masempresas.cea.es
#masempresas

Big Data y Data Science

Una herramienta para la Caracterización de Mercados

Financiado por:

CEA
Confederación de Empresarios de Andalucía

Junta de Andalucía
Consejería de Transformación Económica, Industria, Conocimiento y Universidades

Fomento de la Cultura Emprendedora

Breve Trayectoria Ponente

- José Ignacio Peláez Sánchez.
  - Universidad de Granada, Jaén y Málaga.
- Actualidad. Universidad de Málaga:
  - Catedrático en Computación. Universidad de Málaga.
  - Subdirector de Ecosistemas Digitales del Centro de Investigación Social Aplicada (CISA). www.cisa.uma.es
  - Director del Aula de Métricas y Gestión de Intangibles.
- Miembro del Comité Científico de la Fundación Corporate Excellence. Centrer for Reputation Leadership.
  - Approaching the Future. Informes INTED.
- Fellow Senior en el Programa Fellowship CENTRA.
- Formación Empresas:
  - ITAIPU Binacional.
- Desarrollo de Sistemas:
  - Smart Data Turístico. Turismo Andaluz. Junta de Andalucía.

Objetivos del Webinar

- Qué es el:
  - Big Data,
  - Data Science,
  - Business Intelligence,
  - Business Analysis ,
  - Machine Learning,
  - Inteligencia Artificial…?

- Elaboración de un Proyecto de Data Science.
- Herramientas.
- Caracterización de la sociedad/consumidores.

DATA MINING
BIG DATA
DATA SCIENCE
MACHINE LEARNING
BUSINESS INTELLIGENCE
DATA WAREHOUSE
DEEP LEARNING
REGRESION LOGÍSTICA
REDES NEURONALES
INTELIGENCIA ARTIFICIAL
DIGITAL MARKETING
OBSERVACIONES
DIMENSIONES
ATRIBUTOS
REPORTES
DASHBOARDS
HECHOS
CUBOS
KPIS
DATOS
MACHINE LEARNING
DATA WAREHOUSE

Inteligencia Artificial & Machine Learning & Deep Learning

- Inteligencia Artificial.
  - Comienzo años 50 del pasado siglo
  - Objetivo de hacer que las máquinas pensarán como los humanos o similar a ellos.
  - IA Simbólica. Sistemas expertos.
  - IA Conexionista. Aprendizaje de patrones.
    - Machine Learning: K-vecinosde decisión, SVM, RNA, Deep Learning.

BIG DATA & DATA

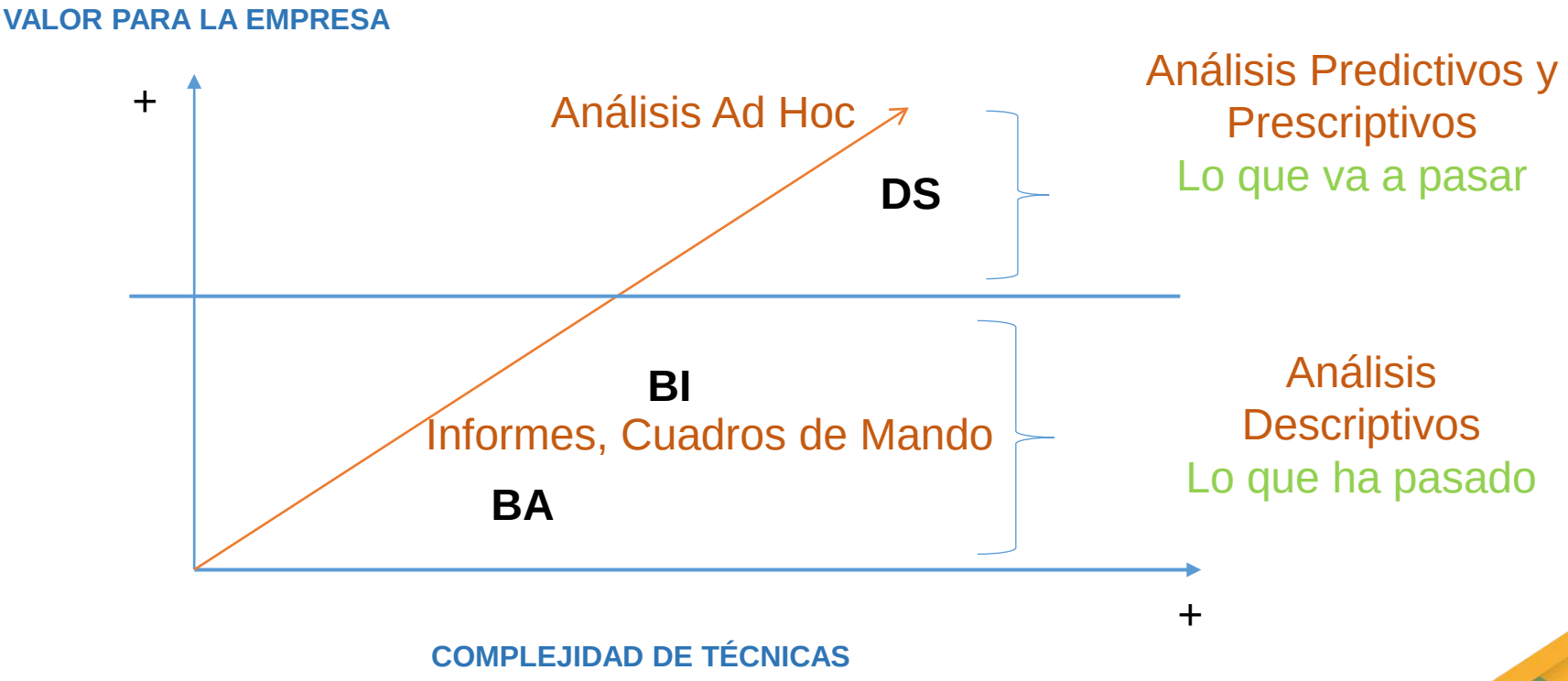
- Big Data.
  - Análisis de grandes volúmenes.
  - Se encarga de analizar grandes volúmenes facilitando las arquitecturas para ello.

