**FORMATO PARA EL DESARROLLO DE COMPONENTE FORMATIVO**

| **PROGRAMA DE FORMACIÓN** | Análisis exploratorio de datos en Python |
| --- | --- |

| **COMPETENCIA** | 220501114. Sistematizar datos masivos de acuerdo con métodos de analítica y herramientas tecnológicas. | **RESULTADOS DE APRENDIZAJE** | 220501114-1. Definir el objetivo del análisis a partir del planteamiento de preguntas que permitan conocer los datos recopilados. |
| --- | --- | --- | --- |

| **NÚMERO DEL COMPONENTE FORMATIVO** | 01 |
| --- | --- |
| **NOMBRE DEL COMPONENTE FORMATIVO** | Análisis de datos: el futuro de las organizaciones |
| **BREVE DESCRIPCIÓN** | Este componente proporciona información para aprender a implementar proyectos de ciencia de datos desde las metodologías existentes para su análisis y a partir de la identificación de las necesidades y los requerimientos de una organización.  Además, permite comprender los conceptos básicos de la estadística y medidas de tendencia central, que son fundamentales para desarrollar un análisis exploratorio de datos de calidad. |
| **PALABRAS CLAVE** | Datos, estadística, gráficos, ideación, metodología, variables |

| **ÁREA OCUPACIONAL** | 6 - Ventas y servicios |
| --- | --- |
| **IDIOMA** | Español |

1. **TABLA DE CONTENIDOS:**

**Introducción**

**1. Importancia del análisis exploratorio de datos**

**2. Metodología de la ciencia de datos**

**3. Identificación de problemas y oportunidades de negocio**

**4. Estadística**

4.1 Media

4.2 Mediana

4.3 Moda

4.4 Tablas de frecuencia

4.5 Varianza

4.6 Desviación estándar

**5. Preparación de los datos**

1. **DESARROLLO DE CONTENIDOS:**

**Introducción:**

Bienvenido a este componente formativo denominado **“Análisis de datos: el futuro de las organizaciones”,** el cual tiene como fin dar a conocer los elementos, metodologías y procedimientos necesarios para llevar a cabo un análisis de datos efectivo dentro de las organizaciones.

Como actividad inicial, se recomienda visitar el recurso didáctico que se propone a continuación:



1. **Importancia del análisis exploratorio de datos**

Para comenzar es necesario recalcar que en la actualidad los datos se han convertido en el activo más importante para cualquier tipo de organización (grande, mediana, pequeña), pues su análisis es fundamental en la toma de decisiones estratégicas que beneficien a todos los implicados:



Para conocer algunas de las razones por la cuales el análisis de datos resulta indispensable para el éxito de las organizaciones, se invita a revisar con atención el siguiente recurso educativo:



Para realizar el análisis exploratorio de datos, se utilizan diversas herramientas tecnológicas entre las que se encuentran Anaconda y Júpiter Lab, desde allí se despliega el lenguaje de programación Python y las librerías necesarias para facilitar su lectura, manipulación y aplicación de métodos estadísticos que permiten la visualización de los datos en forma de tablas y gráficos y por ende su interpretación.

Por otro lado, existen diversas formas para realizar el análisis de datos exploratorio, entre las que se encuentran:

**Tabla 1**

*Formas para realizar el análisis exploratorio*

| Forma | Descripción | |
| --- | --- | --- |
| Univariado | Se analiza una sola variable. |  |
| Bivariado | Permite verificar el comportamiento de dos variables, se puede estudiar una variable en función de la otra |  |
| Multivariada | Cuando se combinan más de dos datos, aborda todos los indicadores disponibles para el análisis de un fenómeno determinado. |  |

En este orden de ideas, se puede afirmar que el objetivo principal del análisis exploratorio de datos es tener una visión clara de los datos que maneja la organización para, posteriormente, indicar la técnica de ciencia de datos más adecuada para su tratamiento.

