Tabla de contenido

[Introducción 3](#_Toc141137833)

[Softwares Claves en el Análisis de Datos 3](#_Toc141137834)

[Plataformas y Herramientas Transversales 4](#_Toc141137835)

[Planteamiento del problema 5](#_Toc141137836)

[Objetivo General 6](#_Toc141137837)

[Objetivos específicos 6](#_Toc141137838)

[Fases del Proceso de Análisis de Datos 6](#_Toc141137839)

[Mapa de Ruta 8](#_Toc141137840)

[Análisis Exploratorio de Datos 9](#_Toc141137841)

[Implementación Modelo de clasificación Logística 10](#_Toc141137842)

[Evaluación del modelo 10](#_Toc141137843)

[Visualización de correlaciones absolutas 11](#_Toc141137844)

[Visualización de categorías de análisis en Power BI 12](#_Toc141137845)

[Plan de Acción 12](#_Toc141137846)

[Conclusiones 13](#_Toc141137847)

[Recomendaciones 14](#_Toc141137848)

[Modelo CRISP-DM 16](#_Toc141137849)

****

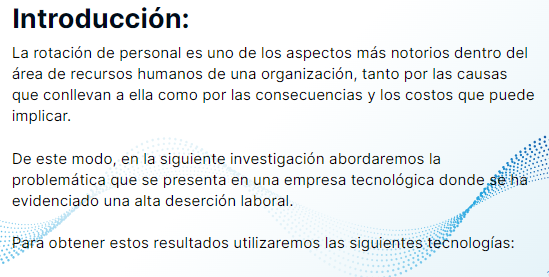
**Muy buenas tardes para todos, nos da gusto presentarles este proyecto sobre el análisis de la rotación de personal de una empresa tecnológica, destacando que la empresa realizó la selección de nuestros servicios luego de evaluar distintas opciones del mercado y particularmente por la agrupación profesional de quienes la conformamos**

****

**En ese sentido pasamos a presentarnos, Jorge Luis Arias, licenciado en turismo, enamorado del análisis de datos, autodidacta y con ganas de seguir formándose, Jholman Andrés Cuartas, ingeniero forestal con especialización en gestión ambiental , convencido de que la tecnología hace parte fundamental de mi día a día, seguidamente Diana María Moncada, Administradora de Empresas, con diplomado en innovación empresarial, atraída por la buena gestión de datos empresariales ya que considero que son la clave del éxito para cualquier empresa. Arístides Bermúdez Consultor Financiero, inclinado a que la informática permite soluciones idóneas, oportunas y eficientes en términos económicos….. y Tamara Omaña, Técnica superior en administración y mención informática, convencida de que la tecnología nos ayuda a enfocarnos mejor en nuestros objetivos empresariales.**

