

Low-cost data science

Pedro Nevado Raja



Madrid, 28th March

- ➔ **Eight** different projects.
- ➔ **Seven** different clients.
- ➔ In **B2B** and **B2C** channels.
- ➔ Domains:
 - Interactions with web catalogues.
 - Live chats.
 - Communication of promotions: messages, channels, timing.
 - Transactions in loyalty and motivation programs.

What does low-cost mean?

➔ Low-cost budget

- From 5.000 to 18.000 euros.

➔ Low-cost software

- Statistical libraries in programming languages such as Julia, R or Java.

➔ Low-cost equipment

- A desk-top computer with a basic configuration.

➔ Low-cost cloud

- For our *Marketing Workbench* project: a cluster of two servers and a shared relational database in AWS (42 euros/month).

What does low-cost mean?

➔ Low-cost data: we use what the customer has available.

- Customer's ERP data: invoices, reservations, ...
- Google Analytic files from content management systems.
- Transactions in loyalty and motivation programs applications: accumulation, redemption or cancellation.
- Tickets of live chats.

Low-cost means no generation of new data for the sake of an analytical project.

→ Quantities

- Up to a maximum of **1 million** records in a file.

→ Complexity

- Up to **54** different **kinds** of data in the same file.
- Up to **7** different file **formats** in the same project.
- Up to **49** different **files** in a project.

➔ Statistical learning

- We try to identify **patterns** in the data using statistical techniques.
- We aim for patterns with **predictive power**.
- We anticipate probable outcomes: knowledge as a **prelude to action**.

➔ Techniques

- Linear and logistic regression, K-clusters and association rules.
- Regression trees, random forests and gradient boost.
- NLP: Word2Vec and Azure Text Analytics.

Low-cost projects

Problems, results, decisions

→ The Set-Up

- A web catalogue of + 2000 pages, for +1000 sales operators, with +30 different professional profiles and + 0.5 million of visits per month.

→ The Problem

- How to shorten the response times of the operators to the final clients making the enquiries.

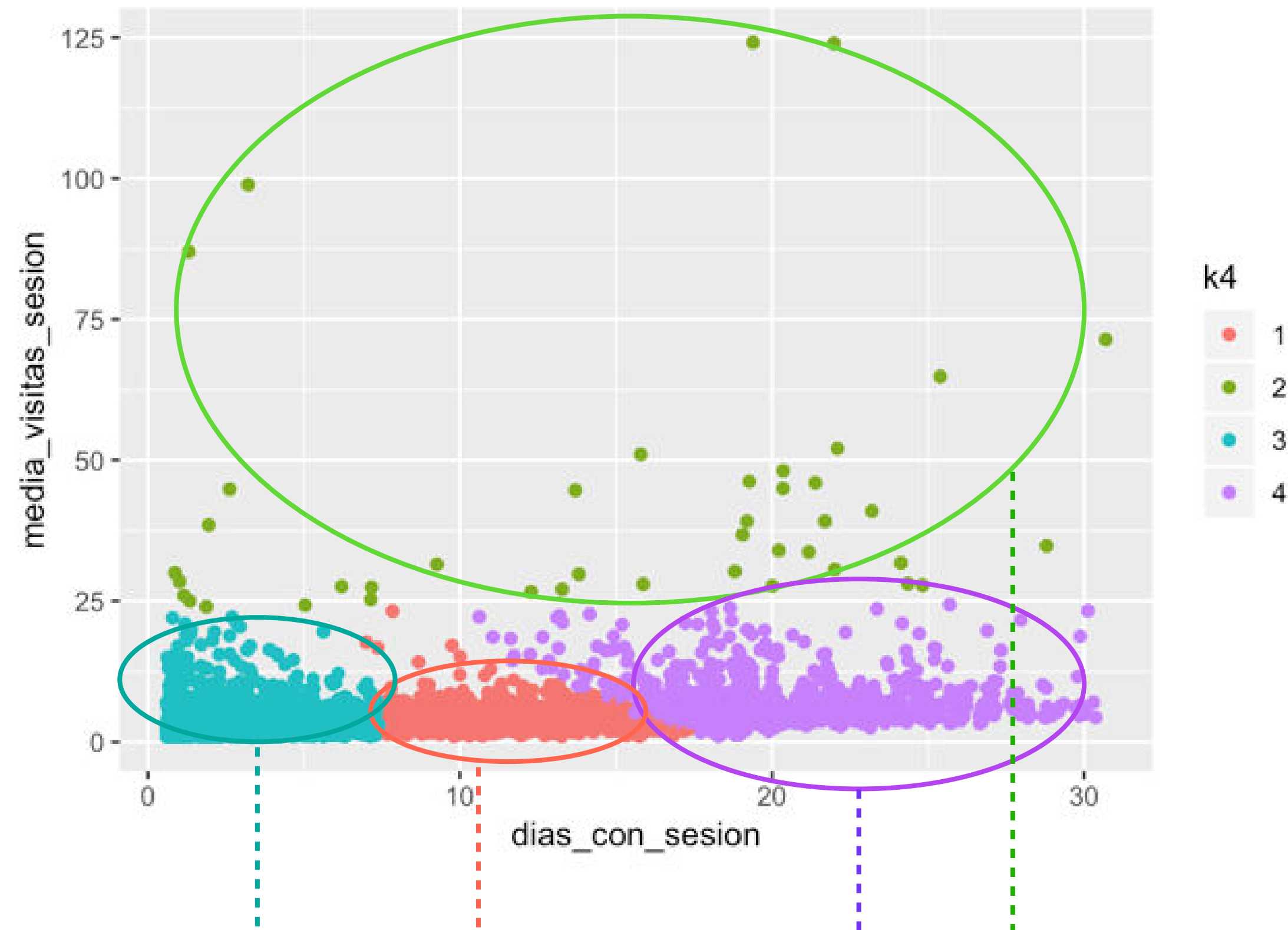
→ The Results

- We identify patterns of use of the catalogue for the different profiles and estimate the most probable pages to be visited for each of them.

