

Analítica Predictiva para Compensaciones y Equidad de Género en Organizaciones de Chile

Alumno: Luis Delgado López

Profesor Guía: Broderick Crawford

Magíster de Informática

Universidad Bernardo O'Higgins

Abstract

La presente investigación titulada “Analítica Predictiva para el Estudio de Compensaciones y Equidad de Género”, busca explorar la aplicación de modelos de analítica predictiva como herramienta para identificar y abordar las disparidades de género en las estructuras de compensación laboral. A través de un enfoque cuantitativo, se analizarán conjuntos de datos de diversas organizaciones para detectar patrones y tendencias que puedan indicar sesgos de género en la asignación de salarios y beneficios. El objetivo es desarrollar un marco analítico que permita a las empresas, colegios y universidades realizar ajustes proactivos y fundamentados hacia una mayor equidad de género en sus políticas de compensación, contribuyendo así a la construcción de entornos laborales más justos e inclusivos.

Palabras Clave: Aprendizaje Automático, Análisis Predictivo, Estudio de Compensaciones

This research, titled “Predictive Analytics for the Study of Compensation and Gender Equity,” seeks to explore the application of predictive analytics models as a tool to identify and address gender disparities in labor compensation structures. Through a quantitative approach, data sets from various organizations will be analyzed to detect patterns and trends that may indicate gender biases in the allocation of salaries and benefits. The objective is to develop an analytical framework that allows companies to make proactive and informed adjustments towards greater gender equity in their compensation policies, thus contributing to the construction of fairer and more inclusive work environments.

Keywords: Machine Learning, Predictive Analytics, Compensation Study

Introducción

En la actualidad, la equidad de género en el ámbito laboral es un tema de gran relevancia social y económica. A pesar de los avances significativos en la lucha por la igualdad de género, persisten brechas salariales y de compensación que dificultan la retención del talento desproporcionadamente a las mujeres en el mercado de trabajo. Esta investigación se centra en la utilización de la analítica predictiva como una herramienta innovadora para examinar y entender las complejidades de las estructuras de compensación y su impacto en la equidad de género.

La analítica predictiva, aplicada al estudio de grandes volúmenes de datos laborales, tiene el potencial de revelar patrones ocultos y tendencias que contribuyen a entender la disparidad de género. Al identificar estos factores, las organizaciones podrán implementar estrategias más efectivas para promover un entorno de trabajo equitativo. Este estudio tiene como objetivo principal desarrollar un modelo predictivo que sirva de guía para las decisiones de compensación, asegurando que las prácticas de remuneración sean justas y no sesgadas.

Handwritten notes:
[X] Inadice
estructuras
mejor
x capitales
...
[+]

Con este propósito, se analizarán datos históricos y actuales de compensaciones para aproximadamente 300 empresas, colegios y universidades participantes del estudio. La meta es proporcionar a las empresas un marco de referencia que les permita avanzar hacia la igualdad de género, no solo como un imperativo ético, sino también como una ventaja competitiva en el mercado global.

1. Problema u oportunidad

Identificación y antecedentes

La equidad de género en el ámbito laboral es un desafío global que ha persistido a lo largo del tiempo. A pesar de los esfuerzos legislativos y organizacionales, las brechas salariales entre géneros continúan siendo una realidad palpable (Smith, 2020; López & Hernández, 2021). ✓

La problemática central radica en la necesidad de comprender las causas subyacentes de la inequidad de género en las compensaciones y desarrollar estrategias basadas en datos para su mitigación.

