THE POTENTIAL OF DATA & PERSONALIZATION FOR REAL ESTATE WEBSITES

A CO-LIBRY E-BOOK

Co-libry

Introduction

We embarked on a multi-year project where we spoke to over one hundred online real estate professionals and conducted in-depth interviews with industry leaders. We've done in-depth interviews with the REA group, Idealista, Fotocasa, Casa.it, Immoscout24, Newhome, Immoweb, Juwai, Funda, and more.

These interviews, combined with insights from the *global real estate portal report* from Mike Delprete and our mentor, Joe Hanna, paint a picture of today's online real estate industry. The **challenges today's real estate portals** and websites are facing are rapidly evolving, and companies need to action, proactively engaging with nascent technology to overcome these challenges.

Traditional online tactics like attracting users to the site and providing search functionality are no longer enough to maximize engagement. In the digital age, **users have higher expectations**. They want a **personalized customer experience**, such as tech giants like Netflix, Amazon, and Spotify. On the other side, advertisers expect more value than "being visible" e.g. lead qualification, insights into their ads' performance, insights into their audience and the market…

In this e-book, we share our insights into today's personalization impact and how the future landscape will look. You'll learn more about data quality, user experience struggles, and opportunities in the industry right now and beyond.

Personalization at top brands

Whether you've noticed it or not, Artificial Intelligence (AI) is already everywhere. It's working behind the scenes, quietly powering personalized experiences in all aspects of modern life. From Spotify to fashion moguls like Tommy Hilfiger, AI is leading the way.

Before we head onto what personalization will bring online real estate in 2021, it will help look at the trends for top general and e-commerce brands since they form how users want to search online.
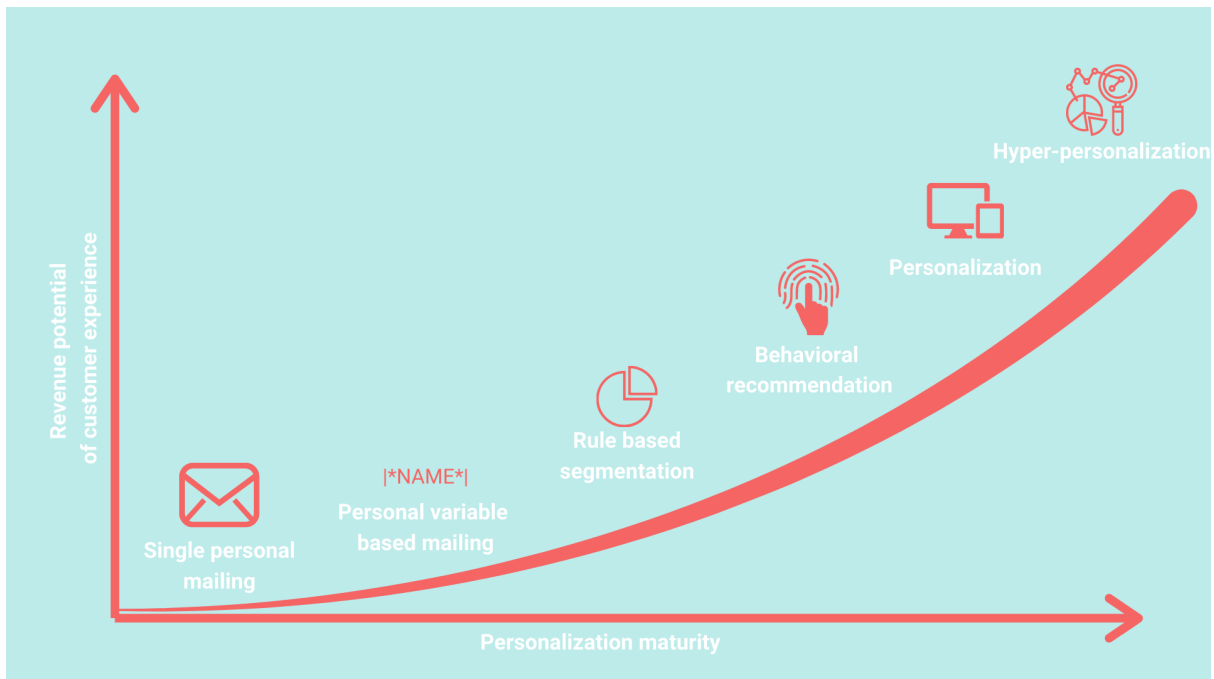
Why personalization matters

" 75% of consumers are more likely to purchase from companies that provide personalized offers "

Brands like Amazon, Spotify, and Starbucks have pioneered the use of AI personalization. The reason why those brand shifted from their traditional approach is simple:

- User engagement with written content has gone down by 60%[1], and information overload is making consumers tune out.

- According to Accenture[2], **75% of consumers are more likely to purchase from companies that provide personalized offers** based on individual preferences.

Online user behavior defines preferences and intent of users. Along with demographic segmentation, websites can deliver a richer, personalized, and more intuitive experience.

Next step, hyper-personalization

The leading brands even push it further and move to **hyper-personalization**. Below are some of the ways tech giants Netflix and Amazon **leverage consumer behavior to deliver better user experiences**.

While traditional personalization might include advertising a customer's name, location, or purchase history, hyper-personalization also **considers browsing, purchasing, and other real-time behavioral data** to hone in on consumer wants or needs. It's more involved, complex, and useful than traditional because it goes beyond necessary customer data.

Netflix

There are over 7k TV shows and movies you can watch on Netflix. A viewer cannot find movies they like to watch on their own. Netflix uses AI and machine learning to power its recommendation engine, which automates its users' search process.

---

[1]
https://business.linkedin.com/marketing-solutions/blog/content-marketing-thought-leaders/2016/the-attention-economy---the-impact-of-attention-scarcity-on-mode

[2]
https://newsroom.accenture.com/news/consumers-welcome-personalized-offerings-but-businesses-are-struggling-to-deliver-finds-accenture-interactive-personalization-research.htm

The recommendation engines are developed by hundreds of engineers that analyze the habits of millions of users. These are **some of the factors and data they consider**:

- Viewer interactions with Netflix services like viewer ratings, viewing history, and more.
- Information about the categories, year of release, title, genres, and more.
- Other viewers with similar watching preferences and tastes.
- The time duration of a viewer watching specific content.
- The device on which a viewer is watching.
- The time of the day a viewer watches.

Every time a user spends time on Netflix watching, it collects data. The more a viewer watches, the more up-to-date and accurate the algorithm is.

Netflix gives their users a **world-class personalized user experience by**:

- Showing movies and shows that might interest them is continually changing.
- Showing different front images and other trailers from the same content to other users.
- Offering the complete description of a movie or just a small summary.

