

Datatón de Análisis de Recursos Humanos (HR Analytics)

Jorge de Jesús Hernández*
Gabriel Solano de Sixto*
Emmanuel de Atocha Zamora Hoil*
jorge.hernandez.2@train.ia.center
gabriel.solano@train.ia.center
emmanuel.atocha@train.ia.center
Centro de Inteligencia Artificial
Cd. Juárez, Chihuahua, Mexico

ABSTRACT

Cite as :

Jorge de Jesús Hernández, Gabriel Solano de Sixto, and Emmanuel de Atocha Zamora Hoil. 2021. Datatón de Análisis de Recursos Humanos (HR Analytics). En *Proceedings of* . Cinvestav Unidad Tamaulipas, Ciudad Victoria, Mexico, 9 páginas.

1 INTRODUCCIÓN

En un mundo empresarial cada vez más dinámico y competitivo, la gestión efectiva del talento se ha convertido en un factor determinante para el éxito de las organizaciones. En este contexto, nos embarcamos en la misión de descubrir *insights* significativos y desarrollar modelos clasificativos que no solo optimicen la gestión de talento, sino que también impacten positivamente en la satisfacción laboral, fomenten la diversidad y impulsen la eficiencia operativa. A través de nuestro enfoque innovador, buscamos ofrecer soluciones que no solo se adapten a las necesidades cambiantes del mercado, sino que también impulsen el crecimiento sostenible de las organizaciones, permitiéndoles alcanzar su máximo potencial. En este camino hacia la excelencia empresarial, nos comprometemos a proporcionar herramientas analíticas avanzadas y perspicaces que guíen a las organizaciones hacia un futuro donde la gestión del talento se convierte en una ventaja estratégica innegable.

2 TRABAJO RELACIONADO

El *African Centre for Data Science & Analytics Ltd.* [1] analiza cómo el análisis de recursos humanos está revolucionando la gestión del talento y estableciendo nuevos estándares para la excelencia organizacional. Cubre temas como la mejora de la contratación y la elección de candidatos, el análisis predictivo para la retención y la planificación de la fuerza laboral y la gestión de la sucesión.

En el entorno empresarial dinámico y cambiante de hoy, las empresas son cada vez más conscientes de la contribución esencial de los recursos humanos (RRHH) para lograr el éxito. Las responsabilidades convencionales de RRHH están experimentando un cambio significativo debido a la introducción del análisis de RRHH. Estos conocimientos derivados de los datos están causando un cambio revolucionario en las estrategias para manejar el talento, ayudar a las empresas en las elecciones bien informadas, refinar la preparación

de la fuerza laboral y aumentar la participación de los empleados. Este artículo profundiza en el ámbito del análisis de RRHH, investigando su impacto en la forma en que las organizaciones supervisan su recurso más invaluable: su personal.

El análisis predictivo se ha convertido en una potente herramienta en el ámbito de la gestión del talento, especialmente para abordar la cuestión crucial de la retención de empleados. Al aprovechar el potencial de los datos y los algoritmos avanzados, las organizaciones ahora pueden pronosticar qué empleados tienen más probabilidades de salir y tomar medidas preventivas para retenerlos. Este enfoque revolucionario supera las estrategias reactivas convencionales, lo que permite a las empresas descubrir tendencias, asociaciones y variables que influyen en el desgaste. A través del análisis predictivo, los equipos de recursos humanos pueden identificar indicadores tempranos, como cambios en el desempeño, la participación o el avance profesional, y desarrollar intervenciones precisas para mejorar la retención. Esto no solo conserva valiosos recursos que de otra manera se invertirían en la contratación y capacitación, sino que también nutre una cultura de apoyo proactivo de los empleados, cultivando niveles elevados de compromiso y lealtad.

Engadly[2] explica cómo el software de gestión del talento proporciona a los líderes información práctica para la toma de decisiones estratégicas en la adquisición, desarrollo y retención de talento. Cubre temas como la identificación de empleados de alto potencial, el seguimiento de la participación de los empleados y la medición de la eficacia de los programas de capacitación.

Al consolidar grandes volúmenes de datos de los empleados, una solución de gestión del talento revela patrones, tendencias e ideas que de otro modo podrían permanecer ocultas. Estos conocimientos son la base sobre la que se construyen estrategias eficaces de gestión del talento.

Osch *et al.*[7], nos especifica que dentro de la literatura sobre la diversidad en los grupos de trabajo, el enfoque principal se ha centrado en las consecuencias de la diversidad estructural para la comunicación, el conflicto o la integración social dentro del grupo. Sin embargo, dado que los grupos de trabajo se centran cada vez más en conectar a los miembros de diferentes ubicaciones geográficas y funciones, y la creciente necesidad de que los equipos se involucren en el traspaso de fronteras, las investigaciones recientes han puesto de relieve que la diversidad estructural desempeña un papel más importante a la hora de exponer a los miembros del grupo a diferentes fuentes de información, conocimientos y reacciones.

Concretamente, la literatura sobre diversidad estructural distingue cuatro fuentes de diversidad estructural, a saber:

- Ubicaciones geográficas: dado que los "ojos y oídos" de los miembros del grupo se encuentran en entornos diferentes, pueden acceder a una mayor variedad de información y conocimiento.
- Asignaciones funcionales: los miembros del grupo pertenecen a diferentes dominios funcionales, lo que les permite acceder a diversas redes sociales.
- Gestores de informes: la diversidad en las estructuras jerárquicas de presentación de informes de los miembros del grupo también aumenta la oportunidad de acceder a distintas redes sociales.
- Unidades de negocio: similares a las asignaciones funcionales, los miembros del grupo que trabajan en diferentes unidades de negocio pueden acceder a conocimientos diversos y únicos fuera del grupo.

