Modelos predictivos para retail

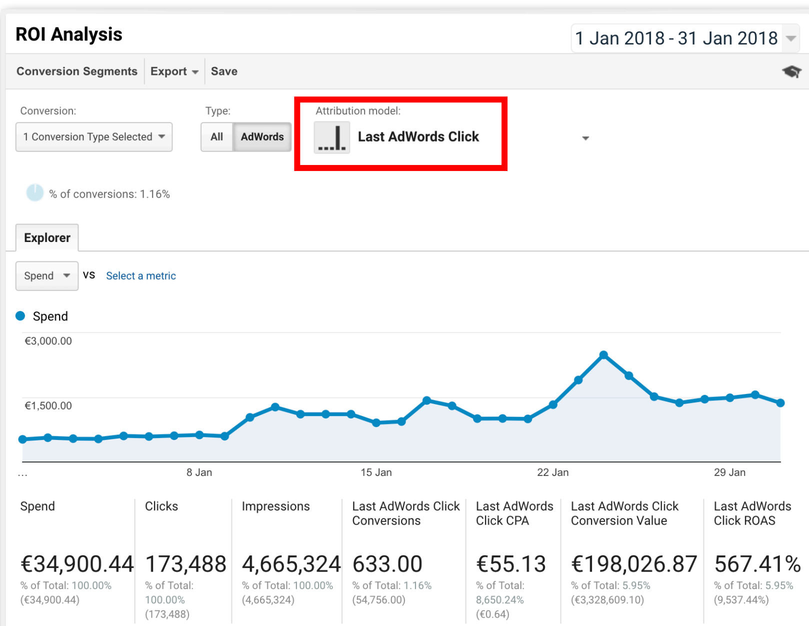
José Ramón Cajide

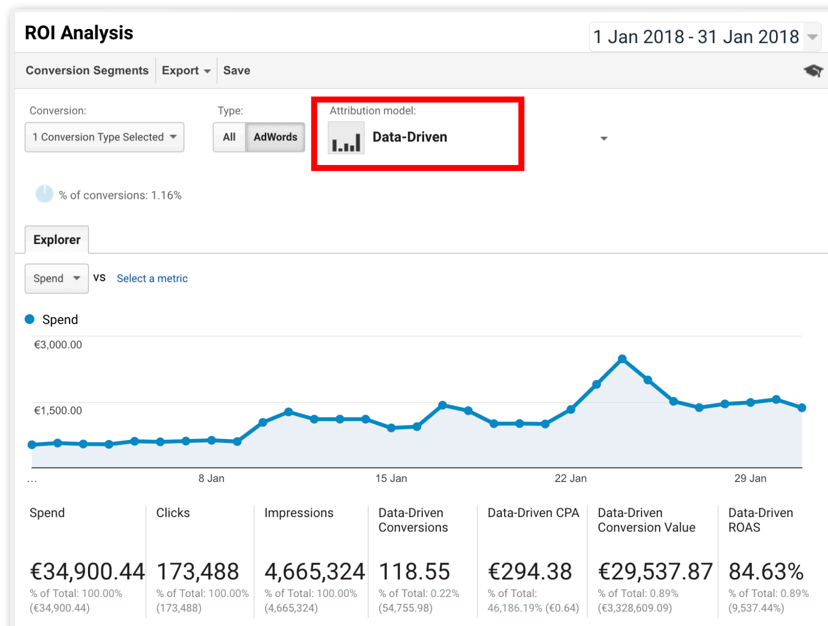
DATA SCIENTIST en **El Arte de Medir**

El arte de medir
PASIÓN Y PRECISIÓN

DATA KONFERENCES

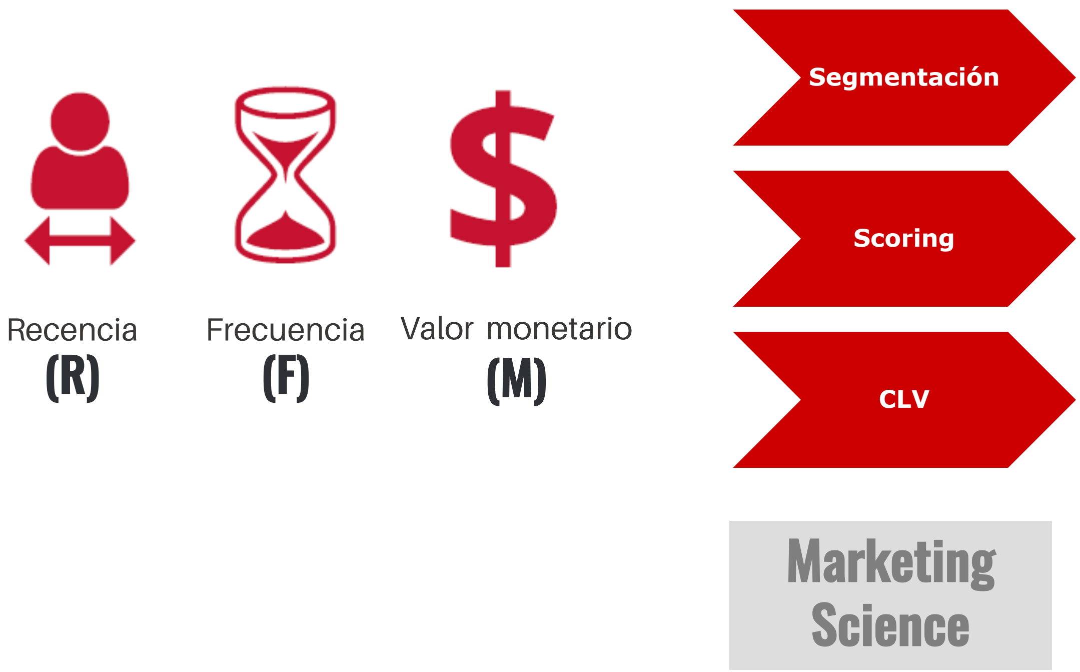
Analítica Digital





Valor de vida del cliente:

Modelos predictivos en situaciones no contractuales



Fuente de datos

```sql
SELECT
  MD5(CAST(client_id AS STRING)) AS client_id,
  orders_id,
  timeplaced,
  amount
FROM
```

Valid: This query will process 5.85 GB when run.

Standard SQL Dialect

RUN QUERY | Save Query | Save View | Format Query | Show Options

Results | Details

Row	client_id	orders_id	timeplaced	amount
1	IX3dlXtEL5UNnhAuFDMdYA==	3.28942586E8	2017-03-18 04:15:17	208.0
2	r2dxlUYgA5Yw3SQLGMPAkA==	3.28957977E8	2017-03-18 05:23:48	10260.0
3	tnz97sF8hJj+MuXQqpFDMw==	3.28500184E8	2017-03-18 00:20:53	129.0
4	J9hyCaxgdPskoBwvHqNTVw==	3.28934039E8	2017-03-18 01:38:43	129.0
5	tKRhDXacG9GIMDLv1pXe8A==	3.28965652E8	2017-03-18 03:06:05	129.0
6	ByEckdCEU+/ae1yNKbM3MQ==	3.28907361E8	2017-03-18 04:35:32	129.0
7	Pl4ZttBhagmsmnH2x6hvug==	3.28490388E8	2017-03-18 07:33:02	129.0
8	aUslbUGkG9CXEs/cs5sBsQ==	3.28984002E8	2017-03-18 08:12:16	418.5
9	asqiRQtkiSgiA4mvHMYu3g==	3.28920727E8	2017-03-18 03:43:35	259.0
10	UHDt2cvfOz8kQttFWzajyQ==	3.28918219E8	2017-03-18 01:16:54	69.0
11	NsxgOBFqKulCmRffFEGLEg==	3.28486209E8	2017-03-18 01:43:54	69.0

El arte de medir
PASIÓN Y PRECISIÓN

Variables RFM

```
SELECT
  MD5(CAST(client_id AS STRING)) AS client_id,
  PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)) AS first_purchase,
  TIMESTAMP_DIFF(TIMESTAMP("2017-12-31 00:00:00"),PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)), DAY) AS days_since,
  COUNT(*) AS frequency,
  ROUND(AVG(amount),2) AS amount
FROM (
  SELECT
    client_id,
    orders_id,
    timeplaced,
    amount
  FROM
```

Valid: This query will process 1.49 GB when run.

Standard SQL Dialect ✕

RUN QUERY | Save Query | Save View | Format Query | Show Options | Query complete (47.3s elapsed, 1.49 GB processed)

Results | Details | Download as CSV

Row	client_id	first_purchase	days_since	frequency	amount
1	pzVe+0vtY2WsWF7O1RUx3g==	2017-05-17 11:47:43.000 UTC	227	23	737.83
2	TVIfFjRvxX2kdYLbt3MgBQ==	2017-01-03 21:47:45.000 UTC	361	74	50.89
3	sks6DwnBncN4tDCpYhU6TQ==	2017-01-25 18:53:18.000 UTC	339	27	295.23
4	6t78VadGNLX7AximW1miNw==	2017-05-03 14:38:59.000 UTC	241	10	288.3
5	mXELRuoiRCHliS9VAqmQDw==	2017-04-03 14:22:59.000 UTC	271	11	222.3
6	N4KR+xpKsQTEbYOjVlc/BQ==	2017-01-04 23:03:10.000 UTC	360	14	122.27

Table | JSON

First < Prev Rows 1 - 6 of 23356398 Next > Last

El arte de medir
PASION Y PRECISION

Modelo RFM básico

```
SELECT
  client_id,
  NTILE(4) OVER (ORDER BY last_purchase DESC) AS rfm_recency,
  NTILE(4) OVER (ORDER BY frequency DESC) AS rfm_frequency,
  NTILE(4) OVER (ORDER BY amount DESC) AS rfm_monetary,
  CONCAT(CAST(NTILE(4) OVER (ORDER BY last_purchase DESC) AS STRING),"-",CAST(NTILE(4) OVER (ORDER BY frequency DESC) AS STRING),"-",CAST(
FROM (
  SELECT
    MD5(CAST(client_id AS STRING)) AS client_id,
    PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)) AS first_purchase,
    PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)) AS last_purchase,
    TIMESTAMP_DIFF(TIMESTAMP("2017-12-31 00:00:00"),PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)), DAY) AS days_since,
    COUNT(*) AS frequency,
```

Valid: This query will process 4.81 GB when run.