Amazon

Similar to Netflix, amazon uses personalized recommendations to show visitors products they are most likely to buy. Recommendations are visible across every page, channel, and device. According to Dynamic Yield's research, 45 different recommendation widgets were visible on the app homepage alone.

Every recommendation is based on a couple of variables such as:

- Location
- Recent purchases
- Saved items or lists
- User reviews
- Purchases of similar customer profiles

Here are some of the unique ways Amazon is implementing recommendations to **leverage sales and increase average order value**:

- Showing visitors similar products that help site visitors narrow down their purchase decision and immediately add the most suitable product to their cart.
- Recommending products on category pages such as 'recommended for you' and 'more top picks for you.'
- Recommendations of time-sensitive products.

Summary

" hyper-personalization also considers browsing, purchasing, and other real-time behavioral data "

These are just a few examples of how some of the biggest companies leverage personalization to become more successful.

You may look at big giants like Netflix and Amazon, thinking this is not for you. Because of resources, timing, or budget. But in reality, there are endless possibilities for improving the customer experience and getting more engagement.

Leading role of e-commerce

Before we have a more in-depth look at how real estate portals and websites can use data and personalization, we can ask ourselves the following question: **how can we learn from e-commerce and its personalization wave?**

There is no doubt that e-commerce has made giant leaps over the last couple of years in personalization. Last year, trends within e-commerce showed that personalization is something consumers expect when searching online.

We used data from Yieldify's comprehensive survey on 400 e-commerce leaders[3] across the US and UK. Research on this scale provides us an accurate representation of the evolving priorities within the industry.

Below, you can find the current state-of-play regarding website personalization, the challenges it faces, and how it's impacting online real estate businesses.

Figures on personalization

" 54% of the interviewed companies currently use data-driven segments "

- Over **74%** of companies surveyed already have a website **personalization program in place**.

- In 2020, **retention[4] (58%)** had overtaken conversion (55%) and acquisition (45%) as the **key goal** for website personalization.

- **54%** of the interviewed companies currently use **data-driven segments**. However, this was identified as an area of high potential growth, as 89% expected to be using it by the end of next year.

- The three most significant obstacles that stand in the way of scaling a personalization strategy are a **lack of expertise (37%)**, **limited functionality and tools (36%)**, and a **lack of time (35%)**.

Winning components

Currently experiencing a surge in popularity, personalization enables today's consumers to feel their wants and needs are heard and understood. When their wants and needs are

---

[3] https://www.yieldify.com/news/ecommerce-website-personalization-adoption-new-yieldify-report/
[4] Customers or visitors returning to your website multiple times.

incorporated into the shopping experience, becoming an integral part of the e-commerce experience, **consumers feel validated and engaged**.

For brands, personalization offers a way to **contextualize and individualize** the messaging, offers, and experiences they deliver using unique data retrieved from each visitor's profile.

When surveyed, 76% of the respondents said that the most popular option for driving website personalization is **real-time behavioral data**. This answer beats out historical cookie data, third-party data, and individual user profiles ingested from customer data platforms (CDPs).
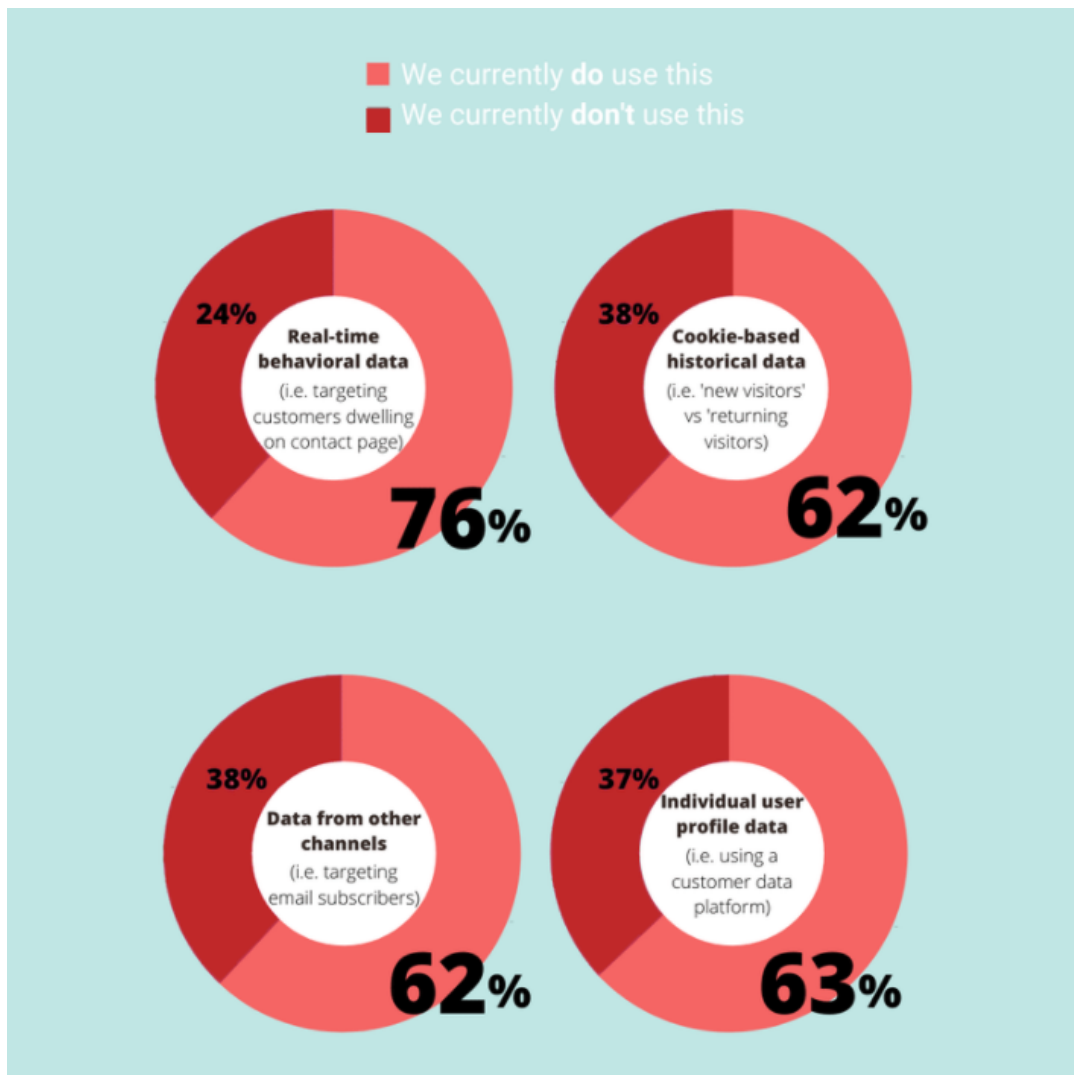
Behavioral data provides critical insights into what users seek most from their online experience. Armed with this understanding, companies can tailor **personalization strategies to align with individual customer preferences**. Product managers, CTOs, developers, and marketers, can then use this data to increase user retention and reduce customer churn.