Para que una organización alcance lo objetivos que se ha propuesto es necesario que organice sus datos, entienda su contenido, comprenda cuáles son las variables más importantes y cómo estas se relacionan una con otra, comenzar a identificar patrones, establecer qué hacer con los datos atípicos y, finalmente, llegar a conclusiones acerca de dicho análisis que permitan la toma de decisiones acertadas frente a los procesos organizacionales.

Todo lo anterior rescata la importancia del análisis exploratorio de datos, que se resume en la forma de entender, visualizar y extraer información relevante de un conjunto de datos para establecer la ruta o técnica más adecuada para procesarlos.

A continuación, se podrá conocer a más detalle los pasos necesarios para el análisis exploratorio de datos.



1. **Metodología para la ciencia de datos**

El objetivo principal de esta metodología consiste en la planeación, de manera organizada, de las metas, proyectos o implementaciones necesarias para que cualquier tipo de tarea, que requiera de un tratamiento específico, funcione de manera correcta.

Desde esta perspectiva, una metodología enfocada en la ciencia de datos permite definir la ruta a seguir para la debida organización y el procesamiento adecuado de los datos.



Una metodología de ciencia de datos permite llevar de manera organizada cada una de las etapas necesarias para desarrollar un proyecto de ciencia de datos y, de esta manera, garantizar el cumplimento de los objetivos, que se han propuesto sobre este tema.

Esta se enmarca en tres acciones principales, las cuales se observan en la siguiente figura

**Figura 1**

*Acciones para la metodología de la ciencia de datos*



Dentro de la implementación de dicha metodología se deben tener en cuenta las actividades propias de gestión de proyecto, entre las que se encuentran:

* Inicio.
* Planeación.
* Ejecución.
* Cierre.

Para implementar la metodología ASUM es necesario tener en cuenta las siguientes fases:

1. **Fase 1: inicio del proyecto**. Lo primero es realizar un levantamiento de la información, se debe hacer un diagnóstico respecto a qué tan preparada está la compañía para iniciarse en un proyecto de analítica, dentro de esta etapa se deben identificar factores como:
   * Tecnologías utilizadas.
   * Fuentes de datos.
   * Acceso a la información.
   * Volumen de información.
   * Instalaciones.

Para conocer los roles necesarios con miras al inicio con el desarrollo de un proyecto de análisis de datos, se debe revisar el siguiente recurso educativo:



1. **Fase 2: entender el negocio**. En esta etapa es importante entender cuáles son esos objetivos, pero, además, es fundamental reconocer los requerimientos y las necesidades organizacionales, al identificar este aspecto se podrán reconocer aquellos problemas que podrían resolverse con la implementación de la analítica:

Para conocer las principales características de esta etapa, se invita a explorar el siguiente recurso educativo:



1. **Fase 3: descubrimiento y entendimiento de los datos.** Se debe realizar la recolección de datos iniciales para determinar la consistencia de la información, se espera poder describirlos, realizar todas las exploraciones planteadas y verificar la calidad de los mismos.

Los datos iniciales pueden provenir de diferentes fuentes de información, como se observa en la siguiente figura:

**Figura 2**

*Fuentes de información de los datos*



Posteriormente, es necesario describir los datos presentes con cada colección analizada, generalmente estos deben estar organizados en filas y columnas, para esto se debe realizar la identificación de las columnas y determinar cuáles de estas son datos categóricos y cuales son numéricos, también es importante establecer algunas medidas de tendencia central como la moda, media, mediana, varianzas, máximos y mínimos.

Por otro lado, es fundamental verificar la calidad de los datos teniendo en cuenta las siguientes preguntas:

| Gente plana haciendo preguntas ilustración vector gratuito |
| --- |



1. **Fase 4: preparando los datos.** Esta es una de las etapas que ocupa mayor cantidad de tiempo, pues, normalmente, los datos no cumplen con los parámetros requeridos para empezar a trabajar con ellos de manera adecuada, por ende, es necesario que en la fase anterior se realice un muy buen trabajo de investigación que permita avanzar rápidamente.

