

COLEGIO UNIVERSITARIO DE ESTUDIOS FINANCIEROS

GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

Trabajo Fin de GRADO



Analítica de RR.HH: Machine Learning

**Agrupación de empleados en función de su
desempeño y potencial. Propuestas de actuación
frente a dichos grupos**

Autor: Navarro Meseguer, Guillermo

Tutor: Aguilar Pastor, Eva

Madrid, abril 2022

ÍNDICE

RESUMEN.....	3
1. INTRODUCCIÓN.....	3
2. MARCO TEÓRICO	4
2.1. DEFINICIÓN DE RR.HH., PRINCIPALES OBJETIVOS Y FUNCIONES.....	4
2.2. ANALÍTICA DE DATOS EN RR.HH.	9
2.2.1. Usabilidad y aplicación de HRA	10
2.2.2. Técnicas empleables en HRA: <i>Machine Learning</i>	16
3. CASO PRÁCTICO	20
3.1. EXPLICACIÓN Y JUSTIFICACIÓN DE LOS DATOS SIMULADOS.....	20
3.1.1. Variables demográficas o situacionales del empleado	21
3.1.2. Variables sobre el compromiso y la satisfacción del empleado y su paquete retributivo. Formación dentro y fuera de la empresa.....	22
3.1.3. Variables sobre la personalidad y su relación con el puesto de trabajo.....	23
3.1.4. Variables indicativas del rendimiento. Potencial y desempeño: Matriz 9 cajas	24
3.2. ANÁLISIS DESCRIPTIVO DE LOS DATOS	25
3.3. CLUSTERING	30
3.3.1. Modelo Supervisado: Reducción dimensional	30
3.3.2. Modelo No Supervisado: Agrupación.....	31
3.4. ANÁLISIS CLÚSTER: PROPUESTAS DE ACTUACIÓN	32
4. CONCLUSIONES.....	39
BIBLIOGRAFÍA.....	41
ÍNDICE DE ILUSTRACIONES.....	43
ÍNDICE DE GRÁFICOS.....	44
ÍNDICE DE ANEXO	45
ANEXO	46

RESUMEN

En este Trabajo de Fin de Grado se realiza una investigación empírica en la que se aplica la analítica de datos a los Recursos Humanos. El objetivo principal partirá de diversas hipótesis que se irán contrastando a lo largo del estudio. La primera parte de que las empresas podrían agrupar a sus empleados, en distintos grupos, lo más heterogéneos posibles, siendo al mismo tiempo cada grupo homogéneo entre sí, utilizando unas variables independientes (retribución, formación, compromiso, satisfacción, rotación, etc.). El objetivo de esta agrupación es ayudar en la toma de decisiones sobre variables dependientes (potencial y desempeño). La segunda hipótesis en la que se fundamenta este estudio es que, dentro del ámbito de la evaluación de los empleados, los departamentos de RR.HH. pueden aplicar políticas concretas en función del grupo o clúster asignado a cada empleado.

1. INTRODUCCIÓN

Este trabajo de investigación pretende analizar, desde una perspectiva empírica, uno de los aspectos más actuales de la gestión del capital humano dentro de las empresas: la analítica de los recursos humanos, en adelante usaremos sus siglas en inglés HRA.

La elección de la temática de este estudio reside en que en la actualidad existen diversos estudios en los que se explican los principales aspectos que se deben tener en cuenta para realizar una buena investigación; los tipos de análisis que se pueden realizar; las variables que se deben tomar en consideración; etc., para una correcta aplicación de HRA, para tomar decisiones más informadas y con ello conseguir que éstas sean más eficientes y eficaces. Sin embargo, son pocos los estudios que abordan desde una perspectiva práctica dichas cuestiones.

Siendo los Recursos Humanos, un área que abarca múltiples funciones, en los siguientes epígrafes se pretende desarrollar una propuesta de gestión de empleados en función de su desempeño y de su potencial, aplicando una de las ramas de la Inteligencia Artificial (AI) que, a su vez, es una de las herramientas más utilizada dentro de la analítica de datos: el *Machine Learning*.

El objetivo principal partirá de diversas hipótesis que se irán contrastando a lo largo del estudio. La primera hipótesis parte de que las empresas podrían agrupar a sus empleados, en distintos grupos, lo más heterogéneos posibles, siendo al mismo tiempo cada grupo homogéneo entre sí, utilizando unas variables independientes (retribución, formación, compromiso, satisfacción, rotación, etc.). El objetivo de esta agrupación es ayudar en la toma de decisiones sobre variables dependientes (potencial y desempeño). La segunda hipótesis en la que se fundamenta este estudio es que, dentro del ámbito de la evaluación de los empleados, los departamentos de RR.HH. pueden aplicar políticas concretas en función del grupo o clúster asignado a cada empleado.

Para ello, en primer lugar, se establecerá el marco teórico que afecta directamente al objeto de estudio, abordando, por un lado, las principales definiciones, objetivos y funciones de los RR.HH. Por otro, la analítica de RR.HH., su usabilidad y aplicación y las técnicas que se emplearan para realizar el análisis.

En segundo lugar, se explicará la metodología aplicada: la simulación de los datos y las variables escogidas; los distintos análisis realizados junto con sus aspectos más técnicos. Obtenidos los resultados, se propondrán distintas actuaciones o propuestas que los encargados del departamento de RR.HH. podrían aplicar de forma genérica a los empleados de los grupos formados, con el objetivo de mejorar su rendimiento.

2. MARCO TEÓRICO

2.1. DEFINICIÓN DE RR.HH., PRINCIPALES OBJETIVOS Y FUNCIONES

Desde que, a mediados del siglo XIX, con la Revolución Industrial se empezó a hacer patente la necesidad de gestionar a los empleados, por parte de las empresas, hasta el día de hoy, el término conocido como Recursos Humanos (RR.HH.) se ha visto sometido a múltiples definiciones y cambios dependiendo, en cada momento, de las principales necesidades y funciones que a éste se le han atribuido.

Partiendo de una visión teórica de la Economía de la Empresa, los RR.HH. pueden definirse como “el área de gestión empresarial responsable de las decisiones y acciones que afectan a la relación entre la empresa y los trabajadores que prestan sus servicios en

ella; decisiones y acciones éstas que son adoptadas para la consecución de los objetivos empresariales” (Albizu y Landeta, 2011: 25). Es importante no perder de vista el objetivo principal de este departamento: alinear sus actuaciones con la consecución de los objetivos fijados por la empresa, sin importar el tamaño, la estructura o la actividad de ésta puesto que es lo que permitirá a las empresas gestionar una de sus principales fuentes de ventaja competitiva: el capital humano.

Desde esta aproximación, los autores mencionados hacen referencia a tres objetivos básicos que los directores de RR.HH. deben perseguir: la eficiencia económica, la eficiencia social y la integración o congruencia de sus actuaciones.

- La eficiencia económica busca optimizar la productividad de los empleados, junto con el coste económico y el grado de flexibilidad y adaptabilidad para conseguir una eficiencia y eficacia sostenible en el tiempo.
- “La eficiencia social trata de mejorar el grado de satisfacción de las expectativas, necesidades e intereses de las personas de la organización”.
- La congruencia de sus actuaciones hace referencia especialmente a la necesidad de estar alineados con los objetivos del resto de la empresa y a ser coherentes en las propias políticas y objetivos de los recursos humanos (Albizu y Landeta, 2011: 27).

Por otra parte, interesa recordar cuáles son las principales funciones que desarrollan para llevar a cabo dichos objetivos. Empresas del sector, como Randstad, indican que las funciones clave de esta área son la planificación, selección y administración de personal, evaluación y desarrollo, formación y relaciones laborales (Randstad, n.d.).

Sin embargo, Albizu y Landeta, sugieren dar un paso más allá y proponen una clasificación de las principales áreas de actuación: dirección estratégica; administración de personal; gestión de RR.HH.; desarrollo de RR.HH.; área económica; seguridad y salud laboral; relaciones laborales; empleo (Ilustración 1).

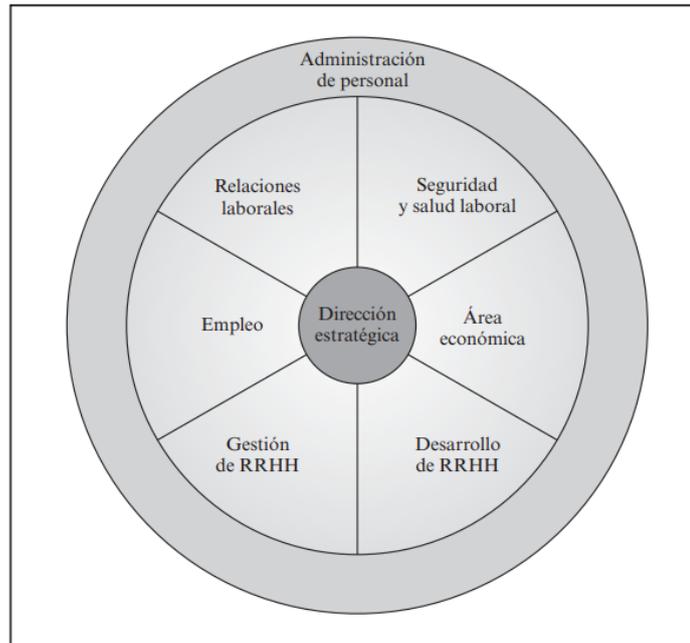


Ilustración 1: Áreas de actuación de la FRRHH.

Fuente: Albizu y Landeta (2011:29)

Señalan que dichas áreas no deben tenerse en cuenta de una forma absoluta puesto que, al tener relaciones directas entre sí, se pueden encajar correctamente de manera flexible en diferentes áreas. Sin entrar en el análisis exhaustivo de cada una de las ocho áreas mencionadas, únicamente nos detendremos en aquellas que son más relevantes para la elaboración de esta investigación, entre otras: el área de dirección estratégica, el área de administración de personal y el área de desarrollo de RR.HH.

Respecto al área de dirección estratégica permite entender que toda actuación dentro de este departamento debe ir unida a la persecución de unos objetivos compartidos por todos los directores funcionales de la organización, jugando así no sólo un papel administrativo sino también uno estratégico. Las demás funciones de RR.HH. partirán de dichos objetivos para desarrollar sus líneas de actuación.

Por otro lado, el área de administración de personal engloba el resto tratándose, como indican los citados autores, de un área transversal que tiene distintas interfaces con las demás, representadas en la figura, de tal modo que muestran su función de apoyo al resto.

En el área de desarrollo de los RR.HH., los autores engloban aquellas actividades que están destinadas a la “adecuación y actualización de las competencias (conocimiento, capacidades y actitudes) del personal en función de las necesidades de la empresa, así como del crecimiento profesional de los trabajadores” (Albizu y Landeta, 2011: 32).

Encargándose por tanto de las siguientes tareas: gestión de competencias; diagnóstico de necesidades de formación y planes de formación; diseño e impulso de planes de carrera; promociones y rotaciones; valoración del rendimiento; valoración del potencial; y gestión del conocimiento (Albizu y Landeta, 2011). Partiendo de esta definición y de los conceptos que involucran a esta área, se puede hablar de la matriz de desempeño y potencial, pues todos ellos son elementos que se relacionan directa o indirectamente con los ejes de coordenadas de dicha matriz.

La matriz de desempeño potencial o también conocida como la matriz de Odiorne, inspirada en la de producto del BCG, propone lo siguiente:

		Desempeño	
		Alto	Bajo
P o t e n c i a l	Alto	Empleados estrella	Empleados problema
	Bajo	Empleados esforzados	Empleados de bajo perfil

Ilustración 2: Matriz de Odiorne contrapuesta con la propuesta por BCG.

Fuente: Elaboración propia a partir de Rothwell y Kazanas (2003)

En ella se observa una posible clasificación de los empleados en función de sus niveles de desempeño y de potencial. “De esta forma, se puede situar a cada trabajador (o grupo de trabajadores) en el cuadrante que mejor le corresponde, y en función de ello determinar las actuaciones de DRR.HH. (desarrollo de RR.HH.) más acordes con su valor actual y futuro, y con los intereses estratégicos de la empresa, así como valorar en términos de fortaleza o debilidad el grupo humano con el que cuenta la empresa” (Albizu y Landeta, 2011: 369). Dentro las actuaciones de DRR.HH. que suelen llevarse a cabo con esta herramienta, una vez distribuidos los empleados en sus respectivas categorías, se encuentra la toma de medidas que permitan diseñar correctamente los planes de carrera con el objetivo doble de evitar problemas de cobertura de vacantes, causadas por el reclutamiento interno y, a su vez, reducir la falta de motivación y de compromiso.

De esta matriz se obtiene la que actualmente se aplica con mayor frecuencia en la gestión de recursos humanos, denominada como la matriz de las 9 cajas, y que proviene de una adaptación que hizo la empresa General Electric para su uso propio (Lee, 2018).

Lo que caracteriza a esta matriz es que transforma la de Odiorne -2x2- en una matriz 3x3. Pasando así de tener cuatro tipos de empleados en función de su desempeño y potencial a poder dividirlos en nueve grupos. En la doctrina académica no hay un consenso sobre qué nombres otorgarles, pero sí se muestra acuerdo sobre la función que desempeña cada grupo. Atendiendo a las definiciones y nombres aportados por la *Academy to Innovative HR (AIHR)*, se encuentran los siguientes tipos: empleados a los que se debe considerar despedir (*bad hires*); profesionales dilema (*up or out dilemmas*); profesionales de buen desempeño (*up or out grinders*); profesionales enigma (*dysfunctional geniuses*); profesionales confiables (*workhorses*); empleados clave (*core players*); empleados estrella en su área (*high performers*); empleados futuras estrellas (*high potentials*); empleados estrella (*stars*) (AIHR, 2021):

- Empleados a los que se debe considerar despedir: la idea principal de estos trabajadores es que en un primer momento no debieron ser contratados, pero si aparecen en la organización deben ser despedidos inmediatamente.
- Empleados dilema: los que teniendo algo de potencial no tienen un buen desempeño; habrá que acudir a la causa de esta situación, y en función de ella tomar decisiones.
- Empleados con buen desempeño o eficientes: es un grupo algo conflictivo, puesto que hacen su trabajo suficientemente bien como para no tener que ser despedidos, pero con muy bajo potencial. El objetivo principal será conseguir aumentar su potencial y rendimiento en el desempeño de su puesto de trabajo.
- Empleados enigma: aquellas personas que, teniendo un gran potencial, no están siendo capaces de rendir correctamente, esto lleva a pensar que en el corto plazo su desempeño aumentará, pero habrá igualmente que localizar la causa de su escaso rendimiento.
- Profesional confiable: buenos trabajadores, pero con bajo potencial. El objetivo de la organización será conseguir que sigan haciendo correctamente su trabajo.
- Empleados clave: aquellos profesionales confiables debido a su desempeño en el puesto y con potencial suficiente para desarrollarse profesionalmente.

- Empleados estrella en su área: empleados alineados con su puesto, la organización se deberá esforzar en retenerles.
- Futuras estrellas: se espera que consigan adecuarse al puesto correctamente, pues teniendo altos niveles de potencial, en un primer momento se presuponen que no han terminado de crecer en su máximo nivel de desarrollo.
- Empleados estrella: forman el grupo de empleados en su área preparados para promocionar a nuevos puestos.



Ilustración 3: Matriz de las 9 cajas.

Fuente: Elaboración propia a partir de AIHR (2021)

2.2. ANALÍTICA DE DATOS EN RR.HH.

Explicados algunos de los conceptos básicos y tradicionales de los RR.HH. que atañen a este estudio, a continuación, se profundizará en aquellos conceptos más novedosos de esta área.

Conviene partir del dato de que al término "análisis de datos en RR.HH.", es frecuente que se le atribuyan diferentes significados y, por tanto, no termina de haber un consenso claro al respecto. Para algunos autores sólo significa un proceso para informar sistemáticamente sobre una serie de métricas de RR.HH. En el otro extremo, se encuentran aquellos otros que consideran la analítica de RR.HH. en un sentido más enriquecido; es decir, las únicas actividades y/o procesos que constituyen la analítica de

RR.HH. son las que implican un modelo predictivo de "alto nivel" (por ejemplo, escenarios "*what-if*" que pronostican las consecuencias de los cambios en las políticas o en las condiciones) (Bassi, 2011).

Sin embargo, Fitz-Enz, uno de los principales promotores de la aplicación de métricas en la gestión del capital humano, sostiene que para responder a la pregunta de qué es realmente la analítica, es necesario acudir "al encuentro del arte y la ciencia. Las artes nos enseñan cómo mirar el mundo. Las ciencias nos enseñan cómo hacer algo. (...) La analítica es un marco mental, una progresión lógica primero y un conjunto de herramientas estadísticas después. (...) Para resolver problemas organizativos, necesitamos una estructura lógica para analizar las numerosas variables que pueden afectar al rendimiento humano. Una vez que hayamos identificado esas variables, podemos emplear la estadística según sea necesario" (Fitz-Enz, 2010: 4-5). Otros autores destacados del campo indican que "los líderes de dentro y fuera de la profesión de RR.HH. a menudo desconocen las formas existentes con apoyo científico para medir y evaluar las implicaciones de las decisiones sobre recursos humanos. Un pilar esencial de cualquier ciencia de la decisión es un sistema de medición que mejore las decisiones mediante principios científicos sólidos y relaciones lógicas" (Cascio *et al*, 2019: 5-6).

Se puede observar, por tanto, una clara tendencia a la afirmación de que una de las principales funciones del HRA es poder tomar decisiones mejor informadas, combinando razonamientos lógicos con métodos estadísticos, sin olvidar que debe ir dirigido a una gestión estratégica y, por tanto, tener un impacto positivo en la creación de valor para la empresa.

2.2.1. Usabilidad y aplicación de HRA

2.2.1.1. Razonabilidad del uso de HRA

Si bien parece que tiene sentido incluir en el manejo de los RR.HH. un sistema que permita realizar un mayor análisis, cabe preguntarse si realmente merece la pena, puesto que no cabe duda de que se trata de una inversión en un campo muy amplio y con diversas alternativas en cuanto a la usabilidad.