AI has a vast range of applications, but its benefits are clear no matter the application. Whether it be live chatbots or product recommendation engines, AI is **effectively converting visitors into customers**, and critically, retaining them.

Impact of E-commerce

In the future, merely being a favorite won't be enough. Don't bank on loyalty! In the wake of the COVID-19 pandemic, fundamental changes have been observed in consumer behavior.

The online shopping space is changing what it means to search online. These stores often offer better and personalized shopping experiences. By engaging with these stores, consumers are starting to get used to this type of experience.

According to recent reports[5], Yieldify indicates that between 65% and 85% of consumers intend to keep their new digital shopping habits in the post-pandemic world. With this in mind, the pressure is on for all online businesses. They have to optimize the entire customer life cycle and make the personalization so effective and intelligent that customers keep coming back.

What does this mean for real estate?

It is essential to understand where the e-commerce market is and where it is heading in the future. **Customers are expecting a better and better experience all the time.** We want to show that e-commerce businesses are massively investing in improving user experience and personalization methods.

This is not only happening for e-commerce but also in the online real estate space. As explained previously, **merely being a favorite won't be enough in the future.**

" Consumers start to get used to personalized experiences and pick real estate portals that offer a better experience "

Real estate portals

Personalization is taking the world by storm. Consumers are getting used to it and starting to expect it, so there's no turning back. In 2021, it's incredibly apparent that we can expect to see personalization become a top trend within the real estate industry. Much like in the e-commerce space, **users will prefer real estate websites that offer personalized listings and other user-tailored experiences.** The real estate platforms that are the most user-friendly and engage the most heavily with personalization can expect to see **conversions improve** and more houses sold or rented.

---

[5] https://www.yieldify.com/blog/ecommerce-personalization-trends-after-covid-19/

E-commerce vs real estate portals

There are a lot of plug-and-play e-commerce recommendations and personalization engines available today. We've noticed that some online real estate businesses have tried to use e-commerce AI engines. Unfortunately, attempts to do so have been unsuccessful. Here are the four most prominent reasons why:

- **Small scale user interactions:** in comparison with e-commerce, visitors don't convert or buy frequently on a real estate website. They usually browse through the website for a long time, maybe saving a couple of listings and having a few contact requests. Besides this, checking whether someone bought the house or apartment isn't possible in most cases.
- **One-time purchases:** visitors (usually) only make one' purchase' on a real estate portal. After that the purchase is finalized, they will not return, at least not for a very long time. Because the users are continually changing, it's more challenging to create clusters/profiles, necessary for personalization.
- **Short lifetime listings:** the recommendation engine needs to adapt because the portal's listings are always changing continually. According to Zillow, the average time it takes to sell a house is only 65 to 93 days. By contrast, in general e-commerce, a product can stay the same for years.
- **Unstructured data and hidden information:** Many property characteristics are hidden in the descriptions and images and not explicitly mentioned in the metadata or machine searchable data.

Put, e-commerce recommendation engines are trained differently to meet specific e-commerce goals. They rely on explicit buyer conversions, static products, returning visitors, and standardized data. This is why **an e-commerce recommendation engine will underperform on a real estate portal.**

Every website characterized by the challenges above needs the following three things to ensure their website is ready for personalization.
- A clear understanding of what you want to accomplish with the data or AI project from a business perspective. Recommendation engines are hype, but they might not be the best fit for your goals.
- Based on these goals, end-to-end tracking of each step in the customer journey and all necessary user interactions.
- Good listing data quality, enriched with the hidden data from the descriptions and images.

Further, we will discuss the user tracking and listing data quality in-depth.

Data issues

Throughout our years of experience working with data from multiple real estate portals and websites, we have identified the **four biggest data struggles**:

- **Limited user tracking**

- **Indirect user preferences**
- **Poor listing data quality**
- **Cold start problem**

Let's look at each of these more closely.

" 40% of business objectives fail due to inaccurate data[6] "

Limited user tracking

A majority of real estate portals and websites, either currently or in the past, use little or too general user tracking. Of course, most companies use Google Analytics or even tools like Hotjar to track users for marketing or UX purposes. And everybody claims to be capturing all user data. But usually, the actual in-depth raw data necessary for personalization isn't available. **Portals heavily underutilized the user tracking techniques that form the basis of personalization**.

So, what is effective user tracking? Ideally, this means **tracking every user interaction necessary** for your business case in real-time. In most cases, this translates into contact requests, search queries, clicks, viewed pages, and scrolling actions, whether anonymous or logged in. It also means collecting data on which elements they interacted with (for example, photos) and for how long. By utilizing this data, you can **start profiling your users and eventually (hyper-)personalize your portal**. You are still able to do this GDPR proof. Read how we do it on our blog. https://co-libry.com/blogs/data-collection-gdpr-real-estate/

To start your personalization journey, depending on the size of your dataset, you'll need approximately 3-6 months' of listings and users data. There are **different tools** you can use to do this:

- Google Analytics
- Google Analytics 360
- Snowplow
- Segment
- ...

We won't be covering all the differences between these tools, but we have an extended blog about Google Analytics vs. Google Analytics 360 vs. Snowplow. You can read it here: https://co-libry.com/blogs/snowplow-analytics-vs-google-analytics-360/

User data, and its practical utilization, is one of the primary reasons behind the success of Netflix, Amazon, and the other tech giants. Data leads to powerful outcomes for both consumers and companies.

---

[6] https://www.data.com/export/sites/data/common/assets/pdf/DS_Gartner.pdf

If you haven't already started harnessing the power in your user data, then now is the time to get started. Your user data not only impacts the user experience but also has much value for your advertisers.

You can give real estate advertisers lead qualification, insights into the profiles of the leads reaching out, make sure their listings are shown to all relevant users and provide them with market insights to improve their business.

Indirect user preferences

Going hand-in-hand with incomplete user tracking is the phenomenon that website visitors seldom state their preferences directly. **Users don't want to fill in a complete form to express their preferences.** In other words, their choices will need to be deducted from their past and current behavior on the website.

One can say while performing a search on a portal, users fill in exactly what they want. This is partially true, and we noticed that **some search fields are barely used**. In a good case, the user will fill in the location, price, and listing type. On average location is the best scoring feature with 81% of the users using this search feature. 42% of the users search on max price and only 7% on min price.

