



PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL
FORMATO GUÍA DE APRENDIZAJE

1. IDENTIFICACIÓN DE LA GUÍA DE APRENDIZAJE

- **Denominación del programa de formación:** Algoritmo de agrupamiento no supervisado *K-means* con Python.
- **Código del programa de formación:** 21710111.
- **Nombre del proyecto:** N.A.
- **Fase del proyecto:** N.A.
- **Actividad de proyecto:** N.A.
- **Competencia:**
220501114 - Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas.
- **Resultados de aprendizaje a alcanzar:**
 - 220501114-01 - Determinar las características y criterios de agrupamiento de datos según el objetivo del análisis.
 - 220501114-02 - Agrupar los datos aplicando un algoritmo no supervisado.
 - 220501114-03 - Validar el agrupamiento de datos de acuerdo con los objetivos definidos para el análisis.
- **Duración de la guía:** 48 horas.

2. PRESENTACIÓN

La presente guía comprende la competencia “Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas” y, a partir de ella, se abordan actividades que le permitirán afianzar los conocimientos adquiridos en los componentes formativos “**Introducción al *Machine Learning***” y “**Aprendizaje no supervisado *K-means***”, relativos a las temáticas *Machine Learning*, aplicaciones del clustering de datos, aprendizaje no supervisado *K-means*, herramientas tecnológicas para el agrupamiento de datos y validación del resultado del análisis, las cuales permiten el desarrollo de habilidades para optimizar procesos y productos en las diferentes áreas de ocupación.



Las estrategias didácticas se centran en el aprendizaje basado en problemas; de manera inicial, se abordarán los conocimientos previos con la participación en el foro social; para continuar con la aplicación de estos conceptos, se desarrollarán actividades que aportan al conocimiento del algoritmo de agrupamiento no supervisado *K-means* con Python, para entornos empresariales de la región que se presentan en la programación de software, multimedia y demás. Es importante mencionar que, para todas las actividades planteadas en esta guía, contará con el acompañamiento del instructor calificado asignado al programa, el cual, de forma continua y permanente, lo orientará con las pautas necesarias para el logro de las actividades de aprendizaje, brindando herramientas básicas de tipo conceptual y metodológico. De igual manera, el instructor programará encuentros sincrónicos para brindar orientaciones específicas relacionadas con las temáticas y actividades, y así ofrecer apoyo en caso de dudas o inconvenientes. La fecha y el horario para esto serán indicados oportunamente.

A continuación, encontrará la descripción de cada actividad y las indicaciones necesarias para su presentación y entrega, cada una está referida a la aplicación de los conocimientos aportados en los componentes formativos: "**Introducción al *Machine Learning***" y "**Aprendizaje no supervisado *K-means***". Recuerde realizar con responsabilidad la presente guía, generando el envío oportuno de las evidencias de aprendizaje y la participación en la realización y socialización de las actividades propuestas.

A continuación, se presentan cada una de las acciones de aprendizaje que le permitirán desarrollar lo anteriormente mencionado.

3. FORMULACIÓN DE LAS ACTIVIDADES DE APRENDIZAJE

En este apartado, se describirán las actividades de aprendizaje para "**Algoritmo de agrupamiento no supervisado *K-means* con Python**", para suplir necesidades contextuales del sector informático.

De este programa de formación, en cada una de las actividades, se detallan las evidencias cuya entrega se realizará durante las cuatro (4) semanas.

Antes de elaborar las evidencias propuestas en esta guía de aprendizaje, es preciso que realice las siguientes actividades iniciales:

- Actualizar los datos personales.
- Participar en el foro social, según indicaciones del instructor.



- Revisar la “Información del programa”, esto le permitirá entender los objetivos, las actividades y la metodología que se llevará a cabo durante en el curso.
- Tener en cuenta que es fundamental entender los resultados de aprendizaje para tener mayor comprensión del proceso.
- Revisar el “Cronograma” que le permitirá conocer la planeación diseñada para lograr de manera secuencial los resultados de aprendizaje.
- Participar de los espacios de comunicación, tales como foros y otros.

Actividad de reflexión Inicial

Python fue creado en 1991, en el Instituto Nacional de Matemáticas, en Holanda, por Guido van Rossum. Le llamó Python por un grupo de comedia denominado así, mas no porque a él le gustaran las serpientes. Este es uno de los lenguajes más populares que existen en el mundo, es idóneo para ciencia de datos y sobrepasa en popularidad a R. Python es lenguaje *Open Source*, muchos programadores desarrollan nuevos componentes y librerías, también es un lenguaje sencillo, en cuanto a su aprendizaje, y es muy popular en desarrollo web; permite el manejo de varios procesadores, ya que es un lenguaje muy común en la construcción de sitios importantes, como YouTube, Instagram, etc.

