

**Universidad Nacional Abierta y a Distancia**  
**Vicerrectoría Académica y de Investigación**  
**Curso: Métodos Estadísticos en la Ciencia de Datos y Analítica**  
**Código: 203008072**

**Guía de actividades y rúbrica de evaluación – Tarea 1**  
**Análisis exploratorio e inferencial**

**1. Descripción de la actividad**

<b>Tipo de actividad: Independiente</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>Puntaje máximo de la actividad: 100 puntos</b>	
<b>La actividad inicia el:</b> miércoles, 4 de septiembre de 2024	<b>La actividad finaliza el:</b> domingo, 29 de septiembre de 2024
<b>Con esta actividad se espera conseguir los siguientes resultados de aprendizaje:</b>  Utilizar el análisis exploratorio e inferencial de los datos que permita la adaptación en modelos de aprendizaje estadístico.	
<b>La actividad consiste en:</b>  <p><b>Paso 1.</b> Descargue en el siguiente enlace, la Base de Datos -BD- (dynamic_pricing.csv), acerca de una empresa de viajes compartidos que desean implementar una estrategia de precios dinámica para optimizar las tarifas en función de las condiciones del mercado en tiempo real. Actualmente, la empresa solo utiliza la duración del viaje para decidir las tarifas. La empresa pretende aprovechar técnicas basadas en estadística para analizar datos históricos:</p> <p><a href="https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset?select=dynamic_pricing.csv">https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset?select=dynamic_pricing.csv</a></p> <p>Se requiere realizar un análisis estadístico de tipo explicativo e inferencial que permitan conocer los patrones que influyen en el costo histórico.</p> <p>Para dar solución a esta problemática usted debe realizar un análisis de datos mediante las siguientes actividades:</p>	

## Paso 2. Desarrollo del Informe de Actividades

### Análisis exploratorio e inferencial.

Realice las siguientes seis actividades, interpretando cada uno de los resultados, recuerda utilizar **solamente** los programas Python o R, no utilizar otro software.

#### Análisis Unidimensional.

- 1) Realizar una tabla de diccionario de datos que tenga tres columnas: nombre de la variable, descripción de la variable y clasificación en numérica o categórica u otra.
- 2) Realizar las tablas de distribuciones de frecuencias y gráficas de cada una de las variables categóricas, colocando para cada una, su correspondiente interpretación descriptiva.
- 3) Calcular las medidas de resumen de número de viajeros, calificación, expectativa de duración del viaje y costo histórico del viaje realizar su correspondiente interpretación descriptiva.
- 4) Analizar la distribución de las variables del punto 3, mediante un Boxplot o Violín plot, realizar su correspondiente interpretación descriptiva.

#### Análisis Bidimensional e Inferencial.

- 5) Analizar la correlación dos a dos entre número de viajeros, calificación, expectativa de duración del viaje contra el costo histórico, siguiendo las indicaciones:
  - Realizar los gráficos scatterplot correspondientes (número de viajeros vs costo histórico, calificación vs costo histórico y expectativa de duración del viaje vs costo histórico), describa las posibles correlaciones de acuerdo con las visualizaciones.
  - Calcular la correlación de Pearson dos a dos (número de viajeros vs costo histórico, calificación vs costo histórico y expectativa de duración del viaje vs costo histórico) y realizar prueba de correlación para determinar si la correlación entre las variables es fuerte, media o débil e interprete cada uno de los resultados.

- 6) Analizar las asociaciones estado de fidelidad del cliente, reservación y tipo de vehículo, siguiendo las indicaciones:
- Realiza una tabla de contingencia para cada una de las asociaciones (deben salir tres tablas), coloque observaciones relevantes frente a las posibles relaciones que encuentre entre categorías.
  - Realice pruebas Chi – Cuadrado para cada una de las tablas de contingencia, puede hacerlo en paralelo e interprete los resultados, determinando que la existencia o no de asociación entre las variables.
- 7) Analizar la relación entre número de viajeros, calificación y expectativa de duración del viaje, siguiendo las indicaciones:
- Diseñe gráficos de boxplot o violín plot para analizar si se detectan diferencias en las distribuciones de viajeros, calificación y expectativa de duración del viaje según la fidelidad del cliente. (Debe obtener tres gráficas)
  - Realice una prueba ANOVA de una sola vía, que compruebe la igualdad o diferencia de medias, interprete la significancia estadística y si encuentra diferencias significativas con algún grupo, implementa una prueba Post Hoc, que permita determinar si la fidelidad del cliente influye en la cantidad de viajeros, su calificación y expectativa del viaje. (Debe obtener tres pruebas)