Además, los grupos de trabajo con altos niveles internos de diversidad geográfica y jerárquica tendrán una gran variedad de relaciones externas (al equipo), lo que requerirá un mayor número de intentos de coordinación. Por lo tanto, anticipan que la diversidad geográfica y jerárquica tendrá un efecto positivo en la coordinación. Sin embargo, como se destacó originalmente en, los grupos de trabajo con alta diversidad funcional interna tendrán que centrarse en superar y coordinar a los miembros del grupo a fin de establecer un entendimiento compartido necesario para el éxito del proyecto antes de iniciar la coordinación externa. Por lo tanto, anticipamos un efecto negativo de la diversidad funcional en la coordinación.

Finalmente, dado que las actividades de búsqueda de información ocurren principalmente entre individuos con el mismo fondo funcional pero que residen en otras partes de la organización, anticipamos un efecto negativo de diversidad funcional y un efecto positivo de diversidad geográfica. No se prevé ningún efecto de la diversidad jerárquica. Por lo tanto:

- (1) La diversidad geográfica de un grupo de trabajo tiene un efecto positivo en el número de actividades de coordinación y búsqueda de información realizadas por un grupo de trabajo.
- (2) La diversidad funcional de un grupo de trabajo tiene un efecto negativo en el número de actividades de coordinación y en el número de actividades de búsqueda de información, pero no en las actividades de representación.
- (3) La diversidad jerárquica de un grupo de trabajo tiene un efecto positivo en el número de actividades de representación y coordinación realizadas por el grupo, y ningún efecto en las actividades de búsqueda de información.

Dawal y Taha [4] hicieron un estudio en dos empresas manufactureras de Malasia, donde se examinaron los efectos de los factores de diseño del trabajo que hacen hincapié en las características del trabajo y los factores ambientales y de organización en la satisfacción del trabajo. Participaron 170 hombres con una edad media de 26,8 años y una experiencia laboral media de 6,5 años (ver Figura 1. Las posibles aplicaciones de estos métodos se desarrollaron, a saber, el diseño de cuestionarios, la recopilación de datos y el análisis estadístico para diagnosticar los factores actuales de diseño del trabajo industrial que afectan la satisfacción de los trabajadores. Se determinaron factores de diseño de trabajo significativos mediante análisis de regresión. El modelo destaca los atributos más significativos en ambas industrias manufactureras que afectan la

Age (Years)	Respondents' Percentage		Work Experience (Years)	Respondents' Percentage	
	Auto 1	Auto 2		Auto 1	Auto 2
18 – 20	3	21	1 – 2	5	48
21 – 25	14	70	3 – 4	5	42
26 – 30	50	9	5 – 6	20	4
31 – 35	26	0	7 – 8	44	6
36 – 40	7	0	9 & above	26	0

(a) Edad

(b) Tiempo en la compañía

Figure 1: Edad y tiempo en la compañía [4]

satisfacción de los trabajadores son la variedad de habilidades, el ruido, la rotación laboral y el método de trabajo.

Su trabajo hace énfasis en la satisfacción en el trabajo es uno de los principales criterios para establecer un entorno organizacional saludable en una organización. Además, los factores relacionados con la satisfacción en el trabajo son relevantes para la prevención de la frustración de los empleados y la baja satisfacción en el trabajo porque los empleados trabajan más duro y se desempeñan mejor siempre que estén satisfechos con su trabajo.

Varios autores han definido las diferencias de motivación como el proceso psicológico que da al propósito y la dirección del comportamiento una predisposición a comportarse de manera intencional para lograr necesidades específicas y no satisfechas, un impulso interno para satisfacer una necesidad insatisfecha y la voluntad de lograr. ¿Por qué necesitamos empleados satisfechos y motivados? La respuesta es la supervivencia. Los empleados motivados y satisfechos ayudan a las organizaciones a sobrevivir y a ser más productivos a cambio.

Sus resultados fueron De los 170 participantes varones entrevistados, el 80% poseía el Certificado de Educación de Malasia (SPM), que es equivalente a los niveles de Cambridge "O", en ambas empresas, mientras que otros poseían certificados de SPM con otros certificados de habilidades. El 69% de los participantes en Auto 1 estaban casados y el 31% eran solteros. Por otro lado, el 87% de los participantes en Auto2 eran solteros y el 13% estaban casados. Los sujetos tenían entre 18 y 40 años con una edad media de 26,8 5,3 años y una experiencia laboral media de 6,5 4,9 años.

El factor edad se repartía normalmente, pero la experiencia laboral no. La experiencia laboral para Auto 1 fue sesgada negativamente mientras que la experiencia laboral para Auto2 fue sesgada positivamente. Las respuestas indicaron que el 83% de los trabajadores de Auto 1 tenían 26 años o más, mientras que el 90% de los trabajadores de Auto 2 tenían menos de 26 años. Solo el 17% de los trabajadores de Auto 1 tenían 25 años o menos, mientras que el 9% de los trabajadores de Auto 2 tenían 26 años o más. En cuanto a la experiencia laboral, el 90% de los trabajadores de Auto 1 había trabajado durante más de cinco años. Otro 10% tenía una experiencia laboral inferior a cinco años. En el caso de Auto 2, el 90% de los trabajadores tenía una experiencia laboral de 4 años o menos. Solo el 10% tenía una experiencia laboral de entre cinco y ocho años. Los encuestados en Auto 2 eran más jóvenes y menos experimentados que los encuestados en Auto 1.

El trabajo de Galván [5], se centra en la alienación laboral, que es un fenómeno complejo que hace referencia a una condición cognitiva y social en la que una persona se desconecta y se separa de suyo interno dentro de su entorno laboral. Al tener en cuenta que dicha problemática ha sido examinada desde diversas perspectivas disciplinarias, las investigaciones sobre esta temática son limitadas. En particular, los trabajadores que se convierten en víctimas de la alienación les otorgan más importancia y atención a las recompensas externas (el salario, por ejemplo) que a su desempeño laboral, por lo que es probable que estos renuncien a su trabajo en un corto plazo.

Identifica dos variables notorias que han sido asociadas con la temática ya mencionada: la satisfacción laboral y el empoderamiento psicológico individual. Esto proporcionó los fundamentos necesarios para la elaboración del objetivo de la investigación, el cual fue el de definir, valorar e interpretar la asociación de la alienación laboral con la satisfacción laboral y el empoderamiento psicológico individual.