Es ante este interrogante, donde aparecen diversos estudios como el realizado por IBM, donde se mostró que existía una fuerte tendencia positiva entre los encuestados que revelaba el importante papel que desempeña la analítica en la gestión eficaz tanto del rendimiento de la plantilla como de la gestión del talento. En la Ilustración 4 se muestra como el 76 por ciento de los encuestados indicaba que presenciaba una mejora en la capacidad de gestión de su fuerza de trabajo, mientras que el 69 por ciento citó la mejora en los niveles de productividad. El 67 por ciento consideraba que la analítica de la fuerza de trabajo es importante para impulsar un mejor retorno de la inversión en la gestión del talento (IBM, 2009).

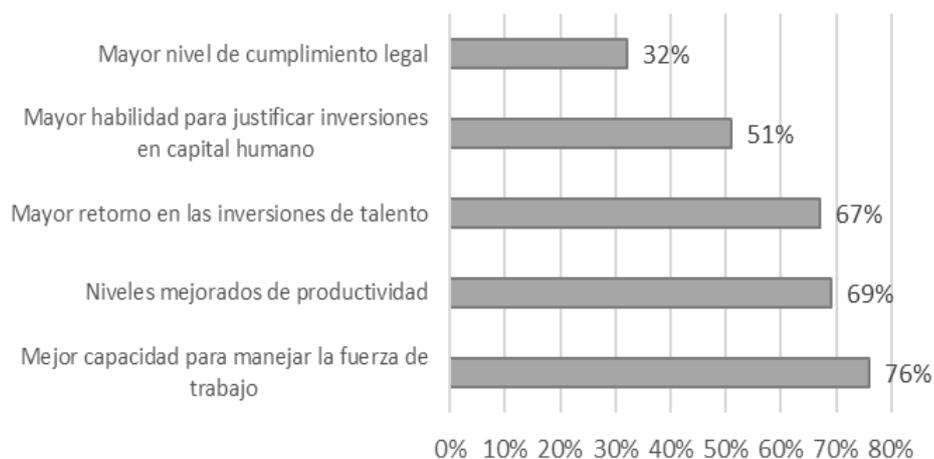


Ilustración 4: Beneficios potenciales de la analítica de personal.

Fuente: Elaboración propia a partir de IBM (2009)

Por otro lado, en la Ilustración 5, se aprecia cómo el 63 por ciento de las empresas de primer orden afirman haber experimentado la influencia del uso de la analítica en la reducción en los costes generados en la toma de decisiones, y el 53 por ciento afirma que le ayudó en el aumento de los ingresos de estas (Roberts, 2009).

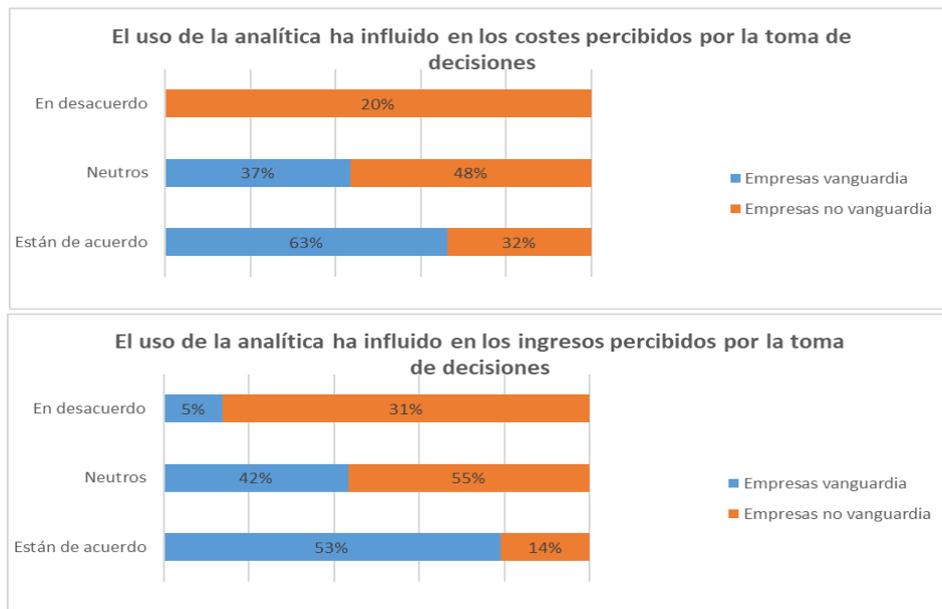


Ilustración 5: Influencia del uso de analítica en la toma de decisiones.

Fuente: Elaboración propia a partir de Roberts (2009)

En otras palabras, se confirma que el empleo de HRA no solamente es algo novedoso, sino que además permite, tomando decisiones de forma más informada, generar un impacto positivo en la cuenta de resultados de la entidad.

A pesar de que estos estudios muestran que la utilización de estas técnicas no es algo baladí en las empresas españolas “solo un 22% de las compañías están trabajando la analítica en RR.HH. para recoger, analizar datos y usar modelos descriptivos, predictivos y prescriptivos. Sólo el 16% de las compañías cuenta con un responsable de HRA en su equipo, un 3% más que el año anterior (...) estos datos muestran claramente que falta aún visión de Data Driven desde las áreas de Recursos Humanos” (Incipy, 2021:18).

2.2.1.2. Niveles de aplicabilidad de HRA

Desde una aproximación teórica de la aplicación de la analítica se detectan cuatro usos principales (Aguado, 2018):

1. **Análisis inicial:** busca dar respuesta a la pregunta ¿qué ha sucedido? Se usan datos básicos, no relacionados entre sí y pueden servir como un primer acercamiento para analizar lo que ha ocurrido recientemente en la organización.

2. **Análisis comparado:** permite obtener valores más concretos que sirvan para comparar los resultados tanto interna como externamente. Por citar algún ejemplo: ¿qué departamentos rinden más?, ¿cuál es la evolución promedio de los salarios?, ¿están a un nivel similar que los de otras empresas del sector? Mientras que el paso anterior no permite basarse en sus resultados para tomar decisiones, este análisis por el contrario aporta abundantes datos comparativos que pueden trasladarse a la empresa en cuestión.
3. **Explicación:** persigue dar respuesta a interrogantes multidimensionales, trata de buscar relaciones entre variables, sin quedarse únicamente en el dato, sino que analiza la razón de ser de este.
4. **Predicción:** el principal objetivo de este último escalón es predecir los comportamientos futuros de aquellas variables que se quieren medir y su impacto en el negocio o en la propia empresa.

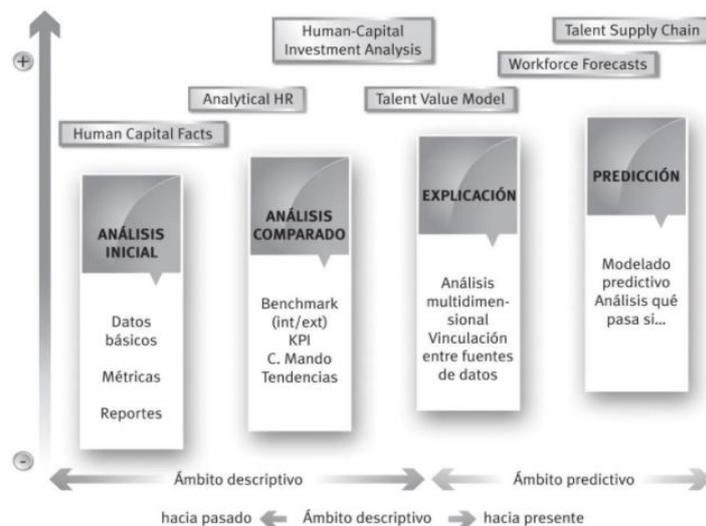


Ilustración 6: Diferentes usos y estados de aplicación de HRA.

Fuente: Aguado (2018:28)

2.2.1.3. Metodologías de aplicación de HRA

Cuando se trata de buscar la forma adecuada de elaborar cualquier proyecto relacionado con la analítica de datos en RR.HH., en la literatura aparecen diversas metodologías, entre las que destacan el modelo LAMP (Cascio y Boudreau, 2010) y el modelo DELTA (Davenport et al, 2010).

El primero, el modelo LAMP, sin duda el más citado por la doctrina académica, se fundamenta en los conceptos de *Logic*, *Analytics*, *Measures* y *Process*. Los autores señalan que a pesar de que el principal objetivo es conseguir las métricas correctas para el análisis, se necesita de las otras tres variables, pues en caso de fallar alguna de ellas no se obtendrían resultados consistentes para la toma de decisiones. En cuanto a la lógica (*Logic*), explican que “el elemento lógico de cualquier sistema de medición proporciona la “historia” detrás de las conexiones entre los números y los efectos y resultados”. Es decir, es el elemento que da cohesión entre lo que se busca, aquello que se mide y los resultados obtenidos. En cuanto a las medidas (*Measures*) se busca especialmente precisión y validez, dentro de la lógica previamente definida. La importancia del análisis (*Analytics*) reside en la correcta interpretación de aquellas variables que se han medido, teniendo en cuenta que dependerá de la profundidad y usabilidad que se le quiera dar a este. Para ello, es necesario la colaboración de expertos en el área que puedan ayudar en la interpretación y toma de decisiones. Por último, los autores hablan del proceso (*Process*) cuyo objetivo es conseguir que, el conocimiento generado con las otras tres patas del modelo realmente pueda ser puesto en práctica dentro de la organización y no se queden solamente en el plano teórico (Cascio y Boudreau, 2010: 10-16).

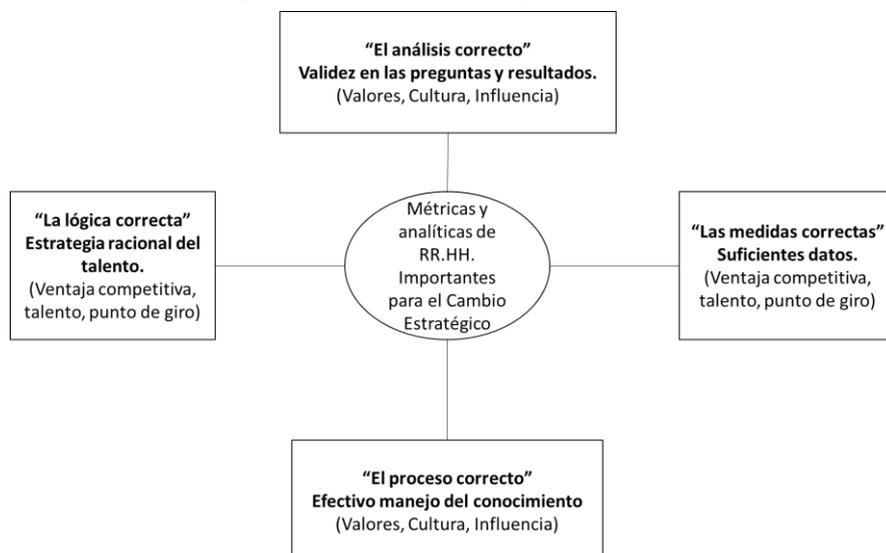


Ilustración 7: Modelo LAMP.

Fuente: Elaboración propia a partir de Cascio y Boudreau (2010:10)

Por otro lado, encontramos el modelo DELTA que ha sido recientemente actualizado por la doctrina, al añadir dos nuevos componentes, como se verá a continuación. El acrónimo, surge de *Data*, *Enterprise*, *Leadership*, *Targets* y *Analysts* (Davenport *et al*, 2010) a los

cuales posteriormente se les unen *Technology* y *Analytics Techniques*, obteniéndose lo que se conoce como el modelo DELTA PLUS (Davenport y Harris, 2017).

Datos (*Data*): hace referencia a la necesidad de obtener datos accesibles y de alta calidad.
Empresa (*Enterprise*): se orienta hacia que las empresas aboguen “por una perspectiva única y coherente para la analítica en toda la organización”.
Liderazgo (*Leadership*): persigue que haya personas líderes dentro de la organización que animen a la incorporación y aceptación de la analítica y el cambio cultural que ello implica dentro de la organización empresarial.
Objetivos (*Targets*): la amplitud del campo de la analítica requiere tener claro el objetivo de dónde se quiere focalizar la organización, pues de lo contrario podrá perderse al no centrarse “en unos pocos casos de uso y aplicaciones iniciales y con propósito”.
Analistas (*Analysts*): las empresas deben conseguir gente con conocimientos tanto técnicos en cuanto a los análisis de datos como conocimientos organizacionales; se debe tener en cuenta que este perfil altamente demandado no es fácil de captar ni de retener.
Tecnología (*Technology*): “a medida que la analítica y la IA se vuelven más críticas para el éxito de una organización, muchas necesitarán desarrollar arquitecturas sofisticadas para ellas”.
Técnicas analíticas (*Analytics Techniques*): hace referencia a que las organizaciones que deseen implementar correctamente y tener éxito con el uso de cualquier tipo de analítica, deben evaluar qué modelos son más convenientes para sus necesidades y si disponen del conocimiento necesario para aplicarlos (Davenport, 2018).

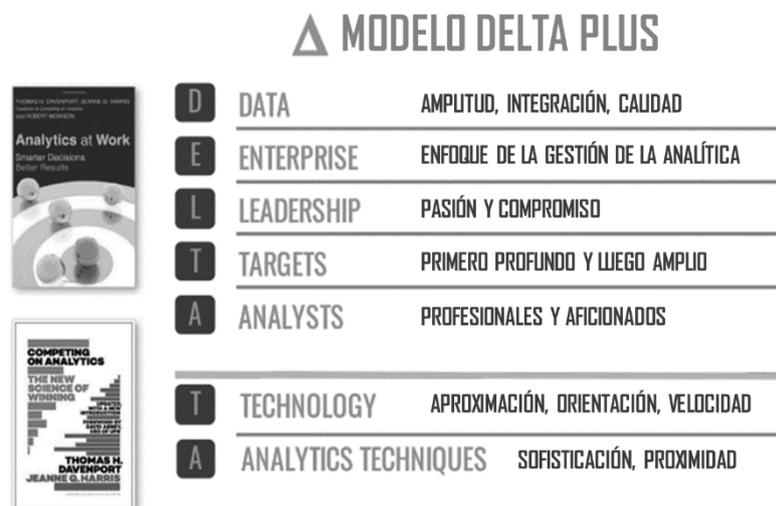


Ilustración 8: Modelo DELTA PLUS.

Fuente: Elaboración propia adaptado de Davenport et al (2010); Davenport y Harris (2017); Davenport (2018)

2.2.2. Técnicas empleables en HRA: *Machine Learning*

Hasta el momento, una de las principales ideas que ha aparecido repetidas veces en este Trabajo de Fin de Grado es que el campo de la analítica es muy amplio, ya sea por el uso que se le quiera dar o por las técnicas y conocimientos que se han de emplear para optimizar sus resultados. Su aplicación al ámbito de los RR.HH., no es diferente. La utilización de unas u otras herramientas, en una buena parte, dependerá del área o áreas en las que la dirección de RR.HH. decida centrarse. Por tanto, el objetivo de este apartado es dar algunas ideas generales sobre cómo podrían ser los primeros pasos a tener en cuenta al comenzar un proyecto de HRA que usa *Machine Learning* como herramienta principal de análisis (Ilustración 9).

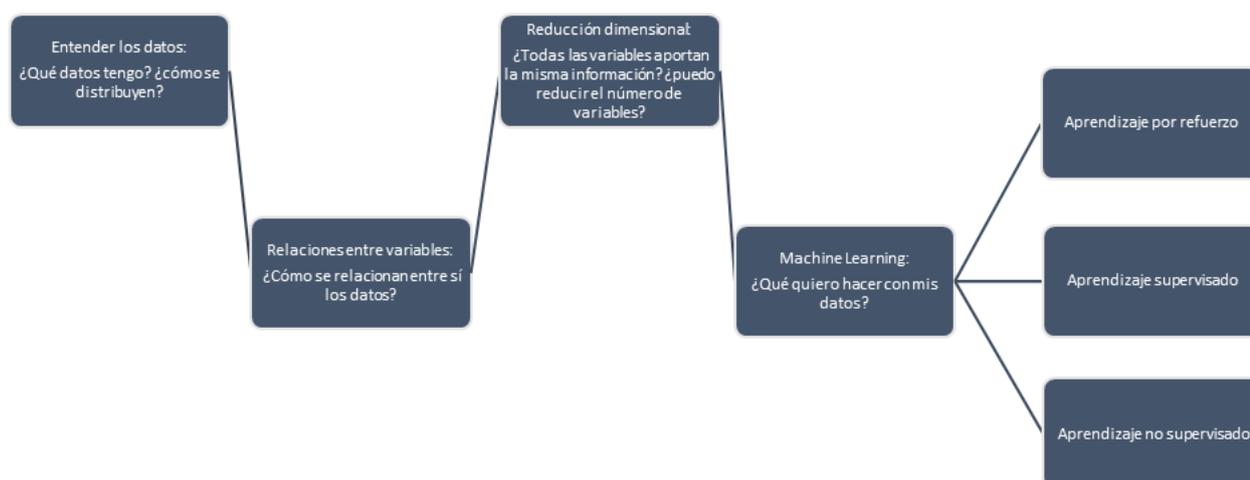


Ilustración 9: Proceso de HRA basado en el Machine Learning.

Fuente: Elaboración propia, adaptado de Aguado (2018)

Aguado, en su libro *HR Analytics: Teoría y práctica para una analítica de recursos humanos con impacto* (Aguado, 2018) propone como comienzo de un análisis de los datos lo que denomina descripción estadística, es decir, partir del análisis de los datos de los que se dispone, de las variables que comprenden estos datos y cómo se distribuyen. El objetivo es dar respuesta a las preguntas: ¿qué datos tengo y cómo se distribuyen? El ejemplo que propone es ver cómo se distribuyen las horas de ausencias en la organización de forma global y por departamentos.

El segundo paso requiere estudiar las relaciones entre las variables. El analista debe desde el principio saber qué función desempeña cada variable, es decir, conocer e interpretar cuáles son las variables explicativas y cuáles son las dependientes. En este nivel se pretende conocer la interrelación de unas variables con otras y para ello resulta muy útil aplicar, entre otras, las técnicas de análisis de correlaciones, de tablas de contingencia, regresiones lineales y logísticas, etc. En función del análisis que se quiera realizar convendrá poner en práctica unas u otras.

