



Revista de Casos de Estudio en HR Analytics

Journal of HR Analytics *Case Studies*

04. Explicando la rotación voluntaria mediante modelos analíticos: el caso de una empresa de seguros española.

Explaining voluntary turnover through analytical models: the case of a Spanish insurance company.

La Revista de Casos de Estudio en HR Analytics

Journal of HR Analytics Case Studies

La Revista de Casos de Estudio en HR Analytics nace con la misión de facilitar el intercambio de conocimiento especializado entre profesionales y académicos en el ámbito de la **analítica de Recursos Humanos**, con el objetivo de mejorar la efectividad de las organizaciones. La entidad responsable de esta revista es la **Asociación para el Desarrollo de la Ingeniería del Conocimiento** (ADIC), siendo esta publicación on-line editada por el **Instituto de Ingeniería del Conocimiento** (IIC) con una periodicidad de un número anual.



Objetivo

La revista tiene como **objetivo** principal ser un vehículo para la reflexión y la difusión de las **buenas prácticas, últimos avances y líneas de investigación** en el ámbito de la analítica aplicada para la toma de decisiones sobre la gestión del capital humano en las organizaciones.

La revista tiene un **carácter científico** y una **vocación divulgativa**, por ello propone artículos fundamentalmente de **carácter aplicado**. Con ellos se pretende que los profesionales de las organizaciones accedan a un conocimiento relevante acerca de cómo otras organizaciones desarrollan HRA. Y, también, acercar a los académicos el conocimiento respecto de cómo se desarrolla HRA en la práctica.



Alcance

El **enfoque de la Revista**, que pretende ser **multidisciplinar**, da cabida (entre otros) a manuscritos que: reflejen **casos prácticos** de aplicación del HRA en las organizaciones; que analicen, comparen y relacionen la utilidad de diferentes **técnicas y/o herramientas** para el abordaje de diferentes objetivos analíticos; que planteen y valoren la efectividad de diferentes **metodologías de trabajo** para el desarrollo de proyectos HRA; que ayuden a entender el **mapa de ruta** por el que transitar desde los niveles básicos del HRA hasta los niveles de excelencia; y que en general ayuden a entender cómo **mejorar la efectividad organizacional** a partir de la analítica de datos referidos a la fuerza de trabajo.

La revista está editada por el Instituto de Ingeniería del Conocimiento y tiene los siguientes órganos de gobernanza.

Editor

David Aguado.

Instituto de Ingeniería del Conocimiento.

Editores Asociados

Jesús de Miguel.

Centro de Investigación para la Efectividad Organizacional,
Universidad Autónoma de Madrid.

Antonio Delgado.

Universidad Autónoma de Madrid.

María Jesús Belizón.

Universidad Pontificia de Comillas.

Beatriz Lucía.

Instituto de Ingeniería del Conocimiento.

Delia Majarín.

Accenture.

Sergio Raja.

Zurich Seguros.

Diseño y Maquetación

Andrés Muñoz Bachiller.

Instituto de Ingeniería del Conocimiento.

Comité Editorial

Magdalena Nogueira.

Universidad Autónoma de Madrid.

Francisco Abad.

Universidad Autónoma de Madrid.

Carmen García.

Universidad Autónoma de Madrid.

José Manuel de Haro.

Universidad de Alicante.

William Ferrando Durán.

Universidad Javeriana.

Carolina Zúñiga.

Universidad Politécnica Salesiana del Ecuador.

José Carlos Andrés.

Viewnext.

Eduardo Páez.

Data4all.

Pablo Haya.

Instituto de Ingeniería del Conocimiento.

Álvaro Barbero.

Instituto de Ingeniería del Conocimiento.

Sonia Rodríguez.

Instituto de Ingeniería del Conocimiento.

Maite Sáez.

Observatorio de Recursos Humanos y
Relaciones Laborales.

.04



.04

Explicando la rotación voluntaria mediante modelos analíticos: el caso de una empresa de seguros española

Explaining voluntary turnover through analytical models: the case of a Spanish insurance company.

Miguel Angel Lorenzo Bilbao

Correo electrónico:
malorenzob@gmail.com

LinkedIn:
<https://www.linkedin.com/in/malorenzob>

Received: 7 june 2023

Received in revised form: 10 october 2023

Accepted: 17 october 2023

Available on-line: 29 november 2023



Resumen

Palabras clave

Rotación voluntaria, retención de talento, modelos predictivos, regresión logística, árbol de decisión.

La rotación voluntaria de empleados es un desafío que enfrentan las organizaciones en la actualidad. Comprender las causas y los factores que influyen en esta rotación se ha convertido en una prioridad para retener talento y garantizar el éxito a largo plazo de las empresas. Este trabajo de investigación fue enfocado en estudiar las bajas ocurridas en una empresa de seguros española durante los años 2021 y 2022. El objetivo principal fue analizar el impacto de los egresos voluntarios e identificar los factores que influyen en la tasa de rotación. Para esto, se utilizaron dos técnicas estadísticas distintas para generar modelos predictivos y evaluarlos en cuanto a su efectividad en la predicción de la rotación voluntaria.

Mediante los resultados obtenidos muestran que, en comparación con el contexto, la rotación voluntaria en la empresa no es alarmante, aunque ha experimentado un aumento. Por medio de los modelos predictivos, se encontró que los dos perfiles con mayor riesgo de egreso voluntario son: los empleados con menos de 2 años de antigüedad, y los empleados entre 2 y 8 años de antigüedad y entre 31 y 38 años de edad, que no percibieron recibido un aumento salarial y no contaban con equidad interna.

Abstract

Keywords

Voluntary employee turnover, talent retention, predictive models, logistic regression, decision tree.

Voluntary employee turnover is a challenge currently faced by organizations. Understanding the causes and influencing factors of this turnover has become a priority for talent retention and ensuring the long-term success of companies. This research work was focused on studying the departures that occurred within a Spanish insurance company during the years 2021 and 2022. The main objective was to analyze the impact of voluntary departures and identify the factors influencing the turnover rate. To achieve this, two distinct statistical techniques were employed to generate predictive models and assess their effectiveness in predicting voluntary turnover.

The results obtained indicate that, in comparison to the broader context, voluntary turnover within the company is not alarming, although it has experienced an increase. Through the predictive models, it was found that the two profiles at the highest risk of voluntary departure are employees with less than 2 years of tenure and employees with tenure between 2 and 8 years, aged between 31 and 38, who did not receive a salary increase and did not have internal equity.

1. Introducción

En la actualidad, para las organizaciones, resulta estratégico hacer uso de los datos para tomar decisiones relacionadas con la gestión para poder reducir costos, aumentar productividad, mejorar la atención al cliente, etc. (ORACLE, 2022). Sin embargo, las empresas tienen que lidiar con cada vez más datos de distintas fuentes, formatos y con mayor velocidad; para poder generar valor a partir de estos datos, resulta imperativo tener las capacidades tecnológicas (tanto infraestructura, como talento humano) necesarias para poder manejar, catalogar y procesar la información, es decir, las capacidades para manejar Big Data (ORACLE, 2022).

HR analytics entra dentro de este dominio y puede ser definida como el uso de la estadística descriptiva, inferencial y otros métodos computacionales para construir y aplicar sistemáticamente modelos predictivos basados en los datos de recursos humanos, con el fin de informar de factores clave que conducen a la optimización de la fuerza laboral (Bossi, et. al., 2022). Según Ben-Gal (cp. Bossi, et. al., 2022), el uso de HR analytics está relacionado con un mejor desempeño económico (comparado con casos de estudio y técnicos) ya que permite identificar oportunidades y riesgos antes de que estos sean claros para los managers. Un tipo de proyecto de HR analytics de interés tanto para recursos humanos como para los negocios, es la predicción de la rotación.

En el contexto actual, tras la pandemia del 2020, tanto la economía global como el empleo se han visto afectados, aumentando la percepción de estrés en el trabajo, así como problemas de insatisfacción y burnout, aumentando a su vez, el índice de rotación voluntaria (Raza, Almutairi, Younas y Fareed, 2022). Según la Oficina de Estadísticas Laborales, en el 2022, en EEUU se reportó que 4.5 millones de trabajadores dejaron su empleo. Esta rotación fue particularmente impactante en las nuevas contrataciones, en donde se reportó que un tercio de estas dejaron su empleo luego de los primeros 6 meses (cp. Raza, Almutairi, Younas y Fareed, 2022). Chatzoudes (2022), agrega que otro factor de riesgo asociado a la rotación en el contexto actual es el grupo generacional, dado a que los Millennials componen la mayor proporción de la plantilla de muchas empresas y tienen una mayor tendencia a rotar que las generaciones anteriores.

Dentro de las empresas del sector asegurador, un estudio realizado en el mercado laboral de la industria aseguradora, solo el 5% de las empresas planean reducir activamente su plantilla, mientras que el 68% de las empresas proyectan aumentar su fuerza laboral, particularmente los perfiles de las áreas de tecnología y siniestros. Estas empresas señalan que el principal racional de incrementar su plantilla, es aumentar el volumen de negocio y hacer cara a las demandas del mercado (The Jacobson Group, c.p. Moorcraft, 2022). Sin embargo, las estadísticas del sector señalan un crecimiento de plantilla del -0,68%, es decir, se observa un estancamiento (ICEA, 2022), por lo que se podría señalar la dificultad en la contratación de nuevos perfiles y la rotación voluntaria como los principales cuellos de botella para conseguir los objetivos estratégicos de planificación de fuerza laboral. En cuanto a la rotación voluntaria reportada en España para el sector asegurador, ha sido del 6,99% (ICEA, 2022); lo cual se encuentra 1,67% por encima del 5,32% reportado para su macro sector (IE University, 2022).

En línea con lo anterior, el objeto de este trabajo es presentar los resultados obtenidos al analizar la rotación en una empresa de seguros y encontrar variables que permitan definir perfiles en riesgo, de manera que la empresa pueda plantearse medidas de retención enfocadas en estos perfiles para gestionar el riesgo.



2. La rotación desde el punto de vista de la Analítica de Recursos Humanos

2.1. ¿Qué es la rotación?

2.1.1. La rotación desde la perspectiva del empleado

Para poder realmente entender este problema, sus ramificaciones y como abordarlo, resulta importante definir la rotación laboral. Desde la perspectiva de la empresa, la rotación es el fin de la relación laboral entre la organización y el empleado (Raza, et. al., 2022). En la bibliografía, se suele diferenciar entre rotación involuntaria y rotación voluntaria. La rotación involuntaria se da cuando la organización es la que toma la decisión de finalizar la relación laboral, mientras que la rotación involuntaria se da cuando la decisión es tomada por el empleado (Raza, et. al., 2022).

2.1.2. La rotación desde la perspectiva de la empresa

Por parte de la empresa, según Allen, Bryant y Vardaman (2010) no solo basta con diferenciar entre rotación voluntaria e involuntaria, sino que es importante diferenciar la rotación voluntaria en función de su impacto como rotación funcional y rotación disfuncional.

La rotación funcional la definen como aquella que no tiene gran impacto para la organización, resulta cuando no se trata de empleados de alto desempeño, son de fácil reemplazo o cuando su reemplazo puede ser una oportunidad para obtener conocimiento/mejores prácticas, puede llevar a más diversidad, dan la oportunidad de promover a otro empleado o la salida puede resultar un ahorro.

La rotación disfuncional es definida como una salida voluntaria que puede ser dañina para la organización, bien sea porque hay pocos perfiles en el mercado, porque el empleado contaba con un buen desempeño o cuando la salida impacta objetivos organizacionales.

2.1.3. Acercamiento metodológico para el estudio de la rotación

Según Ulhas (2019), se recomienda que, para estudiar la rotación y retención, se consideren tres pasos:

1. identificar el costo de la rotación; 2. entender porque se van los empleados; y finalmente 3. implementar estrategias de retención.

1. Identificar el costo de la rotación: Desde la perspectiva de la empresa, el costo de las salidas se puede entender como:

- Interrupción en el servicio al cliente (Ulhas, 2019).
- Costos por separación: tiempo del personal de RRHH y managers en off boarding, empleados temporales, pérdida de diversidad, pérdida de cliente, pérdida de conocimiento organizacional y interrupciones en el trabajo en equipo (Allen, Bryant y Vardaman, 2010).
- Costos por reemplazo: implican pérdida en efectividad y productividad por el tiempo del personal RRHH y el manager en la contratación y on boarding, costos de reclutamiento (postulaciones, outsourcing, bonus de referidos, evaluación) y costos en socialización (Allen, Bryant y Vardaman, 2010).
- Rotación que genera mayor rotación o efecto bola de nieve (Ulhas, 2019).
- Costo estimable a partir del salario bruto anual (SBA), en donde en promedio sería el 25% del SBA, para empleados con conocimiento técnico son 6 meses y para managers son 18 meses (Ulhas, 2019).

2. Identificar los motivos de salida: Son circunstancias que suelen tener en común las personas que se van, estas incluyen:

- El trabajo no cumple con las expectativas que tenía el candidato (proyecto).
- No hay coherencia o match entre la personalidad y los valores organizacionales (misión y valores).
- Falta de oportunidades de carrera.
- Falta de apreciación.
- Falta de soporte y confianza con los pares, superiores y supervisores.
- Estrés y falta de balance vida-trabajo (sobrecarga laboral).
- Compensación.

3. Estrategias de retención: La retención, desde la perspectiva del trabajador, se entiende como una actitud positiva hacia quedarse en la empresa en un periodo de tiempo específico (Bossi, et. al. 2022).

La retención, desde la perspectiva del empleador, es el proceso en el cual se usan recursos para motivar a las personas a quedarse en la organización por un periodo de tiempo (Ulhas, 2019), es decir, son la palanca para intervenir en los factores “stay” dentro del juicio de equilibrio del empleado.

Las tres R de la retención son respetar, reconocer y recompensar, donde las palancas clave para obtenerlas son (Ulhas, 2019):

- Clima: debe ser motivador, donde las personas sientan que pueden contribuir.
- Crecimiento: oportunidades de carrera que satisfagan a la persona.
- Compensación: incluye salario, bonus, variables y otros beneficios.
- Relaciones: con los managers y otros trabajadores.
- Flexibilidad: todo lo que permita tener balance vida trabajo.

Para abarcar estas R, Allen, Bryant y Vardaman (2010) discuten distintas estrategias de retención que cuentan con evidencia empírica y las segmentan por puntos clave de recursos humanos:

- Reclutamiento: Utilizar las ofertas de trabajo realistas. Según Graef (2020), son una estrategia de comunicación donde se muestran los aspectos positivos y negativos del trabajo y la organización, detallando recursos para trabajar, volumen de trabajo, beneficios, potencial de aprendizaje, horario, cantidad de interacción con el cliente, etc.), uso de un programa de referencia de empleados.
- Selección: Medir el fit organizacional durante el proceso de selección y usar weighted application blanks (escala ponderada que permite estandarizar el proceso de selección dándole un peso específico a la experiencia, cualificaciones y características personales).
- Socialización: Hacer uso de personas con experiencia en la organización para que sean mentores, dar feedback temprano a las nuevas contrataciones durante su proceso de adaptación, hacer actividades de orientación agrupando a las nuevas contrataciones para que puedan socializar.
- Aprendizaje y desarrollo: dar una oferta de desarrollo

atadas a la antigüedad. Estas deben ser específicas al trabajo y enlazadas con oportunidades de crecimiento en la organización. De esta manera hay un incentivo claro en mantenerse en la empresa.

