|  |  |
| --- | --- |
| PROGRAMA DE FORMACIÓN | Gestión de analítica de datos |

|  |  |  |  |
| --- | --- | --- | --- |
| COMPETENCIA | 210601026: Procesar datos de acuerdo con procedimiento técnico y metodología estadística. | RESULTADOS DE APRENDIZAJE | 210601026-02: Manejar datos de acuerdo con metodología estadística |

|  |  |
| --- | --- |
| NÚMERO DEL COMPONENTE FORMATIVO | CF4 |
| NOMBRE DEL COMPONENTE FORMATIVO | Estadística descriptiva, gráficas e informe estadísticos. |
| BREVE DESCRIPCIÓN | En este componente se manejan las bases de la estadística descriptiva, se emplean conceptos fundamentales como el de promedio o media de un conjunto con el objetivo de tener la fundamentación teórica para la compresión del análisis estadístico, utilizando diferentes gráficas como de histogramas, cajas y bigotes y gráfico circular de forma dinámica. Todo esto, presentado de forma detallada con contraste y observación de datos. |
| PALABRAS CLAVE | Estadística descriptiva, gráficos estadísticos, informe estadístico, gráficos, tabla de datos |

|  |  |
| --- | --- |
| ÁREA OCUPACIONAL | Ciencias naturales, aplicadas y relacionadas |
| IDIOMA | Español |

**Tabla de contenido**

**Introducción**

**1**. **Estadística descriptiva.**

**2**. **Organiza una tabla de datos y los tipos de variable.**

**3** **Gráficos.**

**4**. **Elaboración del informe estadístico.**

**Introducción**

|  |
| --- |
| Cuadro de texto |
| Estimado aprendiz, reciba una cordial bienvenida a este recurso de aprendizaje orientado a la estadística descriptiva. Durante su desarrollo, se abordarán los principios básicos, la organización de datos y tipos de variables, y finalmente la realización de informes estadísticos. Al finalizar, obtendrá los conocimientos para desenvolverse en el campo de la estadística descriptiva, específicamente será capaz de clasificar datos muestrales, elaborar tablas, crear gráficas y diseñar un informe. Antes de empezar el desarrollo del componente formativo, se sugiere revisar el siguiente video para que amplíe e identifique el panorama de aprendizaje:  ¡Muchos éxitos, que tenga un aprendizaje significativo! |