Usage of search fields

42%	**7%**	**81%**	**11%**
max price	min price	location	min bedrooms
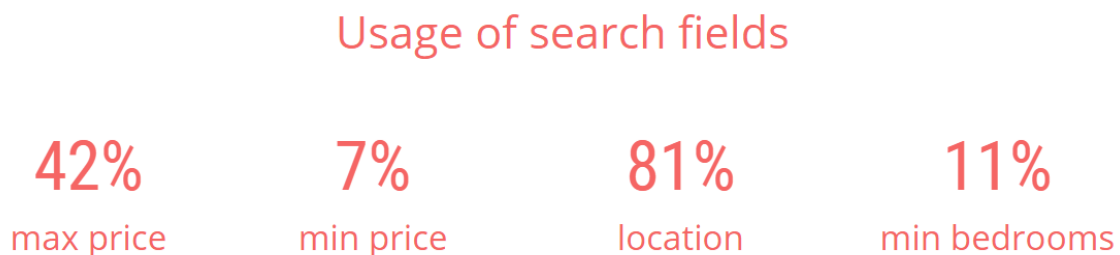
Furthermore, users usually **take their ranges broader** as they don't know yet what they want or fear missing out on the house of their dreams.

The only way to get more insights into users' needs and desires is to deduct their preferences from their behavior on your website.

Poor listing data quality

" 66% of organizations with clean data reported boosted revenue[7] "

Before you start implementing data-driven or AI solutions, you should conduct an Exploratory Data Analysis (EDA) on both your listing as users' data. We'll cover the topic on EDA more in-depth later on. But by doing this, you can **determine the data quality and see what is present in the data and what is missing**. This process will uncover where your data is lacking and which steps are necessary to utilize your data fully.

---

[7] https://blog.zoominfo.com/the-effect-of-dirty-data-on-roi/

From all of the EDAs done over the years, the biggest challenge we saw was that **most portals have lots of blank and empty fields** in the listings data. Often structured data fields such as surface, amount of bedrooms, and more are missing. More 'advanced' structured fields such as elevators, swimming pools, garage, and more were only limited available.

Bear in mind, a missing value can also be a value, meaning that it's merely not present, or we do not have any information regarding this field. But if you have, i.e. 87% of the cases not filled in, your variable's usability will also be quite poor. **This is usually due to real estate agents and sellers, who forget or ignore most optional fields.**

20.7%	55.4%	13.5%
of our client's variables are perfectly filled in	of our client's variables contain 80% blank and empty fields	of our client's extra features are filled in on average

This results in a poor user experience because when visitors use the search engine, listings with missing values won't show. We solved this problem for multiple real estate websites by **looking at the listing descriptions and images**. We found that these were **packed full of valuable information** that wasn't available in the searchable data. So we extracted them with the help of our AI engines.

With one of our most recent clients, we improved their data quality by over 30% in searchable features. More on this in the next section, '3: extra use cases on data & personalization for real estate'.

Cold-start problem

The term derives from cars. When it's really cold, the engine has problems starting up, but it will run smoothly once it reaches its optimal operating temperature.

In our context, it refers to the case where the **amount of available data is limited, the recommendations are poor**, or they lack full coverage over the entire spectrum of possible recommendation combinations.[8]

In terms of recommendations means that a recommendation engine meets a **new visitor for the first time**. Because there is no user history about her, the system doesn't know the user's personal preferences. Getting to know your visitors is crucial in creating a great user experience for them.[9]

There are numerous ways and algorithms to counter the cold-start problem, but one must be aware that this issue is present.

---

[8] Aggerwal, C. C. (2011). Recommender Sytems: The Textbook. Heidelberg New York Dordrecht London: Springer.
[9] https://yuspify.com/blog/cold-start-problem-recommender-systems/

Solving data issues

Analyse data quality

As mentioned above, an EDA is essential to assess the quality of our user and listings data. This is a process of performing initial investigations on data to discover patterns, spot anomalies, test hypotheses, and check assumptions with summary statistics and graphical representations.

" 1 in 3 business leaders don't trust the information they use to make decisions[10] "

At Co-libry, we use this process to get to know the data of new clients. We look at every field the client has in their database, check for anomalies, and map it into our dedicated template. The EDA is always the first thing we do when working with a new client. By doing the EDA first, we can identify the data's potential and fix any data quality issues we encounter in the early stages. Doing this ensures we have everything we need to implement effective AI solutions.

EDAs fulfill an important internal function for creating AI solutions, saving us time on implementation and troubleshooting later down the road. Despite being an internal process, we share the results with our clients. Why? Because it empowers them to identify problems in their data processing pipelines across their business. It enables them to address issues with a data-driven approach. Which, in most cases, gives them already a quick win for the search engine performance.

In terms of machine learning and AI, the well known saying: `Garbage in is garbage out` also holds true. If the data used to train your AI models is impure, your results will also be lacking.

We recognize that there is no standard approach to collecting and structuring data. Every client will have a different tech-stack and way of managing their database structure. They will have their type of EDA, but following you can find our EDA's typical glossary.

EDA

Listings data

Firstly, we take a look at the **general dataset**. How many **variables** we have, how many **observations**, how many **duplicates**, how many variables are barely filled in (+95% NA values), how many variables perfectly filled in,...

In a second step, we make a separation between string variables, categorical variables, numeric variables and timestamp variables.

---

[10] https://partner-connect.netapp.com/fr/partner-directory/ibm-belgium-sanv/ibm-infosphere-biginsights

String variables are text strings on which arithmetical operations cannot be performed (i.e., a property description). The categorical variable can take on one of a limited and usually fixed number of possible values (i.e., the transaction type of a listing, for-sale, or for-rent). These kinds of variables can be both digits as texts. Numeric variables are attributes or values described using a number (i.e., price of a property). Lastly, timestamp variables are a sequence of characters or encoded information identifying when a particular event occurred, usually giving date and time of day (i.e., time and date when a listing was created).

Every group of variables has a different approach, yet some checks are standard i.e., check how many times filled in correctly. A variable can be filled 100% of the time but have 99% of the values, equivalent to NA.

For the **string variables**, we take a look at the strings' length, how many words, how many sentences and how many characters. For the **description and title fields**, we also take a look at which words are frequently used. These insights are useful to determine what kind of words we can extract from the description and title using NLP.

ID fields or unique identifiers are often strings. To prevent rounding errors and conversion errors, we look at how many fields are unique in a certain table or subset. This to check for potential data quality issues.

**Distribution of number of characters in description**

Number of occurances listing type
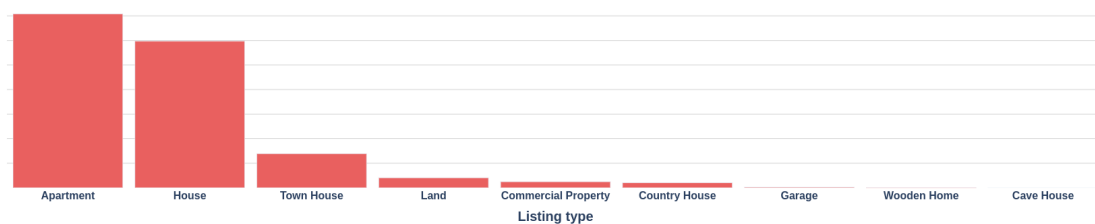
Categorical variables are handled differently. There we take a look at the number of unique values, the most frequent groups, and the frequencies distribution. Another check is whether the categories are exhaustive and mutually exclusive.

