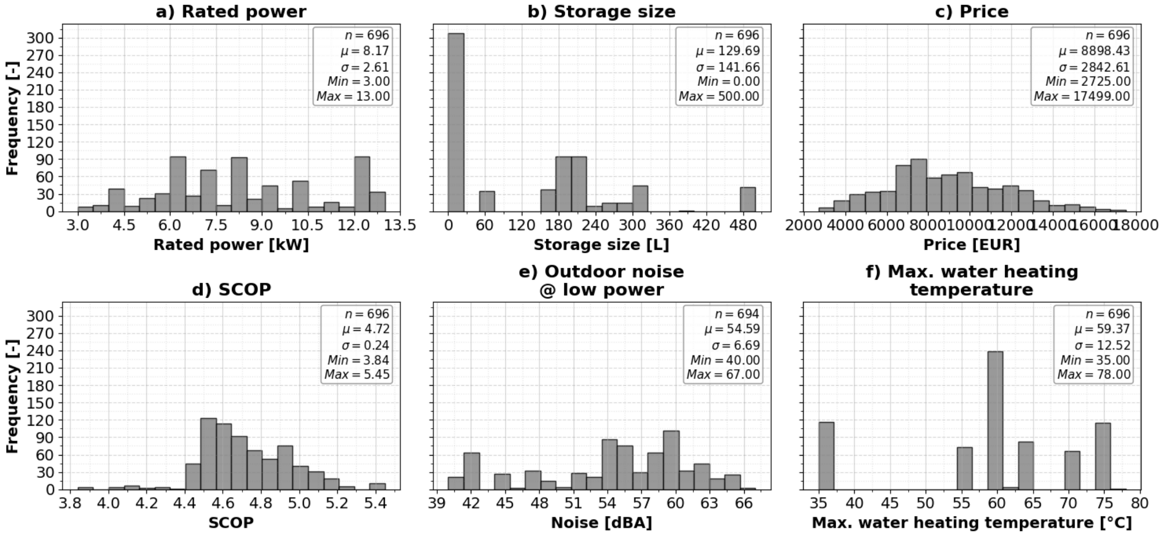

Figure 1: Descriptive statistics of key heat pump features of the scraped heat pump data (n = 696).

Two refrigerants dominated: R32 (n = 446) and R290 (n = 250). R290-based systems had higher prices (mean = €10,746 vs. €7,863) and slightly higher SCOP (4.78 vs. 4.69). Mann-Whitney U tests confirmed significant differences for both - price ($U = 87{,}526$, $p < 0.001$, $r = 0.47$) and SCOP ($U = 68{,}159$, $p < 0.001$, $r = 0.18$). However, refrigerant type is confounded with system architecture: all R290 models were monoblock systems, while R32 appeared in both split and monoblock designs.

3.2. Price modeling with Gradient Boosting model

The Gradient Boosting Regressor achieved the best overall predictive performance, with an R^2 of 0.92 on the training set and 0.83 on the test set, and mean absolute errors of €551 and €887, respectively. These results indicate high predictive accuracy with limited overfitting. The superior performance of gradient boosting compared to linear models suggests that the relationship between heat pump technical specifications and retail prices is predominantly non-linear, reflecting complex interactions between system characteristics and configurations. Although capacity strongly determines price, it is included explicitly as a model feature, allowing the influence of SCOP to be estimated conditional on capacity. The weak relationship between SCOP and price therefore reflects a limited additional price effect beyond what is already explained by system size.

The SHAP analysis revealed that rated power, storage size, and refrigerant type were the most influential determinants of retail price, while noise level, standby power, and COP at the reference point also contributed meaningfully, as summarized in Table 6.

Table 6: Top 7 features ranked by mean absolute SHAP value, representing average impact on predicted price in €.

Feature	Mean SHAP Value (€)	Value ranges in the dataset
Rated power	1174.61	3.44 - 13.0 kW
Storage size	959.80	0 - 500 Liters
Refrigerant type	686.09	R32: 0, R290: 1
Noise	403.77	40 - 69 dBA
Standby power	398.19	0 - 60 W
COP at reference point	246.29	1.65 – 2.65
Tank configuration	206.89	Hydraulic module:0, with warm water tank: 1

In contrast, SCOP and maximum water temperature showed only marginal effects. Rated power showed the strongest SHAP impact (€1,174.61), indicating that variation across the 3.44–13 kW range contributes more to predicted price differences than any other feature. Across all heat pumps in the dataset, the variation in storage-tank size (from 0 to 500 L) leads, on average, to about €960 difference in predicted price. Larger storage volumes, systems with R290 refrigerants and quieter designs were likewise associated with higher prices. The direction of the feature impact is shown in Figure 2.