Analizados estos datos, y conocidos cuáles son los pesos y cómo influyen las variables dependientes sobre las explicativas, e incluso cómo se relacionan las dependientes entre sí, interesa plantearse la información que aportan, eliminando las que no resulten relevantes para el estudio en cuestión. “En esos casos empleamos técnicas que nos permitan reducir un gran número de variables a unas cuantas que explicando un gran porcentaje de la variabilidad del conjunto nos ayuden a tener una visión más manejable de la realidad” (Aguado, 2018: 149). En la literatura, los dos tipos de técnicas más empleados en la reducción dimensional son: el LDA y el PCA. Siendo ambos conceptos muy técnicos a nivel estadístico y abarcando su aplicación a diversas cuestiones, conviene diferenciar entre ambos conceptos: “Teniendo en cuenta todos los datos, el PCA calculará un vector que tenga la mayor varianza asociada (...). Por otro lado, el LDA calculará un vector que discrimine mejor entre las dos clases” (Martínez y Kak, 2001: 228-229). En la Ilustración 10¹, se puede observar un ejemplo de cómo el LDA clasifica en comparación

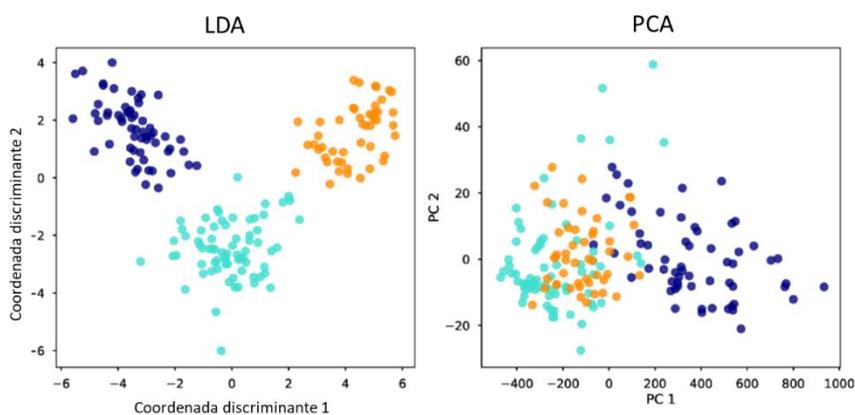


Ilustración 10: Visualización LDA y PCA.

Fuente: Elaboración propia adaptado de Xiaozhou (2020)

¹ Debe señalarse que la Ilustración 10 es un ejemplo. De ahí la importancia de indicar que puede inducir a error el pensar que la discriminación de los valores siempre se observará de forma más clara con el uso del LDA frente al PCA y esto no siempre será así.

con el PCA, dos formas igual de validas, el uso de uno u otro dependerá del objetivo principal del análisis.

En cuarto lugar, tomando como punto de partida el objetivo de aplicar el *Machine Learning*, se deberá dar respuesta a la pregunta: ¿qué quiero hacer con mis datos? Puesto que, en función de la respuesta, se aplicarán unos u otros algoritmos. En cuanto al término de *Machine Learning* (ML), en la literatura encontramos diversas definiciones:

- Una habilidad indispensable para hacer sistemas capaces de identificar patrones entre los datos para hacer predicciones. Es la capacidad que tienen las máquinas de aprender a partir de los datos y tiene como base fundamental la estadística (BBVA, 2019).
- “El Machine Learning es una disciplina del campo de la Inteligencia Artificial que, a través de algoritmos, dota a los ordenadores de la capacidad de identificar patrones en datos masivos y elaborar predicciones (análisis predictivo). Este aprendizaje permite a los computadores realizar tareas específicas de forma autónoma, es decir, sin necesidad de ser programados” (Iberdrola, n.d.).

Dentro del ML, aparecen tres tipos muy concretos de aprendizaje automático: aprendizaje por refuerzo, aprendizaje supervisado y aprendizaje no supervisado.

“Los algoritmos de aprendizaje por refuerzo definen modelos y funciones enfocadas en maximizar una medida de “recompensas”, basados en “acciones” y al ambiente en el que el agente inteligente se desempeñará” (Luna, 2018).

MODELO DE APRENDIZAJE POR REFUERZO

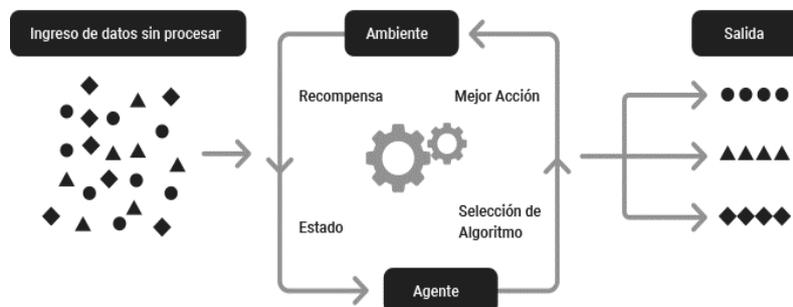


Ilustración 11: Modelo de aprendizaje por refuerzo (Luna, 2018)

Se entiende por aprendizaje supervisado aquel donde se conocen tanto las entradas como las salidas deseadas (etiquetas) y el sistema aprende a asignar las entradas a las salidas.

La clasificación y la regresión son ejemplos de aprendizaje supervisado: en la primera los resultados toman valores discretos (etiquetas de clase) mientras que en la regresión los resultados son continuos. Algunos ejemplos de algoritmos de clasificación son: el vecino más cercano, la regresión logística y la máquina de vectores de apoyo (SVM), mientras que los ejemplos de regresión incluyen la regresión de vectores de apoyo (SVR), la regresión lineal y la regresión polinómica (L'Heureux *et al*, 2017).

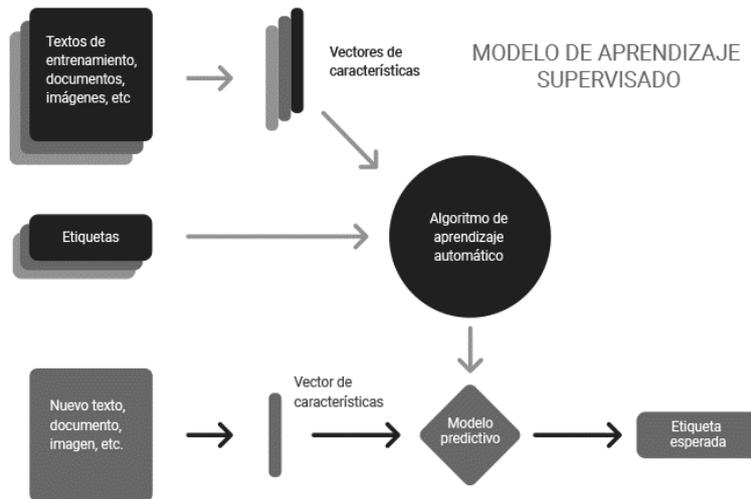


Ilustración 12: Modelo de aprendizaje supervisado (Luna, 2018)

El aprendizaje no supervisado aparece cuando no se conocen los resultados deseados y el propio sistema descubre la estructura de los datos. El aprendizaje no supervisado incluye el *clustering*, que agrupa objetos basándose en criterios de similitud establecidos; *k-means* es un ejemplo de este tipo de algoritmo (L'Heureux *et al*, 2017).

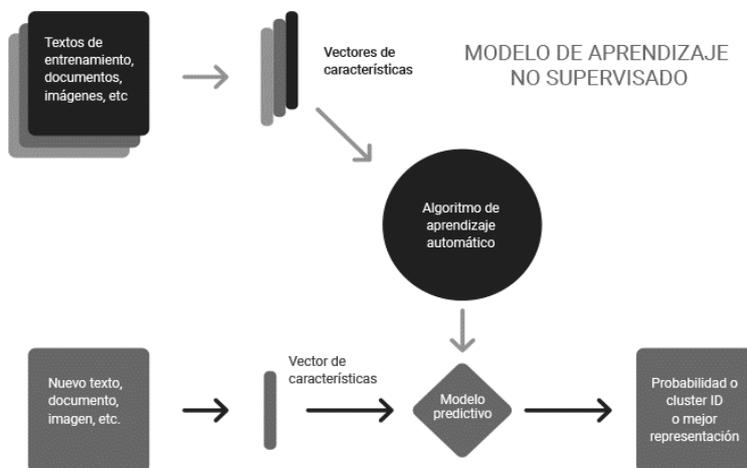


Ilustración 13: Modelo de aprendizaje no supervisado (Luna, 2018)

3. CASO PRÁCTICO

En este apartado se explicará el procedimiento empleado para la realización del caso empírico, junto con los resultados obtenidos. El primer paso será explicar la procedencia de los datos usados en el análisis, cómo estos se han completado con otros necesarios para el estudio presente y la justificación de la composición de cada variable. En segundo lugar, se tratará de explicar qué método de reducción dimensional se va a emplear, seguido por el análisis clúster, tratando de comprender las principales diferencias entre los grupos generados. Y por último se analizará si los resultados obtenidos tienen sentido y qué posibilidades de actuación tendrían los encargados de RR.HH. frente a dichos datos.

3.1. EXPLICACIÓN Y JUSTIFICACIÓN DE LOS DATOS SIMULADOS

Inicialmente, de cara a la realización del análisis, se intentó tener acceso a datos reales de empleados, solicitándolos para ello en algunas empresas que, si bien entendieron la utilidad de estos, la rigurosa política de protección de datos les impidió facilitarlos. Ante esta situación, lo que se planteó fue la posibilidad de generar unos datos simulados; es decir, partiendo de una investigación inicial crear unos datos que, de alguna forma, se acercaran lo máximo posible a la realidad. Con este fin, se eligió como plantilla la base de datos propuesta por Huebner y Patalano, diseñada precisamente para suplir la ausencia de estos en el ámbito de los RR.HH. La base se diseñó con fines académicos y, de hecho, se utiliza en el curso de *HR Metrics and Analytics* que desarrollan en la New England College of Business, para enseñar a profesionales de RR.HH. a trabajar con datos y el análisis de estos (Huebner, 2021).

Con el objetivo de explicar de la forma más clara posible, cómo se han elaborado las modificaciones en los datos mencionados, se dividirá este bloque en diversos subapartados los cuales componen el total del Excel en el que se ha trabajado para el posterior análisis. Para una mayor comprensión, las variables que provengan de los datos propuestos por Huebner y Patalano, se denominarán variables originales, mientras que las demás serán las variables complementarias. En todo caso, conviene señalar que las variables y especialmente las complementarias, fueron revisadas y cotejadas *a posteriori*

por el departamento de RR.HH. de una entidad bancaria europea, con sede en España, que insistió en proteger su anonimato en este estudio.

3.1.1. Variables demográficas o situacionales del empleado

Estas columnas nos permiten tener una primera aproximación a las características de los empleados y su situación dentro de la empresa:

- Empleado ID (variable complementaria): es el número de identificación del empleado. Consiste en un número comprendido entre el 3.040 y el 3.248.
- Sexo (variable original): clasifica entre hombre (1) y mujer (0).
- Casado (variable original): indica si la persona está casada (1) o no (0).
- Edad y años en la empresa (variables originales).
- Años puesto anterior, años puesto actual, años fuera de la empresa y años en el sector (variables complementarias): el objetivo de estas columnas es tratar de calcular un valor de la rotación de los empleados en esta empresa simulada; no todas ellas son empleadas en el análisis.
- Rotación interna (variable complementaria): devuelve un valor comprendido entre 0 y 1, cuanto más próximo sea este valor a uno, más reciente será su rotación. Se calcula con la siguiente fórmula: $\frac{\text{Años en la empresa} - \text{Años puesto actual}}{\text{Años en la empresa}}$. Siendo una aproximación muy limitada, puesto que no tiene en cuenta si una persona ha rotado varias veces en los años que lleva en la empresa, permite tener una aproximación de cómo es la tasa de rotación interna de la empresa.
- Departamento (variable original): muestra a qué departamento pertenece cada empleado: tecnología de la información (*IT*); ventas (*sales*); producción (*production*); ingenieros de software (*software engineering*); administración de oficinas (*admin office*); altos ejecutivos (*executive office*).

En la siguiente ilustración se muestran las variables mencionadas, las originales señaladas en amarillo, mientras que el resto constituyen las variables complementarias.

Empleado_ID	Sexo	Casado	Edad	Años_empresa	Años_puesto_anterior
3040	1	0	47	5	8
3041	0	1	31	9	3
3042	0	0	34	8	8
3043	1	0	43	8	1
3044	1	0	57	5	2

Años_puesto_actual	Rotación_interna	Años_sector	Años_fuera_empresa	Departamento
3	0,4	17	12	Sales
6	0,33	12	3	Sales
6	0,25	8	0	Sales
3	0,63	9	1	Sales
0	1	12	7	IT/IS

Ilustración 14: Variables demográficas y situacionales.

Fuente: Elaboración propia a partir de Huebner y Patalano (2021)

3.1.2. Variables sobre el compromiso y la satisfacción del empleado y su paquete retributivo. Formación dentro y fuera de la empresa

Un segundo bloque de variables que encontramos en nuestro Excel final tiene que ver con la satisfacción y compromiso del empleado y el paquete retributivo que recibe. Las principales variables originales que se encuentran en este apartado son el grado de compromiso, el grado de satisfacción y el salario. Partiendo del supuesto de que el salario, para los creadores de los datos originales, se corresponde con el sueldo fijo, se ha complementado incluyendo el salario variable relacionado directamente con la evaluación recibida por el inmediato superior, y el salario flexible compuesto por otro tipo de retribuciones en especie que comprende los tickets restaurant, ticket guardería, abono de transporte, y acciones específicas de formación para el empleado. Para llevar a cabo esta simulación se establecieron los límites de estas cuantías en los máximos permitidos para estar exentos de tributación, y sin superar el 30 por ciento del salario fijo.

Por otro lado, directamente relacionadas con este tipo de variables, aparecen aquellas que tienen que ver con la formación del empleado (variables complementarias), y que, a su vez, se pueden dividir en dos tipos: externas o internas. La formación externa, indica si el empleado está recibiendo una formación académica de especialización y cuál es el porcentaje económico que la empresa asume de este Máster o Curso. Para establecer un valor promedio en la cuantía de estos cursos, se ha tomado como referencia los ofrecidos por CUNEF en su página web (CUNEF, 2021). La formación interna, hace referencia al

porcentaje cumplimentado por el empleado de la formación que recibe dentro de la empresa, y que se relaciona con su grado de compromiso.

Grado_compromiso	Satisfacción	Ticket_restaurant	Ticket_guarderia	Abono_viaje	Acciones	€_formacion
1,93	2	1	0	0	1	0
2,81	4	0	1	0	0	17943,3
4,77	5	0	0	1	1	15396,5
4,3	4	0	0	1	0	17062,2

%_formacion	Total_flexible	variable	Salario_fijo	Salario_total	Coste_formación_externa	Formación_propia	Formación_empresa
0	1473	616	61568	63657	0	0	0,4
0	1475	579	57859	53922	0	0	0,5
0,7	20825	7424	74241	102490	20883	1	1
0	2347	3266	65310	70923	0	0	0,7

Ilustración 15: Compromiso, satisfacción, retribución y formación de los empleados.

Fuente: Elaboración propia a partir de Huebner y Patalano (2021)

3.1.3. Variables sobre la personalidad y su relación con el puesto de trabajo

Estas variables se incorporaron porque entendemos que la personalidad de los empleados podría encajar en mayor o menor medida, en función del puesto de trabajo en el que se encuentre el empleado, para acercarnos más a la realidad empresarial y poder entender en conjunto todas las variables manejadas. Se partió del modelo de las Cinco Grandes, que trata de conciliar y unificar las teorías acerca del efecto de la personalidad en el mundo laboral. En concreto, una buena parte de autores coinciden que en la personalidad confluyen cinco factores, aunque parece ser que no hay acuerdo en la teoría acerca de los nombres que reciben y en la proporción en la que influye cada uno de ellos. Entre los que se utilizan más frecuentemente destacan, siguiendo a Jaén (2010): extraversión, conformidad, tesón, estabilidad emocional y apertura a la experiencia. Estas variables puestas en relación con el grado de diversidad del puesto y en función del trabajo en equipo, permitirán obtener una puntuación según el departamento al que pertenezcan. Se encontró que el tesón, la estabilidad emocional y la conformidad correlacionan positivamente con la productividad. Además, se observa que las personas extrovertidas y con apertura a la experiencia tienen mayor facilidad para tareas con alto grado de diversidad, y que la estabilidad emocional suele ser un buen parámetro para predecir la capacidad de trabajo en equipo (Jaén, 2010).

En definitiva, la creación de estas variables complementarias persigue conseguir obtener una variable que podemos denominar “ajuste de personalidad con el puesto de trabajo”, la cual nos indica si la personalidad de un empleado encaja de forma excelente (3), bien

o de forma adecuada (2), o no satisfactoriamente (1) con su puesto de trabajo. Al mismo tiempo, esta puntuación se correlaciona con el desempeño.

Además, y a pesar de que no está directamente relacionada, se ha incluido la variable original denominada “ausencias”, que nos indica el número de veces que un empleado ha faltado a la jornada laboral.

Personalidad				
Extraversión	Conformidad	Tesón	Estabilidad_emocional	Apertura_experiencia
0	1	1	1	0
0	1	1	1	0
1	1	1	1	1
0	1	1	1	1
Relación con el puesto de trabajo				
Grado_diversidad_tarea	Grado_trabajo_equipo	Ausencias	Ajuste_personalidad_puesto	
2	2	5	3	
2	2	16	3	
2	2	14	2	
2	2	13	2	

Ilustración 16: Personalidad y su relación con el puesto de trabajo.

Fuente: Elaboración propia a partir de Huebner y Patalano (2021)

3.1.4. Variables indicativas del rendimiento. Potencial y desempeño: Matriz 9 cajas

En este grupo de variables encontramos diferentes medidas de la evaluación de los empleados. En primer lugar, el cumplimiento de los objetivos trata de una variable original que puntúa del 1 al 4 en función del grado de cumplimiento por parte del empleado. Por otro lado, se añadió una columna que corresponde con la evaluación del *manager*, y que toma en consideración las variables que comprenden el cumplimiento de objetivos, ajuste personalidad y puesto, formación interna y externa, el grado de satisfacción del empleado y de compromiso con la entidad. Tratando así de simular una puntuación (del 1 al 10) recibida por parte del inmediato superior del empleado evaluado.