When taking a look at **numeric variables**, the distribution comes in handy. Looking at the distribution of a given numeric variable, one can learn great things. I.e., a negative price will point out that this variable has data quality issues. Longitude and latitude have a range where they're in, so when an observation falls outside of this range, the value is unusable.

Price boxplot
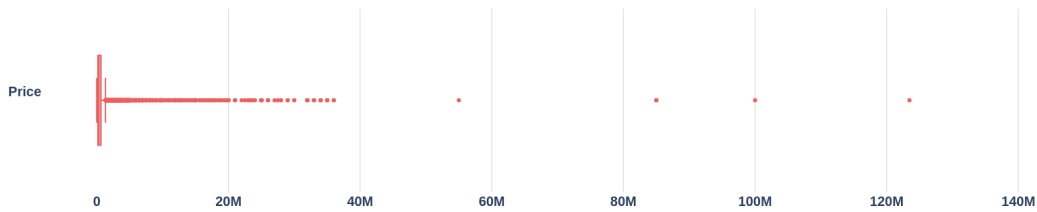
Now the graph above isn't that readable as there are many outliers. If we remove the top 1% of the observations, we get a more readable graph. That being said, the first boxplot is useful and gives you an excellent indication that there are a certain amount of outliers present.

Price boxplot with 99% quantile removed
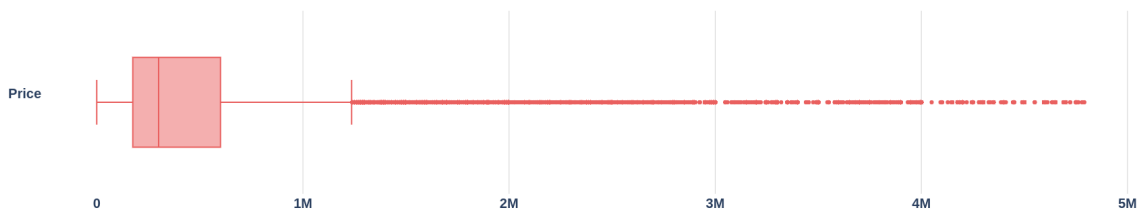
Lastly, the **datetime variables**, here the minimum and maximum, are already a good starting point. The next step is to look at the different dates of the variables and spot whether a specific date has considerably more or fewer occurrences.

Number of listings created in function of date
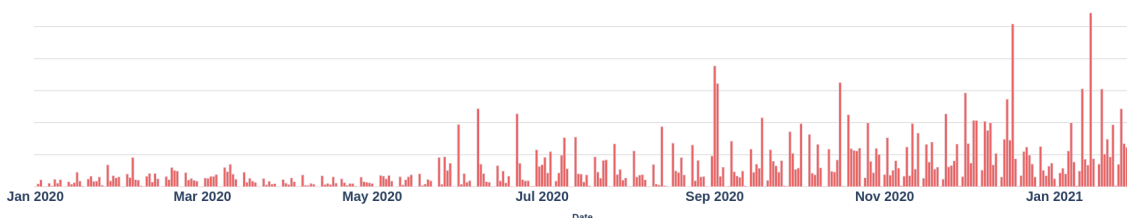
User tracking data

These kinds of datasets are more challenging to analyze uniformly as every portal is different. But we usually start with some fundamental insights such as **number of sessions, number of events, and number of users**.

Secondly, we take a look at the **timestamp** variables. We determine the range, minimum and maximum. Furthermore, as the datetimes of the listings data points, it is to look at the different dates of the variables and spot whether a specific date has considerably more or fewer occurrences. It's not uncommon to find out that certain dates are missing. This imposes a data quality issue in the historical data.

Last but not least, we take a look at the **different events that occur** on a portal. In the first phase of this step, we look at picture events, detail pages visited, search events, contact events, and content events. If you're already able to deduct these, that's already a big step. This is expanded to portal specific events and looking at the filled-in search queries in the next phases.

Conclusion

We typically find that the minimum data requirements of fields are mostly there for all clients. However, the differences in data management can present challenges. There are often anomalies in the data that can skew results or give undesirable outcomes.

" Clients often respond: 'We've never looked at it in this way', in other words in a data-driven way "

A simple example of this is negative pricing. Negative pricing often happens as a result of human error (typos or a moment of inattentiveness). Other times, you could be dealing with fraud or spam. It's always useful to spot the causes of these kinds of irregularities and tackle them at their root.

To conclude this chapter, we'd recommend every company starting with AI to do an EDA at the beginning of your AI projects. But since your database isn't a fixed object, we suggest doing this on a quarterly or half-year basis.

Enrich your data

The more you know about your listings, the better. Every feature and data point within your listings have the potential to drive personalized experiences. For example, if you know which type of images (kitchen, facade, living room, and more) a user is looking at, and for how long and how frequently, **you can predict the probability the user will buy or rent** that house. Please note that we're not advocating quantity over quality here. Data quality is paramount in AI, and you should make no compromises in this area.

By leveraging ML and AI, you can **enrich and validate your unstructured data (descriptions and images)** and take your first step towards making your data robust and AI future proof. The next chapter will take a deeper dive into how Natural Language Processing and Image recognition can help you do this.

Summary

Real estate companies that start using data-driven solutions must first overcome the **most significant data struggles: limited user tracking, indirect data, and low-quality listing data.** Extensive **user-tracking** needs to be conducted to discover insights and define customer segments and profiles. Tools like Google Analytics 360, Segment, and Snowplow can help you achieve this. Data from the search and the users' web-behavior should be captured to offer more comprehensive personalized experiences.

And before you start processing your data or start training your AI models, doing an EDA will save you a lot of time and resources.

Use cases for real estate portals

In this chapter, you will discover three use cases **you can start with today to personalize your real estate portal and improve your data quality**. Following, you can first find a list of the most common use cases.

- Natural language processing (NLP)
- Computer Vision (CV)
- Personalized content
- Lead qualification
- Recommendation engines

Natural language processing

The **text description** of a real estate listing often contains **important information** about the listing. This information can be the **basic features** like the property's price, how many bedrooms it has, or the more **specific ones**, such as a detached, garden, shower, open kitchen or the advanced ones whether it is new or needs renovation, child friendly, and more. This information can often make or break a sale depending on what the prospect is looking for.