- Data Science.
  - Parte analítica de datos.
  - Puede ser vista como evolución del término Minería de Datos.

Business Analysis & Business Intelligence & Data Science

- Datos:
  - Análisis Descriptivos.
    - BA: Trabaja directamente en las áreas de negocio; análisis no demasiado complejos; respuesta a las preguntas del día a día.
      - Pe. Análisis de precios respecto a la competencia, etc.
      - Usuario muy pegado al negocio sin excesivo conocimiento técnico.
      - Utiliza herramientas como hojas de cálculo, powerpoint, QlikView, MicroStrategy etc.
    - BI: Conocimiento del negocio, pero con una componente más técnica.
      - No tiene que dar respuestas del día a día.
      - Tiene una visión más a largo plazo de los datos, automatizando los procesos, creando los cuadros de mando.
      - Conocimientos de SQL, Qlikview y MicroStrategy en más profundidad.
  - Análisis Predictivos y Prescriptivos.
    - DS: Conocimientos técnicos superiores.
      - Análisis predictivos y prescriptivos en función de los datos.

Business Analysis & Business Intelligence & Data Science



Tipos de proyectos

TIPOS DE PROYECTOS

- Podemos agruparlos en dos tipos:
  - Automatización.
    - Conocemos el sector, las variables, etc.,
    - Objetivo aumentar la productividad implantando un proceso.
    - Modelos Predictivos o Simulación.
  - Discovery.
    - Se localizan mejoras.
    - Búsqueda de resultados estratégicos para el negocio.
    - Genera un impacto cualitativo. Ejemplo. Utilización de hospital.

Fases de Proyectos Discovery

- Fase 1. Filtrado. Aglutinar toda la información posible y posteriormente filtrar la importante.
  - En esta fase existe un proceso muy importante que es denominado “Desambiguación”,
- Fase 2. Estructuración.
  - Estructuración de la información en tablas para el análisis.
- Fase 3. Análisis inside.
  - Se hacen preguntas relevantes para el negocio.
  - Se obtienen conclusiones o patrones de negocio.
- Fase 4. Comunicación.
  - Visualización e informes.
    - No hacer copiar y pegar de las gráficas.

Pasos para un Proyecto de Data Science

CALIDAD DE LOS DATOS

- La calidad de los datos es un punto crítico para el éxito del proyecto.
- Es preciso hacer un análisis de calidad de los datos.
- Pasos para realizar un análisis de Calidad de Datos:
  a. Identificar Datos Nulos (faltantes o no informados) y Atípicos.
  b. Estadísticos Básicos: Media, mínimo, máximo, mediana, valores negativos, rango de la variable, etc.
  c. Análisis Longitudinal de Variables.
     i. En negocios con clientes con relación continuada en el tiempo. Analizamos los datos en el tiempo, para ver variaciones en las mismas.
  d. Coherencia entre Variables.
     i. Pe. Importe de compra del cliente respecto al número de compras.

Transformación de Datos

- Es uno de los aspectos que mejor distingue a un Científico de Datos ya que supone el 80% del trabajo.
- Transformar los datos se hace en dos dimensiones:
  - Estructurar los datos para adaptarlo a los algoritmos.
  - Creación de variables sintéticas/derivadas ajustadas al negocio.

Variables Sintéticas

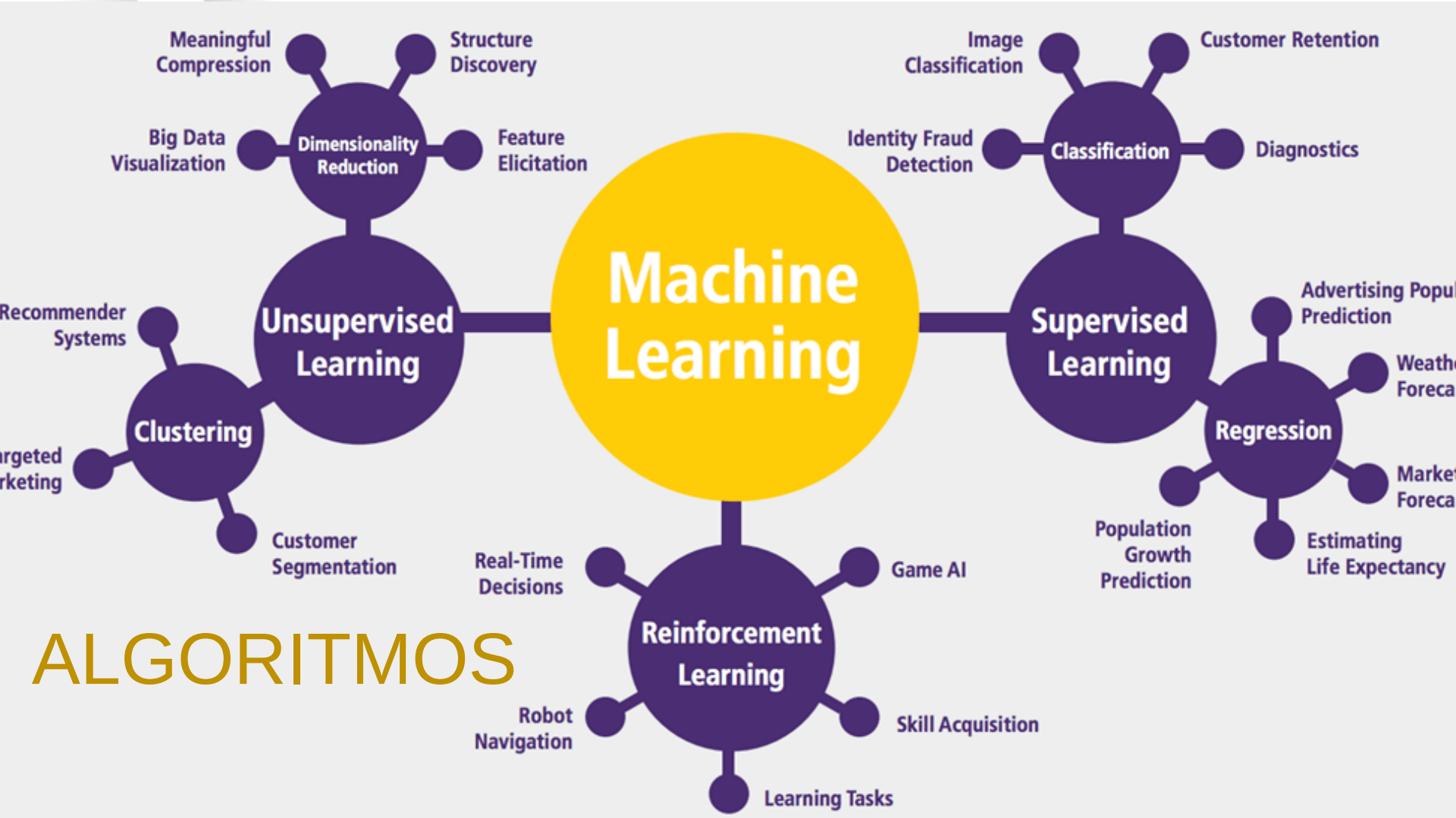
- Operadora Telefónica.
  - Disponemos de una variable “Número de Minutos a Móviles de cada cliente”.
  - Variables Sintéticas que se pueden construir:
    - Media de minutos hablados en la última semana, mes, semestral o anual.
      - Disponemos de datos más estables, no dependientes de un momento puntual por los minutos hablados.
    - Tendencia minutos hablados:
      - Se puede detectar el abandono por falta de uso, u ofertar una nueva tarifa por el aumento del uso.
      - Indicador reducción 50% minutos hablados. Desvinculación del cliente.
      - Indicador incremento 50% minutos hablados. Una vinculación más fuerte del cliente.