**Video introductorio**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Tipo de recurso** | Video spot animado | | | |
| **NOTA** | **La totalidad del texto locutado para el video no debe superar las 500 palabras aproximadamente** | | | |
| **Título** | Estadística descriptiva, gráficas e informe estadísticos. | | | |
| **Escena** | **Imagen** | **Sonido** | **Narración (voz en off)** | **Texto** |
| **1** | Ilustración de concepto informe de datos  228131\_i43 | NA | Actualmente con el auge de las nuevas tecnologías y la digitalización, las personas generan datos diariamente de forma masiva |  |
| **2** | Ilustración del concepto de información de pago  228131\_i434 | NA | mediante transferencias electrónicas, compras en línea, interacciones en las redes sociales, entre otros. Esto es la base para los diferentes análisis en las empresas. | Transferencias electrónicas, pagos y compras en línea |
| **3** | Icono de comunicación plana  228131\_i435 | NA | Las compañías hoy en día han puesto mucha atención a ese flujo inmenso de información producido a nivel global | Existen empresas interesadas |
| **4** | Ilustración del concepto de análisis  228131\_i436 | NA | con la finalidad de obtener esos datos crudos, limpiarlos, procesarlos y analizarlos para la toma de decisiones | En el análisis y procesamiento de datos |
| **5** | Líder empresarial de pie en la flecha y sosteniendo la bandera ilustración vectorial plana. gente de dibujos animados entrenando y haciendo un plan de negocios. concepto de liderazgo, victoria y desafío  228131\_i437 | NA | Teniendo como principales referentes las mejoras de su negocio tomando siempre presente que esto puede distinguirlos de la competencia. | Para innovar en sus procesos |
| **6** | Hombre con laptop analizando infografías. diagrama, gráfico de barras, informe ilustración vectorial plana. análisis, marketing, director de proyectos  228131\_i438 | NA | Para ello, buscan personas que se introduzcan a este mundo de los datos | Apoyados en científicos de datos |
| **7** | 228131\_i439 | NA | con la finalidad de comprenderlos analizarlos y visualizarlos en forma interactiva, | Que los transformen en gráficos |
| **8** | Composición isométrica de análisis de ciencia de big data  228131\_i440 | NA | utilizando las diferentes herramientas analíticas para la futura toma de decisiones. | Utilizando la analítica |
| **9** | Probador de qa. kit de desarrollo. analizando código binario. inspección cercana, codificación, verificación de script abierto. administración de sitios web. reafirmando la calidad. ilustración de metáfora de concepto aislado.  228131\_i441 | NA | Por el volumen de datos masivos, el gobierno facilita esta información libre a la comunidad | Datos abiertos para crear soluciones |
| **10** | Granjero de sexo femenino feliz trabajando en la granja para alimentar a la población ilustración vectorial plana. granja de dibujos animados con tecnología de automatización.  228131\_i442 | NA | con la finalidad de resolver problemas del sector estatal y con el objetivo de solucionar problemas del país | Resolver problemas |
| **11** | 228131\_i443 | NA | pensando en el futuro, mejorar las ciudades a partir de la información y su respectivo análisis de la problemática | Ciudades inteligentes |
| **12** | Probador de qa. kit de desarrollo. analizando código binario. inspección cercana, codificación, verificación de script abierto. administración de sitios web. reafirmando la calidad. ilustración de metáfora de concepto aislado de vector.  228131\_i444 | NA | Las industrias guardan este volumen de información y la estructura en una base de datos | Obtener los datos de bases de datos |
| **13** | Concepto de análisis de datos comerciales  228131\_i445 | NA | mediante proceso de extraer, transformar y cargar los datos | Extracción y transformación |
| **14** | Programadores con laptops trabajando en código y big data. desarrollo de software, procesamiento y análisis de datos, aplicaciones de datos y concepto de gestión. vector ilustración aislada.  228131\_i446 | NA | en una base de datos que contendrá esta información procesada | Datos procesados y organizados |
| **15** | Computadora de uso humano para controlar los brazos del robot que trabajan en producción convocados en la industria de fábricas inteligentes 4  228131\_i447 | NA | Utilizando ese insumo se realizan los estudios y análisis con su resolución | Analizar y tomar decisiones para la solución de problemas |
| **16** | Jury voting flat vector illustration. Social satisfaction poll concept. Multicultural people holding signs with likes and dislikes. Customer feedback, clients giving marks cartoon characters  228131\_i448 | NA | Lo cual conviene a la industria con sus diferentes matices como fuga de clientes |
| **17** | Ilustración del concepto de generación de nuevos clientes potenciales  228131\_i449 | NA | Con la que las compañías buscan predecir si unos clientes se van a ir con la competencia | como la fuga de clientes con la competencia |
| **18** | Ilustración del concepto de consultoría de ventas  228131\_i450 | NA | y tomar previsiones antes de que ocurra, mediante promociones, descuentos, | Para tomar medidas de fidelización |
| **19** | Ilustración del concepto de marketing móvil  228131\_i451 | NA | premios o algún tipo de decisión que tome la directiva junto con el equipo de *marketing.* | Estrategias de marketing |
| **20** | Análisis del rendimiento empresarial con gráficos.  228131\_i452 | NA | El propósito de este componente es instruir al aprendiz a diferenciar los conceptos más relevantes en la estadística descriptiva | Importancia de la estadística descriptiva |
| **21** | Ilustración del concepto de analítica de crecimiento  228131\_i453 | NA | para la fundamentación teórica en analítica de datos | Necesaria para el análisis de datos |
| **22** | Big data en la ilustración del concepto abstracto de salud  228131\_i454 | NA | para su aplicación de manera efectiva en la solución de problemas reales | NA |
| **23** | Trabajador de oficina escalando montañas o acantilados y moviéndose al punto de destino final. concepto de desarrollo empresarial gradual, pasos sucesivos para el logro de objetivos. ilustración de vector de dibujos animados plana.  228131\_i455 | NA | El componente se desarrollará de forma gradual iniciando por las bases estadísticas, | Recursos interactivos transmedia |
| **24** | Fondo de educación online  228131\_i456 | NA | la metodología; es interactiva utilizando diferentes herramientas audiovisuales con sus respectivas explicaciones teóricas |
| **25** | Cursos y tutoriales en línea.  228131\_i457 | NA | es importante el trabajo individual de cada alumno y de este modo amplificar la curva de aprendizaje. | Trabajo autónomo |
| **26** | Ilustración del concepto de tablero  228131\_i458 | NA | En este componente se trabajará una herramienta de nombre *Plotly* | Herramientas modernas para el procesamiento, análisis y visualización de datos |
| **27** | Ilustración del concepto de estadísticas del sitio  228131\_i459 | NA | para generar visualizaciones con gráficos dinámicos o interactivos dentro de Google Colab |
| **28** | Ilustración del concepto de programación  228131\_i460 | NA | la cual permite ejecutar código en el lenguaje Python sin tener que instalarlo |
| **29** | Ilustración de sql de tecnología abstracta  228131\_i461 | NA | Se utilizarán los datos en formato csv tomados de portales de datos abiertos MEData de la alcaldía de Medellín. | Aprenderá con datos reales |
| **30** | Ilustración del concepto de datos visuales  228131\_i462 | NA | Así que en este apartad también aprenderá conceptos fundamentales como el promedio o media de un conjunto de datos | Promedio y la media |
| **31** | Concepto de estrategia de estadísticas de informe de información de datos  228131\_i463 | NA | con la finalidad de tener la fundamentación teórica suficiente para la compresión de análisis estadístico | Comprensión de análisis estadístico |
| **32** | Gráfico de barras  228131\_i464 | NA | utilizando diferentes gráficas histogramas, cajas y bigotes de forma dinámica, y de este modo presentarlo en un informe estadístico. | Histogramas, cajas y bigotes |
| **33** | Gerente priorizando tareas en la lista de tareas  228131\_i465 | NA | Se utilizarán términos como: estadística descriptiva, organización de datos en una tabla, gráficos e informe estadístico. | Informe estadístico |
| **Nombre del archivo** | **228131\_v01** | | | |