En este paso se deben tener en cuenta las siguientes operaciones:

* Seleccionar los datos.
* Crear la estrategia de limpieza de datos.
* Completar los datos.
* Integrar los datos (cuando se obtienen datos de diferentes tipos de fuentes).
* Es primordial que al integrar los datos se maneje el mismo tipo de formato.

Finalmente, dentro de esta fase es importante realizar un informe en el que se describen las diferentes situaciones que se encontraron y se inscriban, si es necesario, dentro del marco de riesgo en caso de ser así, de lo contrario igualmente se deben reportar los avances y los resultados positivos.

1. **Fase 5: construir el modelo.** Al inicio del proyecto se plantearon unos objetivos o metas, con la construcción de un modelo se pretende estructurar todo lo necesario para lograr alcanzarlos, en esta fase se pueden requerir varias interacciones que permite ir afinando el modelo, es muy probable que sea necesario regresar hacia atrás y volver a la fase de preparación de los datos para realizar los ajustes necesarios.

Se invita a revisar con atención el siguiente recurso educativo para identificar los pasos que se deben tener en cuenta para la construcción del modelo:



1. **Fase 6. evaluar el modelo.** Con los criterios de evaluación definidos y los resultados que se obtuvieron del modelo generado anteriormente, se realiza un diagnóstico que permite determinar si el modelo utilizado es el indicado para el negocio.

**En la evaluación de resultados, se deben tener en cuenta las siguientes consideraciones:**

* El resultado responde a las metas definidas por el negocio.
* Se realizar un mejor entendimiento de los datos de la compañía.
* Se genera nueva información de valor que antes del proyecto de analítica no existía
* no responde a las expectativas.

Finalmente, el siguiente recurso educativo brindará claridad respecto a los procesos que se deben tener en cuenta para realizar una revisión final de todas las etapas.



**3. Identificación de problemas y oportunidades de negocio**

Para comenzar con el análisis de datos exploratorio es primordial, primero, conocer el negocio o la organización para la cual se está desarrollando el análisis, además, deben ser muy claros los objetivos, requerimientos y metas de la compañía.

*Design Thinking* es una metodología que utiliza herramientas creativas para identificar las necesidades y requerimientos de los usuarios o de los clientes y, desde allí, desarrollar ideas innovadoras para lograr satisfacerlas.

A continuación, se define de forma específica las fases que contempla esta metodología, las cuales se describirán más adelante, así:



* **Empatizar**: cuando se **empatiza** lo que se quiere es tratar de estar en la misma situación de la otra persona, tratar de entender su situación, de pensar y sentir como esa persona lo haría.

| Ejercicio reflexivo  Haga de cuenta que se tiene una tienda de café, se intenta saber qué piensan los clientes o potenciales clientes al visitar el sitio, qué les llama la atención de visitar su tienda y no otra: puede ser el precio, la atención, la calidad del producto, el lugar en donde está situado, el ambiente que allí se disfruta. |  |
| --- | --- |

* **Definir**: en la etapa de **definir**, lo que se quiere es depurar toda esa información recopilada al momento de empatizar y quedarse con lo que realmente tiene un valor significativo.

| Ejercicio reflexivo  ¿La edad de las personas influye en la tienda de café?  ¿La calidad de atención a la clientela, es lo que se considera de mayor valor intangible en su negocio?  . |  |
| --- | --- |

Cada negocio realiza sus juicios de valor y dará mayor importancia a esas variables que van encaminadas con los objetivos y metas del negocio.

