Proyecto Final Data Set Encuesta Habilidades de Poder estudiantes UCC

Juan Camilo Reyes - 466514

Juan Sebastián Hernández – 808042

Ing. Fernando Portela Gutiérrez

Universidad Cooperativa de Colombia

Sede Ibagué

Año 2024

Introducción

En la actualidad, el desarrollo de habilidades blandas como la creatividad, la innovación, la autoconciencia y la comunicación se ha convertido en un factor clave para el éxito académico y profesional de los estudiantes. Estas habilidades no solo contribuyen al rendimiento académico, sino que también preparan a los estudiantes para enfrentar los retos del entorno laboral, caracterizado por su dinamismo y constante evolución. Conscientes de la importancia de estas competencias, se ha diseñado una encuesta que busca evaluar la percepción de los estudiantes respecto a la relevancia y aplicación de dichas habilidades en su formación.

El propósito de este estudio es comprender cómo estas habilidades se desarrollan dentro del contexto académico, identificando áreas de mejora en el currículo, posibles barreras al desarrollo de competencias, y estrategias efectivas para fomentar el crecimiento personal y profesional de los estudiantes. La información recopilada permitirá a la institución educativa diseñar programas de apoyo más efectivos, optimizar la asignación de recursos, y fomentar un entorno que promueva la creatividad, la innovación, la autoconciencia y la comunicación, aspectos fundamentales para la formación integral de los futuros profesionales.

Los resultados de este estudio contribuirán a generar un impacto positivo en la experiencia educativa, mejorando tanto el compromiso de los estudiantes con su formación como la calidad de los servicios educativos ofrecidos. Además, este documento servirá como guía para la toma de decisiones estratégicas que permitan asegurar una formación académica adaptada a las necesidades reales del entorno moderno, promoviendo así el éxito integral de los estudiantes en sus trayectorias académicas y profesionales.

COMPRENSION DEL NEGOCIO

REVISIÓN DEL ESTADO DEL ARTE

El desarrollo de habilidades blandas, también conocidas como competencias transversales, ha cobrado una relevancia considerable en el ámbito educativo y profesional en los últimos años. Diversos estudios destacan la importancia de estas competencias como factores clave para el éxito en el ámbito académico y laboral, ya que permiten a los estudiantes adaptarse a los cambios, resolver problemas de manera creativa y comunicarse eficazmente con sus pares. A continuación, se presenta una revisión de investigaciones previas que abordan la creatividad, la innovación, la autoconciencia y la comunicación en el contexto académico.

En términos de **creatividad e innovación**, estudios recientes sugieren que estas habilidades no son inherentes, sino que pueden ser fomentadas y desarrolladas mediante metodologías de enseñanza adecuadas y actividades específicas. Según López y Gómez [1], la implementación de proyectos colaborativos y prácticas lúdicas en el currículo académico promueve el pensamiento creativo y la resolución de problemas, lo cual resulta esencial para que los estudiantes enfrenten los desafíos del entorno laboral. Asimismo, investigaciones como la de Ramírez y Silva [2] señalan que la falta de oportunidades para la exploración creativa y la existencia de metodologías de enseñanza rígidas son barreras significativas que limitan el desarrollo de la creatividad en los estudiantes universitarios.

La **autoconciencia** también ha sido ampliamente reconocida como una competencia clave para el crecimiento personal y académico. Tal como indica la investigación de Fernández y Torres [3], el autoconocimiento y la automotivación influyen positivamente en la capacidad de los estudiantes para gestionar el estrés, organizar su tiempo y mejorar su rendimiento académico. Además, se menciona que las actividades de autoevaluación y reflexión personal, tales como prácticas de mindfulness y sesiones de desarrollo personal, pueden ser efectivas para mejorar estos aspectos. A pesar de ello, los mismos autores destacan la necesidad de una mayor integración de estos programas en el currículo universitario para que los estudiantes desarrollen estas competencias de manera más efectiva.

En cuanto a la **comunicación**, tanto oral como escrita, así como la **expresión corporal**, estas habilidades son fundamentales para el desarrollo profesional de los estudiantes. Según un estudio realizado por González y Martínez [4], las habilidades

comunicativas permiten a los estudiantes expresarse de manera clara y coherente, un requisito esencial tanto en el ámbito académico como en el entorno laboral. Sin embargo, los autores también señalan que uno de los mayores desafíos para los estudiantes es la falta de oportunidades para practicar dichas habilidades en un entorno seguro y estructurado, lo cual limita su capacidad para mejorar en este aspecto. La promoción de clubes de debate, actividades orales y talleres de comunicación se presenta como una estrategia efectiva para superar estos obstáculos.

Finalmente, el desarrollo de las **habilidades blandas** se ha posicionado como una necesidad en el ámbito académico, debido a la creciente demanda del sector profesional de egresados que sean capaces de adaptarse y resolver problemas complejos. Sánchez y Pérez [5] argumentan que la integración de actividades prácticas, como hackathons, visitas a empresas y programas de mentorías, tiene un impacto significativo en el desarrollo de habilidades clave para la innovación y la creatividad. Estas actividades no solo motivan a los estudiantes, sino que también les proporcionan una experiencia práctica que mejora su capacidad para enfrentar desafíos reales.

En conclusión, la revisión del estado del arte indica que el desarrollo de competencias blandas, como la creatividad, la autoconciencia y la comunicación, resulta esencial para la formación integral de los estudiantes. Sin embargo, para lograr un impacto significativo, es fundamental que las instituciones educativas incorporen metodologías y actividades que permitan a los estudiantes explorar y fortalecer estas competencias de manera continua.

EVALUACIÓN DE LA SITUACIÓN

La situación actual en el contexto educativo de la institución revela la necesidad de una evaluación profunda sobre el desarrollo de habilidades blandas entre los estudiantes, especialmente aquellas relacionadas con la creatividad, la innovación, la autoconciencia y la comunicación. A través de la encuesta realizada, se ha identificado un conjunto de puntos críticos que afectan tanto el rendimiento académico como el desarrollo personal y profesional de los alumnos.

En primer lugar, los datos recopilados indican que, si bien los estudiantes reconocen la importancia de habilidades como la creatividad y la innovación, existen obstáculos significativos que limitan su aplicación durante el proceso de formación académica. Los estudiantes señalan como principales barreras la falta de espacios para la experimentación y la implementación de métodos de enseñanza tradicionales que no fomentan el pensamiento crítico y creativo. Esta situación sugiere que la institución educativa debe considerar la revisión y actualización de los planes de estudio para integrar metodologías más activas y participativas, que ofrezcan oportunidades reales para el desarrollo de estas competencias.

Por otro lado, la autoconciencia, que incluye elementos como el autoconocimiento, el autocontrol y la automotivación, se presenta como un aspecto crítico para la gestión del rendimiento académico y el bienestar emocional de los estudiantes. Muchos de ellos reportan dificultades para controlar sus emociones y la falta de motivación como factores que obstaculizan su desarrollo académico. Esta situación destaca la necesidad de implementar programas de apoyo que incluyan sesiones de desarrollo personal, actividades de reflexión y estrategias como mindfulness para mejorar la capacidad de autogestión emocional de los estudiantes.

En cuanto a las habilidades comunicativas, se observa que, aunque los estudiantes comprenden la relevancia de la comunicación oral, escrita y la expresión corporal, la falta de oportunidades para practicar dichas habilidades limita su progreso. Las actividades tradicionales no son suficientes para que los estudiantes se enfrenten a situaciones reales que requieran la aplicación efectiva de estas competencias, lo cual impacta negativamente tanto en su rendimiento académico como en su preparación para el mundo laboral. La encuesta también reveló que los estudiantes perciben que existen iniciativas institucionales limitadas para el desarrollo de estas habilidades blandas. Esto se refleja en una baja disponibilidad de talleres, actividades extracurriculares y recursos que permitan potenciar las competencias transversales de los estudiantes. A su vez, se evidenció que los estudiantes se

beneficiarían de una oferta más amplia de actividades, tales como hackáthones, mentorías personalizadas, y programas de liderazgo.

JUSTIFICACIÓN

La realización de una encuesta enfocada en evaluar la importancia, aplicación y barreras relacionadas con la creatividad, la innovación, la autoconciencia y la comunicación tiene un valor significativo para la institución educativa. A través de este estudio se busca obtener información directa de los estudiantes sobre cómo perciben estas competencias y cómo las aplican durante su formación académica. Esta comprensión es clave para el desarrollo de estrategias efectivas que promuevan el aprendizaje integral, más allá de los conocimientos técnicos.

La encuesta permitirá a la institución identificar las necesidades específicas de sus estudiantes, mejorando la calidad de los programas y talleres ofrecidos, y ajustando los métodos pedagógicos para favorecer el desarrollo de estas habilidades esenciales. Asimismo, con esta información se podrán diseñar actividades y recursos que eliminen o mitiguen las barreras al desarrollo de estas competencias, promoviendo un entorno de aprendizaje más dinámico, inclusivo y adaptativo.

En resumen, la encuesta es una herramienta fundamental para diagnosticar el estado actual de las competencias blandas de los estudiantes, identificar brechas y generar iniciativas que enriquezcan el proceso educativo, preparándolos para los desafíos tanto académicos como profesionales. Este proceso no solo mejorará el desempeño y la satisfacción de los estudiantes, sino que también contribuirá a la reputación de la institución como un espacio comprometido con el desarrollo integral de sus alumnos.

OBJETIVOS DE LA MINERÍA DE DATOS

1. **Identificar Patrones y Tendencias en la Aplicación de Habilidades Blandas**
    - Utilizar técnicas de visualización de datos (empleando *seaborn*, *matplotlib.pyplot*) y análisis exploratorio (*pandas*) para identificar patrones en la percepción y aplicación de habilidades blandas (creatividad, innovación, autoconciencia, comunicación) entre los estudiantes. Esto permitirá visualizar cómo varían estas habilidades según el sexo, semestre, edad, y programa académico.
2. **Segmentación de Estudiantes según Niveles de Competencias**
    - Utilizar métodos de *clustering* y escalado (*StandardScaler*, *MinMaxScaler*) para crear segmentos de estudiantes basados en sus niveles de desarrollo de habilidades blandas, frecuencia de aplicación, y percepciones sobre la importancia de estas competencias. Esto ayudará a la institución a identificar grupos de estudiantes con características similares y personalizar actividades de apoyo.
3. **Descubrir Factores que Influyen en el Desarrollo de Competencias**
    - Analizar, mediante técnicas de minería de datos, los factores internos y externos que afectan el desarrollo de habilidades blandas, tales como barreras percibidas y el impacto de ciertos métodos de enseñanza. Se busca determinar correlaciones y relaciones significativas entre diferentes variables del cuestionario.
4. **Predicción del Impacto de Intervenciones Educativas**
    - Implementar técnicas de *machine learning* para predecir cómo intervenciones específicas, como la implementación de talleres o actividades de desarrollo personal, podrían afectar el rendimiento académico y el desarrollo de habilidades de los estudiantes. Esto permitirá diseñar programas más efectivos y orientados a los resultados deseados.
5. **Clasificación de Estudiantes con Necesidades Específicas**
    - Utilizar algoritmos de clasificación (*OneHotEncoder* para preprocesamiento) para identificar a aquellos estudiantes que reportan mayores dificultades en el desarrollo de habilidades blandas, como el

autocontrol o la comunicación, de modo que puedan ser priorizados para programas de apoyo específicos.