Junto a ello, se encuentran las medidas de potencial y desempeño; sin duda, las más relevantes del estudio puesto que permiten generar la matriz de las 9 cajas. El potencial se calculó considerando las siguientes variables: la edad, los años en el puesto actual, los años en el sector y la formación externa e interna de la empresa. Sin embargo, el desempeño se midió observando la evaluación del *manager*, el cumplimiento de los objetivos, el ajuste de personalidad con el puesto, el porcentaje de formación interna realizado, la satisfacción en el puesto y la rotación reciente del empleado. Ambas

columnas devuelven un valor entre 1 y 3, permitiendo la combinación de ambas elaborar la matriz de las 9 cajas.

Calificación rendimiento		9 cajas			
Cumplimiento_objetivos	Evaluación_manager	Potencial	Desempeño	Resultado_eval	Eval_ID
1	5	3	1	Enigma	7
3	5	2	2	Empleado clave	5
4	8	3	3	Futuro lider	9
1	6	2	2	Empleado clave	5

Ilustración 17: Rendimiento en el puesto de trabajo; potencial y desempeño.

Fuente: Elaboración propia a partir de Huebner y Patalano (2021)

3.2. ANÁLISIS DESCRIPTIVO DE LOS DATOS

Explicados los elementos que se han considerado en la elaboración de los datos, a continuación, se analizan los principales resultados obtenidos en la simulación. Debido a las limitaciones de espacio y tiempo, este análisis descriptivo trata de ser lo más conciso posible partiendo de la distribución de los empleados en la matriz de las 9 cajas para obtener algunas conclusiones acerca de las principales características que se derivan de la muestra, sin embargo, para una mejor comprensión y un análisis más exhaustivo, en el anexo del estudio pueden consultarse los gráficos resultantes.

A través de este gráfico (Gráfico 1), se puede apreciar una primera distribución de los empleados en función de su desempeño y de su potencial. Se observa como de los 209 empleados evaluados, el 61 por ciento de estos no debería dar ningún problema inicialmente puesto que tienen una puntuación mayor o igual a 2 tanto en potencial como en desempeño. En cambio, al otro 39 por ciento habrá que prestarles una mayor atención puesto que de la puntuación obtenida se desprende que algunos no rinden lo suficiente y que otros están más acomodados en su puesto de trabajo al no tener potencial suficiente para promocionar. Otro aspecto relevante es que, si comparamos el número de empleados con un desempeño mayor o igual a 2 y con potencial menor a 3 frente al número de empleados con estos mismos resultados respecto al potencial, se puede observar cómo la suma de empleados “efectivos”, “confiables” y “futuras estrellas” frente a la de “dilemas”, “enigmas” y “estrellas en crecimiento” es el doble, indicándonos que en esta compañía se prioriza el desempeño frente al potencial. A grandes rasgos, podríamos concluir que la empresa analizada no tiene problemas relevantes con los resultados de sus empleados, aunque es destacable la falta de potencial; sin embargo, para asegurarnos de

esta afirmación merecería la pena observar cómo se distribuyen las variables que pueden explicar en mayor medida este resultado acerca del desempeño y potencial de la muestra.

Con los siguientes párrafos lo que se busca es profundizar (fijándonos en el salario, la formación externa recibida y la interna cumplimentada, la rotación interna y el grado de satisfacción de los empleados) en el análisis de la matriz anterior.

Matriz de las 9 cajas

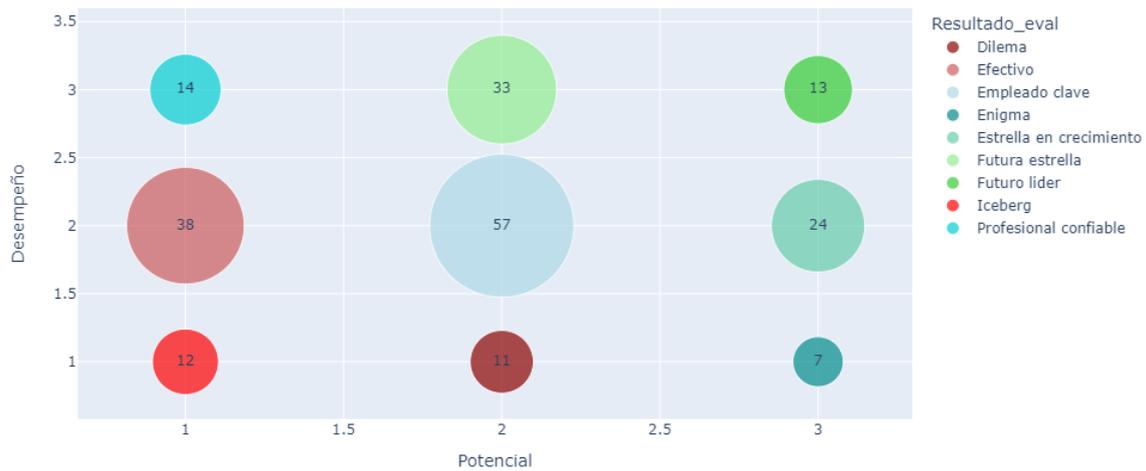


Gráfico 1: Matriz de las 9 cajas de los datos simulados.

Fuente: Elaboración propia.

En el Gráfico 2, podemos observar como la distribución del salario total sigue una distribución normal, cuya media toma el valor 82.910€. Por otro lado, se aprecia una serie de valores atípicos especialmente en la cola derecha de la distribución, debido a que suponen salarios muy superiores a la media, correspondientes a los puestos directivos; sobre todo, llama la atención la presencia de empleados “efectivos” y “dilema”. Sin embargo, en el lateral izquierdo se encuentran aquellos empleados con salarios de menor cuantía, a pesar de encontrar alguno que corresponde a los grupos “profesionales confiables”, “enigma”, “empleados clave” y “futuras estrellas”, predominan especialmente los empleados “efectivos” e “iceberg”. Esto nos puede llevar a pensar que en los puestos más bajos esta entidad no está invirtiendo mucho en el potencial de sus empleados, sino que prefiere que tengan un desarrollo aceptable. Respecto a los valores situados en torno a la media, si nos fijamos en el lateral derecho, destaca la presencia de empleados con medio y alto desempeño (efectivos, clave, confiables, futuras estrella y futuros líderes) lo que nos puede estar indicando que la empresa está tratando de incentivar a estos empleados a aumentar su potencial.

Distribución del salario total de los empleados:

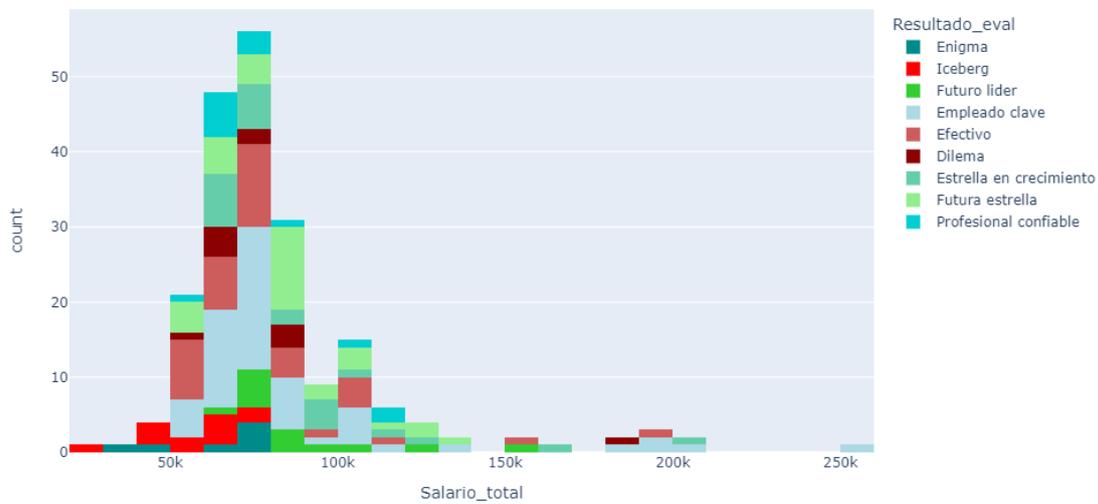


Gráfico 2: Distribución del salario total dentro de la empresa.

Fuente: Elaboración propia.

Sin embargo, el Gráfico 3 nos muestra el porcentaje de la formación externa, que reciben los empleados, pagada por la empresa. Aquellos que reciben el valor cero nos indica que no están recibiendo ninguna formación (*master*) y como se puede apreciar el número de empleados, en esta situación, es muy elevado. Por otro lado, entre aquellos que sí reciben formación externa, sobresale el número de “empleados clave” y de “futuras estrellas”. De los cuales, quienes aprovechen correctamente esta formación recibida pasarán a ser “estrellas en crecimiento” y “futuros líderes” respectivamente.

Porcentaje de formación externa (Master) pagado por la empresa a los empleados:

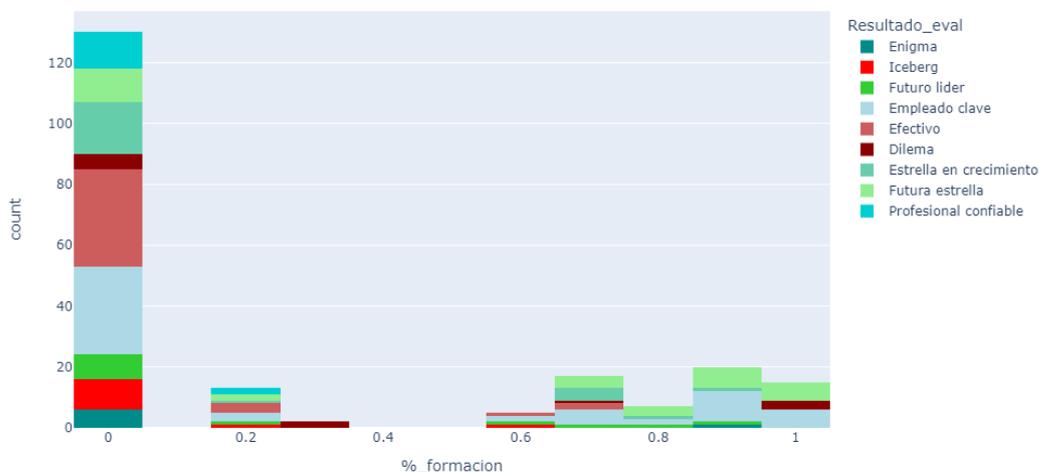


Gráfico 3: Porcentaje de la formación externa que paga la empresa a cada empleado.

Fuente: Elaboración propia.

Por otro lado, si nos fijamos en el grado de compromiso y satisfacción de los empleados (Gráficos 4 y 5) podemos observar que ambos valores obtienen puntuaciones medias altas, pero hay margen de mejora pues aparece el peligro de que empleados con alto potencial decidan no quedarse en la empresa. A este respecto, se debe señalar la insatisfacción y bajo compromiso de empleados “enigma” los cuales, teniendo un alto potencial, en una empresa donde aparentemente se valora, en mayor medida, el desempeño, pueden llegar a no verse lo suficientemente valorados como para quedarse.



Gráfico 4: Grado de compromiso del empleado con la empresa.

Fuente: Elaboración propia.

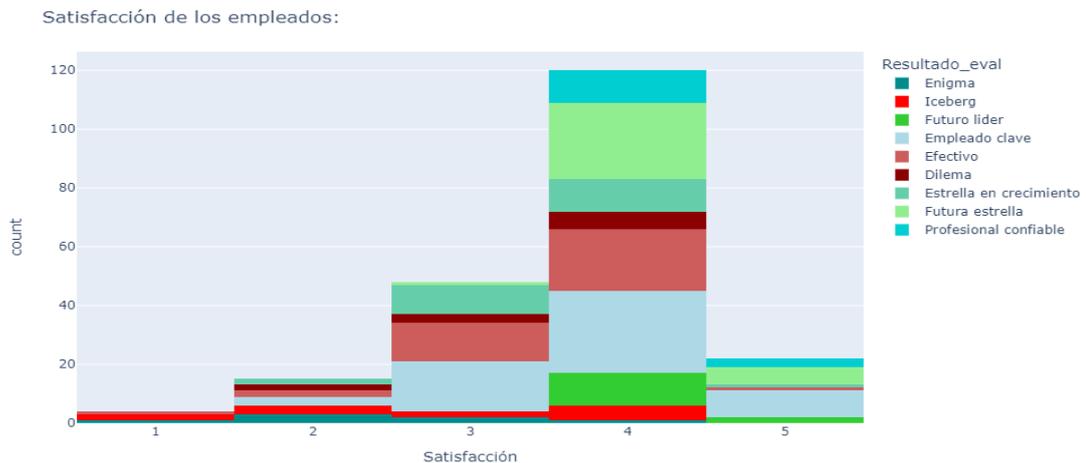


Gráfico 5: Puntuación del 1 al 5 de la satisfacción de los empleados.

Fuente: Elaboración propia.

Sin embargo, puede verse como un síntoma positivo, que los empleados con medio y alto desempeño se encuentren bastante satisfechos y comprometidos, pues de alguna forma permite, sin ser muy laxos, respirar tranquilamente al departamento de RR.HH. pues no existe un gran peligro de que estos empleados vayan a decidir marcharse.

Finalizando el análisis descriptivo, conviene observar las tendencias en la rotación interna de la empresa, puesto que nos permitirá entender si en esta organización simulada, se está usando correctamente la matriz de las 9 cajas, puesto que, si se toma esta matriz como sustitutiva de la matriz de desarrollo y potencial será una de las herramientas que permita desarrollar de forma óptima y coherente los planes de desarrollo de carreras dentro de la organización.

Siendo esto así, el Gráfico 6 nos da a entender dos posibles conclusiones compatibles entre sí. En primer lugar, debido al origen diverso, aleatorio y simulado de los datos se observa como al diseñar esta variable, no se ha tenido en cuenta el desarrollo y potencial de los empleados, causando esto posibles incoherencias como la reciente rotación de empleados “iceberg” o la no rotación interna de empleados con alto potencial y desempeño. Es cierto que, si una persona con estas características (alto potencial y desempeño) no ha rotado desde su incorporación en la compañía, no es un indicador absoluto de que los datos estén mal desarrollados, esta situación también podría derivarse de que lleve poco tiempo en la entidad. Por ello, la segunda conclusión que se puede deducir es que esta empresa, al poner en práctica el uso de la matriz de las 9 cajas, está



Gráfico 6: Rotación reciente de los empleados.

Fuente: Elaboración propia.

teniendo algunas dificultades en el desarrollo de los planes de carrera de sus empleados. Es decir, a través del análisis descriptivo, se pueden identificar sesgos (por ejemplo: empleados mal promocionados) e incoherencias en la aplicación de políticas de RR.HH.

3.3. CLUSTERING

Para el proceso de clusterizado de nuestros datos, los pasos que seguiremos serán una combinación empleando un modelo supervisado en la reducción dimensional y de un modelo no supervisado en la agrupación de los individuos. Es decir, la dimensión de las variables las reduciremos teniendo en cuenta como variable objetivo la posición que ocupa cada individuo en la matriz de las 9 cajas. Por el contrario, en el agrupamiento de los individuos dejaremos que sea el algoritmo quien establezca los grupos en función de las características globales de los individuos.

3.3.1. Modelo Supervisado: Reducción dimensional

Como se explicó en el epígrafe 3.2.2 sobre el marco teórico, una vez realizados los principales análisis descriptivos de las variables más importantes que constituyen nuestros datos, y antes de pasar a realizar los clústers y su posterior análisis, es aconsejable realizar una reducción dimensional. En términos generales, con la reducción dimensional de las variables lo que se busca es “convertir un conjunto de datos de dimensiones elevadas en un conjunto de datos de dimensiones menores, asegurando que la información que proporciona es similar en ambos casos” (Softtek, 2021). Sin duda, este es nuestro caso, puesto que contamos con un número elevado de variables, alrededor de cuarenta, y siendo conscientes de que no todas las variables aportan la misma información, hemos decidido realizar una reducción considerable de las variables de nuestro estudio.

Señalados brevemente en epígrafes anteriores, cuando se habla de reducción dimensional en *machine learning* destacan el LDA y el PCA. Para este estudio aplicaremos el primero puesto que “se trata de un método de clasificación supervisado de variables en el que dos o más grupos son conocidos *a priori* (en nuestro caso, la posición en la matriz de las 9 cajas), y las observaciones se clasifican en uno de ellos en función de sus características.

Haciendo uso del teorema de Bayes, LDA estima la probabilidad de que una observación (un empleado), dado un determinado valor de los predictores, pertenezca a cada una de las clases de la variable cualitativa² (posición en la matriz de las 9 cajas) (Amat, 2016). Cabe señalar que a través del empleo de este algoritmo hemos conseguido reducir de 40 variables a 3 las cuales consiguen explicar un 92,28 por ciento de la varianza. En la práctica se suele tomar el 80 por ciento como un buen indicador de que no se ha perdido mucha información con la reducción dimensional, lo cual es un buen indicativo de la correcta reducción dimensional.

3.3.2. Modelo No Supervisado: Agrupación

Llegados a este punto se debe explicar que cuando se habla de técnica de agrupamiento o clusterización hay diversos algoritmos que pueden ser empleados para agrupar un conjunto de datos similares (*kmeans* o k-medias, método gaussiano, clúster por espectros, agrupamiento aglomerativo, por densidad, etc.). En este estudio, se aplicaron inicialmente los métodos *kmeans*, método gaussiano y clúster por espectros, finalmente se decidió usar solamente el método *kmeans* debido a que los tres algoritmos empleados eran muy parecidos en sus resultados y sólidos en las agrupaciones elaboradas. La principal razón por la que se optó por este método fue que “es uno de los algoritmos de aprendizaje no supervisado más simples para resolver el problema de la clusterización. El procedimiento aproxima por etapas sucesivas un cierto número (prefijado) de clústers haciendo uso de los centroides de los puntos que deben representar” (Sancho, 2015). Uno de los puntos que no debe pasar por desapercibido es que el número de clústers asignados, es algo prefijado por el usuario del algoritmo. Para saber qué número de grupos determinar, se empleó la teoría del codo que consiste en la aplicación de “un método que examina el porcentaje de varianza explicada en función del número de conglomerados. Este método se basa en la idea de que hay que elegir un número de conglomerados tal que la adición de otro no mejore mucho la modelización de los datos (Bholowalia y Kumar, 2014). Es por ello, que en este caso se opta por seleccionar el número 3 como valor óptimo para la creación de los conglomerados o clústers. En los Gráficos 7 y 8 se puede observar cómo

² Debe señalarse que lo que aquí se denomina como “clases de la variable cualitativa” es comúnmente conocido como “etiquetas”. Es decir, aquel parámetro que, en *machine learning*, hace que un modelo sea o no supervisado.

es dicha agrupación y como efectivamente se aprecia una división clara entre los grupos generados. En el siguiente epígrafe se entrará en detalle a analizar las principales diferencias y posibles actuaciones entre los grupos desde una perspectiva de RR.HH.