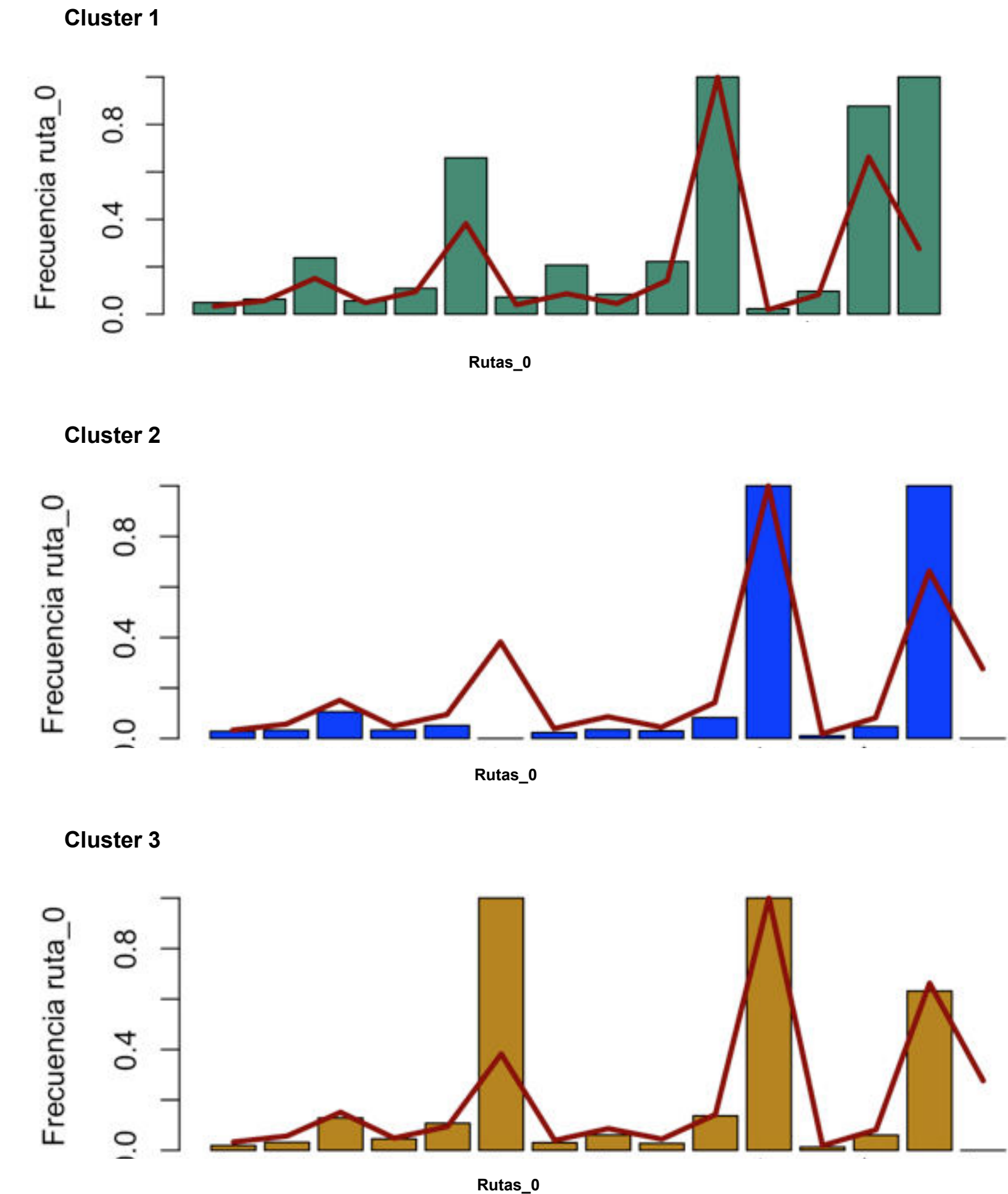
→ The Decisions

- *Short-cuts* and *Favourites* navigation options were defined for each of the profiles in the Catalogue.

Problems, results, decisions (1)



Clusters:
different visited pages in a navigation session
vs.
days of the month with navigation sessions.



Frequency, by number of visits in a month, of the main routes of the catalog.

➔ The Set-Up

- 180.000 bills invoiced in a period of 42 months in a tourist resort.

➔ The Problem

- How to design a loyalty program with that information.