6. **Análisis de la Relación entre Rendimiento Académico y Habilidades Blandas**
    - Explorar la relación entre el rendimiento académico reportado y el nivel de habilidades blandas desarrolladas. Esto incluirá técnicas de regresión y visualización para entender cómo diferentes habilidades influyen en los resultados académicos y la satisfacción del estudiante.
7. **Optimizar la Oferta de Talleres y Recursos Formativos**
    - A partir de los resultados obtenidos en el análisis de datos, sugerir la optimización de los recursos y talleres ofertados por la institución, identificando cuáles son los talleres más solicitados o necesarios y cómo pueden ser distribuidos para maximizar su impacto.
8. **Evaluar la Satisfacción y Necesidades de los Estudiantes**
    - Utilizar minería de datos para evaluar la satisfacción de los estudiantes con la oferta actual de actividades formativas. Esto permitirá identificar áreas de mejora en la oferta académica y formular estrategias que se alineen mejor con las expectativas de los estudiantes.

PLAN DE PROYECTO: MINERÍA DE DATOS PARA LA EVALUACIÓN DE COMPETENCIAS BLANDAS EN ESTUDIANTES UNIVERSITARIOS

Objetivos del Proyecto

- **General:** Utilizar técnicas de minería de datos para analizar la percepción y desarrollo de habilidades blandas entre los estudiantes, con el fin de proporcionar una base para decisiones estratégicas y de mejora en la oferta académica de la institución.
- **Específicos:**
    - Identificar patrones de comportamiento y aplicación de habilidades blandas.
    - Determinar factores que influyen en el desarrollo de competencias.
    - Segmentar a los estudiantes para personalizar estrategias de intervención.
    - Evaluar el impacto de métodos y talleres sobre el rendimiento académico.

Alcance del Proyecto

- Analizar los datos proporcionados por la encuesta, enfocándose en las habilidades de creatividad, innovación, autoconciencia y comunicación.
- Utilizar herramientas y librerías como Jupyter, *pandas*, *numpy*, *matplotlib.pyplot*, *seaborn*, *StandardScaler*, *MinMaxScaler*, y *OneHotEncoder*.
- Desarrollar visualizaciones e informes que resuman los hallazgos.
- Proporcionar recomendaciones para la mejora del currículo académico y de los recursos formativos.

Estructura del Proyecto

El proyecto se desarrollará en las siguientes fases:

Fase 1: Planificación y Preparación de Datos

- **Actividades:**
    - Definir las preguntas de negocio que se busca responder.
    - Revisión y limpieza de los datos proporcionados por el docente utilizando *pandas* para detectar y corregir valores faltantes o inconsistencias.
    - Preparación de un entorno de trabajo en Jupyter.

Fase 2: Análisis Exploratorio de Datos

- **Actividades:**
    - Utilizar *pandas* y *numpy* para la exploración inicial de los datos.
    - Generar visualizaciones con *seaborn* y *matplotlib.pyplot* para entender la distribución de las variables y relaciones entre ellas.

Fase 3: Preparación de Datos

- **Actividades:**
    - Aplicar técnicas de preprocesamiento de datos, incluyendo el uso de *StandardScaler* y *MinMaxScaler* para normalización.
    - Codificación de variables categóricas usando *OneHotEncoder*.
- **Entregables:**
    - Conjunto de datos preparado para el análisis predictivo y la segmentación.

Fase 4: Análisis Avanzado y Minería de Datos

- **Actividades:**
    - Implementar técnicas de segmentación (agrupamiento con *k-means* o similares) para crear grupos de estudiantes con características similares.
    - Identificar factores determinantes para el desarrollo de habilidades blandas utilizando técnicas de regresión o clasificación.

Fase 5: Desarrollo de Modelos Predictivos (2 semanas)

- **Actividades:**
    - Crear modelos que predigan cómo cambios en el currículo o en actividades formativas podrían influir en el desarrollo de habilidades blandas.
    - Evaluar la precisión de los modelos desarrollados.

Fase 6: Generación de Visualizaciones e Informes

- **Actividades:**
    - Elaborar visualizaciones que resuman los hallazgos del análisis.
    - Crear un informe detallado con las conclusiones y recomendaciones clave.

Recursos Necesarios

- **Herramientas y Software:**
    - *Jupyter Notebook y Google Colab*: Para realizar el análisis y ejecutar el código.
    - *Python* con librerías: *pandas*, *numpy*, *matplotlib.pyplot*, *seaborn*, *scikit-learn*.
- **Personal:**
    - Analista de Datos: responsable del análisis exploratorio y la minería de datos.
    - Especialista en Minería de Datos: Encargado de los modelos predictivos y el análisis avanzado.
    - Especialista en Visualización: Desarrollará los informes y visualizaciones.

Riesgos y Mitigación

- **Riesgo 1: Calidad de Datos Insuficiente**
    - *Mitigación:* Realizar una limpieza exhaustiva de datos antes del análisis, identificando y corrigiendo valores faltantes o inconsistencias.
- **Riesgo 2: Falta de Conocimiento en Minería de Datos**

    - *Mitigación:* Proporcionar capacitación adicional al personal involucrado en el uso de herramientas específicas como *Jupyter*, *scikit-learn*, y técnicas de escalado y normalización.

- **Riesgo 3: Inconsistencias en el Cronograma**
    - *Mitigación:* Realizar reuniones semanales para revisar el progreso y ajustar el cronograma según sea necesario.

Criterios de Éxito

- Análisis exhaustivo de los datos de la encuesta con visualizaciones claras y conclusiones valiosas.
- Segmentación adecuada de los estudiantes que permita diseñar intervenciones personalizadas.
- Desarrollo de modelos predictivos útiles para la toma de decisiones respecto a actividades formativas.
- Entrega de un informe final con recomendaciones concretas para la mejora del currículo académico y la implementación de actividades adicionales.

COMPRENSION DE LOS DATOS

OBTENCIÓN DEL CONJUNTO INICIAL DE DATOS

El presente estudio se basa en un conjunto de datos proporcionado por el profesor, cuyo propósito es puramente educativo y tiene como objetivo facilitar la práctica de técnicas de análisis y minería de datos. El dataset contiene información recopilada a partir de una encuesta realizada a estudiantes universitarios, en la que se evaluaron competencias blandas clave, tales como creatividad, innovación, autoconciencia y habilidades comunicativas. Estos datos servirán como base para aplicar técnicas de preprocesamiento, análisis exploratorio y modelos predictivos, contribuyendo así al aprendizaje práctico de metodologías de minería de datos.

DESCRIPCION DE LOS DATOS

El conjunto de datos proporcionado para este estudio consta de diversas variables que capturan información relevante sobre las competencias blandas de los estudiantes universitarios. A continuación, se describen todas las variables del data set, sus posibles valores y el tipo de dato que representan:

1. **Nombre del estudiante**
    - **Tipo de dato**: *String*
    - **Descripción**: Nombre completo del estudiante que respondió la encuesta.
2. **Sexo**
    - **Posibles valores**: "Masculino", "Femenino"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Género del estudiante.
3. **Estrato**
    - **Posibles valores**: Números enteros del 1 al 6
    - **Tipo de dato**: *Integer*

- **Descripción**: Estrato socioeconómico del estudiante.

4. **Programa académico**
    - **Tipo de dato**: *String*
    - **Descripción**: Programa académico en el cual está matriculado el estudiante, por ejemplo, "Contaduría Pública", "Administración de Empresas".
5. **Semestre que cursa actualmente**
    - **Posibles valores**: Semestre Del 1 al 10
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Semestre académico que el estudiante está cursando.
6. **Edad**
    - **Posibles valores**: "Entre 16 y 18 años", "Entre 19 y 21 años", "Entre 22 y 24 años", "25 años en adelante"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Rango de edad del estudiante.
7. **1. ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?**
    - **Posibles valores**: "Poco importante", "Importante", "Muy importante"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Percepción del estudiante sobre la importancia de la creatividad e innovación en su desempeño académico.
8. **2. ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?**
    - **Posibles valores**: "Nunca", "Algunas veces", "Casi siempre", "Siempre"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Frecuencia con la que el estudiante aplica habilidades de creatividad e innovación.
9. **3. En una escala de 1 a 5, ¿qué tan creativo e innovador te consideras?**

- **Posibles valores**: 1, 2, 3, 4, 5
- **Tipo de dato**: *Integer*
- **Descripción**: Autoevaluación del nivel de creatividad e innovación del estudiante.

10. **4. ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?**
    - **Posibles valores**: "Positivamente", "Neutro", "Negativamente"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Opinión del estudiante sobre cómo su creatividad e innovación han influido en su rendimiento académico.
11. **5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?**
    - **Tipo de dato**: *String*
    - **Descripción**: Espacios donde el estudiante ha aplicado su creatividad e innovación.
12. **6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?**
    - **Tipo de dato**: *String*
    - **Descripción**: Barreras percibidas por el estudiante que limitan su capacidad creativa.
13. **7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?**
    - **Tipo de dato**: *String*
    - **Descripción**: Actividades que el estudiante cree que podrían ayudar a mejorar sus habilidades creativas.
14. **8. ¿Qué tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes?**
    - **Posibles valores**: "Poco necesario", "Necesario", "Muy necesario"
    - **Tipo de dato**: *Categorical (String)*

- **Descripción**: Opinión del estudiante sobre la necesidad de talleres para el desarrollo de la creatividad e innovación.

