

**Universidad Nacional Abierta y a Distancia**  
**Vicerrectoría Académica y de Investigación**  
**Guía única para el desarrollo del componente práctico del curso:**  
**Programación para el Análisis de Datos**  
**Código: 203008069**

**Unidad 3 - Tarea 3 - Componente práctico – Prácticas simuladas**

**1. Información general del componente práctico.**

<b>Estrategia de aprendizaje: Exploración y Visualización de Datos.</b>
<b>Tipo de curso: Metodológico</b>
<b>Momento de la evaluación: Intermedio</b>
<b>Puntaje máximo del componente: 120 puntos</b>
<b>Número de actividades del componente registradas en esta guía: 1</b>
<p><b>Con este componente se espera conseguir los siguientes resultados de aprendizaje:</b></p> <p>Resultado de aprendizaje 2: Implementar diferentes funciones, paquetes y módulos para la computación matemática y ciencia de datos.</p> <p>Resultado de aprendizaje 3: Estructurar diferentes técnicas para la manipulación de datos en un lenguaje de programación.</p> <p>Resultado de aprendizaje 4: Crear informes de exploración y análisis de datos, mediante el uso de un lenguaje de programación en un entorno informático interactivo.</p>

**2. Descripción general actividad(es) del componente práctico.**

<b>Escenarios de componente práctico: Con Apoyo TIC</b>	
<b>Tipo de actividad: Independiente</b>	
<b>Número de actividad: 4</b>	
<b>Puntaje máximo de la actividad: 120 puntos</b>	
<b>La actividad inicia el:</b> viernes, 14 de junio de 2024	<b>La actividad finaliza el:</b> domingo, 14 de julio de 2024
<b>Los recursos con los que debe contar para el desarrollo de la actividad son los siguientes:</b>	
<ul style="list-style-type: none"><li>• Base de Datos seleccionada entre las opciones brindadas por el director de Curso (preferiblemente, la selección se debe basar en su campo de acción).</li><li>• Distribución para ciencia de datos Anaconda de Python 3 instalada.</li></ul>	

- Para el desarrollo de esta práctica es necesario que revise en el Entorno de Aprendizaje (Unidad 3 - Contenidos y referentes bibliográficos), las siguientes referencias:
  - Alex Galea. (2018). Applied Data Science with Python and Jupyter: Use Powerful Industry-standard Tools to Unlock New, Actionable Insights From Your Data: Vol. 1st edition. Packt Publishing. (pp. 29-49).
  - Boschetti, A., & Massaron, L. (2016). Python Data Science Essentials - Second Edition (Vol. 0002). Packt Publishing. (pp. 124-134).
  - Samir Madhavan. (2015). Mastering Python for Data Science: Explore the World of Data Science Through Python and Learn How to Make Sense of Data. Packt Publishing. (pp. 77-106).
  - Thakur, A. (2016). Python: Real-World Data Science. Packt Publishing. (pp. 536 - 563).

### La actividad consiste en:

Construir un Análisis Exploratorio de Datos que genere hallazgos significativos de la base de datos seleccionada, poniendo en práctica los siguientes conceptos:

#### Concepto 1: Parámetros y métricas descriptivas

- Muestre las métricas descriptivas para todas las columnas de su base de datos según el tipo de variable como sigue:
  - *Variables numéricas*: Cantidad de datos, media, desviación estándar, mínimo, máximo y percentiles 25, 50 y 75.
  - *Variables categóricas*: Cantidad, cantidad de valores únicos, moda y frecuencia de la moda.
- Muestre la matriz de correlación de las variables numéricas e identifique las variables que más se correlacionan (visualizarla, ver concepto 3).
- Realice un análisis de los resultados obtenidos.

#### Concepto 2: Escalamiento, estandarización y codificación

- Categorice al menos una variable continua. Defina consistentemente las categorías propuestas y explique por qué las propone.
- Estandarice o escale al menos dos variables numéricas que sean diferentes en magnitud o unidades de medición, de modo que pueda comparar y encontrar similitudes o diferencias en sus distribuciones (visualmente, ver concepto 3).

#### Concepto 3: Gráficos y visualización científica

- Realice al menos un diagrama de barras.
- Realice al menos un diagrama de cajas.
- Realice al menos un diagrama de pie de una variable categórica (debe realizar agrupación por conteo).
- Realice un *scatter plot* (o *pair plot*) de las variables numéricas identificadas en el Concepto 1.
- Realice un diagrama de violín con la categorización de la variable continua que desarrollo en el Concepto 2.
- Realice un gráfico que le permita visualizar la comparación de los histogramas (distribuciones) de las variables que estandarizó o escaló en concepto 2.
- Muestre gráficamente la matriz de correlación solicitada en el concepto 1.
- Todos los gráficos realizados deben tener comentarios y análisis de resultados.
- **Puntuación mejorada:** Si su base de datos contiene datos georeferenciados, presente una visualización de mapa con la librería Folium.