**Desarrollo de contenidos**

1. **Estadística descriptiva.**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | .  La estadística descriptiva, como la estadística inferencial, es una de las dos ramas principales de la estadística. Es la disciplina responsable de recopilar, almacenar, organizar, tabular o graficar y calcular los parámetros básicos de los conjuntos de datos. Para luego describirlos cuantitativamente. En este apartado se abordará en forma amplia la estadística descriptiva partiendo de conceptos básicos, para que logre el aprendizaje necesario que le permitirá llevar a la práctica lo aprendido en el análisis de datos. | |
| Para entrar en materia se dará la explicación de los conceptos teóricos y prácticos referente a la estadística descriptiva, Martínez, (2019).  El promedio o media  se calcula sumando todos los datos y dividiendo entre el número total de los mismos.  El promedio se simboliza con la letra X y una barra lineal  encima de esa letra: x  = promedio = media | |  |
| **Ejemplo** calcular el promedio de los siguientes valores: 3, 4, 5, 6,3 | |  |
| La desviación estándar mide qué tanto se alejan los datos del promedio, es decir, si hay mucha variación entre los datos. Si estos se alejan el valor de la desviación va a crecer, pero si se acercan va a disminuir.  La desviación se simboliza con la letra griega sigma:  La desviación es la raíz cuadrada de la varianza:  La varianza se simboliza con | |  |
| **Moda**: es el dato que más se repite.  En este caso la moda es 3 se repite 2 veces.  Me**dia geométrica:**  Es la raíz enésima del producto de los números.  En el ejemplo 3, 4, 5, 6,3 la media geométrica sería la raíz quinta debido a que son 5 de la multiplicación de los 5 elementos  =  =4.04  Se utiliza generalmente para promediar porcentajes, índices y el incremento gradual distante de un periodo a otro en alguna actividad económica. | | Ilustración del concepto de calculadora  **Imagen:** 228131\_i400 |
| **Media armónica**: es el recíproco de la media aritmética. Se calcula como el número total de observaciones dividido por la suma de los recíprocos    Se emplea cuando la variable viene expresada en unidades que son el cociente de dos magnitudes simples. Ejemplo: para el cálculo del promedio de velocidad que se define como distancia sobre tiempo. | | Media armónica:  = |
| Rango de los datos: se calcula restando el valor máximo de los datos menos el valor mínimo de los mismos.  Rango = valor máximo-valor mínimo = 5-3 =2 | | **Imagen:** 228131\_i401 |
| **Cálculo de cuartiles:**  Debajo del Cuartil 1(Q1) está el 25% de los datos, debajo del cuartil 2(Q2) es la mediana o dato central se ubica el 50% de los datos, debajo del Q3 se obtiene el 75% de los referentes.  Para el cálculo de los cuartiles primero se ordena de menor a mayor.  Cuando el número es par se utiliza la fórmula para la posición del cuartil en la secuencia ordenada: | | Donde k es el número del cuartil K=1, 2, 3  N = es el número de datos |
| En caso de que el número de datos sea impar se coloca la fórmula para la posición del cuartil en la secuencia ordenada: | |  |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | | Acordeón tipo 1 |
| **Introducción** | | En el siguiente ejercicio se aprecia mejor el cálculo de cuartiles.  Dados los siguientes números calcule los cuartiles:  cuartil 1(Q1), cuartil 2 (Q2), cuartil 3 (Q3) |
| **Imagen:** 228131\_i402 | | |
| Se ordena de menor a mayor 3, 3, 4, 5, 6 es impar N=5  La posición del  es | | |
| La posición 1.5 se obtiene promediando el primer elemento y el segundo.  El valor del  es: | | |
| La posición del  es | | |
| El valor del cuartil 2 ( es el elemento que está en la tercera posición de la secuencia  El valor del  es:  La posición del  es | | |
| El valor del  es: el promedio del elemento que está en la posición 4 con el que está en la posición 5.  5.5 | | |
| Tipo de recurso | Cajón de texto de color | |
| **Detección de datos atípicos (*outliers):***  Los valores atípicos o en inglés *outliers* se caracterizan por ser extremadamente grandes o pequeños, diferentes a los comunes de la muestra. los cuales pueden influenciar en la media y desviación de una muestra. Principalmente si el valor atípico es grande referente a los otros. | | |
| **Imagen:** 228131\_i403 | | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | **Ejemplos:**  En la siguiente serie de datos [2, 6, 8, 20, 30, 40 500] el valor atípico es 500, esta cifra puede influenciar en la media y desviación de un muestreo. |
| **Imagen:** 228131\_i404 | |
| **Botón 1** | En la siguiente serie de datos [500, 600,700, 2,5, 1000, 500] los valores atípicos son 2 y 5. |
| **Botón 2** | Para hallar el intervalo de los valores atípicos primero se calcula la diferencia entre Q3(cuartil 3) y Q1(cuartil 1), obteniendo el rango intercuartil iqr = Q3-Q1. Este rango intercuartil abarca el 50% de los datos. |
| **Botón 3** | Luego se calcula el intervalo (Q1-1.5iqr, Q3 +1.5iqr), la parte de la izquierda del intervalo, otro valor que sea inferior se considera atípico. La parte derecha del intervalo cualquier valor que esté por encima se denomina valor atípico. |