* **Idear**: en la etapa de **ideación** se deben contemplar todas las posibilidades que se tienen y plantear soluciones a los problemas identificados, por absurda que parezca la idea no se debe dejar de plantearla, pues las ideas más locas pueden terminar en cambios significativamente positivos para la compañía.

| Ejercicio reflexivo  Es un valor agregado que, en la tienda de café, los clientes puedan moler y llevar a preparación la bebida que van a consumir. |  |
| --- | --- |

* **Prototipo:** realice un **prototipo** de lo que considera sería una solución a su problema, no debe ser nada elaborado, incluso si tiene a mano lápiz y papel proceda a plasmar todas las ideas que tenga en mente.

| Ejercicio reflexivo  Los clientes del negocio, ¿disfrutan la preparación de producto, tanto para ellos mismos como para otros clientes? |  |
| --- | --- |

Para la etapa de empatía se puede hacer uso de la herramienta denominada mapa de empatía la cual ayuda a identificar las posibles variables que afectan el negocio. El siguiente recurso educativo permite visualizar cómo se puede aplicar dicha herramienta:





Cada negocio tendrá su estrategia para la recopilación de los datos que le permitan realizar cuadros de empatía, la generación de dolores y necesidades de sus actuales o futuros clientes.

Anteriormente se vio que el paso a seguir debe ser idear las posibles soluciones y crear prototipos que propendan con la puesta en marcha de planes de mejoramiento.

Siguiendo el ejemplo anterior, el negocio de café tendrá que poner mucha atención a dos factores que pueden estar afectando significativamente el negocio como:

* El mal servicio.
* Las instalaciones.

Por ende, se tendrá que realizar un análisis para validar si estas dos variables están ligadas y una desencadena en la otra, estos análisis se pueden determinar mediante un análisis observatorio de los datos y algunos cálculos estadísticos, o se pueden implementar modelos más complejos cuando la relación entre variables no es fácil de detectar.

1. **Estadística**

Para realizar un análisis de datos efectivo es fundamental reconocer diferentes medidas que se utilizan en la estadística y que resultan fundamentales para llevar a cabo dicha labor:

* 1. **Media**

Consiste en realizar la sumatoria de todos los elementos de una lista de datos que sean numéricos y dividirlos sobre el total de elementos, por ejemplo, si se quiere saber la media de las calificaciones de un curso de la materia de Python cuya lista se muestra a continuación.

**Tabla 2**

*Media de calificaciones*

| N.° | Nombre | Nota |
| --- | --- | --- |
| 1 | Estudiante 1 | 3 |
| 2 | Estudiante 2 | 3,5 |
| 3 | Estudiante 3 | 4,5 |
| 4 | Estudiante 4 | 5 |
| 5 | Estudiante 5 | 4 |
| 6 | Estudiante 6 | 5 |
| 7 | Estudiante 7 | 3 |
| 8 | Estudiante 8 | 4,5 |
| 9 | Estudiante 9 | 3 |
| 10 | Estudiante 10 | 4 |
|  | SUMA TOTAL 39,5 | |

La sumatoria total arroja **39,5** si ese resultado se divide entre el total de estudiantes que son **10**, el resultado no dará la media.

**Media = 39,5 / 10 = 3,95**

Con el análisis de la media se debe estar muy atentos, ya que es muy sensible a arrojar datos altos, por ejemplo:

**Tabla 3**

*Media: datos altos*

| N.° | Nombre | Nota |
| --- | --- | --- |
| 1 | Estudiante 1 | 1 |
| 2 | Estudiante 2 | 2 |
| 3 | Estudiante 3 | 2 |
| 4 | Estudiante 4 | 2 |
| 5 | Estudiante 5 | 2 |
| 6 | Estudiante 6 | 2,9 |
| 7 | Estudiante 7 | 2,9 |
| 8 | Estudiante 8 | 5 |
| 9 | Estudiante 9 | 5 |
| 10 | Estudiante 10 | 5 |
|  | SUMA TOTAL 30 | |

Como se observa en el cuadro anterior, a pesar de que el **70%** presenta una valoración baja y con solo **30%** de los datos con valoración alta, el comportamiento de la media tiende a subir.

**Si se divide media = 30/10 = 3,0**

Y si el parámetro de análisis valida que de 3,0 en adelante es una medida aceptable, entonces muy probablemente se incurrirá en errores.

* 1. **Mediana**

Con el cálculo de la mediana se mejora la relación del problema presentado en la media. Esta medida consiste en organizar los datos y tomar el valor que divide los datos en dos partes igual, para esta acción se deben tener en cuenta dos consideraciones:

* **Cuando los datos son impares:** se toma el valor central dejando igual número de registros a ambos lados.