Variables Sintéticas

- Compañías de Seguros.
  - Mediador.
    - Evoluciones de las comisiones, Tendencia rating, Incremento anulaciones, evolución negocio mediador, incremento nueva producción, evolución negocio mediador, rentabilidad del cliente.
  - Cliente.
    - Tenencias de ramos o tipos de pólizas (hogar, auto, vida, etc).
    - Tendencia de pólizas.
    - Anulación de pólizas. Si anula la de auto hace tres meses, es probable que anule hogar.
    - Cambio de dirección.
  - Póliza.
    - Agrega segundo conductor.
    - Prima relación competencia.
    - Incremento de prima.
    - Indicador de rehúse (el cliente se dirige al abandono).

Variables Sintéticas

- Análisis en Redes Sociales.
  - Dominancia.
  - Cajas de resonancia.


Machine Learning
Unsupervised Learning
Meaningful Compression
Structure Discovery
Dimensionality Reduction
Big Data Visualization
Feature Elicitation
Clustering
Recommender Systems
rgeted rketing
Customer Segmentation
Supervised Learning
Image Classification
Customer Retention
Classification
Identity Fraud Detection
Diagnostics
Regression
Advertising Popul Prediction
Weath Foreca
Marke Foreca
Population Growth Prediction
Estimating Life Expectancy
Reinforcement Learning
Real-Time Decisions
Game AI
Robot Navigation
Skill Acquisition
Learning Tasks