- Compensación y recompensa: Personalizar la recompensa en función a necesidades y preferencias. Promover la justicia y equidad de la remuneración y tener un presupuesto de fidelización (hacer explícita la relación entre la retención y la recompensa).
- Supervisión: Entrenar a los supervisores en cómo ser líderes y como desarrollar relaciones efectivas.
- Engagement: Asegurar que el trabajo ofrezca un “significado”, sentido de autonomía y variedad además de objetivos retadores. Apalancar las relaciones entre empleados (construyendo puentes, aumentando el número de relaciones), dar orientación para comunicar como las tareas contribuyen a la misión organizativa, ofrecer oportunidades de desarrollo continuo, sistemas de remuneración por competencia. Ofrecer feedback y reconocimiento por todos los tipos de contribuciones.

A esto, se puede agregar como estrategia, la gestión de la movilidad interna, definida como cambios laterales y/o promociones internas dentro de la organización, se relaciona con la satisfacción del empleado, un mejor desempeño y un menor índice de rotación que las personas contratadas externamente (Bossi, et. a., 2022).



2.2. Técnicas analíticas para explorar la rotación

Dentro de la bibliografía revisada de HR analytics para predecir la rotación, algunas de las técnicas de machine learning utilizadas son:

- **Regresión Logística:** Algoritmo para predecir un resultado binario a partir de una o varias variables. Genera un modelo probabilístico de la ocurrencia de un evento en base a una ecuación de regresión.
- **Árbol de Decisión:** Modelo de aprendizaje supervisado. Sirve para clasificar resultados basados en nodos que se van separando en función a atributos clave.
- **Clasificador de Árboles:** Modelo supervisado que genera predicciones en un esquema de árboles agregados, eligiendo la predicción más común realizada por los árboles.
- **Máquina de Vectores de Soporte:** Modelo supervisado, útil en casos no lineales. Busca el mejor criterio diferenciando entre grupos o clases, ayuda a generar clasificaciones útiles para análisis de regresión (crea un hiperplano con n-dimensiones y las divide entre dos de manera iterativa hasta encontrar el ajuste que minimiza el error y permite discriminar entre dos grupos).
- **Neural Networks:** Algoritmo de aprendizaje automático, se inicia con datos entrenados y la red ajusta sus pesos y genera patrones encontrados en los datos. Constan de tres capas, la capa de entrada (datos de entrada), la capa oculta (procesamiento) y la capa de salida que produce las predicciones o resultados.
- **Análisis de Redes Organizacionales:** Comprende conceptos y métodos de análisis específicos para extraer información valiosa de mapas de organización, estudiando interacciones entre grupo de individuos, donde cada red está formada por nodos (habitualmente personas) interconectados por enlaces que representa relaciones existentes en el mundo real (Haya y García, 2018).

Raza, et. al. (2022) realizaron un estudio utilizando donde utilizaron y compararon varias de estas técnicas en la predicción de la rotación voluntaria. Ellos encontraron que, para los empleados, las variables que tendían a estar relacionadas, y permiten predecir la rotación voluntaria, son la edad, posición (nivel de la posición), los años con el mánager y la antigüedad; con respecto a esta última variable, encontraron que, hasta los tres años de antigüedad, el salario también tiene un efecto sobre la decisión de salida voluntaria.

En cuanto a las distintas técnicas, concluyeron que el mejor modelo para predecir fue el Clasificador de árboles

(extra tres classifier) con una exactitud del 93%, a este le sigue la máquina de vectores de soporte (support vector machine) con un 88% de exactitud, el árbol de decisiones con un 84% y, por último, el modelo de regresión logística con una exactitud del 74%.

Por su parte, Chatzoudes (2022) realizó una investigación con un modelo de ecuaciones estructurales, por medio del cual logró explicar el 57% de la retención a partir de dos niveles explicativos. El primer nivel de relaciones directas con la retención incluye las variables: identificación con la organización (predice el 27%), la satisfacción con el trabajo (predice el 26%) y el compromiso (mediada por las condiciones de trabajo y prácticas de RRHH, predice el 30% de la retención). En segundo nivel, las variables indirectas o mediadoras encontradas son:

- **Condiciones de trabajo:** incluye el clima, relaciones entre empleados, actitud o valoración subjetiva de la organización, niveles organizativos y la calidad de los espacios físicos.
- **Misión y orientación:** Conexión subjetiva entre la dirección y misión de la empresa y valores personales.
- **Prácticas de RRHH:** formación, remuneración, oportunidades de carrera, feedback de desempeño, empoderamiento, y estilos de liderazgo.

En un estudio aún más reciente, Younis, Ahsan y Chatteur (2023), discuten un modelo usando un análisis de redes organizacionales (ONA), en donde consideraron características de las redes organizativas como posición, cercanía y tipo de red, y discutieron estos indicadores en el contexto de distintos escenarios favorables y no favorables que median entre la relación de indicadores de ONA y la rotación voluntaria. En sus hallazgos, concluyen que:

La posición de la red será clave para entender quien está en riesgo de egreso (en escenarios favorables, entre mayor cantidad de redes menor es la probabilidad de egreso, en escenarios no favorables, entre mayor cantidad de redes mayor es la probabilidad de egreso).

La centralidad y el tipo de red se deben considerar en conjunto, considerando la percepción de equidad salarial y el clima como factores clave. Se espera que en situaciones de estrés (pobre clima laboral), la centralidad en redes de amistad sea un factor de protección, reduciendo la probabilidad de egreso.

Si la red es de consejo, se espera que la rotación sea inferior siempre que exista equidad interna y una percepción de justicia retributiva (especialmente entre las personas que ofrecen consejo/expertos).

Y finalmente, en todas las redes, pero principalmente en la de trabajo, es necesario tener en el radar a las personas con potencial de influencia (muchas relaciones y centralidad) y personas de alto desempeño, ya que estas son las que pueden provocar un efecto de bola de nieve en sus pares (personas con las que trabajan y personas con un rol similar).

3. Objetivo

Este trabajo de investigación fue desarrollado con el objetivo principal de estudiar las bajas ocurridas durante el período 2021 y 2022 en una empresa de seguros española. Para ello, se llevó a cabo un análisis detallado de los registros de empleados de la empresa para identificar (a) el impacto que tienen los egresos voluntarios de personal en la empresa, (b) los factores que influyeron en la tasa de rotación y (c) las causas específicas que motivaron la salida de los empleados.

Además, se estableció como objetivo secundario diseñar de un modelo predictivo de rotación voluntaria comparando dos técnicas estadísticas (regresión logística múltiple y el árbol de decisión) términos de capacidad predictiva, efectividad y aspectos éticos.



4. Marco Metodológico

4.1. Obtención de datos y muestra

La información utilizada para este análisis fueron los datos de empleados disponibles en el sistema, los cuales incluyen:

- Datos identificativos y demográficos: sexo, fecha de nacimiento y antigüedad.
- Datos organizativos: dirección organizativa, rol, estado del empleado (activo o baja) y motivo de baja.
- Datos económicos: salario bruto anual (SBA).
- Datos de formación: número de horas de formación entre el 2021 y 2022.
- Datos de absentismo: número de casos y días de ausencia entre el 2021 y 2022.

Además de estos datos, se incluyó la información obtenida a partir de la encuesta de salida realizada a los egresos voluntarios.

4.2. Procesamiento de datos

Luego de recopilar los datos, para facilitar su análisis, se realizaron cálculos para procesar la información de absentismo y SBA y se generaron agrupaciones para hacer comparativas. Para comprender y contextualizar el absentismo, se calculó el Factor Bradford. Este factor es una puntuación que se otorga a cada empleado en función a patrones de absentismo por ausencias no planificadas. Se calcula con la fórmula:

$$\text{Bradford} = (\text{casos de absentismo})^2 * \text{días de ausencia}$$

Para este análisis, dado a que no todos los empleados estuvieron presentes durante el mismo período; se ajustó al 100% del período tanto el número de casos, como la cantidad de días de ausencia de cada empleado. La escala estándar para interpretar este se muestra en la tabla 1.

Tabla 1. Escala de Interpretación del Factor Bradford

Puntuación	Interpretación	Acción recomendada
<51 puntos	Bajo	Advertencia verbal
≤201 puntos	Medio	Advertencia escrita
≤401 puntos	Alto	Advertencia final
≥601 puntos	Muy alto	Despido

Para el SBA, se calcularon (1) percentiles considerando las agrupaciones por rol; (2) la variación del salario entre los dos años; y (3) compa-ratios considerando el rol y departamento de pertenencia de cada empleado. Los grupos artificiales, comprenden tres agrupaciones realizadas mediante la categorización a los empleados en los siguientes criterios:

1- Grupos por movimiento en plantilla: Los criterios para discriminar están en función al estado en plantilla de cada empleado y el tipo de movimiento:

- Grupo control: Está compuesto por los empleados que no presentan ningún movimiento en los últimos dos años.
- Altas (todas): Incluye todas las nuevas contrataciones ocurridas entre enero 2021 y diciembre 2022.
- Altas (solo): Subgrupo, que incluye solo las altas que permanecen en plantilla al cierre del año 2022.
- Altas-Bajas: Subgrupo que solo incluye las personas que se dieron de alta y posteriormente causaron baja.
- Bajas: Grupo conformado por empleados que formaban parte de la plantilla antes del 2021 y han causado baja. Este a su vez se subdivide por bajas voluntarias e involuntarias.
- Rotación Interna: Incluye a los empleados que se mantienen en la misma organización, pero cambian de empresa dentro del grupo. Se pueden subdividir en rotación positiva y rotación negativa.

2- Grupos por generación: Se generan a partir del año de nacimiento del empleado:

- Generación Z: Año de nacimiento ≥ 1996 .
- Millennial: Año de nacimiento entre 1980 y 1995.
- Generación X: Año de nacimiento entre 1965 y 1979.
- Baby Boomer: Año de nacimiento < 1965 .

3- Grupos por rangos de antigüedad:

- Menor o igual a 1 año.
- Entre 2 y 4 años.
- Entre 5 y 10 años.
- Entre 11 y 25 años.
- Mas de 25 años.

4.3. Metodología de análisis estadístico

Para la metodología de este análisis, se realizó análisis exploratorio de datos con el fin de describir la muestra en general y los grupos basados en los movimientos en planilla.

De acuerdo con lo propuesto por Ulhas (2019), el siguiente paso fue identificar el costo de la rotación, utilizando la inversión en formación estimada de las personas que se han dado de baja y sumado a esto una proporción del salario bruto anual (según Ulhas es un estimado del 0,5 para empleados con conocimiento técnico y 1,5 para managers).

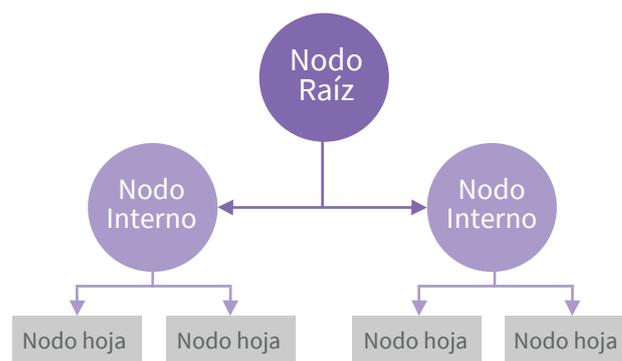
Para entender porque se han ido los empleados, para esto, se realizaron contrastes de hipótesis (por medio de la prueba t) para encontrar en que variables hay diferencias significativas entre el grupo control y el grupo de egreso voluntario (bajas voluntarias) y a su vez, se analizó la información obtenida en la encuesta de salida.

Dado a que el 81% de las bajas voluntarias ocurridas corresponden a empleados de rol técnico, se replicó este contraste controlando el tipo de rol (homogenización de la muestra). Al enfocar en este grupo y posteriormente encontrar un modelo predictivo de la rotación voluntaria, que contemple las variables que permiten diferenciar de manera estadísticamente significativa entre el grupo control y el grupo que se da de baja voluntaria.

La primera técnica estadística aplicada para conseguir este modelo predictivo fue un modelo de regresión logística, el cual es un algoritmo que, por medio de una ecuación, permite predecir la probabilidad de ocurrencia de un resultado binario; en este caso la permanencia (0) o salida voluntaria (1) de un empleado (Raza, Munir, Younas y Fareed, 2022).

La segunda técnica aplicada es el árbol de decisión (decision tree). Un árbol de decisión es una técnica de aprendizaje supervisado no paramétrico, que por medio de un algoritmo genera una estructura jerárquica en forma de árbol, que consta de un nodo raíz, ramas, nodos internos y nodos hoja (IBM, s.f.).

Figura 1. Diagrama ejemplo de un árbol de decisión.



Los árboles de decisión empiezan con un nodo raíz (círculo azul), del cual salen ramas (representado por las flechas) que van hacia los nodos internos (círculo verde), también llamados nodos de decisión; estos representan subconjuntos de evaluaciones basados en atributos clave incorporados al modelo. Por último, están los nodos hoja (rectángulo gris), que representa los posibles resultados de un conjunto de datos (IBM, s.f.).

La estructura del árbol, depende de la cantidad de evaluaciones de atributos necesarios, estas evaluaciones el algoritmo identifica de puntos de división de forma recursiva de arriba hacia abajo, hasta que los registros hayan sido clasificados con etiquetas de clase específica (Raza, Munir, Younas y Fareed, 2022 & IBM, s.f.).

El tamaño del árbol permite evaluar visualmente la complejidad del modelo y el ajuste a los datos: Los árboles más pequeños permiten clasificar los datos en grupos grandes, más homogéneos y con menor cantidad de clases, siendo así más simples. Por lo contrario, los árboles grandes tienen una gran cantidad de clases con muy pocos datos dentro de cada subárbol, cuando esto ocurre, se conoce como fragmentación de datos y esto puede provocar un sobreajuste a los registros, impidiendo la generalización del modelo generado a otros datos.

Entre los principales beneficios de los árboles de decisión, esta que permiten manejar datos continuos y categóricos, no son sensibles a la multicolinealidad (si dos variables atributo tienen una correlación alta, el algoritmo solo seleccionará una para las clasificaciones) y es fácil de interpretar, mostrando que atributos son más importantes para la predicción (IBM, s.f.). Las desventajas de estos modelos son la tendencia a sobreajustarse a los registros y su sensibilidad a pequeñas variaciones (IBM, s.f.).

En nuestro caso, el propósito de generar estos modelos es establecer un sistema de alarmas con el que se identifiquen los empleados con mayor riesgo de egreso voluntario, que, al cruzar contra listados de empleados a retener, sea información clave para los managers y el departamento de recursos humanos, para la gestión adecuada del talento.

Para la comparación de estos modelos, se utilizaron las matrices de confusión en la que se comparan los resultados pronosticados por los modelos con los resultados reales. Una matriz de confusión, esta consiste en contabilizar los falsos-positivos, falsos-negativos, verdaderos-positivos y verdaderos-negativos.

A partir de esto, se calculan los indicadores que se muestran en la Tabla 2.

Tabla 2. Criterios para la Comparación del Rendimiento de los Modelos de Predicción

Medida	Descripción	Cálculo
<i>Sensibilidad</i>	Medida de la habilidad del modelo de identificar los casos positivos de manera correcta	$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$
<i>Especificidad</i>	Medida de la proporción de casos negativos correctamente identificados.	$Specificity = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$
<i>Precisión</i>	Medida de la habilidad del modelo para evadir falsos positivos en la identificación de verdaderos-positivos.	$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$
<i>Exactitud</i>	Medida de la proporción de predicciones correctas (tanto positivas como negativas) del modelo.	$Accuracy = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative}}$
<i>Puntuación F1</i>	Basada en la media armónica de la precisión y sensibilidad. Resume el desempeño general del modelo.	$F1 \text{ score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}}$

5. Resultados

5.1. Describiendo la rotación: Análisis Exploratorio

5.1.1. Descripción de la muestra

Para este estudio, se recopilaron los registros de 980 empleados, que constituyen las personas que formaron parte de la empresa durante los años 2021 y 2022 (ver Gráfico 1).