**Tabla 4**

*Mediana: datos impares*

| N.° | Nombre | Nota |
| --- | --- | --- |
| 1 | Estudiante 1 | 1 |
| 2 | Estudiante 2 | 2 |
| 3 | Estudiante 3 | 2 |
| 4 | Estudiante 4 | 2 |
| 5 | Estudiante 5 | 2 |
| 11 | Estudiante 11 | 2 |
| 6 | Estudiante 6 | 2,9 |
| 7 | Estudiante 7 | 2,9 |
| 8 | Estudiante 8 | 5 |
| 9 | Estudiante 9 | 5 |
| 10 | Estudiante 10 | 5 |
|  | SUMA TOTAL 32 | |

El valor de la mediana para este conjunto de datos sería **2**, teniendo en cuenta que los datos son impares y así quedarían la misma cantidad de registros al realizar la división.

* **Cuando los datos son pares:** por otro lado, lo que se debe realizar cuando los registros son pares es realizar la suma de los dos datos centrales y dividirlos por 2.

**Tabla 5**

*Mediana: datos pares*

| N.° | Nombre | Nota |  |
| --- | --- | --- | --- |
| 1 | Estudiante 1 | 1 |
| 2 | Estudiante 2 | 2 |
| 3 | Estudiante 3 | 2 |
| 4 | Estudiante 4 | 2 |
| 5 | Estudiante 5 | 2 | 2,45 |
| 6 | Estudiante 6 | 2,9 |
| 7 | Estudiante 7 | 2,9 |  |
| 8 | Estudiante 8 | 5 |
| 9 | Estudiante 9 | 5 |
| 10 | Estudiante 10 | 5 |
|  | SUMA TOTAL 30 | |

Como se observa en la tabla, se toman los dos datos centrales dejando igual cantidad de registros en ambas partes, se suman los números centrales y se divide entre 2 es decir se les aplica la **media.**

**Mediana = (2 + 2,9) / 2 = 4,9 / 2 = 2,45**

* 1. **Moda**

Es el valor con mayor frecuencia en la lista de datos, se pueden agrupar todas las posibilidades presentes en la lista y se cuenta el número de veces que se repite:

**Tabla 6**

*Moda*

| No. | Nombre | Nota |
| --- | --- | --- |
| 1 | Estudiante 1 | 1 |
| 2 | Estudiante 2 | 2 |
| 3 | Estudiante 3 | 2 |
| 4 | Estudiante 4 | 2 |
| 5 | Estudiante 5 | 2 |
| 6 | Estudiante 6 | 2,9 |
| 7 | Estudiante 7 | 2,9 |
| 8 | Estudiante 8 | 5 |
| 9 | Estudiante 9 | 5 |
| 10 | Estudiante 10 | 5 |
|  | SUMA TOTAL 30 | |
| Nota | Cantidad |
| 1 | 1 |
| 2 | 4 |
| 2,9 | 2 |
| 5 | 3 |



* 1. **Tablas de frecuencias**

Estas permiten agrupar información de manera que se puedan entender mejor los volúmenes de datos.

* En ese sentido, se encuentran las **frecuencias absolutas** que sirven para representar el número de veces que se repiten los elementos o grupos dentro de una serie de datos.

**Ejemplo**: se requiere saber la edad de los empleados que trabajan para la compañía a nivel nacional, para ellos se extrae la información del sistema de información de recursos humanos. Se debe tener en cuenta la edad por rangos de la siguiente manera:

**Tabla 7**

*Tablas de frecuencia*

| Edad | Total de empleados |
| --- | --- |
| 18 a 25 |  |
| 25 a 35 |  |
| 35 a 40 |  |
| 40 a 50 |  |
| Más de 50 |  |

Los resultados arrojan la siguiente frecuencia absoluta, después de procesar los datos y agrupar en el rango indicado:

**Tabla 8**

*Frecuencia absoluta*

| Edad | Total de empleados |
| --- | --- |
| 18 a 25 | 50 |
| 25 a 35 | 110 |
| 35 a 40 | 15 |
| 40 a 50 | 20 |
| Más de 50 | 5 |
| Total | 200 |

* **La frecuencia absoluta acumulada** corresponde a sumar cada uno de los elementos o grupos de la siguiente manera, en el último elemento, grupo o rango se muestra el total de registro, en este caso de 200, que coincide con el total general, para el primer registro se deja el mismo valor.