**Por Arístides**

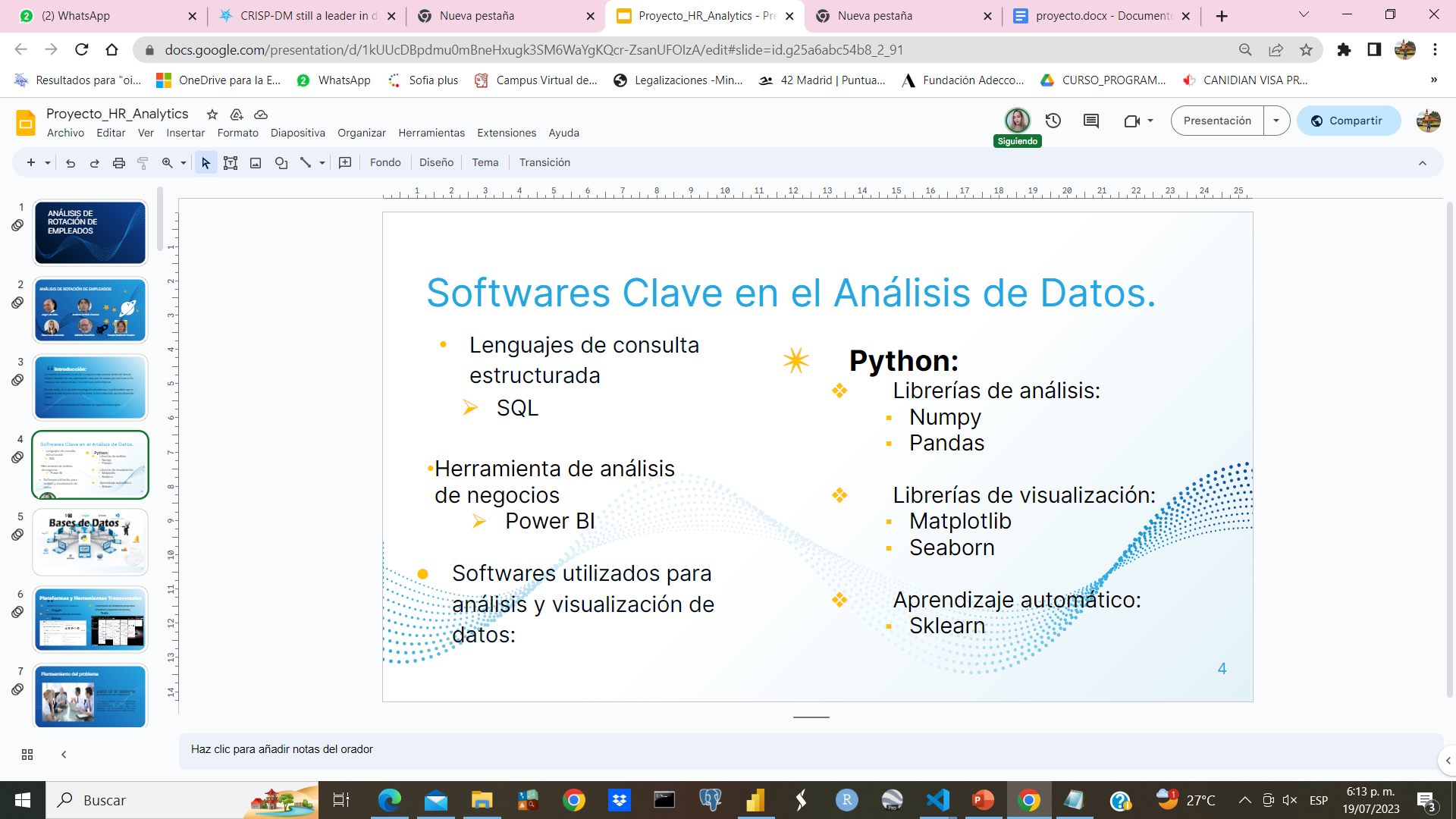
# Introducción

****

**Contenido sugerido**

**“ La rotación de personal se convierte en una incidencia de alto impacto en los resultados cualitativos y cuantitativos de la empresa, de ese modo abordamos la problemática con herramientas tecnológicas que, como se detalla a continuación “**

## Softwares Claves en el Análisis de Datos



## Plataformas y Herramientas Transversales

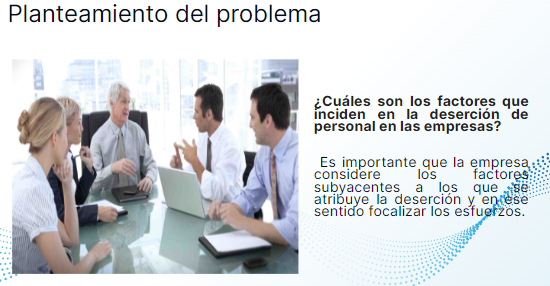
****

**Las plataformas y herramientas utilizadas:**

* **Kaggle, para la obtención de dataset**
* **GitHub, aplicación de control de versiones permitiéndonos gestionar los cambios que surgieron, y así podernos colaborar en el desarrollo.**
* **Trello: Herramienta de gestión de proyectos para planificar y organizar las tareas llevando un perfecto orden, ya que teníamos rápidamente una visualización clara acerca del estado del proyecto y las tareas pendientes.**

**Por Tamara**

# Planteamiento del problema

****

**El equipo se plantea la siguiente pregunta para el desarrollo de este proyecto: ¿Cuáles son los factores que inciden en la deserción de personal en las empresas?**

**En este sentido es importante que las empresas consideren los factores subyacentes a los que se atribuye la deserción laboral y tomar medidas que eviten pérdidas innecesarias**

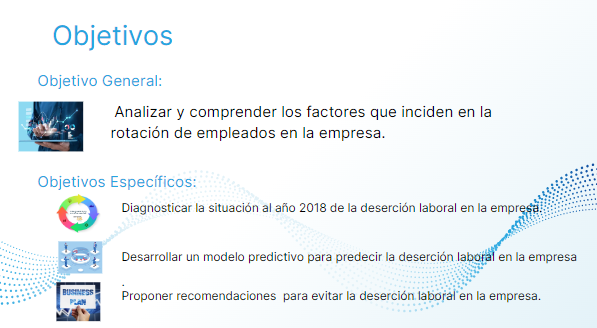
**Es por esto que decidimos plantearnos los siguientes objetivos:**