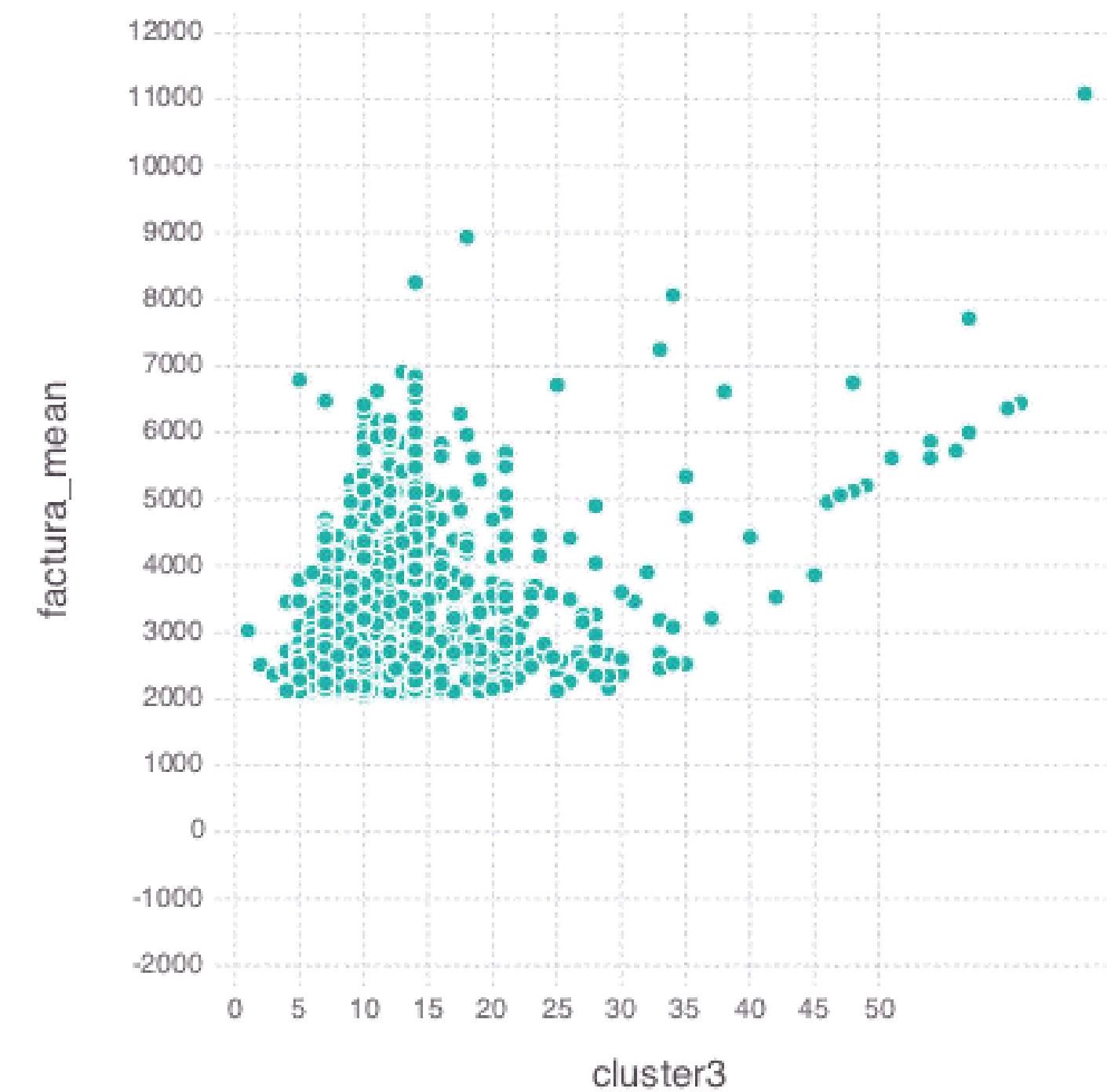
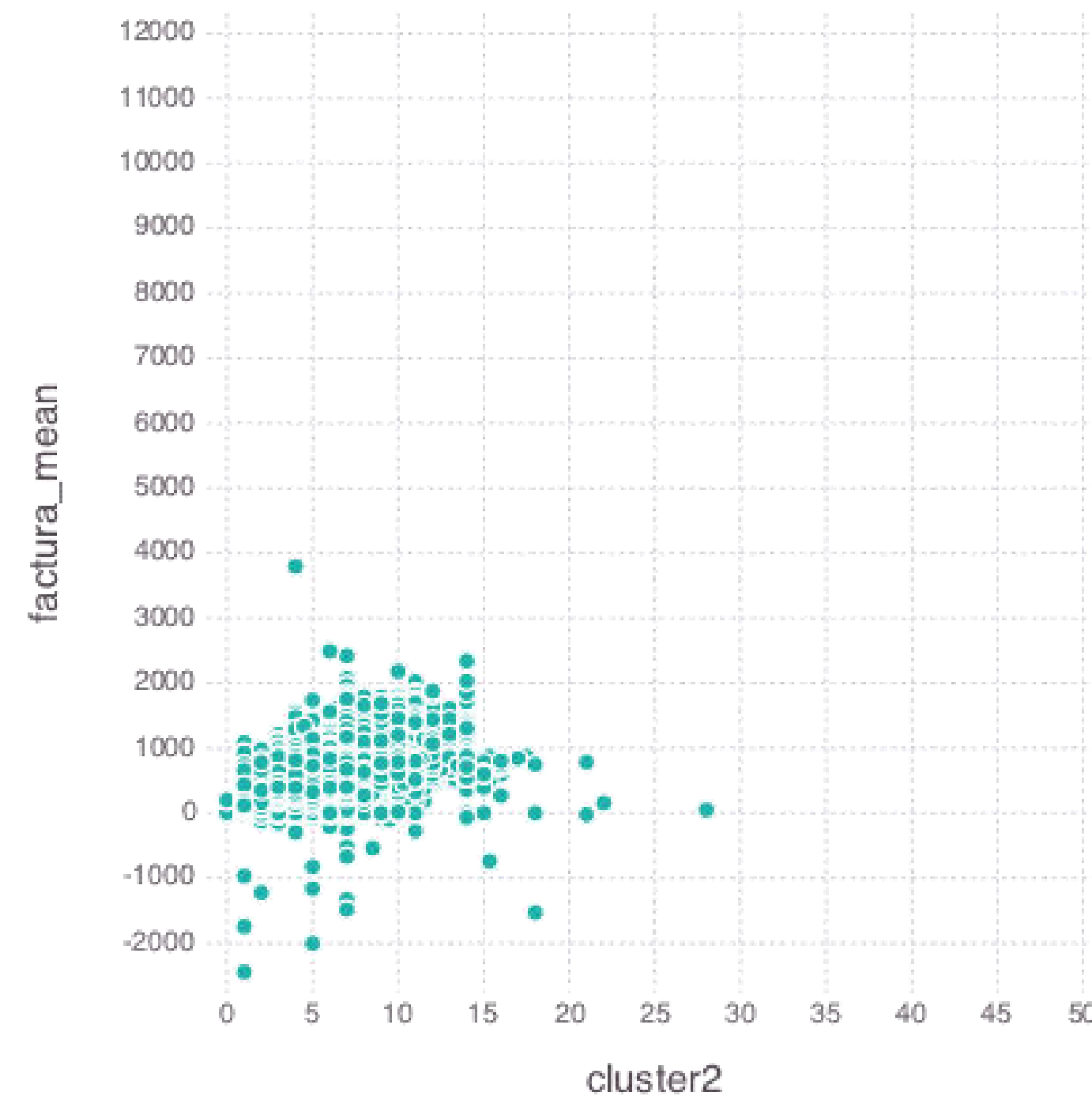
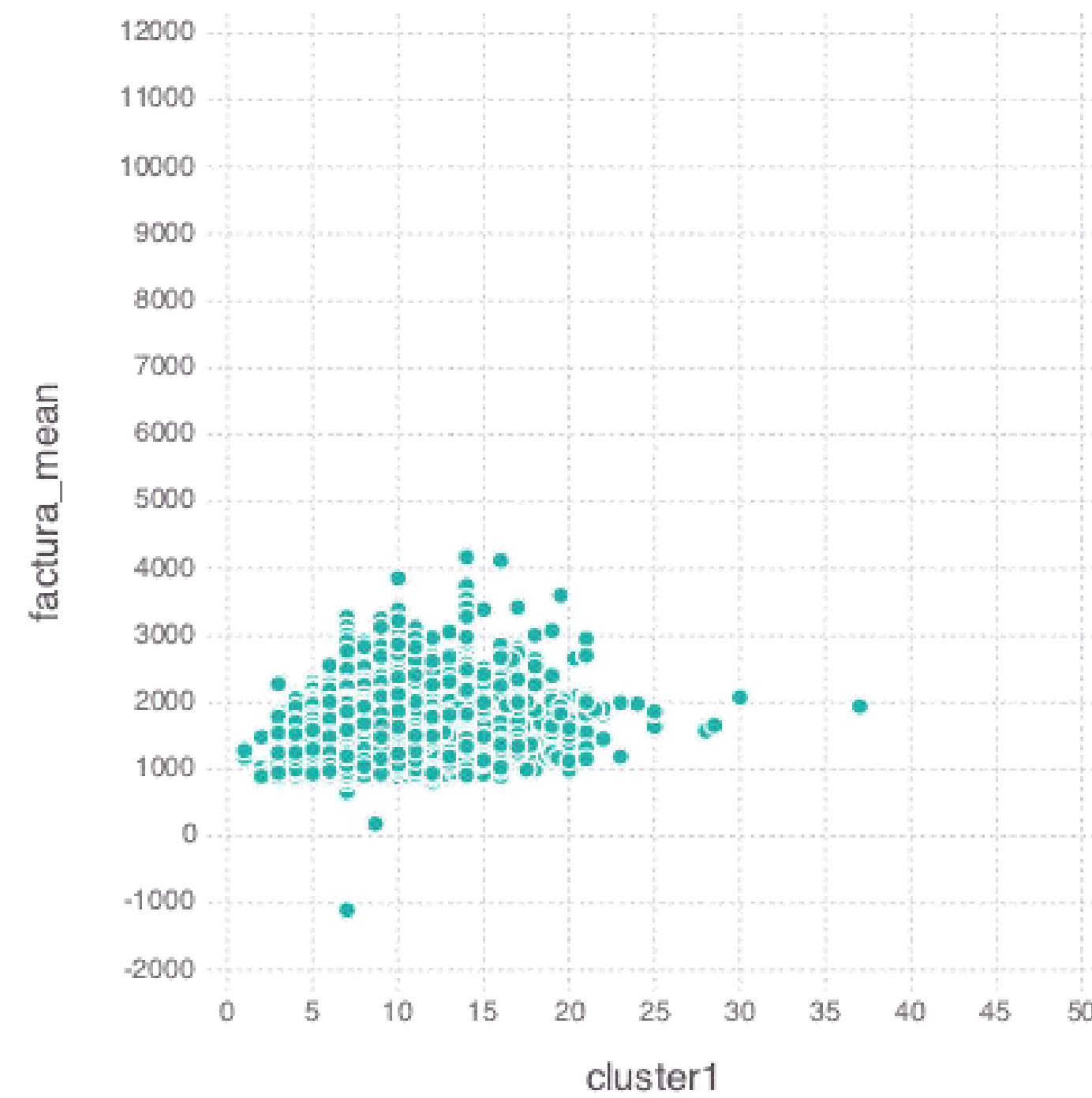
➔ The Results