Esto incluye la plantilla a cierre de diciembre 2022 (800 personas), las desincorporaciones ocurridas desde el 2021 (131 personas), y las personas que rotaron internamente entre empresas del grupo (65 personas).

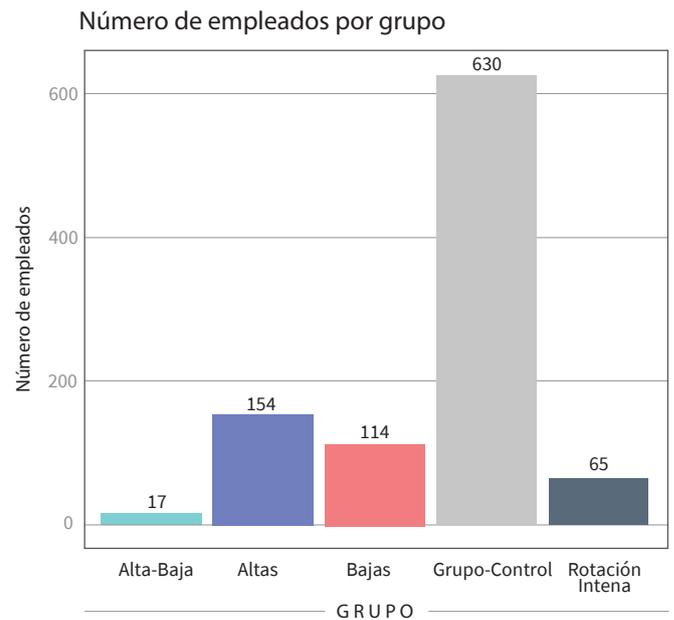


Figura 1. Número de empleados por grupo.

Tabla 3. Estadísticos descriptivos

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	0	2	10	12.6	20	50	11.8	93.6%
EDAD_ANNOS	23	39	46	45.2	52	73	8.8	19.4%
DESEMPEÑO	1	2	2	2.4	3	4	0.7	29.1%
PERCENTIL_SBA	1	3	6	5.5	8	10	2.9	52.1%
DIF_SALARIAL	0	0	3	3.9	4	90	6.7	172.7%
ABS_AJUSTADO_DIAS	0	0	0	20.5	8	709	61.7	300.9%
ABS_AJUSTADO_CASOS	0	0	0	0.8	1	12	1.5	181.3%
BRADFORD	0	0	0	179.3	8	36579	1520.0	847.9%
EQUIDAD_PCT	0	1	1	1.0	1	4	0.3	24.4%
HORAS_FORMACION	0	12	32	47.2	57	1156	79.6	168.5%
FORMACION_INVERSION	0	175	584	813.0	1062	10894	973.2	119.7%

Para describir la muestra, se graficaron los principales demográficos (ver Gráfico 2) y se calcularon estadísticos descriptivos de tendencia central y dispersión (ver Tabla 1). A partir de estos, se puede destacar los siguientes elementos.

Las personas que han sido empleados en la empresa son en su mayoría hombres (62%) y tienen una edad promedio de 45,2 años, por lo que la mayoría pertenecen a la generación X (54%). En cuanto al rol, un poco más de la mitad son técnicos (53%), seguidos luego de administrativos (22%). Según la Dirección Organizativa, dos quintos pertenecen a la dirección de Tecnología (40%); la segunda dirección más grande es Negocio (29,3%), seguida de Finanzas (11,2%) y Personas (8,4%).

En cuanto a la antigüedad, casi un tercio tienen entre 11 y 25 años en la empresa (32,6%; Md= 12,6). Según el factor Bradford, al menos hasta el 75% de los empleados tienen un absentismo bajo. En cuanto al salario, al considerar el rol y departamento, se destaca que el promedio de las personas cuenta con equidad interna ($\bar{X}=1$; Dt=0,3). Por último, en cuanto a la formación: en los últimos dos años, cada empleado ha recibido Md=32 horas de formación, con una inversión estimada de €584 (mediana) por empleado.

5.1.2. Descripción de los grupos

5.1.2.1. Demográficos

Entre el 2020 y 2021, ingresaron a la empresa 171 empleados, de los cuales 17 presentaron baja y 154 se mantienen en plantilla; es decir, hay una rotación del 9,94% de los nuevos ingresos.

Estos nuevos empleados desempeñan en su mayoría de rol técnico (63%), el 45% son mujeres y un poco más de la mitad pertenecen a la generación millennial (53%). De estos, los que se mantienen en plantilla (Grupo Altas. Ver Anexo 1), son 46% mujeres, 60% de rol técnico y 53% Millennial (ver Anexo 1), pertenecen principalmente a la dirección de Negocio (42,2% y Tecnología (31,2%); tienen una edad promedio de 40 años. Dentro de la empresa, su desempeño ha sido catalogado como “medio” (2,3/4), solo 12 personas han presentado algún caso de absentismo y en cuanto a su remuneración, su mediana corresponde al percentil 7, por lo que se puede concluir que las nuevas contrataciones suelen estar mejor remuneradas que las personas que los empleados de mayor antigüedad, probablemente como resultado de una estrategia de atracción de talento. Los empleados que

fueron contratados y posteriormente se dieron de baja (Grupo Alta-Baja. Ver Anexo 2), son en su mayoría hombres (59%), de rol técnico (88%) y en su mayoría, de las generaciones Millennial (53%) y Generación X (41%), con una edad promedio de 39 años y pertenecían a las direcciones de Negocio (35,3%) y Tecnología (29,4%).

Comparativamente con los nuevos ingresos que se mantienen en plantilla, los empleados que egresaron tienen un desempeño superior de 2,7/4 y su remuneración, es un poco inferior, pero aún superior a la de empleados con mayor antigüedad (Md=6). El resto de las personas que se dieron de baja (Grupo Bajas. Ver Anexo 3), tienen un desempeño promedio de 2,5 y salarialmente, tienden a estar igual que el grupo Alta-Baja. Sin embargo, se distinguen de los grupos anteriores en que en su mayoría tienen una antigüedad de 8 años (mediana); más de dos tercios son hombres (68%), su edad promedio es de 47,7 años, en donde el 39% pertenece al grupo generacional Baby Boomer, el 35% a Millennial y el 25% a Generación X y la mayor parte de estos pertenecen a la dirección de Tecnología (41,2%). Al analizar las bajas por motivo de baja, se puede ver que el 55,7% de las bajas son involuntarias, y el 44,2% restantes son bajas voluntarias. Considerando las bajas voluntarias, se calculó el índice de rotación voluntaria acumulado anual, y se encontró que, para el cierre del 2022, este fue del 4,8% ($\Delta 2,4\%$ superior al cierre del 2021), y al calcularlo solo para los incorporados en los últimos dos años, el índice al cierre del 2022 fue del 12,5% ($\Delta 4,8\%$ superior al cierre del 2021). Las bajas voluntarias (ver Anexo 5), son principalmente de rol técnico (81%), hombres (71%) y con una antigüedad inferior a 5 años (77,6%); su edad promedio es de 37,3 años, donde el 67% son de la generación Millennial y el 31% de la Generación X, y pertenecen principalmente a las direcciones de Negocio (34,5%) y Tecnología (29,3%).

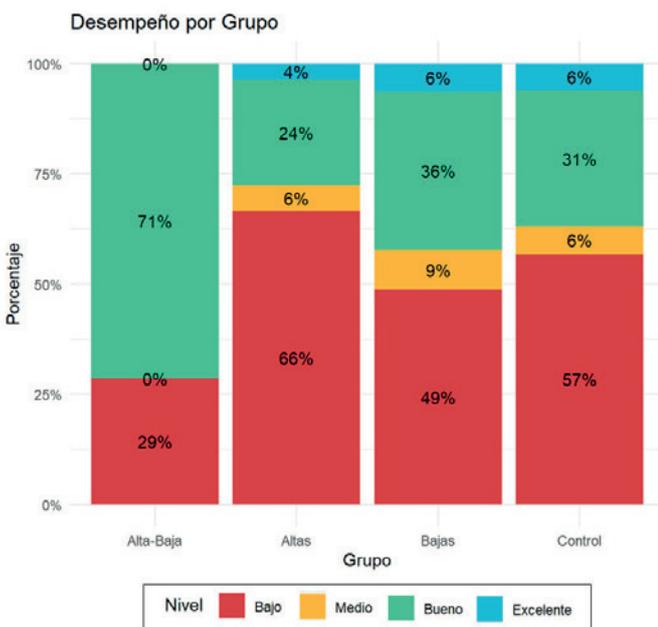
Las bajas involuntarias (ver Anexo 6) se diferencian en que tienen una mayor proporción de mujeres (37%), tienen empleados principalmente con el rol de administrativo (45%) y técnico (41%), su antigüedad supera los 25 años (52,1%) y son mucho mayores en cuanto a edad ($\bar{X}=54$ años) en donde el 60% pertenecen a la generación Baby Boomer y son más que todo de la dirección de Tecnología (47,9%).

5.1.2.2. Desempeño

El 92% (901) de los empleados tienen algún registro de evaluación de desempeño, en esta evaluación se le asigna a cada empleado una categoría de bajo (1) a excelente (4).

En la gráfica 2, podemos observar el porcentaje de personas por categoría dentro de cada grupo de análisis. Mediante esta gráfica, se puede apreciar que los grupos con más empleados de desempeño “excelente”, son el grupo Control (6%) y el grupo de Bajas (6%); mientras que para la categoría de desempeño “bueno” son el grupo de Alta-Baja (71%) y Bajas (36%). Por parte, los grupos con más empleados en la categoría “bajo”, son el grupo de Altas (66%) y Control (57%). Estas observaciones también se contrastan con los promedios, en donde pareciera que las personas mejor calificadas en cuanto a su desempeño son las nuevas contrataciones que se dan de baja ($\bar{X}=2,7$), seguido del resto de las bajas ($\bar{X}=2,5$); mientras que las personas que se mantienen en plantilla son las que cuentan con una menor valoración de desempeño (grupo control $\bar{X}=2,4$; altas $\bar{X}=2,3$). En base a esto, se puede señalar la existencia de un riesgo de fuga de talento de alto desempeño.

Figura 2. Distribución de desempeño por grupo.

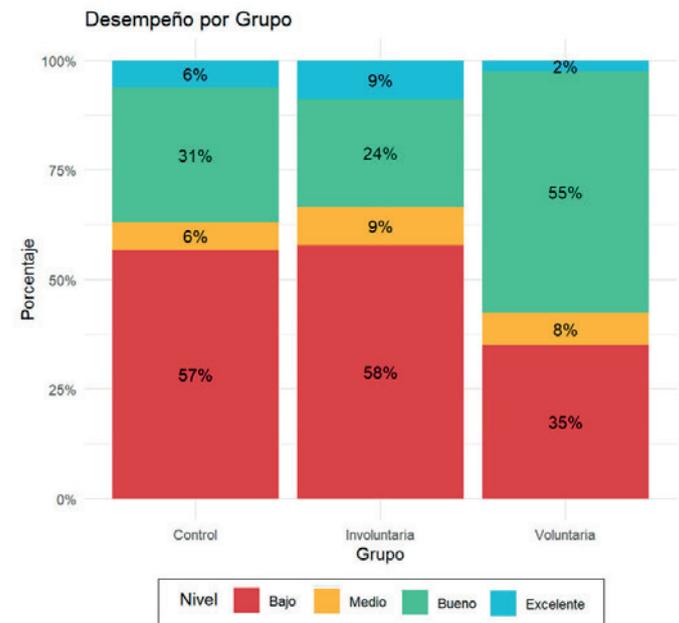


Al observar el desempeño de las bajas según tipo de baja (ver Gráfica 3), se puede observar que la distribución del grupo de baja involuntaria es muy similar a la del grupo control, con la diferencia de que este grupo, tiene un mayor porcentaje de personas en la categoría de “excelente” (diferencia neta $\Delta 3\%$).

Por el contrario, el grupo de bajas voluntarias, tienen una menor proporción de personas en el grupo de excelente, pero también tienen menos personas en las categorías de desempeño “medio” y “bajo”, agrupando un 55% en la

categoría de desempeño “bueno”; lo cual parece confirmar que estos egresos son un riesgo de fuga de talento, ya que pareciera que las personas con mejor desempeño que se van de la empresa lo hacen voluntariamente y no por motivos como jubilación o un fin de contrato.

Figura 3. Distribución de desempeño por tipo de baja



5.1.2.3. Remuneración

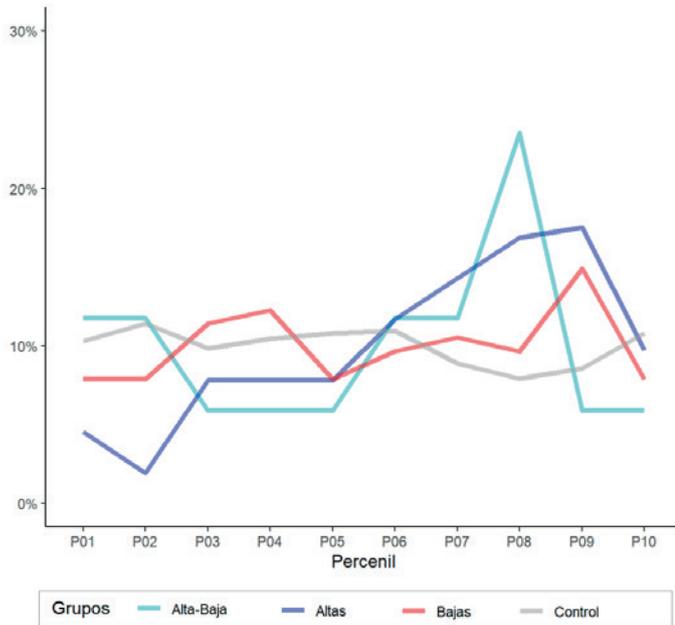
Para analizar la remuneración comparativamente, se calcularon percentiles (controlando por la dirección organizativa de pertenencia) para segmentar a los empleados en 10 grupos compuestos por la misma cantidad de personas (10%) en cada segmento. En la gráfica 4, se pueden apreciar las distribuciones por cada grupo de análisis.

En estos, el grupo Control tiene una distribución similar a la esperada para toda la distribución (todos los segmentos tienden al 10%; $\bar{X}=5,4$; $Dt=2,9$), para el resto, se pueden visualizar modas y asimetrías para distinguir entre las distribuciones. El grupo Alta-Baja ($\bar{X}=6$; $Dt=2,9$), se distingue de los demás en que tiene una asimetría negativa, con la mayor proporción de personas en P08, es decir en los percentiles superiores de toda la distribución y cuentan con un valle entre los percentiles P03 y P04. El grupo de Altas ($\bar{X}=7$; $Dt=2,5$); también tiene una asimetría negativa, agrupando a la mayor proporción de personas entre P07 y P09, mientras que en P01 y P02 no llega al 10%. La distribución de este grupo de empleados nuevo ingreso, es más homogénea (menos dispersa) que la de los

que egresan y su promedio es superior. En base a esto, se podría concluir que este grupo de personas son las mejor remuneradas.

Por último, el grupo que agrupa el resto de las personas que causan baja, ($\bar{X}=6$; $Dt=2,9$) tienen una distribución más parecida al del grupo control, tendiendo al 10% en todos los grupos, con excepción de un pico marcado en P09.

Figura 4. Distribución de percentil salarial por grupo.