Standard SQL Dialect ✕

RUN QUERY | Save Query | Save View | Format Query | Show Options

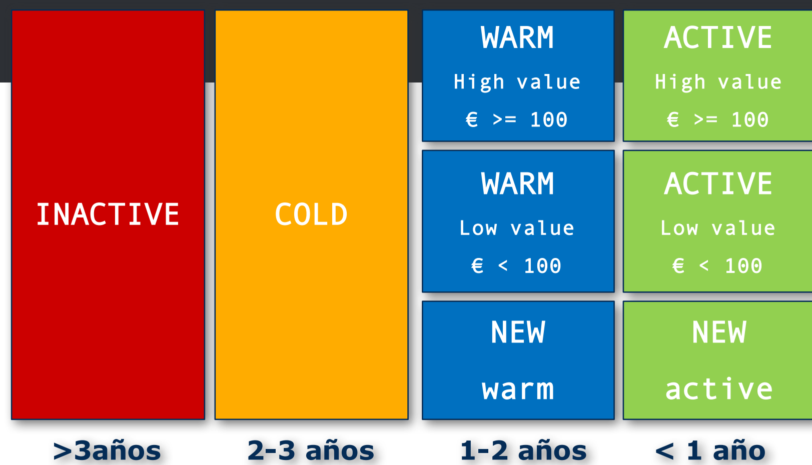
Results | Details

Row	client_id	rfm_recency	rfm_frequency	rfm_monetary	rfm
1	fyMRrVkL9j6LeTOoAxzVUA==	4	2	1	4-2-1
2	PBNROOmrevpXCchqJhGP0A==	2	4	1	2-4-1
3	s2bx6Pn3zTXpPKwstJA9NA==	4	2	1	4-2-1
4	VRPh2rwPZFlwwp4GBgWmAQ==	3	1	1	3-1-1
5	QkrDsDBYrhBtIGVoLvLMng==	1	3	1	1-3-1
6	SjXZyObpenqkETBv7Zm33w==	3	1	1	3-1-1
7	lo/VOZ/wHvhGW5wNzeTwmw==	2	4	1	2-4-1
8	xyr9mnN75LrHMIPP5FOvWA==	4	2	1	4-2-1

		Monetary Value Quartiles			
Recency Quartiles	Frequency Quartiles	1	2	3	4
1st	1	lost			
	2				
	3				
	4				
2nd	1	at risk			
	2				
	3				
	4				
3rd	1	slipping away			
	2				
	3				
	4				
4th	1	active			
	2				
	3				
	4				

El arte de medir
PASIÓN Y PRECISIÓN

Segmentación RFM heurística

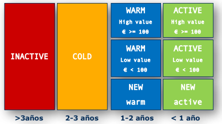



```r
customers_2015 <-customers_2015 %>%
  mutate(segment = case_when(recency > 365*3 ~ "inactive",
                             recency <= 365*3 & recency > 365*2  ~ "cold",
                             recency <= 365*2 & recency > 365*1 ~ "warm",
                             recency <= 365 ~ "active"
                             ))

customers_2015 <- customers_2015 %>%
  mutate(segment = case_when(segment == "warm" & first_purchase <= 365*2 ~ "new warm",
                             segment == "warm" & amount < 100 ~ "warm low value",
                             segment == "warm" & amount >= 100 ~ "warm high value",
                             segment == "active" & first_purchase <= 365 ~ "new active",
                             segment == "active" & amount < 100 ~ "active low value",
                             segment == "active" & amount >= 100 ~ "active high value",
                             TRUE ~ segment))



# Convertimos el segmento, en una variable categórica y la ordenamos
customers_2015$segment = factor(x = customers_2015$segment,
                              levels = c("inactive", "cold",
                                         "warm high value", "warm low value", "new warm",
                                         "active high value", "active low value", "new active"))
```

	segment	recency	first_purchase	frequency	amount
1	inactive	2178.06917	2546.12671	1.814479	48.11277
2	cold	857.73973	1432.07552	2.303205	51.73989
3	warm high value	455.08438	2015.31127	4.714286	327.40746
4	warm low value	474.33569	2063.59762	4.531632	38.59193
5	new warm	509.26324	516.58093	1.044776	66.59903
6	active high value	88.77858	1985.86758	5.888307	240.04574
7	active low value	108.31934	2003.76033	5.935406	40.72452
8	new active	84.94907	89.97222	1.045635	77.13385

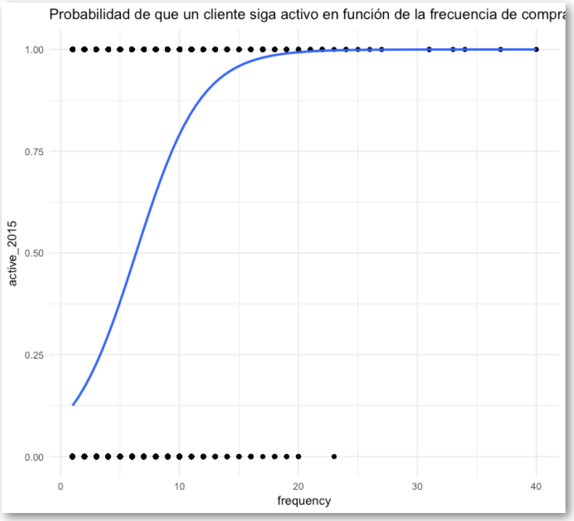
El arte de medir
PASION Y PRECISION

Modelo RFM heurístico



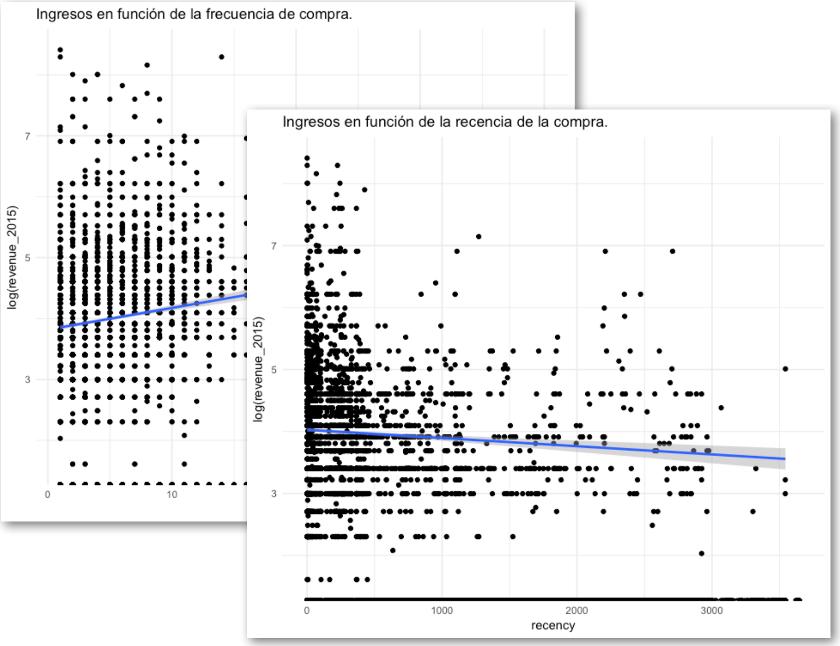
¿Cuál es la probabilidad de que cada cliente siga "vivo"?

$$\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}.$$

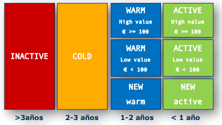


¿Cuánto dinero generará cada uno de nuestros clientes?

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$$



Modelo RFM heurístico



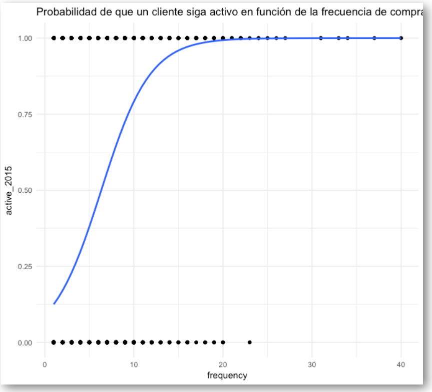
Modelamos los datos a partir del histórico

	customer_id	recency	first_purchase	frequency	avg_amount	max_amount	revenue_2015	active_2015
1	10	3463.9583333	3463.958	1	30.00000	30.00	0	0
2	80	301.9583333	3385.958	6	70.00000	80.00	80	1
3	90	392.9583333	3417.958	10	115.80000	153.00	0	0
4	120	1035.9583333	1035.958	1	20.00000	20.00	0	0
5	130	2604.9583333	3344.958	2	50.00000	60.00	0	0
6	160	2597.9583333	3211.958	2	30.00000	30.00	0	0
7	190	1845.9583333	3324.958	5	68.00000	100.00	0	0
8	220	1692.9583333	3298.958	2	25.00000	30.00	0	0
9	230	3619.9583333	3619.958	1	50.00000	50.00	0	0
10	240	97.9583333	3449.958	4	16.25000	20.00	0	0

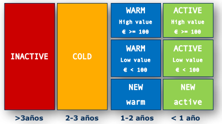
¿Cuál es la probabilidad de que cada cliente siga “vivo”?

```r
prob.model <- multinom(formula = active_2015 ~ recency + first_purchase + frequency + avg_amount + max_amount,
                       data = in_sample)
```

```
Coefficients:
                      Values    Std. Err.
(Intercept)    -5.332175e-01 4.410550e-02
recency        -1.965114e-03 6.000518e-05
first_purchase -1.163276e-05 3.925582e-05
frequency       2.195100e-01 1.479178e-02
avg_amount      4.159627e-04 3.634840e-04
max_amount     -1.558099e-04 2.713806e-04

Residual Deviance: 12368.93
AIC: 12380.93
```