- We identified 80.000 customers, characterised by 22 variables based on 54 different billing items.
- We clustered the customers in three homogenous groups by their patterns of consumption and certain demographic indicators.

➔ The Decisions

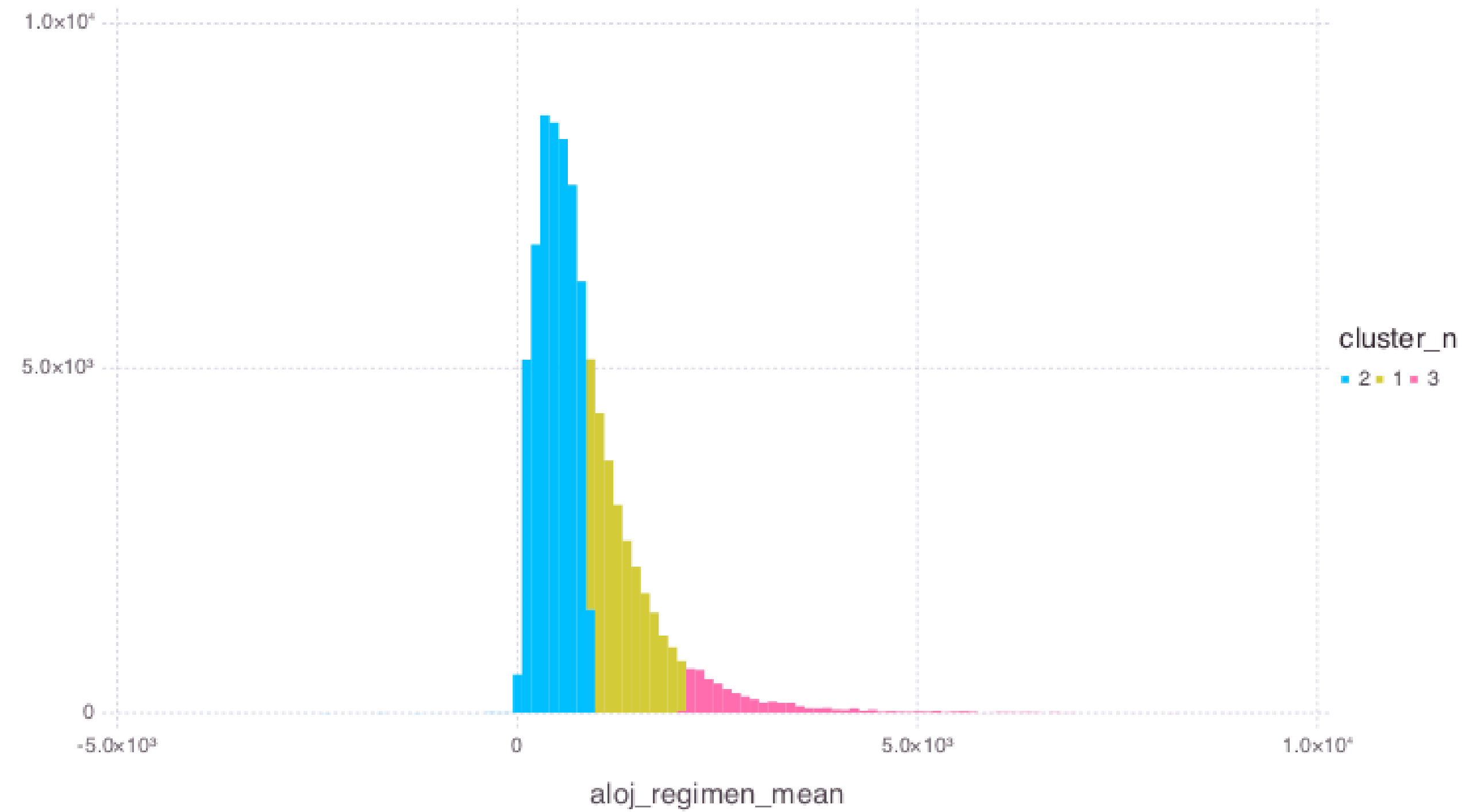
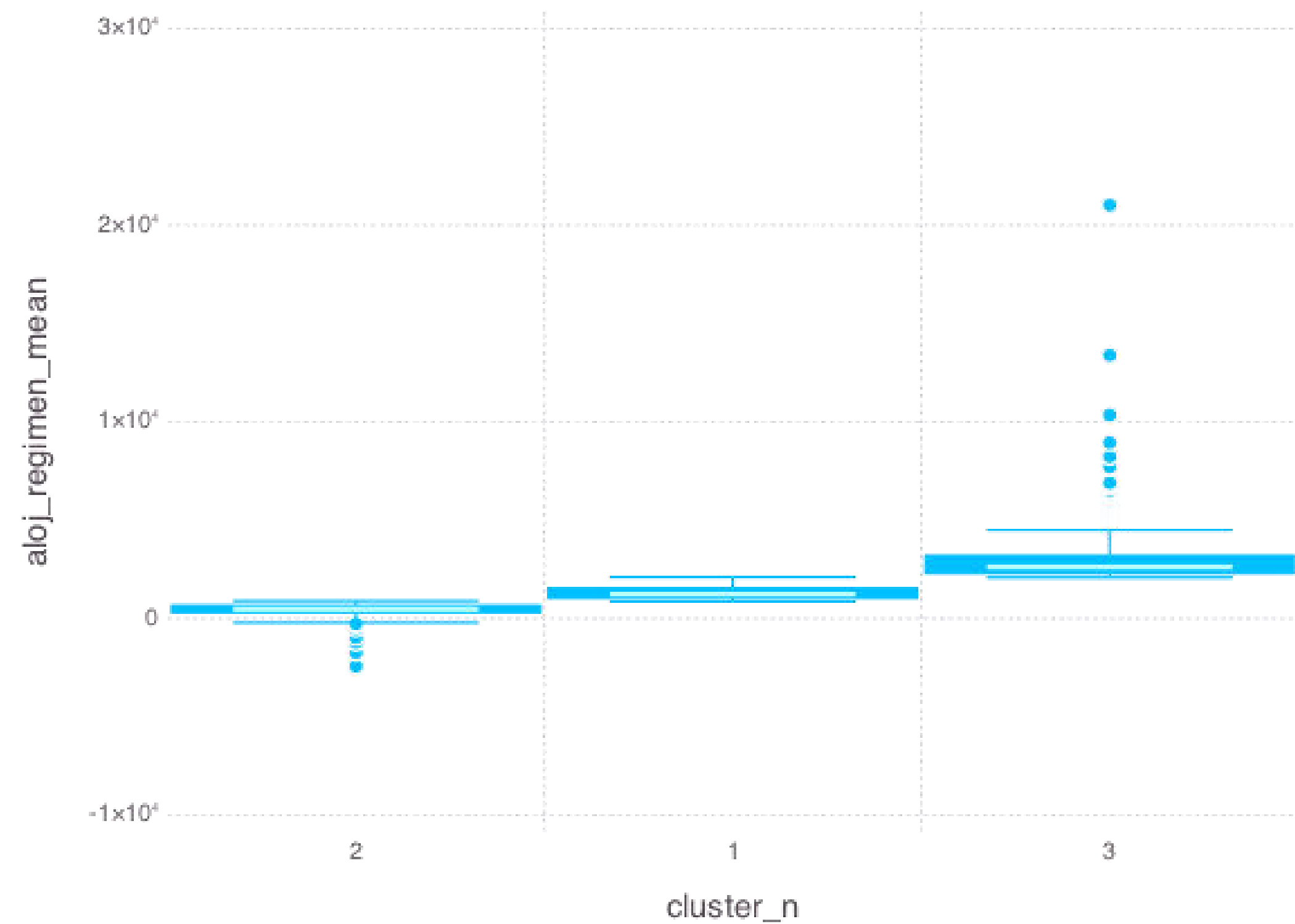
- A loyalty program was designed with incentives taken from the services catalogue of the resort.