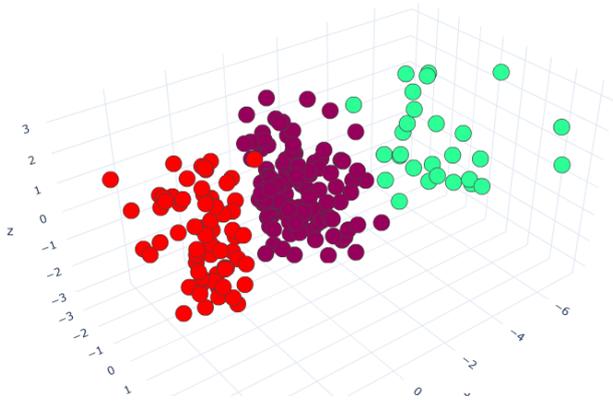


Gráfico 7: Gráfico tridimensional de agrupación de los individuos.

Fuente: Elaboración propia.

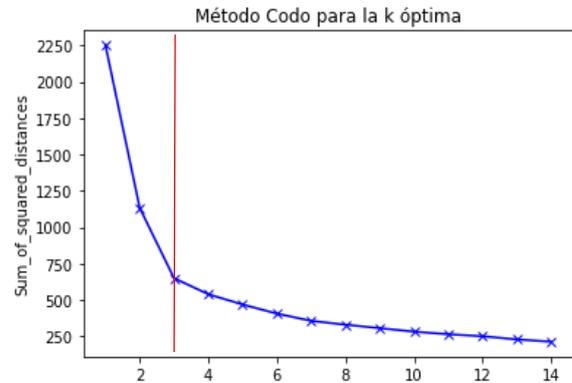


Gráfico 8: Método del Codo, para obtención del número de clúster óptimo.

Fuente: Elaboración propia.

3.4. ANÁLISIS CLÚSTER: PROPUESTAS DE ACTUACIÓN

Una vez determinado el número de clústers que se debe sacar para la muestra y agrupados los individuos en cada uno de estos grupos, el objetivo de este epígrafe es analizar cómo están compuestos y tratar de ver como se distribuyen las características de los empleados en función del clúster al que pertenecen y si estas son lo suficientemente heterogéneas entre empleados que pertenecen a distintos grupos. El propósito de esto será analizar las distintas políticas que debería aplicar el departamento de RR.HH. de la institución para conseguir mejorar el rendimiento de sus empleados. Cabe señalar que las políticas que se proponen solo son aplicables en un escenario ideal, es decir no teniendo en cuenta límites presupuestarios para la actuación del departamento de RR.HH.

En el Gráfico 9 podemos observar cómo se han formado los grupos. En general, se han clasificado en función de su desempeño, lo que nos permite confirmar lo que apuntábamos en epígrafes anteriores de que, en este caso concreto, la empresa simulada tiende a dar una mayor importancia al desempeño que al potencial. Sin embargo, se aprecia que hay determinados individuos en cada clúster, que no pertenecen al mismo nivel de desempeño que los otros empleados pertenecientes a su propio grupo. La razón de esto puede ser que el algoritmo, a pesar de no coincidir en el valor del desempeño,

identifica que las características de esos empleados son muy similares a los de su grupo, y este dato puede ser de gran utilidad. Por ejemplo, en el clúster número 0 ($kmeans = 0$), encontramos que todos sus empleados tienen un desempeño igual a 2, salvo tres empleados, los cuales tienen un alto potencial. Podríamos suponer que, a pesar de tener una puntuación insuficiente en desempeño, al tener unas características similares a aquellos empleados con desempeño de nivel 2, si la empresa incentiva con las políticas correctas a estos empleados, podrán llegar sin dificultad a incrementar su nivel de desempeño; este mismo razonamiento se puede aplicar al clúster 2 ($kmeans = 2$). Sin embargo, con el clúster número 1 ($kmeans = 1$) pasa lo contrario, pudiéndose identificar a un empleado que, aun teniendo un desempeño aceptable, ha sido agrupado con los empleados de peor desempeño, sugiriendo así que comparte más características con empleados de desempeño bajo que con los de desempeño medio.

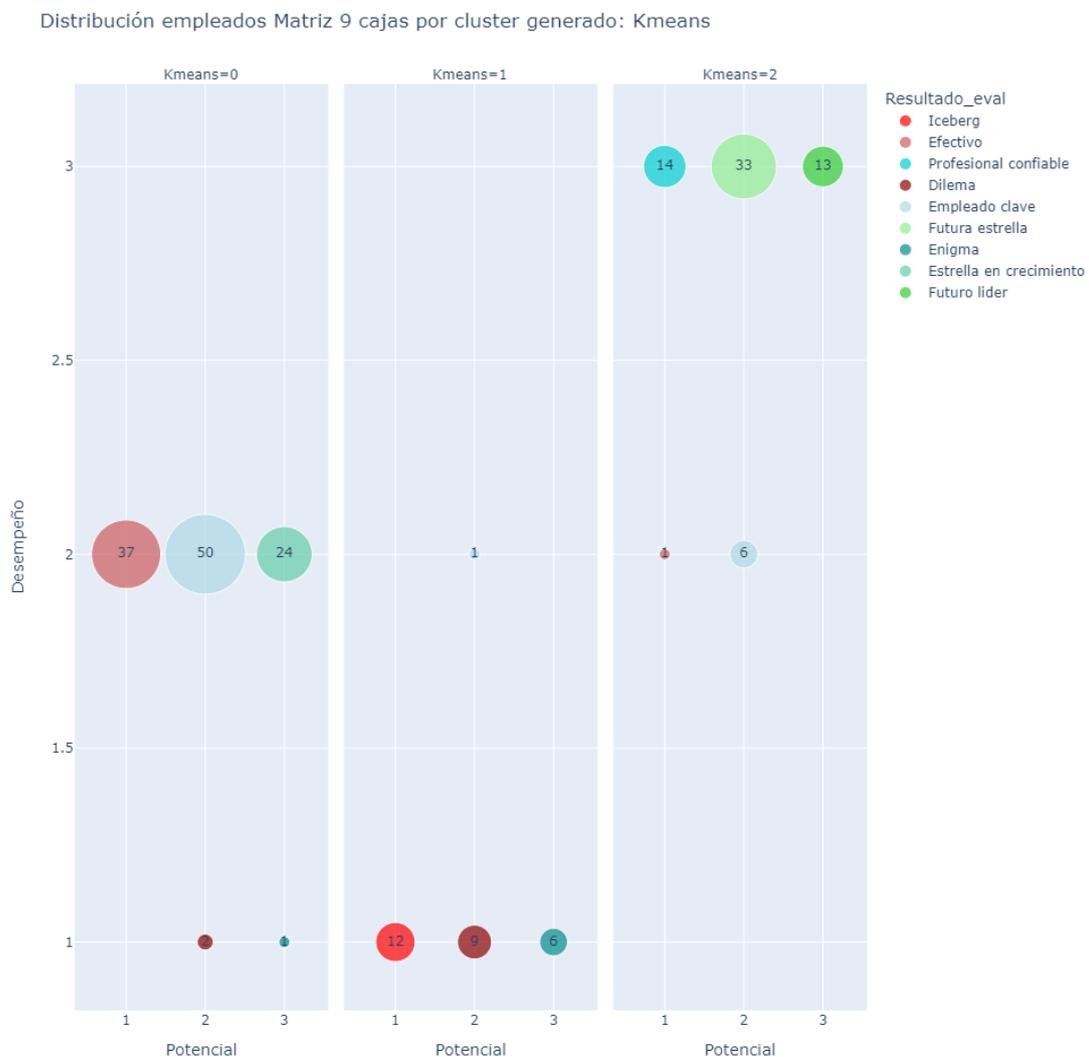


Gráfico 9: Matriz de las 9 cajas de cada clúster generado.

Fuente: Elaboración propia.

Para poder comparar las principales diferencias entre los grupos, haremos un análisis de la distribución de las variables; similar al que hicimos previamente con el total de los datos (salario total, formación externa de la empresa, satisfacción y rotación). En primer lugar, fijándonos en el Gráfico 10 y habiendo calculado la media de la retribución total de cada grupo, se puede observar como el grupo 0, al cual denominaremos grupo de “desempeño aceptable”, es el grupo con mayores sueldos. Es cierto que la principal razón de esto es que los altos ejecutivos que se encuentran situados en este grupo, pero al ser el más numeroso también es el que tiene la media más robusta. En cambio, el grupo 1 o de “bajo desempeño” es el que tiene menores salarios, esto sumado a que, como veremos más adelante, también es el grupo menos satisfecho nos puede dar unas pistas respecto a los planes de actuación por parte de los RR.HH. En cambio, el grupo 2 o de “alto desempeño”, tiene un salario total ligeramente inferior al grupo de “desempeño medio”.

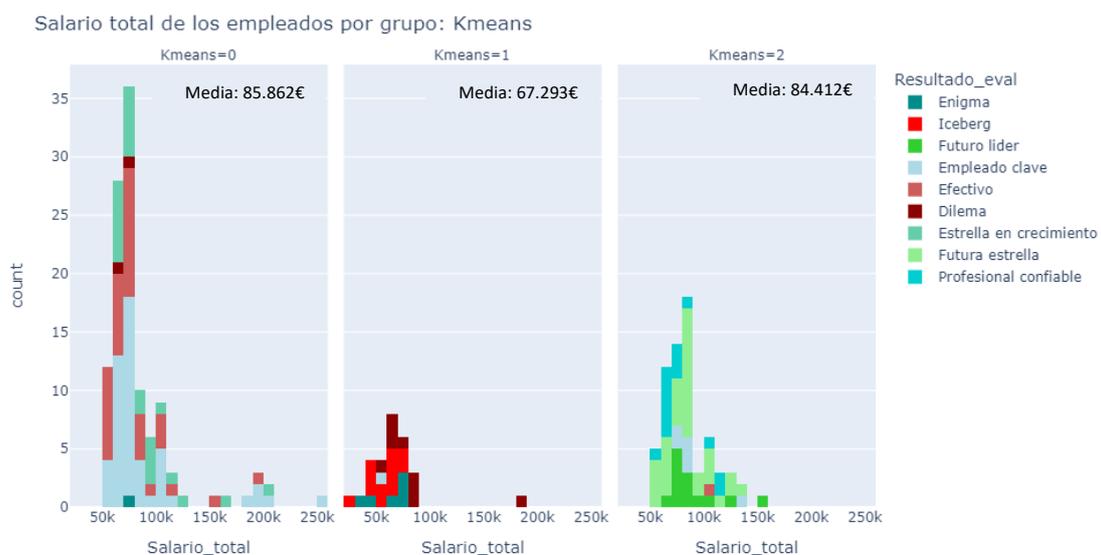


Gráfico 10: Distribución salarios totales por cada clúster generado.

Fuente: Elaboración propia.

Por tanto, lo primero que debería hacer un encargado de recursos humanos de la empresa es ver si esta diferencia salarial, entre el grupo de medio y el de alto desempeño, es causada por diferencia de puestos o si ocurre también para puestos similares, lo que implicaría que nos encontraríamos ante un relevante problema de salarios, puesto que personas con peor desempeño y mismos puestos estaría cobrando sueltos significativamente más altos. Para resolver esta situación se podría incrementar el porcentaje de retribución variable al grupo de “alto desempeño”, con el peligro de que al

no ver consolidado su sueldo total se desmotive y, por otro lado, se podría crear un plan de compensación a medio plazo (2-3 años) con el objetivo de progresivamente establecer unos salarios mayores a estos empleados.

El siguiente gráfico, nos da un poco de luz de cómo es la satisfacción de los empleados en función al clúster que pertenecen. Llama la atención que el grupo de “alto desempeño” tiene una satisfacción mucho mayor, puesto que solamente un empleado tiene una puntuación menor a 4, esto nos puede llevar a pensar que lo que planteábamos anteriormente respecto a los salarios no sea, en el corto plazo y sin haber entrado en un mayor análisis, un problema para la empresa.

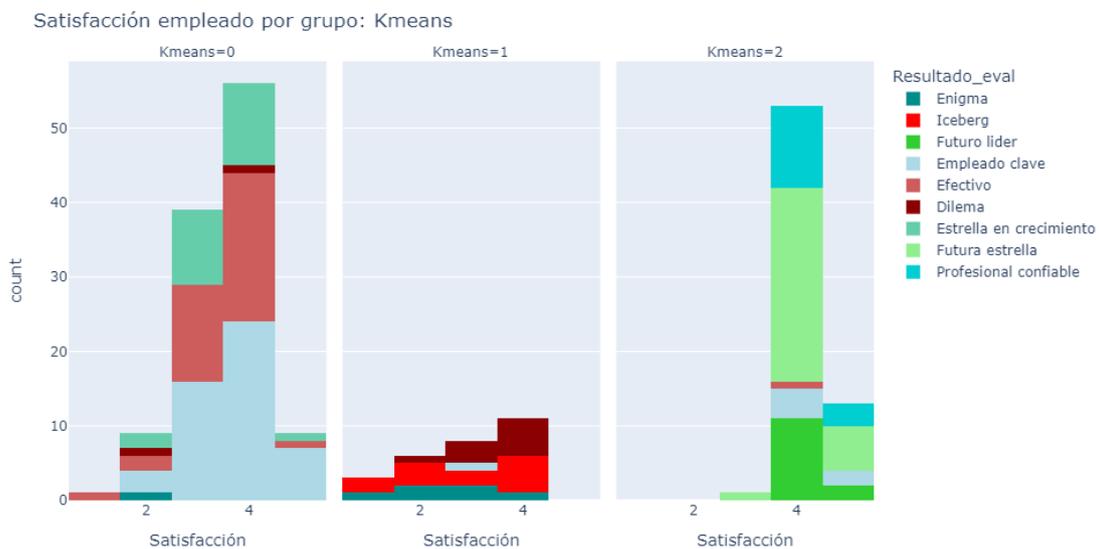


Gráfico 11: Distribución satisfacción por clúster generado.

Fuente: Elaboración propia.

El grupo de “desempeño aceptable” tiene una distribución algo más dispersa. Se ha de prestar especial atención a aquellos empleados con menor satisfacción puesto que podrían reducir su nivel de desempeño siendo esto peligroso especialmente en empleados efectivos que pasarían a ser “icebergs”, y en empleados “estrella en crecimiento” que decayesen a posibles “enigmas”, es decir, una buena parte de este grupo podría tener el peligro de pertenecer al grupo de “bajo rendimiento”. Se sugiere, mantener entrevistas con los empleados menos satisfechos y tratar de llegar a las causas de esta insatisfacción: remuneración; ajuste persona-puesto; ambiente; desarrollo de carrera; etc. Lo más destacable de este tercer grupo serían aquellos empleados “enigma” que, teniendo un alto potencial, no se sienten satisfechos.

En estos casos, convendría hablar con ellos para investigar las causas de dicha insatisfacción y tratar de resolverla pues sería muy favorable para la empresa retener personas con alto potencial y animarlas a mejorar su desempeño. Por otro lado, en este mismo grupo, debe contemplarse la actuación rápida de despido o sustitución de los empleados enigmas poco satisfechos, pues son empleados que están suponiendo un alto coste a la compañía; algo similar ocurre con los empleados “dilema”, en los que conviene valorar si es conveniente esforzarse por aumentar su desempeño o si es preferible mantenerlos hasta encontrar un sustituto, en la propia empresa o al margen de ella.

Sobre la formación recibida externamente, en el Gráfico 12 se observa que la mayoría de los empleados no reciben ninguna formación externa que permita desarrollar su desempeño y su potencial; aunque es cierto que el grupo de “alto desempeño” está más o menos nivelado en cuanto quiénes reciben o no la formación externa. Sin embargo, el grupo de “desempeño aceptable” tiene un escaso número de personas recibiendo formación fuera de la empresa, lo que les limita notablemente su desarrollo profesional. Es aconsejable que la entidad realice un esfuerzo por ofrecer mejores programas formativos externos a este grupo, sobre todo, teniendo en cuenta que es el núcleo más importante, numéricamente hablando, de la organización.

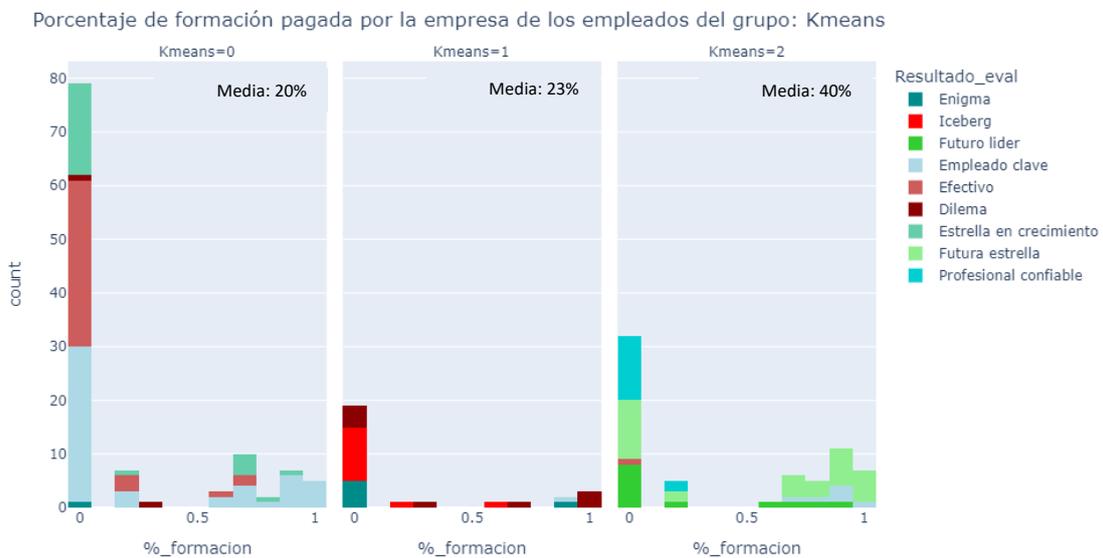


Gráfico 12: Distribución de los porcentajes pagados por la empresa en formación externa.