**Tabla 9**

*Frecuencia absoluta acumulada*

| Edad | Total de empleados | Acumulado |
| --- | --- | --- |
| 18 a 25 | 50 | 50 |
| 25 a 35 | 110 | 160 |
| 35 a 40 | 15 | 175 |
| 40 a 50 | 20 | 195 |
| Más de 50 | 5 | 200 |
| Total | 200 |  |

*Nota*: autoría propia

* Se realizan los cálculos de los datos en porcentajes lo que permitirá entenderlos mejor, para ello se divide cada uno de los elementos de la frecuencia absoluta y acumulada sobre el total de los registros, a este paso se le denomina **frecuencia relativa**.

El procedimiento es dividir cada uno de los registros entre 200 que corresponde al total de los registros así:

| Para total empleados | | |
| --- | --- | --- |
| 50/200 | = |  |
| 110/200 | = |  |
| 15/200 | = |  |
| 20/200 | = |  |
| 5/500 | = |  |

| Y para el acumulado | | |
| --- | --- | --- |
| 50/200 | = |  |
| 160/200 | = |  |
| 175/200 | = |  |
| 195/200 | = |  |
| 200/500 | = |  |

**Tabla 10**

*Frecuencia relativa*

| Edad | Total de empleados | Acumulado | % Total de empleados | Acumulado |
| --- | --- | --- | --- | --- |
| 18 a 25 | 50 | 50 | 0,25 | 0,25 |
| 25 a 35 | 110 | 160 | 0,55 | 0,8 |
| 35 a 40 | 15 | 175 | 0,075 | 0,875 |
| 40 a 50 | 20 | 195 | 0,1 | 0,975 |
| Más de 50 | 5 | 200 | 0,25 | 1 |
| Total | 200 |  |  |  |

Como se puede observar la mayor concentración con 80% de los datos se encuentran en los dos primeros grupos.

* 1. **Varianza**

Se utiliza cuando se quiere saber qué tan dispersos están los datos con respecto a la media, como se observa en la fórmula se eleva al cuadrado lo que evita datos negativos

Texto

Descripción generada automáticamente

* 1. **Desviación estándar**

Se emplea cuando se quiere saber qué tan disperso está un grupo de datos con respecto a la media se representa mediante la siguiente fórmula:

Texto

Descripción generada automáticamente con confianza media

1. **Preparación de los datos**

La preparación de los datos es un paso muy importante para el análisis exploratorio, después de conocer muy bien los objetivos del negocio e identificar las metas de la organización. Es muy importante establecer el camino que se debe seguir para esta preparación.



Normalmente las colecciones de datos se trabajan en archivos con extensión .CSV (valores separados por comas), este tipo de archivo permite que se manejen grandes volúmenes de información, la separación por comas se utiliza para la identificación de las filas y las columnas, como se observa en la siguiente imagen:

**Figura 3**

*Colección de datos*

Texto

Descripción generada automáticamente

Nota. Captura de pantalla, imagen de un archivo CVS vista desde un editor de notas.

También se pueden ver y editar en programas como Excel, para visualizar la información en filas y columnas, es muy común que los datos suelen presentarse en formato de Excel y sea necesario realizar la exportación a CSV.

Si el archivo está en formato CSV y se quiere visualizar a Excel, es necesario seguir algunos pasos, por lo que se invita a revisar el siguiente recurso educativo para conocerlos.