15. **1. ¿Qué tan importante consideras la autoconciencia en tu desempeño académico y profesional?**
    - **Posibles valores**: "Poco importante", "Importante", "Muy importante"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Opinión sobre la importancia de la autoconciencia (autoconocimiento, autocontrol, automotivación, autoimagen) en su desempeño.
16. **2. ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?**
    - **Posibles valores**: "Nunca", "Algunas veces", "Casi siempre", "Siempre"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Frecuencia con la que el estudiante aplica la autoconciencia.
17. **3. En una escala del 1 al 5, califica los aspectos relacionados con tu autoconciencia [Autoconocimiento]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de autoconocimiento del estudiante.
18. **3. En una escala del 1 al 5, califica los aspectos relacionados con tu autoconciencia [Autocontrol]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de autocontrol del estudiante.
19. **3. En una escala del 1 al 5, califica los aspectos relacionados con tu autoconciencia [Automotivación]**
    - **Posibles valores**: 1, 2, 3, 4, 5

- **Tipo de dato**: *Integer*
- **Descripción**: Autoevaluación del nivel de automotivación del estudiante.

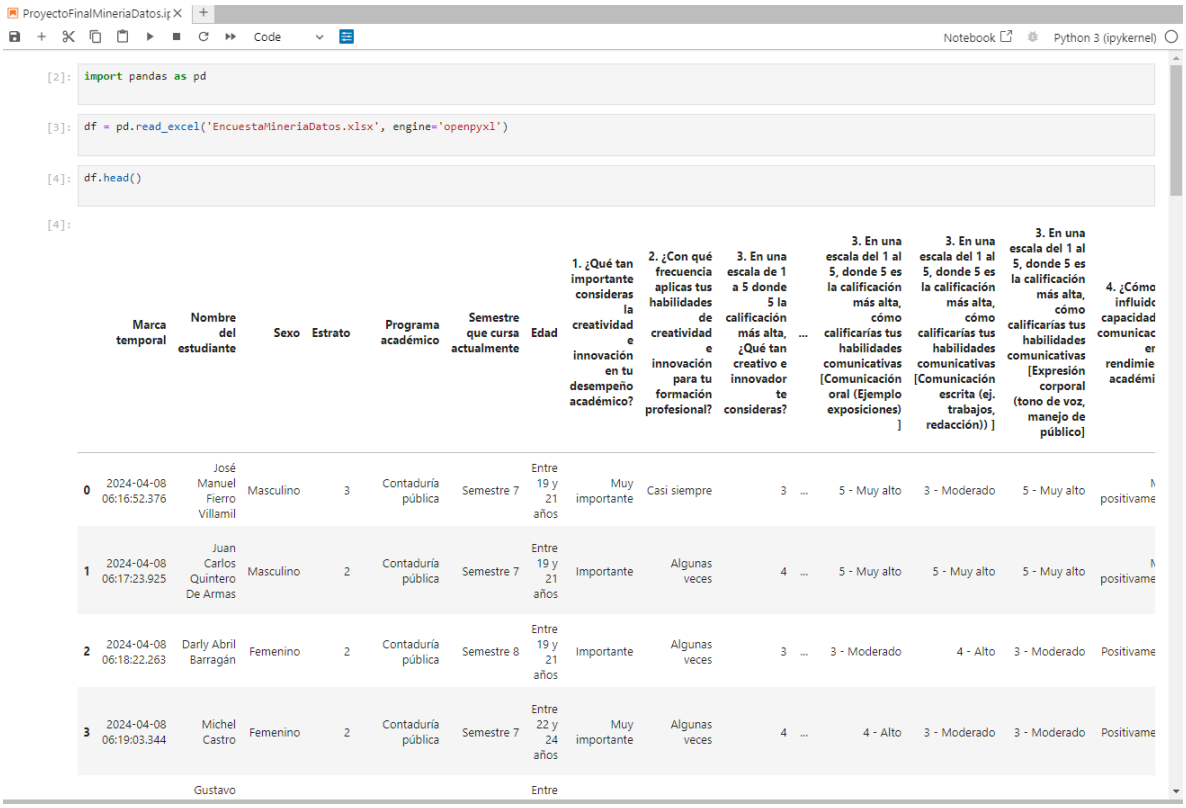
20. **3. En una escala del 1 al 5, califica los aspectos relacionados con tu autoconciencia [Autoimagen]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de autoimagen del estudiante.
21. **4. ¿Cómo consideras que ha influenciado tu nivel de autoconciencia en tu rendimiento académico?**
    - **Posibles valores**: "Positivamente", "Neutro", "Negativamente"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Opinión sobre la influencia de la autoconciencia en el rendimiento académico.
22. **5. ¿En qué momentos de tu formación crees que más has aplicado tu autoconciencia?**
    - **Tipo de dato**: *String*
    - **Descripción**: Momentos específicos donde el estudiante ha aplicado la autoconciencia.
23. **6. ¿Qué obstáculos o situaciones crees que pueden limitar tu capacidad de autoconciencia?**
    - **Tipo de dato**: *String*
    - **Descripción**: Obstáculos percibidos para desarrollar la autoconciencia.
24. **7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu nivel de autoconciencia en tu formación profesional?**
    - **Tipo de dato**: *String*
    - **Descripción**: Actividades que podrían ayudar a mejorar la autoconciencia, como talleres de desarrollo personal.

25. **1. ¿Qué tan importante consideras la comunicación (oral, escrita y expresión corporal) en tu desempeño académico y profesional?**
    - **Posibles valores**: "Poco importante", "Importante", "Muy importante"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Opinión del estudiante sobre la importancia de la comunicación.
26. **2. ¿Con qué frecuencia aplicas la comunicación oral, escrita y expresión corporal en tu formación profesional?**
    - **Posibles valores**: "Nunca", "Algunas veces", "Casi siempre", "Siempre"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Frecuencia de aplicación de habilidades comunicativas.
27. **3. En una escala del 1 al 5, califica tus habilidades comunicativas [Comunicación oral]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de habilidad en comunicación oral (e.g., exposiciones).
28. **3. En una escala del 1 al 5, califica tus habilidades comunicativas [Comunicación escrita]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de habilidad en comunicación escrita (e.g., redacción de trabajos).
29. **3. En una escala del 1 al 5, califica tus habilidades comunicativas [Expresión corporal]**
    - **Posibles valores**: 1, 2, 3, 4, 5
    - **Tipo de dato**: *Integer*
    - **Descripción**: Autoevaluación del nivel de expresión corporal (e.g., tono de voz, manejo del público).

30. **4. ¿Cómo ha influido tu capacidad de comunicación en tu rendimiento académico?**
    - **Posibles valores**: "Positivamente", "Neutro", "Negativamente"
    - **Tipo de dato**: *Categorical (String)*
    - **Descripción**: Opinión sobre la influencia de la capacidad de comunicación en el rendimiento académico.
31. **5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de comunicación con mayor frecuencia?**
    - **Tipo de dato**: *String*
    - **Descripción**: Espacios donde el estudiante ha aplicado sus habilidades comunicativas.
32. **6. ¿Qué obstáculos o situaciones crees que pueden limitar tus habilidades de comunicación durante tu formación profesional?**
    - **Tipo de dato**: *String*
    - **Descripción**: Obstáculos percibidos para mejorar las habilidades comunicativas.
33. **7. ¿Qué actividades consideras que pueden ayudarte a mejorar tus habilidades de comunicación en tu formación profesional?**
    - **Tipo de dato**: *String*
    - **Descripción**: Actividades que podrían ayudar a mejorar las habilidades comunicativas, como talleres de oratoria.
34. **Talleres que le gustaría realizar para fortalecer sus habilidades de poder**
    - **Tipo de dato**: *String*
    - **Descripción**: Preferencia del estudiante sobre los talleres específicos que le gustaría realizar para mejorar sus habilidades.

EXPLORACION DEL CONJUNTO DE DATOS

Importamos la librería pandas por medio del comando **import pandas as pd** y cargamos el data Set al notebook por medio del comando: **df = pd.read_excel('EncuestaMineriaDatos.xlsx', engine='openpyxl')**



ProyectoFinalMineriaDatos.ip

Code

Notebook

Python 3 (ipykernel)

```
[2]: import pandas as pd
```

```
[3]: df = pd.read_excel('EncuestaMineriaDatos.xlsx', engine='openpyxl')
```

```
[4]: df.head()
```

[4]:

	Marca temporal	Nombre del estudiante	Sexo	Estrato	Programa académico	Semestre que cursa actualmente	Edad	1. ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?	2. ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?	3. En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?	...	3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas [Comunicación oral (Ejemplo exposiciones)]	3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas [Comunicación escrita (ej. trabajos, redacción))]	3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas [Expresión corporal (tono de voz, manejo de público]	4. ¿Cómo influido capacidad comunicac er rendimie académi
0	2024-04-08 06:16:52.376	José Manuel Fierro Villamil	Masculino	3	Contaduría pública	Semestre 7	Entre 19 y 21 años	Muy importante	Casi siempre	3	...	5 - Muy alto	3 - Moderado	5 - Muy alto	N positivame
1	2024-04-08 06:17:23.925	Juan Carlos Quintero De Armas	Masculino	2	Contaduría pública	Semestre 7	Entre 19 y 21 años	Importante	Algunas veces	4	...	5 - Muy alto	5 - Muy alto	5 - Muy alto	N positivame
2	2024-04-08 06:18:22.263	Darly Abril Barragán	Femenino	2	Contaduría pública	Semestre 8	Entre 19 y 21 años	Importante	Algunas veces	3	...	3 - Moderado	4 - Alto	3 - Moderado	Positivame
3	2024-04-08 06:19:03.344	Michel Castro	Femenino	2	Contaduría pública	Semestre 7	Entre 22 y 24 años	Muy importante	Algunas veces	4	...	4 - Alto	3 - Moderado	3 - Moderado	Positivame
		Gustavo					Entre								