**El objetivo principal y su alcance es analizar y comprender los datos suministrados**

**y como trazados específicos los siguientes:**

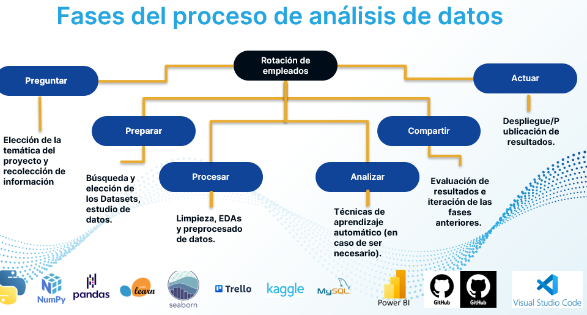
## Objetivo General

## Objetivos específicos

****

# Fases del Proceso de Análisis de Datos

**Jholman:**

****

**Teniendo en cuenta lo anterior,** el equipo quiso utilizar los datos para mejorar la retención de los empleados. Para ello, tuvimos que dividir este proyecto en tareas manejables y las organizamos en torno a las seis fases del proceso de análisis de datos.

la primera fase fué:

PREGUNTAR: aquí nos hicimos las preguntas tanto para definir el problema a resolver, como lo que equivaldría a un resultado exitoso.

PREPARAR: Luego nos preparamos y encontramos el dataset de estudio en la plataforma Kaggle, el cúal teniendo en cuenta sus variables, consideramos nos iba a arrojar los resultados que esperábamos.

PROCESAR: Procesamos los datos, los limpiamos, y nos aseguramos de que estuvieran completos, correctos, pertinentes y sin errores ni valores atípicos.

ANALIZAR: luego analizamos los datos limpios y sacamos nuestras primeras visualizaciones que más adelante les vamos a mostrar, se decide por recomendación de nuestros mentores crear una columna fecha la cual, nos iba a permitir analizar mejor la deserción en la empresa y obtener mejores resultados. Adicionalmente, teniendo en cuenta la correlación de variables con la columna attrition, se obtienen las categorías de análisis.

Por otra parte, se aplican técnicas de aprendizaje automático para predecir la deserción laboral, y este caso utilizamos un algoritmo de clasificación como lo es la regresión logística, la cuál es una opción común y eficiente para problemas de clasificación binaria pues es un modelo lineal que se utiliza para predecir la probabilidad de pertenecer a una clase determinada.

COMPARTIR: aquí evaluaremos los resultados, los cuales compartiremos con ustedes en la siguiente parte de la presentación, mostraremos también las interacciones de las fases anteriores.

Por último está la fase ACTUAR: En función de los resultados, nos centramos en dar nuestras recomendaciones para mejorar las áreas claves y así bajar el porcentaje de deserción laboral en la compañía.

# Mapa de Ruta

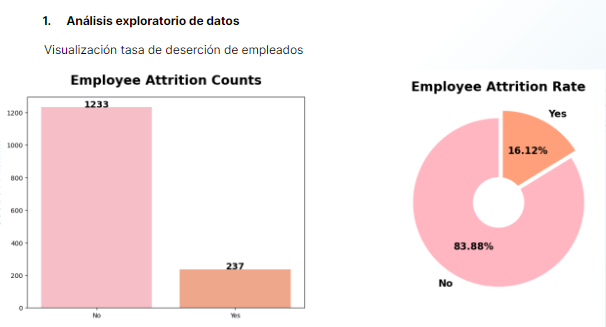


igualmente les queremos mencionar cuál fue el mapa de ruta para el desarrollo del proyecto:

1. análisis exploratorio de los datos
2. creación de la columna año
3. análisi de la deserción por correlación con otras variables
4. análisis estadístico, importancia de las características
5. visualizaciones
6. conclusiones y recomendaciones.