Problems, results, decisions (2)



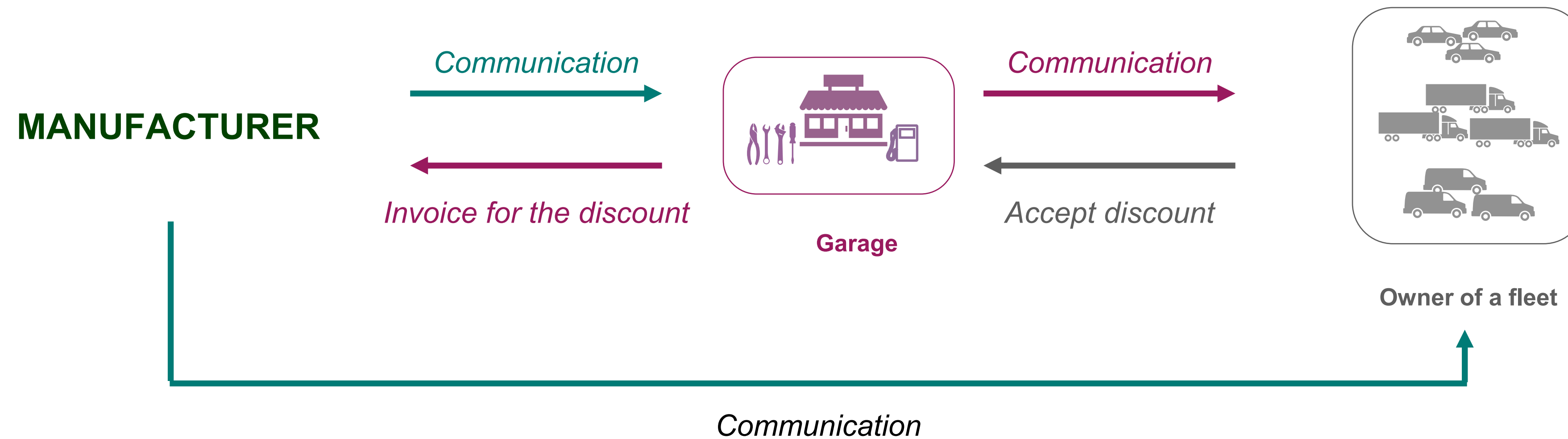
Clusters of customers by the average expenditure
(The cluster with less amount in the middle)

Problems, results, decisions (2)



Clusters of customers by the average expenditure
(The cluster with less invoiced amount: number 2)

Problems, results, decisions (3)



➡ The Set-Up

- 40.000 invoices over a period of 30 months: the garage invoices the manufacturer for the discount made to the final buyer.

➔ The Problem

- What factors explain the promotional behaviour of garages and fleets.

➔ The Results

- We identified approx. 1000 garages and 6000 fleets.
- For garages, the main motivation for participating was to increase their customer base, rather than increasing the loyalty of their present customers.
- For fleets, the communication of garages seemed to outweigh the direct communication from the manufacturer.