ALGORITMOS

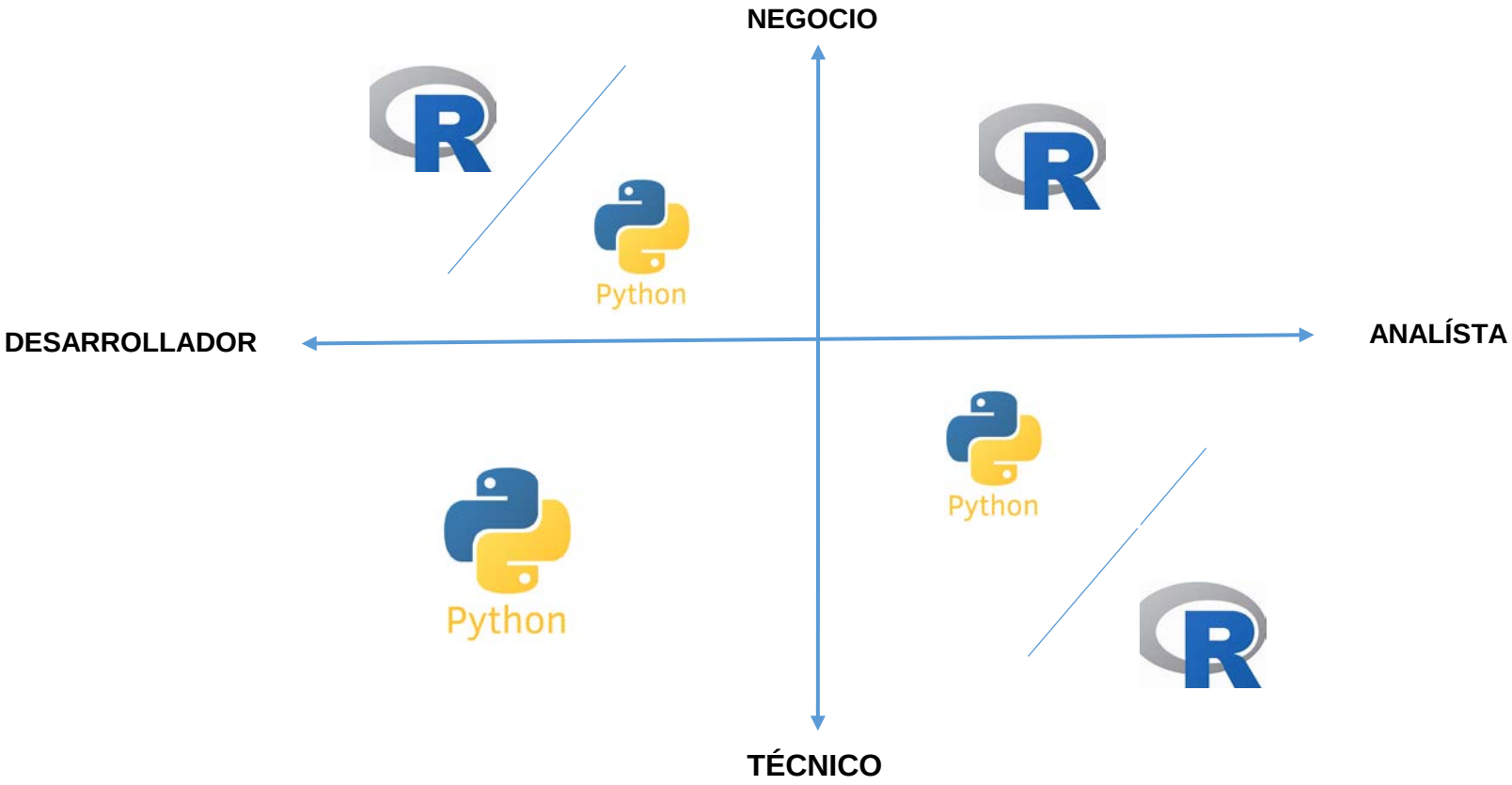
Algoritmos para DataScience

- Existe multitud de algoritmos.
  - Investigación.
  - Negocio.
- Tipos de Algoritmos:
  - Agrupación/Segmentación.
  - Predicción.
  - Análisis de texto.

¿Qué Herramienta Aprendo?

- Idea. Aprender los dos.
- Practico. Comenzar con uno y aprender el proceso, y posteriormente aprender el segundo.
- Preguntas para tomar la decisión del primer lenguaje:
  - Origen:
    - Técnico: Python.
    - Snálisis: R, DataMiner, etc.