Fuente: Elaboración propia.

En cuanto a la rotación interna dentro de la empresa, se confirma que la variable tiene un error de creación, es decir, la fórmula que se empleó para calcularla no fue la más adecuada, pues nos devuelve lo reciente que es la rotación de un empleado respecto a los años que lleva en la empresa, no teniendo en cuenta que un mismo empleado puede rotar varias veces. De todas formas, el análisis de esta variable sirve para detectar que esta organización es poco dinámica en cuanto a la rotación interna debido a que en el grupo de “desempeño aceptable” son muchos los que no han rotado desde que llegaron a la empresa. Algo parecido ocurre con el grupo de “alto desempeño”, siendo esto algo que afecta a la empresa puesto que se percibe la presencia de empleados con el potencial y desempeño suficiente para promocionar y que todavía no lo han hecho. Es cierto que se aprecia cierta rotación interna, aunque no muy coherente con la teoría de la matriz de las 9 cajas. Por otro lado, es muy llamativo que en el grupo de “bajo desempeño” haya alguna reciente rotación de empleados “iceberg”, lo que nos puede dar a entender que, o bien es un fallo en la simulación de los datos, o se trata de empleados que han sido recolocados dentro de la organización tratando de aumentar su nivel de desempeño. Se trata de una apuesta arriesgada, habría que vigilar de cerca a este grupo para actuar con rapidez en caso de que no se observe una mejoría inmediata en dichos empleados.

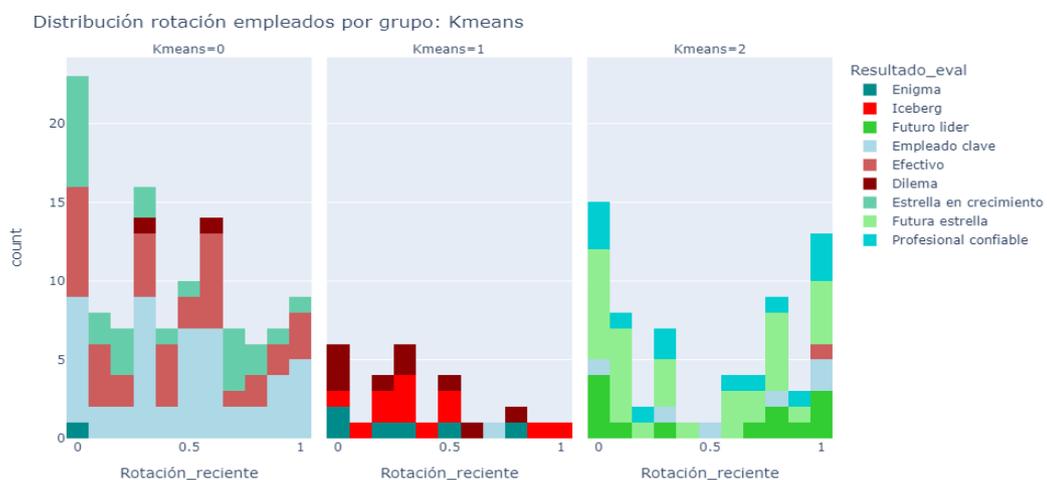


Gráfico 13: Distribución de la rotación interna de los clusters generados.

Fuente: Elaboración propia.

Debe señalarse la importancia de entender que las propuestas de actuación aquí indicadas no deberían ser aplicadas con carácter absoluto para todos los individuos de cada clúster, sino que se tratan de actuaciones genéricas que en la medida de lo posible deberían adaptarse a la situación individual de cada empleado.

En definitiva, en el siguiente cuadro se resumen algunas propuestas de posibles políticas a aplicar en función de los resultados obtenidos a partir de las variables analizadas:

	Grupo bajo rendimiento	Grupo desempeño aceptable	Grupo alto desempeño
Retribución total	Se recomienda despedir a los empleados “iceberg” no satisfechos con la empresa.	Se propone mantener salarios totales. Buscar alternativas de mejora de rendimiento en formación y rotación interna.	Se debería analizar si hay desproporción salarial frente a puestos similares del grupo de desempeño aceptable, en caso de ser así nivelar los salarios totales. Esto se podría hacer con una mayor retribución variable.
Satisfacción	Es conveniente hablar con los empleados con medio o alto potencial y ver las causas de insatisfacción.	Es necesario hablar con empleados con medio o alto potencial y ver las causas de insatisfacción.	Se entiende que con el empleo de las demás políticas debería ser suficiente.
Formación externa	Se debería aumentar la formación de empleados con alto potencial. El propósito de esto es ver si conseguimos aumentar su desempeño.	Se propone aumentar la formación externa en empleados con potencial alto. Prioridad frente al grupo de alto desempeño.	Sería conveniente reducir la formación externa, en la medida que se pueda. Se hace notar que se trata de empleados con alto desempeño y potencial.
Rotación interna³	Se propone la opción de valorar la posibilidad de que empleados con alto potencial, “enigma”, puedan tener mejor desempeño en otro puesto. Habrá que tener en cuenta, si procede, la causa de insatisfacción del empleado.	N.A. ⁴	Se les debe dar prioridad frente a los empleados del grupo de desempeño medio.
Comentarios	<ul style="list-style-type: none"> - Respecto a las políticas aquí aplicadas, se debe advertir del peligro de acabar ofreciendo “café para todos”. Por eso se sugiere que, junto con una perspectiva global de cada grupo generado, también se intente individualizar las políticas a los distintos empleados. - Debemos ser conscientes de los límites presupuestarios a la hora de llevar a cabo las políticas propuestas. 		

³ Debemos ser conscientes del error de medida y simulación que alberga dicha variable.

⁴ N.A: No aplica.

4. CONCLUSIONES

Observando las hipótesis planteadas al principio de este Trabajo de Fin de Grado, y a través del desarrollo de este, se puede concluir que el empleo del *Machine Learning* con el objetivo de analizar los datos de los empleados dentro de la empresa, sí tiene sentido y usándolo correctamente es de enorme utilidad para la empresa, pues permite tomar decisiones de una forma más informada y eficaz. Debe apuntarse igualmente, la importancia que tiene realizar un análisis descriptivo del conjunto de los datos antes de abordar cualquier tipo de algoritmo, pues permite identificar fallos, incoherencias y sesgos en las distintas variables, como lo que se ha podido observar en la elaboración de la variable rotación interna.

La primera hipótesis plantea la posibilidad de agrupar empleados en función de unas variables independientes con el objetivo de tomar decisiones sobre otras dependientes. A través de este estudio se han elaborado unos datos simulados con los que se ha tratado de corroborar esta hipótesis. Así fue al aplicar el algoritmo de *clustering*, el cual agrupó a los individuos de tal forma que segmentó a los empleados en función de su nivel de desempeño: alto, aceptable y bajo. A partir de ahí se pudo desarrollar un análisis descriptivo de cada clúster, compararlos y observar las diferencias entre ellos. A continuación, se propusieron una serie de actuaciones, quedando estas resumidas en el cuadro final del epígrafe anterior, que buscan, de una forma genérica, mejorar el rendimiento de los empleados pertenecientes a cada grupo.

La segunda hipótesis se liga con el párrafo anterior. Se evidencia que las decisiones que se pueden tomar con tres grupos no son lo suficientemente específicas como para ser aplicables a todos los empleados. Por ello, para posteriores estudios académicos se recomienda aplicar un mayor número de clústers que permitan clasificar a los empleados de una forma más precisa pudiendo así aplicar medidas más concretas. Es decir, cuantos más grupos de individuos con características similares haya en la empresa, más reducidos serán estos y, por tanto, las políticas de actuación para la mejora de su rendimiento se podrán diseñar “a medida” para cada grupo de empleados.

Por otra parte, a lo largo del estudio han ido apareciendo diversas limitaciones o dificultades que han impedido realizar un análisis con mayor profundidad y unas conclusiones más específicas. Desde un punto de vista operativo, el análisis realizado en

este estudio ha sido muy genérico puesto que pretendía arrojar algo de luz acerca de la utilidad del manejo de la analítica de datos dentro del ámbito concreto de los Recursos Humanos.

En todo caso, siendo consciente de que el estudio se podía haber acotado algo más, e incluso que se podía haber profundizado en el análisis de las distintas variables que causan los diversos comportamientos de los empleados, y con ello su agrupación, cabe señalar que estas carencias han sido causadas, entre otras razones, por la ausencia de datos reales generados por empresas concretas por la obligada protección de datos. Esta limitación, como ya se explicó, obligó a crear una simulación la cual tiene tantas limitaciones como variables.

Junto a ello, las limitaciones de espacio y tiempo igualmente han sido relevantes para este estudio; sin ellas, probablemente se podrían haber incorporado y analizado un mayor número de clústers que habría permitido generar políticas más específicas para cada grupo de empleados.

Con todo, se ha conseguido aplicar un proceso, un análisis y un desarrollo inicial a un caso empírico del empleo de análisis de datos aplicado a una empresa con el objetivo de mejorar la toma de decisiones orientadas a un rendimiento óptimo de los empleados.

BIBLIOGRAFÍA

- Academy to Innovate HR (2021) *The 9 Box Grid: A Practitioner's Guide* [en línea] disponible en <https://www.aihr.com/blog/9-box-grid/#:~:text=9%20box%20grid%3F-.The%209%20box%20grid%20is%20a%20well%2Dknown%20talent%20management.and%20bringing%20those%20two%20together> [consulta: 04 marzo 2022].
- Aguado, D. (2018) *HR Analytics. Teoría y práctica para una analítica de recursos humanos con impacto*. 1ª ed. España: ESIC.
- Albizu, E. y Landeta, J. (2011) *Dirección Estratégica de los Recursos Humanos*. 2ª ed. País Vasco: Ediciones Pirámide.
- Amat, J. (2016) *Análisis discriminante lineal (LDA) y análisis discriminante cuadrático (QDA)* [en línea] disponible en https://www.cienciadedatos.net/documentos/28_linear_discriminant_analysis_lda_y_quadratic_discriminant_analysis_qda [consulta: 02 abril 2022].
- BBVA (2019) 'Machine learning': ¿qué es y cómo funciona? [en línea] disponible en <https://www.bbva.com/es/machine-learning-que-es-y-como-funciona/> [consulta: 06 marzo 2022].
- Bassi, L. (2011) Raging Debates in HR Analytics. *McBassi & Company*, 34, 15-16.
- Bholowalia, P. y Kumar, A. (2014) EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN. *International Journal of Computer Applications*, Volumen 105 (9), 17-24.
- Bonache, J.; Cabrera, A. (2010) *Dirección de personas. Evidencias y perspectivas para el siglo XXI*. 3ª ed. España: Prentice Hall.
- Cascio, W. F. (2019) *Investing in people*. 3rd ed. United States of America: Society for Human Resource Management.
- Cascio, W. F. y Boudreau, J. (2010) *Investing in People: Financial Impact of Human Resource Initiatives*. 2ª ed. Hoboken: Pearson FT Press.
- Colegio Universitario de Estudios Financieros (2021) *Proceso de admisión y reserva de plaza de Postgrado* [en línea] disponible en <https://ces.cunef.edu/admision-becas-y-ayudas-al-estudio/admision-postgrado/> [consulta: 19 marzo 2022].
- Devenport, T. H. (2018) *DELTA Plus Model & Five Stages of Analytics Maturity: A Primer*. Portland: International Institute for Analytics.
- Devenport, T. H.; Harris, J. (2017) *Competing on Analytics: Updated, with a New Introduction: The New Science of Winning*. 1a ed. United States of America: Harvard Business Review Press.
- Devenport, T. H.; Harris, J.; Morison, R. (2010) *Analytics at Work: Smarter Decisions, Better Results*. 1a ed. United States of America: Harvard Business Review Press.
- Dolan, S. L.; Cabrera, R.; López, A. (2014) *La gestión de personas y del talento*. 1a ed. España: McGraw-Hill
- Fitz-Enz, J. (2010) *The New HR Analytics*. United States of America: Amacon.

- Huebner, R. (2021) Human Resources Data Set [Kaggle] 19 Abril. Disponible en <https://www.kaggle.com/datasets/rhuebner/human-resources-data-set> [consulta: 19 marzo 2022].
- IBM (2009) *Getting Smart About Your Workforce: Why Analytics Matter*. Canadá: IBM GLOBAL BUSINESS SERVICES.
- Iberdrola (n.d.) *Descubre los principales beneficios del 'Machine Learning'* [en línea] disponible en <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico> [consulta: 06 marzo 2022].
- INCIPY (2021) *Transformación digital en RRHH. 6º Estudio*. [en línea] disponible en https://f.hubspotusercontent00.net/hubfs/410258/INCIPY_6%C2%BA%20ESTUDIO_Transformacion_Digital_RRHH%202021.pdf [consulta: 19 marzo 2022].
- L'Heureux, A.; Grolinger, K.; Elyamany, H.; Capretz, M. (2017) Machine Learning With Big Data: Challenges and Approaches. *IEEE ACCESS*, 5, 7776-7797.
- Luna, J. (2018) *Tipos de aprendizaje automático* [en línea] disponible en <https://medium.com/soldai/tipos-de-aprendizaje-autom%C3%A1tico-6413e3c615e2> [consulta: 06 marzo 2022].
- Martínez, A. y Kak, A. (2001) PCA versus LDA. *IEEE transactions on pattern analysis and machine intelligence*, 23 (2), 228-233.
- Randstad (2016) *Funciones y responsabilidades del departamento de rrhh* [en línea] disponible en <https://www.randstad.es/tendencias360/como-se-organiza-un-departamento-de-recursos-humanos/> [consulta: 04 marzo 2022].
- Roberts, B. (2009) *HR Analytics: Gaining Insights for the Upturn*. BusinessWeek Research Services. Silicon Valley: The McGraw-Hill Companies Inc.
- Sancho, F. (2015) Clustering por K-medias [en línea] disponible en <http://www.cs.us.es/~fsancho/?e=43> [consulta 02 abril 2022].
- Softtek (2021) Reducción de dimensionalidad en el Machine Learning [en línea] disponible en <https://softtek.eu/tech-magazine/artificial-intelligence/la-reduccion-de-dimensionalidad-en-el-machine-learning/#:~:text=Por%20tanto%2C%20la%20t%C3%A9cnica%20de, en%20similar%20en%20ambos%20casos> [consulta: 02 abril 2022].

ÍNDICE DE ILUSTRACIONES

Lista de ilustraciones

Ilustración 1: Áreas de actuación de la FRRHH.	6
<i>Ilustración 2: Matriz de Odiorne contrapuesta con la propuesta por BCG.</i>	<i>7</i>
Ilustración 3: Matriz de las 9 cajas.	9
Ilustración 4: Beneficios potenciales de la analítica de personal.....	11
Ilustración 5: Influencia del uso de analítica en la toma de decisiones.	12
Ilustración 6: Diferentes usos y estados de aplicación de HRA.....	13
Ilustración 7: Modelo LAMP.	14
Ilustración 8: Modelo DELTA PLUS.	15
Ilustración 9: Proceso de HRA basado en el Machine Learning.	16
Ilustración 10: Visualización LDA y PCA.....	17
Ilustración 11: Modelo de aprendizaje por refuerzo (Luna, 2018)	18
Ilustración 12: Modelo de aprendizaje supervisado (Luna, 2018)	19
Ilustración 13: Modelo de aprendizaje no supervisado (Luna, 2018).....	19
Ilustración 14: Variables demográficas y situacionales.....	22
Ilustración 15: Compromiso, satisfacción, retribución y formación de los empleados.	23
Ilustración 16: Personalidad y su relación con el puesto de trabajo.	24
Ilustración 17: Rendimiento en el puesto de trabajo; potencial y desempeño.	25