Either this information is not included in the searchable data, it's incomplete or incorrect. When this happens, users will have a poor experience. They **might entirely miss listings that would be a perfect fit** for them since the data isn't searchable. Or they'll get listings that aren't relevant. Either way, the user has a poor experience.

In a recent article of Online Marketplaces (Property Portal Watch), real estate portals such as Wikicasa, Zoopla, Bien'ici, and Zillow reveal **new user search habits**. Here's what they say:

- **Wikicasa (Italy):** "a general increase in searches for terms related to an increase in living space such as 'detached' and 'garden.' "

- **Zoopla (UK):** "properties with features such as gardens and balconies, search terms ranked #1 and #8 respectively on Zoopla's 2020 list of most searched features."

- **Bien'ici - France:** "an uptick of 78% for those searching for outdoor space with the balcony/terrace tick box option being particularly popular among users."

- **Zillow - USA:** "The desire for increased space seen elsewhere is present, with Zillow predicting that homes with a 'backyard oasis' will be a trend in the coming year. Also predicted to be important in the new year are amenities conducive to intergenerational living as people seek to maximize time with relatives by moving

in with them as well as 'homecation' amenities such as swimming pools, firepits and a lakeside dock."

" Detecting keywords such as 'home office,' 'renovated,' or 'garden' significantly impacted our user experience and business goals "

In this particular use case of NLP, we want to **recognize specific keywords in the descriptions**. These are determined using a join of 2 techniques. In a first step, the business speaks and indicates which keywords are essential to know. In a second step, the data speaks and means which keywords are present in the data. These steps are joined into one glossary of words the NLP will need to recognize.

Textual Description $\xrightarrow{\text{NLP}}$ **Detected Keywords**

Live on a quiet tree lined street in a classic pre-war brownstone located in prime Park Slope. This two bedroom house with home office has dark hardwood floors, original moldings, high tin ceilings, and windows in every room. The bathroom of was recently renovated and has a shower. The property also has a small garden.

Quiet	✓	Pool	
Two bedrooms	✓	Garage	
Office	✓	Fitness	
Garden	✓	Renovated	✓
Hardwood	✓	Shop	

You can find many open-source models that can 'interpret' a given text, but these models are trained so generally that they will not have the performance a real estate portal wants and needs. That's why we've trained NLP models that are specific to real-estate (jobs and cars), so these models have only seen real estate data and thus not be confused by all the general text they can see.

These models do spot not only the one-on-one matches but also writing errors. In addition, they also take into account similar words. i.e., we want to determine whether a description would be classified as 'quiet.' **The model will recognize the following terms as 'quiet.'**

['quiet_residential', 'peaceful_street', 'quiete', 'calmness', 'calm_brightness', 'quiet_and_wooded', 'quietude', 'quiet_and_peaceful', 'calmth', 'quietly', 'calmly', 'quiet', 'peacefuul', 'aquiet', 'quieton', 'little_traffic', 'andquiet', 'calms', 'peacefull', 'quiet_neighborhood', 'quietresidential', 'acalm', 'calm_and_tranquility', 'quiett', 'quiet_and_green', 'peaceful', 'calm', 'calm_street']

Improving user experience

Once the NLP is done, you can use this data to improve your real estate advertisers' user experience as your website users.

To help out your real estate advertisers, you make sure that their listings show up in all the relevant search queries. **To improve the user experience for your website visitors, you can do three things:**
1. You can enable them to fill in more search criteria.
2. You can present them with more filter options.
3. You can show them a summary of all the important features on the detail page of a listing.

In general, this will help to improve user experience and brand satisfaction. Because you help your website users to **search more precisely and easily find the listings they want**. And your real estate advertiser is happy because he will get **more and better leads**.

Improving organic search results

Short-tail keywords typically consist of one or two words, whereas long-tail keywords consist of three to five or more words. Short-tail keywords often get the most traffic, even thousands of searches a day, but they are incredibly competitive and have a low conversion rate. (Example: 'Flat rent London')

When users are looking for more specific results, they tend to use long queries to get more relevant results. **Long-tail keywords are easier to score for in Google and have a higher conversion rate** because they are so specific.

Nowadays, it's much more common for users to perform search queries in google such as **'2 bedroom flat to rent in London all bills included'** or . By creating tons of SEO pages built for short and long-tail key**'1 bed flat short term rent'**words, you're able to increase your organic traffic from Google.

Based on the extra NLP data, you can auto-generate short and long-tail SEO listing pages combined with effective URLs
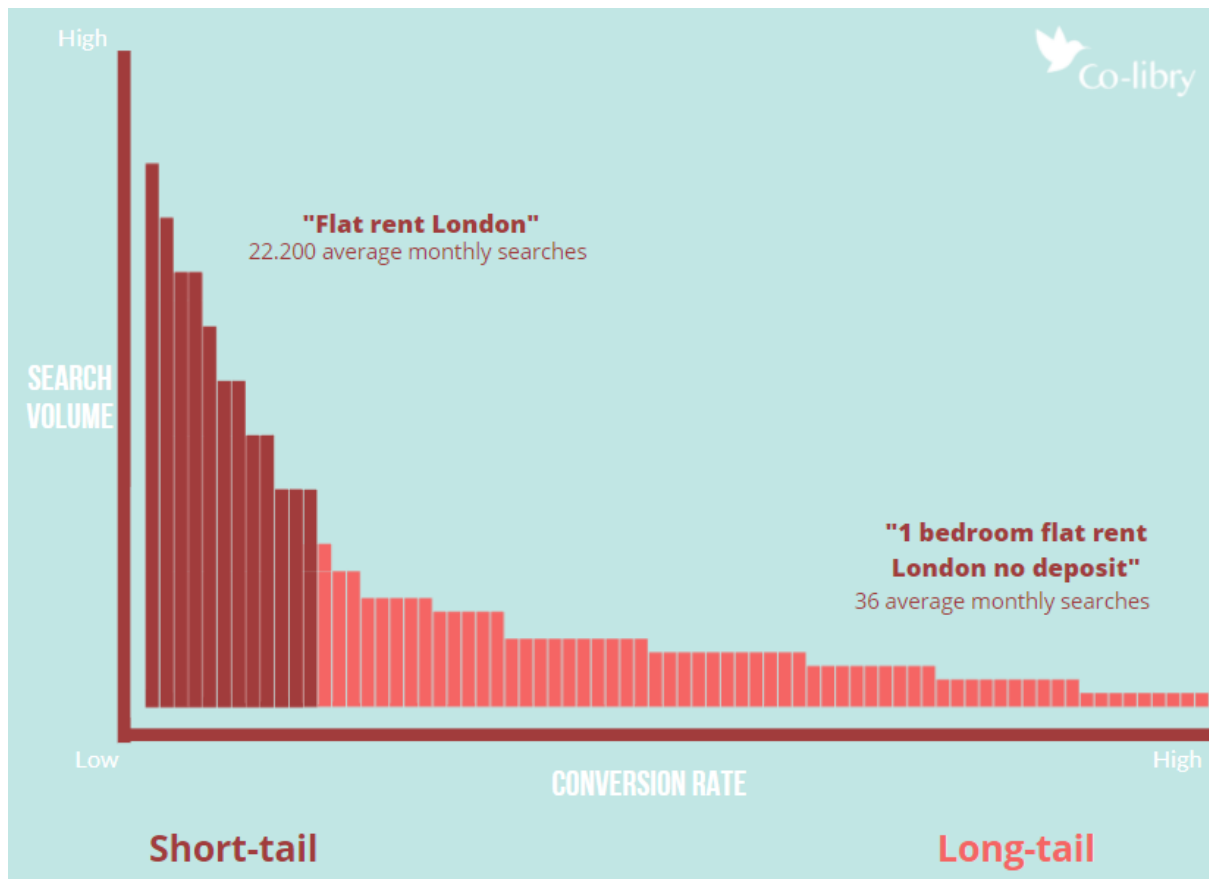
If you have thousands of listings on your portal, manually doing this work is impossible. Thanks to NLP, you're able to achieve this feasibly.