Se identifican los siguientes desafíos u oportunidades:

- **Desigualdad Salarial:** Las mujeres suelen recibir salarios inferiores a sus colegas masculinos, incluso cuando desempeñan roles similares o idénticos. Esta disparidad económica afecta la calidad de vida, la motivación laboral y la equidad en general.
- **Sesgos Inconscientes:** Los sesgos de género pueden influir en las decisiones de compensación. Estos prejuicios sutiles pueden afectar la asignación de salarios, promociones y oportunidades de desarrollo profesional.
- **Falta de Transparencia:** La falta de transparencia en las políticas de compensación dificulta la identificación y corrección de las desigualdades salariales. Las empresas a menudo carecen de sistemas robustos para monitorear y abordar estas brechas.
- **Impacto Social y Económico:** La equidad salarial no solo es un asunto de justicia, sino también un factor clave para el desarrollo sostenible y la productividad económica. Las desigualdades afectan negativamente a las familias y la sociedad en su conjunto.
- **Evolución hacia una gestión de riesgo compartido:** Las empresas están abandonando enfoques tradicionales y adoptando estrategias más flexibles que consideran tanto los intereses de la organización como los de los empleados.
- **Profesionalización del manejo:** Se busca una gestión más rigurosa y basada en datos. El uso de sistemas probados de evaluación de cargos garantiza la equidad interna.
- **Mejora en las encuestas de remuneración de mercado:** Las organizaciones

están invirtiendo en obtener datos precisos sobre las tendencias salariales externas para mantener su competitividad.

- **Métodos estadísticos para estructuras de cargos y remuneraciones:** El análisis predictivo ayuda a diseñar estructuras salariales óptimas y a identificar patrones de compensación.
- **Introducción gradual de sistemas de renta variable:** Las empresas buscan vincular la remuneración al desempeño y resultados.
- **Políticas diferenciadas según nivel jerárquico:** Las compensaciones deben adaptarse a las distintas posiciones dentro de la organización.
- **Estrategias de retención y gestión de talento:** La retención de ejecutivos y empleados clave es fundamental en un mercado competitivo.
- **Negociaciones colectivas y ajustes históricos:** Las empresas están revisando sus prácticas históricas y adaptándose a las nuevas realidades económicas.

Líneas Temáticas sobre Compensaciones y Análisis Predictivo

El estado del arte en modelos predictivos se está desarrollando activamente, con empresas e instituciones aplicando estas técnicas para mejorar la gestión de compensaciones y remuneraciones.

En el análisis de compensaciones evalúan los sistemas de recompensa y beneficios que se ofrecen a los empleados en las organizaciones. Esto incluye la revisión de modelos de compensación, estrategias remunerativas, y cómo estos influyen en la motivación y satisfacción laboral.

Los elementos clave a considerar son:

- **Remuneración y Recompensa:** Análisis de sueldos, salarios, incentivos y prestaciones.
- **Estrategias de Compensación:** Evaluación de planes para mejorar la productividad y satisfacer necesidades de los colaboradores.
- **Impacto en Productividad:** Relación entre sistemas de compensación y eficiencia laboral.
- **Tendencias Actuales:** Estudio de prácticas modernas y su efectividad en el contexto laboral.

Por otro lado, el Machine Learning se encuentra en constante evolución y en sus aplicaciones tenemos:

- **Análisis Predictivo:** Utilización de algoritmos de machine learning para predecir tendencias de compensación y ajustar estrategias salariales.
- **Personalización:** Desarrollo de modelos que permiten personalizar paquetes de compensación basados en datos individuales y de rendimiento.

- **Optimización de Recursos:** Implementación de sistemas que ayudan a optimizar la distribución de compensaciones y beneficios.
- **Detección de Disparidades:** Uso de técnicas de aprendizaje automático para identificar y corregir disparidades salariales dentro de las organizaciones.

2. Objetivos del Proyecto

Objetivo General

Desarrollar un modelo de analítica predictiva, que permita identificar y analizar las disparidades de género en las estructuras de compensación laboral, con el fin de promover la equidad de género y la toma de decisiones basadas en datos.