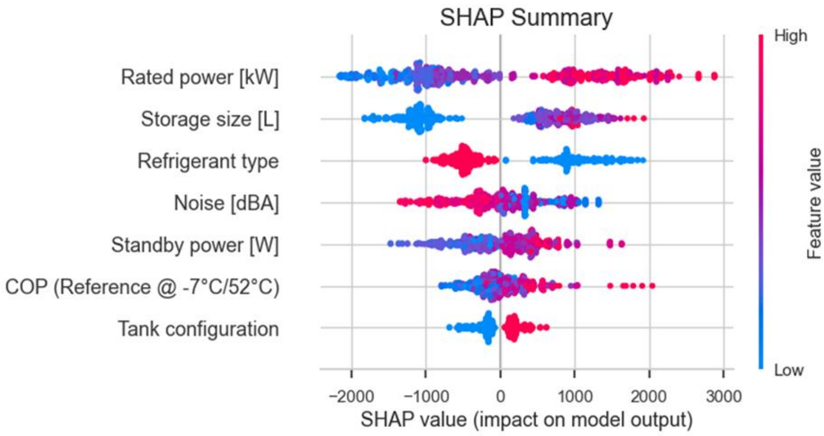

Figure 2: SHAP summary plot showing both global and local feature effects from the gradient boosting model. Each dot represents an individual sample; color indicates feature value, and position indicates the SHAP impact on predicted price.

3.3. Market segmentation

Unsupervised Gaussian Mixture clustering applied to standardized heat pump features (SCOP, noise level, price per kW, rated power, and maximum water temperature) revealed five distinct product groups within the dataset of 696 certified air-to-water heat pumps. Cluster validation using the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) confirmed the selection of five clusters as the best trade-off between compactness and separation (Figure 3).

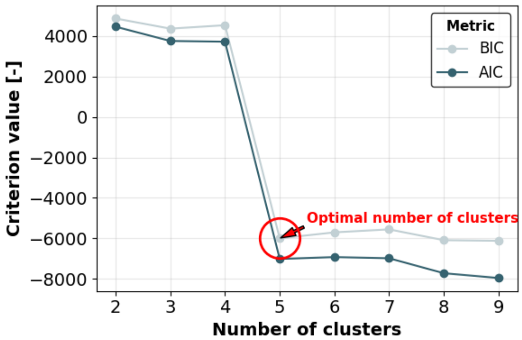

Figure 3: Optimal number of clusters for Gaussian Mixture clustering based on BIC and AIC. Both metrics reach their lowest values at k = 5, indicating that five clusters provide the best balance between model fit and parsimony.

The resulting clusters are primarily structured by manufacturer origin (EU vs. non-EU) and system configuration (monoblock vs. split). Despite no prior labelling, the algorithm naturally grouped European and non-European manufacturers, highlighting systematic differences in product design and market positioning. The radar plots in Figure 4 further illustrate these distinctions, showing that non-EU clusters emphasize affordability and standardization, while EU clusters focus on retrofit capability, quiet operation, and higher supply temperatures - reflecting divergent value propositions within the European heat pump market.

The non - EU clusters (Monoblock and Split, jointly 59.3 % of models) emphasize affordability and standardized configurations, with average prices below €9,000 and typical rated powers around 8-9 kW. These systems predominantly use R32 refrigerant and target the mass residential segment. In contrast, EU clusters (40.7 %) exhibit greater design diversity and specialization. The EU-Monoblock-Retrofit and Retrofit-Quiet groups feature higher maximum supply temperatures (66-71 °C) and lower noise levels (down to 41 dBA), catering to retrofit and comfort-oriented markets. The EU-Split cluster balances moderate efficiency (SCOP $\approx$ 4.8) with compact form factors and mid-range prices. The cluster features are summarized in Table 7.