### Paso 3. Presentación del informe de actividades

- Realizar una presentación en Genially, PowerPoint u otra herramienta similar, en donde muestre en cada diapositiva los ítems solicitados en el paso 2, en diez diapositivas, así:
  - I. Primera: portada, título, autor, fecha.
  - II. Segunda: debe diseñar y presentar un objetivo general que sea coherente con la problemática propuesta en el paso 1 y los análisis estadísticos hecho en el paso 2 y en la misma diapositiva redactar tres objetivos específicos del análisis estadístico realizado en el paso 2, donde el primer objetivo se relacione con el análisis unidimensional, el segundo objetivo con el análisis bidimensional y el tercer objetivo con el análisis inferencial. Para el diseño de los objetivos (general y específico) es obligatorio seguir las indicaciones, a continuación:

### Verbo en Infinitivo + ¿Qué? + ¿Cómo? + ¿Para qué?

Estructura de un Objetivo	Ejemplo
Verbo en Infinito	Determinar
¿Qué?	El número de errores en la elaboración de un objetivo general
¿Cómo?	Mediante la revisión de su estructura gramatical
¿Para?	Para corregir los errores que pueda presentar

Además, debe revisar el verbo en infinitivo que va a utilizar de acuerdo con los niveles de la taxonomía de Bloom, que van de izquierda a derecha, es decir desde el más bajo hasta el más alto:

**Imagen 1.** Tabla periódica de Actividades Digitales según la Taxonomía de Bloom.

# Tabla periódica de Actividades Digitales según la Taxonomía de Bloom

De Describir										Blg Bloguear	
Li Listar	Mc Marcar										
Lo Localizar	Co Conectar										
Ds Destacar	Re Resumir	Com Comentar	Reg Registrar en diario	Ac Actuar	Il Ilustrar	Exa Examinar	Pl Planear	Va Validar	Op Expresar Opiniones	Mdf Modificar	Pdc Podcastear
Me Memorizar	Pa Parafrasear	Pe Perfilar	Exp Explicar	Ed Editar	En Entrevistar	Ev Evaluar	Dm Desmenuzar	Rcm Recomendar	Rfl Reflexionar	Pb Publicar	Rol Gamificar con roles
Fa Marcar Favoritos	Cmp Comparar	Pr Predecir	Ca Cargar Online	Co Compartir	Rv Revisión y corregir	In Inspeccionar	Dec Deconstruir	Inf Informar	Gr Graduar	Vbg Video Blogear	Wk Construir Wiki
Nu Numerar	Ds Demostrar	Id Identificar	Exp Explicar	Con Construir	Ar Articular	Ded Deducir	Ctg Categorizar	En Enlazar	Es Aplicar Escalas	Hip Hacer Hipótesis	Prd Producir

Descripción: La imagen muestra una "Tabla periódica de Actividades Digitales según la Taxonomía de Bloom", que clasifica actividades de aprendizaje digital en seis niveles cognitivos, cada uno representado por un color distinto: Recordar (azul) incluye actividades como "Listar" y "Memorizar"; Entender (celeste) abarca actividades como "Resumir" y "Explicar"; Aplicar (verde) contiene actividades como "Actuar" y "Demostrar"; Analizar (amarillo)



comprende actividades como "Categorizar" e "Informar"; Evaluar (rojo) abarca actividades como "Criticar" y "Reflexionar"; y Crear (morado) incluye actividades como "Bloggear" e "Inventar". La tabla facilita la selección de actividades digitales para alcanzar diversos objetivos de aprendizaje.  
Fuente: <https://tuguia de aprendizaje.co/taxonomia-de-bloom-actualizada/>

Es obligatorio que el objetivo general y los objetivos específicos estén en la misma categoría, es decir en el mismo color, por ejemplo, si escoge de objetivo general aplicar, puede escoger de específicos, explicar, examinar, construir, etc.