Luego del anterior preámbulo, deberá instalar Python a través de un *toolkit* para ciencia de datos llamado Anaconda, el cual se baja del sitio <http://www.anaconda.com> y reflexione sobre las siguientes preguntas:

- ¿Qué funciones de la aplicación Anaconda encontró al instalarla?
- ¿Qué ofrece Anaconda Navigator y Anaconda Prompt?
- ¿Qué aplicaciones contiene Anaconda Navigator que permiten usar Python de forma gráfica?
- Describa la ruta que ha utilizado para instalar Python a través de pantallazos del procedimiento.

Esta actividad formativa es de carácter no sumativo, ya que no tendrá una calificación. Es una actividad de reflexión previa al desarrollo de las actividades propuestas.

Recuerde: este ejercicio tiene como finalidad encaminarle y motivarle en la elaboración de las evidencias que se plantean para esta guía de aprendizaje, por tal razón, no es calificable, pero sí es el punto de inicio para lo que se le solicitará en las demás evidencias.



3.1. Actividad de aprendizaje de la competencia: Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas. (220501114).

Las actividades planteadas para la competencia permiten fortalecer en el aprendiz la capacidad de uso de una de las técnicas de agrupamiento más conocidas en el uso de máquinas de aprendizaje no supervisado que es el algoritmo *K-means* o K-medias, usando lenguaje de programación Python; el aprendiz, por tanto, construirá particiones de un conjunto de datos, identificará patrones y reforzará el aprendizaje del algoritmo.

Todas las actividades y evidencias de esta guía se trabajarán con base en el siguiente caso:

Entendiendo el comportamiento de los consumidores

El comportamiento del consumidor se refiere a la forma en que los consumidores seleccionan, deciden, usan y disponen de bienes y servicios. Este comportamiento ayuda a comprender mejor las actitudes y preferencias de los consumidores que influyen en el comportamiento de compra, y ayuda a los vendedores a comprender cómo satisfacer las necesidades de los clientes, proporcionarles valor y monetizar su negocio.

Para predecir el comportamiento de los consumidores, se usará la fuente de datos (*dataset*)

Mall_Customers.csv, anexa en la presente guía o puede ser bajada de la ruta

<https://www.kaggle.com/datasets/vjchoudhary7/customer-segmentation-tutorial-in-Python>

El *dataset* consta de los siguientes campos:

- **CustomerID:** Identificación del cliente.
- **Gender:** Genero del cliente.
- **Age:** Edad del cliente.
- **AnnualIncome:** Ingresos anuales del cliente.
- **SpendingScore:** Puntuación del gasto, valor entre 1 y 100, que indica cuánto ha gastado un cliente en el centro comercial; cuanto más valor, más puntaje.



3.1.1. Actividad de aprendizaje AA1 - Identificar técnicas avanzadas de aprendizaje automático y reconocimiento de patrones según el objetivo del análisis.

En esta actividad, el aprendiz tendrá la oportunidad de llevar a cabo una exploración inicial de datos del problema y, de esta forma, comprender los datos que se van a utilizar; los objetivos son examinar, explicar, resumir y visualizar los datos en cuestión, además de tener una primera aproximación de identificación de patrones en los datos del problema.

Para cumplir con esto, se deben realizar las siguientes tareas:

- Importar de las librerías de Python necesarias para realizar la exploración usando Jupyter lab.
- Cargar los datos del caso propuesto.
- Identificar los tipos de variables.
- Obtener estadísticas de cada una de las variables.
- Identificar datos nulos o datos perdidos.
- Obtener una distribución de datos de cada una de las variables.
- Realizar visualización de relación entre las múltiples variables del *dataset*.
- Realizar una matriz de correlación de las variables.

Duración: 16 horas.

Materiales de formación a consultar:

Para el desarrollo de esta actividad, es importante la lectura y el análisis del material de formación del componente formativo **Introducción al *Machine Learning***.

Evidencias: a continuación, se describen las acciones y la correspondiente evidencia que conforman la actividad de aprendizaje:

- **AA1-EV01. Presentación interactiva: características y criterios de agrupamiento de datos con Python.**

Esta evidencia, la cual es **una presentación interactiva** que muestra el desarrollo de las tareas planteadas inicialmente, usando Python en Jupyter lab, debe contener lo siguiente:

- 1) Mostrar la importación de las librerías necesarias.



- 2) Cargue de datos relacionados.
- 3) Identificación de tipos de variables.
- 4) Obtención de estadísticas de cada una de las variables.
- 5) Generación de gráficos de distribución de las variables.
- 6) Visualización de la relación entre las múltiples variables del *dataset*.