Modelo RFM heurístico



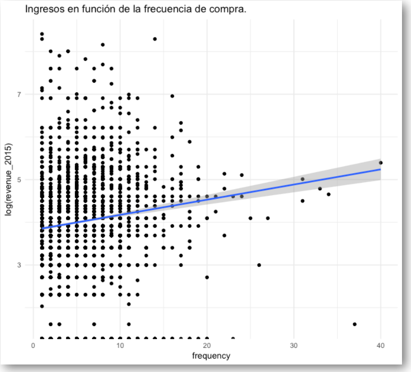
Modelamos los datos a partir del histórico

	customer_id	recency	first_purchase	frequency	avg_amount	max_amount	revenue_2015	ctive_2015
1	10	3463.9583333	3463.958	1	30.00000	30.0	0	0
2	80	301.9583333	3385.958	6	70.00000	80.0	80	1
3	90	392.9583333	3417.958	10	115.80000	153.0	0	0
4	120	1035.9583333	1035.958	1	20.00000	20.0	0	0
5	130	2604.9583333	3344.958	2	50.00000	60.0	0	0
6	160	2597.9583333	3211.958	2	30.00000	30.0	0	0
7	190	1845.9583333	3324.958	5	68.00000	100.0	0	0
8	220	1692.9583333	3298.958	2	25.00000	30.0	0	0
9	230	3619.9583333	3619.958	1	50.00000	50.0	0	0
10	240	97.9583333	3449.958	4	16.25000	20.0	0	0

¿Cuánto dinero generará cada uno de nuestros clientes?

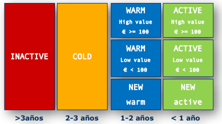
```
amount.model = lm(formula = log(revenue_2015) ~ log(avg_amount) + log(max_amount), data = in_sample_purchased_2015)
summary(amount.model)
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-2.7866 -0.1811 -0.0770  0.1852  3.5656 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      0.37000    0.04003   9.242   <2e-16 ***
log(avg_amount)  0.54881    0.04167  13.171   <2e-16 ***
log(max_amount)  0.38813    0.03796  10.224   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4781 on 3883 degrees of freedom
Multiple R-squared:  0.6927,	Adjusted R-squared:  0.6926 
F-statistic:  4377 on 2 and 3883 DF,  p-value: < 2.2e-16
```



Modelo RFM heurístico

Aplicamos ambos modelos predictivos

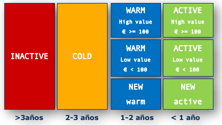


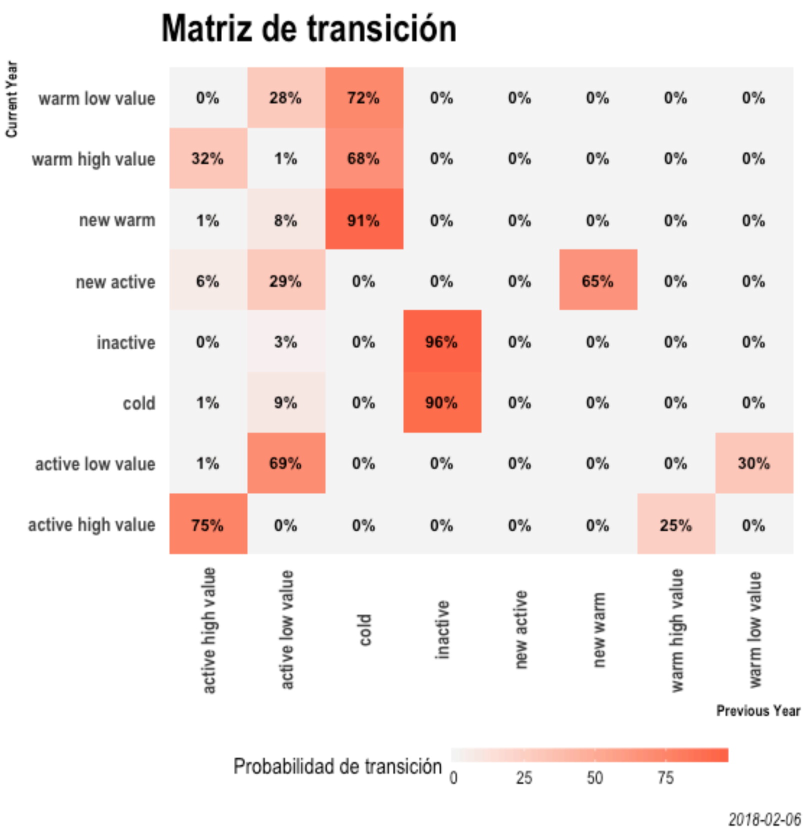

```
customers_2015 <- customers_2015 %>%
  mutate(prob_predicted = predict(object = prob.model, newdata = customers_2015, type = "probs"),
         revenue_predicted = exp(predict(object = amount.model, newdata = customers_2015)),
         score_predicted = prob_predicted * revenue_predicted)
```

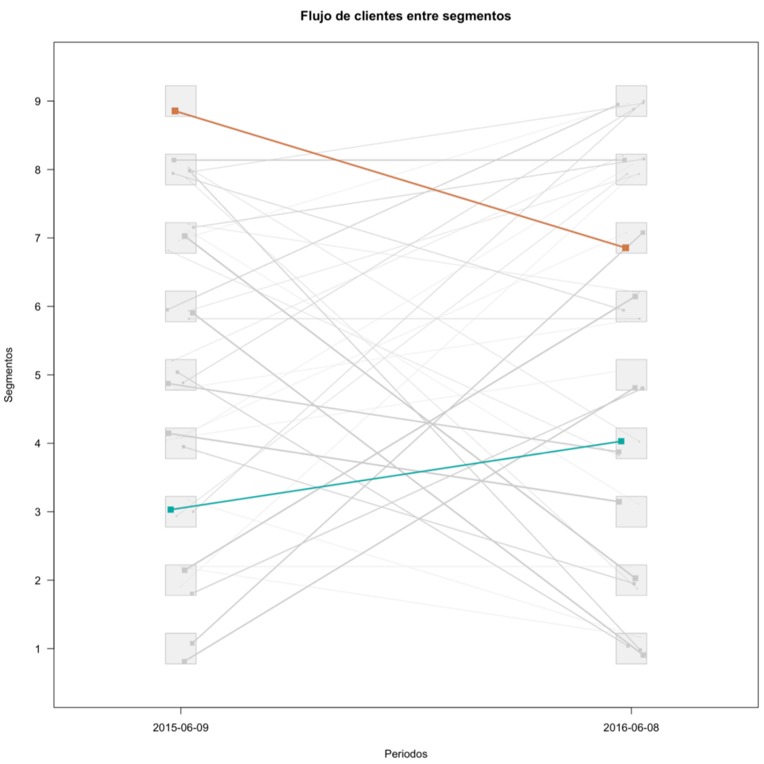
	customer_id	recency	first_purchase	frequency	avg_amount	max_amount	prob_predicted	revenue_predicted	score_predicted
1	10	3828.9583333	3828.9583333	1	30.00000	30.00	0.0003800663	35.04801	0.01
2	80	342.9583333	3750.9583333	7	71.42857	80.00	0.5751788847	82.55809	47.49
3	90	757.9583333	3782.9583333	10	115.80000	153.00	0.5381086855	138.42614	74.49
4	120	1400.9583333	1400.9583333	1	20.00000	20.00	0.0440304521	23.97045	1.06
5	130	2969.9583333	3709.9583333	2	50.00000	60.00	0.0025676010	60.70928	0.16
6	160	2962.9583333	3576.9583333	2	30.00000	30.00	0.0025976311	35.04801	0.09
7	190	2210.9583333	3689.9583333	5	68.00000	100.00	0.0216540011	87.62916	1.90
8	220	2057.9583333	3663.9583333	2	25.00000	30.00	0.0151391262	31.71083	0.48
9	230	3984.9583333	3984.9583333	1	50.00000	50.00	0.0002806970	56.56173	0.02
10	240	462.9583333	3814.9583333	4	16.25000	20.00	0.3530455399	21.38878	7.55

Modelo RFM heurístico

¿Cómo cambian los clientes de segmento?

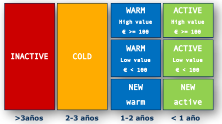






Modelo RFM heurístico

La matriz de transición como método para estimar el CLV



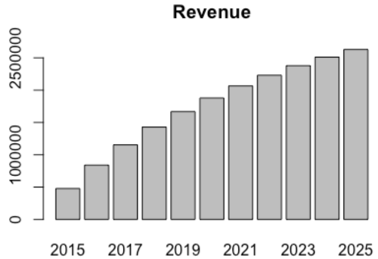
REVENUE

	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
inactive	9158	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
cold	1903	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
warm high value	119	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
warm low value	901	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
new warm	938	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
active high value	573	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
active low value	3313	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
new active	1512	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

?