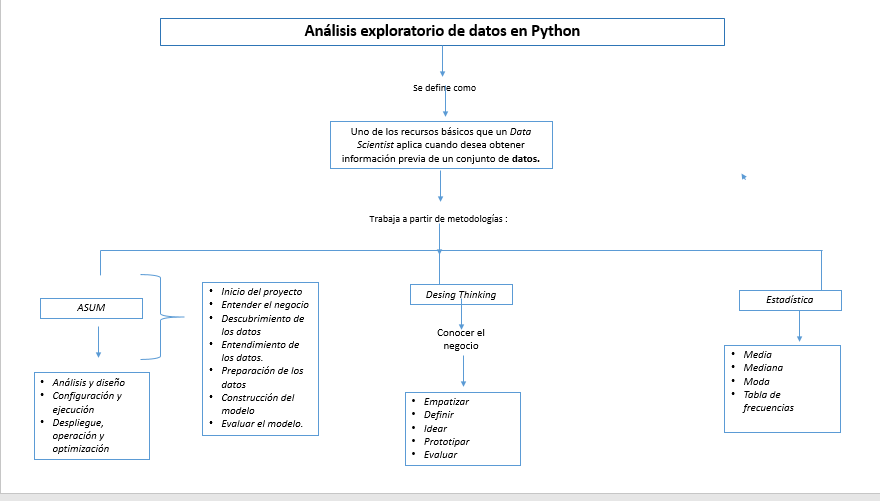
Finalmente, después de preparar los datos se debe realizar la revisión detallada de los mismos y determinar:

* El total de registros.
* Cuáles son las variables categóricas.
* Las variables numéricas.
* Establecer si faltan datos en alguna de las variables.
* Establecer si existen inconsistencia en los registros, por ejemplo, que dentro de la variable edad que es numérica se tenga valores de tipo carácter.
* Establecer si es necesario eliminar los registros que presenta valores vacíos nulos de las variables seleccionadas para el análisis.
* Establecer que datos necesitan ser remplazados.
* Diseñar un plan con las diferentes opciones que, de implementación, puede ser que sobre los mismos datos se puedan realizar dos tipos de análisis.

1. **Síntesis**

En la actualidad las empresas, independientemente de su tamaño, manejan gran cantidad de datos e información que se han ido convirtiendo en la herramienta base para la toma de decisiones estratégicas. Desde esta perspectiva, este componente resulta fundamental al brindar las bases para desarrollar un proyecto de análisis exploratorio de datos. En este componente se aclaran conceptos, se identifican metodologías de aplicación, además se describen estrategias puntuales para la preparación de datos y el análisis estadístico de datos.

Una breve revisión de los temas vistos, se encuentra en el siguiente esquema:



1. **ACTIVIDADES DIDÁCTICAS**

| **DESCRIPCIÓN DE ACTIVIDAD DIDÁCTICA** | |
| --- | --- |
| **Nombre de la Actividad** | Afianzando conceptos |
| **Objetivo de la actividad** | Recordar lo aprendido sobre los pasos a seguir en la preparación del análisis exploratorio de los datos. |
| **Tipo de actividad sugerida** | Sopa de letras |
| **Archivo de la actividad**  **(Anexo donde se describe la actividad propuesta)** | Actividad didáctica 1. CF01 |

| **DESCRIPCIÓN DE ACTIVIDAD DIDÁCTICA** | |
| --- | --- |
| **Nombre de la Actividad** | Relacionando conceptos |
| **Objetivo de la actividad** | Apropiarse de los conceptos necesarios para el análisis exploratorio de datos. |
| **Tipo de actividad sugerida** | Arrastrar y soltar |
| **Archivo de la actividad**  **(Anexo donde se describe la actividad propuesta)** | Actividad didáctica 2. CF01 |