Lineamientos generales para la entrega de la evidencia:

- **Producto por entregar:** presentación interactiva: características y criterios de agrupamiento de datos con Python.
- **Formato:** Prezi o mp4.
- Para presentar la evidencia, remítase al área de la actividad correspondiente y acceda al espacio: **Presentación interactiva: características y criterios de agrupamiento de datos con Python. AA1-EV01**

3.1.2. Actividad de aprendizaje AA2 - Relacionar los criterios de agrupamiento de datos según el objetivo del análisis en un algoritmo no supervisado.

En esta actividad, ejecutará el algoritmo, que se clasifica como algoritmo de aprendizaje no supervisado, para los datos del caso planteado y cuyo objetivo principal es obtener clústeres, identificando variables independientes, obtener una muestra de prueba, obtener el número de clústeres del modelo, aplicar el algoritmo k basado en características, identificar coordenadas de los centroides, realizar entrenamiento del modelo, predecir a qué clúster pertenece un dato de muestra.

Duración: 16 horas.

Materiales de formación a consultar:

Para el desarrollo de esta actividad, es importante la lectura y el análisis del material de formación del componente formativo **Aprendizaje no supervisado K-means**.

Evidencias: a continuación, se describen las acciones y la correspondiente evidencia que conforman la actividad de aprendizaje:



- **AA2-EV01. Informe de agrupamiento: datos en Markdown aplicando *K-means* en Python.**

De acuerdo con el caso planteado inicialmente, deberá entregar un informe formal en Markdown mostrando el código fuente completo con comentarios sobre su funcionalidad.

- 1) En el informe, se debe reflejar carga de datos.
- 2) La obtención de una muestra de prueba aleatoria a partir de la información inicial.
- 3) Eliminación de variables que no aportan al modelo, tanto en la muestra como en el conjunto inicial.
- 4) Realización de escalamiento (normalización) de datos si es necesario.
- 5) Obtención de valor óptimo de K, aplicación de algoritmo *K-means*.
- 6) entrenamiento del modelo, obtención de las coordenadas de los centroides.
- 7) realización de predicción con la muestra obtenida.

Lineamientos generales para la entrega de la evidencia:

- **Producto por entregar:** informe de agrupamiento de datos en Markdown aplicando *K-means* en Python.
- **Formato:** PDF
- Para presentar la evidencia, remítase al área de la actividad correspondiente y acceda al espacio: **Informe de agrupamiento: datos en Markdown aplicando *K-means* en Python. AA2-EV01.**

3.1.3. Actividad de aprendizaje AA3 - Interpretar los resultados obtenidos en el agrupamiento de datos de acuerdo con los objetivos definidos para el análisis.

En esta actividad, realizará un video mostrando gráficamente los clústeres obtenidos de acuerdo con el caso propuesto; para la elaboración del video, se deben interpretar y analizar los resultados obtenidos, probar el algoritmo con otros valores de K y mostrar los resultados, mostrar cuál es el valor óptimo de K.

Duración: 16 horas.

Materiales de formación a consultar: para el desarrollo de esta actividad, es importante la lectura y el análisis del material de formación del componente formativo **Aprendizaje no supervisado *K-means*.**



Evidencias: a continuación, se describen las acciones y la correspondiente evidencia que conforman la actividad de aprendizaje:

- **AA3-EV01. Video: interpretación de clústeres encontrados en un algoritmo no supervisado.**

Para la elaboración del video, tendrá en cuenta el caso presentado inicialmente y deberá:

1. Mostrar los gráficos de los clústeres obtenidos.
2. Explicar la interpretación de los resultados obtenidos.
3. Presentar los resultados que se obtuvieron al cambiar el valor de K.
4. El video puede elaborarse en cualquier plataforma de videos, como Vimeo, Youtube, etc.
5. Tomar la URL donde haya quedado alojado el video y anexarla al espacio de entrega de la evidencia.

Lineamientos generales para la entrega de la evidencia:

- **Producto por entregar:** video de interpretación de clústeres encontrados en un algoritmo no supervisado.
- **Formato:** video.
- Para presentar la evidencia, remítase al área de la actividad correspondiente y acceda al espacio:
Video: interpretación de clústeres encontrados en un algoritmo no supervisado AA3-EV01

4. ACTIVIDADES DE EVALUACIÓN

Tome como referencia la técnica e instrumentos de evaluación citados en la guía de Desarrollo Curricular

Evidencias de aprendizaje	Criterios de evaluación	Técnicas e instrumentos de evaluación
Evidencia de conocimiento: Presentación interactiva: características y criterios de agrupamiento de datos con Python.	<ul style="list-style-type: none">• Selecciona el conjunto de datos requerido para el caso de análisis solicitado según especificaciones técnicas.• Prepara los datos del conjunto de datos de	Lista de chequeo: IE-AA1-EV01.