Utilizing keywords in the right way can have a significant impact. For example, we conducted some **research** for London (UK) on long-tail keywords, which you can find here: https://docs.google.com/spreadsheets/d/1joQG0qimjDhDO7GLS0wVE2ytRG4UVQZVAVuA A3ZhyK0/edit?usp=sharing (note: this is just a small subset of all possible keywords for London)

Now, Google some of these keywords and see for yourself. Notice how **almost none of those keywords result in specific landing pages** containing only those listings? What an opportunity, right?

Lead qualification

We define a lead as a contact request via the contact form or chat directly to the real estate private or professional advertiser.

We noticed advertisers typically received the email, phone number, name, and sometimes a short message from the user through our interviews, but there is **hardly any lead qualification**.

However, by tracking user behavior on the website, you can gain a lot more information about its visitors and potential buyers. You can uncover information about the **types of**

listings they prefer and what **stage in the buying or renting process** they are in. You can determine whether they are a casual browser or somebody looking for Pinterest inspiration. You can discover what **category of user** they fall into, such as young first-time buyers, investors, DIYers, and families. You can even determine how likely they are actually to buy or rent a place.

By building these profiles, you can provide our advertisers deeper insights into the leads you send them. With the example shown in the image below, your advertisers can personalize the interactions they have with their leads.
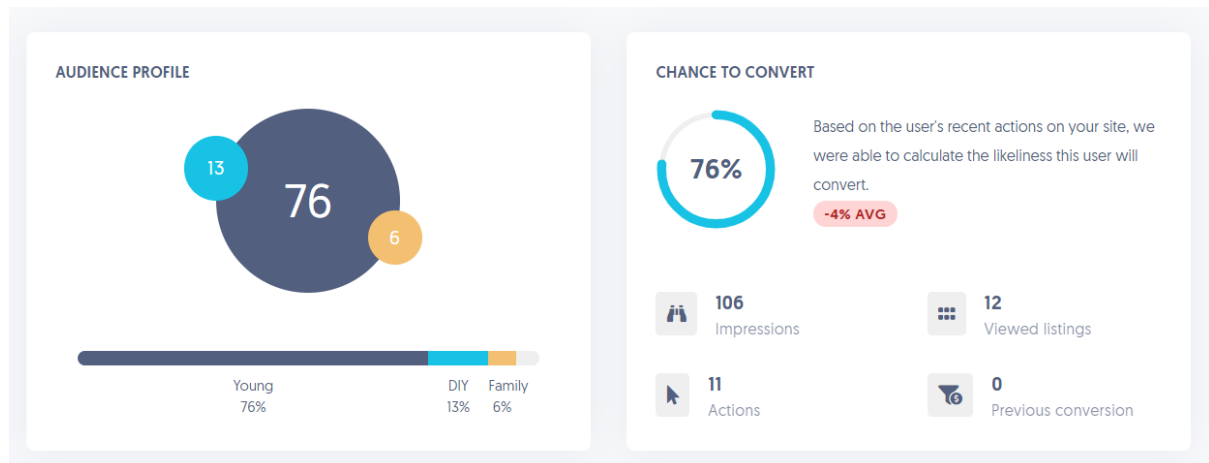
These more personalized interactions start with the response time. The advertiser will want to interact immediately with a user with a high probability of buying or rent. Another form of personalization is that a young family will be interested because there are a couple of playgrounds and schools nearby. If the kitchen needs to be refurbished, the advertiser can sell this to the DIY'ers as an excellent opportunity to save money. As for the non DIY'ers the fact that this opens opportunities for a personal touch can be pointed out.

Recommendation engines

" 38% revenue of product growth when customers receive transparent ads based on their activity on a site[11] "

As we've moved into an era of big data, it is increasingly important to find ways to sift through the vast amount of data. Recommendation engines are incredibly adept at this task.

They'll **enable website visitors to see the most relevant listings based on their taste, style, and other preferences.** They also ensure that the user gets the right information right away and they're not presented with information that doesn't align with their wants and needs. By leveraging these benefits, recommendation engines maximize conversion rates and improve user experiences.

---

[11] Harvard Business - https://hbr.org/2018/01/ads-that-dont-overstep?autocomplete=true

The three most relevant recommendations systems are:

- Content-based recommendations
- Collaborative recommendations
- Hyper-personalized recommendations

Content-based recommendations
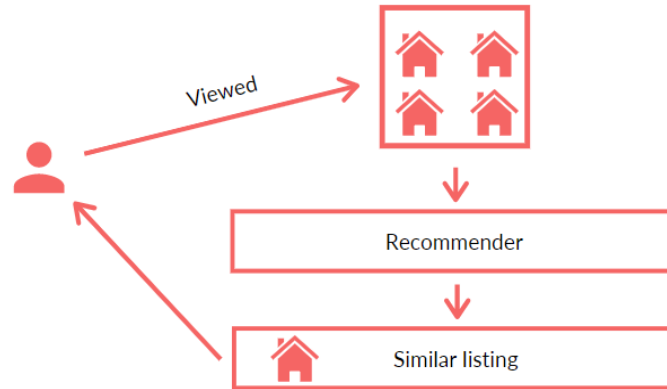
Content-based recommendation systems work by combining data from each listing or web element to create a profile. We can then match this profile to user preferences. For example, these could be the price, address, amount of bedrooms, and year it was built for a house. A list of features for each home is used to create specific profiles.

User profiles are generated by tracking their behavior on the website or their user history.