1. **MATERIAL COMPLEMENTARIO:**

| **Tema** | **Referencia APA del Material** | **Tipo de material**  **(Video, capítulo de libro, artículo, otro)** | **Enlace del Recurso o**  **Archivo del documento o material** |
| --- | --- | --- | --- |
| Importancia del análisis exploratorio de datos. | Comunicación numérica. (s.f.). *¿Qué es el análisis exploratorios de datos?* [Video]. YouTube. <https://www.youtube.com/watch?v=UeMpYEktLfU> | Video | <https://www.youtube.com/watch?v=UeMpYEktLfU> |
| Metodología de la ciencia de datos. | Torres, C., Holman, J., y Méndez, J., (2022). *Metodología para la analítica de datos*. <https://ideca.gov.co/sites/default/files/MetodologiaAnaliticaDatos.pdf> | Artículo | <https://ideca.gov.co/sites/default/files/MetodologiaAnaliticaDatos.pdf> |
| Preparación de los datos. | Kaggle.com. (2022). *Height of Male and Female by Country 2022*. <https://www.kaggle.com/majyhain/height-of-male-and-female-by-country-2022> | Página Web | <https://www.kaggle.com/majyhain/height-of-male-and-female-by-country-2022> |

1. **GLOSARIO:**

| **TÉRMINO** | **SIGNIFICADO** |
| --- | --- |
| **Data Sets** | colección de datos organizados en filas y columnas. |
| **Dato** | valor con el que se forman series y colecciones numéricas o categóricas. |
| ***Design Thinking*** | metodología utilizada para realizar ideación en búsqueda de solución a problemas. |
| **Frecuencias** | cantidad con la que se repite un elemento, grupo o rango dentro de una colección. |
| **Media** | valor que se calcula para determinar el promedio de una serie de datos numéricos. |
| **Mediana** | valor central de una serie de datos numéricos ordenados. |
| **Moda** | valor que se presenta con mayor frecuencia en una serie de datos. |

1. **REFERENCIAS BIBLIOGRÁFICAS:**

Arias, H., Rojas, M., Sepúlveda, D., Rojas, C., y Rodríguez, Y. (2019). *Modelo de negocio Design Thinking.* UNAD.

Bustos, M. (2021). Técnicas e instrumentos para recoger datos del hecho social educativo. *Revista Científica Retos de la Ciencia, 5*(10), 50-61.

Cordoví, V., Benito, V., Pruna, L., Muguercia, A., y Antúnez, J. (2018). Aprendizaje de las medidas de tendencia central a través de la herramienta EXeLearning. *Medisan, 22*(3), 257-263.

Ortigoza, A., y Mateus, C. (2021). Aprendizaje basado en proyectos de las medidas de dispersión en busca de la cultura estadística en un contexto rural. *Innovaciones Educativas, 23*(35), 73-87.

Torres, C., Holman, J., y Méndez, J., (2022). *Metodología para la analítica de datos.* Ideca. <https://ideca.gov.co/sites/default/files/MetodologiaAnaliticaDatos.pdf>

1. **CONTROL DEL DOCUMENTO**

|  | **Nombre** | **Cargo** | **Dependencia** | **Fecha** |
| --- | --- | --- | --- | --- |
| **Autor (es)** | Ronald Alexander Vacca Ascanio | Experto Temático | Regional Distrito Capital. Centro de Diseño Metrología | Marzo 2022. |
| Jeimy Lorena Romero Perilla | Diseñador Instruccional | Regional Norte de Santander Centro de la industria, la empresa y los servicios | Marzo 2022. |
| María Inés Machado López | Diseñador Instruccional | Regional Norte de Santander Centro de la industria, la empresa y los servicios | Marzo 2022. |
| Carolina Coca Salazar | Metodóloga | Regional Distrito Capital-Centro de Diseño y Metrología | Marzo 2022. |
|  | José Gabriel Ortiz Abella | Corrector de estilo | Regional Distrito Capital -Centro de Diseño y Metrología. | Abril del 2022. |

1. **CONTROL DE CAMBIOS**

|  | **Nombre** | **Cargo** | **Dependencia** | **Fecha** | **Razón del Cambio** |
| --- | --- | --- | --- | --- | --- |
| **Autor (es)** |  |  |  |  |  |