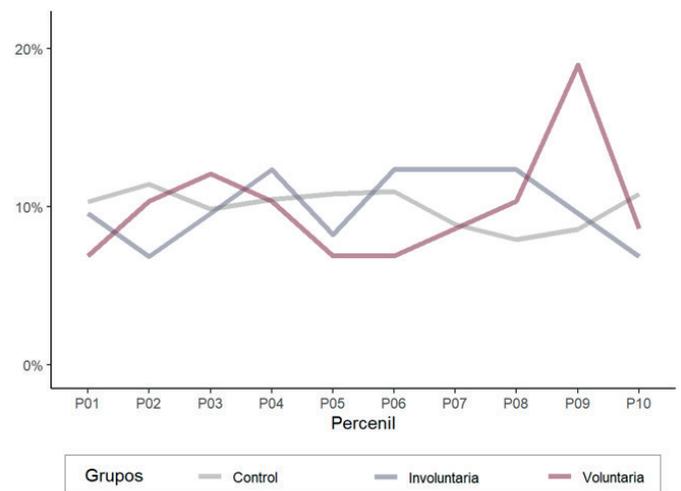


En cuanto a la remuneración según el tipo de baja, en el gráfico 5, se observa que los empleados que egresan de manera involuntaria tienen una distribución más parecida a la del grupo control, que a la de los empleados que egresan voluntariamente. La distribución de las bajas voluntarias muestra un pico en P09, donde un 20% de las personas que se dan de baja voluntaria, están en el percentil 90% de su distribución correspondiente; sin embargo, también se observa un valle entre los percentiles medios (P05 y P06).

En cuanto al promedio, las bajas voluntarias están un poco por encima ($\bar{X}=5,8$; $Dt=2,9$) que las involuntarias ($\bar{X}=5,6$; $Dt=2,7$); sin embargo, debido a que el coeficiente de variación del grupo de voluntarias es de 50,6%, este resultado debe tomarse con precaución. Al comparar el porcentaje de diferencia salarial entre 2020 y 2021, se destaca que el promedio de las bajas voluntarias e involuntarias es el mismo ($\bar{X}=1,6\%$), mientras que para el grupo control, es del $\bar{X}=4,5\%$; esto podría indicar que las personas que se dan de baja se diferencian de las que se

mantiene en plantilla, en cuanto a que estas últimas han recibido en promedio, un mayor aumento salarial en el último año.

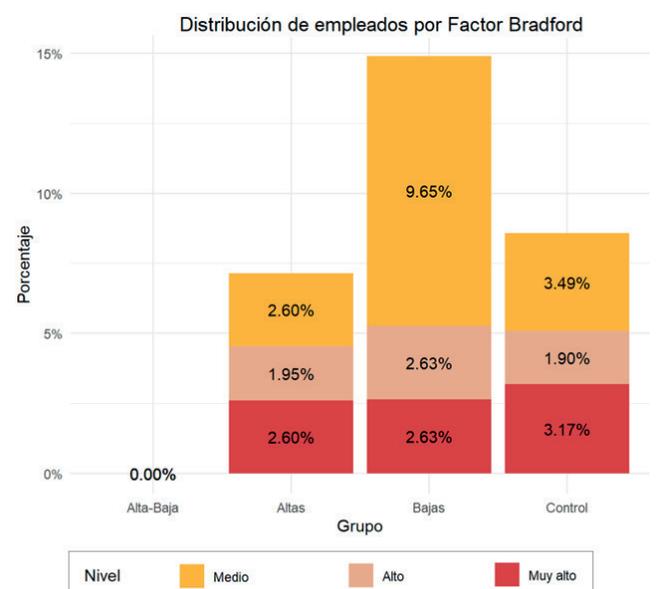
Figura 5. Distribución de percentil salarial por tipo de baja.



5.1.2.4. Absentismo

En cuanto al absentismo, al categorizar por el Factor Bradford, se puede observar que, en todos los grupos, al menos el 91% de los empleados se encuentran en la categoría “bajo”, con la excepción del grupo de Bajas, donde el 85,1% de los empleados entra en esta categoría y el grupo Alta-Baja, en donde, por el contrario, todos los empleados tienen un absentismo bajo (ver Gráfica 6). El grupo control es el segundo que más empleados tiene en una categoría distinta a bajo (8,6%), y es además el grupo que cuenta con más empleados en el grupo de absentismo “muy alto”.

Figura 6. Distribución de Factor Bradford por grupo.



Al comparar las diferencias de esta distribución entre el grupo control y las bajas por tipo, se puede observar que las bajas voluntarias son el grupo con menor proporción de personas en absentismo “muy alto” (1,72%) y medio (1,72%); también se puede señalar que existen más diferencias entre el grupo de bajas voluntarias y el de bajas involuntarias que contra el grupo control (ver Gráfico 7). Por su parte, el grupo de baja involuntaria es el que presenta una mayor proporción de personas con alguna categoría de absentismo (distinta a bajo), teniendo un 13,7% en la categoría de absentismo “medio”.

5.1.2.5. Formación

En cuanto a la formación y la inversión estimada por empleado, se puede observar que los grupos conformados por persona que han egresado de la empresa (Bajas y Alta-Baja) son los que tienen el menor promedio tanto en horas, como en inversión (ver Gráfica 8 y Gráfica 9). Por su parte, el grupo de Altas son el grupo que cuenta con mayor inversión por empleado, mientras que el grupo control cuenta con mayor cantidad de horas. Esto podría ser explicado porque las personas que han egresado han estado menos tiempo dentro de la empresa y por lo tanto contaron con menos oportunidades de recibir formación. Por otro lado, la hora por formación del 2022 resulto más costosa que el 2021 y esto podría explicar porque las altas, aunque tienen menos horas que el grupo control, tienen una inversión superior.

Figura 7. Distribución de Factor Bradford por tipo de baja.

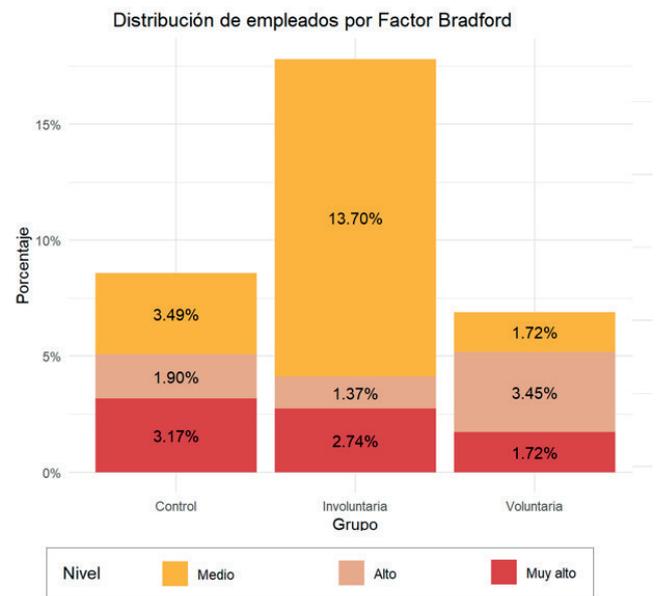


Figura 8. Distribución de inversión estimada en formación por grupo.

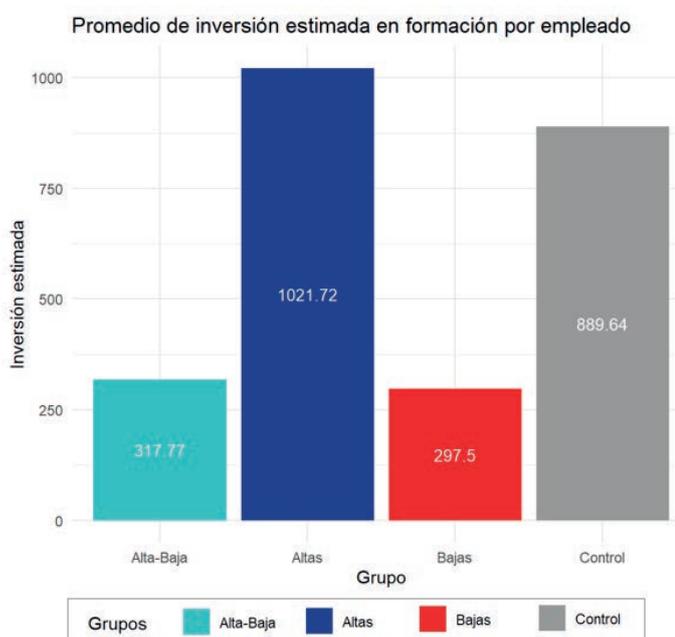
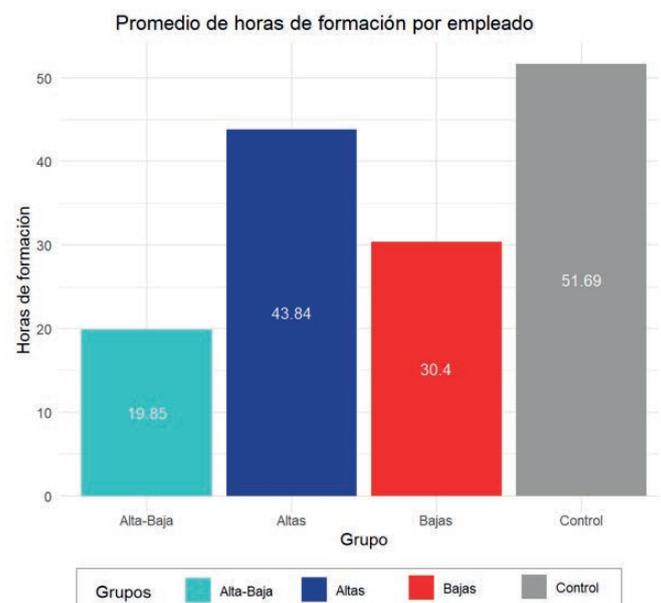


Figura 9. Distribución de horas de formación por grupo.



Finalmente, al comparar los tipos de baja por las horas (ver Grafica 10) e inversión en formación (ver Grafica 11), se pueden observar que el grupo control cuenta con mayor cantidad de horas e inversión que las bajas, sin importar el tipo de baja, y además, se puede señalar que las personas que egresan de forma involuntaria, están por debajo de los que se dan de baja de forma voluntaria, tanto en las horas (Δ -45%; diferencia neta=180,67 horas) como en la inversión estimada en formación (Δ -53%; diferencia neta=-21,94€).

Figura 10. Distribución de inversión estimada en formación por tipo de baja

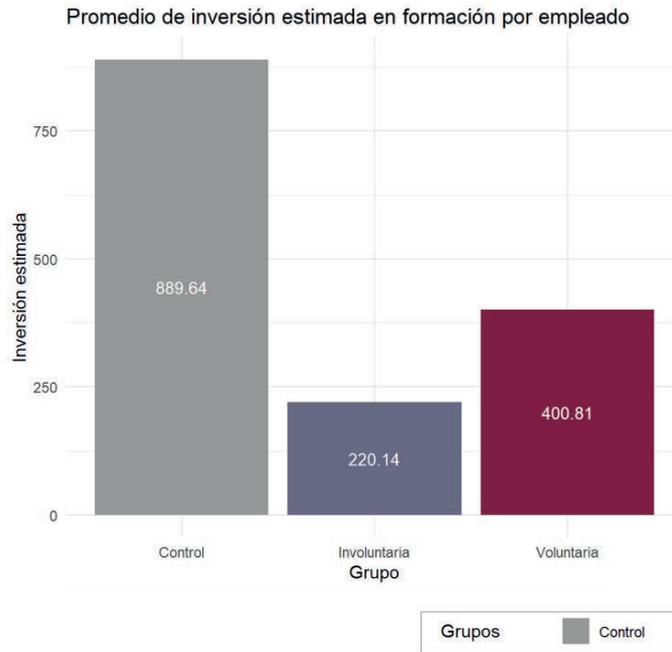
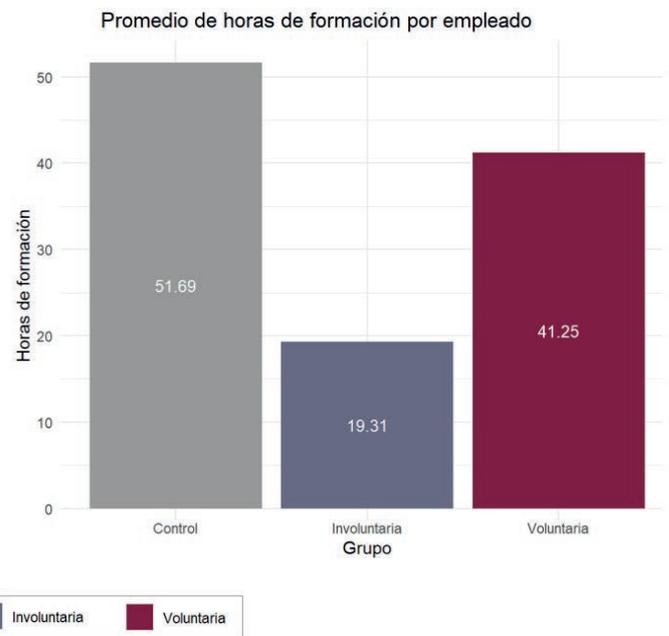
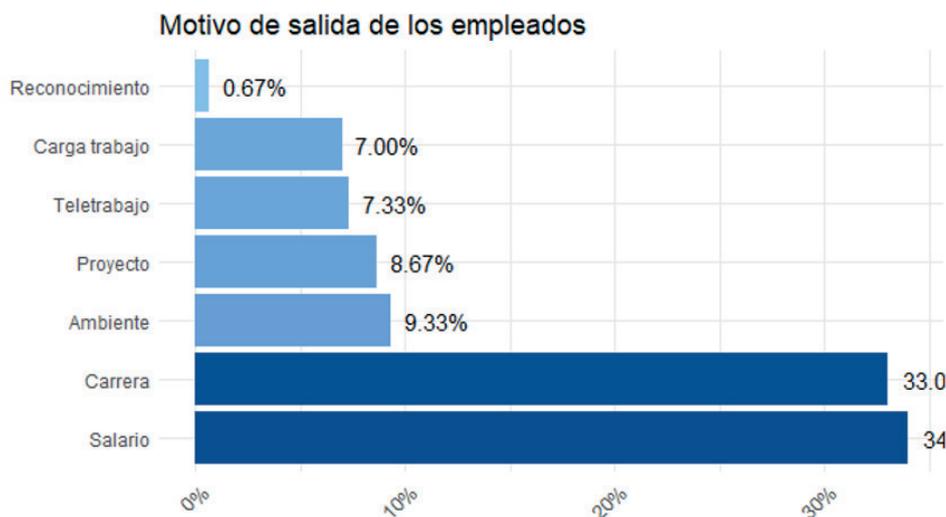


Figura 11. Distribución de horas de formación por tipo de baja



Para las bajas voluntarias, también se consta de información sobre su motivo de baja, según lo expuesto en la entrevista de salida. Estos motivos fueron ponderados para entender cuál tiene mayor peso para los trabajadores (ver Gráfica 12). Los principales motivos de salida señalados fueron el salario (34%) y la carrera profesional (33%). En segundo plano, también se indican como importantes el ambiente de trabajo (9,33%), el proyecto (8,67%), tener más teletrabajo (7,33% y la carga de trabajo (7%). Esto señala la retribución y plan de carrera como principales puntos de dolor asociados al egreso voluntario de los empleados en la empresa.

Figura 12. Motivo de baja voluntaria expuesto en la encuesta de salida.



5.2. Describiendo la rotación: contrastando las diferencias entre grupos

Se realizó un primer contraste de hipótesis para investigar las posibles diferencias entre los empleados que abandonan la empresa de forma voluntaria y aquellos que permanecen en ella. Se compararon ambos grupos en términos de variables demográficas, organizativas, económicas, de formación y de absentismo.

La hipótesis nula planteada en estos contrastes es que no existen diferencias significativas entre los dos grupos con relación a estas variables. Para evaluar esta hipótesis, se empleará la prueba t de Student, la cual permitirá identificar aquellas variables en las que se encuentren diferencias estadísticamente significativas.