Los hallazgos tienen fuertes implicaciones organizacionales en el sector automotriz de Ciudad Juárez, ya que a un empleado alienado no le importa personalmente la participación en los procesos de trabajo, su objetivo primordial es únicamente el de ganar dinero. Por lo que tal situación debe ser considerada por las empresas que integran al sector automotriz en Ciudad Juárez, ya que es predominante la utilización de estrategias enfocadas en recompensar (monetariamente) el desempeño de los trabajadores sin contemplar sus necesidades de desarrollo psicológico. Es a través del otorgamiento de esas recompensas extrínsecas lo que permite que las organizaciones impidan que sus trabajadores se sientan insatisfechos, haciéndoles caer dentro de un estado neutro dado que tampoco están satisfechos. De esa forma se evita la presencia de disgustos en los trabajadores.

Este estudio se desarrolló en Ciudad Juárez, situada en la frontera norte de México, el cual es un polo de desarrollo económico del país; su concentración industrial es de 332 empresas instaladas, las cuales se dedican a la manufactura de diferentes productos, teniendo una ocupación de 298,262 personas (INDEX, 2018). Considerando que las empresas manufactureras que pertenecen al sector automotriz tienen un gran impacto en la economía de Ciudad Juárez, se optó por desarrollar una investigación dentro de dicho contexto, ya que se pretende aportar más evidencia sobre las variables antes descritas, pues el problema que se aborda es el de definir y a la vez valorar cómo se asocia la alienación laboral con la satisfacción laboral y con el empoderamiento psicológico a nivel individual de los trabajadores que laboran dentro de los mandos medios y superiores de las empresas manufactureras del sector automotriz, en Ciudad Juárez, Chihuahua, México

La migración es un factor que estudia Boisen [3], y sobre todo de la proveniente de Veracruz. Durante la última década, la población de los estados del sureste de México se ha incorporado de manera creciente tanto a los flujos migratorios internos como a los flujos internacionales. Uno de los nuevos destinos principales de la migración interna es la Industria Maquiladora de Exportación (IME) situada en la frontera norte de la república. La IME ubicada en dicha frontera ha incorporado un elevado número de operadores provenientes de los estados del sur de México, como Veracruz y Tabasco, así como de otros estados del sureste del país, que antes no formaban parte del panorama laboral y cultural de las ciudades

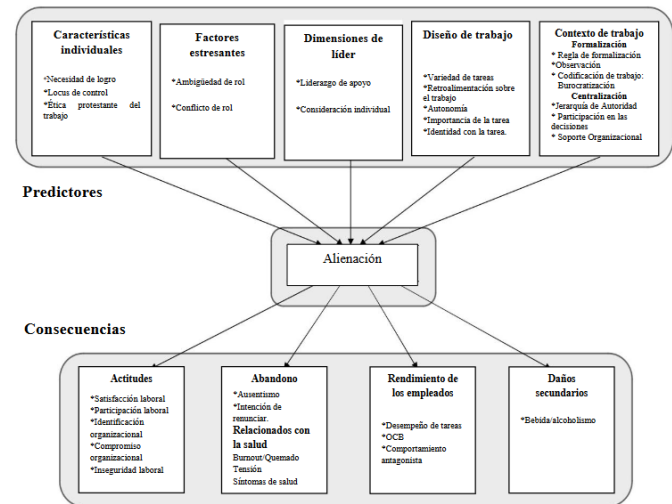


Figure 2: Modelo teórico de la alienación laboral: Predictores y consecuencias [5]

fronterizas (lo que se demuestra en el análisis exploratorio de los datos provistos con respecto a lugar de origen de los trabajadores).

A partir de la década de los años sesenta, con el crecimiento de los centros urbanos y de la industria petroquímica, comenzó a cambiar la dinámica ocupacional de Oteapan (donde se centra el estudio) y de algunos otros pueblos, hacia una mayor vinculación e inserción en el mercado laboral regional. Este cambio estuvo estrechamente vinculado al crecimiento del sector de la industria, de la construcción y del comercio, asociado al auge petrolero y al crecimiento urbano. Se desarrolló de esta forma una marcada tendencia hacia el abandono progresivo de la agricultura, acompañado por una creciente urbanización del tipo de empleo en la comunidad.

Posteriormente, entre los años 1988 y 1993, sobrevino una importante reestructuración de la empresa paraestatal Petróleos Mexicanos (PEMEX) que condujo al despido de la mitad de los empleados de dicha empresa, por lo que alrededor de 120 mil personas perdieron su trabajo (Estrada, 1996: 61). Los despidos de PEMEX afectaron severamente la economía de los hogares en los pequeños pueblos que se habían vuelto crecientemente dependientes de los ingresos petroleros. A ello se tiene que añadir el cierre simultáneo, en 1993, del paraestatal Azufrero Panamericano, ubicado en el vecino pueblo de Jáltipan, donde también trabajaba mucha gente de Oteapan.

A esta situación adversa sobrevino, en diciembre de 1994, la crisis económica nacional, caracterizada por una súbita salida de capital, así como por una acelerada inflación del peso mexicano que redujo su valor adquisitivo. Esta circunstancia agravó aún más la situación económica de las familias que habían sido golpeadas previamente por los despidos de la paraestatal, sin posibilidad de encontrar alternativas en un mercado laboral local, que se encontraba severamente contraído. Esta situación se expresaba en 1995 en una tasa de desempleo de 9.8% en la ciudad de Coatzacoalcos, el más alto del país (INEGI, 2005).

A raíz de esta situación adversa, a principios de 1995, unos pocos hombres salieron de Oteapan rumbo a la frontera norte alentados

por las noticias que ahí había abundante trabajo. Con ello comenzó la migración a Ciudad Juárez. Las condiciones precarias de vida en el pueblo se mantuvieron durante los siguientes años, con el 85% de la población económicamente activa de Oteapan recibiendo menos de 2 salarios mínimos (INEGI, 2008).