ÍNDICE DE GRÁFICOS

Lista de gráficos

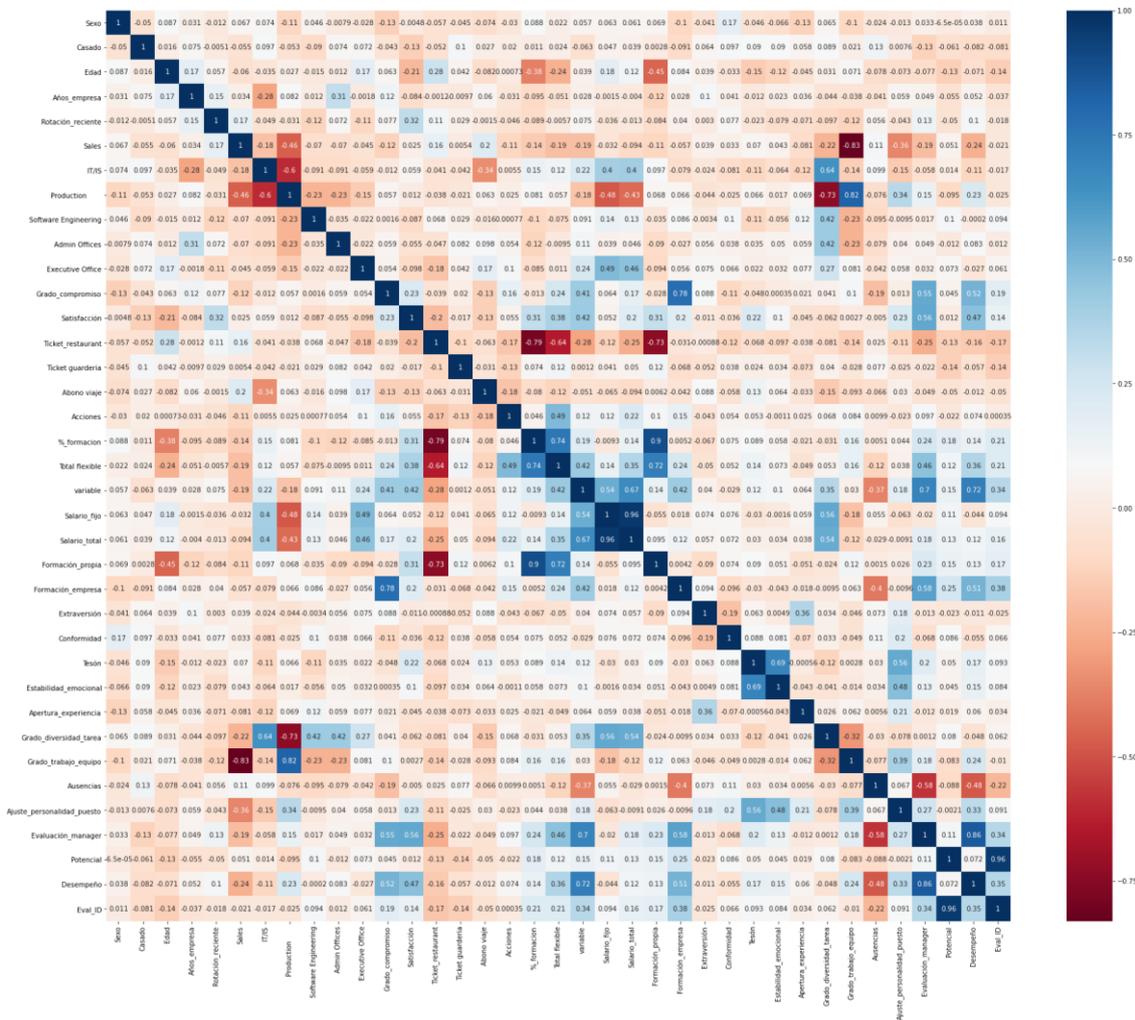
Gráfico 1: Matriz de las 9 cajas de los datos simulados.....	26
Gráfico 2: Distribución del salario total dentro de la empresa.....	27
Gráfico 3: Porcentaje de la formación externa que paga la empresa a cada empleado.....	27
Gráfico 4: Grado de compromiso del empleado con la empresa.....	28
Gráfico 5: Puntuación del 1 al 5 de la satisfacción de los empleado.....	28
Gráfico 6: Rotación reciente de los empleados.....	29
Gráfico 7: Grafico tridimensional de agrupación de los individuos.....	32
Gráfico 8: Método del Codo, para obtención del número de clúster óptimo.....	32
Gráfico 9: Matriz de las 9 cajas de cada clúster generado.....	33
Gráfico 10: Distribución salarios totales por cada clúster generado.....	34
Gráfico 11: Distribución satisfacción por clúster generado.....	35
Gráfico 12: Distribución de los porcentajes pagados por la empresa en formación externa.....	36
Gráfico 13: Distribución de la rotación interna de los clústers generados.....	37

ÍNDICE DE ANEXO

Lista de gráficos dentro del Anexo

Anexo 1: Gráfico de correlaciones.	46
Anexo 2: Gráfico proporción de hombres y mujeres dentro de la muestra.	47
Anexo 3: Gráfico distribución de la edad dentro de la muestra.	47
Anexo 4: Gráfico distribución años dentro de la empresa.....	47
Anexo 5: Gráfico distribución de empleados por departamento.	47
Anexo 6: Gráfico ajuste personalidad-puesto de los empleados.....	47
Anexo 7: Gráfico distribución del número de ausencias de los empleados.	47
Anexo 8: Gráfico distribución de la retribución variable.	47
Anexo 9: Gráfico distribución de la retribución flexible.....	47
Anexo 10: Gráfico distribución de la evaluación de los managers a los empleados.	47
Anexo 11: Gráfico distribución de la edad dentro de los clústers.	47
Anexo 12: Gráfico distribución de años en la empresa por cada clúster.....	47
Anexo 13: Gráfico distribución de los departamentos en cada clúster.	47
Anexo 14: Gráfico ajuste personalidad-puesto en cada clúster.	47
Anexo 15: Gráfico ausencias de los empleados en cada clúster.....	47
Anexo 16: Gráfico retribución variable en cada clúster.....	47
Anexo 17: Gráfico retribución flexible en cada clúster.....	47
Anexo 18: Gráfico distribución de la evaluación de los managers a los empleados en cada clúster.....	47

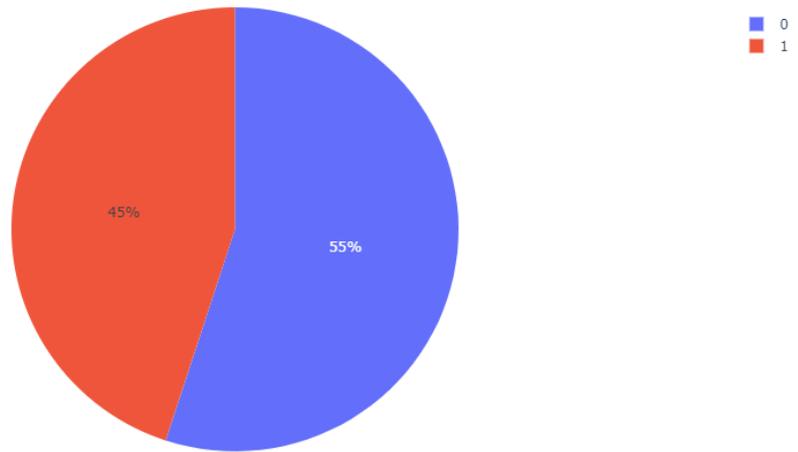
ANEXO



Anexo 1: Gráfico de correlaciones.

Fuente: Elaboración propia.

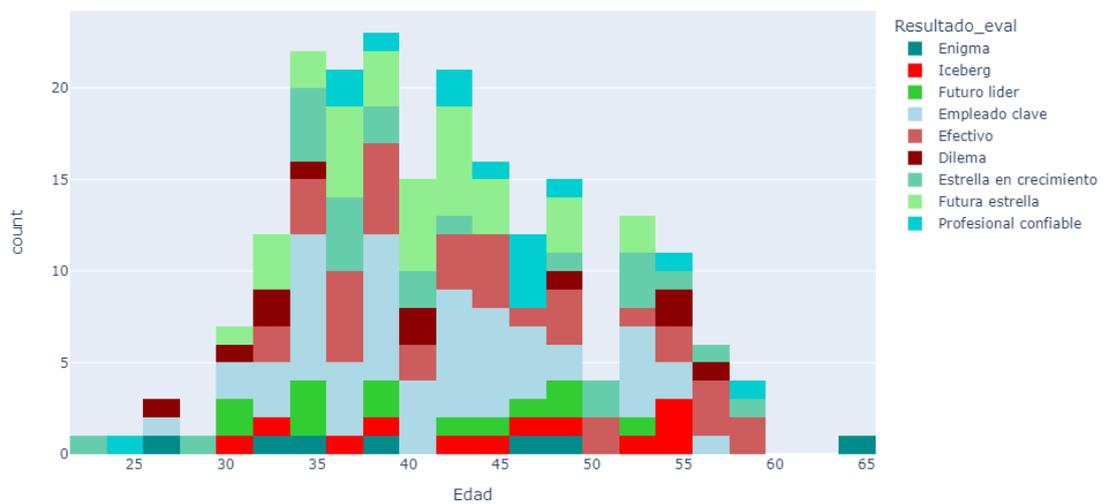
Distribución sexo empleados: Mujer (0); Hombre (1)



Anexo 2: Gráfico proporción de hombres y mujeres dentro de la muestra.

Fuente: Elaboración propia.

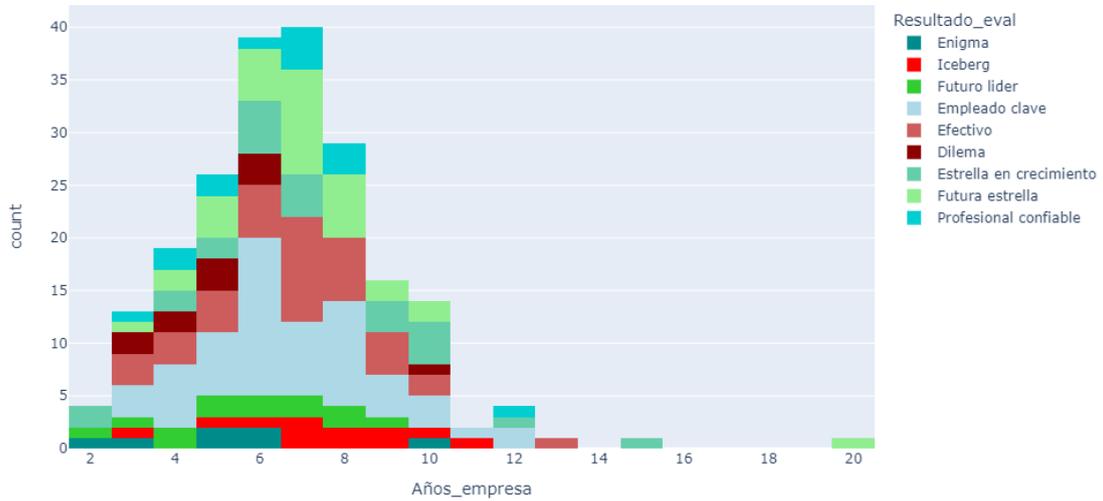
Distribución de la edad de los empleados:



Anexo 3: Gráfico distribución de la edad dentro de la muestra.

Fuente: Elaboración propia.

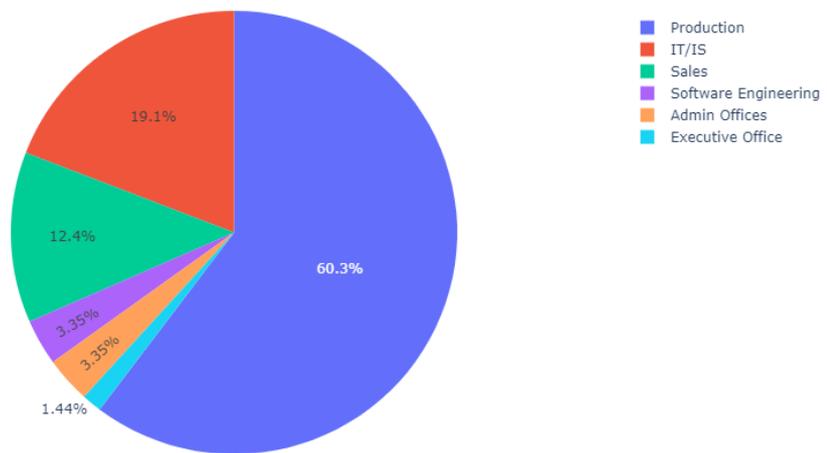
Distribución de los años dentro de la empresa de los empleados:



Anexo 4: Gráfico distribución años dentro de la empresa.

Fuente: Elaboración propia.

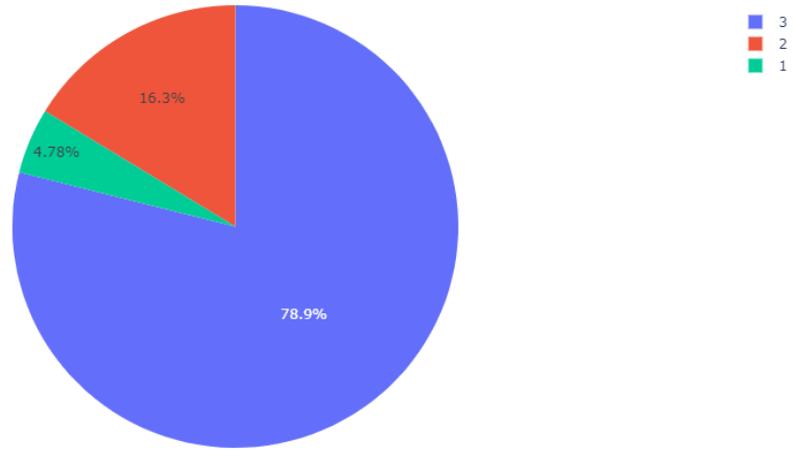
Distribución empleados por departamento:



Anexo 5: Gráfico distribución de empleados por departamento.

Fuente: Elaboración propia.

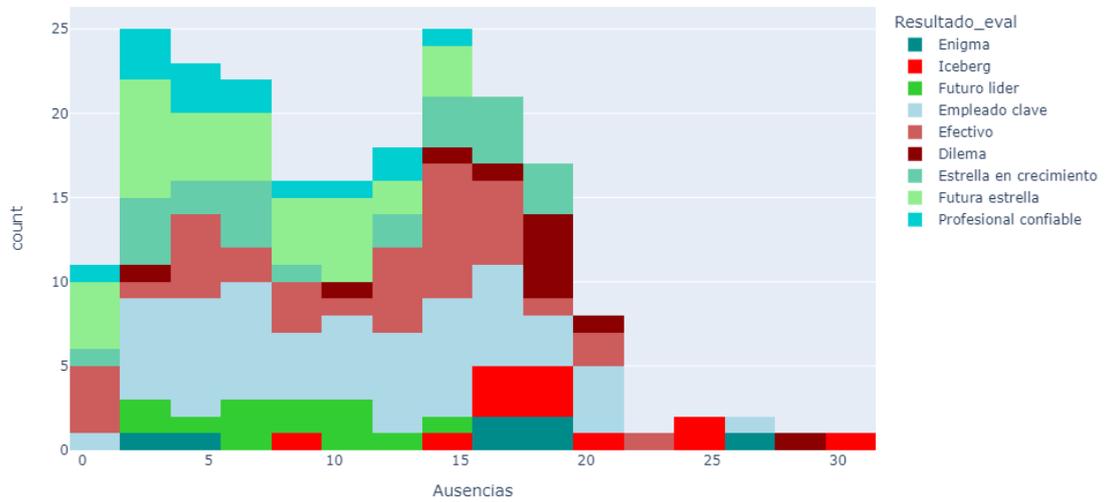
Ajuste personalidad-puesto de los empleados: Excelente (3), Buena (2), Mala (1)



Anexo 6: Gráfico ajuste personalidad-puesto de los empleados.

Fuente: Elaboración propia.

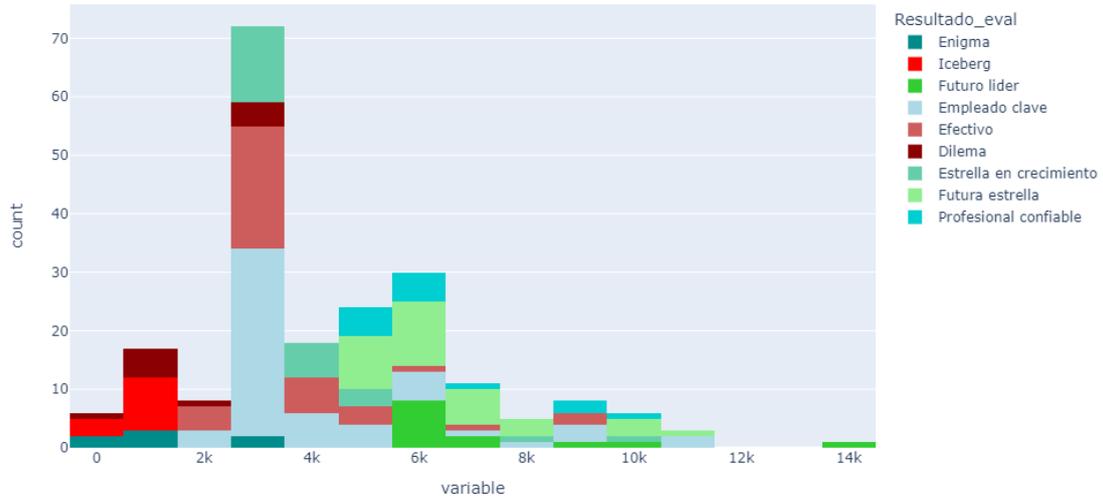
Distribución de las ausencias de los empleados:



Anexo 7: Gráfico distribución del número de ausencias de los empleados.

Fuente: Elaboración propia.

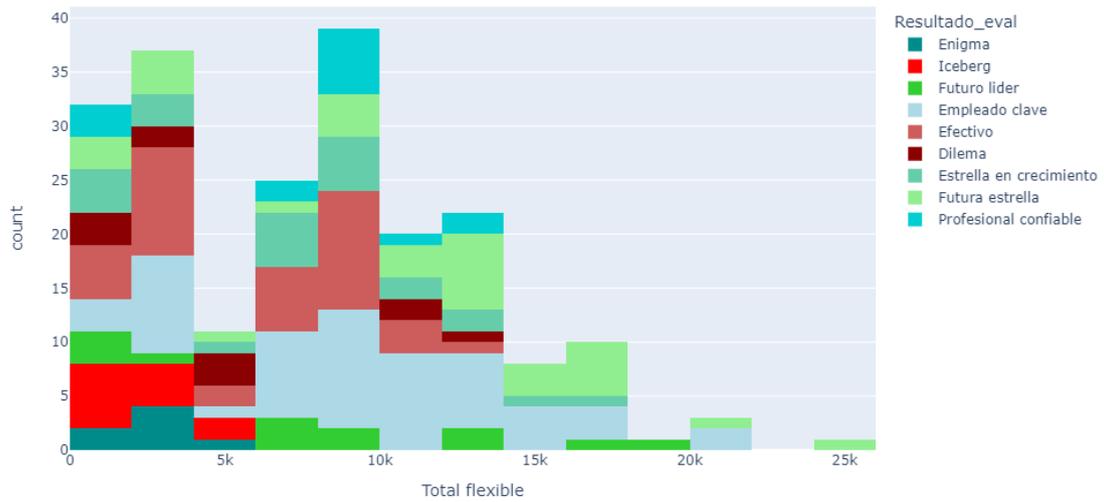
Distribución de la retribución variable de los empleados:



Anexo 8: Gráfico distribución de la retribución variable.

Fuente: Elaboración propia.