Objetivo Específicos

Para cumplir con las expectativas de este estudio, se espera lograr lo siguiente:

- Evaluar la situación actual para establecer una línea de base de la inequidad existente.
- Aplicar modelamiento predictivo para determinar en qué cargos y rubros se identifican brechas salariales entre hombres y mujeres.
- Realizar el desarrollo utilizando herramientas compatibles con Azure Cloud Services.
- Mantener la confidencialidad de la información en la presentación de los resultados.

3. Solución Propuesta

La relevancia de esta investigación se fundamenta en la premisa de que una mayor equidad de género no solo es un imperativo ético, sino que también conlleva beneficios económicos y de productividad para las organizaciones. Por lo tanto, el presente estudio busca contribuir al cuerpo de conocimiento existente, proporcionando evidencia empírica acerca del uso de machine learning para la predicción de compensaciones y como estos resultados afectan las brechas salariales de género en Chile.

Para ello se tomarán algunas consideraciones, que determinarán el alcance de esta iniciativa. A continuación se presenta un flujo de referencia acerca del flujo de la información y cómo esta interactuará con la nube de Azure y sus herramientas.

Metodología de Trabajo

El desarrollo de este estudio se dará de la siguiente manera

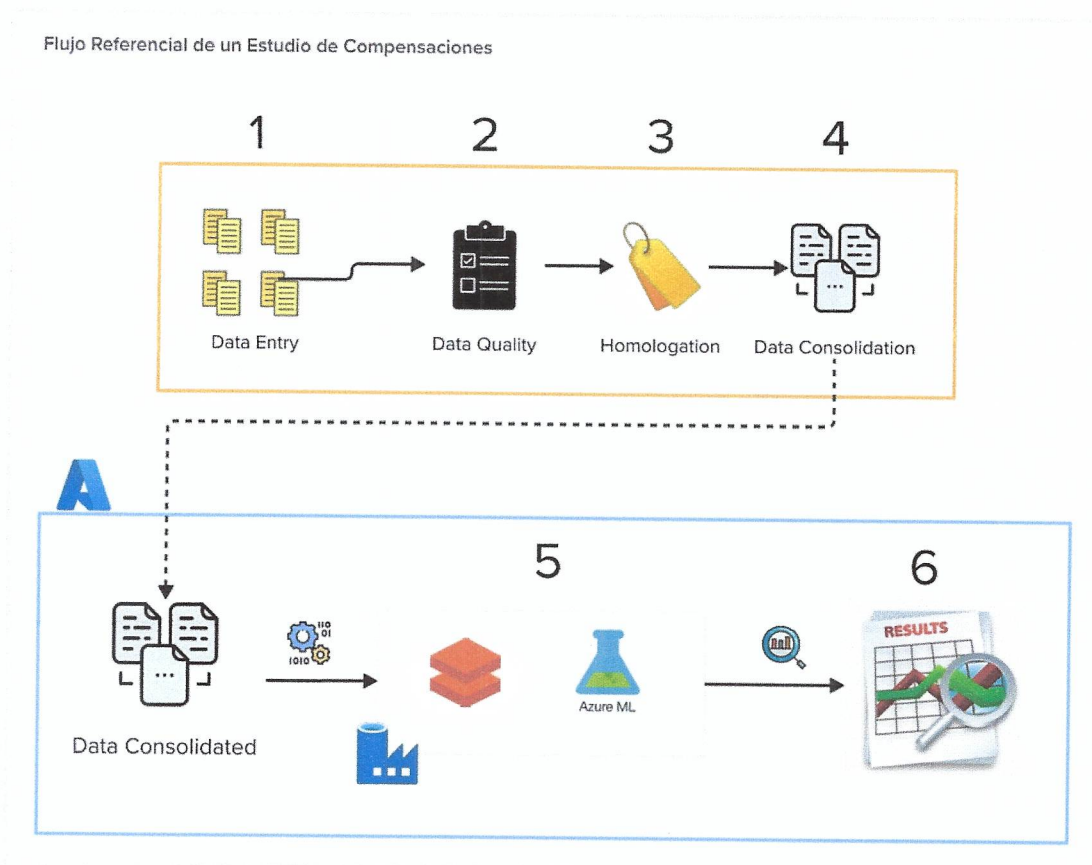


FIG. # _____

Marco Teórico

CRISP-DM (Cross-Industry Standard Process for Data Mining):