Esta información será luego utilizada como filtro para seleccionar las variables a incluir en el modelo predictivo. De esta manera, se considerarán únicamente aquellas variables que muestren una diferencia estadísticamente significativa entre los dos grupos, minimizando al máximo la probabilidad de cometer un error de tipo I (falso positivo o rechazar la hipótesis nula cuando es verdadera).

En la Tabla 4 se pueden identificar las variables que variables muestran diferencias significativas. Estas variables incluyen el grupo generacional ($t=6,5$; $p<0,001$) y la edad ($t=8,24$; $p<0,001$), el grupo de antigüedad ($t=13,2$; $p<0,001$), tener más o tener menos de 2 años de antigüedad ($t=-7,7$; $p<0,001$) y los años de antigüedad ($t=16,38$;

$p<0,001$), la valoración del desempeño ($t=-2,8$; $p<0,001$), la diferencia salarial ó aumento salarial entre 2020 y 2021 ($t=5,4$; $p<0,001$) y la inversión en formación ($t=2,63$; $p<0,001$).

El grupo generacional se utilizó brindar un contexto a la información de la edad y facilitar su interpretación. Sin embargo, dado a que tanto la agrupación generacional como la edad (en años) resultaron significativas, se procederá a incluir únicamente la edad en su escala continua.

De manera similar, se realizaron contrastes entre los años de antigüedad y los grupos de antigüedad. Dado a que ambos fueron significativos, se proseguirá solo con los años de antigüedad en su escala continua. Además, se puede destacar que la clasificación que divide a las personas en más (dummificada con el valor 0) o menos de dos años de antigüedad (dummificada con el valor 1), también generó una diferencia significativa. Esto, en concordancia con los resultados del análisis descriptivo, parece apuntar a que existe un problema de retención del talento en sus primeros dos años en la empresa. Por lo tanto, se recomienda implementar acciones o intervenciones que faciliten la adaptación y satisfacción del empleado desde el onboarding y un seguimiento post incorporación.

Tabla 4. Prueba t-student.

VARIABLE	CONF.LOW	CONF.HIGH	STATISTIC	P.VALUE	SIG
GRUPO_GENERACIONAL	0.308	0.582	6.478	0.000	**
GRUPO_ANTIGUEDAD	1.407	1.906	13.218	0.000	**
ANTIGUEDAD_ANNOS	10.395	13.254	16.380	0.000	**
EDAD_ANNOS	6.984	11.321	8.424	0.000	**
DESEMPEÑO	-0.519	-0.086	-2.816	0.007	*
GRUPO_ROL	-0.359	0.202	-0.558	0.578	
SEXO	-0.051	0.201	1.188	0.239	
DIRECCION_SL	-0.750	0.011	-1.936	0.057	
SBA_PROMEDIO	-2613.944	6216.345	0.809	0.420	
PERCENTIL_SBA	-1.264	0.344	-1.141	0.258	
DIF_SALARIAL	0.019	0.041	5.424	0.000	**
ABS_AJUSTADO_DIAS	-2.918	18.095	1.438	0.154	
ABS_AJUSTADO_CASOS	-0.054	0.624	1.679	0.098	
BRADFORD	-62.267	122.508	0.647	0.519	
ANTIGUEDAD_2_ANNOS	-0.643	-0.380	-7.770	0.000	**
EQUIDAD_PCT	-0.040	0.097	0.820	0.415	
HORAS_FORMACION	-30.124	50.990	0.515	0.609	
FORMACION_INVERSION	117.655	860.007	2.633	0.011	*

En la tabla 5, se muestran las correlaciones entre las variables previamente descritas y el grupo de pertenencia (baja voluntaria y control).

A través de estas correlaciones, podemos observar que todas las relaciones son bajas. Según su dirección, podemos interpretar lo siguiente:

- Antigüedad: Existe una relación negativa, lo que significa que los empleados con menos antigüedad tienden a tener una mayor propensión a abandonar la empresa de manera voluntaria ($r_{xy}=-0,317$; $R^2=10\%$).
- Edad: Hay una relación negativa, lo que implica que los empleados más jóvenes tienen una mayor tendencia a generar egreso voluntario ($r_{xy}=-0,323$; $R^2=10,4\%$).
- Variación/aumento salarial: Existe una relación negativa, lo que indica que aquellos empleados con una menor variación salarial tienen una mayor tendencia a generar egreso voluntario ($r_{xy}=-0,128$; $R^2=1,6\%$).
- Inversión en formación: Hay una relación negativa, lo que significa que los empleados que reciben una menor inversión en formación tienen una mayor tendencia a generar egreso voluntario ($r_{xy}=-0,14$; $R^2=1,95\%$).
- Desempeño: Existe una relación positiva, lo que implica que los empleados con mayor desempeño tienen una mayor propensión a generar egreso voluntario ($r_{xy}=0,104$; $R^2=1,08\%$).

Tabla 5. Matriz de correlación entre las variables y el grupo de pertenencia (baja voluntaria vs. control)

VARIABLE	RXY	R2
ANTIGUEDAD_ANNOS	-0.317	10.049
EDAD_ANNOS	-0.323	10.433
SBA_PROMEDIO	-0.018	0.032
PERCENTIL_SBA	0.044	0.194
DIF_SALARIAL	-0.128	1.638
EQUIDAD_PCT	-0.03	0.090
ABS_AJUSTADO_DIAS	-0.044	0.194
ABS_AJUSTADO_CASOS	-0.061	0.372
BRADFORD	-0.015	0.023
HORAS_FORMACION	-0.035	0.123
FORMACION_INVERSION	-0.14	1.960
DESEMPENNO	0.104	1.082

5.3. Prediciendo la rotación: Modelo de regresión múltiple

Se desarrolló un primer modelo de regresión logística múltiple, utilizando como variable dependiente o predicha el estado del empleado (baja voluntaria o activo en plantilla), y como variable independiente o predictora, los años de antigüedad, el desempeño, edad del empleado, inversión en formación estimada y la diferencia o aumento salarial percibido.

En el anexo 7 se presenta la salida del cálculo del modelo de regresión logística. con las variables descritas anteriormente. Por medio de este modelo de regresión logística múltiple, podemos observar que, aunque previamente las variables por separado explicaban muy poco, en conjunto, dentro del modelo, pueden explicar hasta el $R^2=28.1\%$ (Chi^2 p valor = 0.000) de que un empleado pertenezca al grupo control o al grupo de bajas voluntarias. A su vez, se puede observar el desempeño ($z=1,23$; $p=0,225$) y la inversión en formación ($z=-1.56$; $p=0,12$) son las variables de menor peso dentro del modelo.

Debido a que la mayor parte de las bajas voluntarias son de perfil técnico, se decidió replicar el mismo procedimiento (pruebas t y correlaciones en el anexo 8), controlando el rol por medio de la homogenización de la variable; es decir, excluyendo al resto de las observaciones cuyo rol es distinto a "Técnico", con el fin de comprobar si esto permite mejorar la capacidad predictiva del modelo. Este segundo modelo permite predecir un $R^2=33.8\%$ (Chi^2 p valor = 0.000) de la varianza de egresar voluntariamente. Además, incluye los casos de absentismo ocurridos, debido a que se presenta como una variable que genera una diferencia significativa ($t=2$; $p<0,05$) y se priorizó por encima de los días de absentismo y el factor Bradford por su coeficiente (ver Anexo 7). Esto señala que, para los empleados de rol técnico, el absentismo es una variable diferenciadora (especialmente los casos ocurridos), en donde las personas con menos casos de absentismo tendrán una mayor propensión a generar un egreso voluntario (ver anexo 9).

5.3.1. Comparación de modelos de regresión logística múltiple

Al comparar los modelos, se puede apreciar ambos permiten hacer una predicción, pudiendo diferenciar de manera estadísticamente significativa cuando un empleado causa baja voluntaria de se mantiene activo en plantilla (Chi^2 p valor = 0.000). Sin embargo, estos modelos difieren en varios aspectos (ver tabla 4):

- La capacidad predictiva, es inferior en el modelo 1 ($R^2=28.1\%$ vs $R^2=33.8\%$), por lo cual se puede concluir que, al controlar el rol, se pueden llegar a conclusiones más acertadas.
- En cuanto a la bondad de ajuste, o la forma en la que los modelos se ajustan a las observaciones; se comparó utilizando el valor del AIC (Criterio de Información de Akaike), por medio del cual, se considera que el modelo será aquel con el valor más bajo.
- El modelo 1 tiene un AIC de 221.7, y al controlar el rol, el modelo 2 cuenta con un AIC de 160. En función a este criterio, se concluye que el modelo 2 es más confiable, ya que tiene un mayor ajuste.

	Modelo 1	Modelo 2
	R^2 0.281	R^2 0.338
	AIC 221.7	AIC 160
	Chi^2 p valor <0.001	Chi^2 p valor <0.001
	Beta (p)	Beta (p)
Intercepto	2.06* (0.039)	1.51 (0.128)
Antigüedad	-3.70** (<0.001)	-3.34* (0.008)
Desempeño	1.21 (0.225)	1.66 (0.098)
Edad	-3.17* (0.002)	-2.42* (0.0155)
Diferencia salarial	-2.65* (0.008)	-1.148 (0.2510)
Formación inversión estimada	-1.55 (0.1200)	-1.84 (0.065)

Tabla 6. Comparación de modelos de regresión



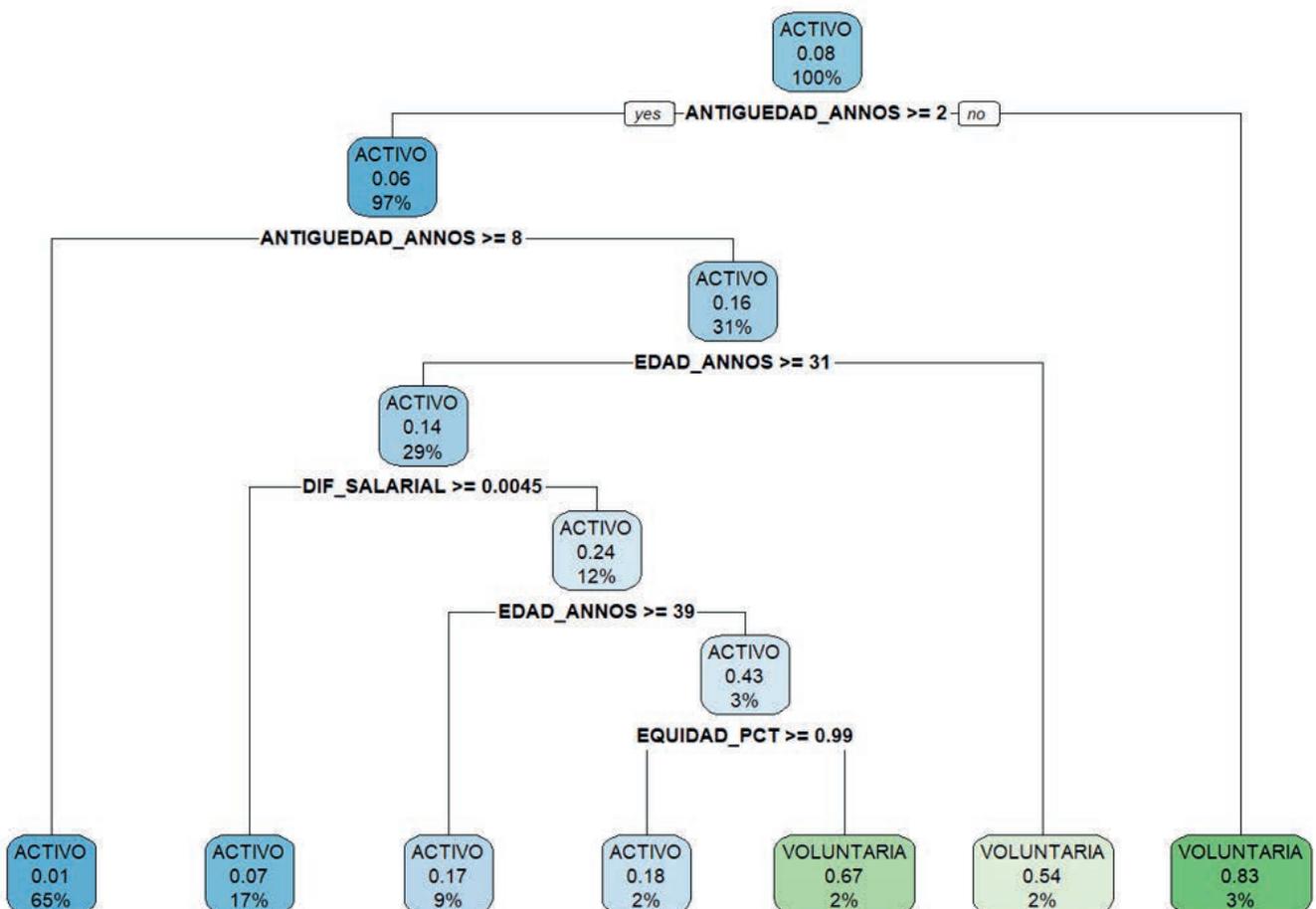
5.4. Prediciendo la rotación: árbol de toma de decisiones (decision tree)

Como segunda técnica estadística, se ejecutó un árbol de toma de decisiones. En este, se introdujeron variables continuas y categóricas (dirección organizativa, sexo, grupo de rol, grupo de antigüedad, grupo generacional y categoría Bradford) y continuas (edad, antigüedad, salario, percentil salarial, diferencia salarial, compara-ratio, días y casos de absentismo y horas e inversión en formación).

A partir de este modelo (ver Gráfico 13), podemos confirmar la observación del análisis exploratorio de datos, de que las personas con menos de dos años en la organización son la que cuentan con un mayor riesgo de

generar una baja voluntaria. Seguido a este segmento de personas, están los empleados con una antigüedad entre 2 y 8 años y entre 31 y 38 años de edad, que no han recibido un aumento salarial y no tienen equidad interna, es decir, salarialmente no se les ha aumentado en los últimos dos años, y reciben un ingreso inferior al de sus pares (considerando el departamento y posición). Esta información, parece ser coherente con lo encontrado en el modelo de regresión logística, pudiendo dar incluso más contexto de los perfiles en riesgo de rotación voluntaria.

Figura 13. Árbol de toma de decisiones.



5.5. Comparación de técnicas analíticas: modelos de regresión logística vs árbol de decisiones

En la tabla 7, se muestra una matriz de confusión con los valores calculados para comparar los modelos de regresión y el árbol de decisiones.

	Modelo Regresión 1	Modelo Regresión 2	Árbol de Decisiones
Sensibilidad	11%	27%	36%
Especificidad	76%	49%	99%
Precisión	24%	18%	81%
Exactitud	60%	45%	95%
Puntaje F1	0.15	0.22	0.5

Tabla 7. Matriz de confusión

Al comparar los modelos de regresión logística, se observa que:

- La sensibilidad fue superior en el modelo 2, es decir, este modelo es mejor para identificar correctamente los casos de egreso voluntario.
- La especificidad fue superior en el modelo 1, en donde este modelo es mejor a la hora de identificar los casos de permanecer en plantilla.
- La precisión es superior en el modelo 1, en donde este modelo tiene una mayor habilidad de evadir falsos positivos, es decir, señalar como con probabilidad de egreso a alguien que no la tiene.
- La exactitud es superior en el modelo 1, en donde este modelo demuestra una mayor proporción de predicciones correctas.
- El F1 es superior en el modelo 2, lo cual es un indicativo de que este modelo tiene un mejor desempeño que el modelo 1 a la hora de predecir.