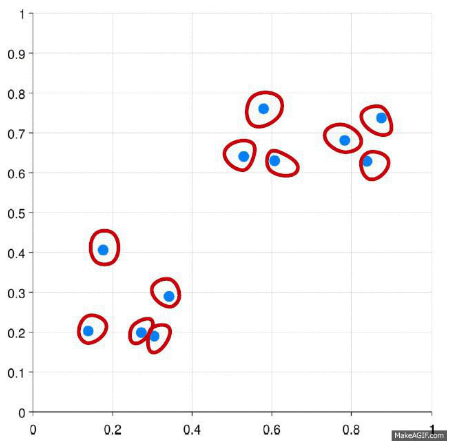
0%	28%	72%	0%	0%	0%	0%	0%
32%	1%	68%	0%	0%	0%	0%	0%
1%	8%	91%	0%	0%	0%	0%	0%
6%	29%	0%	0%	0%	65%	0%	0%
0%	3%	0%	96%	0%	0%	0%	0%
1%	9%	0%	90%	0%	0%	0%	0%
1%	69%	0%	0%	0%	0%	0%	30%
75%	0%	0%	0%	0%	0%	25%	0%

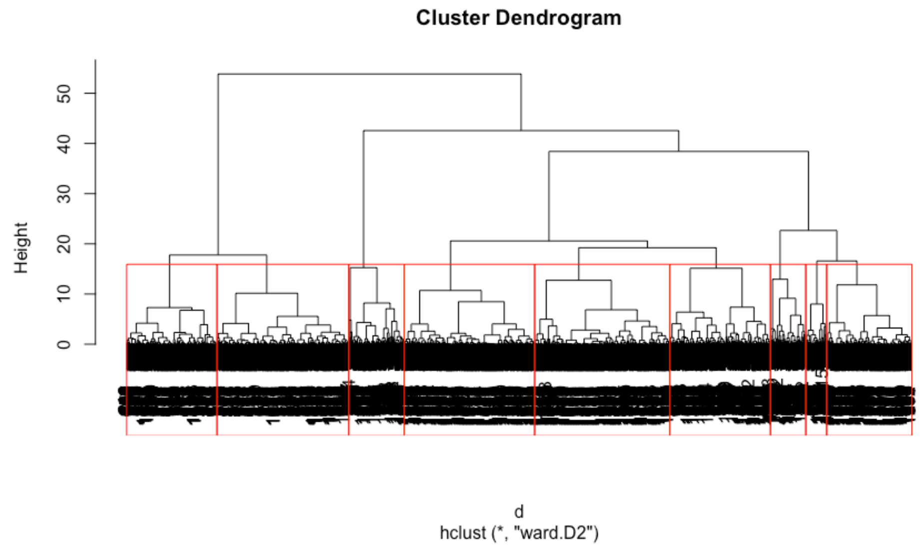
	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
inactive	9158	10517	11539	12636	12940	13186	13386	13542	13664	13759	13834
cold	1903	1584	1711	874	821	782	740	709	684	665	650
warm high value	119	144	165	160	156	152	149	146	143	141	139
warm low value	901	991	1058	989	938	884	844	813	789	771	756
new warm	938	987	0	0	0	0	0	0	0	0	0
active high value	573	657	639	624	607	593	581	571	562	554	547
active low value	3313	3537	3305	3134	2954	2820	2717	2637	2575	2527	2490
new active	1512	0	0	0	0	0	0	0	0	0	0

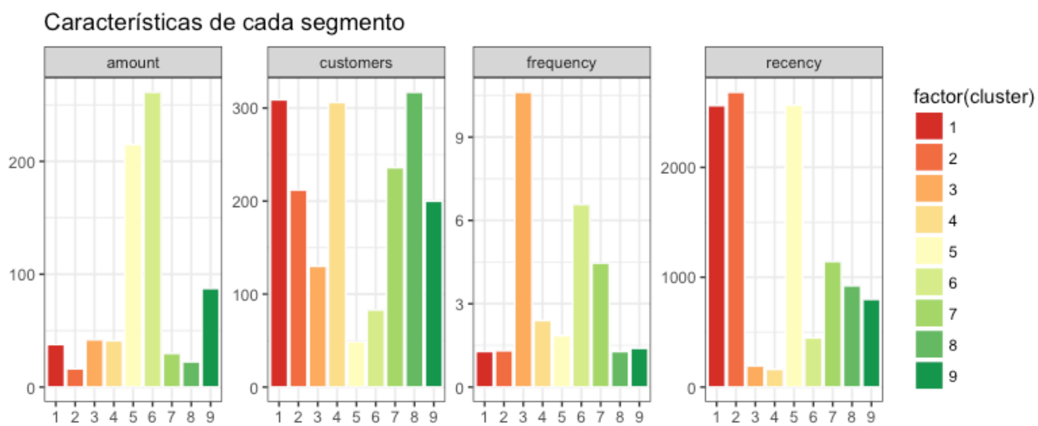


El arte de medir
PASIÓN Y PRECISIÓN

Segmentación jerárquica

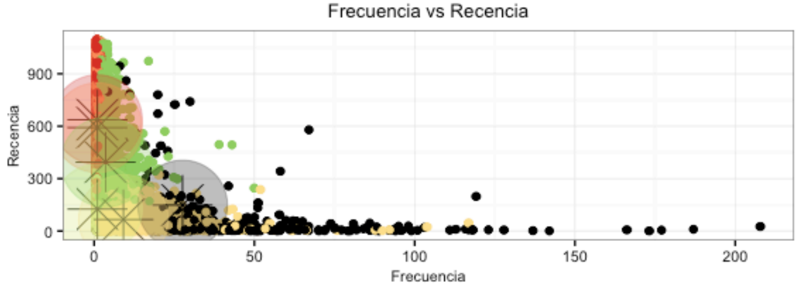




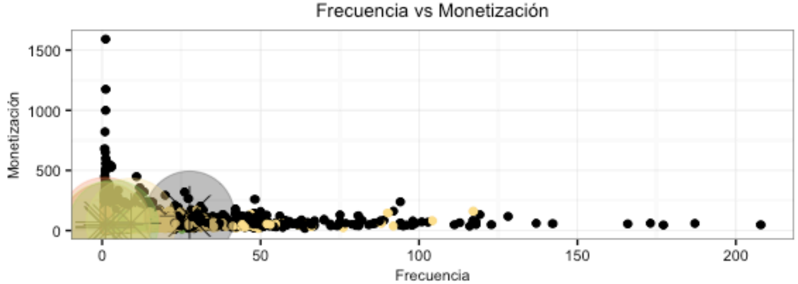


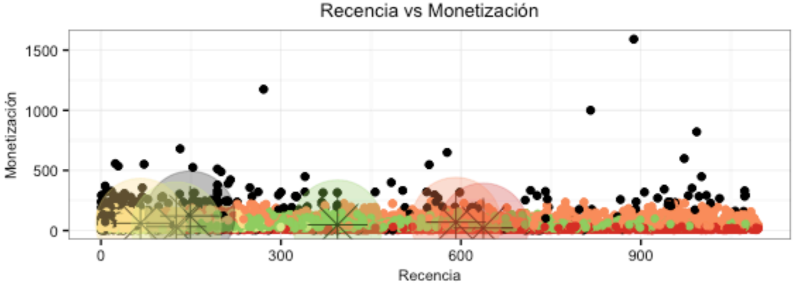
Segmentación k-means

Segmentación RFM-R (Todos los clientes)

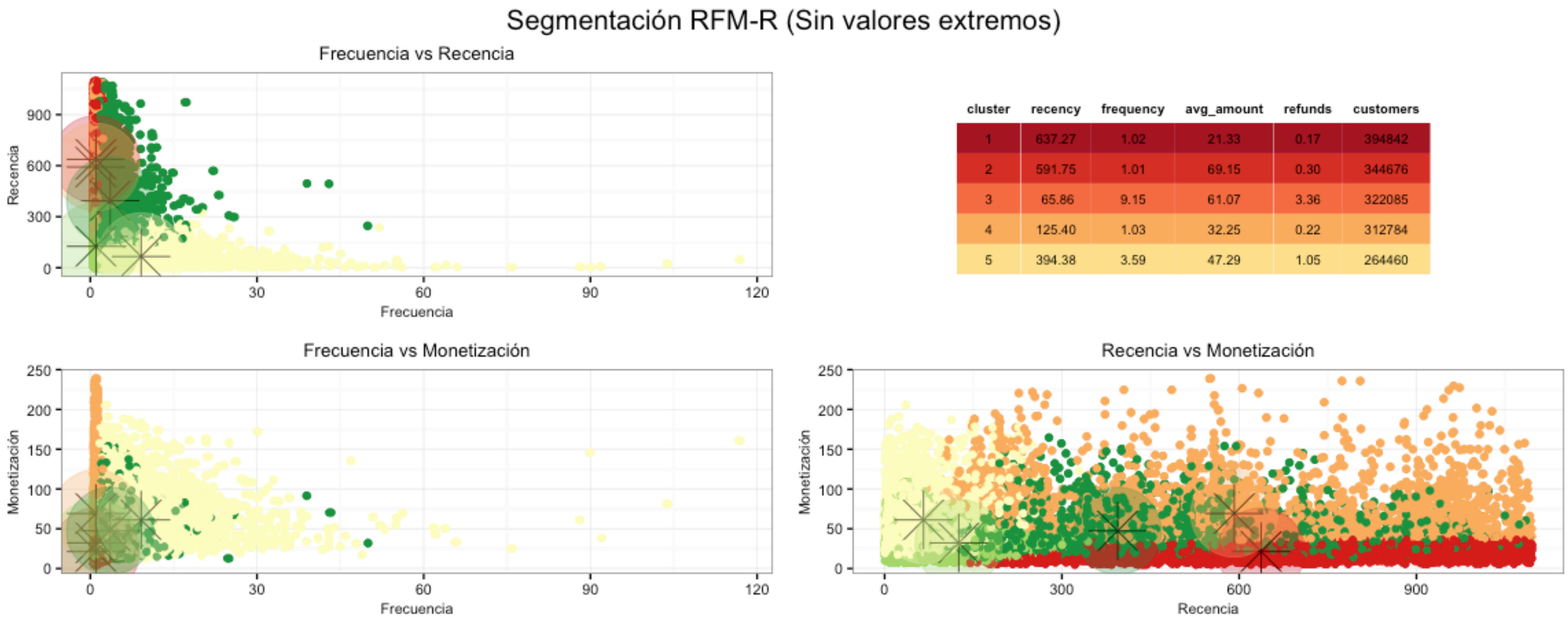


cluster	recency	frequency	avg_amount	refunds	customers
0	148.57	27.59	121.64	16.25	68286
1	637.27	1.02	21.33	0.17	394842
2	591.75	1.01	69.15	0.30	344676
3	65.86	9.15	61.07	3.36	322085
4	125.40	1.03	32.25	0.22	312784
5	394.38	3.59	47.29	1.05	264460