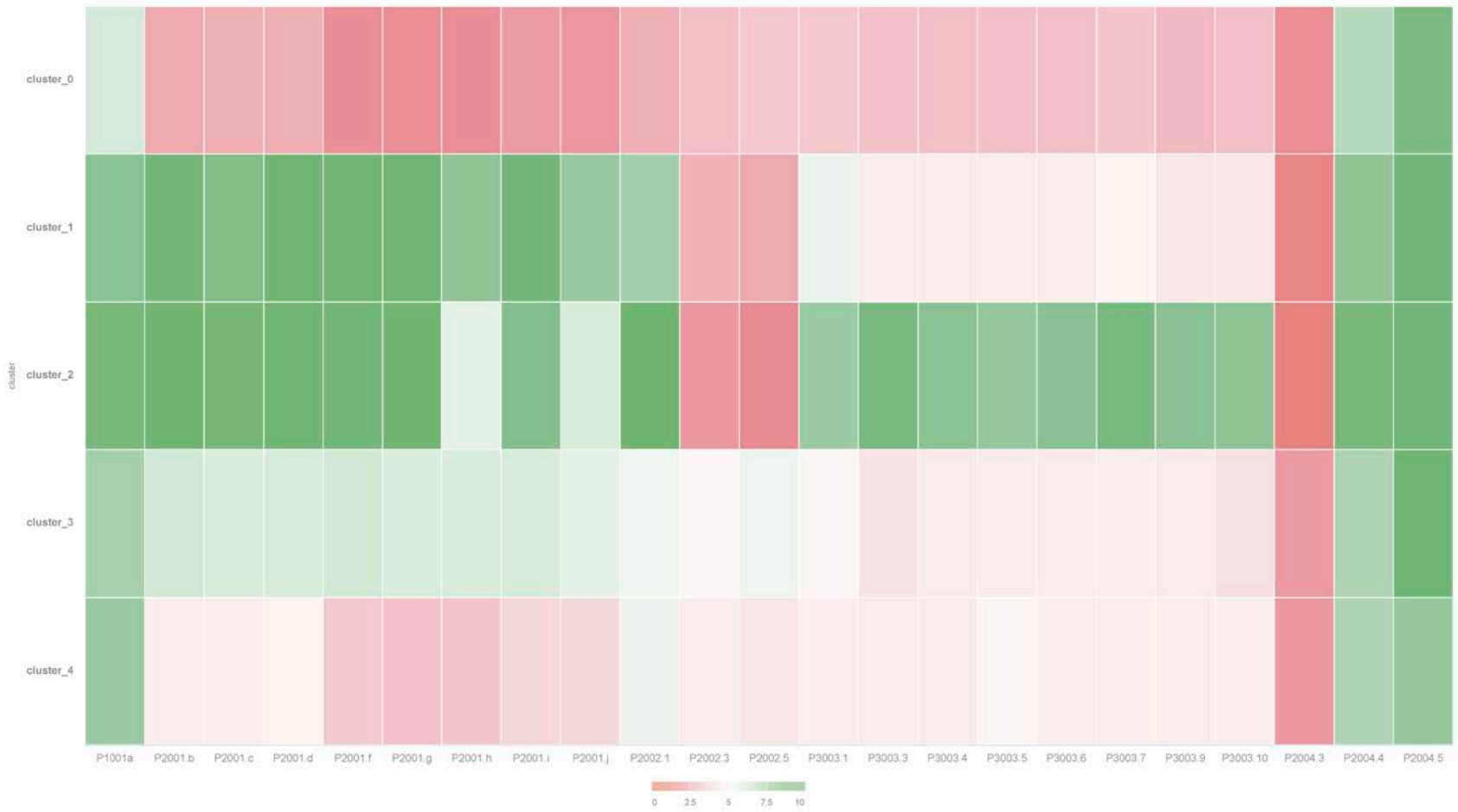
Gráfica de Elección



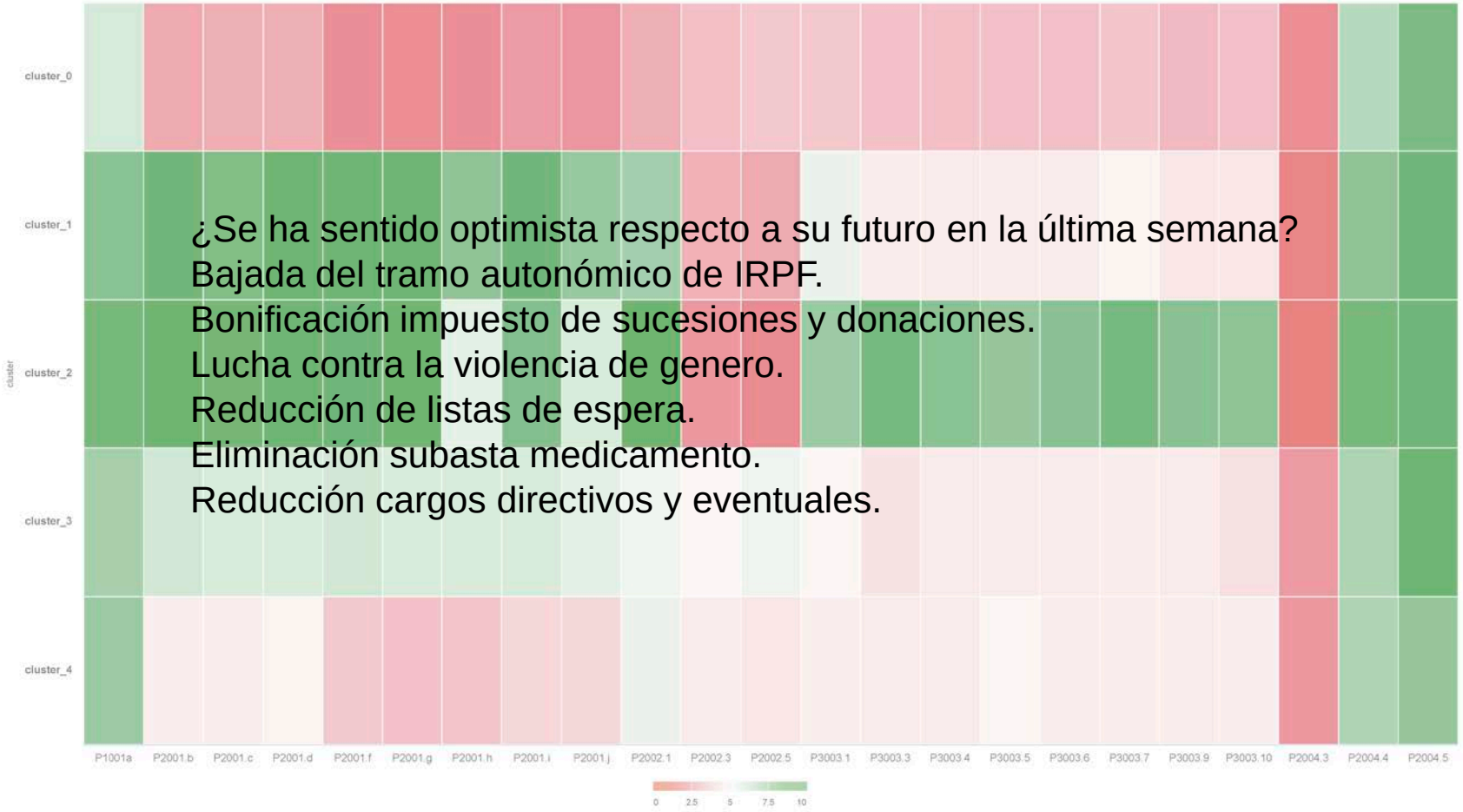
LOS EJEMPLOS

Análisis Social

- Datos demoscópicos.
- Se ha realizado un proceso de Transformación y limpieza de datos.
- Se han aplicado técnicas de machine learning:
  - Selección de características: Se han aplicado un experimento de múltiples comparaciones y selección del óptimo utilizando el método del frente de Pareto.
  - Clusters: Se han aplicado técnicas de sectorización no supervisada, aplicando algoritmos basados en funciones de densidad, x-mean y x-meioid.
  - Predicción. Se han aplicado técnicas de IA, en concreto de Deep learning.