Tercera a la novena: presentar cada una de las actividades del paso 2, resumidas y organizadas. Cada actividad se debe colocar en una diapositiva, en este caso serían siete diapositivas.

- III. Décima: presentar mínimo seis conclusiones con los resultados de las actividades propuestas, dos para cada análisis estadístico realizado (unidimensional, bidimensional e inferencial), que sean coherentes y estén respaldadas por los resultados estadísticos más relevantes obtenidos en el desarrollo de las actividades propuestas

**En total deber realizar diez diapositivas, no debe hacer más, ni menos.**

- Sustentar la presentación del informe de actividades, mediante una grabación con una duración máxima de 8 minutos, es decir máximo 50 segundos por diapositivas, solamente hable los hallazgos más importantes, prepare su presentación, sea preciso, evite la repetición o la redundancia, no lea diapositivas, solo diseñelas como guía para los exponer los resultados de su análisis estadístico.
- Generar un enlace en plataformas como youtube o Loom con acceso público y dejarlo en el Anexo 1 – Plantilla Entrega Tarea, en el orden indicado en el paso 5.

#### **Paso 4. Aportes en el foro**

El estudiante procede con el desarrollo de las actividades y presenta los aportes en el foro de los avances que vaya realizando durante los pasos 2 y 3, teniendo en cuenta los siguientes parámetros:

- Se definen como aportes en el foro a los documentos adjuntos en Word o PDF con screenshots, donde se presentan avances del desarrollo de las actividades

propuestas en los pasos 2 y 3 **(No subir ningún tipo de archivo ipynb o r)**, para proteger la originalidad de su código.

- Colocar los aportes dentro del **Anexo 1 – Plantilla Entrega Tarea** que se encuentra en la carpeta: Entorno de aprendizaje-->Guías de actividades y rúbrica de evaluación - Tarea 1, siguiendo las indicaciones de esta plantilla. Cada aporte se define como una actividad desarrollada descritas en los pasos 2 y 3.
- Recuerde que para recibir retroalimentación de sus aportes por parte del tutor/a en relación con esta actividad, es necesario que sus aportes sean realizados con **al menos 3 días hábiles de anticipación al cierre, para alcanzar a recibir retroalimentación oportuna de su tutor/a en el foro, en caso contrario no será posible dejar esta retroalimentación.**

### Paso 5 – Compilación trabajo final – Informe de actividades

Cada estudiante de forma **individual** debe compilar el desarrollo del informe de actividades mediante **un documento en formato Word o PDF solamente** en **Anexo 1 – Plantilla entrega Tarea**, que se encuentra en la carpeta del campus siguiendo la ruta Entorno del aprendizaje -> Momento intermedio -> Guía de actividades y rúbrica de evaluación - Unidad 1 - Tarea 1 – Análisis Exploratorio e Inferencial y debe contar con los siguientes elementos:

- Portada.
- Objetivo General.
- Objetivos Específicos.
- Desarrollo del informe de actividades (Paso 2): colocar los códigos utilizados en Python o R para generar el desarrollo de cada una de las actividades propuestas.
- Conclusiones. (Paso 3)
- Enlace de la sustentación del informe (Paso 3)
- Referencias Bibliográficas en normas APA.

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

En el entorno de Aprendizaje debe: Consultar el material bibliográfico de la primera unidad del curso.

En el entorno de Evaluación debe: Entrega del informe de actividades en el Anexo 1 en formato Word o PDF únicamente.

### Evidencias de trabajo independiente:

Las evidencias de trabajo independiente para entregar son:

- Aportes de las actividades en el foro de la unidad.

- Entrega del informe de actividades en el Anexo 1 en Word o PDF únicamente.
- Enlace de la presentación del informe de actividades.

**Evidencias de trabajo grupal:**

En esta actividad no se requieren evidencias de trabajo grupal.

## Lineamientos generales para la elaboración de las evidencias de aprendizaje a entregar.

Para evidencias elaboradas **independientemente**, tenga en cuenta las siguientes orientaciones

1. De forma individual cada estudiante debe subir el desarrollo del informe de actividades mostrando los procedimientos e interpretación de manera clara y concisa de cada uno de los resultados.
2. Recuerde realizar la grabación de la presentación subir el informe de actividades en la plantilla Anexo 1.
3. realizar la grabación de la presentación, compartir un enlace, compruebe el funcionamiento correcto del mismo, que sea de acceso público usando su navegador en modo incognito.