Segmentación k-means



cluster	recency	frequency	avg_amount	refunds	customers
1	637.27	1.02	21.33	0.17	394842
2	591.75	1.01	69.15	0.30	344676
3	65.86	9.15	61.07	3.36	322085
4	125.40	1.03	32.25	0.22	312784
5	394.38	3.59	47.29	1.05	264460

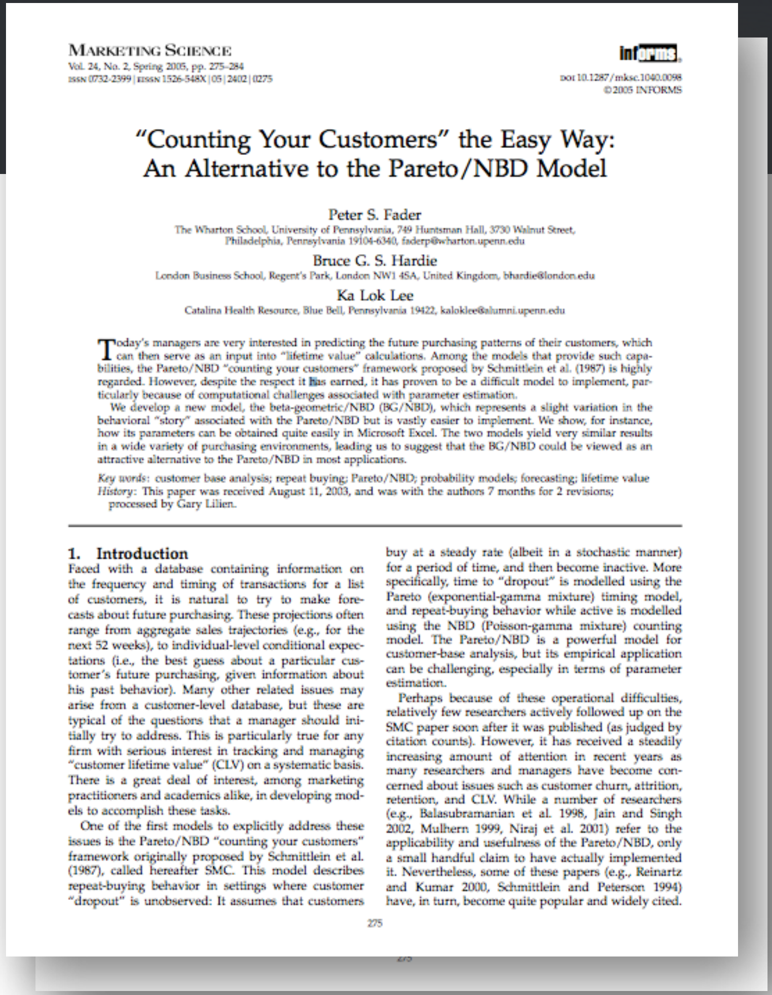
El arte de medir
PASIÓN Y PRECISIÓN

Modelos probabilísticos

- Válido para situaciones **no contractuales**

- Válido bajo las siguiente hipótesis:

    1. Los clientes realizan pedidos mientras están activos
    2. Los clientes pueden abandonar tras un periodo de tiempo



MARKETING SCIENCE
Vol. 24, No. 2, Spring 2005, pp. 275–284
ISSN 0732-2399 | EISSN 1526-548X | 05 | 2402 | 0275
informs.
DOI 10.1287/mksc.1040.0098

"Counting Your Customers" the Easy Way: An Alternative to the Pareto/NBD Model

Peter S. Fader
The Wharton School, University of Pennsylvania, 749 Huntsman Hall, 3730 Walnut Street, Philadelphia, Pennsylvania 19104-6340, faderp@wharton.upenn.edu
Bruce G. S. Hardie
London Business School, Regent's Park, London NW1 4SA, United Kingdom, bhardie@london.edu
Ka Lok Lee
Catalina Health Resource, Blue Bell, Pennsylvania 19422, kaloklee@alumni.upenn.edu

Today's managers are very interested in predicting the future purchasing patterns of their customers, which can then serve as an input into "lifetime value" calculations. Among the models that provide such capabilities, the Pareto/NBD "counting your customers" framework proposed by Schmittlein et al. (1987) is highly regarded. However, despite the respect it has earned, it has proven to be a difficult model to implement, particularly because of computational challenges associated with parameter estimation.

We develop a new model, the beta-geometric/NBD (BG/NBD), which represents a slight variation in the behavioral "story" associated with the Pareto/NBD but is vastly easier to implement. We show, for instance, how its parameters can be obtained quite easily in Microsoft Excel. The two models yield very similar results in a wide variety of purchasing environments, leading us to suggest that the BG/NBD could be viewed as an attractive alternative to the Pareto/NBD in most applications.

Key words: customer base analysis; repeat buying; Pareto/NBD; probability models; forecasting; lifetime value
History: This paper was received August 11, 2003, and was with the authors 7 months for 2 revisions; processed by Gary Lilien.

1. Introduction

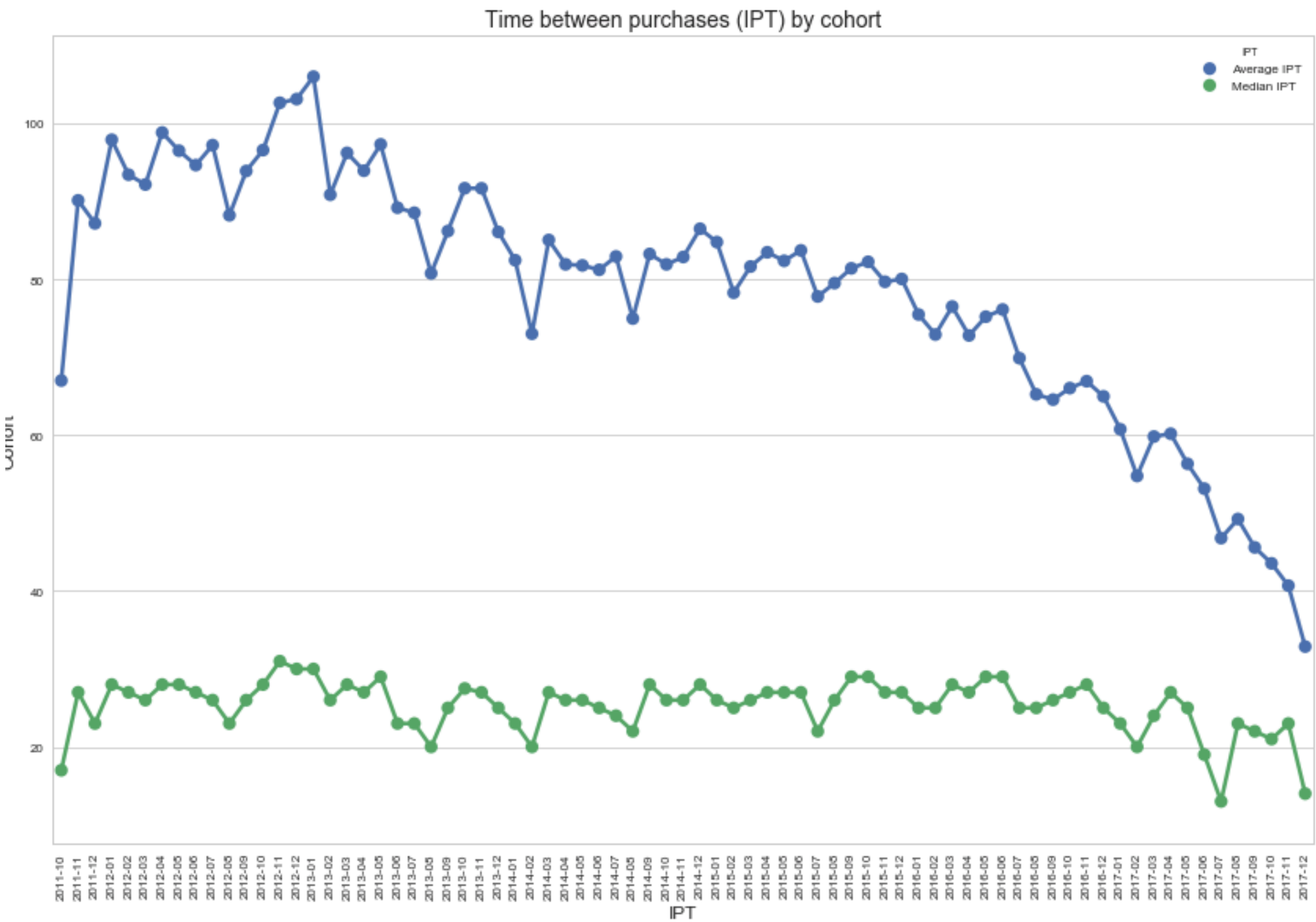
Faced with a database containing information on the frequency and timing of transactions for a list of customers, it is natural to try to make forecasts about future purchasing. These projections often range from aggregate sales trajectories (e.g., for the next 52 weeks), to individual-level conditional expectations (i.e., the best guess about a particular customer's future purchasing, given information about his past behavior). Many other related issues may arise from a customer-level database, but these are typical of the questions that a manager should initially try to address. This is particularly true for any firm with serious interest in tracking and managing "customer lifetime value" (CLV) on a systematic basis. There is a great deal of interest, among marketing practitioners and academics alike, in developing models to accomplish these tasks.

One of the first models to explicitly address these issues is the Pareto/NBD "counting your customers" framework originally proposed by Schmittlein et al. (1987), called hereafter SMC. This model describes repeat-buying behavior in settings where customer "dropout" is unobserved: It assumes that customers buy at a steady rate (albeit in a stochastic manner) for a period of time, and then become inactive. More specifically, time to "dropout" is modelled using the Pareto (exponential-gamma mixture) timing model, and repeat-buying behavior while active is modelled using the NBD (Poisson-gamma mixture) counting model. The Pareto/NBD is a powerful model for customer-base analysis, but its empirical application can be challenging, especially in terms of parameter estimation.