1. **Definición del Problema:** Especificar claramente el problema de equidad de género que se desea abordar con la modelación predictiva, como las disparidades salariales entre hombres y mujeres.
2. **Recolección de Datos:** Obtener datos relevantes que puedan incluir salarios, roles laborales, educación, experiencia laboral, y otros factores que puedan influir en las compensaciones.
3. **Preprocesamiento de Datos:** Limpiar y preparar los datos para el análisis, lo que puede incluir la eliminación de datos faltantes o irrelevantes, la codificación de variables categóricas y la normalización de rangos de datos.
4. **Selección de Variables:** Identificar las variables más relevantes que afectan la equidad de género en las compensaciones. Esto puede requerir análisis exploratorios y técnicas de selección de características.
5. **Construcción del Modelo:** Elegir y aplicar algoritmos de aprendizaje automático para crear modelos predictivos. Los modelos pueden ser supervisados o no supervisados, dependiendo de la naturaleza de los datos y los objetivos del estudio.
6. **Entrenamiento del Modelo:** Utilizar un conjunto de datos de entrenamiento para enseñar al modelo a reconocer patrones y hacer predicciones precisas.
7. **Evaluación del Modelo:** Probar el modelo con un conjunto de datos de prueba para evaluar su precisión y capacidad de generalización. Se pueden utilizar métricas como la precisión, la curva ROC, el área bajo la curva (AUC), entre otras.
8. **Ajuste del Modelo:** Basado en los resultados de la evaluación, ajustar y optimizar el modelo para mejorar su rendimiento.
9. **Implementación:** Aplicar el modelo a nuevos datos para hacer predicciones y tomar decisiones informadas sobre políticas de compensación para promover la equidad de género.
10. **Monitoreo y Actualización:** Continuar monitoreando el rendimiento del modelo y actualizarlo conforme se disponga de nuevos datos o cambien las condiciones del mercado laboral.

Estudio de Compensaciones

1. Acceso a Encuestas de Sueldos y Salarios: RECOPIACION DE DATOS
 - Obtener información es el primer paso para entender cómo se compara la remuneración de tu empresa con la de tus competidores.
2. Asignación de Responsabilidades: PROCESO DE HOMOLOGACION
 - Define claramente las responsabilidades y roles de los empleados.
 - Categoriza los puestos según niveles jerárquicos y funciones específicas.

3. Definición de Categorías Laborales: PROCESO DE HOMOLOGACION
 - Agrupa los puestos en categorías según similitudes en habilidades, responsabilidades y nivel de experiencia.
 - Esto facilitará la comparación y análisis.
4. Recopilación de Datos: RECOPIACION DE DATOS
 - Recolecta datos sobre salarios, beneficios y otros componentes de compensación.
 - Considera factores como antigüedad, educación y desempeño individual.
5. Análisis Comparativo: ANALITICA DESCRIPTIVA
 - Utiliza técnicas de analítica predictiva para analizar los datos recopilados.
 - Compara los salarios internamente y con el mercado.
6. Comparación con la Media del Mercado: ANALITICA DE MEDIAS
 - Evalúa si tus salarios están por encima o por debajo de la media del mercado.
 - Ajusta según la competitividad y las metas de retención.
7. Toma de Decisiones sobre la Estructura Salarial: DECISIONES POR EMPRESA
 - Basándote en los análisis, toma decisiones informadas sobre la estructura salarial.
 - Define rangos salariales para cada categoría laboral.
8. Alineación con los Resultados:
 - Asegurar de que la estructura de compensaciones esté alineada con los objetivos estratégicos de la empresa.