Los primeros migrantes de Oteapan que salieron con rumbo a Ciudad Juárez en el año 1995 eran hombres solos. A pesar de las dificultades, estos pocos migrantes lograron insertarse en la IME y comenzaron pronto a mandar dinero a sus familiares en el pueblo. A través de las remesas, de las visitas y de la información que llegaba del norte, poco a poco se difundió en el pueblo la idea de que en la frontera había abundante trabajo y buenos ingresos.

Después de un tiempo de radicar en Ciudad Juárez, los migrantes comenzaron a traer sus esposas y familias nucleares. Luego trajeron a parientes consanguíneos y afines de primer grado colateral y descendente³. De este modo, poco a poco, se fueron trasladando las densas redes sociales del pueblo a la frontera, donde se reprodujeron y extendieron en función de la migración. Actualmente, dichas redes sociales, basadas en el parentesco y en la vecindad, se reproducen tanto en el lugar de origen, como en el lugar de destino, así como entre estos dos espacios.

En Ciudad Juárez, a dos mil quinientos kilómetros de distancia de su lugar natal, los migrantes veracruzanos viven en un número limitado de colonias en la zona surponiente de la ciudad, en las cercanías de los parques industriales. En sus formas de inserción en esta ciudad fronteriza, la gente de Oteapan reproduce importantes elementos del conjunto residencial tradicional. Aquí, lo más común es que una familia nuclear rente una casa conjuntamente con las familias nucleares de los hermanos, y a veces las de las hermanas. El resultado final es una unidad residencial donde conviven bajo un mismo techo las unidades domésticas de los hermanos adultos casados, bajo una lógica de organización muy parecida a la que rige el conjunto residencial en Oteapan.

Y por último, cabe destacar la discriminación a los migrantes. De acuerdo al censo del 2000, la población en el municipio de Juárez era de 1 218 817 habitantes, siendo prácticamente la mitad hombres y la mitad mujeres. Más exactamente, el 50.28% mujeres y el 49.72% hombres. La población de 5 años y más originaria del municipio de Juárez ("no migrante municipal") era de 862, 890 habitantes; mientras que la población originaria de otros estados ("migrante estatal" con "residencia actual en otra entidad"), era de 106, 922 (INEGI, 2003). A los 106, 922 migrantes con residencia actual en otros estados, nosotros sumamos, para fines estadísticos, los 12,041 migrantes de otros municipios del estado de Chihuahua; así como los 10, 004 migrantes internacionales que habitaban esta ciudad. Es decir, un total de 128, 967 habitantes de esta ciudad no nacieron en ella, por lo que podemos sostener que poco más del 10% del conjunto de sus habitantes eran migrantes, es decir, provenían de otros lugares del estado, de otros estados del país, o de otros países [6].

Así, un elemento importante que hace a Ciudad Juárez una entidad con características propias frente a otras en el estado, es que ella vive cotidianamente el arribo de migrantes, los cuales, si bien "sólo" son un poco más del 10% del total de sus habitantes, significan un referente importante en la ciudad. En efecto, en los medios de comunicación se alude cotidianamente a esos migrantes; los gobernantes en turno también lo hacen; y el imaginario ciudadano está poblado

por estos "fuereños", para bien o para mal. Para los originarios de la ciudad, Ciudad Juárez es un lugar de "oportunidades", abierto al migrante. Pero paradójicamente, al lado de esta actitud abierta y amistosa, puede observarse un conjunto de prejuicios por parte de los originarios de la entidad hacia los que llegan a este lugar efectivamente en busca de trabajo y de condiciones de vida no tan precarias como las de sus lugares de origen. Son prejuicios hacia los migrantes que, sin duda, contienen una fuerte connotación regionalista, clasista y hasta sexista, pues se centra hacia las obreras que trabajan en las maquiladoras, y que incluye a todos los "fuereños", incluidos los hombres, por supuesto [6].

3 METODOLOGÍA

La metodología para resolver la satisfacción laboral con respecto a la variable de lugar de origen es:

- **Análisis de Diversidad en el Lugar de Trabajo:** Evaluar el estado actual de la diversidad en la organización y proponer estrategias para mejorarla.
- **Instrumentos de Recopilación de Datos:** Utilizaremos encuestas detalladas y entrevistas estructuradas para recopilar información sobre la satisfacción laboral, identificando aspectos específicos como el equilibrio entre vida laboral y personal, las oportunidades de desarrollo profesional y la calidad del entorno laboral.
- **Análisis de Datos:** Implementaremos técnicas estadísticas avanzadas para analizar la relación entre la ubicación geográfica y la satisfacción laboral con el tiempo de residencia. Esto incluirá la identificación de patrones, la elaboración de perfiles de satisfacción y la evaluación de posibles correlaciones.

Resultados Esperados:

- **Diferencias Regionales:** Se espera identificar diferencias significativas en la satisfacción laboral entre empleados de distintas regiones geográficas.
- **Factores Clave:** La investigación buscará identificar los factores específicos relacionados con la ubicación que más influyen en la satisfacción laboral, como el acceso a servicios, el clima laboral regional y las oportunidades de desarrollo profesional.
- **Implicaciones para la Gestión de Talentos:** Con los resultados, se proporcionarán recomendaciones prácticas para adaptar estrategias de gestión de talento según la ubicación geográfica, con el objetivo de mejorar la satisfacción y retención de empleados.