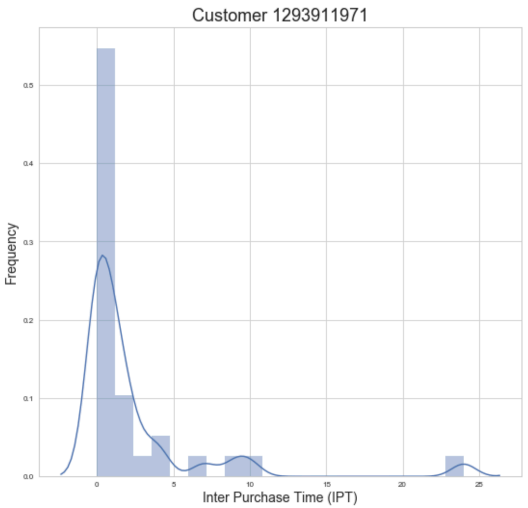
Perhaps because of these operational difficulties, relatively few researchers actively followed up on the SMC paper soon after it was published (as judged by citation counts). However, it has received a steadily increasing amount of attention in recent years as many researchers and managers have become concerned about issues such as customer churn, attrition, retention, and CLV. While a number of researchers (e.g., Balasubramanian et al. 1998, Jain and Singh 2002, Mulhern 1999, Niraj et al. 2001) refer to the applicability and usefulness of the Pareto/NBD, only a small handful claim to have actually implemented it. Nevertheless, some of these papers (e.g., Reinartz and Kumar 2000, Schmittlein and Peterson 1994) have, in turn, become quite popular and widely cited.

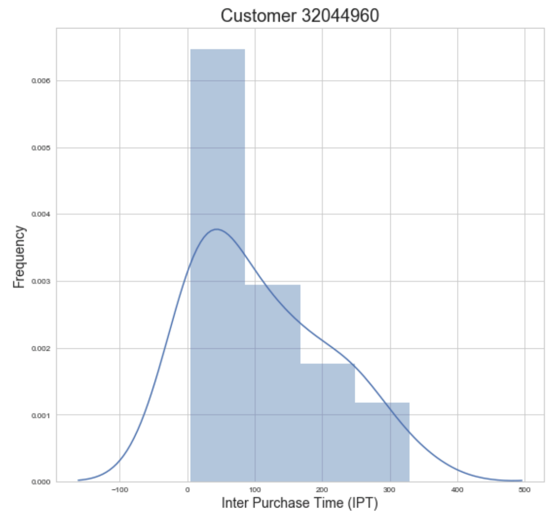
https://goo.gl/S9fFPm

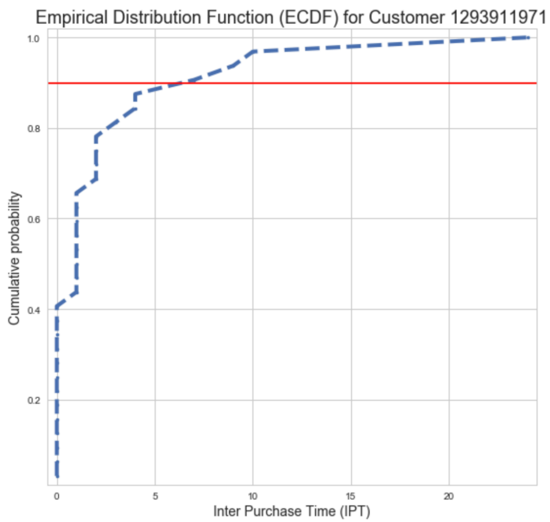
Frecuencia de compra

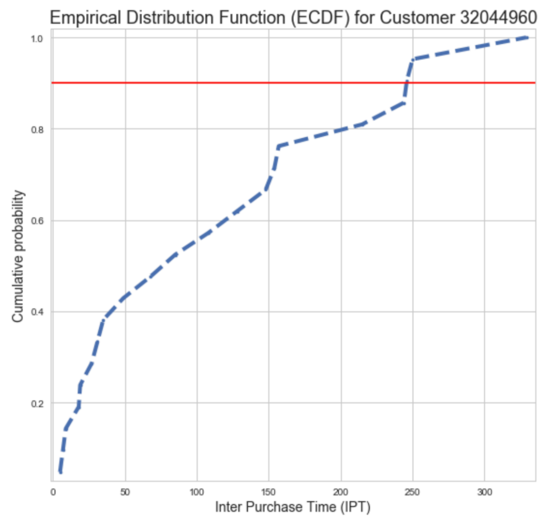


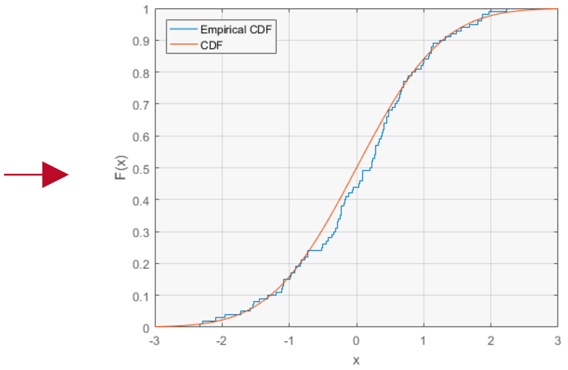
Frecuencia de compra y abandono





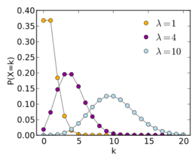


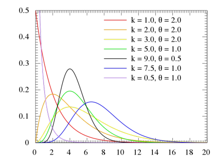




Suposiciones del modelo

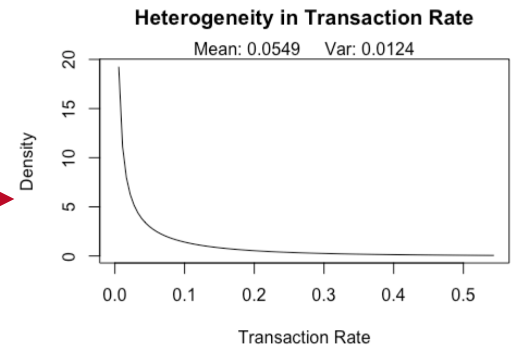
[1] Los clientes realizan sus compras de acuerdo a una distribución de probabilidad discreta de Poisson con una frecuencia de compra λ





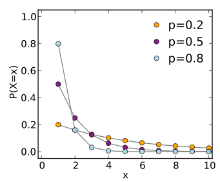
$\mathrm{Gamma}(r, \alpha)$

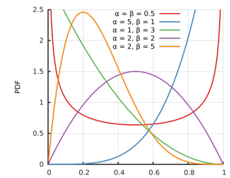
$f(\lambda \mid r, \alpha)$



Propensión a la compra

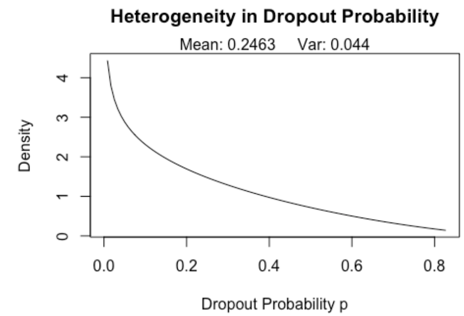
[2] Tras cada compra, un cliente tiene un probabilidad ***p*** de convertirse en inactivo, la cual sigue una distribución geométrica





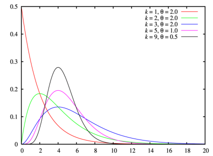
$\mathrm{Beta}(a. b)$

$f(p \mid a, b)$



Probabilidad de abandono

[3] EL importe de cada compra que realiza un cliente sigue una distribución Gamma



CLV

El arte de medir
PASIÓN Y PRECISIÓN

Fuente de datos

5.549.938
transacciones

1.727.844
clientes

42 *países*

38 *meses*

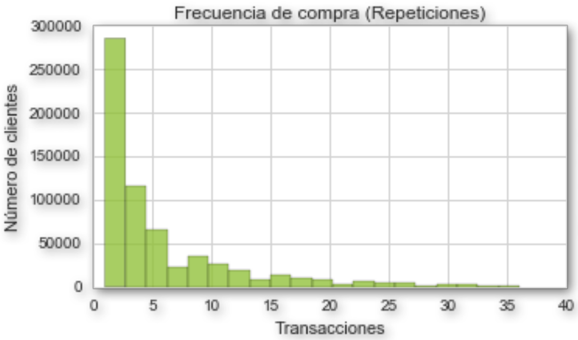
```sql
SELECT
  MD5(CAST(client_id AS STRING)) AS client_id,
  COUNT(distinct timeplaced) - 1 as frequency,
  TIMESTAMP_DIFF(PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MAX(timeplaced)),PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)), DAY) AS recency,
  TIMESTAMP_DIFF(TIMESTAMP("2017-12-31 00:00:00"),PARSE_TIMESTAMP('%Y-%m-%d %H:%M:%S',MIN(timeplaced)), DAY) AS T

FROM (
    SELECT
      client_id,
      orders_id,
      timeplaced,
      amount
    FROM
```