Evidencias de aprendizaje	Criterios de evaluación	Técnicas e instrumentos de evaluación
AA1-EV01	<p>acuerdo a los criterios del objeto de análisis para el estudio requerido.</p> <ul style="list-style-type: none">• Realiza el análisis exploratorio inicial del conjunto de datos en búsqueda de las características y similitudes entre los datos requeridos para el objeto de estudio o análisis de acuerdo a algoritmos utilizados.• Simplifica los datasets con variables o atributos similares para obtener el conjunto de datos final a ser analizado de acuerdo al conjunto de datos.	
<p>Evidencia de desempeño:</p> <p>Informe de agrupamiento: datos en Markdown aplicando <i>K-means</i> en Python.</p>	<ul style="list-style-type: none">• Prepara el ambiente de trabajo en Python con las herramientas y librerías necesarias de acuerdo al modelo a realizar.• Selecciona el algoritmo a ser utilizado de acuerdo con los requerimientos y objetivos del análisis a realizar.	<p>Lista de chequeo:</p> <p>IE-AA2-EV01.</p>



Evidencias de aprendizaje	Criterios de evaluación	Técnicas e instrumentos de evaluación
AA2-EV01	<ul style="list-style-type: none">• Normaliza el conjunto de datos para garantizar escalas similares en sus características de acuerdo al algoritmo seleccionado.• Selecciona las características y similitudes de los datos objeto de estudio según algoritmo seleccionado.• Determina el número de clústeres a ser construidos durante el análisis del conjunto de datos según técnica de agrupamiento.• Entrena el algoritmo con un conjunto de datos de prueba para lograr un aprendizaje y mejor predicción en los resultados del objeto de estudio.• Realiza el agrupamiento de los datos final de acuerdo a lo requerido para el análisis propuesto.	
	<ul style="list-style-type: none">• Grafica los resultados obtenidos de la	Lista de chequeo:



Evidencias de aprendizaje	Criterios de evaluación	Técnicas e instrumentos de evaluación
Evidencia de producto: Video: interpretación de clústeres encontrados en un algoritmo no supervisado AA3-EV01	clusterización para permitir su posterior análisis según técnica utilizada. <ul style="list-style-type: none">• Analiza los resultados obtenidos del análisis frente a los objetivos propuestos en el inicio del proyecto para su aplicación.• Determina las conclusiones finales del análisis realizado para la toma correcta de decisiones.	IE-AA3-EV01.

5. GLOSARIO DE TÉRMINOS

Aprendizaje automático: Rama de la inteligencia artificial cuyo objetivo es implementar técnicas que permitan a los computadores aprender mediante un proceso de inducción del conocimiento.

Aprendizaje automático supervisado: El algoritmo recibe datos de entrenamiento consistentes en datos etiquetados.

Aprendizaje automático no supervisado: El algoritmo identifica patrones y saca conclusiones de los datos que se le proporcionan.

Inteligencia artificial: Sistemas informáticos que pueden aprender como aprende un ser humano.

Machine Learning: Aprendizaje automático.

K-Means: Algoritmo de agrupamiento o *clustering*.

Python: Lenguaje de alto nivel usado para construir todo tipo de aplicaciones y muy usado en ciencia de datos.



6. REFERENTES BIBLIOGRÁFICOS

Abdulhamit, S. (2020). *Data analysis using Python*. Academic Press.

Akram. (2018). *Mall-customers*. Kaggle. <https://www.kaggle.com/akram24/mall-customers>

Rusell, R. (2018). *Machine Learning. Step-by-Step Guide to implement machine learning algorithms with Python*.

Severance C. (2020). *Python para todos. Explorando la información con Python 3*.

https://www.academia.edu/40030172/Python_para_todos_Explorando_la_informaci%C3%B3n_con_Python_3_Charles_R_Severance

7. CONTROL DEL DOCUMENTO

	Nombre	Cargo	Dependencia	Fecha
Autor	Héctor Henry Jurado Soto	Experto Temático	Regional Cauca – Centro de Teleinformática y Producción Industrial	Abril 2022
Autor	Caterine Bedoya Mejía	Diseñadora Instruccional	Regional Distrito Capital – Centro de Gestión Industrial	Abril 2022
Autor	Rafael Neftalí Lizcano Reyes	Responsable Equipo de Desarrollo Curricular	Regional Santander - Centro Industrial del Diseño y la Manufactura	Abril 2022
Autor	Darío González	Corrector de Estilo	Regional Distrito Capital – Centro de Diseño y Metrología	Mayo 2022



8. CONTROL DE CAMBIOS (diligenciar únicamente si realiza ajustes a la guía)

	Nombre	Cargo	Dependencia	Fecha	Razón del Cambio
Autor (es)					