4 EXPERIMENTACIÓN

Se hizo un análisis exploratorio de los datos:

4.1 Recursos

Se trabajó sobre el dataset *bosch_aic_datathon*, los cuales tiene las siguientes características (después del preprocesamiento y limpieza de datos):

Para atacar la problemática de los motivos de renuncia, se exploraron los datos y los principales son (ver Figura18) :

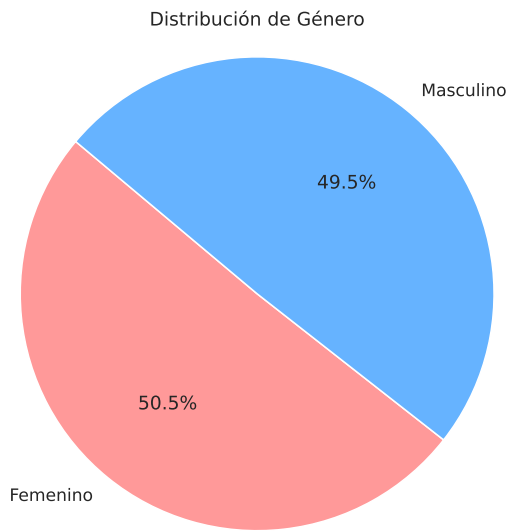


Figure 3: Género

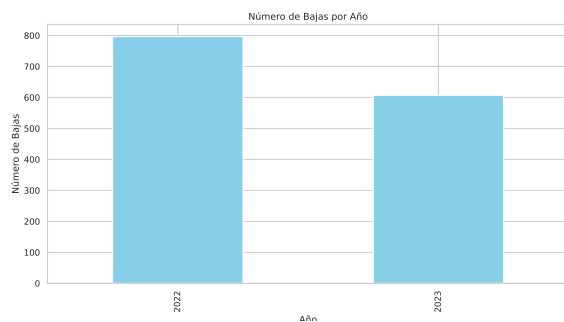


Figure 4: Bajas por año

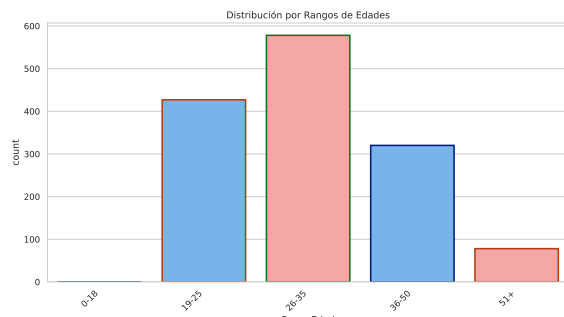


Figure 5: Rango de edades

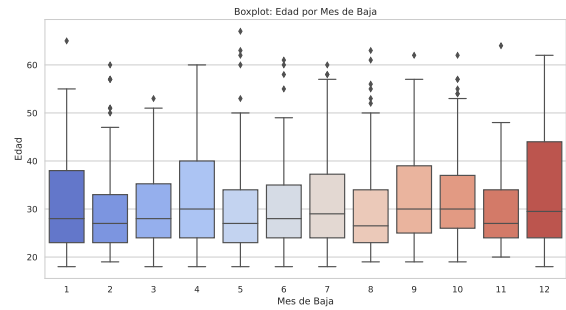


Figure 6: Edad y mes de baja

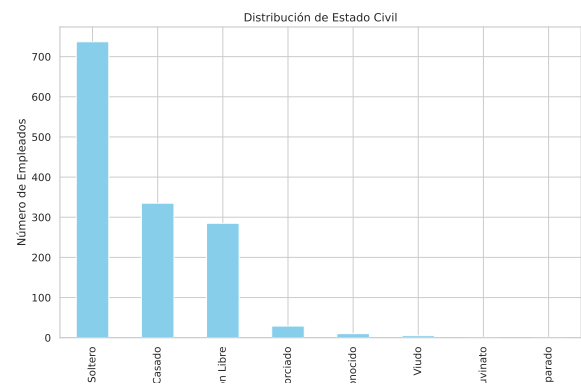


Figure 7: Estado Civil

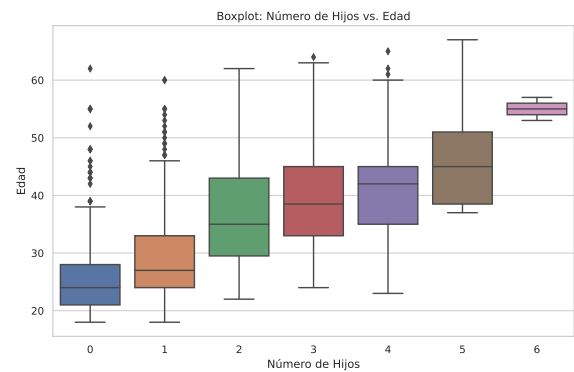


Figure 8: Número de hijos por edad del trabajador

Posteriormente, se hizo la separación de valores numéricos y categóricos y hacer la división de los datos en entrenamiento y prueba.

4.2 Modelos

Los modelos de clasificación que se usaron son los siguientes:

- (1) RandomForestClassifier

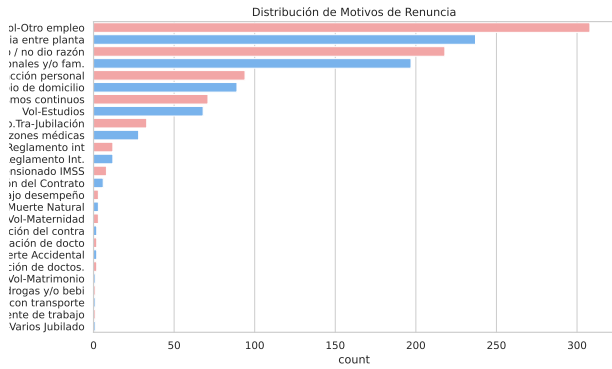


Figure 9: Motivos de renuncia

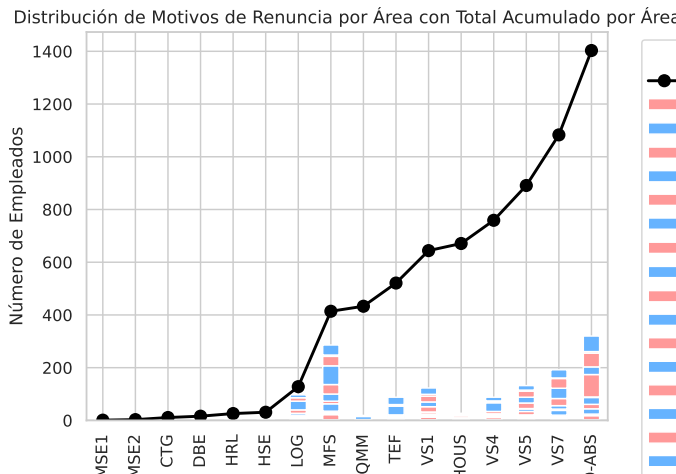


Figure 10: Distribución de motivos de renuncia

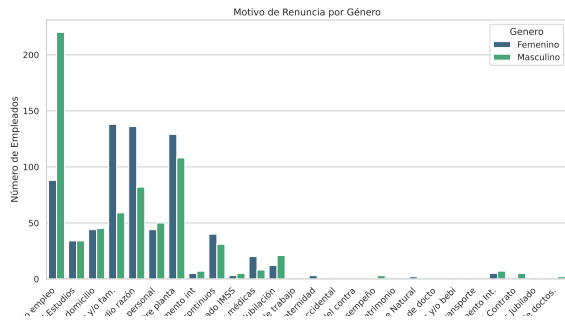


Figure 11: Motivos de renuncia de género

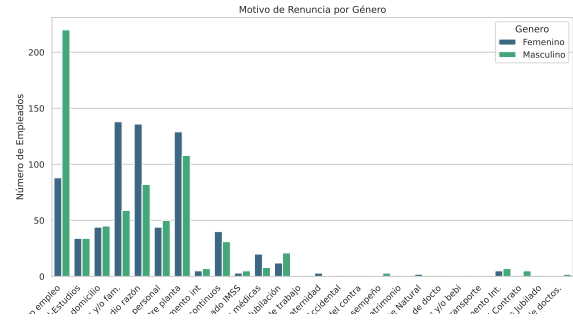


Figure 12: Motivos de renuncia de género

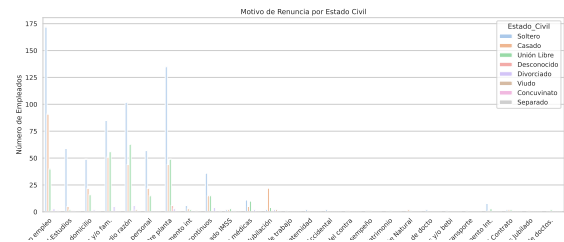


Figure 13: Motivos de renuncia de género

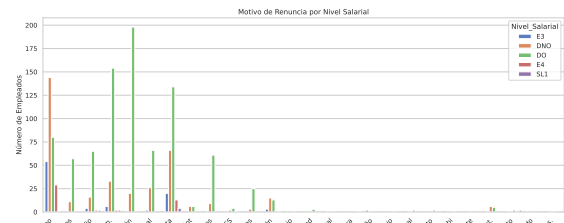


Figure 14: Motivos de renuncia de género

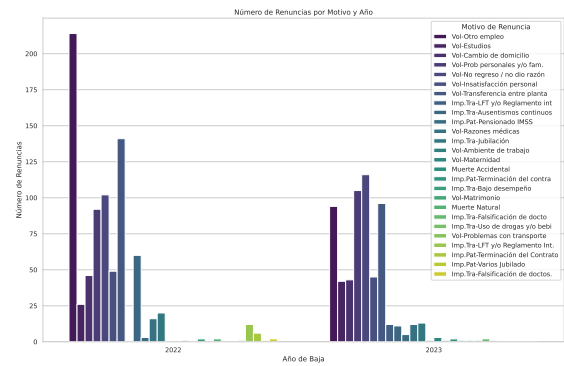


Figure 15: Motivos de renuncia por año

- (2) DecisionTreeClassifier
- (3) LogisticRegression

4.2.1 RandomForestClassifier. Contexto de Uso:

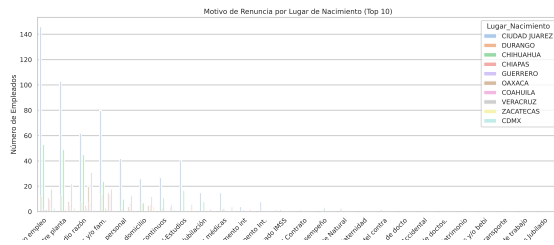


Figure 16: Lugar d origen

```
Data columns (total 23 columns):
# Column Non-Null Count Dtype
---
0 Posición 1403 non-null object
1 Area 1403 non-null object
2 Liquidación 1403 non-null object
3 Grupo 1403 non-null object
4 CODIGO_POSTAL 1403 non-null int64
5 Motivo_Renuncia 1403 non-null object
6 Nivel_Salarial 1403 non-null object
7 Fecha_Baja 1403 non-null object
8 Regla_PHT 1403 non-null object
9 Fecha_Alta 1403 non-null object
10 Años 1403 non-null int64
11 Meses 1403 non-null float64
12 Antigüedad 1403 non-null int64
13 Genero 1403 non-null object
14 Lugar_Nacimiento 1403 non-null object
15 Nacionalidad 1403 non-null object
16 Edad 1403 non-null int64
17 Tiempo_CDJuarez 1403 non-null object
18 Estado_Civil 1403 non-null object
19 Numero_Hijos 1403 non-null int64
20 RangoEdad 1403 non-null object
21 Mes_Baja 1403 non-null int64
22 Año_Baja 1403 non-null int64
dtypes: float64(1), int64(7), object(15)
```

Figure 17: Tipo de datos del dataset

```
Vol-Otro empleo 308
Vol-Transferencia entre planta 237
Vol-No regreso / no dio razón 218
Vol-Prob personales y/o fam. 197
Vol-Insatisfacción personal 94
Vol-Cambio de domicilio 89
Imp.Tra-Ausentismos continuos 71
Vol-Estudios 68
Imp.Tra-Jubilación 33
Vol-Razones médicas 28
Imp.Tra-LFT y/o Reglamento int 12
Imp.Tra-LFT y/o Reglamento Int. 12
Imp.Pat-Pensionado IMSS 8
Imp.Pat-Terminación del Contrato 6
Imp.Tra-Bajo desempeño 3
Name: Motivo_Renuncia, dtype: int64
```

Figure 18: Tipo de datos del dataset

- Ideal para problemas de clasificación, especialmente cuando hay un gran número de características o variables predictivas.
- Útil cuando se busca un modelo robusto y resistente al sobreajuste.
- Capaz de manejar automáticamente la importancia de las características y manejar conjuntos de datos grandes y complejos.