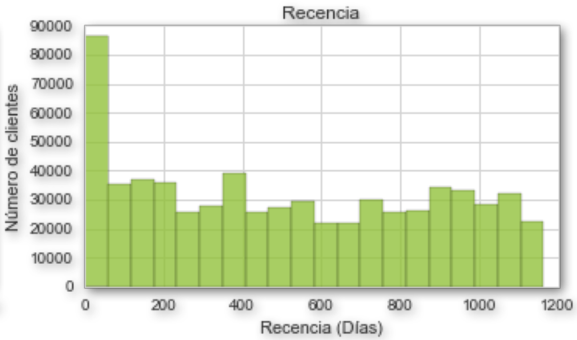
Standard SQL Dialect

RUN QUERY | Save Query | Save View | Format Query | Show Options

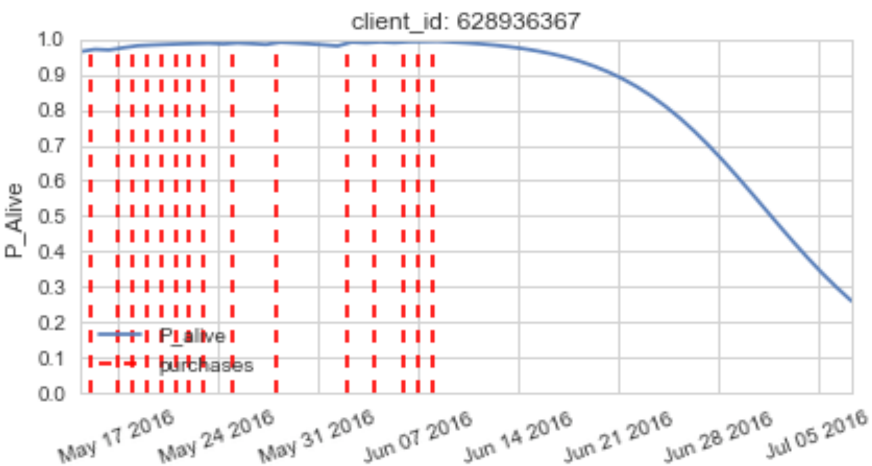
Results | Details | Download as CSV | Dow

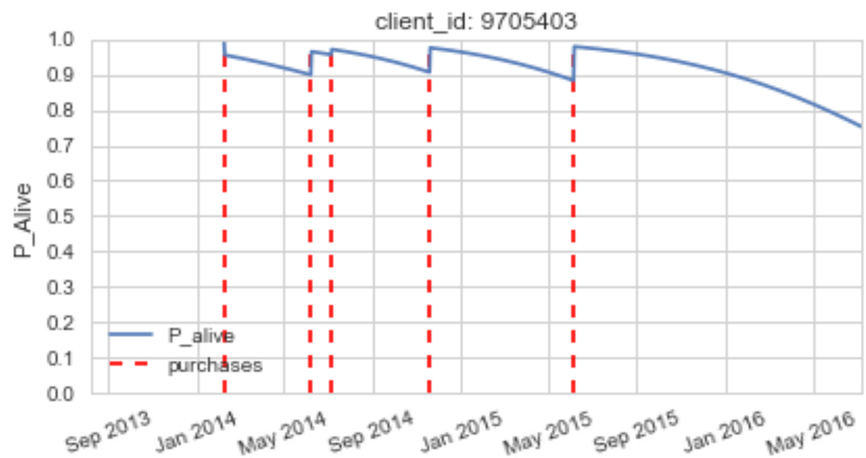
Row	client_id	frequency	recency	T
1	9rtoXFsm2W78qwSwS2rK6w==	85	117	2677
2	u08WSQjCwqg/oDuDGBFe5A==	73	1	2620
3	OHbyvsQQ0LOJ2ZzNmF0DHA==	65	106	2662
4	/Xq/W6VO0c4fnab5VNqveA==	49	119	2676
5	1grlumKuyU6uUdHHsE1Q0g==	42	120	2676
6	GMORTDintePHsIHUA17wAw==	39	68	2668
7	kEOLFD1VPlzTKm4vuqkflQ==	39	99	2665
8	epUnJ2X/KvqxpBqjqxdCHQ==	37	67	2624
9	YaZKFfOooCdu5VliNk3Qfg==	37	108	2666





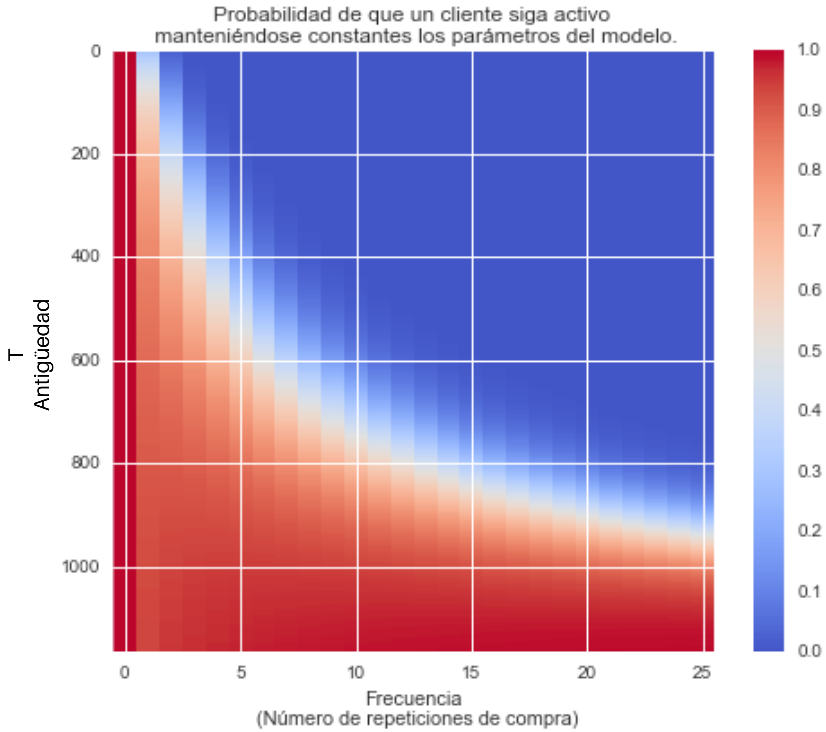
El hábito de compra del cliente

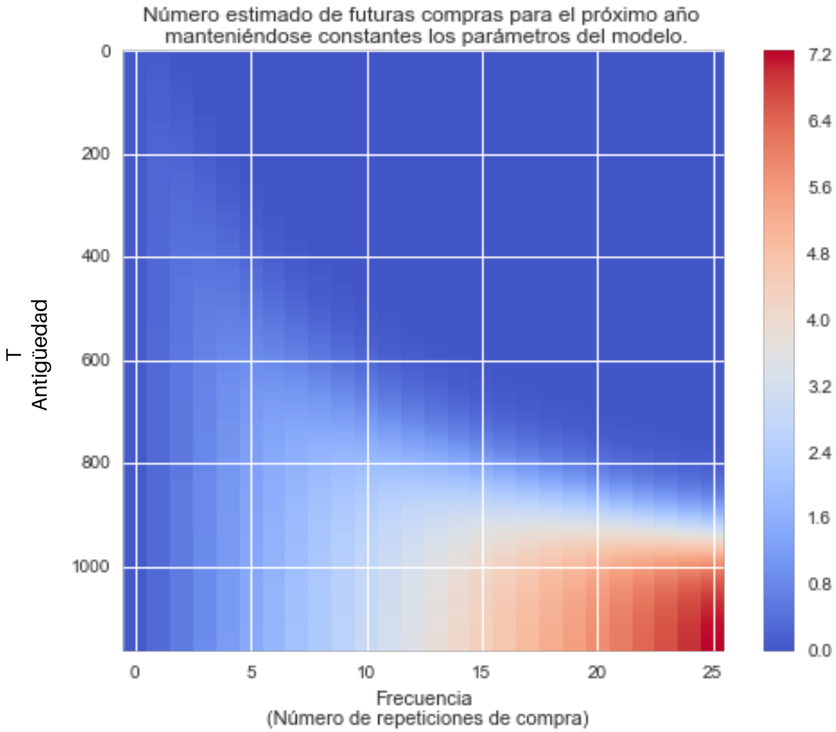




$$P(\tau>T|\lambda,\ \mu,\ x,\ t_x,\ T) = \frac{1}{1+\mu/(\mu+\lambda)[e^{(\lambda+\mu)(T-t_x)}-1]}\ .$$

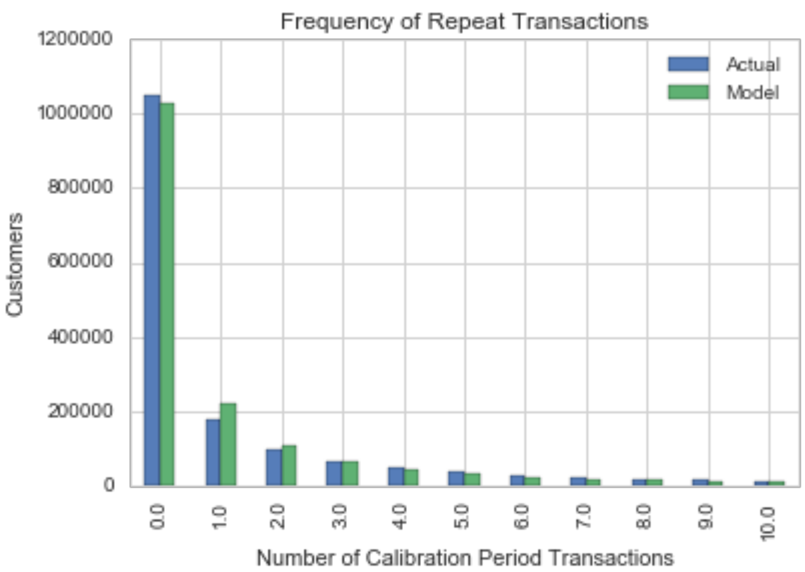
Resultados del modelo

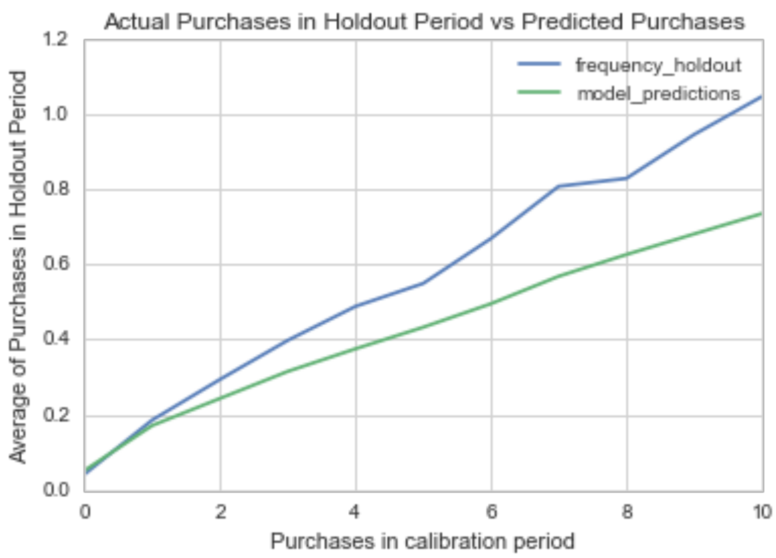




$$\mathscr{L}(\lambda,\ p\,|\,t_1,\ t_2,\ \ldots,\ t_{max(x)}) = (p + (1-p)e^{-\lambda(T - t_{max(x)})})((1-p)^{x-1}\lambda^x e^{-\lambda(t_{max(x)})})$$