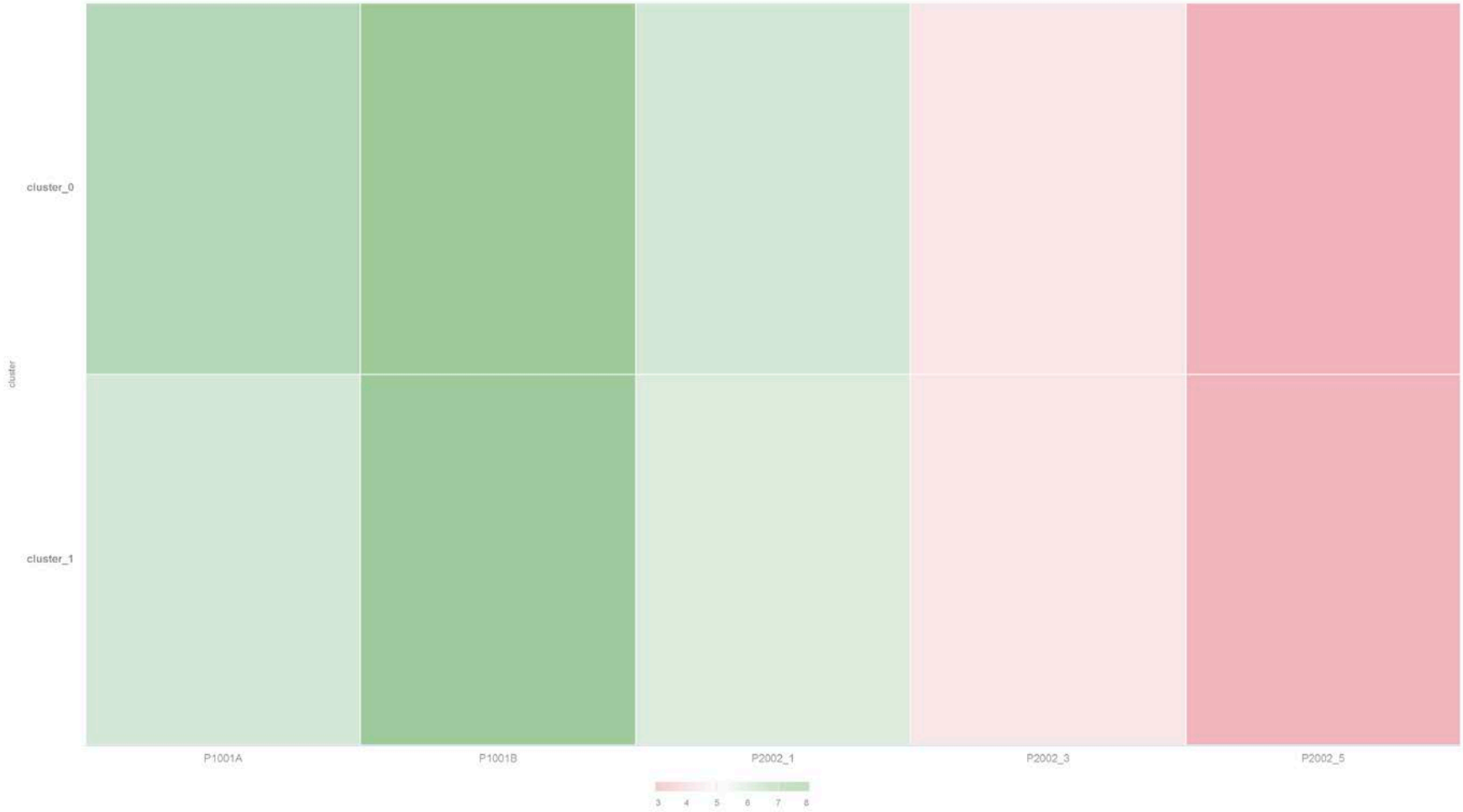

cluster_0
cluster_1
cluster_2
cluster_3
cluster_4
cluster
P1001a
P2001.b
P2001.c
P2001.d
P2001.f
P2001.g
P2001.h
P2001.i
P2001.j
P2002.1
P2002.3
P2002.5
P3003.1
P3003.3
P3003.4
P3003.5
P3003.6
P3003.7
P3003.9
P3003.10
P2004.3
P2004.4
P2004.5
0
2.5
5
7.5
10