En conclusión, el modelo 2 se considera preferible, debido a que tiene un mejor ajuste a los datos (criterio AIC) u un mejor desempeño en general para predecir la rotación voluntaria (mayor R^2 , sensibilidad y F1). Esto señala la necesidad de considerar el rol como variable moderadora, por lo cual resultaría recomendable, contar con un modelo predictivo por cada agrupación de rol existente, ya que el rol parece condicionar o influir entre la relación de las variables incluidas sobre la probabilidad de egreso

voluntario. Al agregar el árbol de decisión a esta comparativa, se puede señalar que este supera a ambos modelos de regresión logística en todas las métricas calculadas con la matriz de confusión, demostrando un mejor desempeño en general, con mayores aciertos y menos errores.

Considerando esto, podemos concluir que el árbol de toma de decisiones resulta una mejor técnica estadística para predecir la rotación voluntaria, y principalmente, para predecir definir un perfil de las personas que se espera que se mantengan en plantilla con mayor probabilidad.



6. Discusión

Por medio del análisis realizado, se puede observar que, comparativamente con el contexto post pandemia descrito, la rotación voluntaria ocurrida en la empresa no resulta alarmante (índice de rotación voluntaria 2022= 4,8%), incluso si se compara con el benchmarking de la industria aseguradora en España (índice de rotación voluntaria= 6,99%) reportado por ICEA el 2022.

No obstante, la rotación voluntaria ha venido en aumento (Δ 2,4% superior al cierre del 2021), particularmente en las nuevas contrataciones (Δ 4,8% superior al cierre del 2021).

El coste de general de estas bajas voluntarias, estimado a partir del salario bruto anual fue de 1.317.589,33€, y el costo por reemplazo fue en promedio, de 89 días (time to hire) por empleado, y una inversión en euros para la selección que ha aumentado un Δ 38,7% en comparación al 2021. El costo por separación, calculado a partir de la formación, fue de 2.392,73 horas, con una inversión estimada de 23.247€.

Por estos costes, la tendencia en aumento del egreso voluntario, y su asociación con satisfacción laboral y salud mental, sigue resultando importante controlar la rotación, especialmente en los casos en los que el riesgo de fuga sea de empleados con alto desempeño o impliquen un alto riesgo.

Entre los resultados encontrados en el análisis descriptivo, se destacó que:

La mayor proporción de la plantilla pertenece a la generación X (54%), considerando que la edad promedio de jubilación para el 2022 es de 66 años (Instituto Santalucía, 2023), estos empleados tienen un estimado de 15 años más de vida laboral antes de su jubilación. Considerando esto, y que la mayor parte de las bajas voluntarias son Millennials (67%), se podría considerar necesario empezar a diseñar planes de retención para este segmento, con el fin de poder suplir las necesidades de talento a futuro.

Además, se destacó que las personas que egresan voluntariamente tenían a estar entre las categorías de desempeño bueno y excelente en mayor proporción que los empleados que se mantienen en plantilla (57% vs 37%), por lo cual se señala que, a la hora de trabajar la rotación voluntaria, es importante ser estratégico y reconocer que empleados se desean retener, ya que, en el estado actual, pareciera que hay una mayor fuga de talento de alto desempeño.

En cuanto a los modelos calculados, por medio de los de regresión logística múltiple podemos concluir que se obtiene un mejor desempeño al hacer predicciones controlando el rol de los empleos, y que las variables que nos permiten predecir de forma más estable, son la antigüedad, la edad, el desempeño, la inversión en formación y la diferencia o aumento salarial recibido.

Por su parte, el árbol de toma de decisiones, resulto ser más sensible, específico, preciso, exacto y en general, tener un mejor desempeño que los modelos de regresión. Destacando dos perfiles con mayor riesgo de egreso voluntario:

- Empleados con menos de 2 años de antigüedad.
- Empleados con entre 2 y 8 años de antigüedad y entre 31 y 38 años de edad, que no percibieron un aumento salarial y no cuentan con equidad interna.

Sin embargo, es necesario señalar que los árboles de toma de decisión tienen una tendencia a sobre ajustarse a los datos, y dado a que los cálculos que realiza y en general la salida producida (output) son una capa oculta, puede resultar éticamente cuestionable el uso de estos resultados y es necesario tener precaución de no incurrir en la discriminación los empleados.

Aunque todos estos modelos resultaron estadísticamente significativos para explicar la rotación voluntaria, ninguna de las variables introducidas es accionable. Es decir, ninguna de las variables permite diseñar una ruta de intervención clara con la cual se pueda apalancar los factores “stay” de la decisión de mantenerse en la empresa por parte del empleado.

Algunas de estas variables, podrían incluir: percepción de la cultura y clima laboral, satisfacción con el crecimiento y oportunidades de carrera, satisfacción con la remuneración, reconocimiento y feedback, satisfacción con las relaciones con los managers y compañeros de trabajo, flexibilidad o balance vida-trabajo (Ulhas, 2019 & Chatzoudes, 2022), posición, centralidad y tipo de red social de pertenencia (Younis, Ahsan y Chatteur, 2023) y variables de personalidad (Bossi, et. e. 2022).

Introducir algunas de estas variables sería clave, ya que esto, al menos según la bibliografía revisada, no solo son variables que permiten trazar un objetivo en el diseño de un plan de retención, sino que este tendría evidencia empírica al contener variables asociadas con la satisfacción y salud mental del empleado, así como en una

reducción del índice de rotación voluntaria para la empresa.

Siguiendo esta línea de ideas, también resulta preponderante diseñar los planes de intervención con estrategias en base a evidencia empírica, como las señaladas por Allen, Bryant y Vardaman (2010), de las cuales se podría hacer énfasis en la personalización de recompensas (compensación), las ofertas de trabajo realistas (reclutamiento), el uso de mentores (socialización) y las ofertas de desarrollo atadas a la antigüedad (aprendizaje y desarrollo).

Aunado a esto, se recomienda tener claro un listado de talento clave/crítico a retener, ya que, de no contar con dicho listado, si se aplica una intervención, esta no solo se estaría aplicando a ciegas, sino que no se tendría un objetivo claro, con lo cual se incurriría al riesgo de que esta intervención genere un coste superior al de la rotación y que, además, de tener un beneficio, este no sea fácilmente medible.

Para esto, se recomienda replicar este estudio, incluyendo variables accionables, relacionadas como la satisfacción subjetiva con la recompensa (beneficios y salario), trabajo colaborativo (con el equipo y entre áreas), cohesión con compañeros de trabajo, desarrollo profesional, el liderazgo organizacional y relación con el supervisor; la sensación de estabilidad, equidad y autonomía y la capacidad de lograr balance o integración vida (Griffeth, Hom y Gaertner, 2000).

Una vez se cuente con esta información y nuevos modelos, se podrá diseñar un plan de intervención, que puede ser basado en estrategias con evidencia empírica, como las

mencionadas por Allen, Byant y Verdaman (2020) y aplicadas al personal que resulte estratégico retener. A la hora de aplicar esta intervención, se podría aplicar un diseño pre-test post-test con grupo control, en donde se compare la satisfacción y rotación de los empleados a los que se les aplica la intervención, versus a los empleados a los que no (grupo control).

A partir de este diseño, se podrá determinar si la intervención aplicada es efectiva, además de que se podría calcular un retorno de inversión ($ROI = \text{ganancia} / \text{inversión} * 100$) en donde la inversión es el costo de la intervención, y la ganancia sea el coste estimado de la rotación que se calculó en este estudio.

Para esto, en el diseño que se plantea, será necesario cumplir con los siguientes contrastes:

- Grupo experimental (empleados a retener) vs grupo control en pre-test: Permite identificar que no hay una diferencia preexistente entre estos dos grupos antes de aplicar la intervención.
- Grupo control pre-test y post-test: Permite verificar que no hay cambios que puedan ser explicados por otras variables distintas a la intervención.
- Grupo experimental pre-test vs post-test: Permite identificar si hay una reducción en su rotación.
- Grupo experimental vs grupo control en post-test: Permite señalar si hay una diferencia entre aplicar o no la intervención.

En caso de que se cumplan estos contrastes, se podrá concluir que se ha influido de manera positiva en la retención de los empleados y subsecuentemente, en su bienestar.



7. Conclusión

Este trabajo de investigación se enfocó en el estudio de las bajas ocurridas entre los años 2021 y 2022 en una empresa de seguros española. El objetivo principal fue analizar el impacto de los egresos voluntarios, los factores que influyen en la tasa de rotación y las causas específicas que motivaron la salida de los empleados. Además, se planteó como diseñar un modelo predictivo de rotación voluntaria utilizando distintas técnicas estadísticas y comparándolas entre sí.

Para entender el impacto de las bajas voluntarias en la empresa, se estimó que el coste general y el coste de separación, en términos de la inversión en formación, inversión en reclutamiento y selección, horas promedio de selección (time to hire) y un estimado a partir del salario bruto anual.

Los resultados obtenidos revelaron que, en comparación con el contexto post pandemia, la rotación voluntaria en la empresa no es alarmante, aunque ha experimentado un aumento, especialmente en las nuevas contrataciones.

También se observó que la mayoría de las bajas voluntarias correspondían a empleados de la generación Millennials, lo cual representa un riesgo, si se considera que la mayor proporción de la plantilla es más madura, y que los nuevos empleados son más difíciles de retener. A su vez, se señaló en función al desempeño de las personas que egresan, existe una fuga de talento.

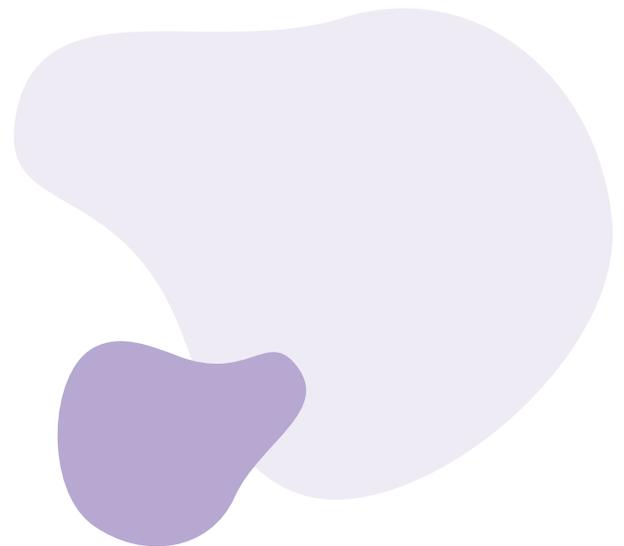
En cuanto a los modelos predictivos, se encontró que el árbol de decisión mostró un mejor desempeño en la predicción de la rotación voluntaria que la regresión logística múltiple. Se identificaron dos perfiles con mayor riesgo de egreso voluntario: empleados con menos de 2 años de antigüedad y empleados con entre 2 y 8 años de antigüedad y entre 31 y 38 años de edad, que no habían recibido un aumento salarial y no contaban con equidad interna.

Sin embargo, se señaló la importancia de tener precaución al utilizar los resultados de los árboles de decisión debido a su tendencia a sobre ajustarse a los datos y la posibilidad de generar discriminación hacia los empleados. Aunque se encontró que las variables utilizadas en los modelos eran estadísticamente significativas para explicar la rotación voluntaria, ninguna de ellas era accionable, lo que dificulta el diseño de intervenciones claras para retener a los empleados.

Se recomienda incluir variables como la percepción de la

cultura y clima laboral, satisfacción con el crecimiento y oportunidades de carrera, remuneración, reconocimiento y feedback, relaciones con los managers y compañeros de trabajo, flexibilidad laboral y variables de personalidad en futuros estudios para poder diseñar intervenciones más efectivas. Se destacó la importancia de basar estas intervenciones en estrategias respaldadas por evidencia empírica, como la personalización de recompensas, ofertas de trabajo realistas, uso de mentores y ofertas de desarrollo atadas a la antigüedad.

Se sugiere replicar este estudio incluyendo las variables mencionadas y diseñar un plan de intervención basado en los resultados obtenidos. Se propone un diseño pre-test post-test con grupo control para evaluar la efectividad de la intervención y calcular un retorno de inversión.



8. Referencias bibliográficas

- Allen, D., Bryant, P. & Vardaman, J. (2010). Retaining Talent: Replacing Misconceptions with Evidence-Based Strategies. **Academy of Management Perspective**, 48-64
- Bossi, F., Di Gruttola, F., Mastrogiorgio, A., D'Arcangelo, S., Lattanzi, N., Malizia, A., & Ricciardi, E. (2022). Estimating Successful Internal Mobility: A Comparison Between Structural Equation Models and Machine Learning Algorithms. **Journal Frontiers in Artificial Intelligence**. 5:848015. doi: 10.3389/frai.2022.848015
- Chatzoudes, D. (2022). Factors Affecting Employee Retention: Proposing an Original Conceptual Framework. **International Journal of Economics and Business Administration**, 10, 49-76
- Graef, M. (2020). Realistic Job Previews. **Umbrella Summaries**. Recuperado de <https://www.qic-wd.org/umbrella/realistic-job-previews>
- Griffeth, R., Hom, P., Gaertner, S. (2000). A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Test, and Research Implications for the Next Millennium. *Journal of Management*. Vol 26, No. 3, 463-488
- Haya, P y García, P. (2018). El Análisis de Redes Organizacionales. **Instituto de Ingeniería del Conocimiento**. 61-7.1
- ICEA (2022). Índices de Personal de las Entidades Aseguradoras: Estadística año 2022. **Investigación Cooperativa entre Entidades Aseguradoras y Fondos de Pensiones**. Recuperado de <https://www.icea.es/es-es/informaciondelseguro/paginas/fichadetexto.aspx?idpublicacion=3241>
- IE University (2022). Informe de Datos Agregados. **HR Club de Benchmarking**. XX Edicion.
- Instituto Santalucía (2023). Edad de jubilación 2023: ¿Cuáles son los cambios?. Recuperado de <https://institutosantalucia.es/edad-de-jubilacion/>
- Moorcraft, B., (2022). Insurers can't hire fast enough to replace the workers that are leaving. **Insurance Business e-magazine**. Recuperado de <https://www.insurancebusinessmag.com/us/news/breaking-news/insurers-cant-hire-fast-enough-to-replace-the-workers-that-are-leaving-420110.aspx>
- ORACLE (2022). The Evolution of Big Data and the Future of the Data Lakehouse. Recuperado de <https://www.oracle.com/big-data/what-is-big-data/>
- Raza, A., Munir, S., Almutairi, M., Younas, F., & Fareed, M. M. S. (2022). Predicting Employee Attrition Using Machine Learning Approaches. **Journal of Business Research**, 143, 1-10. <https://doi.org/10.1016/j.jbusres.2022.08.038>
- Ulhas, S. (2019). A Study of Employee Retention. **Journal of Emerging Technologies and Innovative Research**, 6, 331-337
- Younis, S., Ahsan, L. & Chatteur, F. (2023). An employee retention model using organizational network analysis for voluntary turnover. DOI: 10.1007/s13278-023-01031-w