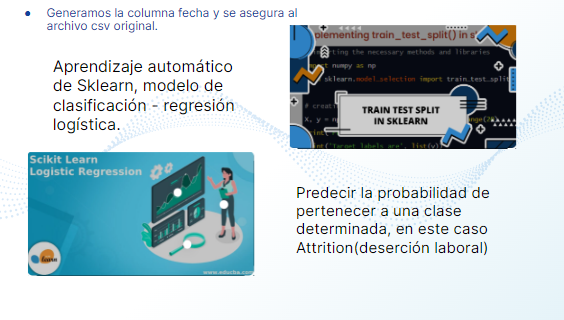
# Análisis Exploratorio de Datos

DIANA



En el análisis exploratorio de los datos, el dataset se basa en 1470 empleados, donde la gráfica de visualización de la tasa de deserción, arroja el 16,12%, porcentaje muy elevado, ya que según los expertos en el campo de rrhh, la tasa de deserción oscila entre 4 al 6%; por lo que la tasa de deserción está en un nivel de alta incidencia principalmente por la repercusión en otras áreas de trabajo. En ese sentido, la organización debe tomar medidas inmediatas para reducir la tasa de deserción.

## Implementación Modelo de clasificación Logística



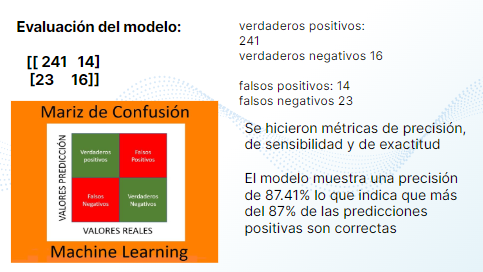
Apoyados en la biblioteca de Python de aprendizaje automático de Scikit-learn, decidimos utilizar el modelo de clasificación de regresión logística

pudiendo predecir la probabilidad de pertenecer a una clase determinada, en este caso la deserción laboral.

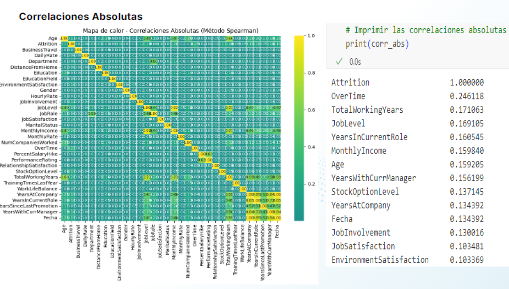
Arrojando con precisión el 87,41%, lo que indica que las predicciones positivas son correctas

En detalle mostraremos el mapa de calor, indicando las correlaciones entre variables y se imprimen las correlaciones absolutas con la variable attrition.

## Evaluación del modelo

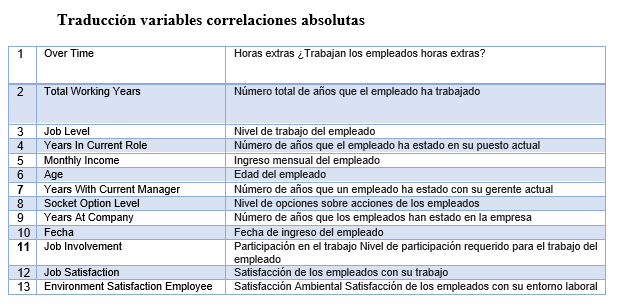


## Visualización de correlaciones absolutas



Según sus resultados de correlación hemos obtenido, las variables consideradas relevantes para analizar la deserción son aquellas que tienen una correlación positiva relativamente alta con "Attrition".

Estas variables indican su relación directa con la probabilidad de deserción y, para nuestro caso son:

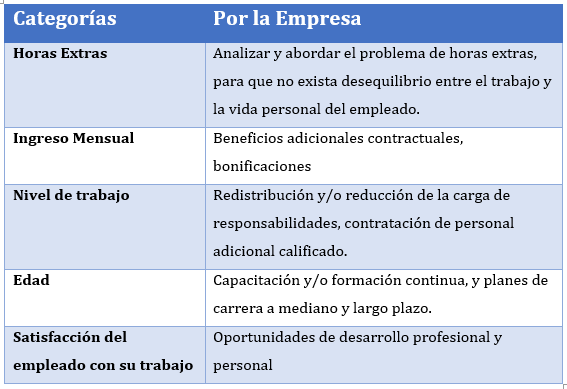


Estas variables las consideramos como clave para analizar la deserción en el conjunto de datos.