Características del Modelo:

- Ensamble de árboles de decisión.

- Cada árbol se entrena con una muestra diferente del conjunto de datos y las predicciones se combinan para obtener una predicción final.
- Puede manejar tanto problemas de clasificación como de regresión.

Consideraciones:

- Puede ser propenso a sobreajuste en conjuntos de datos pequeños con ruido.
- Ofrece una buena interpretabilidad de las características más importantes.

La búsqueda de hiperparámetros nos puede ayudar a encontrar la combinación óptima que mejora el rendimiento del modelo en función de la métrica de evaluación seleccionada (por defecto, la precisión).

La matriz de confusión se aprecia en la Figura 19:

```
Matriz de Confusión:
[[ 0  0  0  0  0  0  0  0  0  0  1  0  0  1]
 [ 0  0  0  0  0  0  0  0  0  0  1  0  0  1]
 [ 0  0  0  0  0  0  0  0  0  0  5  4  2  0  3]
 [ 0  0  0  0  0  0  0  0  0  0  0  1  0  0  0]
 [ 0  0  0  0  1  0  0  0  0  0  0  0  0  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  1  0  0]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  1  0  0]
 [ 0  0  0  0  0  0  0  0  1  1  1  6  5  0  1]
 [ 0  0  0  0  0  0  0  0  1  0  1  4  1  0  3]
 [ 0  0  0  0  0  0  0  1  0  0  0  10  4  0  0]
 [ 0  0  1  0  0  0  0  0  0  0  35  2  5  0  3]
 [ 0  0  0  0  0  0  0  0  1  0  1  46  8  0  12]
 [ 0  0  0  0  0  0  0  1  1  0  10  8  10  0  5]
 [ 0  0  0  0  0  0  0  0  0  0  0  0  1  2  0  1]
 [ 0  1  0  0  0  0  0  0  1  0  3  25  15  0  8]]
```

Figure 19: Matriz de confusión

La matriz de confusión es útil para evaluar el rendimiento del modelo, especialmente en problemas de clasificación. Proporciona información sobre cuántas instancias fueron clasificadas correctamente y cuántas fueron clasificadas incorrectamente para cada clase:

- TN (True Negatives): Instancias que fueron clasificadas correctamente como negativas.
- FP (False Positives): Instancias que fueron clasificadas incorrectamente como positivas.
- FN (False Negatives): Instancias que fueron clasificadas incorrectamente como negativas.
- TP (True Positives): Instancias que fueron clasificadas correctamente como positivas.

Dando para este modelo una exactitud de 0.36462093862815886. La exactitud es una métrica comúnmente utilizada para evaluar modelos de clasificación, pero su interpretación puede verse afectada por el desequilibrio de clases. En casos de clases desequilibradas, puede ser beneficioso considerar otras métricas como precisión, recall o F1-score.

F1-Score es la media armónica de precisión y recall. Es útil cuando hay un desequilibrio entre precisión y recall.

4.2.2 DecisionTreeClassifier. Contexto de Uso:

- Adecuado para problemas de clasificación y regresión.
- Útil cuando se busca un modelo interpretable y fácil de visualizar.
- Puede ser una buena opción para conjuntos de datos pequeños y limpios.

Características del Modelo:

Reporte de Clasificación:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.00	0.00	0.00	2
2	0.00	0.00	0.00	14
3	0.00	0.00	0.00	1
4	1.00	1.00	1.00	1
5	0.00	0.00	0.00	1
6	0.00	0.00	0.00	1
7	0.00	0.00	0.00	15
8	0.20	0.10	0.13	10
9	0.00	0.00	0.00	21
10	0.62	0.76	0.69	46
11	0.41	0.68	0.51	68
12	0.19	0.26	0.22	38
13	0.00	0.00	0.00	4
14	0.17	0.15	0.16	53
accuracy			0.36	277
macro avg	0.17	0.20	0.19	277
weighted avg	0.28	0.36	0.31	277

Figure 20: Reporte de clasificación

- Crea un árbol de decisiones basado en las características del conjunto de datos.
- Cada nodo en el árbol representa una decisión basada en una característica.
- Puede manejar datos categóricos y numéricos.

Consideraciones:

- Puede ser propenso al sobreajuste, especialmente en conjuntos de datos complejos o ruidosos.
- Puede requerir técnicas de ajuste de hiperparámetros para evitar sobreajuste.

Exactitud del modelo DecisionTreeClassifier: 0.259927797833935 (ver Figura 21)

Reporte de Clasificación para DecisionTreeClassifier:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.00	0.00	0.00	2
2	0.10	0.21	0.18	14
3	0.00	0.00	0.00	1
4	1.00	1.00	1.00	1
5	0.00	0.00	0.00	1
6	0.00	0.00	0.00	1
7	0.04	0.07	0.05	15
8	0.00	0.00	0.00	10
9	0.12	0.10	0.11	21
10	0.53	0.54	0.54	46
11	0.38	0.35	0.37	68
12	0.11	0.11	0.11	38
13	0.00	0.00	0.00	4
14	0.27	0.23	0.24	53
accuracy			0.26	277
macro avg	0.17	0.17	0.17	277
weighted avg	0.27	0.26	0.26	277

Figure 21: Reporte de Clasificación para DecisionTreeClassifier

4.2.3 LogisticRegression. Contexto de Uso:

- Adecuado para problemas de clasificación binaria y multi-clase.
- Útil cuando se busca un modelo lineal que funcione bien con características lineales o aproximadamente lineales.
- Puede ser una buena opción para problemas con interpretación de coeficientes.