### Condiciones de entrega:

El archivo Jupyter Notebook presentado debe contar con la siguiente estructura:

- Portada en celda markdown:
  - Curso
  - Tarea
  - Presentado por: Estudiante
  - Grupo
  - Código
  - Presentado a: Tutor
  - Universidad Nacional Abierta y a Distancia – UNAD
  - Fecha
- Introducción del Análisis Exploratorio de Datos propuesto, con objetivos establecidos.
- Presente los resultados que le permitan desarrollar los conceptos (no necesariamente en el mismo orden). Preferiblemente, presenta cálculos y gráficos asociados de forma iterada, así como los comentarios y análisis de resultados.
- Presente conclusiones y discusión general de los hallazgos significativos del Análisis Exploratorio de Datos.
- Presentar referencias bibliográficas con normas APA en una celda markdown.
- Debe ser entregado un archivo en extensión .ipynb nombrado de la siguiente manera en el entorno de Evaluación:  
*G##\_NombreEstudiante\_EDA.ipynb*

donde:

- *G##*: es el número del grupo al que pertenecen
- *NombreEstudiante* es el nombre del estudiante tal como aparece en el foro.
- *EDA*: Siglas de *Exploratory Data Analysis* (Análisis Exploratorio de Datos).

Por ejemplo:

*G21\_MiguelAngelVargasValencia\_EDA.ipynb*

- No se acepta el enlace de Google Colab u otros como Word o PDF.

### **Para el desarrollo de la actividad tenga en cuenta que:**

En el entorno de Información inicial debe: Revisar la agenda del curso y debe tener en cuenta las fechas de inicio y finalización de la actividad.

En el entorno de Aprendizaje debe: Realizar la lectura de las referencias bibliográficas correspondientes a la Unidad 3 y participar semanalmente con aportes significativos en el foro de la Unidad 3 - Tarea 3 – Componente práctico – Practicas simuladas.

En el entorno de Evaluación debe: Entregar un archivo en formato (\*.ipynb) con los ejercicios desarrollados y que cumpla con las condiciones de entrega establecidos en la actividad.

### **Evidencias de trabajo independiente:**

Las evidencias de trabajo independiente para entregar son:

- La construcción de un Análisis Exploratorio de Datos a la base de datos seleccionada; el cual tendrá que realizar de forma individual en un documento con extensión ipynb, como se solicita en las condiciones de entrega en el foro. habilitado para el desarrollo y avances de la tarea en el entorno de aprendizaje.
- En el Entorno de Evaluación – Unidad 3 - Tarea 3 – Componente práctico – Practicas simuladas, subir un único archivo que cumpla con las condiciones de entrega y que contenga las correcciones y sugerencias que el tutor realizó previamente en el foro.

### **Evidencias de trabajo grupal:**

En esta actividad no se requieren evidencias de trabajo grupal.

## **3. Lineamientos generales para la elaboración de las evidencias**

Para evidencias elaboradas **de forma Independiente**, tenga en cuenta las siguientes orientaciones

1. Realice un reconocimiento general del curso y de cada uno de los entornos antes de abordar el desarrollo de las actividades.
2. Identifique los recursos y referentes de la unidad a la que corresponde la actividad.
3. Intervenga en el foro de discusión aplicando las normas de netiqueta Virtual, evidenciando siempre respeto por las ideas de sus compañeros y del cuerpo docente.
4. Antes de entregar el producto solicitado revise que cumpla con todos los requerimientos que se señalaron en esta guía de actividades, rúbrica de evaluación y por parte del tutor en el foro de discusión:
  - Ejercicios desarrollados haciendo uso correcto de la edición de texto en las celdas markdown con correcta escritura en la sintaxis matemática.
  - Estructura del documento:
    - Portada.
    - Introducción.
    - Construcción de Análisis Exploratorio de Datos.
    - Discusión de Resultados.
    - Conclusiones.
    - Referencias.
  - El documento deberá ser nombrado con la siguiente estructura:  
*G##\_NombreEstudiante\_EDA.ipynb*
  - Entregar en el entorno de evaluación un archivo final el cual contenga los cambios y sugerencias realizados previamente por el tutor en el foro de la actividad.
  - No cometa fraudes, ni plagios ni actos que atenten contra el normal desarrollo académico de las actividades.

Tenga en cuenta que todos los productos escritos individuales o grupales deben cumplir con las normas de ortografía y con las condiciones de presentación que se hayan definido.

En cuanto al uso de referencias considere que el producto de esta actividad debe cumplir con las normas **APA**

En cualquier caso, cumpla con las normas de referenciación y evite el plagio académico, para ello puede apoyarse revisando sus productos escritos mediante la herramienta Turnitin que encuentra en el campus virtual.