Junio 2019

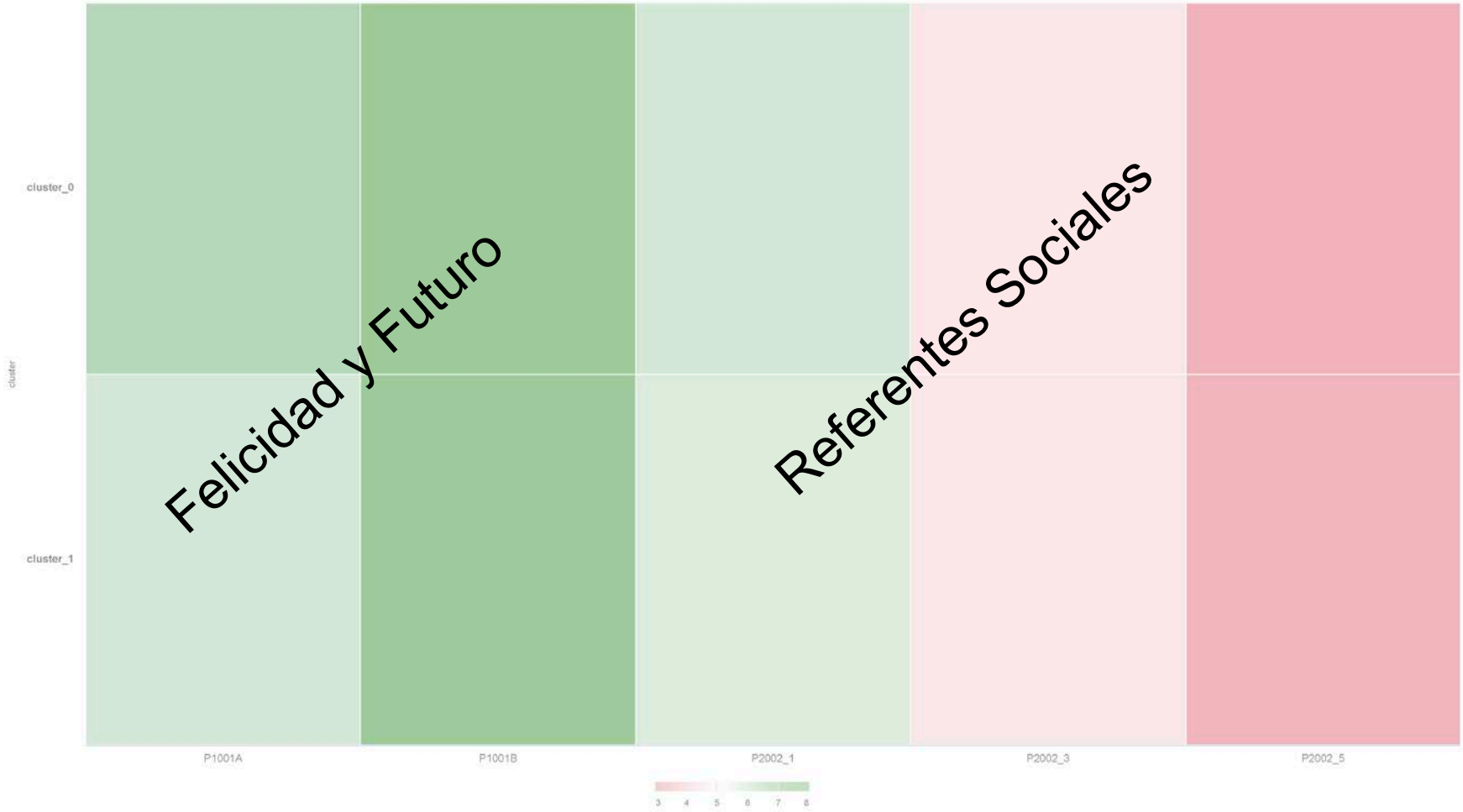


¿Se ha sentido optimista respecto a su futuro en la última semana?
Bajada del tramo autonómico de IRPF.
Bonificación impuesto de sucesiones y donaciones.
Lucha contra la violencia de genero.
Reducción de listas de espera.
Eliminación subasta medicamento.
Reducción cargos directivos y eventuales.

Junio 2019


cluster_0
cluster_1
cluster
P1001A
P1001B
P2002_1
P2002_3
P2002_5
3
4
5
6
7
8

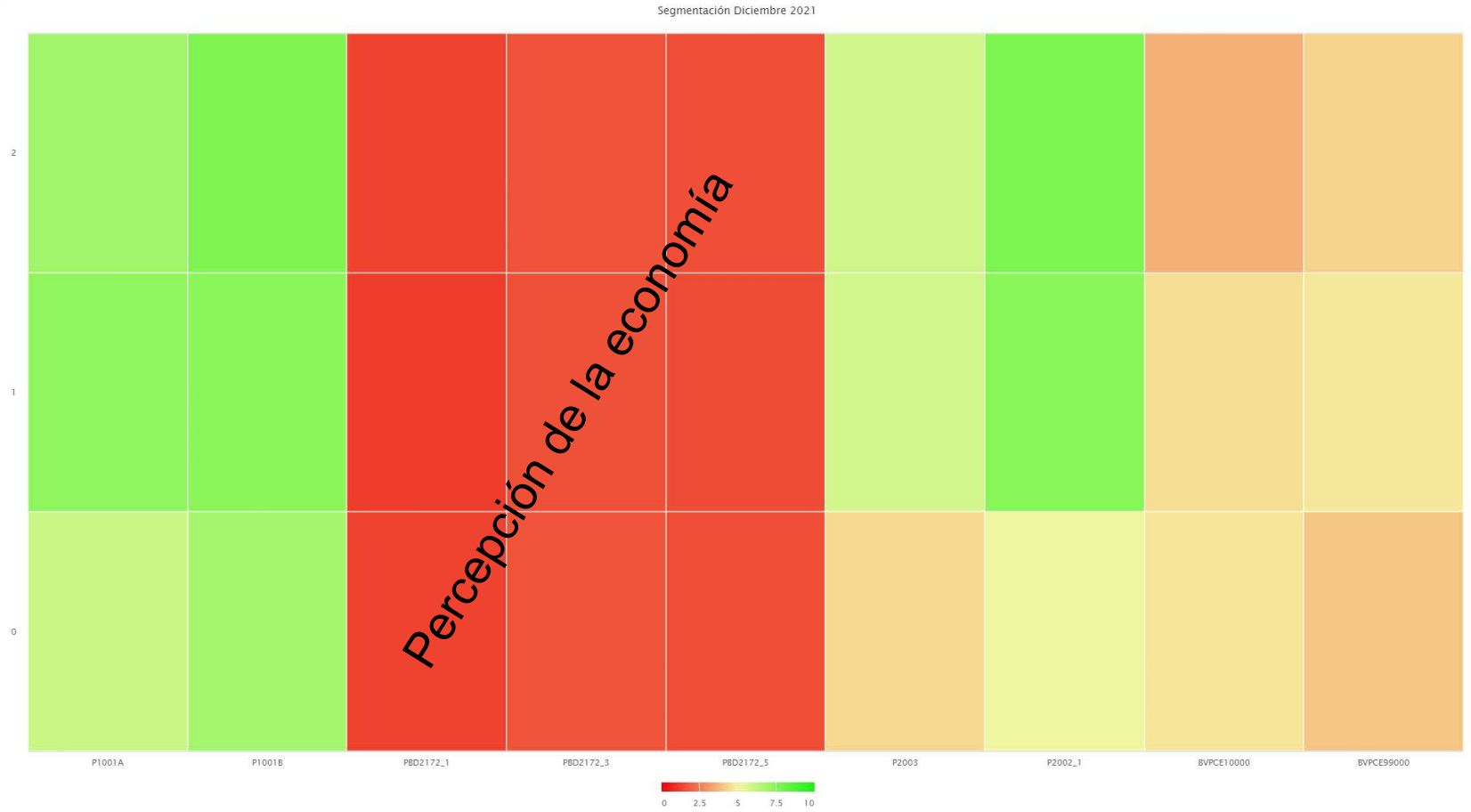
Junio 2020


cluster_0
cluster_1
cluster
Felicidad y Futuro
Referentes Sociales
P1001A
P1001B
P2002_1
P2002_3
P2002_5
3
4
5
6
7
8