Características del Modelo:

- Modelo lineal que utiliza la función logística para producir salidas entre 0 y 1.
- Se puede entender como la probabilidad de pertenecer a una clase en lugar de una predicción directa de la clase.

Consideraciones:

- Sensible a valores atípicos en los datos.

- No es adecuado para relaciones no lineales entre características y resultado.
- Puede requerir preprocesamiento adicional para manejar multicolinealidad u otros problemas.

Exactitud del modelo LogisticRegression: 0.3574007220216607.

Reporte de Clasificación para LogisticRegression:				
	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.00	0.00	0.00	2
2	0.00	0.00	0.00	14
3	0.00	0.00	0.00	1
4	0.00	0.00	0.00	1
5	0.00	0.00	0.00	1
6	0.00	0.00	0.00	1
7	0.00	0.00	0.00	15
8	0.00	0.00	0.00	10
9	0.00	0.00	0.00	21
10	0.38	0.67	0.49	46
11	0.45	0.72	0.55	68
12	0.32	0.24	0.27	38
13	0.00	0.00	0.00	4
14	0.18	0.19	0.19	53
accuracy			0.36	277
macro avg	0.09	0.12	0.10	277
weighted avg	0.25	0.36	0.29	277

Figure 22: Reporte de Clasificación para LogisticRegression:

4.3 Reporte de clasificación

El reporte de clasificación es el siguiente:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	16
1	0.74	0.60	0.67	53
2	0.66	0.73	0.69	70
3	0.31	0.41	0.35	34
4	0.33	0.39	0.36	38
accuracy			0.53	211
macro avg	0.41	0.43	0.42	211
weighted avg	0.52	0.53	0.52	211

Figure 23: Reporte de clasificación

5 CONCLUSIONES

El modelo tiene mejor exactitud al disminuir el número de clases, sin embargo se debe considerar hacer una mejor codificación de las variables categóricas a futuro o considerar optar por redes neuronales para poder entrenar el modelo de manera más adecuada.

A continuación, las conclusiones personales.

Gabriel Solano de Sixto:

Este reto fue una gran fuente de aprendizaje, agradezco a IA Center, BOHS y al gran equipo que me tocó. Esta experiencia fue muy gratificante, porque aplicas muchas habilidades, y sobre todo trabajar en colaboración, compartir ideas, puntos de vista sin duda es algo increíble. En definitiva, el Datathon me ha dado un nuevo panorama de las cosas, en donde he visto que lo más importante, en todo, no es armar un modelo de Machine Learning o limpiar los datos. Si no entender el contexto, porque de nada sirve que tengas un montón de datos, si no sabes cuál es la perspectiva de negocio, o a qué solución quieres llegar. Un ejemplo claro es este reto de Análisis de Recursos Humanos, donde podremos observar que hace que los empleados migren a otras empresas. Hay cosas que no me imaginaba, como que a causa de la inseguridad, algunas personas renuncian a su trabajo o el estado civil afecta a que cambies de empleo. Sin duda, el hacer un análisis de este tipo, te focaliza a ver de donde estás partiendo y que camino tomaste para llegar a

una solución que ayude a la empresa a mejorar en sus áreas de oportunidad.

Jorge de Jesús Hernández: Considero que al realizar el análisis exploratorio de los datos pudimos determinar algunas razones por las cuales las personas pueden optar a otros trabajos, si bien la edad y la forma de pensar afectan a las decisiones que tomamos, también es cierto que mientras más libertad tengamos más posibilidad hay de tomar decisiones arriesgadas, el tener una estabilidad también puede ser importante para las personas tanto familiar, económica, profesional y personal.

La retención de los empleados no siempre depende de las empresas, pero sin duda se puede evitar una rotación excesiva con programas, incentivos un buen ambiente laboral y considerar al empleado una persona más dentro de la organización. Considero que es importante la satisfacción del empleado también, para reducir costos en nuevos procesos de contratación de personal y además con la satisfacción del personal se ha demostrado que la productividad puede aumentar.

Este reto me dio la posibilidad de aplicar mis conocimientos y también desarrollar nuevos, el trabajo en equipo fue fundamental para poder desarrollar la mejor idea y también el poder abrir la

mente a posibilidades que no tenía en mi punto de vista, a futuro me gustaría poder participar en nuevos retos de diferentes áreas.

REFERENCES

- [1] 2023. How HR Analytics is transforming talent management strategies. <https://www.linkedin.com/pulse/how-hr-analytics-transforming-talent-management-strategies-africdsa/>. Accessed: 2023-12-2.
- [2] 2023. Talent management analytics for strategic decision-making. <https://engagedly.com/solutions/talent-management/data-driven-insights-for-strategic-decision-making/>. Accessed: 2023-12-2.
- [3] Hjorth Boisen and Susann Vallentin. 2009. La industria maquiladora y la migración interna en México. *Gaceta Laboral* 15, 1 (2009), 5–28.
- [4] Siti Zawiah Md Dawal and Zahari Taha. 2009. Job satisfaction model for manufacturing industry: Focus on ergonomics and human factors in work design. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*. 239–243. <https://doi.org/10.1109/TIC-STH.2009.5444498>
- [5] Oscar Galván Mendoza. 2019. *Efectos de la alienación laboral sobre la satisfacción laboral y el empoderamiento psicológico*. Ph.D. Dissertation. <http://erecursos.uacj.mx/handle/20.500.11961/5110>
- [6] Sergio Ravelo Blancas, Patricia Sánchez Díaz. 2006. Retroceso laboral, discriminación y riesgo en las maquiladoras: el casode Ciudad Juárez, Chihuahua. *El Cotidiano* (2006). <https://www.redalyc.org/articulo.oa?id=32513510>
- [7] Wietske Van Osch, Charles Steinfield, and Yanjie Zhao. 2016. Team Boundary Spanning through Enterprise Social Media: Exploring the Effects of Group-Level Diversity Using a Data Science Approach. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*. 2176–2185. <https://doi.org/10.1109/HICSS.2016.274>