➔ The Decisions

- An increase of the materials of communication on the point of sale and a shift of the focus of the messages to garages to the acquisition of new customers.

Problem, results, decisions (3)



taller	camapaña	antig	nv_cli	med_nv_cli	hist_nv_cli	vts	med_vts	hist_vts	facturas	fact_deng	denegaciones
185226	x25	126	1	0	0,00	4.122	0	0,00	1	0	0,00
185226	x25 sep	236	2	1	1,00	16.764	4.122	1,00	3	0	0,00
185226	x25 marzo 2017	404	2	2	1,00	30.900	10.443	1,00	7	0	0,00
185226	tarjeta x20 junio 2017	465	0	2	1,00	0	17.262	1,00	0	0	0,00
185226	x30 septiembre 2017	602	0	1	0,75	0	12.946	0,75	0	0	0,00
185226	x25 febrero 2018	762	0	1	0,60	28.280	10.357	0,60	2	2	0,00
185226	x25 junio 2018	861	0	1	0,50	0	13.344	0,67	0	0	0,15
185226	x25 septiembre 2018	966	4	1	0,43	17.506	11.438	0,57	6	0	0,15
185226	x20 marzo 2019	1.133	1	1	0,50	13.744	12.196	0,62	5	0	0,11

Definition of variables based on past behaviour

prob	antig	med_vts	hist_vts	med_nv_cli
0,69	258	123.970,0	1	13,0
0,94	426	272.760,5	1	18,5
0,90	487	240.242,3	1	14,0
0,86	624	223.349,8	1	11,0
0,94	784	292.970,0	1	11,8

Estimations of the probability of active participation in a promotion of each of the garages

prob	antig	med_vts	hist_vts	med_nv_cli
0,20	244	5.744,0	1	2,0
0,66	503	81.966,7	1	12,3
0,79	899	103.253,3	1	12,7
0,16	243	2.039,0	1	0,0
0,25	411	21.395,0	1	2,0

→ The Set-Up

- A manufacturer with a portfolio of +25 product categories, +15 types of franchises, 3 communication channels and +12 different kind of incentives for sales.

→ The Problem

- How to optimise the design of the promotional formula
products-campaign + incentive + channel-message → type of franchise.

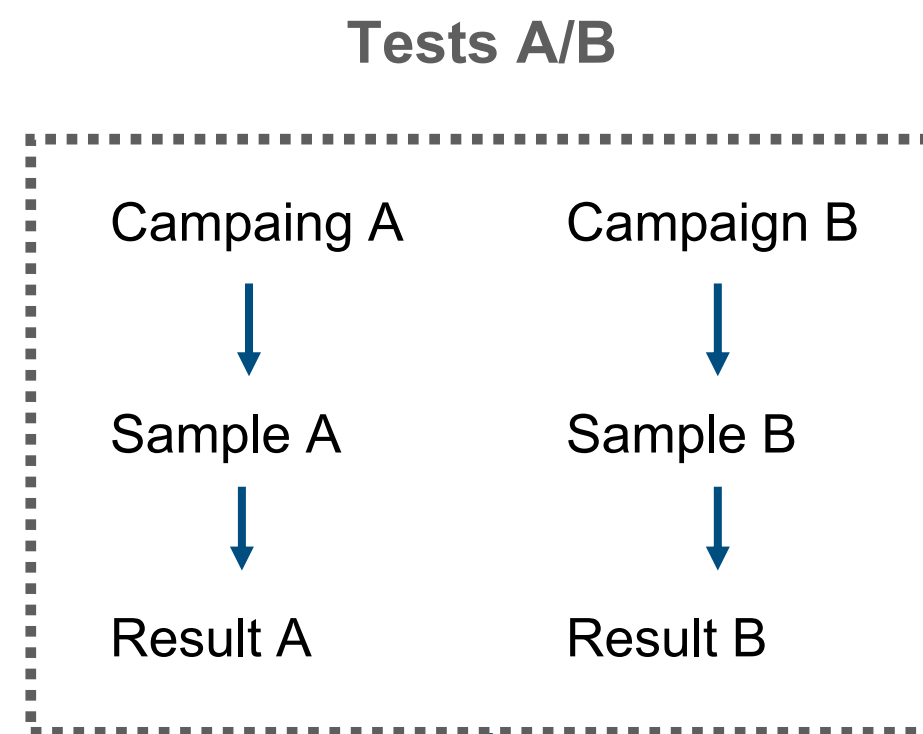
→ The Decisions

- The development of a marketing 'workbench': a cloud application for experimentation and simulation.

Problems, results, decisions (4): experimentation

Marketing workbench

Method



Statistical contrast:

- Average daily sale per participant.
- Is it statistically significant with a t-distribution test?

Campaign A

- Start date
- End date
- Markets
- Concepts
- **Products in campaign**
- **Incentive**
- **E-mail (yes/no)**
- **SMS (yes/no)**
- **Message**

Campaign B

- Shared dates, markets and concepts with Campaign A
- **Products in campaign**
- **Incentive**
- **E-mail (yes/no)**
- **SMS (yes/no)**
- **Message**

Experiment

Download samples

- Random samples stratified by **concepts** and **markets**.

Upload participant files

Upload sales files

Results

- **Average daily sales** per participant and product in each campaign.

Statistical contrast

- **Probability** that the **difference between the means** is not statistically significant.

Problems, results, decisions (4): experimentation

DATOS DE CAMPAÑAS

CAMPAÑA A

Nombre

exp2_febrero_2021_A

Entre 5 y 25 caracteres; letras, números, guión bajo; sin espacios.

Productos

Seleccione hasta 3

Sistemas de frenado
Iluminación
Encendido

Incentivo

Descuento en metálico

Comunicación

sms

Texto comunicación

Texto del mensaje SMS.

Máximo 500 caracteres.

CAMPAÑA B

Nombre

exp2_febrero_2021_B

Entre 5 y 25 caracteres; letras, números, guión bajo; sin espacios.

Productos

Seleccione hasta 3

Sistemas de frenado
Iluminación
Encendido

Incentivo

Tarjetas regalo

Comunicación

email

Texto comunicación

Texto del mensaje para correo electrónico.

Máximo 500 caracteres.