## Visualización de categorías de análisis en Power BI

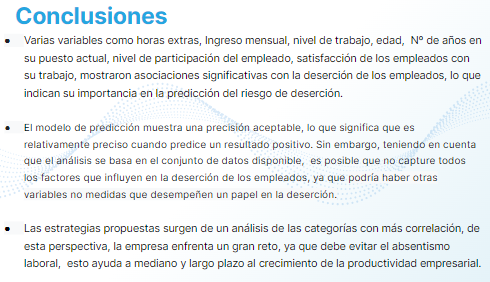
# Plan de Acción

**Aristides:** Teniendo en cuenta lo anterior quisimos propender el siguiente plan de acción, para el cual se tuvo en cuenta las categorías de análisis más relevantes como son:

****

**Tamara:**

# Conclusiones

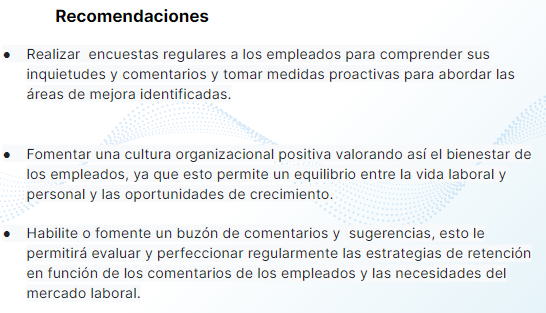
****

**Conclusiones:**

1. Varias variables como horas extras, Ingreso mensual, nivel de trabajo, edad, Nº de años en su puesto actual, nivel de participación del empleado, satisfacción de los empleados con su trabajo, mostraron asociaciones significativas con la deserción de los empleados, lo que indican su importancia en la predicción del riesgo de deserción.
2. El modelo de predicción muestra una precisión aceptable, lo que significa que es relativamente preciso cuando predice un resultado positivo. Sin embargo, teniendo en cuenta que el análisis se basa en el conjunto de datos disponible, es posible que no capture todos los factores que influyen en la deserción de los empleados, ya que podría haber otras variables no medidas que desempeñen un papel en la deserción.
3. Las estrategias propuestas surgen de un análisis de las categorías con más correlación, de esta perspectiva la empresa enfrenta un gran reto, ya que debe evitar el absentismo laboral, ya que esto ayuda a mediano y largo plazo al crecimiento de la productividad empresarial.

# Recomendaciones

**Diana**

****

**Recomendaciones:**

* Realizar encuestas regulares a los empleados para comprender sus inquietudes y comentarios y tomar medidas proactivas para abordar las áreas de mejora identificadas.

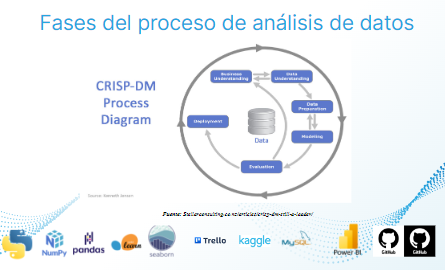
- Fomentar una cultura organizacional positiva valorando así el bienestar de los empleados, ya que esto permite un equilibrio entre la vida laboral y personal y las oportunidades de crecimiento.

Habilite o fomente un buzón de comentarios y sugerencias, esto le permitirá evaluar y perfeccionar regularmente las estrategias de retención en función de los comentarios de los empleados y las necesidades del mercado laboral.

**Andrés**

# Modelo CRISP-DM

**Finalizamos con este esquema que es importante tener en cuenta como analistas de datos.**

****

**Las cuales son el modelo CRISP-DM (Cross-Industry Standard Process for Data Mining) tiene 6 fases**

* **Comprensión del negocio**
* **Comprensión de los datos**
* **Preparación de los datos**
* **Modelado**
* **Evaluación, y**
* **Despliegue**

**Comprensión del negocio: comprender el objetivo y los requisitos del análisis de datos, identificar metas y criterios de éxito del proyecto. analiza situación actual del negocio y define preguntas clave que se pretenden responder con el análisis de datos.**

**Comprensión de los datos: En esta etapa se recopilan los datos necesarios e se exploran para adquirir un conocimiento inicial de su estructura y calidad. Se realiza un análisis descriptivo de los datos, identificando su distribución, outliers y posibles problemas. Además, se busca identificar la relación entre los diferentes atributos y su relevancia para el análisis.**

**Preparación de los datos: En esta fase se realiza la limpieza y transformación de los datos, con el fin de prepararlos para el análisis. Se trata de un proceso iterativo en el que se eliminan los datos incorrectos y se corrigen los errores. También se lleva a cabo la integración de diferentes fuentes de datos, la selección de variables relevantes y la creación de nuevas variables si es necesario.**