Métricas Metodológicas de Evaluación

-
1. **Error Cuadrático Medio (MSE):** Mide el promedio de los cuadrados de los errores, es decir, la diferencia cuadrada entre los valores observados y los valores predichos por el modelo¹.
2. **Error Absoluto Medio (MAE):** Es el promedio de las diferencias absolutas entre los valores predichos y los valores reales, proporcionando una medida de la magnitud de los errores sin considerar su dirección¹.

3. **Coefficiente de Determinación (R-cuadrado)**: Indica la proporción de la varianza de la variable dependiente que es predecible a partir de las variables independientes¹.
4. **Raíz del Error Cuadrático Medio (RMSE)**: Es la raíz cuadrada del MSE y proporciona una medida de la magnitud de los errores con una escala que es comparable a la de los datos originales¹.
5. **R-cuadrado Ajustado**: Ajusta el R-cuadrado basándose en el número de predictores en el modelo, proporcionando una medida más precisa de la bondad de ajuste para modelos con diferentes números de predictores¹.

Métricas de Evaluación de Negocio

1. **Coefficiente de Gini:**
 - a. El coeficiente de Gini es una medida utilizada para evaluar la **desigualdad económica**. Cuanto mayor sea el índice de Gini, mayor será la desigualdad de los ingresos en la población.
 - b. Se calcula en una escala de **0 a 1**, donde 0 representa una distribución perfectamente igualitaria de los ingresos y 1 representa la máxima desigualdad (una persona o grupo concentra todo el ingreso).
2. **Brecha Salarial:**
 - a. La brecha salarial se refiere a la diferencia en los ingresos entre diferentes grupos, como género, raza o nivel de experiencia.
 - b. Para calcular la brecha salarial, puedes restar el salario promedio de un grupo (por ejemplo, mujeres) del salario promedio del otro grupo (por ejemplo, hombres).
 - c. La brecha salarial se expresa como un porcentaje o una cantidad absoluta.
3. **Índice de Disparidad Salarial:**
 - a. El índice de disparidad salarial mide la diferencia entre los salarios de diferentes grupos.
 - b. Puede calcularse como la relación entre los salarios promedio de dos grupos (por ejemplo, hombres y mujeres).
 - c. Un índice superior a 1 indica que el grupo con salarios más altos gana significativamente más que el otro grupo.

PARA EL ANALISIS

1. **Análisis Comparativo de Compensaciones:**
 - Compara los salarios y beneficios de tu empresa con los del mercado en el que operas.
 - Evalúa si tus compensaciones son competitivas y si están alineadas con las tendencias del mercado¹.
2. **Análisis de Equidad Interna y Competitividad Externa:**
 - Examina la equidad salarial dentro de tu organización.
 - Compara las remuneraciones internas para garantizar que sean justas y coherentes con las responsabilidades y habilidades de los empleados.

- [Evalúa cómo se posiciona tu empresa en relación con otras empresas del mismo sector².](#)
3. **Análisis de Políticas de Incentivos Variables:**
 - Revisa las políticas relacionadas con bonos, comisiones y otros incentivos variables.
 - Evalúa su efectividad para motivar y retener a los empleados.
 4. **Análisis de Programas de Beneficios Laborales:**
 - Examina los beneficios no monetarios ofrecidos a los empleados, como seguro médico, vacaciones, flexibilidad laboral, etc.
 - [Evalúa la efectividad de estos programas para atraer y retener talento².](#)
1. **Compensación entre Sesgo y Varianza:**
 - El **trade-off** entre sesgo y varianza es un concepto importante al elegir un algoritmo de aprendizaje automático.
 - [El sesgo es la expectativa en el error, y la varianza es la variabilidad en el modelo¹.](#)
 2. **Modelos Predictivos para Equidad Salarial:**
 - Desarrollo de modelos predictivos que analicen datos de compensaciones.
 - [Identificación de patrones y tendencias relacionados con las diferencias salariales de género².](#)
 3. **Automatización de Procesos de Homologación de Puestos:**
 - Diseño de modelos de IA para automatizar procesos manuales en compensaciones.
 - [Por ejemplo, leer descripciones de trabajo y proporcionar recomendaciones para la homologación de puestos³.](#)
 4. **Optimización de Políticas de Reajuste Salarial:**
 - Utilización de algoritmos de machine learning para ajustar salarios según mérito, desempeño, promoción, equidad externa, etc.

