

# FAR-Trans

An Investment Dataset for Financial Asset Recommendation

Dr. Richard McCreadie  
[IJCAI - 29/07/24](#)



**WORLD  
CHANGING  
GLASGOW**

**A WORLD  
TOP 100  
UNIVERSITY**

# Acknowledgements

---



- This work is an outcome of a collaboration between the University of Glasgow and the National Bank of Greece



Dr. Javier Sanz-Cruzado

University of Glasgow – Financial Informatics



[javier.sanz-cruzadopuig@glasgow.ac.uk](mailto:javier.sanz-cruzadopuig@glasgow.ac.uk)



Nikolaos Droukas

National Bank of Greece – Analytics Centre for Excellence



[droukas.nikolaos@nbg.gr](mailto:droukas.nikolaos@nbg.gr)

- ... and was funded via



Horizon 2020  
European Union Funding  
for Research & Innovation

within the **∞Infinitech** project

- ... with later work supported by



Engineering and  
Physical Sciences  
Research Council



Financial Recommender Research has a **problem...**

# The Pursuit of Profitability

- If we look at other recommender system fields, the goal is to find items that the customer will like
  - ‘Find me movies I would enjoy watching’
  - ‘Recommend me clothing that I would like to wear’
  - ‘Suggest a tasty dish to make tonight’
- ... but in finance, most solutions are customer (investor) agnostic; all customers will be recommended the same assets, regardless of preferences or circumstances
- This is because these solutions optimise for asset **profitability** only
  - There are many reasons for this...
    - Surely the investor’s goal is to make money
    - Profitability is ‘easy’ to measure
    - I don’t have real customers to learn from



# Why is profitability not enough?

---



PROFITABILITY DOES NOT CORRELATE WITH USER PREFERENCES

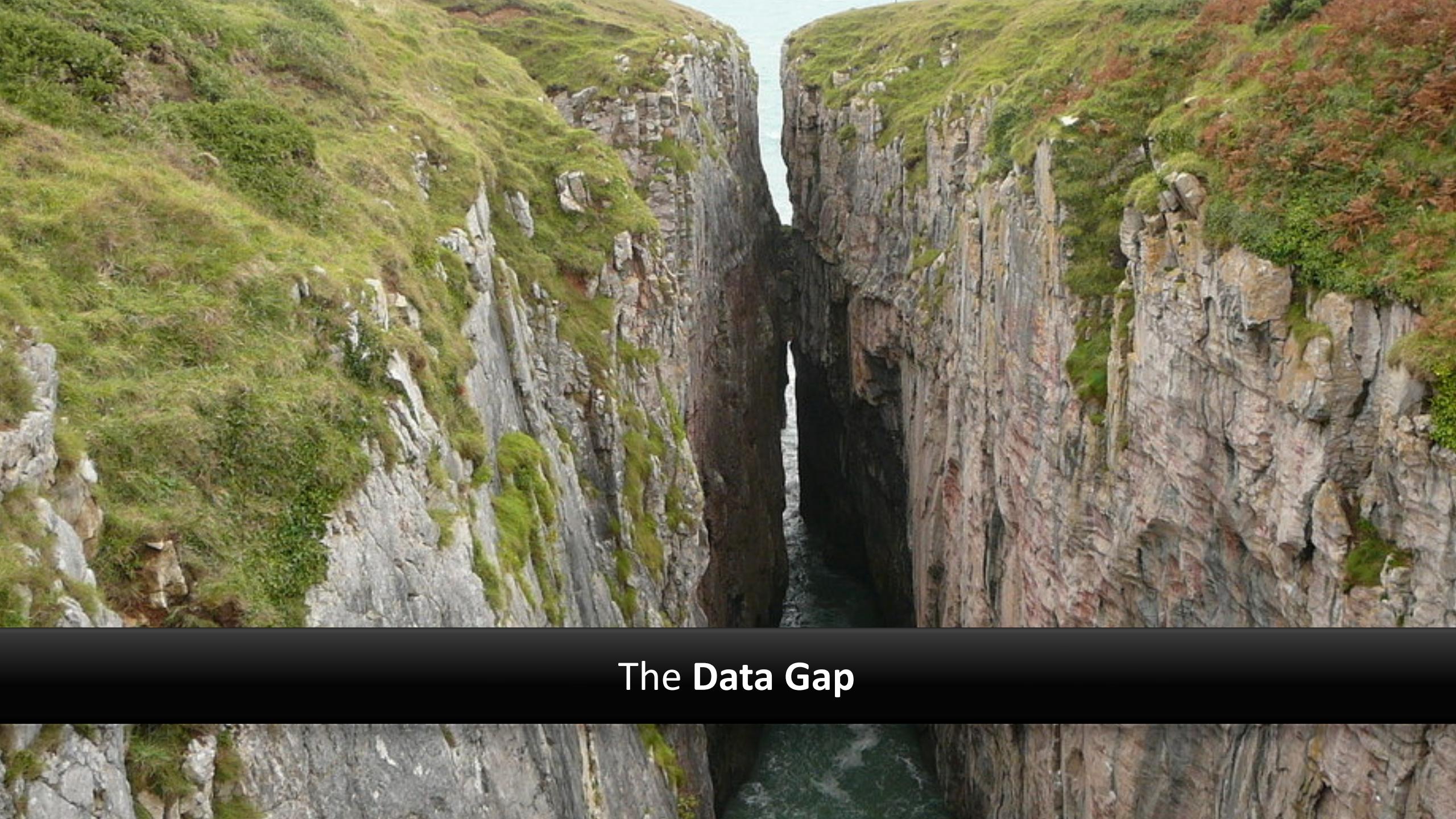


LEGALLY, COMPANIES ARE REQUIRED TO KNOW YOUR CUSTOMER (KYC)



MOST OF THE 'NEW' USE CASES, SUCH AS AI-BASED ASSISTANTS NEED TO PERSONALISE

But if we want to go beyond profitability, we encounter...

A photograph of a narrow, deep sea cave. The walls are made of dark, layered rock, showing vertical sedimentary structures. The top of the cliff is covered in green moss and some brown autumn-colored vegetation. The entrance to the cave is very narrow, creating a sense of depth and confinement. The water inside the cave is dark and still.

# The Data Gap

# What should a FinRecSys dataset contain?



- If we think a bit about what we would want ideally in a financial recommender dataset, at its core we need to have information about assets (to recommend) and customers (looking to invest)



Rich information regarding  
a range of financial assets  
to recommend



Good quality market  
pricing information for  
those assets



Information about a set of  
customers



Historical interaction data  
for those customers

Item Information

User Information

# The Data Gap



- Currently, it is very difficult to undertake any meaningful research into personalised financial recommendation unless you are based in a company with real customers
- This is because companies are actively disincentivised from releasing their data:
  - Concerns over customer data privacy and legal implications
  - The data potentially has commercial value
  - Data is often licenced, and so its unclear what can be shared with third parties
  - The data is usually not in a 'clean' form that is ready to be analysed and so work would be needed
  - What would the company get from doing so?

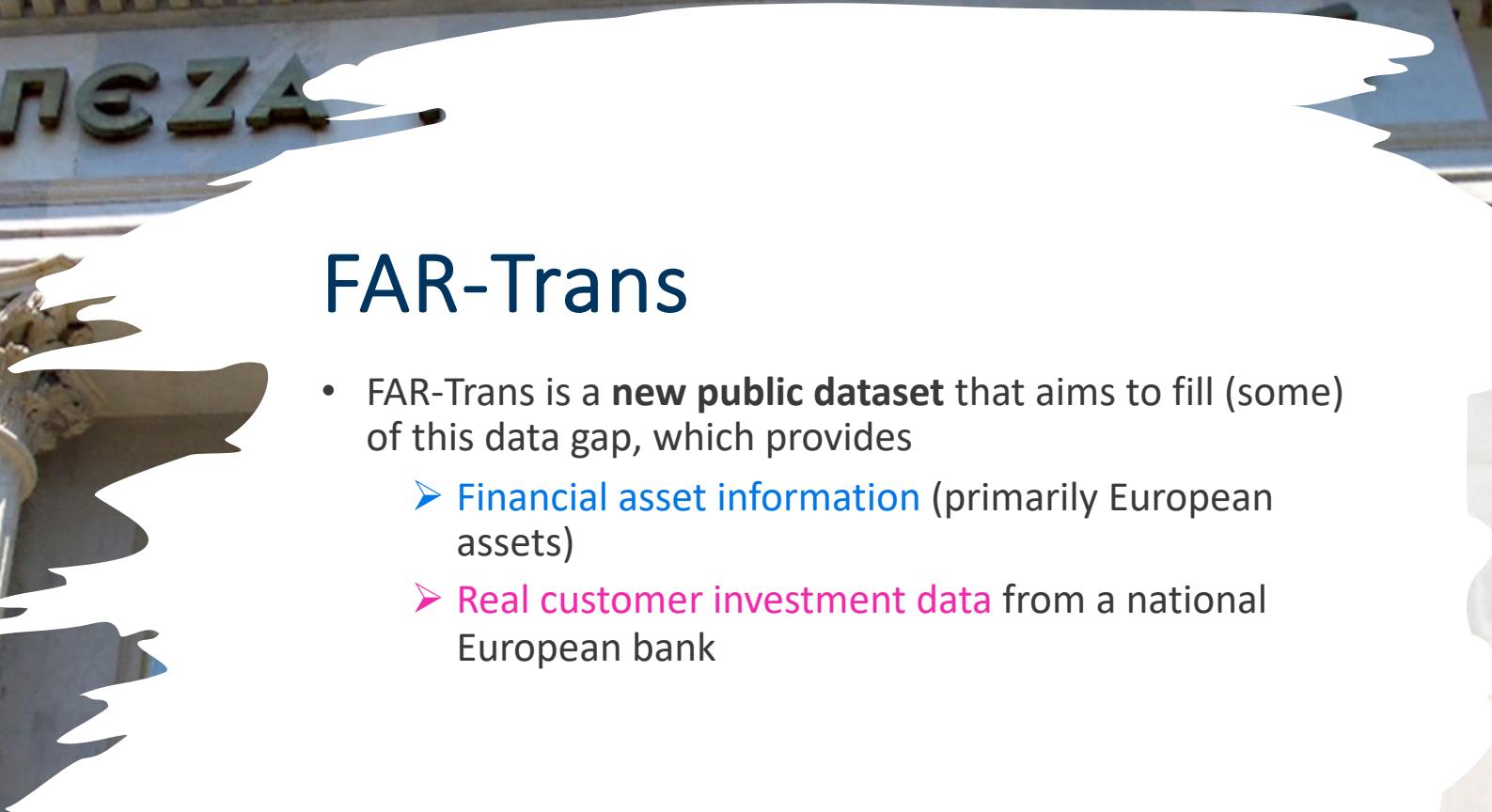
As a result, up till now, no such dataset existed

The closest the community has is a 2011-13 dataset from ObjectWay Financial Software, but this lacks the item information [1]

[1] Musto et al. *Financial product recommendation through case-based reasoning and diversification techniques*. In Poster Proceedings of the 8th ACM Conference on Recommender Systems (RecSys 2014)



Introducing FAR-Trans



ΕΘΝΙΚΗ ΤΡΑΠΕΖΑ

## FAR-Trans

- FAR-Trans is a **new public dataset** that aims to fill (some) of this data gap, which provides
  - Financial asset information (primarily European assets)
  - Real customer investment data from a national European bank

# High-level Statistics



*Dataset Period: January 2018 to November 2022 (~5 years)*

## Assets

### (Mostly) Athens Stock Exchange

|                         |         |
|-------------------------|---------|
| Unique Assets           | 806     |
| Assets with Investments | 321     |
| Price Data Points       | 703,303 |
| Average Return / Period | 37.16%  |
| % Profitable Assets     | 54.28%  |

## Customers

### Bank Investors

|                             |                   |
|-----------------------------|-------------------|
| Unique Customers            | 29,090            |
| Transactions (unique)       | 388,049 (154,103) |
| Acquisitions (unique)       | 228,913 (89,884)  |
| Average Return of Customers | 22.89%            |
| % Customers with Profits    | 54.56%            |

# Dataset Composition



## Name

|                          |
|--------------------------|
| asset_information.csv    |
| close_prices.csv         |
| customer_information.csv |
| limit_prices.csv         |
| markets.csv              |
| questionnaires.csv       |
| transactions.csv         |
| README.txt               |

- Information about the assets (ISIN, Name, etc).
- Contains the daily close prices for every asset in the dataset.
- Customer (anonymized) identifiers and categorization.
- Asset pricing metadata
- Market metadata
- Questions for the risk MiFiD questionnaire (answers not provided)
- Investment transactions by customers.

# FAR-Trans: Assets



- With regard to assets, we provide **asset descriptor information** and **market pricing (closing prices only)** as well as **where the asset was being traded**.  
➤ Note: Pricing data has undergone cleaning – see the paper for more details

| HSBC Global Investment Funds - Indian Equity EC |              |
|---|--------------|
| ISIN  | LU0164858028 |
| Short Name                                      | UNK_574      |
| Category  | MTF          |
| Sub-Category                                    | Equity       |
| Exchange ID                                     | XLUX         |
| Last Updated                                    | 03/01/2018   |

*asset\_information.csv*

Some assets may also have  
a sector and industry label,  
where we could find that  
information online

| ISIN         | Date       | Price (€)   |
|--------------|------------|-------------|
| LU0164858028 | 03/01/2018 | 162.4004235 |
| LU0164858028 | 04/01/2018 | 165.0849151 |
| LU0164858028 | 05/01/2018 | 165.7226286 |
| LU0164858028 | 08/01/2018 | 166.7345024 |
| LU0164858028 | 09/01/2018 | 167.3428186 |
| LU0164858028 | 10/01/2018 | 167.4762868 |

*close\_prices.csv*

|               |                                  |
|---------------|----------------------------------|
| Exchange ID   | XLUX                             |
| Market ID     | CCLX                             |
| Name          | Finesti S.A.                     |
| Description   | Finesti is a Luxembourg-based... |
| Country       | Luxembourg                       |
| Trading Days  | Mon,Tue,Wed,Thu,Fri              |
| Trading Hours | 08:00-16:30                      |
| Market Class  | Public Securities                |

*markets.csv*

# FAR-Trans: Customers



- For each customer we provide programmatically assigned **categories** for those customers as well as their **investment history** with the bank

|                      |                        |
|----------------------|------------------------|
| 4961594D18DCF4D55F6D |                        |
| Type                 | Mass                   |
| Risk Level           | Predicted Conservative |
| Investment Capacity  | Predicted_CAP_LT30K    |
| Last Questionnaire   | 01/01/2000             |
| Last Updated         | 02/01/2018             |

*customer\_information.csv*

| Customer ID | ISIN         | Transaction ID | Type | Date       | Value (€) | Units    | Channel | Market ID |
|-------------|--------------|----------------|------|------------|-----------|----------|---------|-----------|
| 4961594...  | GRF000037009 | -8365          | Buy  | 02/01/2018 | 8800      | 2525.983 | Branch  | AEDAK     |
| 4961594...  | GRF000037009 | 11584908       | Sell | 05/10/2022 | 9375      | 2525.983 | Branch  | AEDAK     |
| 4961594...  | GRF000031002 | -31645         | Buy  | 02/01/2018 | 21067     | 1738.689 | Branch  | AEDAK     |
| 4961594...  | GRF000031002 | 11584904       | Sell | 05/10/2022 | 24504     | 1738.689 | Branch  | AEDAK     |
| 4961594...  | GRF000025004 | -35710         | Buy  | 02/01/2018 | 5050      | 750.107  | Branch  | AEDAK     |
| 4961594...  | GRF000025004 | 11584901       | Sell | 05/10/2022 | 87        | 11.077   | Branch  | AEDAK     |
| 4961594...  | GRF000025004 | 11584898       | Sell | 05/10/2022 | 5828      | 739.03   | Branch  | AEDAK     |

*transactions.csv*

**Note:** If a customer held assets before the start of our dataset (before 2018) we add synthetic 'buy' transactions at the end of day 1 for those customers, this is to avoid instances of a customer selling an asset that you never saw them purchase

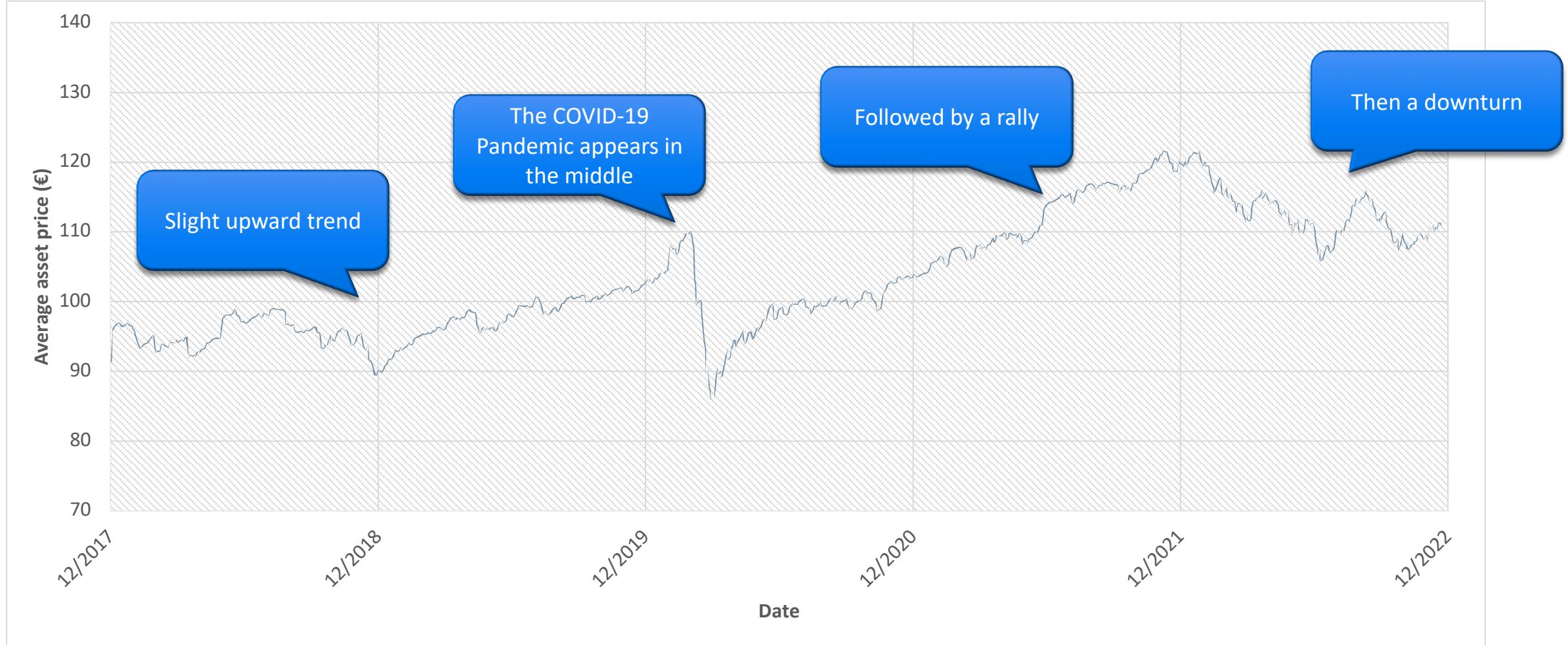
Customer categories were provided by the bank, see the README file for definitions of the categories

A close-up photograph of a man with light brown hair and a well-groomed reddish-brown beard. He is wearing black-rimmed glasses and a dark grey t-shirt. He is looking directly at a laptop screen, which is visible in the lower half of the frame. The background is blurred, showing another person's arm and shoulder, suggesting a collaborative environment like an office or a study group.

# Dataset Analysis

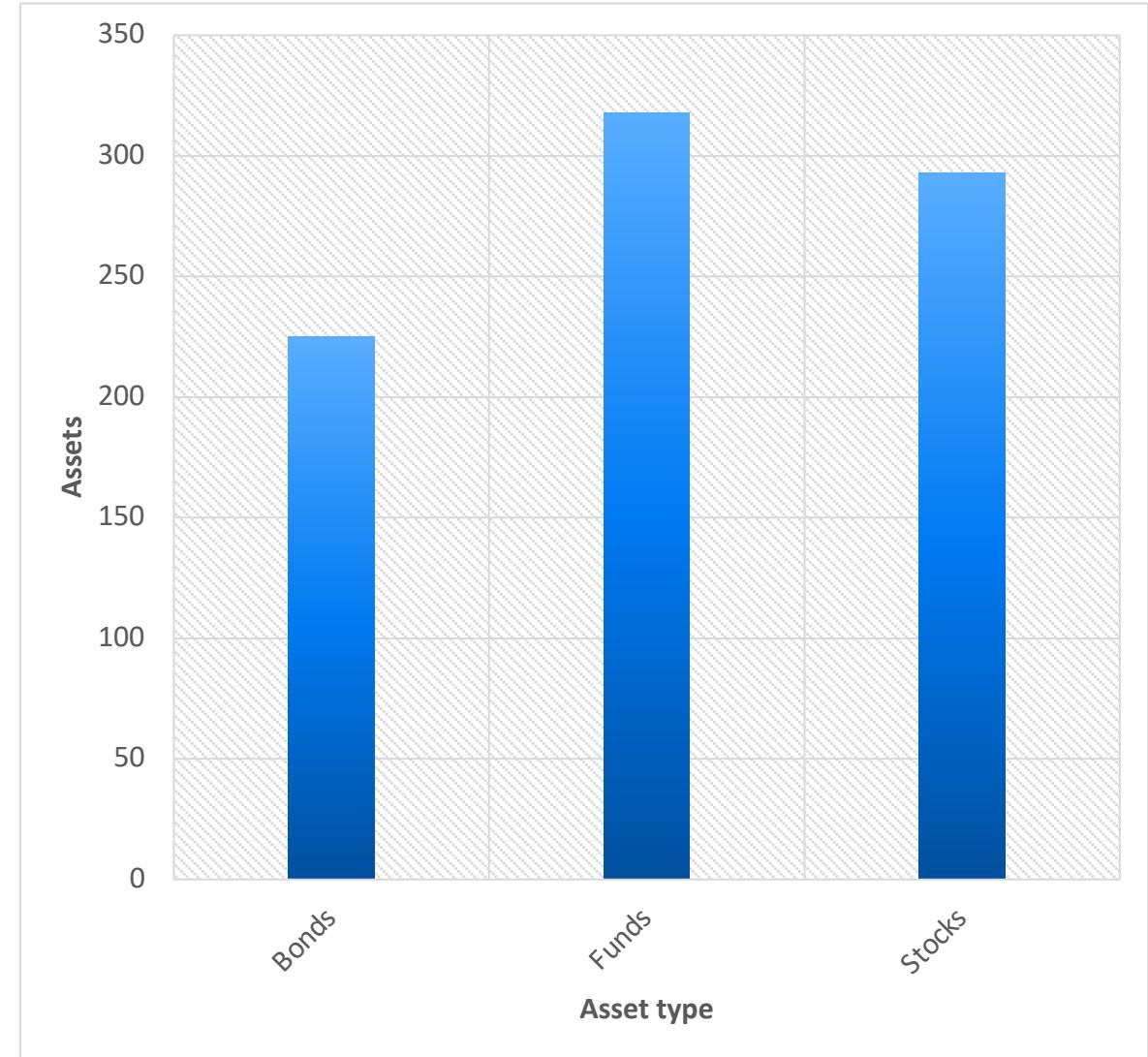
# Market Performance

- January 2018 to November 2022 (~5 years)

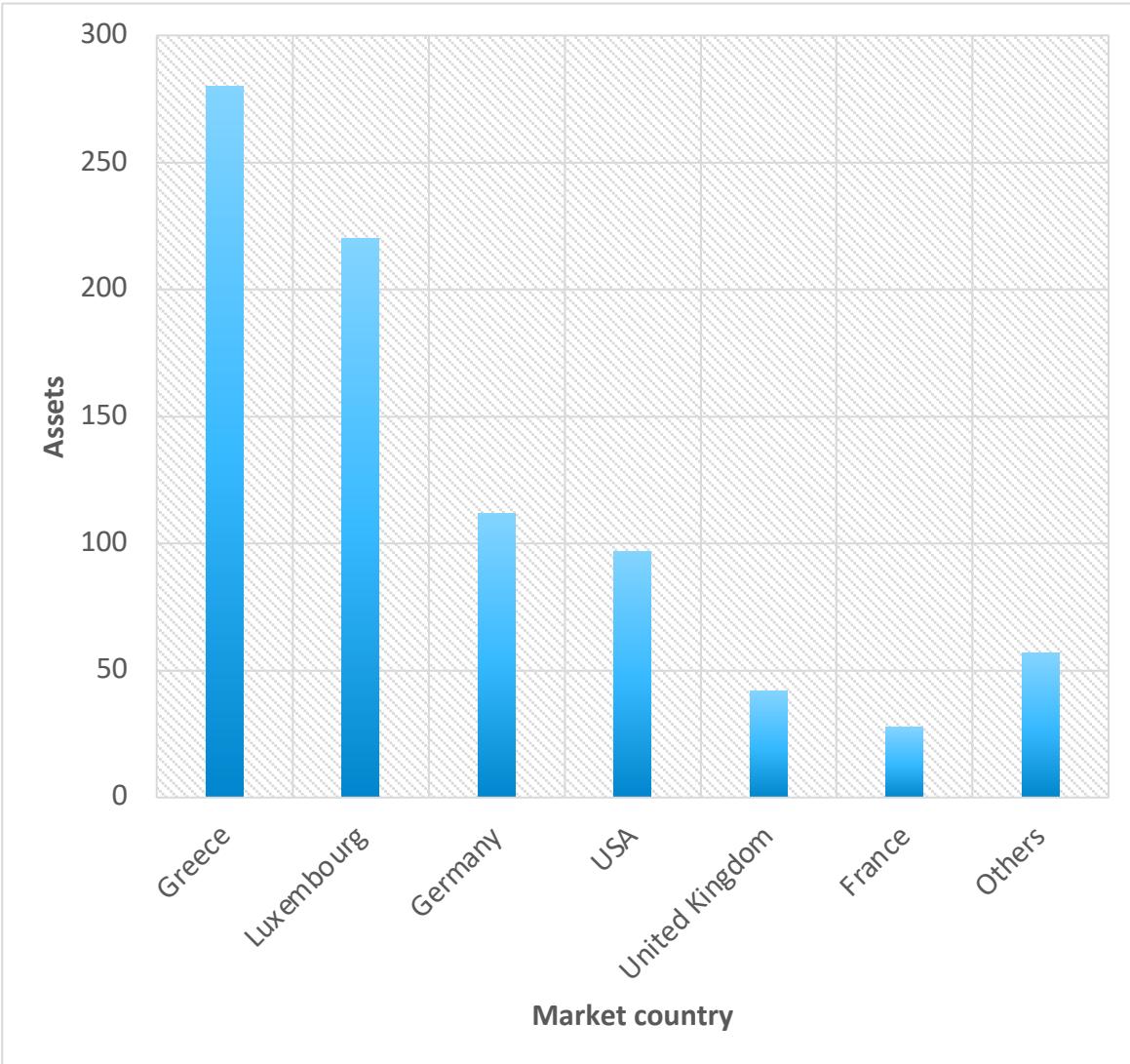


# What is Asset Type Distribution?

- The dataset provides fundamental, pricing and transactions for 806 assets
- A mixture of **Stocks**, **Bonds** and **Mutual Funds**

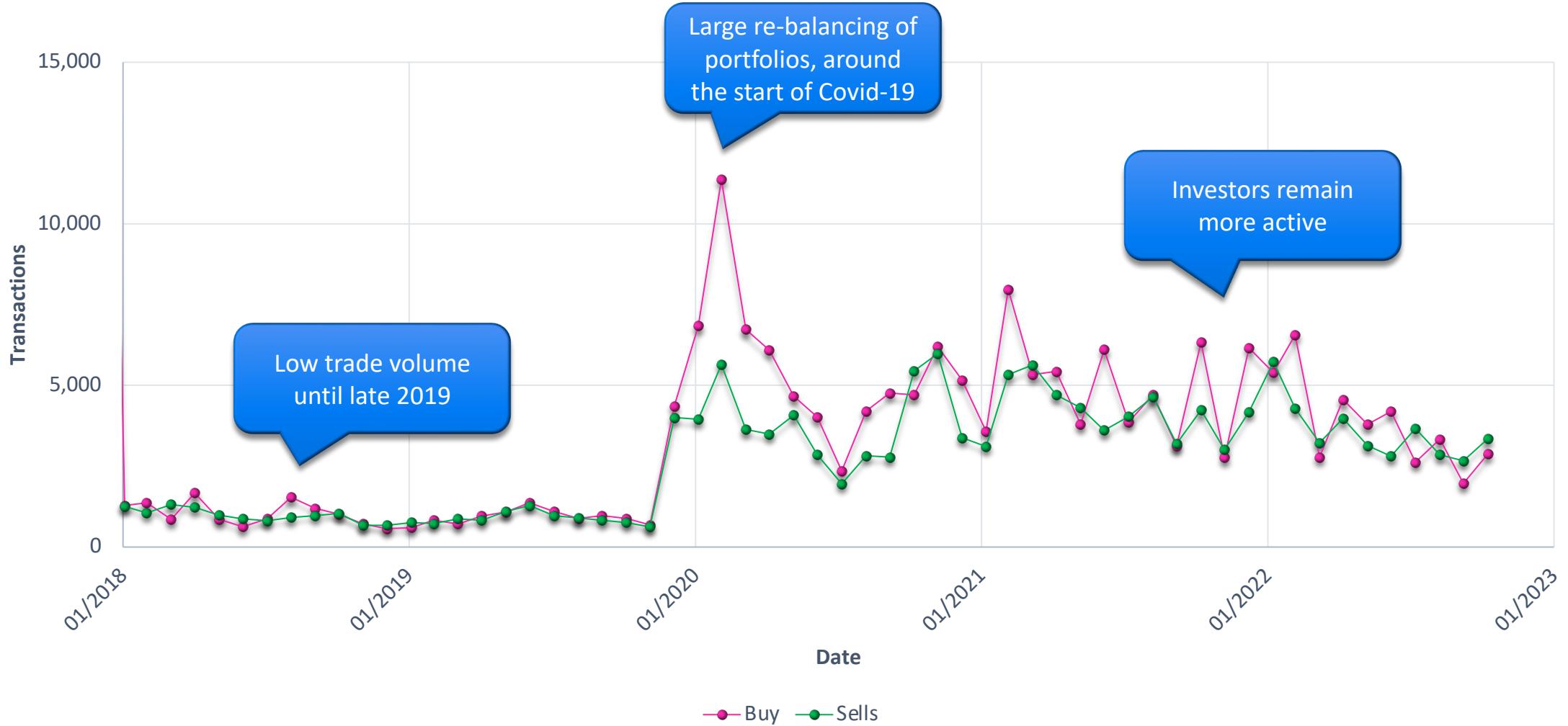


# What is Asset Listing Distribution?

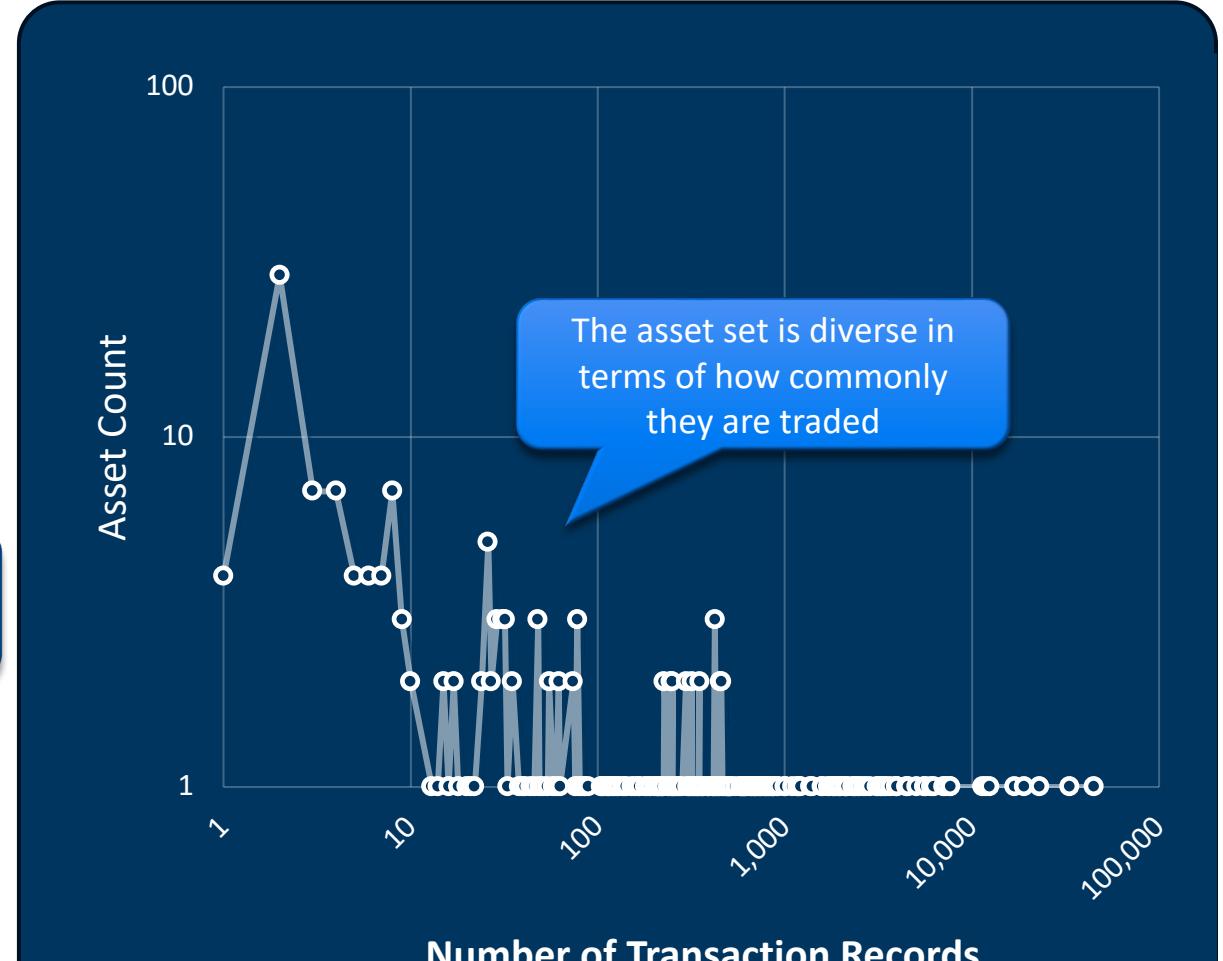
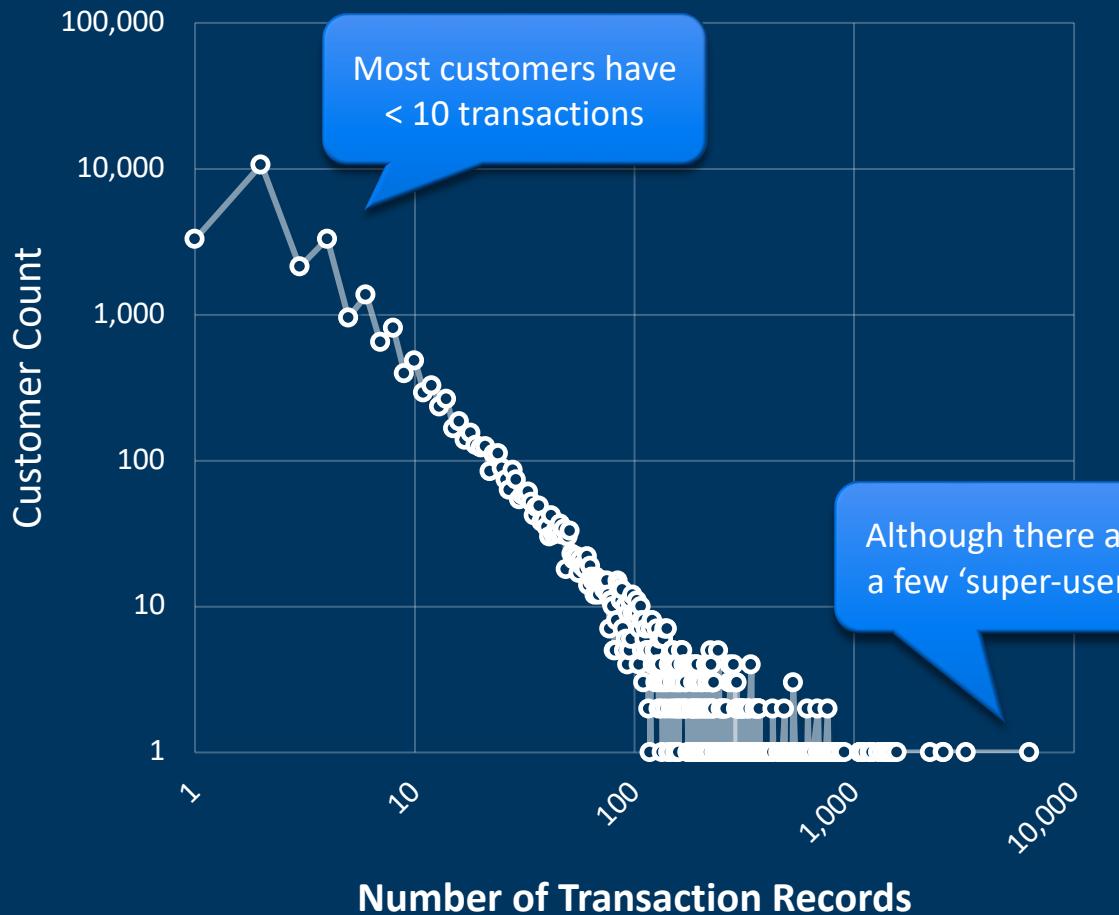


- As this is from a European Bank, most of the assets are listed on European Exchanges
- Primarily, in [Greece](#), [Luxemburg](#) and [Germany](#), with some [U.S. assets](#) (usually big companies)

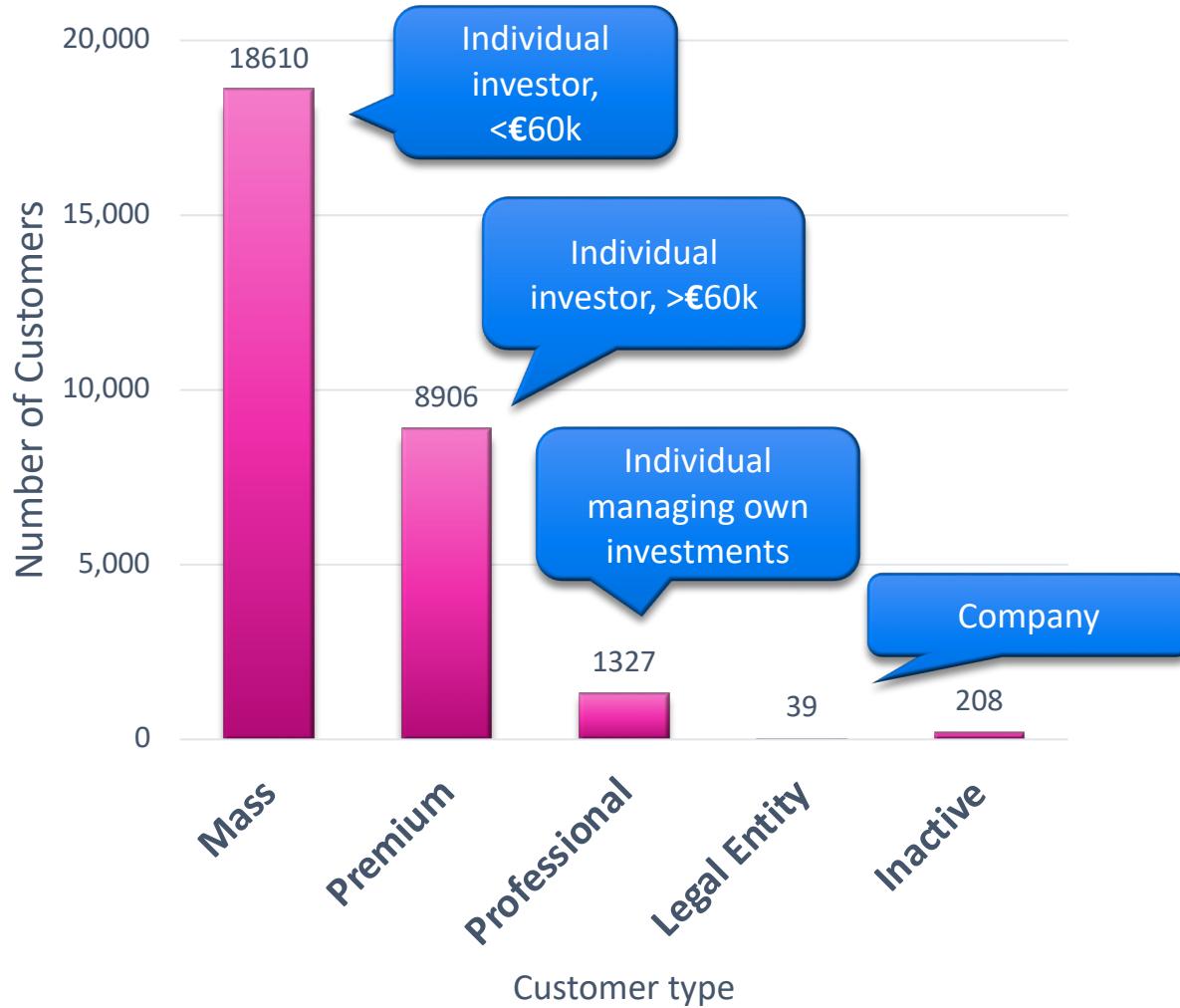
# How does transaction volume vary over time?



# What properties do our Transactions have?



# Who are our Customers?



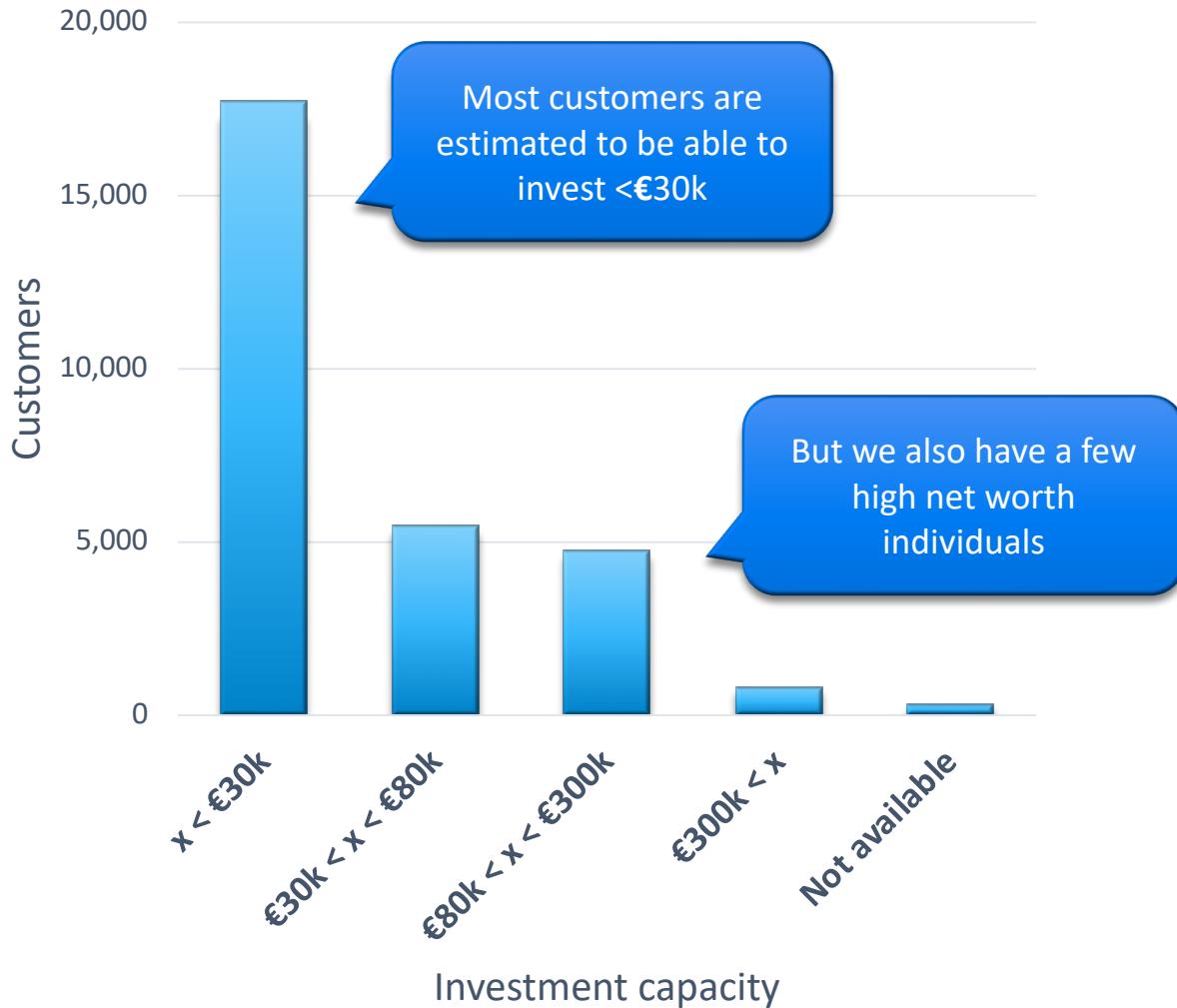
- The Bank provides multiple categorizations of the customers
  - This can be from the questionnaire the customer enters when opening the account
  - ... or based on analysis of their past investments and account activity
- Categories
  - Customer Type

# Who are our Customers?



- The Bank provides multiple categorizations of the customers
  - This can be from the questionnaire the customer enters when opening the account
  - ... or based on analysis of their past investments and account activity
- Categories
  - Customer Type
  - Risk Profile

# Who are our Customers?



- The Bank provides multiple **categorizations** of the customers
  - This can be from the questionnaire the customer enters when opening the account
  - ... or based on analysis of their past investments and account activity
- Categories
  - Customer Type
  - Risk Profile
  - **Investment Capacity**

# Recommendation Baselines

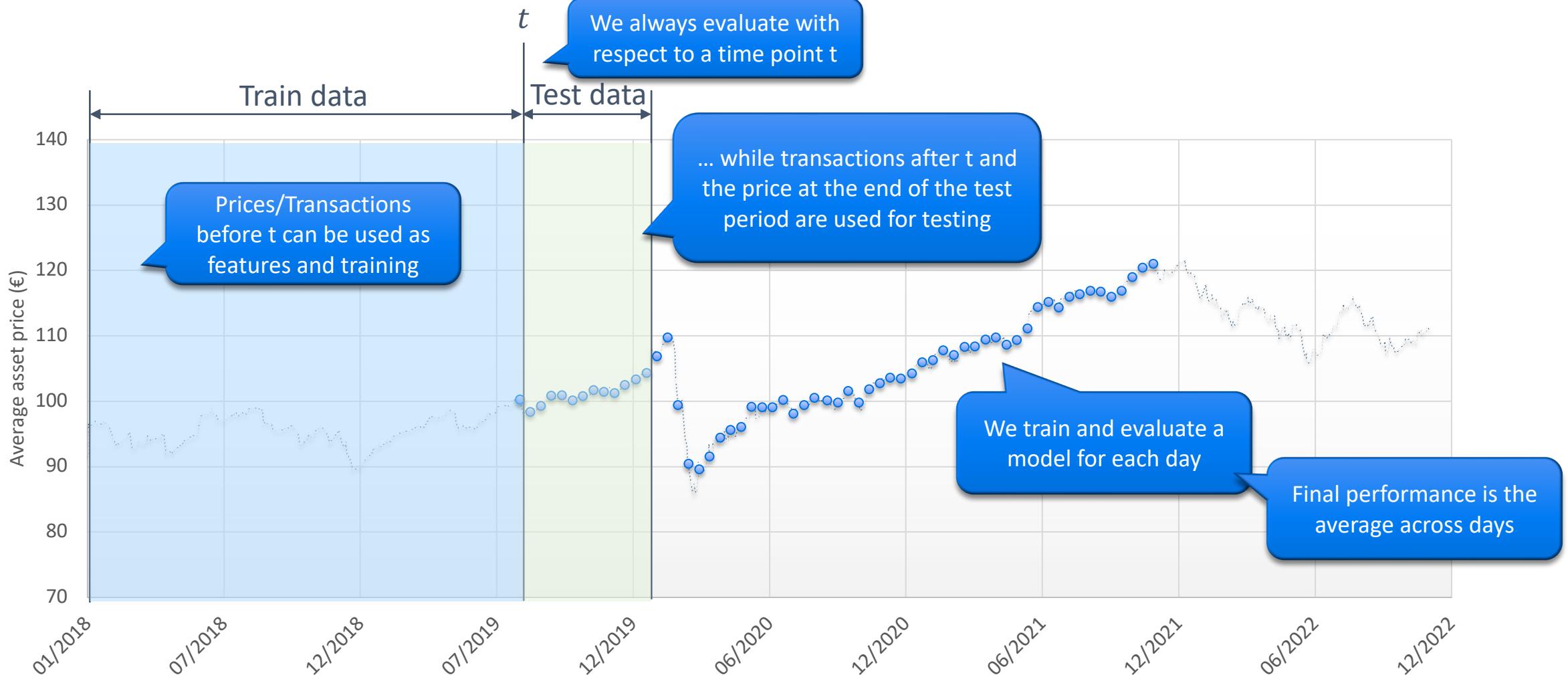
# Evaluation Guidance

---



- For any dataset, it is important to provide some guidance on how that dataset should be used
  - ... there are always pit-falls to be aware of
- This dataset was designed to provide a means to evaluate financial asset recommendation systems, both those that are based on **profitability prediction** and those that **personalise based on customer transactions**
  - If evaluating profitability prediction, reporting the **RoI** or **Sharpe Ratio** for the top-n results is recommended
    - Note: an assumption here is that the user will invest in all top-n assets though!
  - If evaluating recommendation performance, a ranking metric such as **nDCG@n** can be used here, where an asset is considered a good recommendation if the customer invested in it between the date of prediction and the end of the test period
    - Note: this assumes that the customer is satisfied with the assets they invested in (which is not a good assumption)

# ... A quick reminder of experimental setup



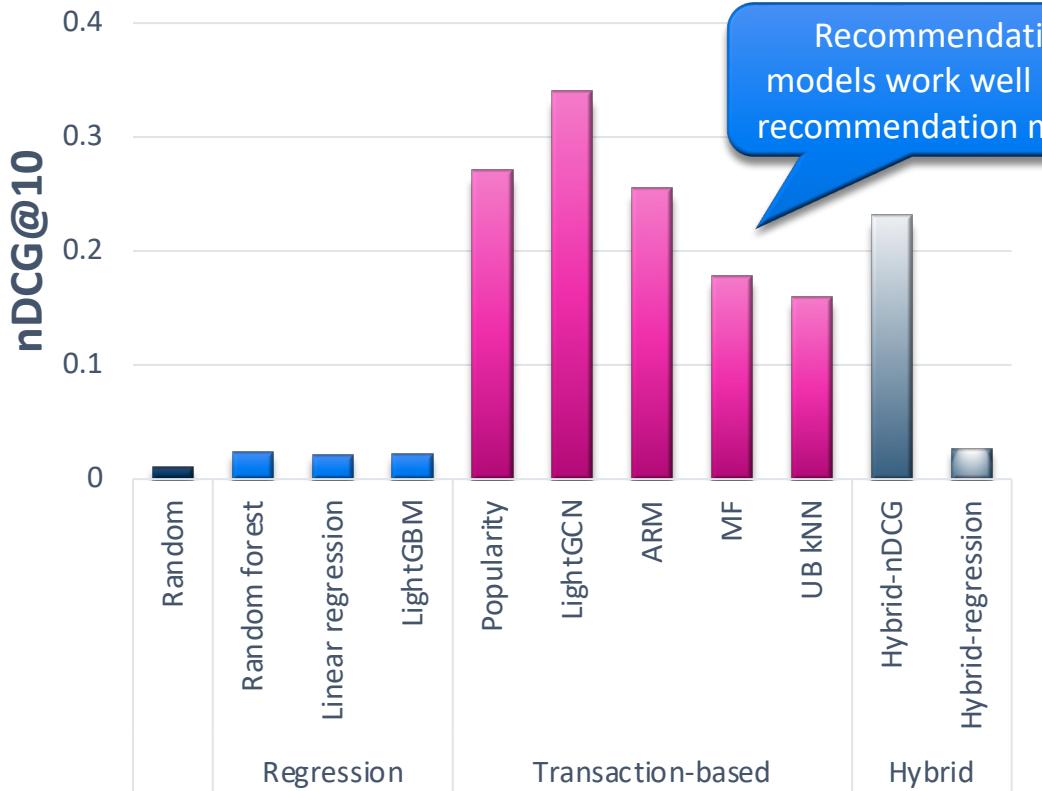
**Do not evaluate only on a single day**, performance will vary greatly across days!

# Have Reasonable Expectations

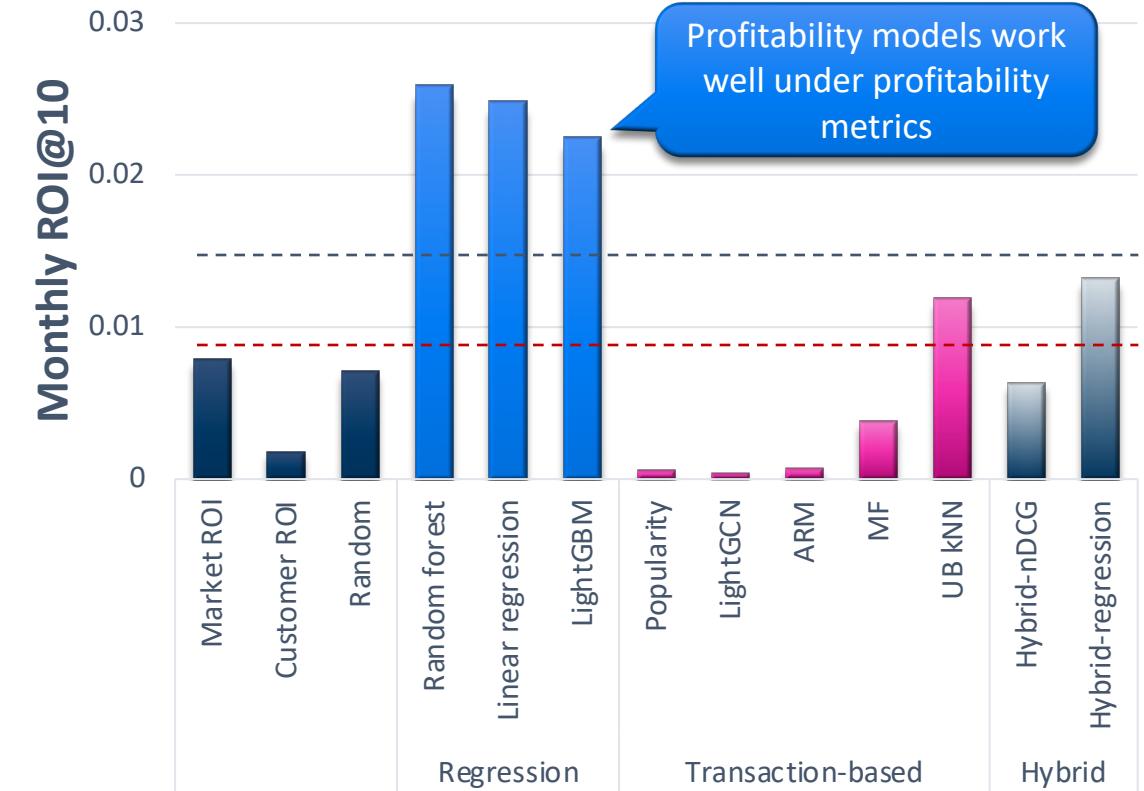


- You should always consider what it is that your model is trying to do, and hence evaluate accordingly – **profitability and customer preference do not correlate strongly**

## Transaction-based evaluation



## Profitability-based evaluation



# Understand the Limitations



- I said earlier that FAR-Trans is a solution to **some** of the data gap – but it is not a perfect solution!
- Consider the following
  - There is no dividend or interest payment information provided for the assets
  - There is no voting rights information
  - It only covers a small number of stocks, and only a 5-year period
  - There is no information on rates at which money could be borrowed by customers (for leverage)
  - We cannot tell if customers were satisfied with their investments
  - We cannot tell why a customer invested in an asset
  - We don't know how long a customer aims hold an asset for (if they have not sold yet)

**As a community, we have a long way to go before we have a really good understanding on how to evaluate financial recommendation systems**

... and go beyond 'line-goes-up'



Download FAR-Trans  
for free!

<https://researchdata.gla.ac.uk/1658/>

# Questions?



Dr. Richard McCreadie

Real-time IR, Machine Learning, Big Data Stream Processing, Evaluation

✉ [Richard.Mccreadie@glasgow.ac.uk](mailto:Richard.Mccreadie@glasgow.ac.uk)



Dr. Javier Sanz-Cruzado

University of Glasgow – Financial Informatics

✉ [javier.sanz-cruzadopuig@glasgow.ac.uk](mailto:javier.sanz-cruzadopuig@glasgow.ac.uk)