Visualizamos nuestro Data Set por medio del comando **df.head()** y hacemos la exploración de nuestra base de datos por medio de comandos como **df.info()**

```
[5]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 626 entries, 0 to 625
Data columns (total 37 columns):
 #   Column
Non-Null Count  Dtype
---  ------
--------------  -----
 0   Marca temporal
626 non-null    datetime64[ns]
 1   Nombre del estudiante
626 non-null    object
 2   Sexo
626 non-null    object
 3   Estrato
626 non-null    int64
 4   Programa académico
626 non-null    object
 5   Semestre que cursa actualmente
626 non-null    object
 6   Edad
626 non-null    object
 7   1.     ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?
626 non-null    object
 8   2.     ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?
626 non-null    object
 9   3.     En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?
626 non-null    int64
 10  4.     ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?
626 non-null    object
 11  5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?
626 non-null    object
 12  6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?
626 non-null    object
 13  7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?
626 non-null    object
 14  8.     Que tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes
626 non-null    object
 15  1. ¿Qué tan importante consideras la autoconciencia (autoconocimiento, autocontrol, automotivación, autoimagen) en tu desempeño académico y profesional?
626 non-null    object
 16  2. ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?
626 non-null    object
 17  3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoconocimiento     ]
626 non-null    object
 18  3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autocontrol  ]
626 non-null    object
 19  3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Automotivación      ]
```

Revisamos si dentro de nuestro Data Set existen campos vacíos por medio del siguiente comando: **df.isnull().sum()**

```
[6]: df.isnull().sum()

[6]: Marca temporal
0
Nombre del estudiante
0
Sexo
0
Estrato
0
Programa académico
0
Semestre que cursa actualmente
0
Edad
0
1.     ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?
0
2.     ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?
0
3.     En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?
0
4.     ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?
0
5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?
0
6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?
0
7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?
0
8.     Que tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes
0
1. ¿Qué tan importante consideras la autoconciencia (autoconocimiento, autocontrol, automotivación, autoimagen) en tu desempeño académico y profesional?
0
2. ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?
0
3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoconocimiento\t]
0
3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autocontrol\t]
0
3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Automotivación\t]
0
3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoimagen]
0
4. ¿Cómo consideras que ha influenciado tu nivel de autoconciencia en tu rendimiento académico?
0
5. ¿En qué momentos de tu formación crees que más has aplicado tu autoconciencia?
0
```

Por medio del siguiente comando **df.isna().sum().sum()** porcedemos a identificar la cantidad de valores nulos que encontramos en nuestro Data Set

```
df.isna().sum().sum()
```

```
np.int64(1284)
```

VERIFICACION DE LA CALIDAD DE LOS DATOS

Utilizar el comando **null_values = df.isna().sum() print(null_values)** para conocer la distribución de los valores nulos

```
null_values = df.isna().sum()
print(null_values)
```

```
Marca temporal
0
Nombre del estudiante
0
Sexo
0
Estrato
0
Programa académico
0
Semestre que cursa actualmente
0
Edad
0
1.     ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?
0
2.     ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?
0
3.     En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?
0
4.     ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?
0
5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?
0
6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?
0
7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?
0
8.     Que tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes
```

Mediante el siguiente código **print(df.columns)** verificamos el nombre de las columnas para proceder a cambiarlos por unos nombre mas cortos

```
print(df.columns)
```

```
Index(['Marca temporal', 'Nombre del estudiante', 'Sexo', 'Estrato ',
       'Programa académico', 'Semestre que cursa actualmente', 'Edad ',
       '1.     ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?',
       '2.     ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?',
       '3.     En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?',
       '4.     ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?',
       '5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?',
       '6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?',
       '8.     Que tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes',
       '1. ¿Qué tan importante consideras la autoconciencia (autoconocimiento, autocontrol, automotivación, autoimagen) en tu desempeño académico y profesional?',
       '2. ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoconocimiento\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autocontrol\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Automotivación\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoimagen]',
       '4. ¿Cómo consideras que ha influenciado tu nivel de autoconciencia en tu rendimiento académico?',
       '5. ¿En qué momentos de tu formación crees que más has aplicado tu autoconciencia?',
       '6. ¿Qué obstáculos o situaciones crees que pueden limitar tu capacidad de autoconciencia?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu nivel de autoconciencia en tu formación profesional?',
       '1. ¿Qué tan importante consideras la comunicación (oral, escrita y expresión corporal) en tu desempeño académico y profesional?',
       '2. ¿Con qué frecuencia aplicas la comunicación oral, escrita y expresión corporal, en tu formación profesional?',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación oral (Ejemplo exposiciones)
]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación escrita (ej. trabajos, redacc
ión)) ]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Expresión corporal (tono de voz, manejo de
público]',
       '4. ¿Cómo ha influido tu capacidad de comunicación en tu rendimiento académico?',
       '5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de comunicación con  mayor frecuencia?',
       '6. ¿Qué obstáculos o situaciones crees que pueden limitar tus habilidades de comunicación oral, escrita y expresión corporal durante tu formación profesiona
l?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tus habilidades de comunicación en tu formación profesional?',
       'Cuáles de los siguientes talleres le gustaría realizar para fortalecer sus habilidades de poder ',
       'Hipotesis', 'Unnamed: 36'],
      dtype='object')
```

Por medio del siguiente comando procedemos a eliminar dos columnas (**'Marca temporal')** que para nosotros son innecesarias ya que una de ellas entrega el tiempo en el que se contestó la encuesta y la otra (**'Hipotesis',**) es una columna que tiene muy pocas respuestas **df = df.drop(columns=['Hipotesis', 'Marca temporal']), print(df.columns)**, Luego verificamos el data Set nuevamente **print(df.columns)**

```
df = df.drop(columns=['Hipotesis', 'Marca temporal'])
```

```
print(df.columns)
```

```
Index(['Nombre del estudiante', 'Sexo', 'Estrato ', 'Programa académico',
       'Semestre que cursa actualmente', 'Edad ',
       '1.     ¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?',
       '2.     ¿Con qué frecuencia aplicas tus habilidades de creatividad e innovación para tu formación profesional?',
       '3.     En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?',
       '4.     ¿Cómo ha influido tu capacidad creativa e innovadora en tu rendimiento académico?',
       '5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de creatividad e innovación?',
       '6. ¿Por qué motivos o situaciones consideras que tu creatividad e innovación se puede ver obstaculizada?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu creatividad e innovación?',
       '8.     Que tan necesario consideras que se ofrezcan talleres y actividades para el desarrollo de la creatividad e innovación para los estudiantes',
       '1. ¿Qué tan importante consideras la autoconciencia (autoconocimiento, autocontrol, automotivación, autoimagen) en tu desempeño académico y profesional?',
       '2. ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoconocimiento\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autocontrol\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Automotivación\t]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos  relacionados con tu autoconciencia  [Autoimagen]',
       '4. ¿Cómo consideras que ha influenciado tu nivel de autoconciencia en tu rendimiento académico?',
       '5. ¿En qué momentos de tu formación crees que más has aplicado tu autoconciencia?',
       '6. ¿Qué obstáculos o situaciones crees que pueden limitar tu capacidad de autoconciencia?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tu nivel de autoconciencia en tu formación profesional?',
       '1. ¿Qué tan importante consideras la comunicación (oral, escrita y expresión corporal) en tu desempeño académico y profesional?',
       '2. ¿Con qué frecuencia aplicas la comunicación oral, escrita y expresión corporal, en tu formación profesional?',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación oral (Ejemplo exposiciones)
]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación escrita (ej. trabajos, redacc
ión)) ]',
       '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Expresión corporal (tono de voz, manejo de
público]',
       '4. ¿Cómo ha influido tu capacidad de comunicación en tu rendimiento académico?',
       '5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de comunicación con  mayor frecuencia?',
       '6. ¿Qué obstáculos o situaciones crees que pueden limitar tus habilidades de comunicación oral, escrita y expresión corporal durante tu formación profesiona
l?',
       '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tus habilidades de comunicación en tu formación profesional?',
       'Cuáles de los siguientes talleres le gustaría realizar para fortalecer sus habilidades de poder ',
       'Unnamed: 36'],
      dtype='object')
```

PREPARACION DE LOS DATOS

SELECCION DE LOS DATOS

Para poder empezar con la selección de los datos cambiamos el nombre de las columnas para que sean más cortos y menos tediosos de entender por medio del siguiente comando

```
# Lista de columnas actuales a renombrar
columnas_actuales_com = [
    '1. ¿Qué tan importante consideras la comunicación (oral, escrita y expresión corporal) en tu desempeño académico y profesional?',
    '2. ¿Con qué frecuencia aplicas la comunicación oral, escrita y expresión corporal, en tu formación profesional?',
    '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación oral (Ejemplo exposiciones) ]',
    '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Comunicación escrita (ej. trabajos, redacción)
    '3. En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas  [Expresión corporal (tono de voz, manejo de púb
    '4. ¿Cómo ha influido tu capacidad de comunicación en tu rendimiento académico?',
    '5. ¿En qué espacios de tu formación has requerido aplicar tus habilidades de comunicación con  mayor frecuencia?',
    '6. ¿Qué obstáculos o situaciones crees que pueden limitar tus habilidades de comunicación oral, escrita y expresión corporal durante tu formación profesional?',
    '7. ¿Qué actividades consideras que pueden ayudarte a mejorar tus habilidades de comunicación en tu formación profesional?',
    'Cuáles de los siguientes talleres le gustaría realizar para fortalecer sus habilidades de poder '
]

# Crear un diccionario que asocia cada columna actual con su nuevo nombre (COM1, COM2, COM3_1, ..., COM3_3, COM4, ...)
nuevos_nombres_com = {
    columnas_actuales_com[0]: 'COM1',
    columnas_actuales_com[1]: 'COM2',
    columnas_actuales_com[2]: 'COM3_1',  # Primera ocurrencia de la pregunta 3
    columnas_actuales_com[3]: 'COM3_2',  # Segunda ocurrencia de la pregunta 3
    columnas_actuales_com[4]: 'COM3_3',  # Tercera ocurrencia de la pregunta 3
    columnas_actuales_com[5]: 'COM4',
    columnas_actuales_com[6]: 'COM5',
    columnas_actuales_com[7]: 'COM6',
    columnas_actuales_com[8]: 'COM7',
    columnas_actuales_com[9]: 'COM8'
}

# Renombrar las columnas en el DataFrame
df.rename(columns=nuevos_nombres_com, inplace=True)

# Verificar los nuevos nombres
print(df.columns)
```

Podemos observar como la tabla se ve más legible al reducir los nombres de las columnas

```
df.head()
```