Junio 2020

cluster_0
cluster_1
cluster_2
cluster_3
cluster_4
cluster
P1001A
BS62000_3
BS62000_4
BS63000_1
BS63000_3
P2003
P2004_2
P2004_3
P2004_4
P2004_6
P2004_7
P2002_2
P2002_5
BVPCE101000
BVPCE102000
BVPCE104000
BVPCE106000_2
BVPCE106000_4
BVPCE106000_5
0
2.5
5
7.5
10
Referentes sociales aparece la empresa o empresarios

Marzo 2021



Diciembre 2021

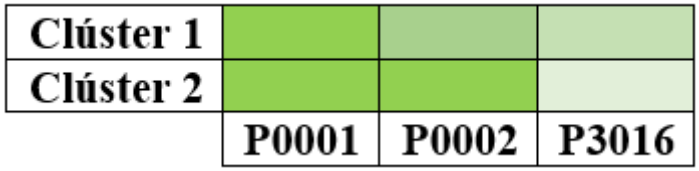
Análisis Empresarial

- Fundación Corporate Excellence. Center for Reputation Leadership.
- Datos de Ecosistemas Digitales.
- Analizar lo que piensan los ciudadanos respecto a temas de interés para el futuro de las empresas visto por los directivos de empresas.
- Proceso.
  - Definición de temas de estudio.
  - Traslado a Ecosistemas Digitales.
  - Scraping.
  - Análisis Semántico y de Sentimiento.

TEMÁTICAS DE ESTUDIO	
1	Confianza hacia las organizaciones
2	Digitalización y tecnología
3	Reputación y riesgo reputacional
4	Futuro del trabajo
5	Comunicación y publicidad
6	Diversidad, equidad e inclusión
7	Cambio climático
8	Agenda 2030 y ODS
9	Ciberseguridad
10	Ética empresarial
11	Métricas de sostenibilidad.
12	Reputación del CEO

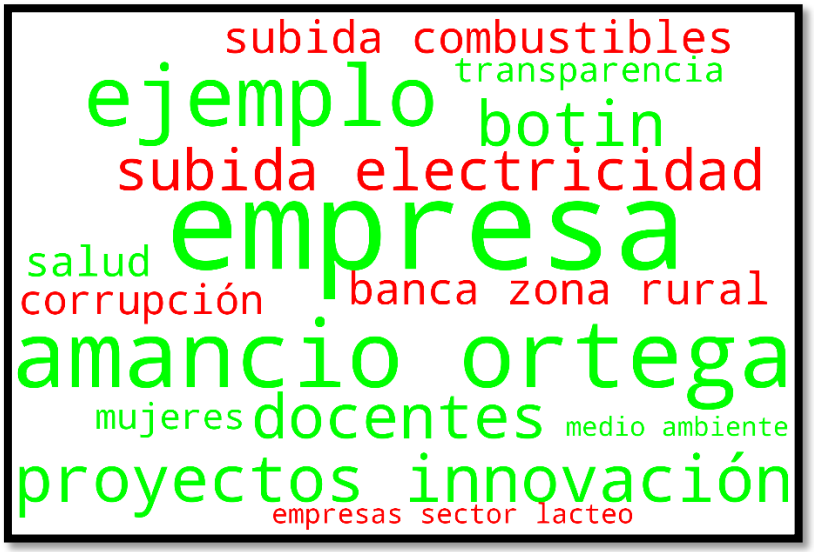
ID de Pregunta	Pregunta
P0001	¿Se siente optimista respecto a su futuro?
P0002	¿Se ha sentido feliz en la última semana?
P0003	¿Cuál es el problema que más le afecta a usted?
P0004	¿Le preocupa la subida de suministros y energía?
P0005	¿Valoración situación económica?
P0006	¿Cree que mejorará en el futuro?
P0007	¿Tiene dificultades para llegar a final de mes?
P1001	¿Se habla de la agenda 2030?
P1002	¿Le preocupan los mares y océanos?
P1003	¿Le preocupan los ríos?
P1004	¿Le preocupan los bosques?
P1005	¿Le preocupa la fauna/animales en general?
P2001	Lucha contra la violencia de genero.
P2002	¿Qué valoración tiene de la igualdad entre hombres y mujeres?
P3001	¿Abusando del teletrabajo?
P3002	¿Sentimiento hacía los empresarios?
P3003	¿Le gusta trabajar en empresas multinacionales?
P3004	¿Trabaja más horas que las que le pagan?
P3005	¿Confía en su empresa?
P3006	¿Respeta la empresa su horario de trabajo?
P3007	¿Es importante para usted la ciberseguridad?
P3008	¿Qué empresarios conoce?

Clúster 1
Clúster 2
Clúster 3
Clúster 4
Clúster 5
Clúster 6
P0001
P0002
P0006
30002
P3003
P3004
P3005
P3011
P3015
P3016
P4002
P5001
P5005
P7001


Clúster 1
Clúster 2
P0001
P0002
P3016

Clúster 1
Clúster 2
Clúster 3
P0001
P0003
P0004
P0005
P0006
P3016

economia circular
igualdad hombre mujeres
planeta
impacto económico
jóvenes trabajo
salud
iniciativa bosques
medioambiente
trabajo
gestión agua
iniciativas conciliación
igualdad oportunidad trabajo
disminuir CO2 Repsol
agua ropa inditex
recompra muebles usados


subida combustibles
transparencia
ejemplo
botin
subida electricidad
empresa
salud
corrupción
banca zona rural
amancio ortega
mujeres
docentes
medio ambiente
proyectos innovación
empresas sector lacteo

Información y Consultas en
masempresas.cea.es

/CEA.es @CEA.es_ /CEA.es

Gracias

Financiado por:

CEA
Confederación de Empresarios de Andalucía

Junta de Andalucía
Consejería de Transformación Económica, Industria, Conocimiento y Universidades