OPCIONES DE PARA EXPERIMENTAR

Venta cruzada de productos

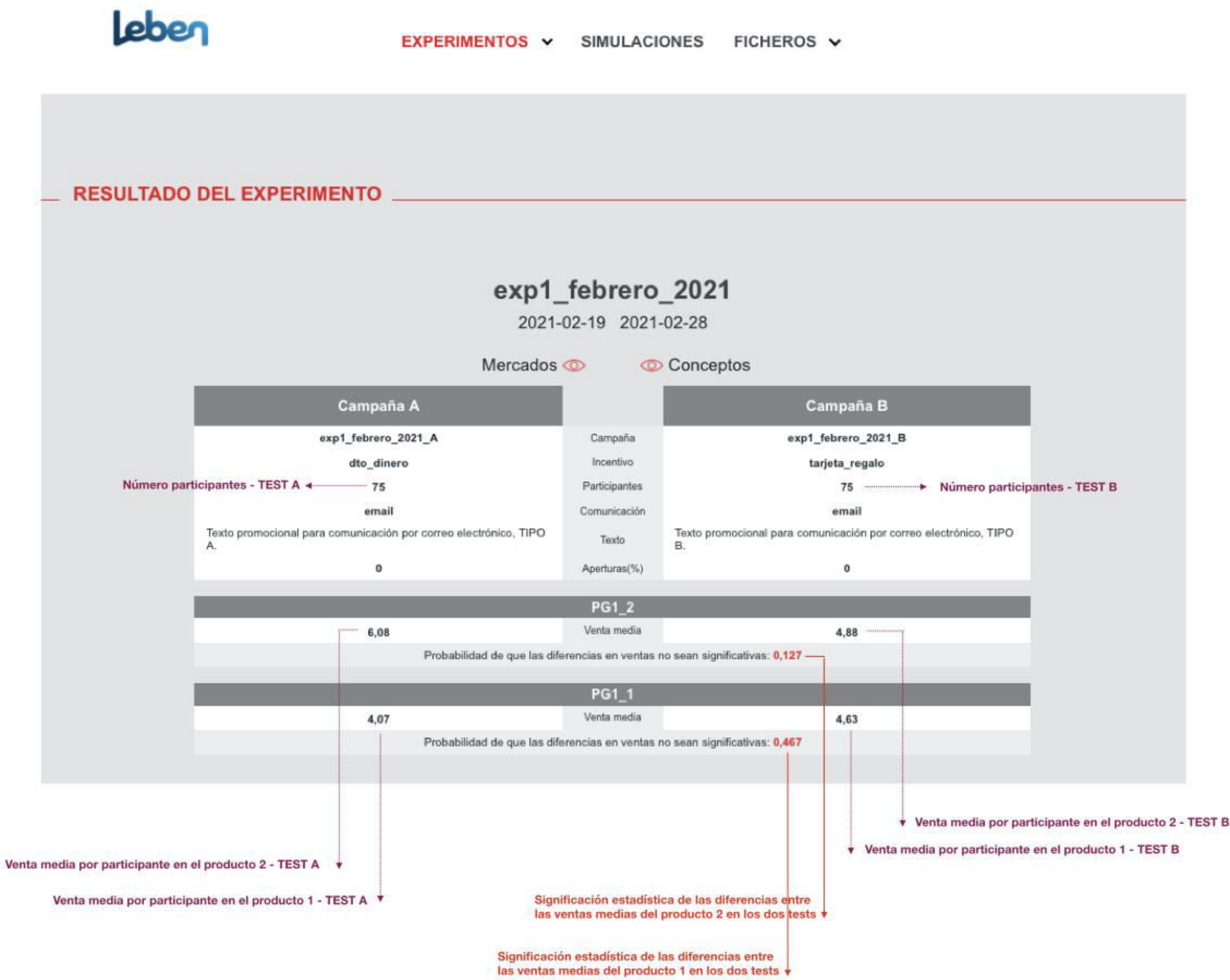
Tipo de incentivo

Canal de comunicación

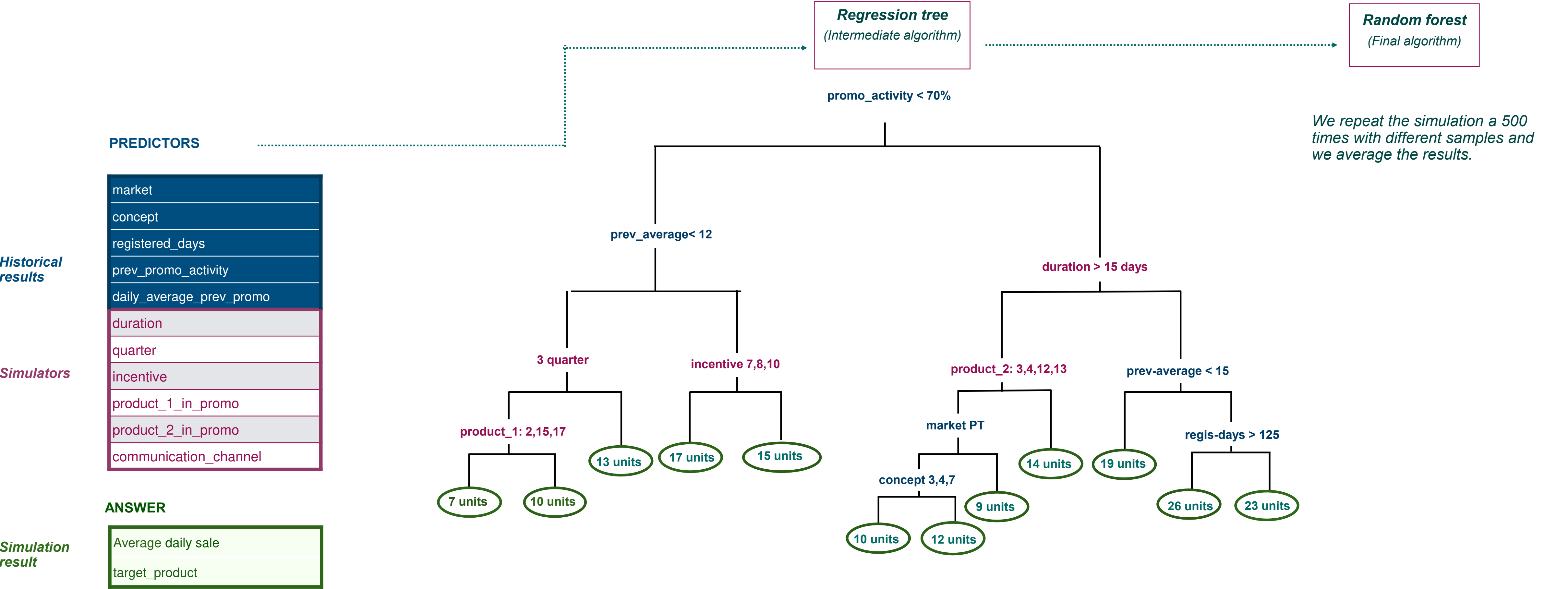
OBTENER MUESTRAS

Muestra test A

Muestra test B



Problems, results, decisions (4): simulation



Problems, results, decisions (4): simulation

SIMULACIÓN DE RESULTADOS DE UNA CAMPAÑA

DATOS SIMULACIÓN

Nombre

promo_simulada_marzo_1

Entre 5 y 25 caracteres: letras, números, guión bajo; sin espacios.