	Nombre	Sexo	Estrato	Carrera	Semestre	Edad	CI1	CI2	CI3	CI4	...	COM1	COM2	COM3_1	COM3_2	COM3_3	COM4	
0	José Manuel Fierro Villamil	Masculino	3	Contaduría pública	Semestre 7	Entre 19 y 21 años	Muy importante	Casi siempre	3	Positivamente	...	Muy importante	Siempre	5 - Muy alto	3 - Moderado	5 - Muy alto	Muy positivamente	prese exp
1	Juan Carlos Quintero De Armas	Masculino	2	Contaduría pública	Semestre 7	Entre 19 y 21 años	Importante	Algunas veces	4	Neutro	...	Muy importante	Siempre	5 - Muy alto	5 - Muy alto	5 - Muy alto	Muy positivamente	ir com prc
2	Darly Abril Barragán	Femenino	2	Contaduría pública	Semestre 8	Entre 19 y 21 años	Importante	Algunas veces	3	Neutro	...	Importante	Casi siempre	3 - Moderado	4 - Alto	3 - Moderado	Positivamente	prese exp
3	Michel Castro	Femenino	2	Contaduría pública	Semestre 7	Entre 22 y 24 años	Muy importante	Algunas veces	4	Positivamente	...	Muy importante	Algunas veces	4 - Alto	3 - Moderado	3 - Moderado	Positivamente	A do
4	Gustavo Andrés cruz romero	Masculino	5	Administración de Empresas	Semestre 3	Entre 19 y 21 años	Importante	Siempre	4	Positivamente	...	Muy importante	Siempre	4 - Alto	4 - Alto	4 - Alto	Muy positivamente	prese exp

5 rows × 34 columns

Una vez tenemos todo más organizado podemos empezar a sacar las primeras afirmaciones en base a la data Set que estamos trabajando como por ejemplo:

Usando el comando **conteo_sexo = df['Sexo'].value_counts()** podemos observar que en este caso de las 625 personas que contestaron la encuesta 363 Fueron hombres, 262 Fueron mujeres y 1 persona prefirió no decirlo.

```
# Contar la cantidad de respuestas de hombres y mujeres
conteo_sexo = df['Sexo'].value_counts()

# Mostrar el resultado
print(conteo_sexo)
```

```
Sexo
Femenino               363
Masculino              262
Prefiero no decirlo      1
Name: count, dtype: int64
```

De igual forma con el comando **conteo_edad = df['Edad '].value_counts()** podemos observar los rangos de edad para los estudiantes que contestaron la encuesta

```
# Contar la cantidad de respuestas por cada categoría de edad
conteo_edad = df['Edad '].value_counts()

# Mostrar el resultado
print(conteo_edad)
```

```
Edad
Entre 19 y 21 años     250
Entre 16 y 18 años     195
Entre 22 y 24 años      95
25 años en adelante     86
Name: count, dtype: int64
```

De igual forma podemos observar cuantos estudiantes pertenecen a cada semestre con el comando **conteo_semestre = df['Semestre'].value_counts()**

```
# Contar la cantidad de respuestas por cada semestre
conteo_semestre = df['Semestre'].value_counts()

# Mostrar el resultado
print(conteo_semestre)
```

```
Semestre
Semestre 1     156
Semestre 3      97
Semestre 5      95
Semestre 6      74
Semestre 7      72
Semestre 2      37
Semestre 8      26
Semestre 4      23
Semestre 10     23
Semestre 9      23
Name: count, dtype: int64
```

Y lo mismo para observar de que carreras son mediante el comando **conteo_carrera = df['Carrera'].value_counts()**

```
# Contar la cantidad de respuestas por cada carrera
conteo_carrera = df['Carrera'].value_counts()

# Mostrar el resultado
print(conteo_carrera)
```

```
Carrera
Medicina Veterinaria y Zootecnia    186
Contaduría pública                  148
Derecho                              95
Administración de Empresas           84
Ingeniería Civil                     66
Ingeniería de Sistemas               47
Name: count, dtype: int64
```

Para cada una de las preguntas podríamos por medio del comando **.value_counts()** saber cuántas personas escogieron entre las opciones existentes como por Ejemplo para la primera pregunta que dice “¿Qué tan importante consideras la creatividad e innovación en tu desempeño académico?” identificada en la tabla como **CI1** tenemos el siguiente resultado

```
print(conteo_ci1)
```

```
CI1
Muy importante     414
Importante         197
Poco importante     15
Name: count, dtype: int64
```

O si queremos la respuesta en terminos de porcentaje usamos el comando **porcentaje_ci1 = df['CI1'].value_counts(normalize=True) * 100**

```
# Contar la cantidad de respuestas por cada carrera en porcentaje
porcentaje_ci1 = df['CI1'].value_counts(normalize=True) * 100

# Mostrar el resultado en porcentaje
print(porcentaje_ci1)
```

```
CI1
Muy importante     66.134185
Importante         31.469649
Poco importante     2.396166
Name: proportion, dtype: float64
```

Y podemos concluir que para el 66% de los estudiantes encuestados la creatividad e innovacion en el desempeño academico es muy importante, para el 31% es importante y para el 2% es poco importante.

TRANSFORMACION DE LOS DATOS

Validamos por medio del comando **df.info()** que tenemos variables categóricas y necesitamos pasar todo a numérico

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 626 entries, 0 to 625
Data columns (total 32 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Sexo      626 non-null    object
 1   Estrato   626 non-null    int64
 2   Carrera   626 non-null    object
 3   Semestre  626 non-null    object
 4   Edad      626 non-null    object
 5   CI1       626 non-null    object
 6   CI2       626 non-null    object
 7   CI3       626 non-null    int64
 8   CI4       626 non-null    object
 9   CI5       626 non-null    object
 10  CI6       626 non-null    object
 11  CI7       626 non-null    object
 12  CI8       626 non-null    object
 13  AC1       626 non-null    object
 14  AC2       626 non-null    object
 15  AC3_1     626 non-null    object
 16  AC3_2     626 non-null    object
 17  AC3_3     626 non-null    object
 18  AC3_4     626 non-null    object
 19  AC4       626 non-null    object
 20  AC5       626 non-null    object
 21  AC6       626 non-null    object
 22  AC7       626 non-null    object
 23  COM1      626 non-null    object
 24  COM2      626 non-null    object
 25  COM3_1    626 non-null    object
 26  COM3_2    626 non-null    object
 27  COM3_3    626 non-null    object
 28  COM4      626 non-null    object
 29  COM5      626 non-null    object
 30  COM6      626 non-null    object
 31  COM7      626 non-null    object
```

Procedemos a importar el **from sklearn.preprocessing import OrdinalEncoder** para poder utilizar el siguiente comando **encoder = OrdinalEncoder()** y asi pasamos todas nuestras variables categoricas a tipo int64

```
Sexo        int64
Estrato     int64
Carrera     int64
Semestre    int64
Edad        int64
CI1         int64
CI2         int64
CI3         int64
CI4         int64
CI8         int64
AC1         int64
AC2         int64
AC3_1       int64
AC3_2       int64
AC3_3       int64
AC3_4       int64
AC4         int64
COM1        int64
COM2        int64
COM3_1      int64
COM3_2      int64
COM3_3      int64
COM4        int64
dtype: object
   Sexo  Estrato   Carrera  Semestre  Edad  CI1  CI2  CI3  CI4  CI8  ...  \
0     1         3        1         7     2    1    2    3    4    2  ...
1     1         2        1         7     2    0    0    4    3    0  ...
2     0         2        1         8     2    0    0    3    3    2  ...
3     0         2        1         7     3    1    0    4    4    3  ...
4     1         5        0         3     2    0    4    4    4    0  ...

   AC3_2  AC3_3  AC3_4  AC4  COM1  COM2  COM3_1  COM3_2  COM3_3  COM4
0      2      2      3    4     1     4       4       2       4     1
1      4      4      4    1     1     4       4       4       4     1
2      3      3      3    3     0     2       2       3       2     4
3      2      3      3    4     1     0       3       2       2     4
4      0      2      2    1     1     4       3       3       3     1

[5 rows x 23 columns]
```

```
df.head()
```

	Sexo	Estrato	Carrera	Semestre	Edad	CI1	CI2	CI3	CI4	CI8	...	AC3_2	AC3_3	AC3_4	AC4	COM1	COM2	COM3_1	COM3_2	COM3_3	COM4
0	1	3	1	7	2	1	2	3	4	2	...	2	2	3	4	1	4	4	2	4	1
1	1	2	1	7	2	0	0	4	3	0	...	4	4	4	1	1	4	4	4	4	1
2	0	2	1	8	2	0	0	3	3	2	...	3	3	3	3	0	2	2	3	2	4
3	0	2	1	7	3	1	0	4	4	3	...	2	3	3	4	1	0	3	2	2	4
4	1	5	0	3	2	0	4	4	4	0	...	0	2	2	1	1	4	3	3	3	1

```python
from sklearn.preprocessing import OrdinalEncoder
```

```python
columnas_categoricas = [
    'CI1', 'CI2', 'CI4', 'CI8',
    'AC1', 'AC2', 'AC3_1', 'AC3_2', 'AC3_3', 'AC3_4', 'AC4',
    'COM1', 'COM2', 'COM3_1', 'COM3_2', 'COM3_3', 'COM4'
]

# Crear una instancia de OrdinalEncoder
encoder = OrdinalEncoder()

# Aplicar el encoder a las columnas categóricas
df[columnas_categoricas] = encoder.fit_transform(df[columnas_categoricas])

# Convertir las columnas codificadas a int64
df[columnas_categoricas] = df[columnas_categoricas].astype('int64')

# Mostrar el DataFrame resultante para verificar la conversión
print(df.dtypes)
print(df.head())
```

HOMOGENIZACIÓN DE LOS DATOS

Para estandarizar los datos podemos utilizar cada uno de los escaladores que tienen características distintas. Dependiendo de la distribución de los datos, algunos pueden ser más útiles que otros. **from sklearn.preprocessing import StandardScaler, MinMaxScaler, MaxAbsScaler, RobustScaler**

```python
from sklearn.preprocessing import StandardScaler, MinMaxScaler, MaxAbsScaler, RobustScaler

# Crear instancias de los diferentes escaladores
scalers = {
    'StandardScaler': StandardScaler(),
    'MinMaxScaler': MinMaxScaler(),
    'MaxAbsScaler': MaxAbsScaler(),
    'RobustScaler': RobustScaler()
}

# Seleccionar columnas numéricas que tengan sentido estandarizar
columnas_numericas = ['Edad', 'Semestre']

# Aplicar los diferentes escaladores a las columnas numéricas y almacenar los resultados en un diccionario
escalados = {}
for nombre, scaler in scalers.items():
    df_escalado = df.copy()  # Crear una copia del DataFrame original
    df_escalado[columnas_numericas] = scaler.fit_transform(df_escalado[columnas_numericas])
    escalados[nombre] = df_escalado

# Mostrar las primeras filas de los DataFrames estandarizados
for nombre, df_escalado in escalados.items():
    print(f'\n{nombre}:')
    print(df_escalado.head())

# Graficar los resultados de la estandarización utilizando StandardScaler y MinMaxScaler
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))

# Graficar distribuciones de 'Edad' después de aplicar StandardScaler
plt.subplot(2, 1, 1)
sns.histplot(escalados['StandardScaler']['Edad'], kde=True, color='blue')
plt.title('Distribución de Edad con StandardScaler')

# Graficar distribuciones de 'Edad' después de aplicar MinMaxScaler
plt.subplot(2, 1, 2)
sns.histplot(escalados['MinMaxScaler']['Edad'], kde=True, color='green')
plt.title('Distribución de Edad con MinMaxScaler')

plt.tight_layout()
plt.show()
```

```
StandardScaler:
   Sexo  Estrato   Carrera  Semestre      Edad  CI1  CI2  CI3  CI4  CI8  ...
0     1        3        1  1.018352 -0.115042    1    2    3    4    2  ...
1     1        2        1  1.018352 -0.115042    0    0    4    3    0  ...
2     0        2        1  1.393125 -0.115042    0    0    3    3    2  ...
3     0        2        1  1.018352  0.885183    1    0    4    4    3  ...
4     1        5        0 -0.480739 -0.115042    0    4    4    4    0  ...