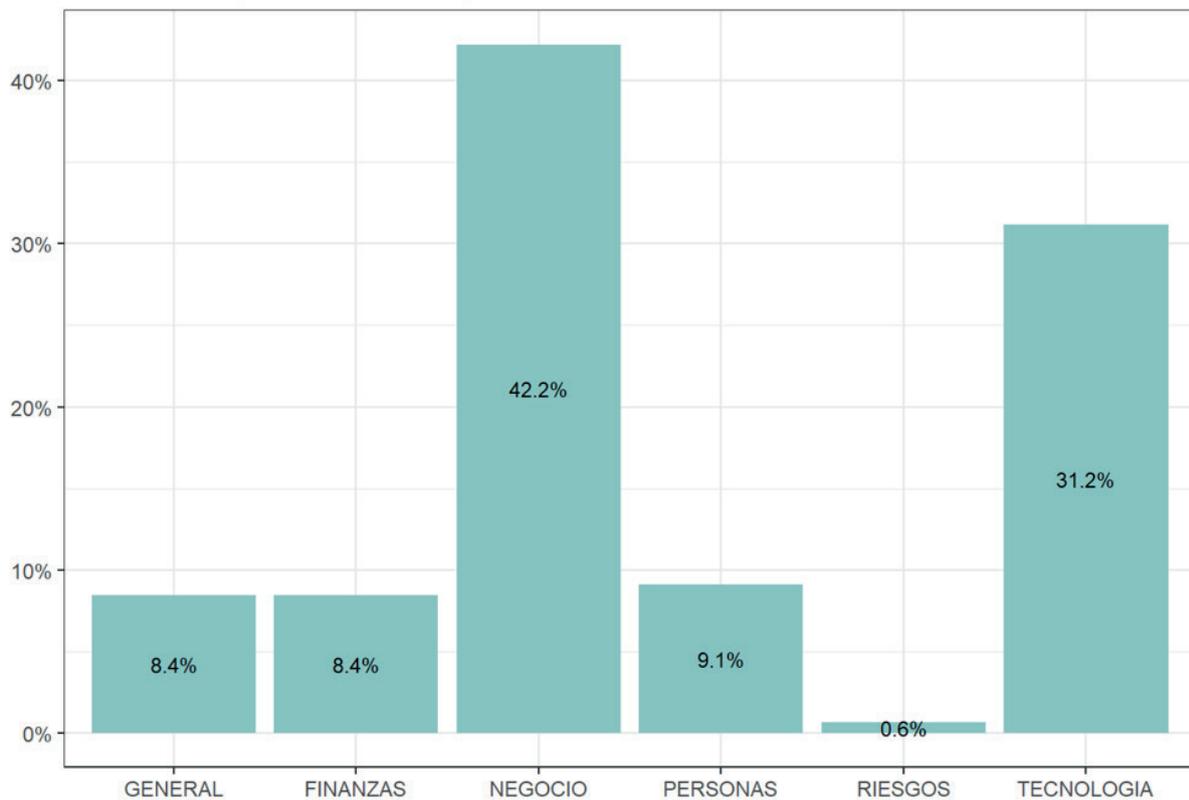
Anexos

Anexo 1. Tabla y gráfico organizativo del Grupo Altas

Estadísticos descriptivos para el Grupo Altas

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	0	0	0	0.1	0	1	0.2	428.60%
EDAD_ANNOS	24	33	41	40	46	58	8.3	20.90%
DESEMPEÑO	1	2	2	2.3	3	4	0.6	27.80%
PERCENTIL_SBA	1	5	7	6.6	9	10	2.5	37.40%
DIF_SALARIAL	0	0	0	1.6	0	49	6.2	400.10%
ABS_AJUSTADO_DIAS	0	0	0	7.6	0	213	27.9	368.50%
ABS_AJUSTADO_CASOS	0	0	0	0.6	0	12	1.7	273.20%
BRADFORD	0	0	0	62.5	0	3387	312.5	500.30%
EQUIDAD_PCT	1	1	1	1	1	2	0.2	20.90%
HORAS_FORMACION	0	18	31	43.8	56	223	42.4	96.70%
FORMACION_INVERSION	0	447	804	1021.7	1328	4447	909.5	89.00%

Distribución por Dirección Organizativa

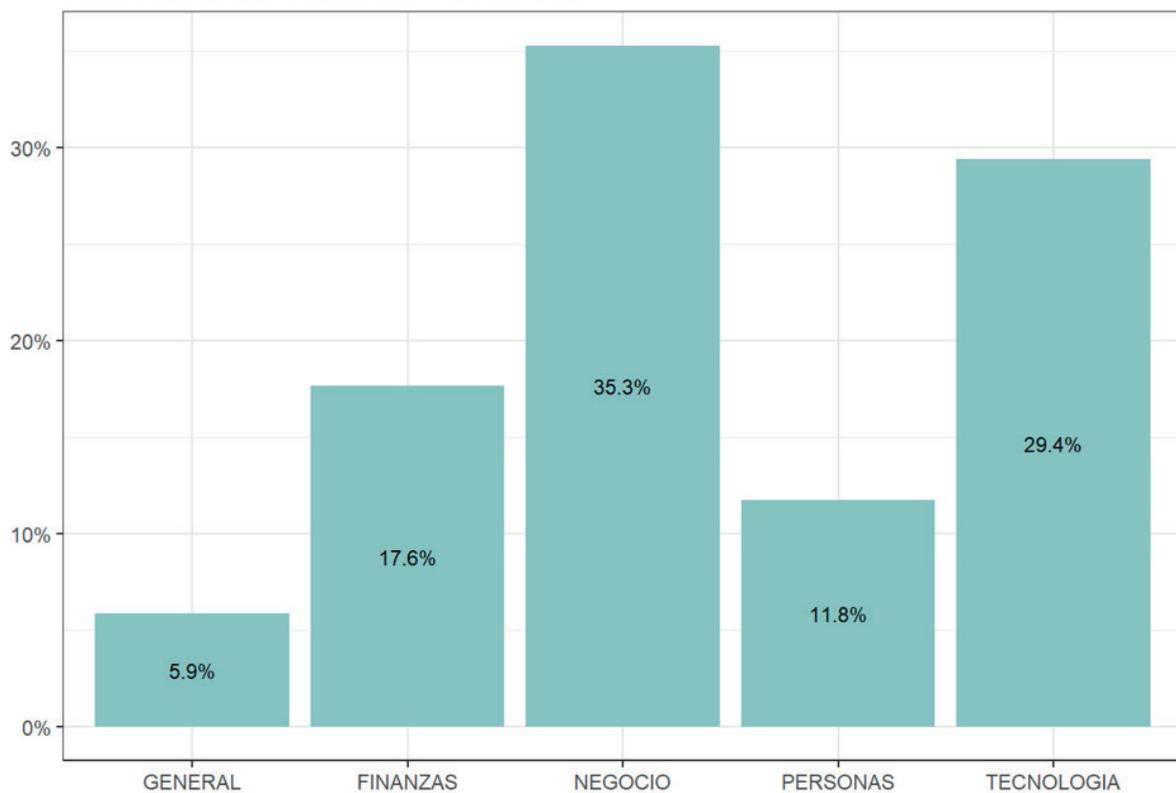


Anexo 2. Tabla y gráfico organizativo del Grupo Al-tas-Bajas

Estadísticos descriptivos para el Grupo Alta-Baja

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	0	0	0	0.4	1	1	0.5	123.20%
EDAD_ANNOS	26	33	37	39	48	50	8.4	21.60%
DESEMPEÑO	2	2	3	2.7	3	3	0.5	18.00%
PERCENTIL_SBA	1	3	6	5.6	8	10	2.9	52.20%
DIF_SALARIAL	0	0	0	0.2	0	4	0.8	412.30%
ABS_AJUSTADO_DIAS	0	0	0	9.1	7	77	20.4	225.20%
ABS_AJUSTADO_CASOS	0	0	0	0.9	1	5	1.6	182.90%
BRADFORD	0	0	0	31.2	7	140	57.2	183.60%
EQUIDAD_PCT	1	1	1	1	1	1	0.2	22.60%
HORAS_FORMACION	0	1	9	19.9	27	81	26.4	132.80%
FORMACION_INVERSION	0	24	129	317.8	300	1476	476.5	149.90%

Distribución por Dirección Organizativa

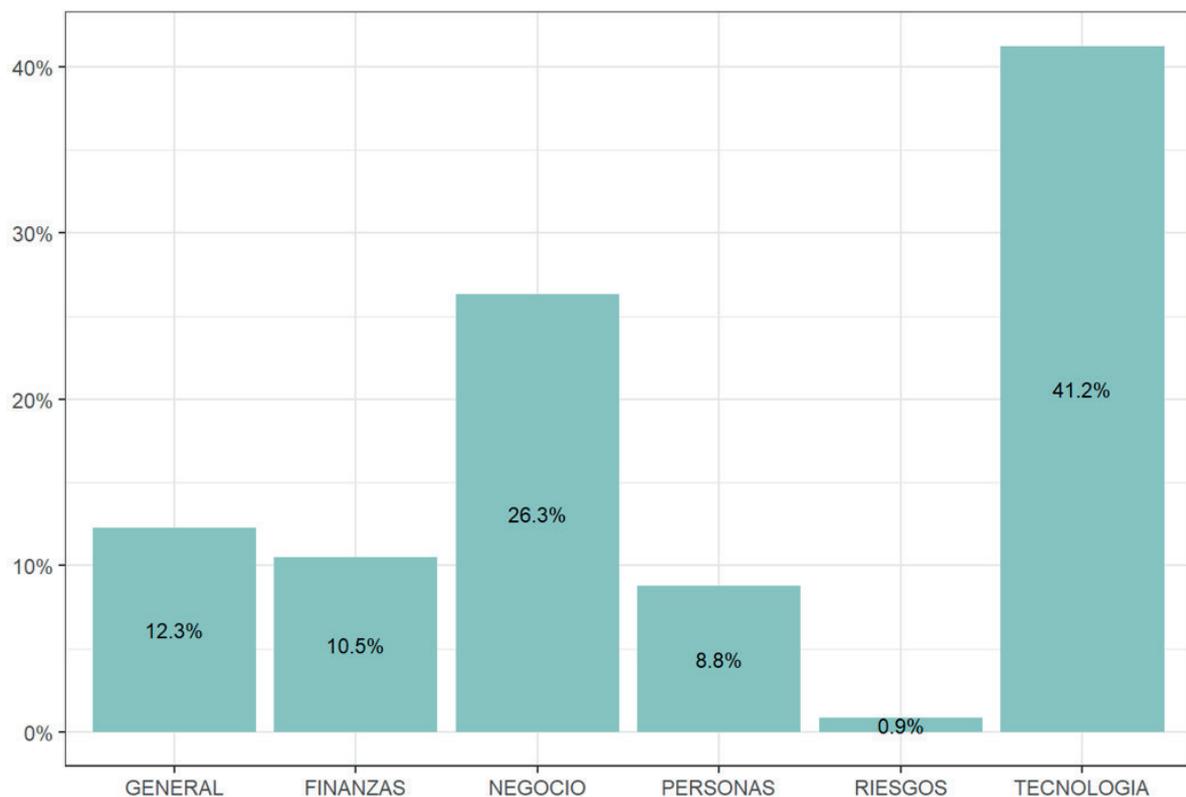


Anexo 3. Tabla y gráfico organizativo del Grupo Bajas

Estadísticos descriptivos para el Grupo Bajas

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	1	2	8	16.9	33	50	16.7	99.10%
EDAD_ANNOS	23	37	48	47.7	59	73	12.6	26.30%
DESEMPEÑO	1	2	2	2.5	3	4	0.8	30.40%
PERCENTIL_SBA	1	3	6	5.7	8	10	2.8	49.40%
DIF_SALARIAL	0	0	0	1.8	3	25	3.8	209.00%
ABS_AJUSTADO_DIAS	0	0	0	46.5	9	709	128.4	276.10%
ABS_AJUSTADO_CASOS	0	0	0	1.1	2	12	2.2	199.80%
BRADFORD	0	0	0	883	16	36579	4177.4	473.10%
EQUIDAD_PCT	0	1	1	1.1	1	2	0.3	24.00%
HORAS_FORMACION	0	0	8	30.4	25	1156	112.2	369.10%
FORMACION_INVERSION	0	0	72	297.5	254	10451	1022.9	343.80%

Distribución por Dirección Organizativa

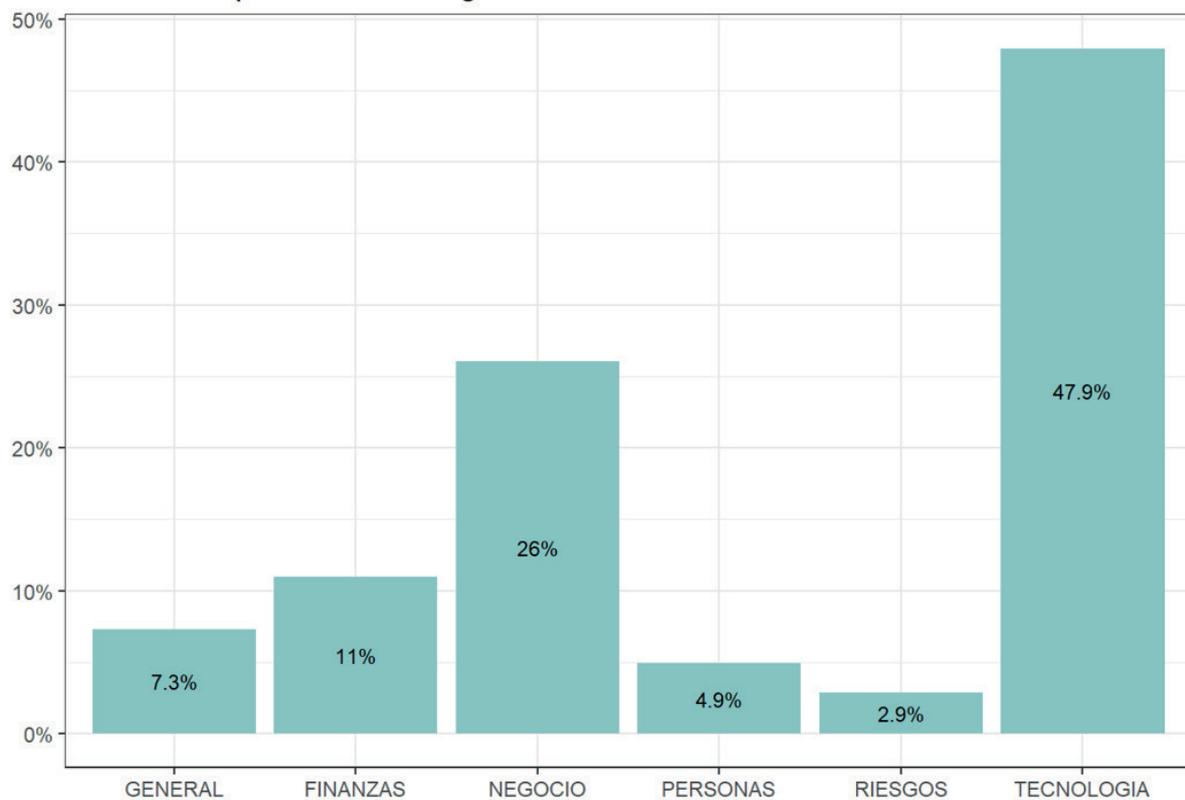


Anexo 4. Tabla y gráfico organizativo del Grupo Control

Estadísticos descriptivos para el Grupo Control

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	2	6	14	15.3	23	43	10.2	66.90%
EDAD_ANNOS	27	41	47	46.4	52	64	7.4	16.00%
DESEMPEÑO	1	2	2	2.4	3	4	0.7	29.30%
PERCENTIL_SBA	1	3	5	5.4	8	10	2.9	53.90%
DIF_SALARIAL	0	2	3	4.5	5	90	6.6	145.00%
ABS_AJUSTADO_DIAS	0	0	0	20	11	374	49.3	247.00%
ABS_AJUSTADO_CASOS	0	0	0	0.8	1	12	1.3	153.30%
BRADFORD	0	0	0	96.8	12	12528	575.1	594.10%
EQUIDAD_PCT	1	1	1	1	1	4	0.3	25.70%
HORAS_FORMACION	0	19	40	51.7	62	1144	73.9	142.90%
FORMACION_INVERSION	0	294	699	889.6	1152	10894	917	103.10%