1. **Análisis de Equidad Salarial con Machine Learning:**
 - **Ejemplo:** Un estudio realizado por **Salesforce** utilizó machine learning para analizar las diferencias salariales entre géneros en su empresa. El modelo identificó patrones y factores que contribuían a las brechas salariales. Como resultado, Salesforce ajustó los salarios para garantizar una mayor equidad.
-
2. **Automatización de Procesos de Homologación de Puestos:**
 - **Ejemplo:** **IBM** implementó un sistema basado en machine learning para homologar puestos de trabajo en su organización. El modelo analizó descripciones de trabajo y recomendó niveles salariales

apropiados para cada puesto. Esto permitió una mayor eficiencia y equidad en la estructura salarial.

- **Homologación de cargos**
- **Inferencia de rentas por carga/puesto de trabajo**

3. Predicción de Aumentos Salariales:

- **Ejemplo: Google** desarrolló un modelo predictivo utilizando machine learning para predecir cuándo y cuánto aumentar los salarios de los empleados. El modelo consideró méritos, desempeño y otros factores. Esto ayudó a tomar decisiones más informadas sobre compensaciones.
- **Modelo de scoring de desempeño para el reajuste de sueldos**

4. Optimización de Programas de Beneficios Laborales:

- **Ejemplo: Amazon** utilizó machine learning para analizar datos de empleados y optimizar sus programas de beneficios. El modelo identificó qué beneficios eran más relevantes para diferentes grupos de empleados, mejorando la satisfacción y retención del personal.
- **Clustering**
- **No supervisado para identificar las tendencias de cada empresa de acuerdo a su programa de beneficios**

1. Compensation Learning:

- **Descripción:** El **Compensation Learning** es un enfoque que se basa en la compensación de pesos en algoritmos de machine learning.
- **Aplicación:** Se ha utilizado para predecir y comprender las diferencias salariales, considerando factores como experiencia, educación y habilidades específicas¹.

2. Random Forest:

- **Descripción:** El **Random Forest** es un algoritmo de ensemble que combina múltiples árboles de decisión.
- **Aplicación:** Se ha aplicado para predecir con alta precisión las compensaciones y evaluar la equidad salarial¹.

3. Análisis de Regresión:

- **Descripción:** Los modelos de regresión, como la regresión lineal o la regresión logística, se han utilizado para relacionar variables independientes (como experiencia o habilidades) con salarios.
- **Aplicación:** Estos modelos ayudan a identificar qué factores influyen en las compensaciones¹.

4. Redes Neuronales:

- **Descripción:** Las redes neuronales son modelos inspirados en el cerebro humano que pueden aprender patrones complejos.

- Aplicación: Se han utilizado para predecir salarios y evaluar la equidad en grandes conjuntos de datos¹.

5. Gradient Boosting:

- **Descripción:** El **Gradient Boosting** es otro algoritmo de ensemble que combina múltiples modelos más débiles.
- Aplicación: Se ha aplicado para predecir compensaciones y evaluar la importancia relativa de diferentes características¹.

Equidad de Género

Analítica Predictiva

El modelado estadístico y el machine learning son dos disciplinas que comparten técnicas de análisis y se utilizan para aprender de los datos. Aunque no hay un único framework universalmente aceptado, existen varias bibliotecas y herramientas populares en Python que son ampliamente utilizadas para trabajar con machine learning y estadísticas.