   AC3_2  AC3_3  AC3_4  AC4  COM1  COM2  COM3_1  COM3_2  COM3_3  COM4
0      2      2      3    4     1     4       4       2       4     1
1      4      4      4    1     1     4       4       4       4     1
2      3      3      3    3     0     2       2       3       2     4
3      2      3      3    4     1     0       3       2       2     4
4      0      2      2    1     1     4       3       3       3     1
```

MinMaxScaler:

	Sexo	Estrato	Carrera	Semestre	Edad	CI1	CI2	CI3	CI4	CI8	...
0	1	3	1	0.666667	0.333333	1	2	3	4	2	...
1	1	2	1	0.666667	0.333333	0	0	4	3	0	...
2	0	2	1	0.777778	0.333333	0	0	3	3	2	...
3	0	2	1	0.666667	0.666667	1	0	4	4	3	...
4	1	5	0	0.222222	0.333333	0	4	4	4	0	...

	AC3_2	AC3_3	AC3_4	AC4	COM1	COM2	COM3_1	COM3_2	COM3_3	COM4
0	2	2	3	4	1	4	4	2	4	1
1	4	4	4	1	1	4	4	4	4	1
2	3	3	3	3	0	2	2	3	2	4
3	2	3	3	4	1	0	3	2	2	4
4	0	2	2	1	1	4	3	3	3	1

MaxAbsScaler:

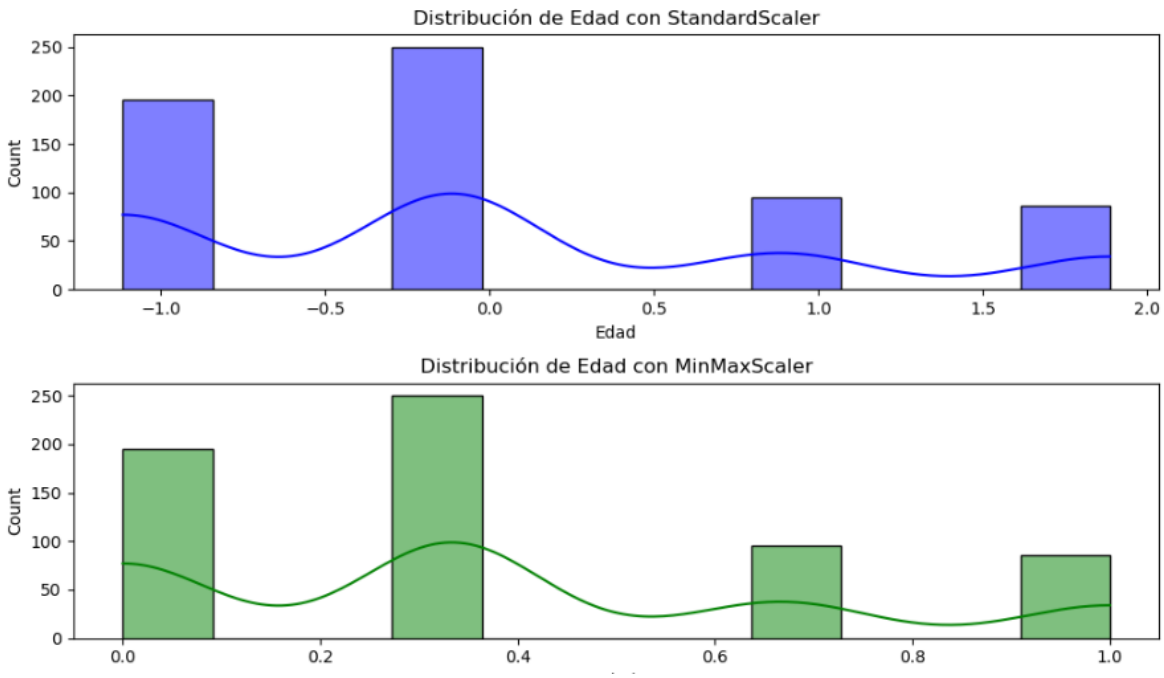
	Sexo	Estrato	Carrera	Semestre	Edad	CI1	CI2	CI3	CI4	CI8	...	\
0	1	3	1	0.7	0.50	1	2	3	4	2	...	
1	1	2	1	0.7	0.50	0	0	4	3	0	...	
2	0	2	1	0.8	0.50	0	0	3	3	2	...	
3	0	2	1	0.7	0.75	1	0	4	4	3	...	
4	1	5	0	0.3	0.50	0	4	4	4	0	...	

	AC3_2	AC3_3	AC3_4	AC4	COM1	COM2	COM3_1	COM3_2	COM3_3	COM4
0	2	2	3	4	1	4	4	2	4	1
1	4	4	4	1	1	4	4	4	4	1
2	3	3	3	3	0	2	2	3	2	4
3	2	3	3	4	1	0	3	2	2	4
4	0	2	2	1	1	4	3	3	3	1

RobustScaler:

	Sexo	Estrato	Carrera	Semestre	Edad	CI1	CI2	CI3	CI4	CI8	...	\
0	1	3	1	0.625	0.0	1	2	3	4	2	...	
1	1	2	1	0.625	0.0	0	0	4	3	0	...	
2	0	2	1	0.875	0.0	0	0	3	3	2	...	
3	0	2	1	0.625	0.5	1	0	4	4	3	...	
4	1	5	0	-0.375	0.0	0	4	4	4	0	...	

	AC3_2	AC3_3	AC3_4	AC4	COM1	COM2	COM3_1	COM3_2	COM3_3	COM4
0	2	2	3	4	1	4	4	2	4	1
1	4	4	4	1	1	4	4	4	4	1
2	3	3	3	3	0	2	2	3	2	4
3	2	3	3	4	1	0	3	2	2	4
4	0	2	2	1	1	4	3	3	3	1



Aplicación de Escaladores: Se aplica cada escalador a las columnas numéricas `Edad` y `Semestre`, donde la estandarización tiene sentido.

Gráficas: Se visualizan las distribuciones de la columna `Edad` con `StandardScaler` y `MinMaxScaler`. Esto te dará una idea de cómo cambia la distribución de los datos con estos métodos.

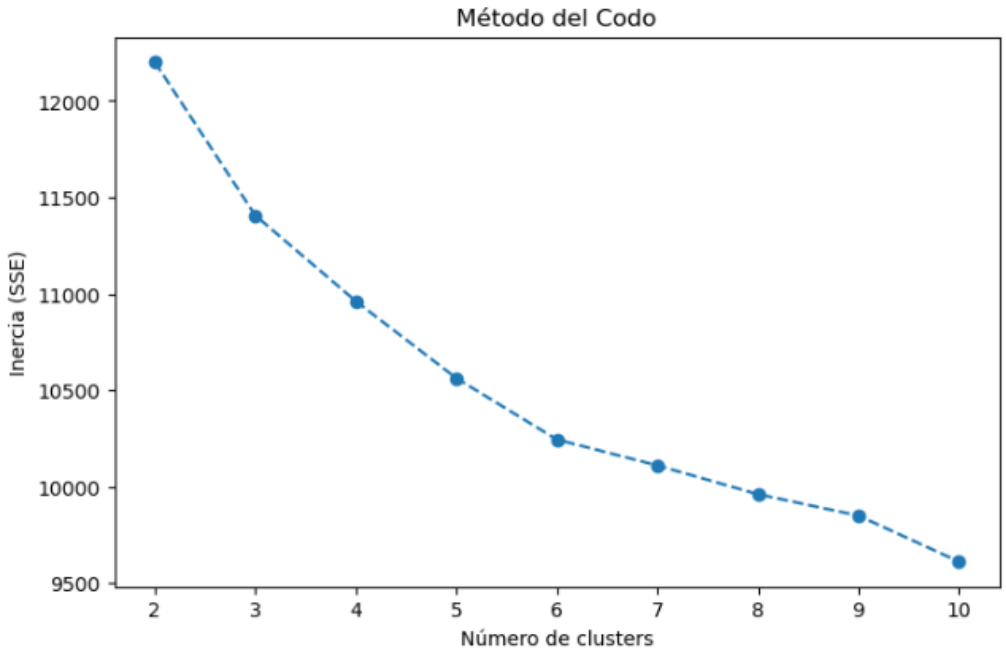
MODELAMIENTO

SELECCIONAR TECNICAS DE MODELO

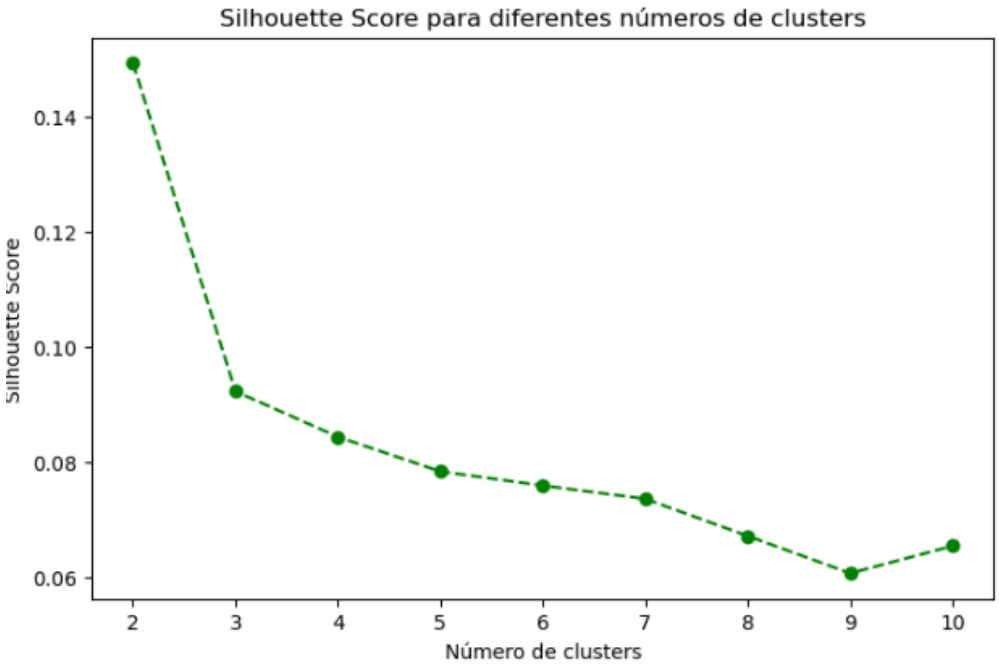
Determinación del número de clusters con el método del codo y silhouette score

```python
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
import numpy as np
import matplotlib.pyplot as plt

# Escalamos los datos con StandardScaler
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df.select_dtypes(include=[np.number]))

# Método del Codo
sse = []
for k in range(2, 11):  # Evaluamos para k de 2 a 10 clusters
    kmeans = KMeans(n_clusters=k, random_state=42)
    kmeans.fit(df_scaled)
    sse.append(kmeans.inertia_)

# Graficar el método del Codo
plt.figure(figsize=(8, 5))
plt.plot(range(2, 11), sse, marker='o', linestyle='--')
plt.xlabel('Número de clusters')
plt.ylabel('Inercia (SSE)')
plt.title('Método del Codo')
plt.show()

# Método del silhouette score para determinar el número óptimo de clusters
silhouette_scores = []
for k in range(2, 11):
    kmeans = KMeans(n_clusters=k, random_state=42)
    labels = kmeans.fit_predict(df_scaled)
    silhouette_scores.append(silhouette_score(df_scaled, labels))

# Graficar el silhouette score
plt.figure(figsize=(8, 5))
plt.plot(range(2, 11), silhouette_scores, marker='o', linestyle='--', color='green')
plt.xlabel('Número de clusters')
plt.ylabel('Silhouette Score')
plt.title('Silhouette Score para diferentes números de clusters')
plt.show()
```

METODO DEL CODO



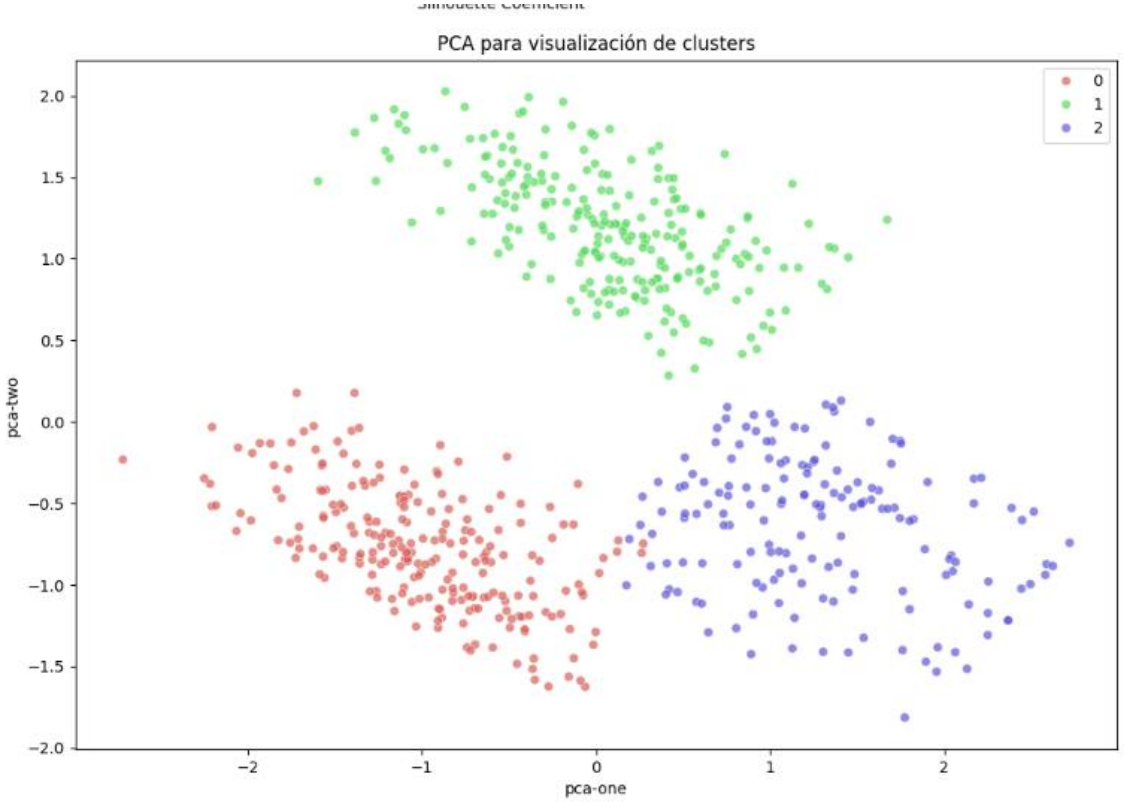
METODO DEL SILHOUETTE



El análisis de componentes principales (**PCA**) se utilizó para reducir la dimensionalidad de los datos y permitir una mejor visualización en 2D, así como para facilitar la aplicación de clustering.

```python
from sklearn.decomposition import PCA

# Reducimos los datos a 2 componentes principales para visualización
pca = PCA(n_components=2)
df_pca = pca.fit_transform(df_scaled)

# Visualizamos la varianza explicada por cada componente principal
explained_variance = pca.explained_variance_ratio_
print(f'Varianza explicada por el primer componente: {explained_variance[0]:.2f}')
print(f'Varianza explicada por el segundo componente: {explained_variance[1]:.2f}')

# Visualización de los clusters en los datos reducidos por PCA
optimal_k = 4  # Supongamos que el número óptimo de clusters es 4 basado en los métodos anteriores
kmeans = KMeans(n_clusters=optimal_k, random_state=42)
labels = kmeans.fit_predict(df_scaled)

# Graficar los datos reducidos por PCA
plt.figure(figsize=(8, 5))
plt.scatter(df_pca[:, 0], df_pca[:, 1], c=labels, cmap='viridis', s=50)
plt.xlabel('Componente Principal 1')
plt.ylabel('Componente Principal 2')
plt.title(f'Clusters Visualizados con PCA (k={optimal_k})')
plt.colorbar()
plt.show()
```

```
Varianza explicada por el primer componente: 0.23
Varianza explicada por el segundo componente: 0.08
```



Determinamos cual son las variables mas influyentes para nuestra Variable depnediente Desempeño Profesional

```python
# Calcula la correlación entre 'DesempeñoProfesional' y las demás variables
correlations = df.corr()['DesempeñoProfesional'].drop('DesempeñoProfesional')

# Obtén las 10 variables más influyentes (en valor absoluto)
top_10_influyentes = correlations.abs().nlargest(10)

print("Las 10 variables más influyentes en el Desempeño Profesional son:")
top_10_influyentes
```

```
Las 10 variables más influyentes en el Desempeño Profesional son:
```

	DesempeñoProfesional
COM3_1	0.615084
AC3_1	0.602840
COM3_3	0.587537
AC3_3	0.580589
COM3_2	0.571378
COM2	0.540723
AC2	0.526806
AC3_2	0.493728
AC3_4	0.481745
CI3	0.457562

COM3_1: En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas Comunicación oral (Ejemplo exposiciones)

AC3_1: En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos relacionados con tu autoconciencia Autoconocimient

COM3_3: En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas Expresión corporal (ton de voz, manejo de público)

AC3_3: En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos relacionados con tu autoconciencia Automotivación

COM3_2: En una escala del 1 al 5, donde 5 es la calificación más alta, cómo calificarías tus habilidades comunicativas Comunicación escrita (ej. trabajos, redacción)

COM2: ¿Con qué frecuencia aplicas la comunicación oral, escrita y expresión corporal, en tu formación profesional?

AC2: ¿Con qué frecuencia aplicas la autoconciencia en tu formación profesional?

AC3_2: En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos relacionados con tu autoconciencia Autocontrol

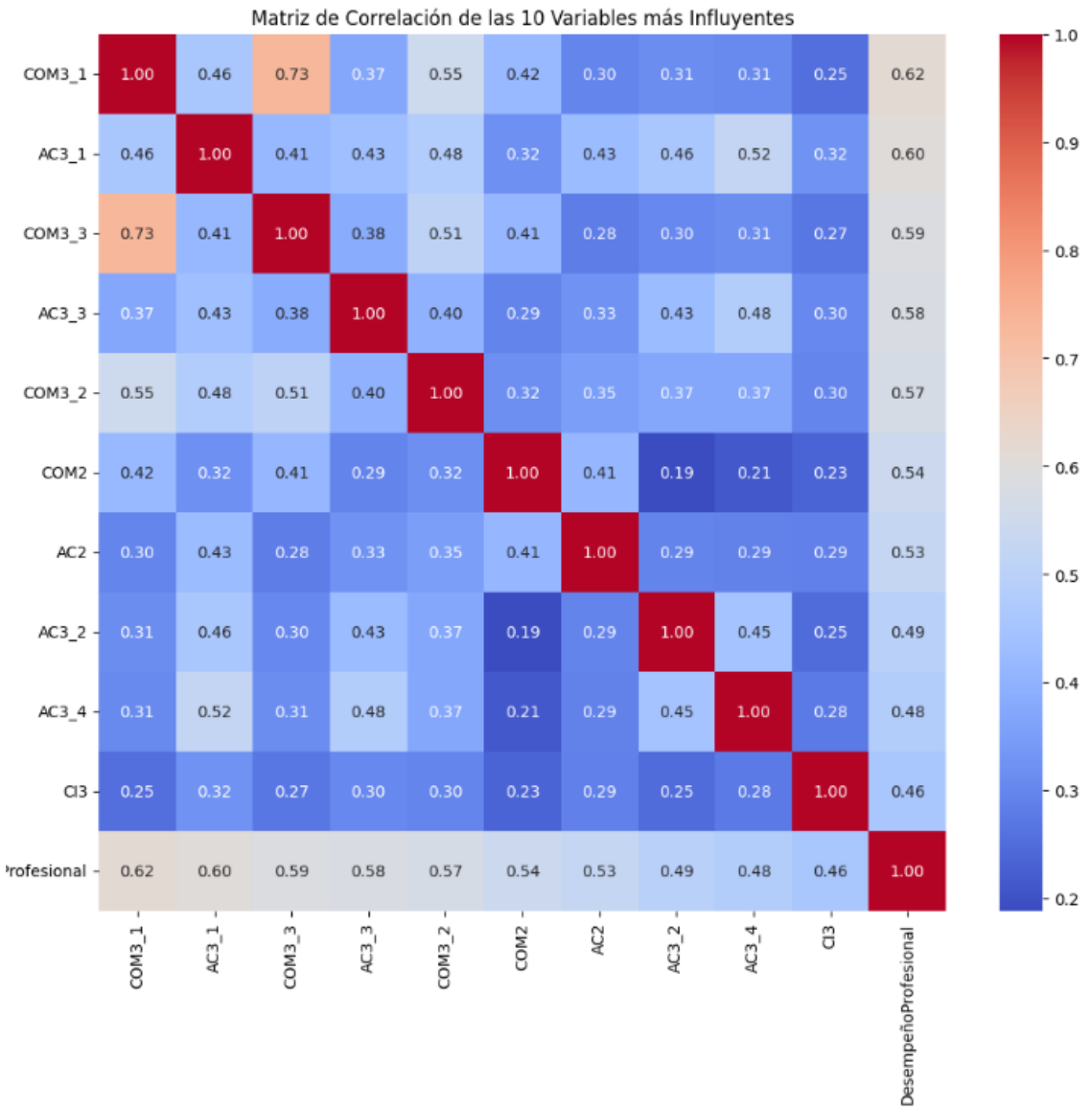
AC3_4: En una escala del 1 al 5, donde 5 es la calificación más alta, califica los aspectos relacionados con tu autoconciencia Autoimagen

CI3: En una escala de 1 a 5 donde 5 la calificación más alta, ¿Qué tan creativo e innovador te consideras?

Por medio del siguiente comando

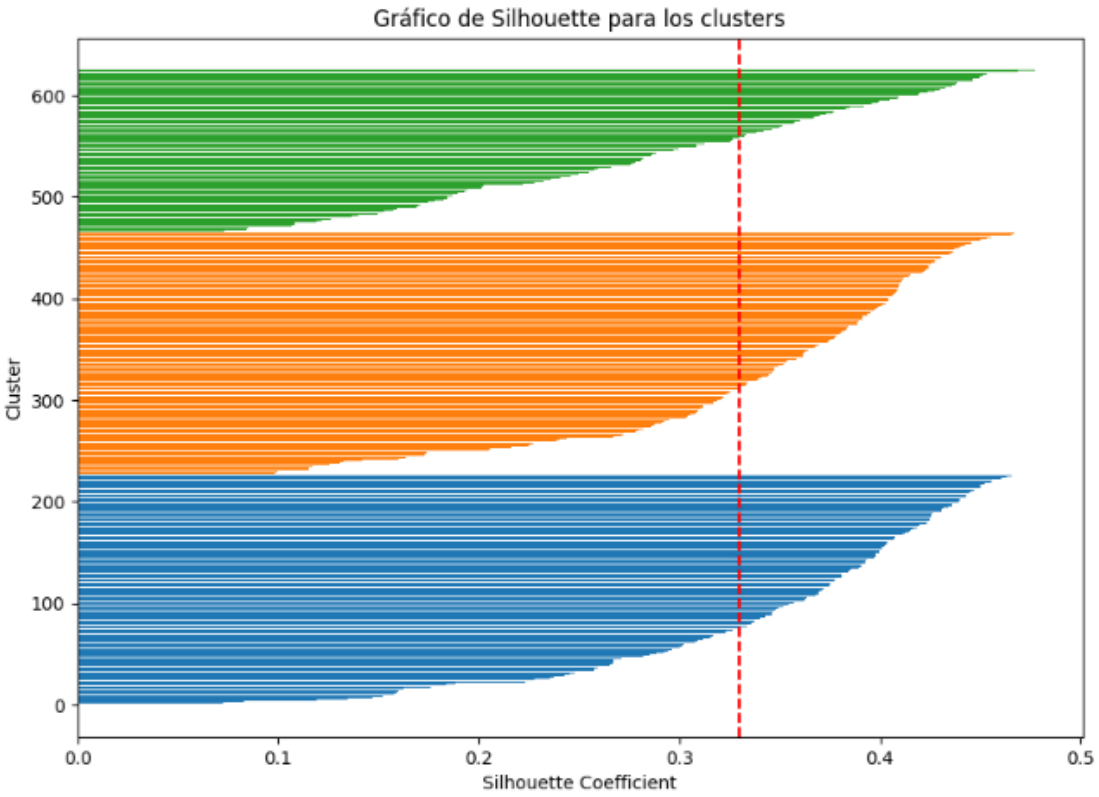
```
# Crea una matriz de correlación solo con las 10 variables más influyentes y 'DesempeñoProfesional'
matriz_correlacion = df[top_10_variables.tolist() + ['DesempeñoProfesional']].corr()

# Crea el mapa de calor
plt.figure(figsize=(12, 10))
sns.heatmap(matriz_correlacion, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Matriz de Correlación de las 10 Variables más Influyentes')
plt.show()
```

Graficamos una matriz de correlación con las variables más influyentes en nuestra variable dependiente “Desempeño Profesional”



Utilizando el siguiente comando podemos visualizar la gráfica de Silhouette y los clústeres generados

```
# Calcular el silhouette score final para los clusters generados
silhouette_avg = silhouette_score(df_scaled, labels)
print(f'Silhouette Score promedio para {optimal_k} clusters: {silhouette_avg:.2f}')

# Visualización del silhouette score por cluster
from sklearn.metrics import silhouette_samples
import numpy as np

silhouette_vals = silhouette_samples(df_scaled, labels)

plt.figure(figsize=(8, 6))
y_lower, y_upper = 0, 0
for i in range(optimal_k):
    ith_cluster_silhouette_vals = silhouette_vals[labels == i]
    ith_cluster_silhouette_vals.sort()
    y_upper += len(ith_cluster_silhouette_vals)
    plt.fill_betweenx(np.arange(y_lower, y_upper), 0, ith_cluster_silhouette_vals)
    y_lower = y_upper

plt.title('Gráfico de Silhouette para los clusters')
plt.xlabel('Silhouette Coefficient')
plt.ylabel('Cluster')
plt.show()
```



Una vez tenemos todo lo necesario para entrenar los modelos y todos nuestros datos estandarizados y etiquetados procedemos a entrenarlos

Para los No supervisado entrenamos dos modelos un DBSCAN Y KMEANS

```python
import pickle
from sklearn.cluster import DBSCAN


# KMeans model training and saving
kmeans = KMeans(n_clusters=optimal_clusters, init='k-means++', max_iter=300, n_init=10, random_state=0)
kmeans.fit(df)

# Save the KMeans model to a pickle file
with open('kmeans_model.pickle', 'wb') as f:
    pickle.dump(kmeans, f)

# DBSCAN model training and saving
dbscan = DBSCAN(eps=0.5, min_samples=5) # Adjust eps and min_samples as needed
dbscan.fit(df)

# Save the DBSCAN model to a pickle file
with open('dbscan_model.pickle', 'wb') as f:
    pickle.dump(dbscan, f)

# Download the files (you might need to adjust the paths if necessary)
from google.colab import files
files.download('kmeans_model.pickle')
files.download('dbscan_model.pickle')
```

Y para los modelos supervisados entrenamos dos modelos tambien un RANDOM FOREST Y UN ARBOL DE DECISION

```python
from sklearn.ensemble import RandomForestRegressor
from sklearn.tree import DecisionTreeRegressor

# Define X (features) and y (target variable)
X = df.drop('DesempeñoProfesional', axis=1)  # Assuming 'DesempeñoProfesional' is your target
y = df['DesempeñoProfesional']

# RandomForestRegressor model
rf_model = RandomForestRegressor(random_state=0)  # You can adjust hyperparameters
rf_model.fit(X, y)

# Save the RandomForestRegressor model
with open('rf_model.pickle', 'wb') as f:
    pickle.dump(rf_model, f)

# DecisionTreeRegressor model
dt_model = DecisionTreeRegressor(random_state=0) # You can adjust hyperparameters
dt_model.fit(X, y)

# Save the DecisionTreeRegressor model
with open('dt_model.pickle', 'wb') as f:
    pickle.dump(dt_model, f)

# Download the models
files.download('rf_model.pickle')
files.download('dt_model.pickle')
```

Y los resultados de estos entrenamientos se pueden visualizar en el aplicativo desarrollado que encontraran en https://proyectofinal-mineriadedatos.streamlit.app/

Referencias

[1] A. López y B. Gómez, "Estrategias educativas para promover la creatividad en el aula universitaria", *Revista Iberoamericana de Educación*, vol. 82, pp. 45-56, 2021.

[2] J. Ramírez y C. Silva, "Factores que obstaculizan el desarrollo de la creatividad en la educación superior", *Educación y Sociedad*, vol. 35, pp. 123-134, 2020.

[3] P. Fernández y R. Torres, "Autoconciencia y rendimiento académico: una revisión de la literatura", *Psicología y Educación*, vol. 29, pp. 87-98, 2022.

[4] M. González y S. Martínez, "Desarrollo de habilidades comunicativas en estudiantes universitarios", *Revista de Comunicación y Educación*, vol. 27, pp. 112-123, 2021.

[5] L. Sánchez y J. Pérez, "Integración de actividades prácticas para el desarrollo de habilidades blandas en estudiantes universitarios", *Revista Innovación Educativa*, vol. 40, pp. 34-46, 2022.