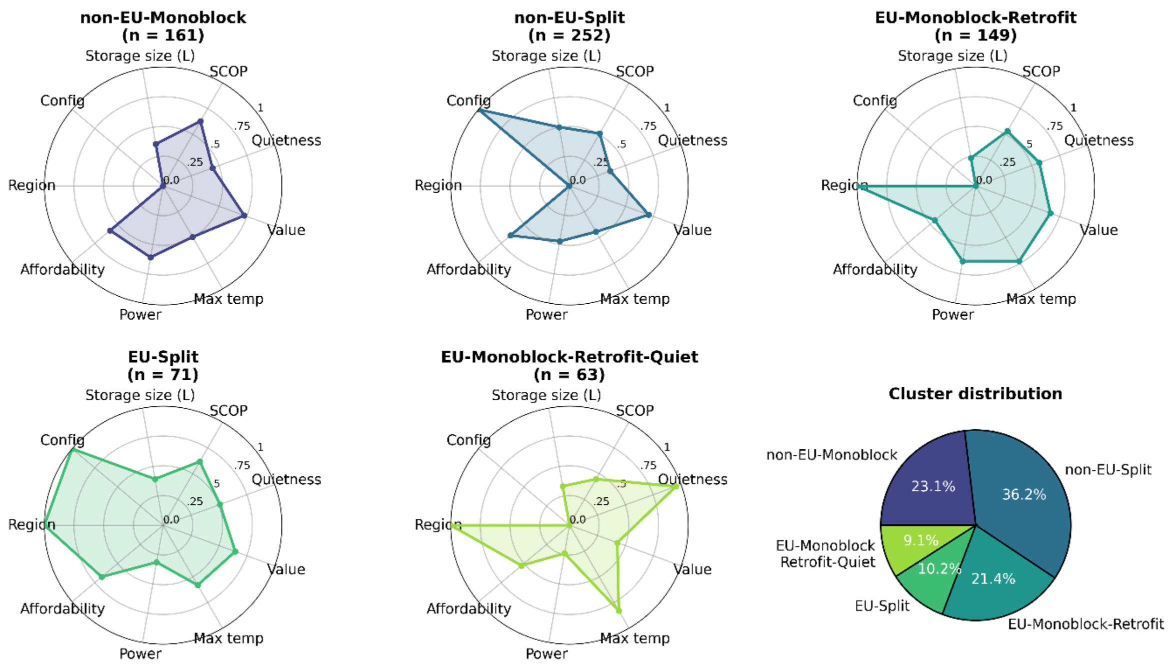

Figure 4: Radar plots illustrating the normalized feature profiles of each cluster (market segment), along with a pie chart showing the market share of these clusters.

Table 7: Detailed segment-level cluster profiles, including number of products, performance indicators, brand, and configuration traits for each cluster.

Attribute	**non-EU-Monoblock**	**non-EU-Split**	**EU-Monoblock-Retrofit**	**EU-Split**	**EU-Monoblock-Retrofit-Quiet**
Number of products	161 (23.1 %)	252 (36.2 %)	149 (21.4 %)	71 (10.2 %)	63 (9.1 %)
Average price [€]	8,894 ± 3,466	7,930 ± 1,710	10,851 ± 3,159	7,529 ± 2,117	9,707 ± 1,752
Price per kW [€/kW]	1,002 ± 410	1,046 ± 270	1,147 ± 287	1,202 ± 323	1,738 ± 418
SCOP	4.85 ± 0.28	4.66 ± 0.17	4.70 ± 0.28	4.84 ± 0.17	4.56 ± 0.01
Noise level [dBA]	56.2 ± 5.0	58.4 ± 3.5	52.5 ± 6.4	54.2 ± 5.5	41.3 ± 1.0
Rated power [kW]	9.3 ± 2.5	8.0 ± 2.3	9.6 ± 2.3	6.5 ± 1.6	5.7 ± 1.0
Max. temperature [°C]	56 ± 16	54 ± 11	66 ± 9	60 ± 0	71 ± 7
Refrigerants	R32 (94), R290 (67)	R32 (252)	R290 (120), R32 (29)	R32 (71)	R290 (63)

Overall, the unsupervised clustering demonstrates that regional manufacturing origin and system configuration dominate market segmentation, while efficiency (SCOP) varies only modestly across clusters. The resulting archetypes - from non-EU-Split budget systems to EU-Monoblock-Retrofit-Quiet premium models - reflect distinct value propositions and consumer priorities within the European heat pump market.

4. Discussion

This study provides the first large-scale, data-driven analysis linking certified technical specifications of AWHPs with real market prices across Europe. By combining Keymark certification data with retail prices obtained through automated web scraping, the research quantifies how product characteristics - rather than nominal efficiency ratings - shape consumer-facing costs. Results from a fine-tuned gradient boosting model (R^2 = 0.87, MAE app. €740) show that rated heating power, configuration, and noise level are the dominant price determinants, while SCOP contributes only marginally. Each additional kilowatt of heating capacity increases price by approximately €535-1,150, and monoblock systems with integrated tanks are on average €2,300 more expensive than comparable configurations.