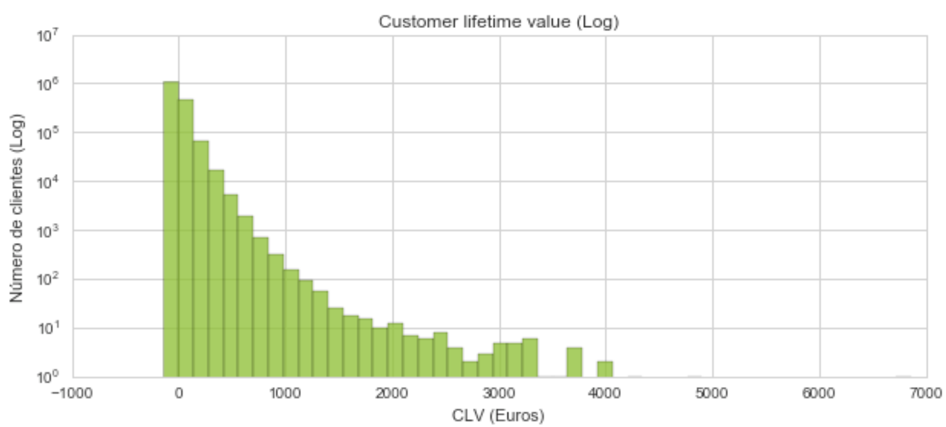
Validación del modelo





Resultado del modelo

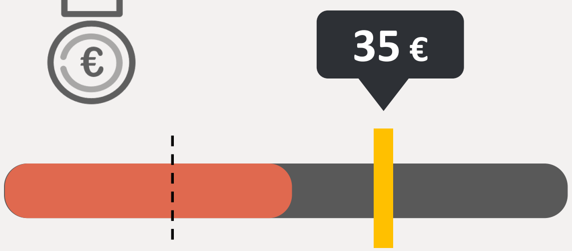
	frequency	recency	T	monetary_value	predicted_purchases	prob_alive	expected_profit	clv
client_id								
628936367	15.0	25.0	25.0	62.597333	50.03	0.989887	60.021445	2965.478239
9705403	5.0	669.0	1037.0	19.400000	1.39	0.822656	17.278515	23.769112



report

Valor de la vida del cliente

26,73€ ▲ 4% YoY



Vida del cliente

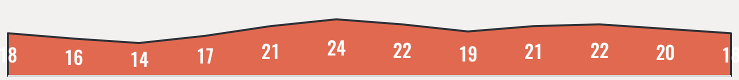
74 semanas ▲ 2% YoY

Ratio abandono clientes

25% ▼ 2% YoY

Repetición de compra media

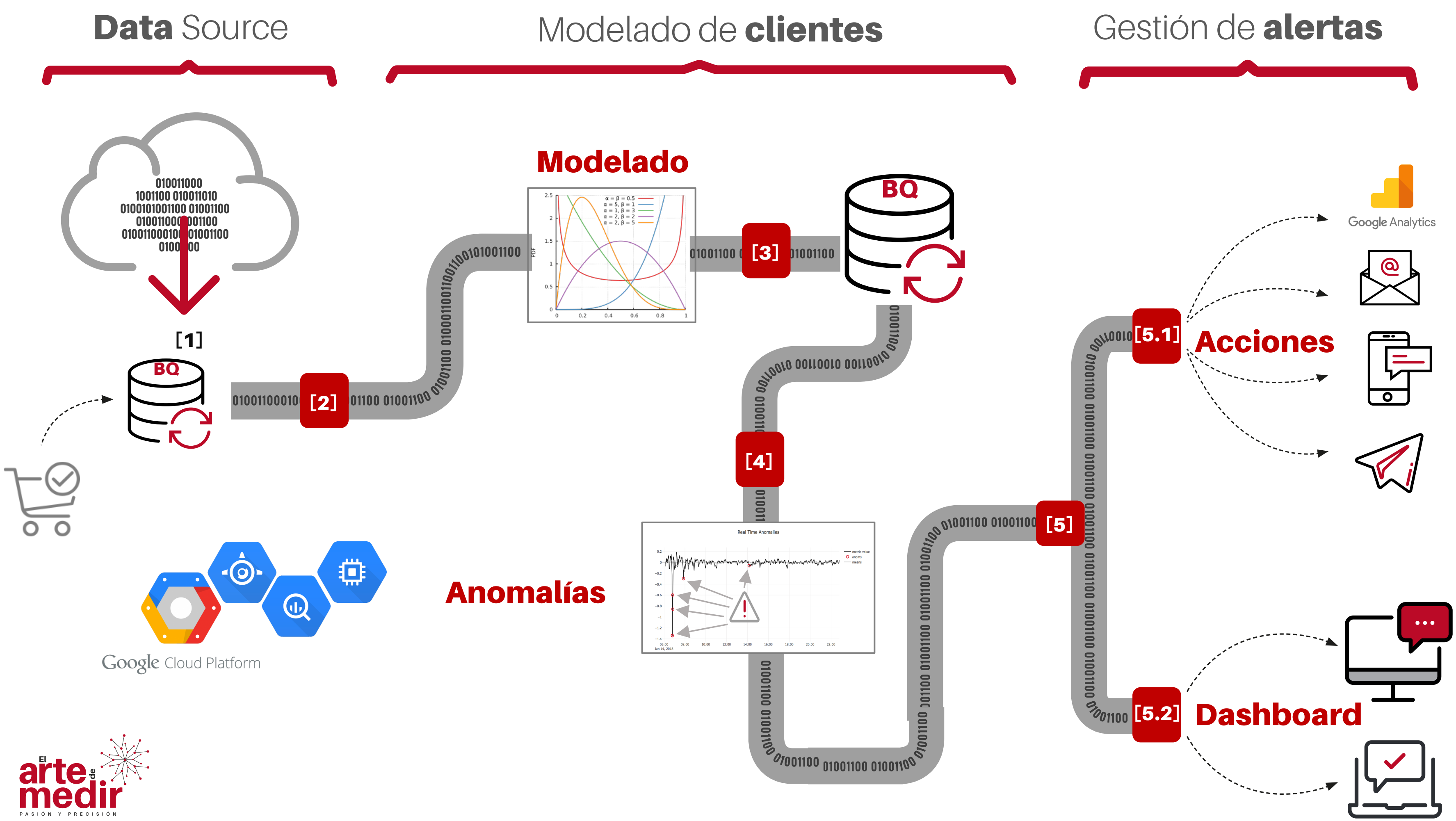
18 semanas ▼ 2% YoY



- Analizar ROI de campañas
- Optimizar presupuestos
- Planificación de medios
- Automatización Mkt
- Análisis de cohortes
- Segmentación

El arte de medir
PASIÓN Y PRECISIÓN

Arquitectura



GRACIAS

Jose Ramón Cajide

DATA SCIENCE & BIG DATA ANALYST

Master en Analitica Web por la Universidad British Columbia. Master en Analitica Web. Web Analytics Certification (Market Motive). Master en Data Science. Google Regional Trainer part of the Google Partner Academy Programme .

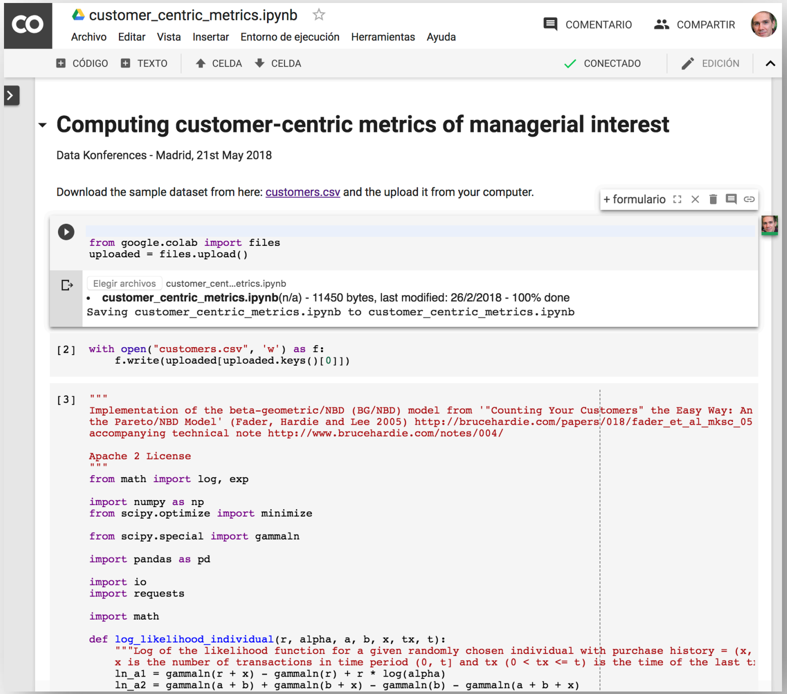
es.linkedin.com/in/jrcajide

@jrcajide

El arte de medir

PASIÓN Y PRECISIÓN

Notebook



https://goo.gl/ceK5bb

El arte de medir

PASIÓN Y PRECISIÓN