Fecha de Inicio

01-03-2021

Formato de fecha dd-mm-aaaa

Fecha de Fin

10-03-2021

Formato de fecha dd-mm-aaaa

Mercados

Seleccione al menos uno

España

Portugal

Andorra

Conceptos

Red_ventas_1

Red_ventas_2

Red_ventas_3

Red_ventas_4

Productos

Seleccione hasta 3

Sistemas de frenado

Iluminación

Encendido

Incentivo

Descuento en metálico

Comunicación

email

Texto comunicación

Texto para mensaje de comunicación por correo electrónico, promoción simulada.

• Duración

• Período (estacionalidad)

Selección de participantes

• Mercado

• Red de ventas

Incentivo

Productos en campaña

Canal de comunicación

Mensaje promocional

SIMULAR



Some insights from our projects

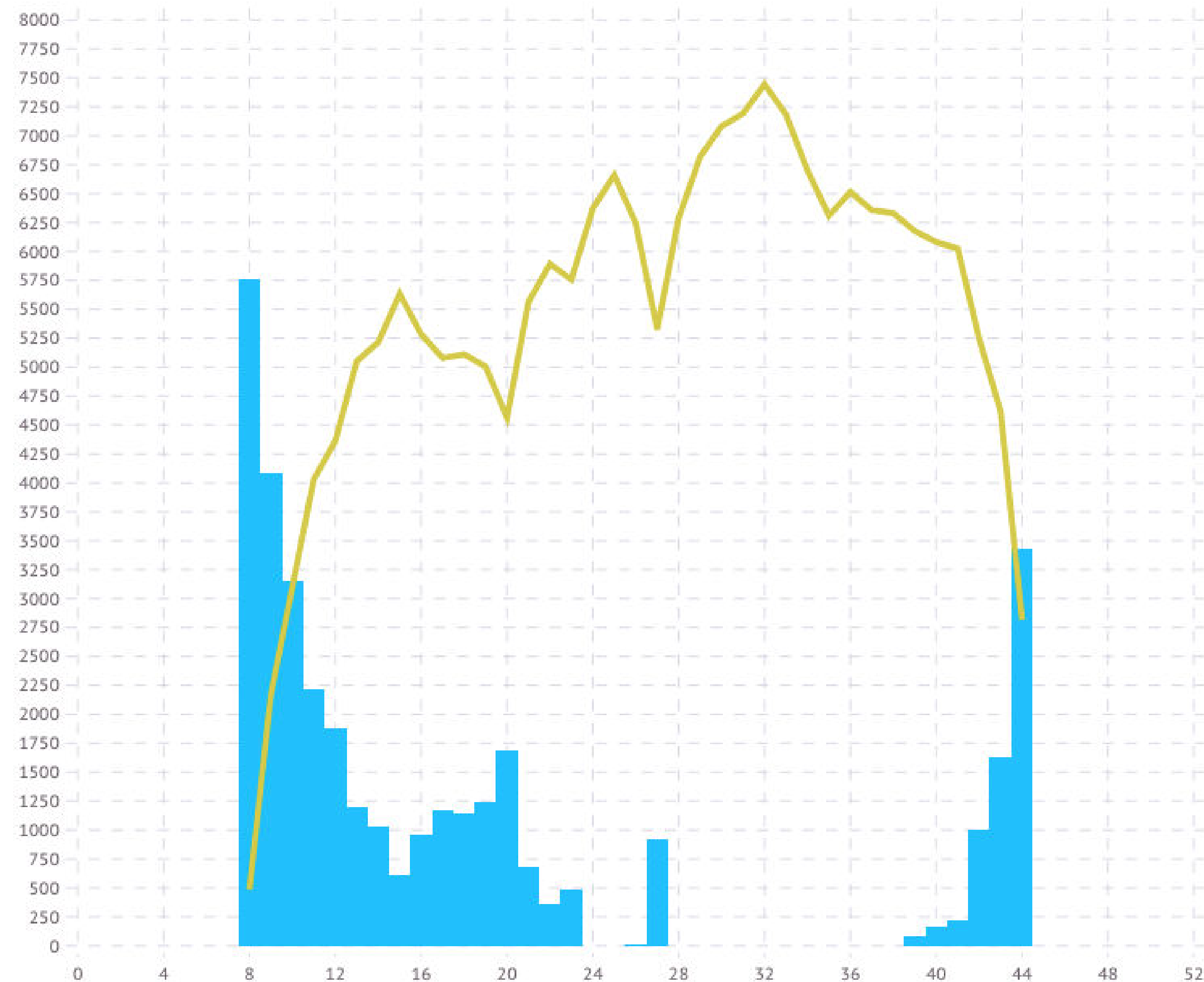
(In friendly auto-imperative form)

Some insights from our projects (1)

- ➔ Integrate vertically your incentives in the value chain of your customer.
 - In a tourist resort, better a free meal, a free drink or a free spa than a bouquet of flowers.
- ➔ Aim for incentives without opportunity cost (the loss of a potential sale/profit for the supplier).
 - At full occupancy:

the cost of a free hotel night = *the maintenance and service cost of the room*
the loss of a probable sale of the room

Some insights from our projects (1)



In **blue**, nights with **zero** opportunity cost
(in each week of the year)

room-nights invoiced vs. room-nights available (per week)

Some insights from our projects (2)

- ➔ Identify periods and products with temporary **capacity slacks** and adjust the accrual and the redemption to them.
 - ◉ Capacity slacks reduce the opportunity costs in the provision of an incentive/gift; specifically in the presence of fixed costs.
- ➔ In complex channels of communication, aim for the **nearest position**, in time and space, to the decision.
 - ◉ The power of communication of a message increases with the nearness to the decision maker.

Some insights from our projects (3)

→ Experiment

- Don't wait for an event to happen spontaneously: make it happen and analyse the results.
- Experimentation is a quick way to learn.

→ Simulate

- Leverage your past investment in experimentation and your accumulated historical knowledge.
- Simulation is even cheaper and quicker than experimenting.

Epilogue

A clue for the quiz

➔ Questions are welcome

⦿ ...

Pedro Nevado Raja

pnevado@grupoleben.com

pedro@didekin.es

www.github.com/didekin