Unsupervised clustering further revealed five distinct market segments primarily differentiated by manufacturer origin (EU vs. non-EU) and configuration type (monoblock vs. split). European clusters emphasize retrofit suitability, higher maximum water temperatures, and lower noise levels, whereas non-EU clusters focus on affordability and standard residential applications. Together, these findings highlight systematic design and pricing differences across global suppliers and contribute to improving market transparency for policymakers and consumers.

4.1. Implications

The analysis provides empirical evidence on how technical and design-related factors influence heat pump prices in Europe. By focusing explicitly on the equipment cost component, the results are directly relevant for policy instruments such as labeling schemes. The weak association between price and SCOP indicates that higher-priced heat pumps do not necessarily offer higher theoretical efficiency. This creates a challenge for manufacturers, as efficiency metrics - particularly SCOP - remain one of the few indicators that consumers can easily access and understand, potentially exerting price pressure in favor of theoretically more efficient models.

At the same time, several researchers have shown that realized efficiency in practice often deviates from laboratory-based SCOP values reported on certification labels [15,16]. Real-world performance is influenced by operational dynamics such as defrost cycles, part-load operation, switching between space heating and domestic hot water, control modes (e.g., eco vs. comfort), and building-specific conditions. As a result, certificate-based efficiency metrics may not fully reflect achievable performance in typical applications. To avoid disadvantaging advanced, well-controlled heat pump systems, energy labeling and policy frameworks could benefit from incorporating more dynamic efficiency indicators that better represent real-world operation. Reporting performance metrics that account for realistic operating conditions would improve transparency, support fairer market positioning, and enable more informed consumer and policy decisions.

The observed variation of configuration and price across manufacturer origin also suggests limited comparability in product information. Broader access to certified data, supported by consistent product identifiers and naming conventions, could improve the reliability of market analyses and consumer comparisons. Methodologically, the study shows that linking certification databases with retail data through open-source tools enables scalable and transparent market monitoring. This approach can be extended to other heating and appliance markets, providing a reproducible framework for evaluating product-level cost and performance drivers.

4.2. Limitations and future research

While this study provides statistically robust insights into price formation in the German-led Western European air-to-water heat pump market, several limitations suggest directions for future work. Expanding retailer coverage to Southern and Eastern Europe would improve geographic representativeness and allow regional price comparisons. Incorporating temporal price data could further reveal seasonal or policy-driven dynamics. Future research may also benefit from text-mined product specifications or manufacturer documentation to capture attributes such as control sophistication, build quality, or design, which are not fully represented in certification data.

Data limitations remain a central challenge. Matching Keymark-certified models with retailer listings was constrained by non-standardized product naming, reducing the number of valid matches. Although this is unlikely to bias results directionally, it may limit representativeness toward models with clearer labeling. Furthermore, the analysis focuses exclusively on equipment prices, excluding installation, accessories, and operational costs. This controlled scope isolates product-level pricing behavior but should be considered when comparing with total system cost studies.

Finally, the absence of detailed metadata for advanced features (e.g., inverter type, defrosting mode, or control strategy) and the lack of real-world efficiency or installation data constrain explanatory power. Combining certified and market data with field measurements of installed costs and performance would allow more comprehensive techno-economic assessments and better inform policy design, payback analysis, and total cost of ownership modeling

5. Conclusion

This study provides the first quantitative, data-driven assessment linking certified technical specifications of residential AWHPs with real market prices across Europe. By combining Keymark certification data with web-scraped retail prices, it offers an evidence-based view of how equipment attributes influence market valuation. The analysis demonstrates that rated power, configuration type, acoustic performance, and refrigerant choice are the dominant factors shaping product pricing, whereas SCOP - often used in marketing and policy metrics - plays only a minor role.

Unsupervised Gaussian mixture clustering revealed five distinct product segments, differentiated primarily by manufacturer origin (EU vs. non-EU) and system configuration (monoblock vs. split). European products tend to emphasize higher water outlet temperatures and lower noise, reflecting a focus on retrofit and comfort applications, while non-EU products are positioned toward affordability and standard residential use. Together, these results highlight structural differences in product design and pricing strategies among global suppliers.

Beyond empirical insights, the study contributes a reproducible open-data framework that integrates certified product databases with market data, helping to improve transparency and comparability in the European heat pump market. Future extensions combining equipment prices with installation, maintenance, and operational cost data could provide a more comprehensive picture of total system economics. Overall, the findings inform evidence-based policy design, support consumer-oriented transparency, and advance empirical understanding of how technology, performance, and value interact in the residential heat pump sector.