**Modelado: En esta etapa se selecciona y aplica una variedad de técnicas de modelado para crear un modelo que pueda predecir o explicar los problemas identificados en la fase anterior. Se prueban diferentes modelos y se evalúan su capacidad predictiva o explicativa. También se lleva a cabo la validación cruzada y la optimización de los modelos.**

**Evaluación: En esta fase se evalúan los resultados obtenidos en el modelo y se comprueban si cumplen los objetivos establecidos en la fase de comprensión del negocio. Se analiza la calidad y significancia de los resultados y se determina si se deben realizar ajustes o mejoras en el modelo.**

**Despliegue: En esta última fase se implementa y documenta el modelo final. Se realiza su integración en la organización y se pone en marcha su uso. Además, se crean y distribuyen informes y se capacita a los usuarios para que puedan interpretar y utilizar los resultados obtenidos.**

**Cada una de estas fases se alimenta y depende de las anteriores, pero también pueden retroalimentar y ajustar las fases posteriores si es necesario. El modelo CRISP-DM proporciona una guía estructurada y sistemática para el proceso de análisis de datos, asegurando una mayor eficacia y eficiencia en el mismo.**



|  |  |  |
| --- | --- | --- |
| **STT** | **Nombre del atributo** | **Significado del atributo** |
| 1 | Age Employee's age | Edad Edad del empleado |
| 2 | Gender Employee's Gender | Género Género del empleado |
| 3 | BusinessTravel Frequency of employees' business trips | Frecuencia de los viajes de negocios de los empleados |
| 4 | DailyRate Daily salary rate for employees | DailyRate Tasa de salario diario para empleados |
| 5 | Department Office of employees | Departamento Oficina de empleados |
| 6 | DistanceFromHome Distance from home in miles to work | DistanceFromHome Distancia de la casa en millas al trabajo |
| 7 |  |  |
|  |  |  |

**Traducción variables correlaciones absolutas**

|  |  |  |
| --- | --- | --- |
| 1 | OverTime Do employees work overtime | Horas extras ¿Trabajan los empleados horas extras? |
| 2 | Total Working Years Total number of years the employee has worked | Número total de años que el empleado ha trabajado |
| 3 | Job Level Employee's level of work | JobLevel Nivel de trabajo del empleado |
| 4 | YearsInCurrentRole Number of years the employee has been in their current role | YearsInCurrentRole Número de años que el empleado ha estado en su puesto actual |
| 5 | Monthly Income Employee's monthly income | Ingreso mensual del empleado |
| 6 | Age | Edad del empleado |
| 7 | YearsWithCurrManager Number of years an employee has been with their current manager | Número de años que un empleado ha estado con su gerente actual |
| 8 | SocketOptionLevel Employee stock option level | StockOptionLevel Nivel de opciones sobre acciones de los empleados |
| 9 | YearsAt Company Number of years employees have been with the company | Número de años que los empleados han estado en la empresa |
| 10 | Fecha | Fecha de ingreso del empleado |
| 11 | Job Involvement Level of involvement required for the employee's job | Participación en el trabajo Nivel de participación requerido para el trabajo del empleado |
| 12 | Job Satisfaction Employee satisfaction with their work | Satisfacción de los empleados con su trabajo |
| 13 | EnvironmentSatisfaction Employee satisfaction with their working environment | Satisfacción Ambiental Satisfacción de los empleados con su entorno laboral |
|  |  |  |

PLAN DE ACCIÓN A IMPLEMENTAR POR EMPRESA:

Vista la incidencia de las categorías de análisis se propone el siguiente plata de acción

* Implementar programas de retención de empleados.

\_ de reconocimiento

* de oportunidades de desarrollo profesional y personal
* Beneficios adicionales contractuales,bonificaciones
* Analizar y abordar problema de horas extras, se aprecia un volumen importante de empleados que aplican a esta manera para incrementar sus ingresos que a su vez desequilibra el trabajo y la vida personal y redunda en la economía de las áreas de trabajo
* Redistribución y/o reducción de la carga de responsabilidades, contratación de personal adicional y/o calificado y con experiencia
* Otras opciones como flexibilidad laboral en distintos aspectos
* Promover el desarrollo profesional, entre otros, capacitación y/o formación continua y planes de carrera a mediano y largo plazo