1. Scikit-Learn (sklearn):

- Scikit-Learn es una biblioteca de Python que proporciona una amplia gama de algoritmos de machine learning y herramientas para el preprocesamiento de datos.
- Es ideal para tareas como clasificación, regresión, clustering y reducción de dimensionalidad.
- Scikit-Learn sigue una interfaz consistente y es ampliamente adoptado en la comunidad de data science.

2. StatsModels:

- StatsModels se centra en el modelado estadístico y proporciona herramientas para realizar análisis de regresión, pruebas de hipótesis y estimación de parámetros.
- Es especialmente útil para investigadores y analistas que desean realizar inferencias estadísticas.

3. TensorFlow y Keras:

- TensorFlow es una biblioteca de código abierto desarrollada por Google para machine learning y deep learning.

- Keras es una API de alto nivel que se ejecuta sobre TensorFlow y simplifica la creación de redes neuronales.
- Ambas herramientas son ampliamente utilizadas para tareas de aprendizaje profundo.

4. PyTorch:

- PyTorch es otra biblioteca popular para deep learning.
- Es conocida por su flexibilidad y facilidad de uso, especialmente para investigadores y experimentación.

5. Pandas y NumPy:

- Aunque no son específicamente frameworks de machine learning, Pandas y NumPy son esenciales para la manipulación y procesamiento de datos.
- Pandas proporciona estructuras de datos como DataFrames, mientras que NumPy es útil para operaciones numéricas y algebraicas.

6. XGBoost y LightGBM:

- Estas bibliotecas se centran en boosting y gradient boosting.
- Son ampliamente utilizadas para tareas de clasificación y regresión.

PROYECCIONES Y TENDENCIAS

El Análisis de Componentes Principales (PCA) es una técnica fundamental para reducir la dimensionalidad de los datos y encontrar patrones ocultos en un conjunto de características.

1. Preparación de Datos:

- Reúne los datos relevantes sobre compensaciones, como salarios, beneficios, antigüedad, nivel de experiencia, etc.
- Asegúrate de que los datos estén limpios y normalizados.

2. Implementación de PCA en Python:

- Utilizaremos la biblioteca scikit-learn para aplicar PCA.

- Instala la librería si aún no lo has hecho: `pip install scikit-learn`.
3. **Ejemplo de Aplicación de PCA:**
- Supongamos que tienes un conjunto de datos con múltiples características relacionadas con compensaciones (salario, bonificaciones, beneficios, etc.).
 - Aquí hay un ejemplo básico de cómo aplicar PCA en Python:

Python

```
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# Carga tus datos (reemplaza con tus datos reales)
data = pd.read_csv('datos_compensaciones.csv')

# Preprocesamiento: normaliza los datos
scaler = StandardScaler()
data_scaled = scaler.fit_transform(data)

# Aplica PCA
pca = PCA(n_components=2) # Elige el número de componentes deseados
pca_result = pca.fit_transform(data_scaled)

# Visualiza los resultados
plt.scatter(pca_result[:, 0], pca_result[:, 1])
plt.xlabel('Componente Principal 1')
plt.ylabel('Componente Principal 2')
plt.title('Visualización PCA')
plt.show()
```

4. **Interpretación:**
- Los componentes principales (PCs) representan combinaciones lineales de las características originales.
 - Puedes examinar los coeficientes de los PCs para entender qué características contribuyen más a cada componente.
5. **Selección de Componentes:**
- Elige el número de componentes según la cantidad de varianza que deseas retener. Puedes visualizar la varianza explicada por cada componente.
6. **Aplicación en Modelos de Machine Learning:**
- Utiliza los componentes principales como características en tus modelos de regresión, clasificación, etc.
 - Los PCs pueden ayudar a reducir la maldición de la dimensionalidad y mejorar la eficiencia computacional.