1. **Organiza una tabla de datos y los tipos de variable.**

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | Se hará mención a la organización de los datos en un esquema con las respectivas variables.  En las filas de una tabla de datos, se debe colocar observaciones o individuos y en las columnas las variables.  En la tabla 1 hay 5 observaciones iniciando en 1 y terminando en 5.  En las columnas se anotan las diferentes variables: edad, sexo, cabeza de hogar, caracterización ciudadana, estrato. |
| **Tabla 1** *Tabla de datos en las columnas las variables vs observaciones en las filas*   |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | Observación o Individuo | Edad | Sexo | Cabeza de hogar | Tipo de trabajo | Estrato | | 1 | 45 | masculino | si | independiente | 1.0 | | 2 | 45 | femenino | no | otro | 3.0 | | 3 | 28 | femenino | no | otro | 3.0 | | 4 | 30 | femenino | si | emprendedor | 3.0 | | 5 | 36 | femenino | no | independiente | 2.0 |   **Nota.**  Muestra en las filas las observaciones y en las columnas las variables. | |
| **Botón 1** | **Edad:**  Esta es una variable numérica discreta toma los valores: 45, 28, 30,36 |
| **Botón 2** | **Sexo o género:**  Es una variable categórica nominal toma los valores: masculino y femenino |
| **Botón 3** | **Cabeza de hogar:**  Es una distinción categórica nominal toma los valores: si, no |
| **Botón 4** | **Tipo de trabajo:**  Dicho ítem es una variable categórica nominal toma los valores: independiente, otro, emprendedor. |
| **Botón 5** | **Estrato:**  En este caso es una variable numérica discreta toma los valores: 1.0, 2.0, 3.0 |
| **Botón 6** | **Variables:**  Las cuantitativas son continuas y discretas.  Ejemplo: el peso, la edad tomando los meses, días  Las cualitativas: categóricas ordinales y categóricas nominales |
| **Botón 7** | **Categórica nominal:**  representan estados sin clasificación intrínseca, ejemplo: marcas de zapatos, género, raza, etnia. |
| **Botón 8** | **Categórico ordinal:**  Hace referencia a calificar valores que representan categorías con datos intrínsecos, tiene características de orden, ejemplo: malo, bueno, excelente. |