Considere que En el acuerdo 029 del 13 de diciembre de 2013, artículo 99, se considera como faltas que atentan contra el orden académico, entre otras, las siguientes: literal e) "El plagiar, es decir, presentar como de su propia autoría la totalidad o parte de una obra, trabajo, documento o invención realizado por otra persona. Implica también el uso de citas o referencias faltas, o proponer citad donde no haya coincidencia entre ella y la referencia" y liberal f) "El reproducir, o copiar con fines de lucro, materiales educativos o resultados de productos de investigación, que cuentan con derechos intelectuales reservados para la Universidad."

Las sanciones académicas a las que se enfrentará el estudiante son las siguientes:

- a) En los casos de fraude académico demostrado en el trabajo académico o evaluación respectiva, la calificación que se impondrá será de cero puntos sin perjuicio de la sanción disciplinaria correspondiente.
- b) En los casos relacionados con plagio demostrado en el trabajo académico cualquiera sea su naturaleza, la calificación que se impondrá será de cero puntos, sin perjuicio de la sanción disciplinaria correspondiente.

#### 4. Formato de Rúbrica de evaluación

<b>Tipo de actividad: Independiente</b>	
<b>Número de actividad: 4</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>La máxima puntuación posible es de 120 puntos</b>	
<b>Criterios</b>	<b>Desempeños</b>
<b>Primer criterio de evaluación:</b>  Del procedimiento: Uso de los conceptos de Parámetros y Métricas descriptivas para su análisis exploratorio de datos.	<p><b>Nivel alto:</b> El estudiante utiliza adecuadamente los conceptos de Parámetros y Métricas descriptivas de forma correcta y como fue solicitado.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 20 puntos y 25 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante desarrolla de forma parcial, o con errores, los conceptos de Parámetros y Métricas descriptivas, o no se ciñe a lo solicitado para su análisis de datos.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 19 puntos</b></p>



<p><b>Este criterio representa 25 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel bajo:</b> El estudiante no desarrolla, o desarrolla de forma incorrecta, los conceptos de Parámetros y Métricas descriptivas solicitados. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 9 puntos</b></p>
<p><b>Segundo criterio de evaluación:</b></p> <p>Del procedimiento: Uso de los conceptos de escalamiento, estandarización y codificación para su análisis exploratorio de datos.</p> <p><b>Este criterio representa 25 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante utiliza adecuadamente los conceptos de escalamiento, estandarización y codificación de forma correcta y como fue solicitado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 20 puntos y 25 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante desarrolla de forma parcial, o con errores, los conceptos de escalamiento, estandarización y codificación, o no se ciñe a lo solicitado para su análisis de datos. <b>Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 19 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no desarrolla, o desarrolla de forma incorrecta, los conceptos de escalamiento, estandarización y codificación solicitado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 9 puntos</b></p>
<p><b>Tercer criterio de evaluación:</b></p> <p>Del procedimiento: Uso de los conceptos de gráficos y visualización científica para su análisis exploratorio de datos.</p> <p><b>Este criterio representa 30 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante utiliza adecuadamente los conceptos de gráficos y visualización científica de forma correcta y como fue solicitado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 20 puntos y 30 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante desarrolla de forma parcial, o con errores, los conceptos de gráficos y visualización científica, o no se ciñe a lo solicitado para su análisis de datos. <b>Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 19 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no desarrolla, o desarrolla de forma incorrecta, los conceptos de gráficos y visualización científica solicitado. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 9 puntos</b></p>

<p><b>Cuarto criterio de evaluación:</b></p> <p>De la participación: Intervención semanal en el foro con aportes significativos.</p> <p><b>Este criterio representa 20 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> El estudiante interactúa de forma oportuna, adecuada y respetuosa en el foro. Realizando aportes significativos de su trabajo en archivos .ipynb, realiza las correcciones a las que haya lugar, de acuerdo con las indicaciones del tutor. Aporta apoyando al aprendizaje de sus compañeros. <b>Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> Aunque el estudiante interactúa en el foro, no responde las participaciones académicas con aportes significativos y su desarrollo no responde con suficiente argumentación frente al referente consultado o no presenta las correcciones indicadas por el tutor. <b>Si su trabajo se encuentra en este nivel puede obtener entre 5 puntos y 9 puntos</b></p> <p><b>Nivel bajo:</b> El estudiante no presenta aportes individuales y no participa activamente en el foro. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 4 puntos</b></p>
<p><b>Quinto criterio de evaluación:</b></p> <p>Del desempeño: Cumplimiento de las condiciones para realización y entrega del documento final.</p> <p><b>Este criterio representa 20 puntos del total de 120 puntos de la actividad</b></p>	<p><b>Nivel alto:</b> Entrega el trabajo con al menos el 90% de las condiciones exigidas para entrega del documento. <b>Si su trabajo se encuentra en este nivel puede obtener entre 10 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> Entrega el trabajo con el 70% a 90% de las condiciones exigidas para entrega del documento. <b>Si su trabajo se encuentra en este nivel puede obtener entre 5 puntos y 9 puntos</b></p> <p><b>Nivel bajo:</b> Entrega el trabajo con menos del 70% de las condiciones exigidas para entrega del documento. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 4 puntos</b></p>