References

[1] European Heat Pump Association (EHPA). EHPA Market Report 2024. Brussels: EHPA; 2024. Available from: https://www.ehpa.org/wp-content/uploads/2024/08/Executive-summary_EHPA-heat-pump-market-and-statistic-report-2024-2.pdf. Accessed 11 November 2025.

[2] Toleikyte A, Roca Reina JC, Volt J, Carlsson J, Lyons L, Gasparella A, et al. The Heat Pump Wave: Opportunities and Challenges. Luxembourg: Publications Office of the European Union; 2023. https://op.europa.eu/en/publication-detail/-/publication/94cf1914-10b1-11ee-b12e-01aa75ed71a1/language-en. Accessed 11 November 2025.

[3] Verbraucherzentrale Rheinland-Pfalz e.V. Luft-Wasser-Wärmepumpen: Eine Auswertung von 160 Angeboten aus Rheinland-Pfalz. Mainz: Verbraucherzentrale Rheinland-Pfalz; 2025. Available from: https://www.verbraucherzentrale-rlp.de/sites/default/files/2025-06/250605_vz-rlp_auswertung_wp_angebote.pdf. Accessed 11 November 2025.

[4] Ban A, Bungău C. (2022). Estimation of the initial investment cost in systems equipped with heat pumps using regression analysis. Journal of Engineering Sciences and Innovation. 7. 293-304. https://jesi.astr.ro/wp-content/uploads/2023/03/2_Ban-Alexandra.pdf

[5] Paiho S, Pulakka S, Knuuti A. Life-cycle cost analyses of heat pump concepts for Finnish new nearly zero-energy residential buildings. Energy and Buildings, Volume 150, 2017, Pages 396-402. https://doi.org/10.1016/j.enbuild.2017.06.034

[6] Bianco V, Scarpa F, Tagliafico LA. Estimation of primary energy savings by using heat pumps for heating purposes in the residential sector. Applied Thermal Engineering, Volume 114, 2017, Pages 938-947. https://doi.org/10.1016/j.applthermaleng.2016.12.058

[7] Miara M, Günther D, Langner R, Helmling S, Wapler J. 10 years of heat pump monitoring in Germany: Outcomes of several monitoring campaigns from low-energy houses to un-retrofitted single-family dwellings. In: Proceedings of the 12th IEA Heat Pump Conference; 2017. https://etkhpcorderapi.extweb.sp.se/api/file/1642

[8] Heat Pump KEYMARK Certificates. Certified Products. Available from: https://keymark.eu/en/products/heatpumps/certified-products. Accessed 12 February 2025.

[9] Tjaden T, Hoops H, Rösken K. RE-Lab-Projects/hplib: Heat Pump Library (v2.0). 2021. https://doi.org/10.5281/zenodo.5521597. Accessed 12 February 2025.

[10] Aste N, Adhikari RS, Manfren M. Cost optimal analysis of heat pump technology adoption in residential reference buildings. Renewable Energy, Volume 60, 2013, Pages 615-624. https://doi.org/10.1016/j.renene.2013.06.013

[11] Bouzghiba H, Géczi G. Comparative analysis of the environmental load of natural refrigerant (R290). Comparative Analysis of the Environmental Load of Natural Refrigerant (R290). Journal of Central European Green Innovation, 10 (Suppl 1), 135-142, 2022. https://doi.org/10.33038/jcegi.3506

[12] Mitchell R. Web Scraping with Python: Collecting More Data from the Modern Web. 2nd ed. Sebastopol, CA: O'Reilly Media; 2018. https://edu.anarcho-copy.org/Programming%20Languages/Python/Web%20Scraping%20with%20Python,%202nd%20Edition.pdf
[13] Hastie T. , Tibshirani R. , & Friedman J.. The Elements of Statistical Learning. Springer Series in Statistics 2009. https://doi.org/10.1007/978-0-387-84858-7
[14] Lundberg S, Lee S. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777, 2017. https://dl.acm.org/doi/10.5555/3295222.3295230
[15] Brudermueller T, Potthoff U, Fleisch E, Wortmann F, Staake T. Estimation of energy efficiency of heat pumps in residential buildings using real operation data. Nature Communications 16, 2834, 2025. https://doi.org/10.1038/s41467-025-58014-y
[16] Nolting L, Steiger S, Praktiknjo A. Assessing the validity of European labels for energy efficiency of heat pumps. Journal of Building Engineering, 18, 476–486, 2018. https://doi.org/10.1016/j.jobe.2018.02.013