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.

### 3. Formato de Rúbrica de evaluación

<b>Tipo de actividad: Independiente</b>	
<b>Momento de la evaluación: Intermedio</b>	
<b>La máxima puntuación posible es de 100 puntos</b>	
<b>Primer criterio de evaluación:</b>  <b>Procedimiento</b> Realiza las actividades propuestas para llevar a cabo un adecuado análisis estadístico unidimensional que permita describir cada una de las variables presentes en la base de datos.  <b>Este criterio representa 15 puntos del total de 100 puntos de la actividad.</b>	<p><b>Nivel Alto:</b> El estudiante realiza todas las actividades propuestas para llevar a cabo un adecuado análisis estadístico unidimensional que permita describir cada una de las variables presentes en la base de datos.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 12 puntos y 15 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante realiza tres de las actividades propuestas para llevar a cabo un análisis estadístico unidimensional que permita describir cada una de las variables presentes en la base de datos.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 11 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante realiza una o dos de las actividades propuestas, lo que hace que el análisis estadístico unidimensional esté incompleto o no entrega ninguna de las actividades propuestas.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 8 puntos</b></p>
<b>Segundo criterio de evaluación:</b>  <b>Procedimiento</b> Realiza las actividades propuestas para llevar a cabo un adecuado análisis estadístico bidimensional que permita describir las	<p><b>Nivel Alto:</b> El estudiante realiza todas las actividades propuestas para llevar a cabo un adecuado análisis estadístico bidimensional que permita describir las correlaciones o asociaciones entre las variables de interés presentes en la base de datos.  <b>Si su trabajo se encuentra en este nivel puede obtener entre 12 puntos y 15 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante realiza tres de las actividades propuestas para llevar a cabo un adecuado análisis estadístico bidimensional que permita describir las correlaciones o</p>

<p>correlaciones o asociaciones entre las variables de interés presentes en la base de datos.</p> <p><b>Este criterio representa 15 puntos del total de 100 puntos de la actividad</b></p>	<p>asociaciones entre las variables de interés presentes en la base de datos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 9 puntos y 11 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante realiza una o dos actividades, lo que hace que no esté completo el análisis estadístico bidimensional o no entrega ninguna de las actividades propuestas.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 8 puntos</b></p>
<p><b>Tercer criterio de evaluación:</b></p> <p><b>Procedimiento</b> Realiza las actividades propuestas para un análisis inferencial adecuado que permita verificar estadísticamente las diferencias de medias y las correlaciones entre las variables de interés en la base de datos.</p> <p><b>Este criterio representa 20 puntos del total de 100 puntos de la actividad</b></p>	<p><b>Nivel Alto:</b> El estudiante realiza todas las actividades propuestas para llevar a cabo un adecuado análisis inferencial que permita verificar estadísticamente las diferencias de medias y las correlaciones o asociaciones entre las variables de interés presentes en la base de datos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 17 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante realiza tres de las actividades propuestas para llevar a cabo un adecuado análisis inferencial que permita verificar estadísticamente las diferencias de medias y las correlaciones o asociaciones entre las variables de interés presentes en la base de datos.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 14 puntos y 16 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante realiza uno o dos de las actividades propuestas, lo que hace que el análisis inferencial esté incompleto o no entrega ninguna de las actividades propuestas.</p> <p><b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 13 puntos</b></p>