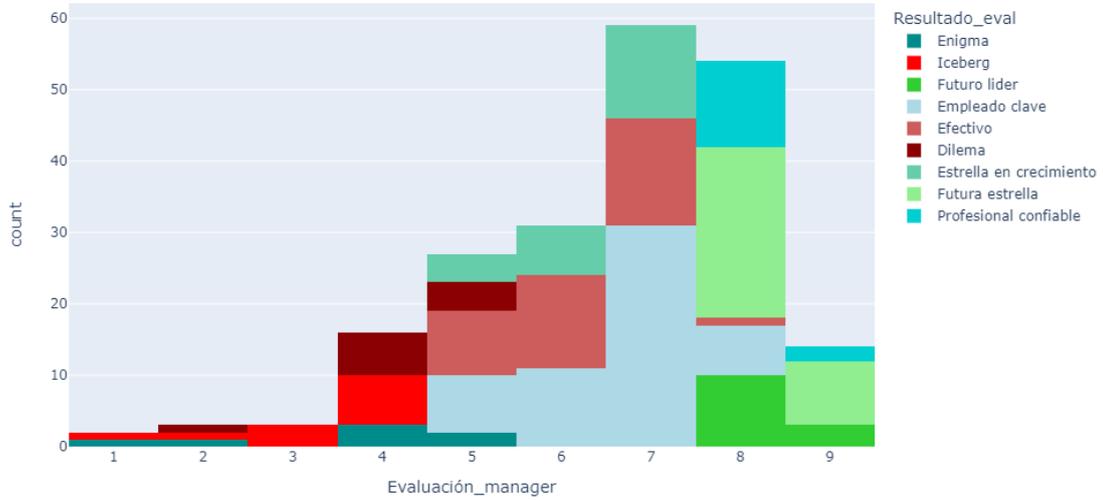
Distribución de la retribución flexible de los empleados:



Anexo 9: Gráfico distribución de la retribución flexible.

Fuente: Elaboración propia.

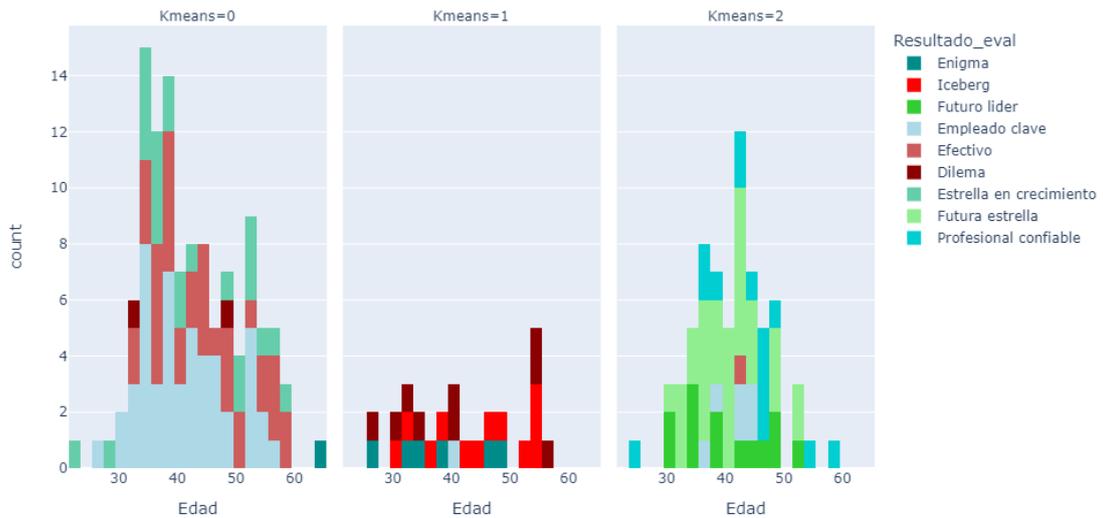
Distribución de la evaluación de los managers a los empleados:



Anexo 10: Gráfico distribución de la evaluación de los managers a los empleados.

Fuente: Elaboración propia.

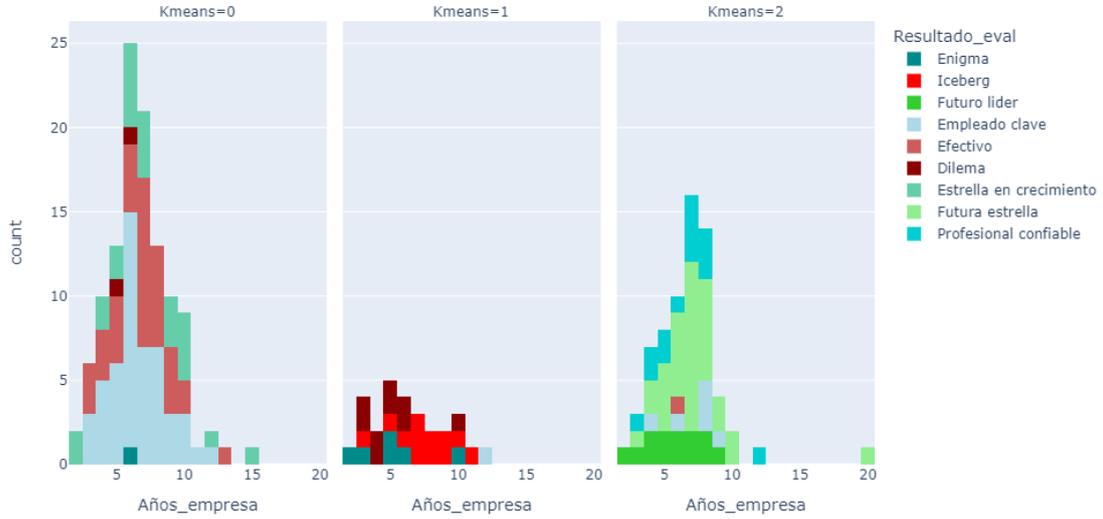
Distribución de la edad función de cada grupo:



Anexo 11: Gráfico distribución de la edad dentro de los clústers.

Fuente: Elaboración propia.

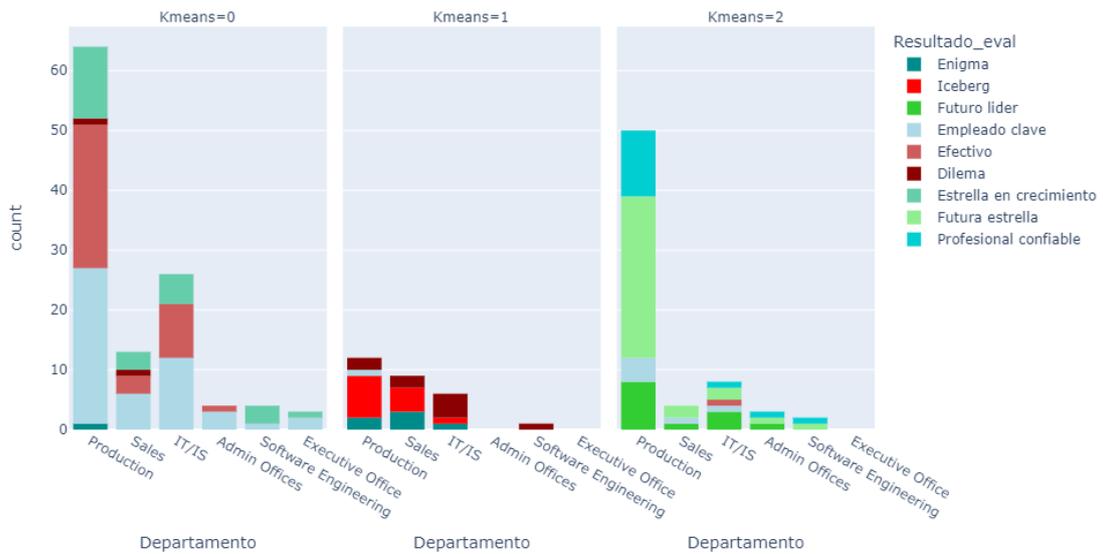
Distribución años dentro de la empresa en función del grupo al que pertenece:



Anexo 12: Gráfico distribución de años en la empresa por cada clúster.

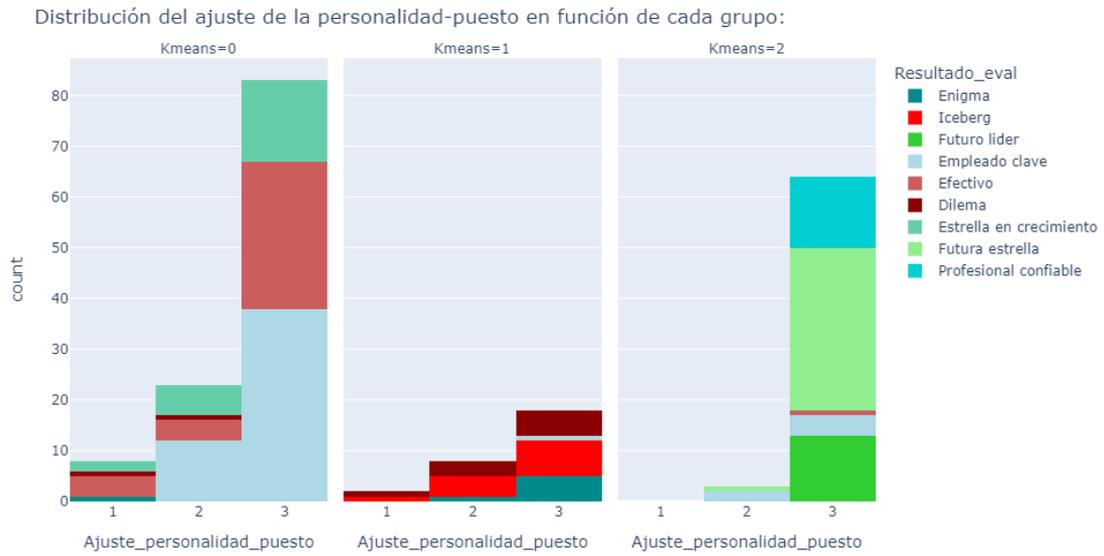
Fuente: Elaboración propia.

Distribución de los departamentos en función de cada grupo:



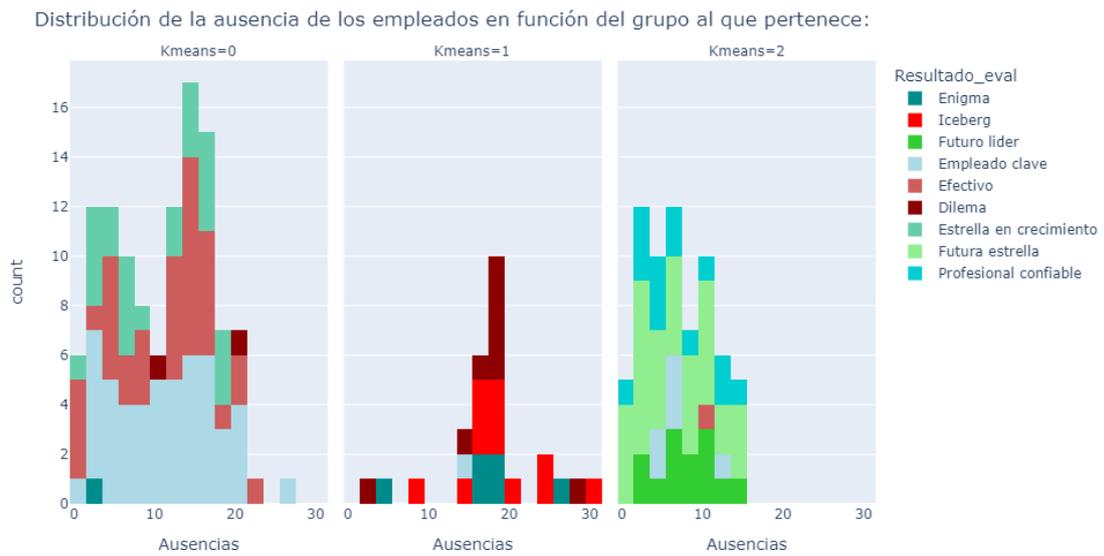
Anexo 13: Gráfico distribución de los departamentos en cada clúster.

Fuente: Elaboración propia.



Anexo 14: Gráfico ajuste personalidad-puesto en cada clúster.

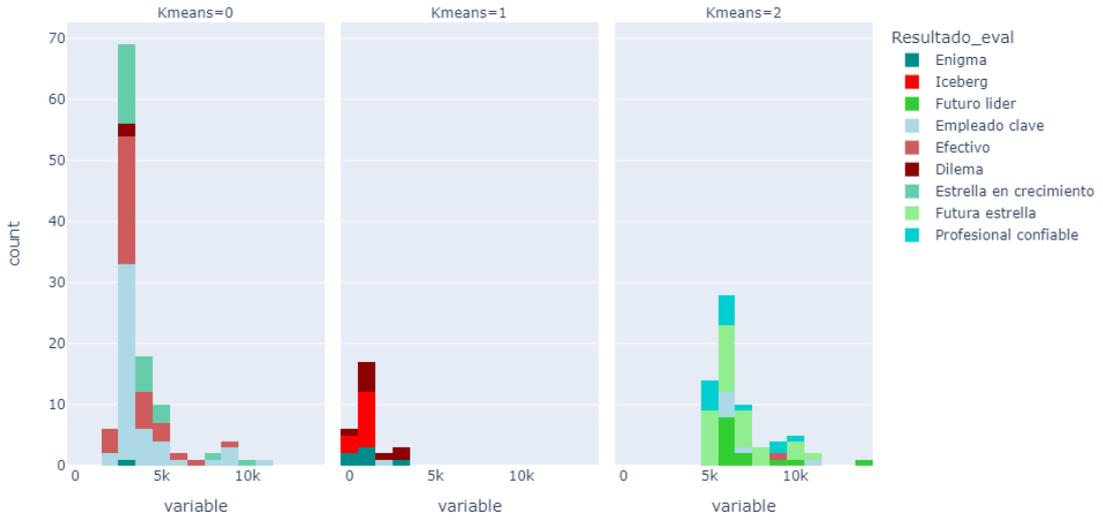
Fuente: Elaboración propia.



Anexo 15: Gráfico ausencias de los empleados en cada clúster.

Fuente: Elaboración propia.

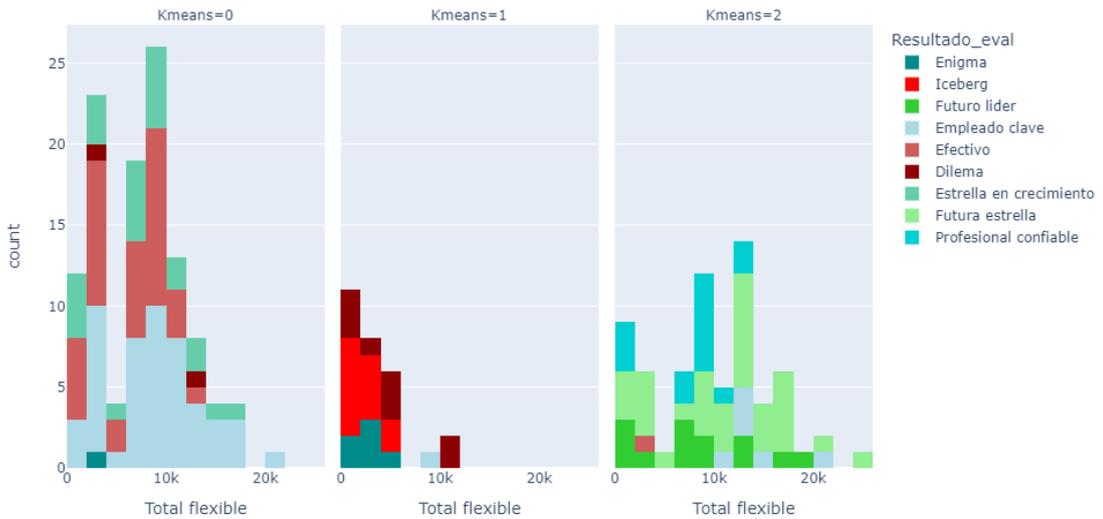
Distribución de la retribución variable en función del grupo:



Anexo 16: Gráfico retribución variable en cada clúster.

Fuente: Elaboración propia.

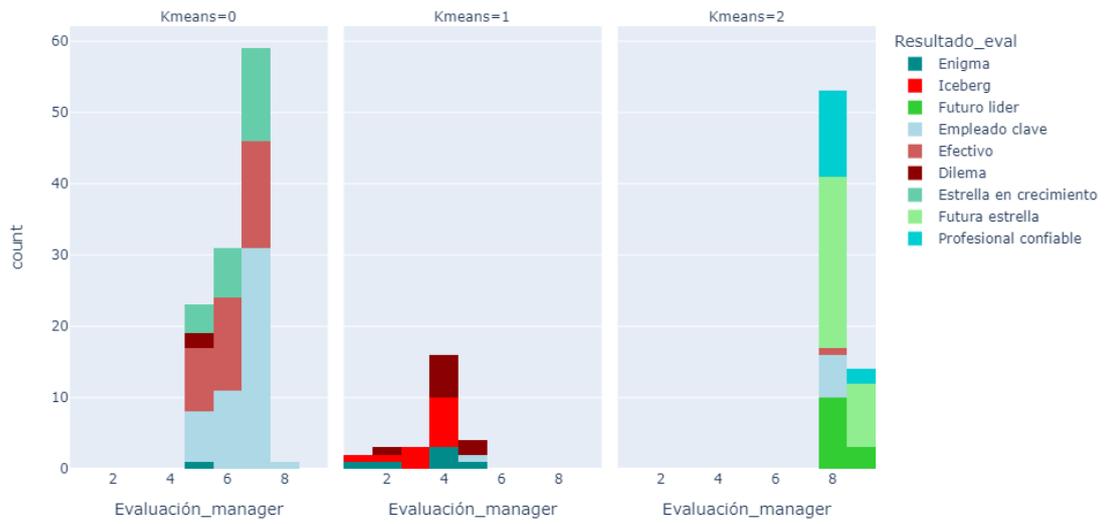
Distribución de la retribución flexible total en función del grupo:



Anexo 17: Gráfico retribución flexible en cada clúster.

Fuente: Elaboración propia.

Distribución de la evaluación otorgada por los managers:



Anexo 18: Gráfico distribución de la evaluación de los managers a los empleados en cada clúster.

Fuente: Elaboración propia.