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| Esos datos deben estar colocados en una tabla organizada  Las tablas deben estar numeradas y explicar brevemente que contienen. Las tablas deben llevar un nombre generalmente en la parte superior.  Las tablas se pueden numerar con números romanos: I, II, III, IV, V, VI, VII, etc.  Las tablas se pueden numerar con números arábigos: 1, 2, 3,4, 5, 6, 7, 8, etc.  La frecuencia absoluta es las veces que se repite el dato.  Dentro de la tabla se debe especificar la variable o dato a estudiar o analizar  Lind et al, (2019) | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | Se emplearon referentes de créditos de consumo otorgados en Medellín y obtenidos de la página de datos abiertos de la alcaldía de la capital de Antioquía en formato csv, Metada alcaldía de Medellín, (2022).  Se construirá una tabla de distribución de frecuencias por año: |
| **Tabla 2** *Año Vs. Frecuencia absoluta de beneficiarios del crédito de consumo.*   |  |  | | --- | --- | | Año | Total, beneficiarios por año  (frecuencia absoluta) | | 2017 | 1921 | | 2018 | 8765 | | 2019 | 1466 | | Total | 12152 |   Nota. Muestra año Vs. Número de beneficiarios del crédito de consumo. | |
| **Botón 1** | **Frecuencia absoluta acumulada**  Es cuando se van sumando las frecuencias por cada año de forma acumulada. |
| **Botón 2** | **Frecuencia absoluta relativa**  Es el resultado de dividir la frecuencia absoluta acumulada entre el total de las repeticiones en este caso 12152. |
| **Botón 3** | **Frecuencia absoluta relativa porcentual**  Se calcula la frecuencia absoluta relativa porcentual: multiplicando por 100 la absoluta relativa. |

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| **Tabla 3** *año vs frecuencia absoluta relativa de beneficiarios del crédito de consumo*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | Año | Total beneficiarios por año (frecuencia absoluta) | Frecuencia absoluta acumulada | Frecuencia absoluta relativa | Frecuencia absoluta relativa porcentual (%) | | 2017 | 1921 | 1921 | 1921/12152 = 0.158 | 15.80% | | 2018 | 8765 | 10686 | 10686/12152=0.8793 | 87.93% | | 2019 | 1466 | 12152 | 12152/12152 = 1 | 100% | | Total | 12152 |  |  |  |   Nota. Muestra año vs frecuencia absoluta relativa y porcentual de beneficiarios del crédito de consumo. | |
| La tabla de datos esta ordenada mediante filas y columnas, las primeras corresponden a la información en forma horizontal, cada uno de los individuos. Las columnas en forma vertical haciendo referencia a cada una de las variables de las reseñas consignadas. La idea fundamental es hacer un tabloide y comparar de forma fácil las diversas características que tienen cada una de las variables de estudio en el problema a resolver. Los datos pueden ser cualitativos o cuantitativos. | |

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| Intervalos de clase: son aquellos en los que se agrupan y ordenan los valores observados. Una variable de la tabla se puede agrupar por intervalo y para ello, se debe tomar los valores máximo y mínimo y dividirlo en cuantas clases o intervalos de clase se va a dividir dicho intervalo. Estos siempre serán cerrados por la izquierda y abiertos por la derecha, excepto el ultimo. La frecuencia o número de casos que cae en ese estaría en una columna al lado llamada frecuencia absoluta o simplemente frecuencia. | |
| **Ejemplo**  **Tabla 4** *Muestra intervalos vs frecuencia absoluta.*   |  |  | | --- | --- | | **Intervalo Vs Frecuencia** | | | Intervalo | Frecuencia | | [0,2) | 2 | | [2,4) | 3 | | [4,6) | 4 | | [6,8) | 5 | | [8,10) | 3 |   *Nota.*: muestra intervalos vs frecuencia absoluta. | |