<p><b>Cuarto criterio de evaluación:</b></p> <p><b>Contenido</b> Diseña un objetivo general coherente con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3.</p> <p><b>Este criterio representa 10 puntos del total de 100 puntos de la actividad</b></p>	<p><b>Nivel Alto:</b> El estudiante diseña un objetivo general coherente con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3. <b>Si su trabajo se encuentra en este nivel puede obtener entre 9 y 10 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante diseña un objetivo general, pero le falta un poco de coherente con el desarrollo de las actividades propuesto, sin embargo, sigue las indicaciones del paso 3. <b>Si su trabajo se encuentra en este nivel puede obtener entre 7 y 8 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante diseña un objetivo general, pero no es coherente con el desarrollo de las actividades propuesto y/o no sigue las indicaciones del paso 3, o no diseña ningún el objetivo general. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 y 6 puntos</b></p>
<p><b>Quinto criterio de evaluación:</b></p> <p><b>Contenido</b> Diseña tres objetivos específicos que sean coherentes con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3.</p> <p><b>Este criterio representa 10 puntos del total de 100 puntos de la actividad</b></p>	<p><b>Nivel Alto:</b> El estudiante diseña tres objetivos específicos coherente con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3. <b>Si su trabajo se encuentra en este nivel puede obtener entre 9 y 10 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante diseña dos objetivos específicos coherente con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3. <b>Si su trabajo se encuentra en este nivel puede obtener entre 7 y 8 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante diseña un objetivo específico coherente con el desarrollo de las actividades propuesto, siguiendo las indicaciones del paso 3. El estudiante diseña los objetivos específicos, de forma incoherente con el desarrollo de las actividades propuesto y/o no sigue las indicaciones del paso 3, o no diseña ningún objetivo específico. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 6 puntos</b></p>

<p><b>Sexto criterio de evaluación:</b></p> <p><b>Contenido</b> Elabora seis conclusiones, dos para cada análisis (unidimensional, bidimensional e inferencial), coherentes y respaldadas por los resultados estadísticos más relevantes obtenidos en las actividades propuestas.</p> <p><b>Este criterio representa 10 puntos del total de 100 puntos de la actividad</b></p>	<p><b>Nivel Alto:</b> El estudiante elabora seis conclusiones, una para cada análisis estadístico realizado (unidimensional, bidimensional e inferencial), que sean coherentes y estén respaldadas por los resultados estadísticos más relevantes obtenidos en el desarrollo de las actividades propuestas. <b>Si su trabajo se encuentra en este nivel puede obtener entre 9 y 10 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante elabora entre cuatro y cinco conclusiones, que sean coherentes y estén respaldadas por los resultados estadísticos más relevantes obtenidos en el desarrollo de las actividades propuestas. <b>Si su trabajo se encuentra en este nivel puede obtener entre 7 y 8 puntos</b></p> <p><b>Nivel Bajo:</b> El estudiante elabora entre dos y tres conclusiones, que sean coherentes y estén respaldadas por los resultados estadísticos más relevantes obtenidos en el desarrollo de las actividades propuestas. El estudiante elabora una sola conclusión bien realizada o elabora conclusiones que carecen de incoherente con el desarrollo de las actividades propuestas o no establece ninguna conclusión. <b>Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 6 puntos</b></p>
<p><b>Séptimo criterio de evaluación:</b></p> <p><b>Contenido</b> Sustenta e interpreta los resultados estadísticos obtenidos en el desarrollo de las actividades propuestas a través de la presentación en diapositivas de un informe de actividades, siguiendo</p>	<p><b>Nivel Alto:</b> El estudiante sustenta e interpreta los resultados estadísticos obtenidos en el desarrollo de las actividades propuestas a través de la presentación en diapositivas de un informe de actividades, siguiendo las indicaciones del paso 3. <b>Si su trabajo se encuentra en este nivel puede obtener entre 17 puntos y 20 puntos</b></p> <p><b>Nivel Medio:</b> El estudiante sustenta e interpreta los resultados estadísticos obtenidos en el desarrollo de las actividades propuestas a través de la presentación en diapositivas de un informe de actividades, siguiendo las indicaciones del paso 3, tiene algunas imprecisiones en la interpretación. <b>Si su trabajo se encuentra en este nivel puede obtener entre 14 puntos y 16 puntos</b></p>



las indicaciones del  
paso 3.

**Este criterio  
representa 20  
puntos del total de  
100 puntos de la  
actividad**

**Nivel Bajo:** El estudiante sustenta e interpreta los resultados estadísticos obtenidos en el desarrollo de las actividades propuestas a través de la presentación en diapositivas de un informe de actividades, pero tiene incorrecta la interpretación de los resultados. El estudiante sustenta e interpreta los resultados estadísticos obtenidos en el desarrollo de las actividades propuestas a través de la presentación en diapositivas de un informe de actividades, de manera incompleta o no entrega la sustentación.

**Si su trabajo se encuentra en este nivel puede obtener entre 0 puntos y 13 puntos**