With this information, **content-based recommendations can serve your visitors with personalized listings based on their previous user behavior, also known as 'similar listings.'**

Collaborative recommendations
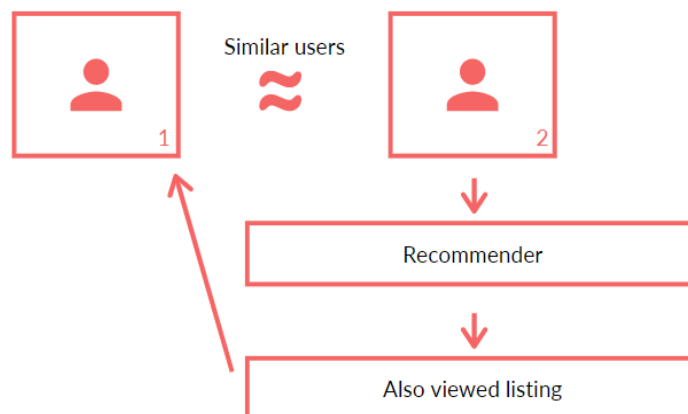
Collaborative recommendations suggest other listings based on similar user profiles.
If you've ever seen **"other users also viewed"** on a website, then you've witnessed collaborative recommendations in action. This technique is a powerful way of providing users

with new types of listings and exposing them, for example, to new neighborhoods that they otherwise wouldn't have considered but could be a good fit.————
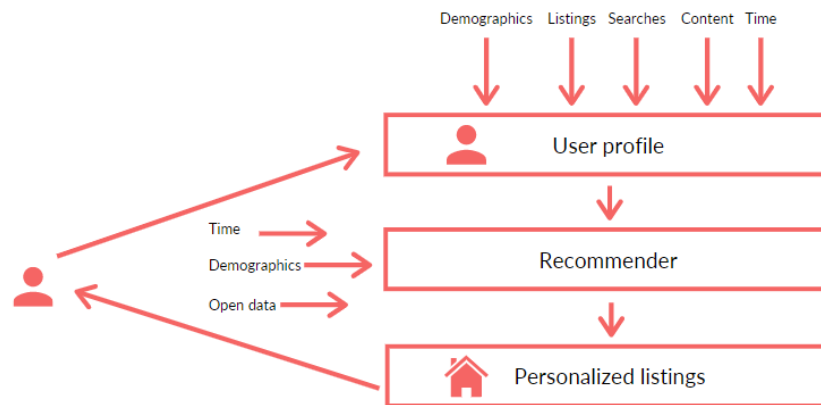
Personalized recommendations put the user central. The recommendations engine will gather as much information as possible on the user. We are talking about which **searches has he/she done** when the user is browsing on your portal, which **content has he/she viewed**, open data, and more. The recommender will aggregate all these data points and construct a thorough user profile. **Using this user profile, it'll know what to send to the user, where to send it, and when to send it.**

For example, a user is mostly browsing on his mobile device during his lunch break for his office in Brussels. Sometimes he looks at a few more listings at home in Ghent at 8 PM.

He is searching for apartments for sale in Ghent with, says, two bedrooms at a max price of €350.000, given a specific neighborhood. The recommender will take the user's history, alongside his current interactions, the time of browsing, and the open data it can find. Using past interactions, the recommender noticed that he's always searching and looking at listings in a given neighborhood, but his price range differs from €300.000 to €400.000.

The user is looking at apartments with a minimum number of bedrooms of 2, so he looked at apartments with 3 and 4 bedrooms. Furthermore, given the user tracking, the recommender also noticed that the user is always looking long at the living room and kitchen while going quickly through the other pictures. He goes back to work at 1 PM. At 5 PM, there is a perfect match for the user. A 3 bedroom apartment in Ghent with three bedrooms, in that neighborhood he's searching and for a price of €359.000. Now the recommender notices that this is a good match. He will send a pop-up or an email to our user with the listing at 8PM at his home in Ghent, showing first the living room picture.

The user opens it. The recommender celebrates as it got it right. The feedback is stored in his memory. At 8 PM, another listing comes online, not really in the neighborhood, two-bedroom apartment in Ghent for €325.000. This is a little more fuzzy match, so the recommender waits until the user arrives for his daily visit during the lunch break.

The recommender recommends the less fuzzy match. The user clicks on it but quickly changes to another listing. Okay, the recommender now got it. Price is less an issue if the other parameters match.

These kind of recommendation engines are often a hybrid between several recommendation algorithms together with state of the art real-time data processing pipelines.

Real estate portal example

We came across this interesting article from Compass[12]. They successfully implemented similar listing and personalized recommendations in their system.

Here's what they have to say:

" The response has been overwhelmingly positive: a 153% increase in homepage clickthrough rate and a 107% increase in engagement actions (such as sharing or saving a listing, or contacting the listing agent) on pages referred by these features."

Summary

Data is extremely powerful, and you will get even more powerful results if you use better data. **Much of the data of real estate engines is underutilized**—for example, the details hidden within images and listing descriptions are typically absent in the searchable data.

By **profiling your user**, you can provide better-qualified leads to estate agents and private sellers. User insights can help you ascertain how likely someone is to buy and when, and what demographic they fall into. This information is invaluable when it comes to closing sales.

To provide compelling, personalized experiences, **content must be relevant to wants and needs of your website visitors**. Recommendation engines achieve this by taking user preferences (for example, in style, taste, budget, age of property) and providing listings that align with these preferences. Ultimately, this results in the user having a better experience because they spend less time searching for the right properties and more time viewing the properties relevant to them.

What's next?

We hope that by now you have some more insights into the potential of personalization for online real estate and where the market is heading.

Implementing some of these techniques and following the road of personalization is not always easy. But it is where the consumer market is heading, which sets companies apart from their competitors and eventually gets more revenue.

---

[12]
https://medium.com/compass-true-north/similar-homes-and-homepage-recommendations-new-frontiers-of-ai-in-real-estate-1102330561eb

This ebook is only the tip of the iceberg of what is possible. There's still a lot more to discover in terms of data and personalization.

If you are curious about what data and personalization can do for you, your team, or your company – don't hesitate to schedule a free 30-minute call with our experts.

https://calendly.com/co-libry_wendy/let-s-get-to-know-each-other-co-libry-2

Fire all your questions at us. We're here to help you.

Who's Co-libry

Co-Libry enables real estate, car and job websites to personalise their customer approach through a variety of data and AI solutions.

Our mission is to accompany those companies to grow their revenue, market share and build a future proof business. We focus on a seamless user and customer experience.

Co-Libry captures the value of data and turns it into a personalised experience. Resulting in increased website traffic, higher conversion rates, faster deal cycles and new data-driven revenue streams.

Our clients choose Co-libry because of our know-how and focus on real estate, cars and job websites instead of general e-commerce. But also to reduce their risk, time of implementation and costs of such AI projects.

Some of our clients: