

# Measuring the Price of Data in Commercial Data Marketplaces

Santiago Andrés Azcoitia

IMDEA Networks Institute, Universidad Carlos III de Madrid

Costas Iordanou, Cyprus University of Technology

Nikolaos Laoutaris, IMDEA Networks Institute

[Developing the  
Science of Networks]

# A quiz

How valuable is this?



How much is it?



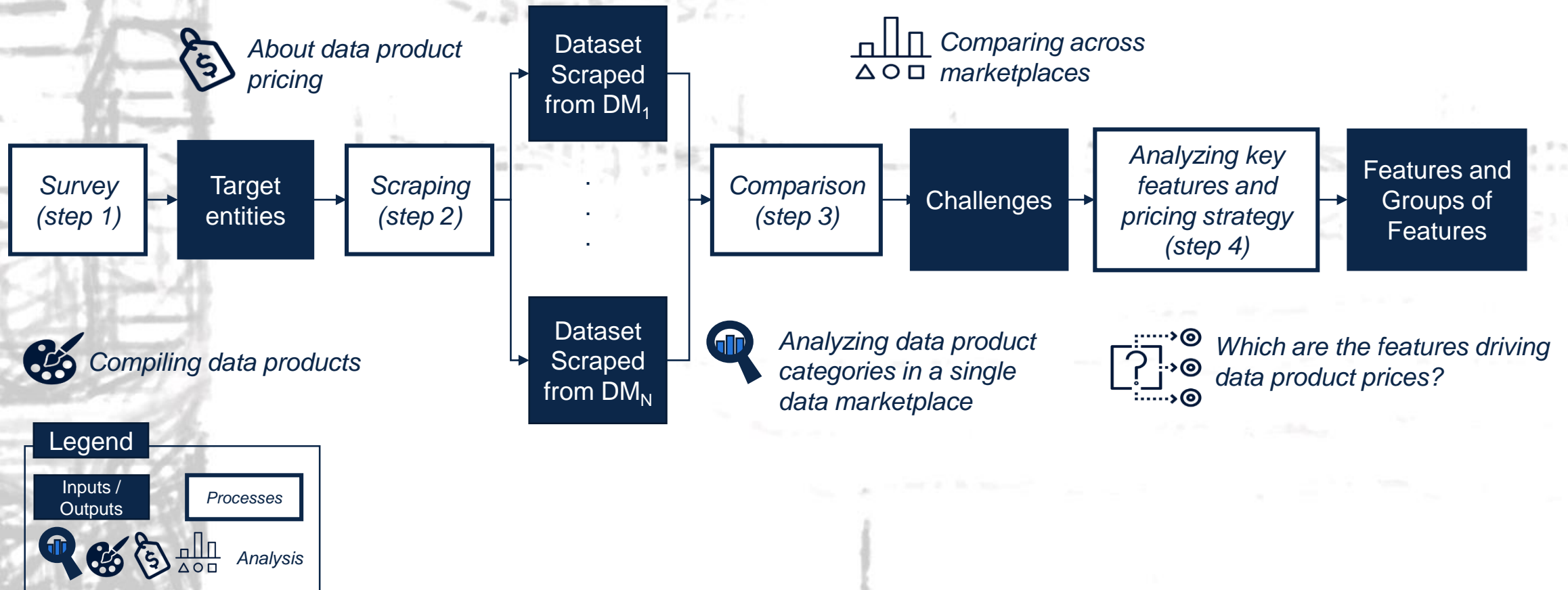


# A quiz

And what's the price of this?

id	Self-emp-not-inc	educ	educnum	married-civ-spouse	exec-managerial	relation	race	sex	age	age	age	country
38	Private	215646	HS-grad	9 Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40	United-States
53	Private	234721	11th	7 Married-civ-spouse	Handlers-cleaners	Husband	Black	Male	0	0	40	United-States
28	Private	338409	Bachelors	13 Married-civ-spouse	Prof-specialty	Wife	Black	Female	0	0	40	Cuba
37	Private	284582	Masters	14 Married-civ-spouse	Exec-managerial	Wife	White	Female	0	0	40	United-States
49	Private	160187	9th	5 Married-spouse-absent	Other-service	Not-in-family	Black	Female	0	0	16	Jamaica
52	Self-emp-not-inc	209642	HS-grad	9 Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	45	United-States
31	Private	45781	Masters	14 Never-married	Prof-specialty	Not-in-family	White	Female	14084	0	50	United-States
42	Private	159449	Bachelors	13 Married-civ-spouse	Exec-managerial	Husband	White	Male	5178	0	40	United-States
37	Private	280464	Some-college	10 Married-civ-spouse	Exec-managerial	Husband	Black	Male	0	0	80	United-States
30	State-gov	141297	Bachelors	13 Married-civ-spouse	Prof-specialty	Husband	Asian-Pac-Islander	Male	0	0	40	India
23	Private	122272	Bachelors	13 Never-married	Adm-clerical	Own-child	White	Female	0	0	30	United-States
32	Private	205019	Assoc-acdm	12 Never-married	Sales	Not-in-family	Black	Male	0	0	50	United-States
40	Private	121772	Assoc-voc	11 Married-civ-spouse	Craft-repair	Husband	Asian-Pac-Islander	Male	0	0	40	?
34	Private	245487	7th-8th	4 Married-civ-spouse	Transport-moving	Husband	Amer-Indian-Eskimo	Male	0	0	45	Mexico
25	Self-emp-not-inc	176756	HS-grad	9 Never-married	Farming-fishing	Own-child	White	Male	0	0	35	United-States
32	Private	186824	HS-grad	9 Never-married	Machine-op-inspct	Unmarried	White	Male	0	0	40	United-States
38	Private	28887	11th	7 Married-civ-spouse	Sales	Husband	White	Male	0	0	50	United-States
43	Self-emp-not-inc	292175	Masters	14 Divorced	Exec-managerial	Unmarried	White	Female	0	0	45	United-States
40	Private	193524	Doctorate	16 Married-civ-spouse	Prof-specialty	Husband	White	Male	0	0	60	United-States
54	Private	302146	HS-grad	9 Separated	Other-service	Unmarried	Black	Female	0	0	20	United-States
35	Federal-gov	76845	9th	5 Married-civ-spouse	Farming-fishing	Husband	Black	Male	0	0	40	United-States
43	Private	117037	11th	7 Married-civ-spouse	Transport-moving	Husband	White	Male	0	2042	40	United-States
59	Private	109015	HS-grad	9 Divorced	Tech-support	Unmarried	White	Female	0	0	40	United-States

# So, what is the price of data in the B2B market? What are the features that are driving the prices of data products?





**We found 15 scrapable DMs + 30 individual providers that allowed us to collect information about 215,075 data products from 2,115 data sellers**

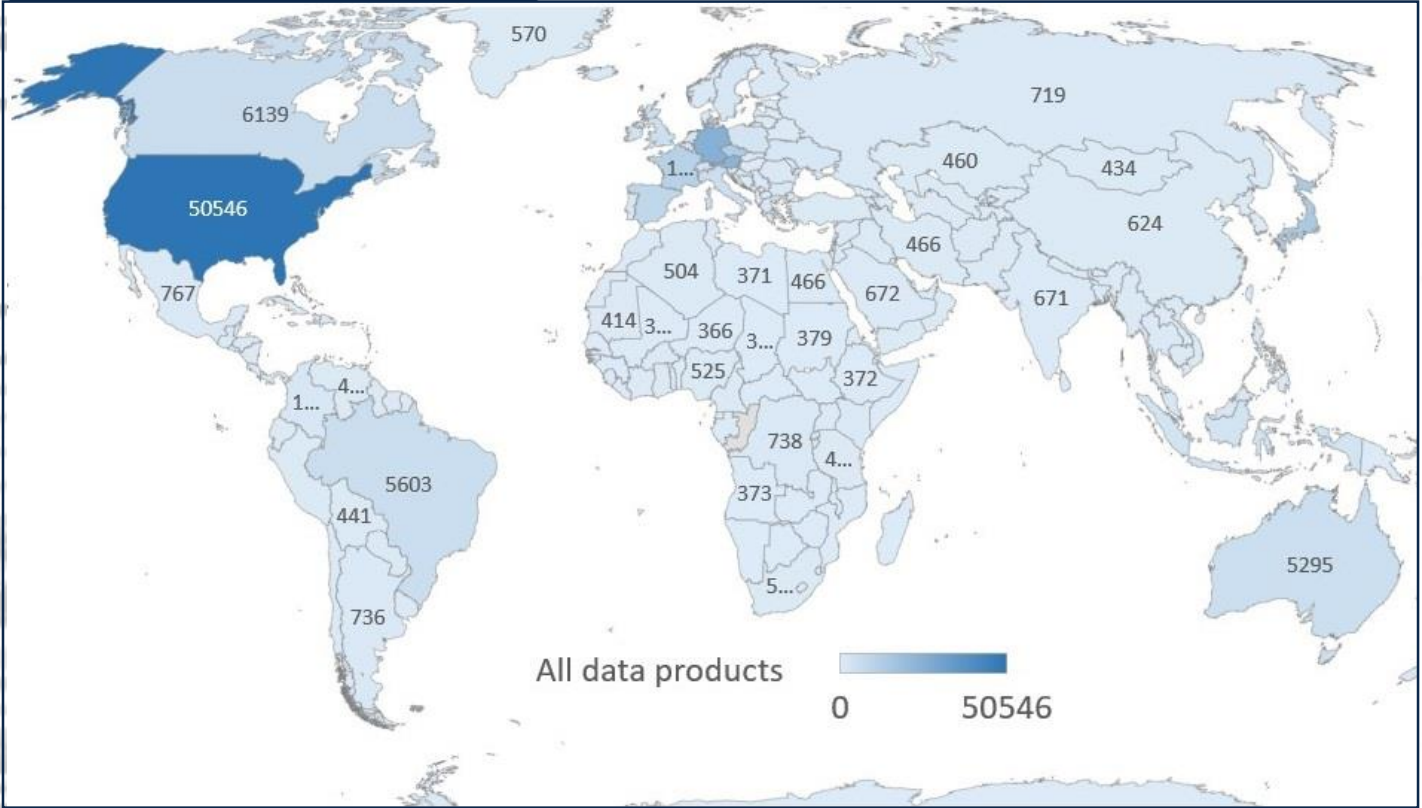
Marketplace	#Data Products	#Paid products	#Sellers
Advaneo	198,743	1	N/A
AWS	4,263	2,674	262
DataRade	1,592	1,592	1,262
Snowflake	889	889	200
Knoema	158	158	142
DAWEX	160	160	79
Carto spatial data catalog	8,182	5,283	42
Crunchbase Marketplace	9	9	15
Veracity	115	95	38
Refinitiv	187	187	76
Liveramp	N/A	N/A	135
LOTAME	16	16	126
TheTradeDesk	N/A	N/A	122
Quandl	N/A	N/A	49
Factset	N/A	N/A	84
Other data providers	777	775	30





14,472 (7%) of data products did not inform about scope and 1,177 claim to have global scope

Nº products by country



Top countries by paid products

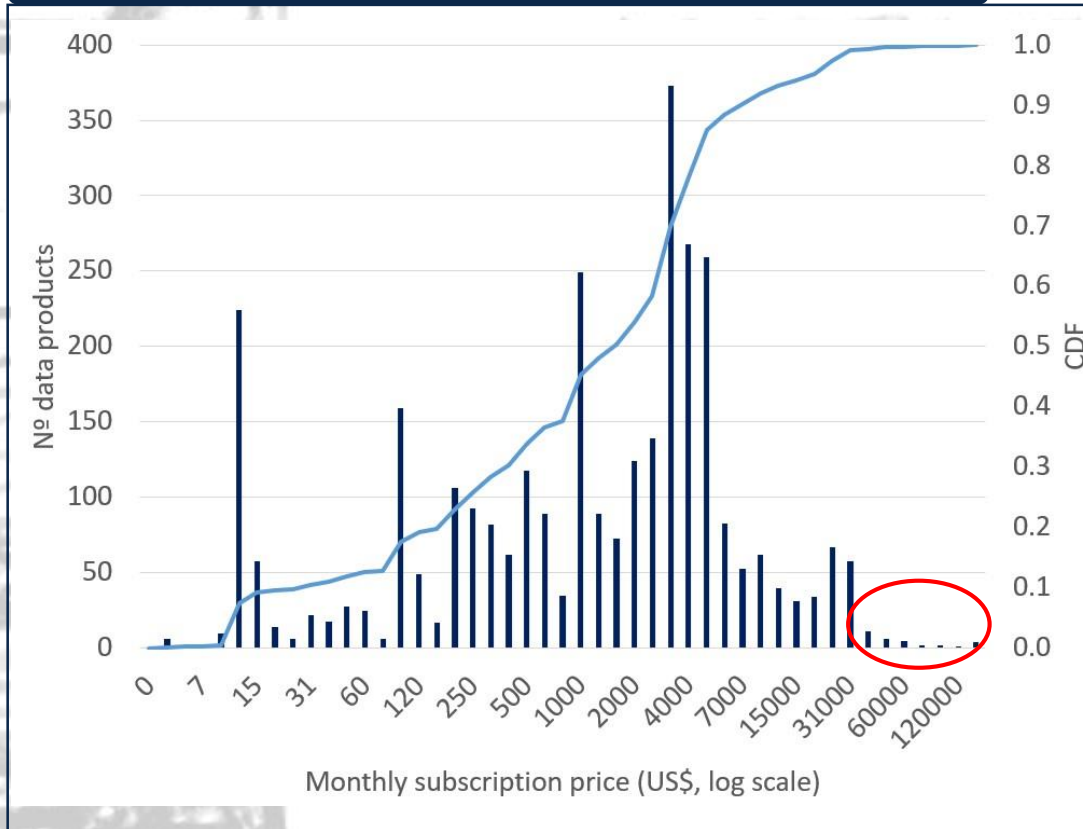
Country	% paid products
United States	30%
Canada	9.3%
United Kingdom	9.2%
Germany	7.6%
France	7.4%
Spain	7.1%

Only 11,823 data products are paid, and 4,162 from 443 distinct providers of them disclose information about their prices

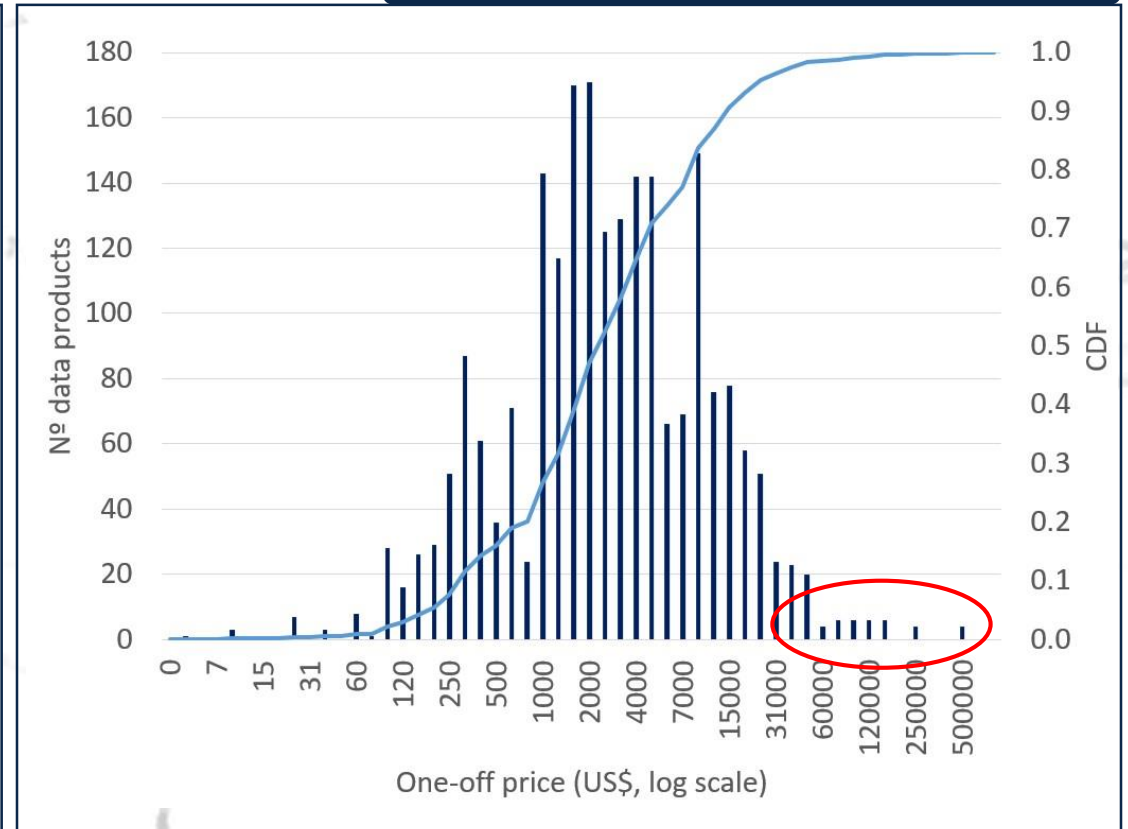


# Data sells at an immensely wide range of prices

## Subscription-based data product prices



## One-off data product prices

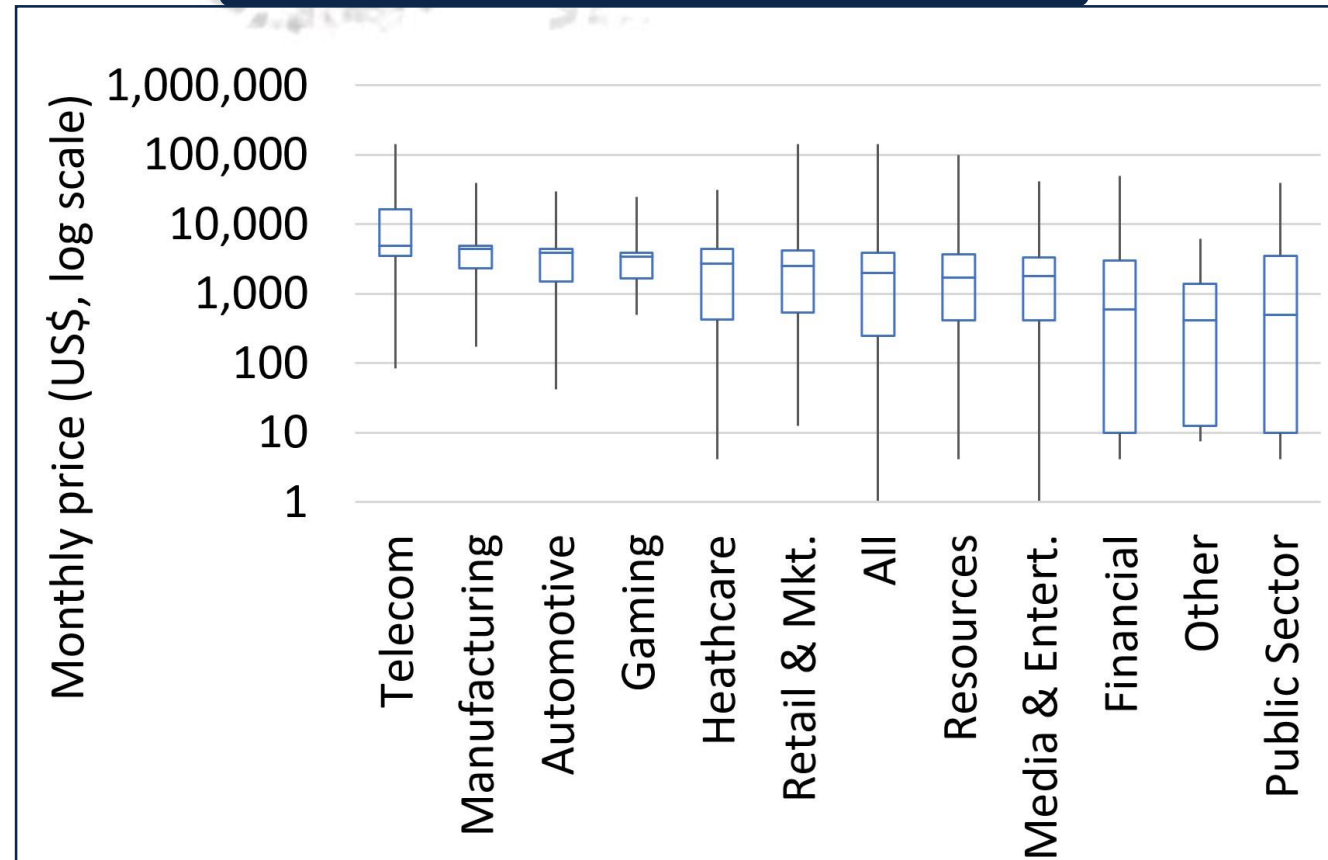


Most valuable data products in the long tail are related to marketing. We will come back to them later.



Focusing on AWS marketplace, we observe that prices depend heavily on the category, which in this DMs resemble “industries”


Box-plot data product prices by industry







Cross DMs analysis is more challenging, since DMs i) provide different detail, and ii) use different categorisation and criteria to assign categories to products



### Consumer transaction and payment data

Provided By: [Alliant](#)

Alliant consumer transaction and payment data, sourced from Alliant's proprietary cooperative database of billions of transactions. Examples: Credit card transactions, dollar and number broken out by block group Alliant's proprietary payment score metric

[Continue to subscribe](#)

[Product offers](#) | [Overview](#) | [Usage](#) | [Support](#)

#### Product offers

The following offers are available for this product. Choose an offer to view the pricing and access duration options for the offer. Select an offer and continue to subscribe. Your subscription begins on the date that your request is approved by the provider. Additional taxes or fees might apply.

##### Public offer

Payment schedule: Upfront payment | Offer auto-renewal: Supported

☒ \$5,500 for 1 month

☐ \$35,000 for 12 months

#### Overview

Consumer transaction and payment data, aggregated at the geographic block group level. Data is sourced from Alliant's proprietary cooperative database which aggregates hundreds of leading DTC brand's 1st party detailed transactional CRM data. Deterministic view into U.S. geographic block groups transaction and payment data. Example data points include: -total number and dollar amount of credit card transactions by block group in last 5 years -total number and dollar amount of write offs by block group in last 5 years -Alliant's proprietary payment score metric (grouped 1-20)

Overview one sheet: [https://info.alliantinsight.com/hubfs/Downloadable%20Content%20Alliant%20AWS\\_Geo\\_Performance.pdf](https://info.alliantinsight.com/hubfs/Downloadable%20Content%20Alliant%20AWS_Geo_Performance.pdf)

Provided By  
[Alliant](#)





Cross DMs analysis is more challenging, since DMs i) provide different detail, and ii) use different categorisation and criteria to assign categories to products

Explore / Consumer Transaction Data

# Yodlee's 4M Panel (US Consumer Transaction Data, de-identified)

Starts at \$400,000 / year

A dataset by [Envestnet](#) | [Yodlee](#)

	SECONDARY_MERCHANT_NAME	PRIMARY_MERCHANT_NAME	TRANSACTION_CATEGORY_NAME	TRANSACTION_BASE_TYPE	TX MONTH
1	Paypal	7-Eleven	Entertainment/Recreation	Debit	
2					

Request Free Data Sample →

3K Merchants

99% High precision mapping for 600 tickers

USA covered

9 years of historical data

Get a Quote →

Contact Provider →

"Our most granular offering providing line-by-line transactions for 4 millions US consumers."

Access to Consumer spend data of de-identified 4M users over 9 years. Clean tagged consumer transaction data on millions of merchants public and private. Suitable for all investment use cases - Fundamental, Quant, Private Equity, Venture Capital.

### Data Attributes

Attribute & Description	Example
-------------------------	---------

Envestnet | Yodlee  
Powering Dynamic Innovation for Financial Services

Verified Provider  
100% Response rate

Trusted by

COMMERCE DATA



**So, which are the features actually driving the prices of data products?**





## Top 33 datasets worth > US\$30K per month share some characteristics...

- ▶ All of them provide huge amounts of data for MM people, 10k locations or companies
- ▶ 61% of them offer daily updates
- ▶ They provide fresh rather than long-historical data
- ▶ 67% of them are US-focused, 21% are global
- ▶ 73% of them relate to retail and marketing
- ▶ Most valuable B2B marketing products include precise enterprise data, and key contact information
- ▶ At least 48% of them enable a very granular location-based analysis, and 27% of them geo-located data
- ▶ 21% B2C marketing products allow for *session reconstruction* (i.e., connecting the different data points of individuals and entities)

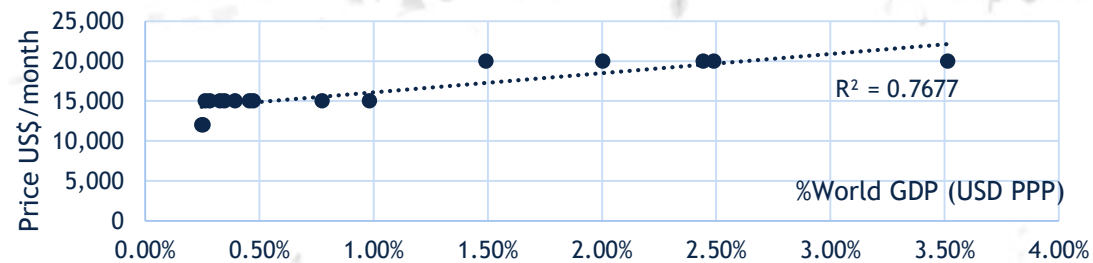


## Top 33 datasets worth > US\$30K per month share some characteristics...

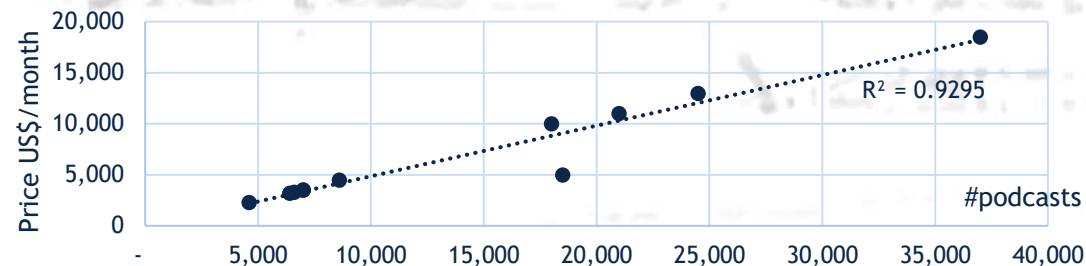
- ▶ All of them provide huge amounts of data for MM people, 10k locations or companies
- ▶ 61% of them offer daily updates
- ▶ They provide fresh rather than long-historical data
- ▶ 67% of them are US-focused, 21% are global
- ▶ 73% of them relate to retail and marketing
- ▶ Most valuable B2B marketing products include precise enterprise data, and key contact information
- ▶ At least 48% of them enable a very granular location-based analysis, and 27% of them geo-located data
- ▶ 21% B2C marketing products allow for *session reconstruction* (i.e., connecting the different data points of individuals and entities)

## ... and we are able to find clear simple seller-specific pricing strategies

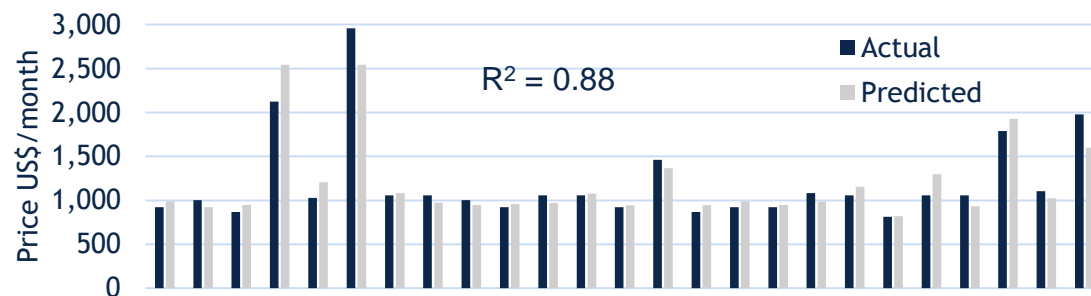
Price vs country GDP in mobile coverage pricing



Price vs #podcasts in podcast metadata pricing



Pricing of customer segmentation products





Based on this analysis we derived a list of “desirable” features of data that data prices seem to depend upon

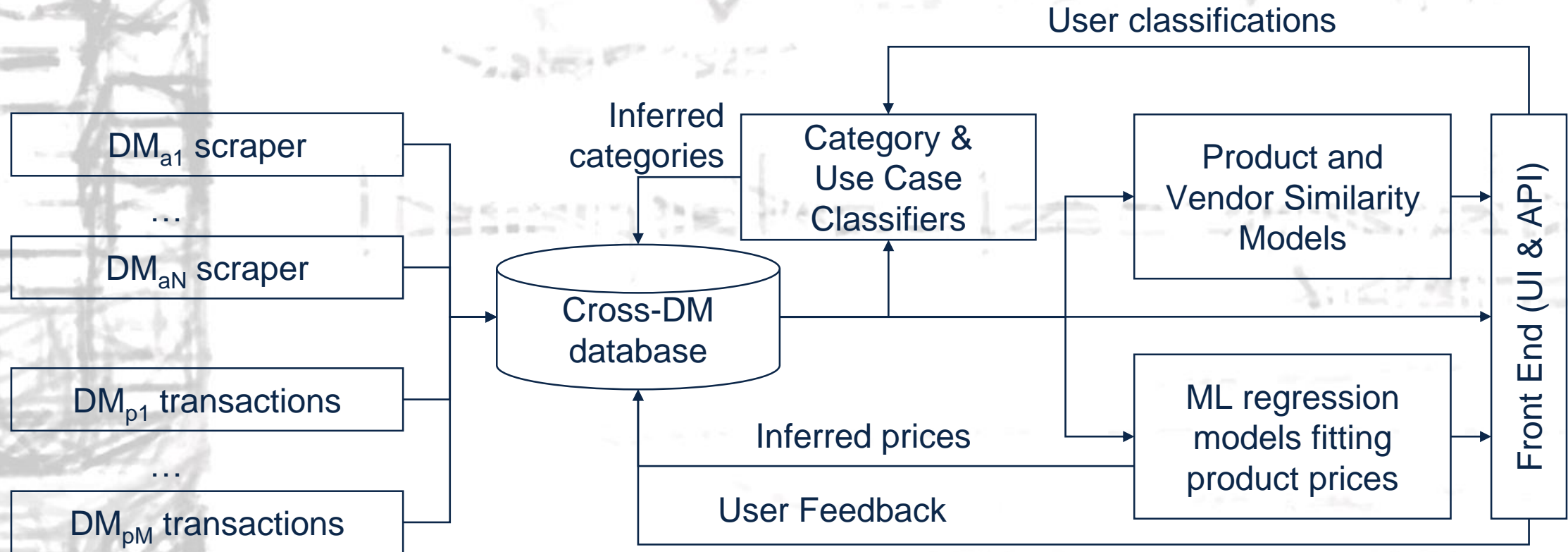
Question	Group	Definition	N° features	Example of features
What?	Category	Labels attached to the product that define the type of data it contains	custom	'Weather', 'Gaming', 'Financial'
	Description	Stem-like features obtained from data product descriptions	custom	'wordmarket', 'wordidenti', 'wordlist'
	Identifiability	Tells whether the product allows the buyer to recognize the activity of individuals or to identify specific companies	2	'idSessions', 'IdCompanies'
How much?	Volume	Normalized n° units covered broken down by the nature of such units	14	'people', 'entities'
	Update rate	Defines the frequency between data updates as announced by the seller	11	'real time', 'monthly', 'hourly'
	Granularity	Defines the resolution and potential aggregations of data	3	'aggLevel', 'spatialResol', 'timeResol'
How?	Delivery method	Defines how the buyer can have access to data	8	'S3Bucket', 'Download', 'FeedAPI'
	Format	Defines the way in which data is arranged	17	'txt', 'shapefile', 'xls'
	Add-ons	Tells whether the product attaches any add-on or has any limitations	2	'ProfServices', 'Limitations'
When?	History	Time scope included	1	'History'
Where?	Geo scope	Metrics about countries included in the data product	up to 249	'N° Countries', 'USA', 'Canada'



## Bottomline...

- ▶ We have provided a first glimpse into the growing market of B2B data.
- ▶ We found that data sells at an immensely wide range of prices up to US\$500k or US\$150k monthly.
- ▶ We have pointed to the challenges of comparing across DMs and potential solutions for that.
- ▶ We were able to list groups of features that are determining prices of real products in the market.
- ▶ We came to the conclusion that there is a need of an overarching solution to provide transparency on how much a piece of data might be worth in the market, and why.

**We are working on such a data quotation tool to be able to predict the prices of a data product out of its metadata based on market prices and transactions**



**Such a tool will have limitations, since it does not consider: i) the usability for the buyer, ii) the quality of the data, iii) the specific value for a buyer.**

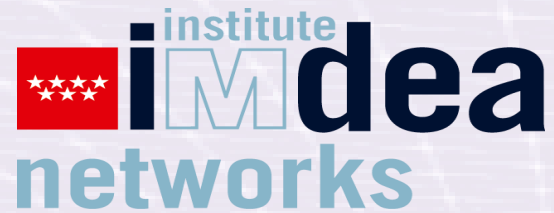
Thank you!

Now it is Q&A time!

Acknowledgement to:

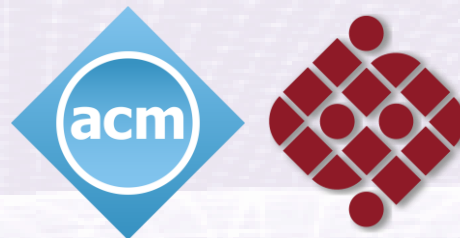
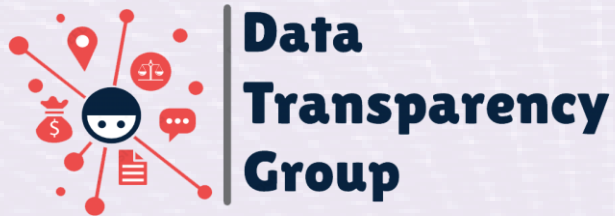
Horizon 2020

For more information please contact:



**Santiago Andrés Azcoitia**

santiago.azcoitia@imdea.org



**CoNEXT 2022**