Distribución por Dirección Organizativa

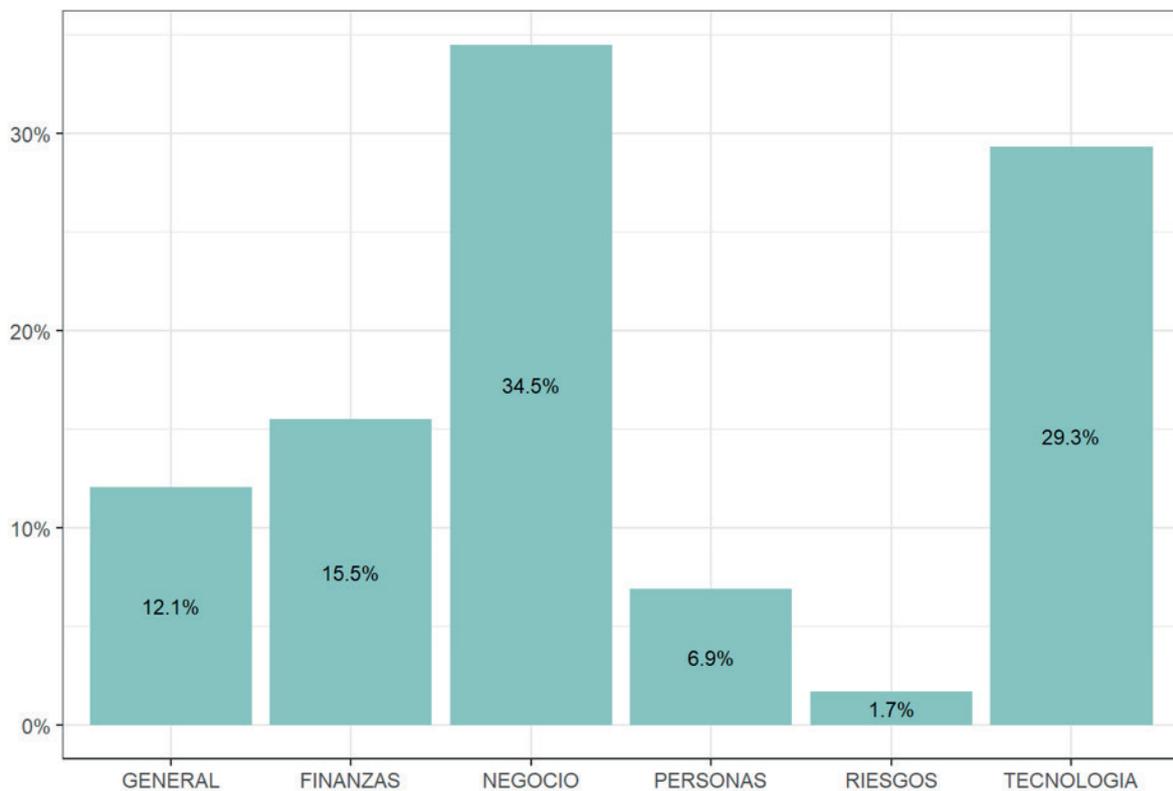


Anexo 5. Tabla y gráfico organizativo de bajas voluntarias.

Estadísticos descriptivos de las bajas voluntarias

Variable	min	Q1	mediana	Media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	0	1	2	3.4	4	26	4.5	132.40%
EDAD_ANNOS	26	31	36	37.3	43	56	8	21.40%
DESEMPEÑO	1	2	3	2.7	3	4	0.7	24.50%
PERCENTIL_SBA	1	3	6	5.8	9	10	2.9	50.60%
DIF_SALARIAL	0	0	0	1.6	0	17	3.7	234.40%
ABS_AJUSTADO_DIAS	0	0	0	12.4	0	183	37.3	301.00%
ABS_AJUSTADO_CASOS	0	0	0	0.6	0	5	1.2	220.30%
BRADFORD	0	0	0	66.7	0	2271	308.7	462.80%
EQUIDAD_PCT	0	1	1	1	1	2	0.2	24.70%
HORAS_FORMACION	0	1	11	41.3	30	1156	152.8	370.30%
FORMACION_INVERSION	0	15	121	400.8	298	10451	1386.3	345.90%

Distribución por Dirección Organizativa

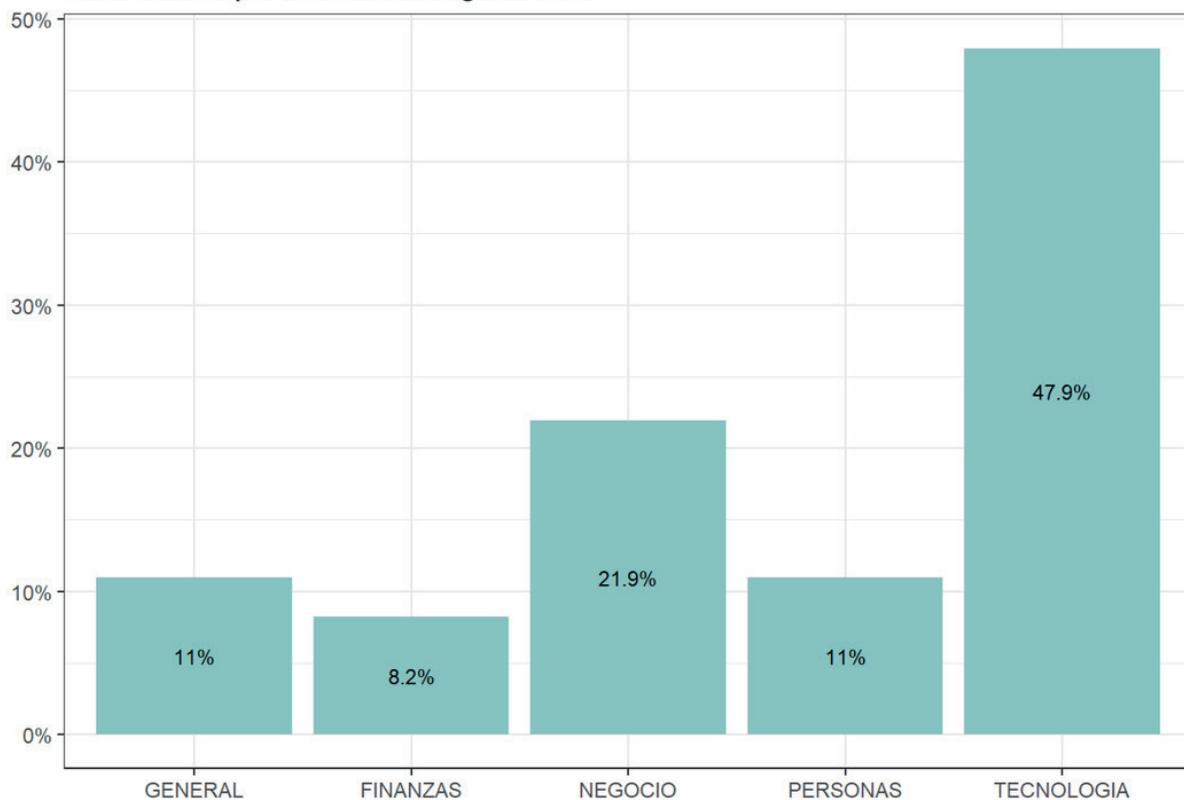


Anexo 6. Tabla y gráfico organizativo de bajas involuntarias.

Estadísticos descriptivos de las bajas involuntarias

Variable	min	Q1	mediana	media	Q3	max	sd	coef_var
ANTIGUEDAD_ANNOS	0	7	27	23.7	37	50	17.1	72.20%
EDAD_ANNOS	23	48	59	54	60	73	10.2	18.80%
DESEMPEÑO	1	2	2	2.3	3	4	0.8	32.90%
PERCENTIL_SBA	1	3	6	5.6	8	10	2.7	49.00%
DIF_SALARIAL	0	0	0	1.6	3	25	3.5	216.80%
ABS_AJUSTADO_DIAS	0	0	0	64.9	20	709	154.7	238.30%
ABS_AJUSTADO_CASOS	0	0	0	1.4	2	12	2.5	174.30%
BRADFORD	0	0	0	1333.1	88	36579	5171.3	387.90%
EQUIDAD_PCT	0	1	1	1.1	1	2	0.2	22.70%
HORAS_FORMACION	0	0	7	19.3	24	199	35.3	182.60%
FORMACION_INVERSION	0	0	70	220.1	217	1994	395.7	179.80%

Distribución por Dirección Organizativa



Anexo 7. Tabla de modelo de regresión logística 1.

```
##
## Call:
## glm(formula = MOVIMIENTO ~ +ANTIGUEDAD_ANNOS + DESEMPENNO + EDAD_ANNOS +
##     DIF_SALARIAL + FORMACION_INVERSION_ESTIMADA, family = binomial("logit
##     "),
##     data = DF_TABLA_MODELO)
##
## Coefficients:
##
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.755e+00  1.339e+00   2.057 0.039652 *
## ANTIGUEDAD_ANNOS -1.714e-01  4.630e-02  -3.702 0.000214 ***
## DESEMPENNO       2.935e-01  2.418e-01   1.213 0.224949
## EDAD_ANNOS      -9.591e-02  3.022e-02  -3.174 0.001503 **
## DIF_SALARIAL    -1.630e+01  6.137e+00  -2.656 0.007899 **
## FORMACION_INVERSION_ESTIMADA -4.883e-04  3.141e-04  -1.554 0.120079
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 301.79  on 659  degrees of freedom
## Residual deviance: 209.72  on 654  degrees of freedom
## (28 observations deleted due to missingness)
## AIC: 221.72
##
## Number of Fisher Scoring iterations: 8
```

Anexo 8. Tabla del modelo de regresión logística 2.

Prueba t-student para empleados con rol “Técnico”

variable	conf.low	conf.high	statistic	p.value	sig
ANTIGUEDAD_ANNOS	8.3475365	11.5208863	12.4011178	0.0000	*
EDAD_ANNOS	6.2379334	11.2158997	7.0187112	0.0000	*
DESEMPEÑO	-0.6272211	-0.1213509	-2.9934144	0.0048	*
SEXO	-0.0812555	0.2093204	0.8810451	0.3817	
DIRECCION_SL	-0.7175198	0.1552773	-1.2883965	0.2025	
SBA_PROMEDIO	-3607.9599592	3570.6772494	-0.0104154	0.9917	
PERCENTIL_SBA	-0.8039517	0.9333675	0.1493119	0.8819	
DIF_SALARIAL	0.0144116	0.0436322	3.9363386	0.0001	*
ABS_AJUSTADO_DIAS	6.8764152	23.2451993	3.6362731	0.0004	*
ABS_AJUSTADO_CASOS	0.0039337	0.7175507	2.0202994	0.0476	*
BRADFORD	20.6601792	97.4602815	3.0284405	0.0027	*
HORAS_FORMACION	-44.2888943	55.6919919	0.2294321	0.8195	
FORMACION_INVERSION	11.0229917	929.1223452	2.0570712	0.0449	*

Anexo 9. Modelo de regresión logística 2.

```
##
## Call:
## glm(formula = MOVIMIENTO ~ +ANTIGUEDAD_ANNOS + DESEMPENNO + EDAD_ANNOS +
##     DIF_SALARIAL + FORMACION_INVERSION_ESTIMADA + ABS_AJUSTADO_CASOS,
##     family = binomial("logit"), data = DF_TABLA_MODELO_TECNICO)
##
## Coefficients:
##
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)      2.534e+00  1.666e+00   1.521 0.128353
## ANTIGUEDAD_ANNOS -1.949e-01  5.834e-02  -3.340 0.000836 ***
## DESEMPENNO       5.063e-01  3.056e-01   1.657 0.097506 .
## EDAD_ANNOS      -9.021e-02  3.727e-02  -2.420 0.015515 *
## DIF_SALARIAL    -1.261e+01  6.154e+00  -2.050 0.040379 *
## FORMACION_INVERSION_ESTIMADA -3.509e-04  3.057e-04  -1.148 0.251028
## ABS_AJUSTADO_CASOS -3.633e-01  1.971e-01  -1.844 0.065202 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 217.84  on 345  degrees of freedom
## Residual deviance: 146.03  on 339  degrees of freedom
## (16 observations deleted due to missingness)
## AIC: 160.03
##
## Number of Fisher Scoring iterations: 7
```

.04



instituto de ingeniería
del conocimiento

INSTITUTO DE INGENIERÍA DEL CONOCIMIENTO (IIC)

PIONEROS EN INTELIGENCIA ARTIFICIAL DESDE 1989

TECNOLOGÍA Y ANÁLISIS DE DATOS AL SERVICIO DE RR. HH.



HR ANALYTICS

El análisis de los datos de RR. HH. permite obtener información de valor para una mejor gestión del talento. El IIC aplica **analítica descriptiva y predictiva** para optimizar procesos de selección, predecir el absentismo o la rotación e identificar a los profesionales con más potencial, entre otros proyectos.



EVALUACIÓN DE COMPETENCIAS

Dentro de la **plataforma online eValue**, desarrollamos pruebas objetivas y fiables para evaluar las competencias transversales, el nivel de inglés o las motivaciones de candidatos y empleados. Además de tomar mejores decisiones, se obtienen **datos de calidad** para analizar, por ejemplo, sus necesidades de formación.



ANÁLISIS DE REDES ORGANIZACIONALES

Los proyectos AROS permiten analizar las relaciones de trabajo y las interacciones entre los profesionales. Representadas visualmente en un grafo, se pueden identificar **redes informales, referentes ocultos o cuellos de botella** en la organización, para emprender acciones de mejora.

El Instituto de Ingeniería del Conocimiento (IIC) es un centro privado de I+D+i fundado en 1989 y pionero en Inteligencia Artificial. Llevamos más de 30 años trabajando en análisis Big Data y diferentes técnicas de **Machine Learning** y **Procesamiento del Lenguaje Natural (PLN)**, entre otras.

Nuestra apuesta de valor es el desarrollo de algoritmos y soluciones tecnológicas adaptadas a diferentes sectores y basadas siempre en la investigación aplicada. Como centro sin ánimo de lucro, comercializamos los productos y servicios fruto de nuestro trabajo para reinvertir los beneficios en más investigación e innovación.

En el IIC estamos especializados en tecnologías de analítica avanzada y en el desarrollo de **modelos de Inteligencia Artificial** para las áreas de Banca, RR. HH., Energía, Salud, Seguros e Inteligencia de Cliente.

Además, destacamos por ser expertos en la preparación y tratamiento de todo tipo de datos para su análisis y por la metodología que utilizamos para acompañar a nuestros clientes durante todo el proceso.



ENTORNO
BANCARIO



ENTORNO
RR.HH.



ENTORNO
SEGUROS



ENTORNO
ENERGÍA



ENTORNO
SALUD



ENTORNO
INTELIGENCIA
DE CLIENTE



NUESTROS ASOCIADOS:



Instituto de Ingeniería del Conocimiento

C/ Francisco Tomás y Valiente, 11 EPS,
Edificio B, 5ª planta UAM Cantoblanco.
28049 Madrid

<http://www.iic.uam.es/empleo-iic/>

contacto.iic@iic.uam.es

(+34) 91 497 2323



www.linkedin.com/company/instituto-de-ingenier-a-del-conocimiento---iic



www.twitter.com/IIConocimiento



www.youtube.com/IIConocimiento



INNOVADATA



iic
instituto
de ingeniería
del conocimiento



www.iic.uam.es

Instituto de Ingeniería del Conocimiento

C/ Francisco Tomás y Valiente, nº 11
Escuela Politécnica Superior (EPS),

Edificio B, 5ª planta
Universidad Autónoma de Madrid (UAM).

28049 Cantoblanco, Madrid

T. (+34) 91 497 23 23