**Otro ejemplo**

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Se tienen las siguientes 12 notas calificadas del número 1 al 10 en una evaluación de matemáticas correspondiente a la primera evaluación, siendo 1 la menor y 10 la mayor. Se quieren agrupar por intervalos mediante 5 clases. Las 12 notas son:  2,3,5,6,7,8,9,3,7,9,10,9  Este conjunto de datos se quiere agrupar en forma de intervalo.  Primero se extraerá el valor menor y el mayor. En este caso la estimación mínima es 2 y la máxima es 10, se calcula el tamaño del intervalo, posteriormente se mira cuantos datos caen en ese intervalo, generalmente no se trabajan con más de 10, se puede buscar que la longitud o amplitud del intervalo sea un numero entero, aunque no es estricto, podría ser decimal |
| **Imagen:** 228131\_i405 | |
| **Amplitud**  Para calcular la amplitud o longitud del intervalo se utiliza la siguiente fórmula:  Amplitud o longitud del intervalo = (valor máximo-valor mínimo) /número de clases  Suponiendo que se quieren 4 clases por lo tanto el número de clases = 4  Valor máximo es 10, por lo tanto, el máximo = 10  Valor mínimo es 2, por lo tanto, mínimo = 2  La diferencia entre el valor máximo y mínimo es 8, por ende, valor máximo-valor mínimo = 8  Para hallar la amplitud o longitud del intervalo el valor anterior lo divididos entre el número de clases que escogimos, en este caso 4    Por lo tanto, seria 8 dividido en 4 resultando 2, entonces, la amplitud o longitud del intervalo = 2  Se inicia con el valor mínimo y el intervalo cerrado a la izquierda y abierto a la derecha con una longitud de 2, es decir, [2,4) (ver tabla 5) | |
| **Datos que caben en el intervalo**  Se observa cuantas notas caen en cada uno de los intervalos ordenados de forma creciente y se coloca el número en la columna de la frecuencia**.**  La suma de las frecuencias debe ser igual al número total de notas en este caso 12  **Tabla 5** *Intervalo Vs Frecuencia notas*   |  |  | | --- | --- | | **Intervalo vs frecuencia notas** | | | Intervalo | frecuencia | | [2,4) | 3 | | [4,6) | 1 | | [6,8) | 3 | | [8,10] | 5 | | Total | 12 |   Nota. Muestra intervalos vs frecuencia absoluta para las 12 notas de matemáticas.  Se observa que el primer intervalo [2,4) tiene una frecuencia de 3 que son los datos: 2, 3,3  Se evidencia que el segundo intervalo [4,6) tiene una frecuencia de 1 que es el dato: 5  Para el tercer intervalo se aprecia que [6,8) tiene una frecuencia de 3 que son los datos: 6, 7, 7  Se examina que el cuarto intervalo [8,10] tiene una frecuencia de 5 que son los datos: 8, 9, 9 ,9, 10 | |
| **Promedio**  Si se tienen los datos agrupados como en la tabla 5. Ahora se pide calcular el promedio, tomando un punto medio de cada intervalo, se llama marca de clase, luego se debe multiplicar por la frecuencia y dividirlo por el número total de la frecuencia.  En el primer intervalo el punto medio o marca de clase es 3  En el segundo intervalo el punto medio o marca de clase es 5  En el tercer intervalo el punto medio o marca de clase es 7  En el cuarto intervalo el punto medio o marca de clase es 9  Promedio = (suma (dato\*frecuencia) / (total frecuencia))  Promedio = (3\*3+5\*1+7\*3+9\*5) / 12 = (3+5+21+45)/12 = 74/12 = 6.166  El valor anterior es una aproximación del promedio.  El verdadero valor del promedio se puede calcular si no se tienen los datos agrupados por intervalos. En este caso se tiene los valores exactos, por lo tanto, se podría calcular.  Promedio = suma de datos / número de datos  Promedio = (2+3+5+6+7+8+9+3+7+9+10+9)/12 = 78/12 = 6.5 | |

1. **Gráficos.**

|  |  |
| --- | --- |
| Tipo de recurso | Cajón de texto de color |
| La parte visual es importante en cualquier tipo de análisis o estudio mediante un gráfico se facilita la comprensión de los datos y la relevancia en la investigación planteada. Existen representaciones animadas dinámicas y estáticas en esta sección se trabajará con dinámicos. Los dinámicos con la librería plotly, plotly, (2022). | |

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | Se abordarán diferentes gráficos estadísticos:  Para estas se emplearon referentes de créditos de consumo otorgados en Medellín y obtenidos de la página de datos abiertos de la alcaldía de la capital de Antioquia en formato csv, Metada alcaldía de Medellín, (2022). |
| **Imagen:** 228131\_i405 | |
| **Gráfico de histograma:**  Se representa en forma horizontal o vertical, también a manera de barras muestra el dato y las veces que se repite o su frecuencia  **Figura 1** *Muestra un histograma cantidad de hombres vs. cantidad de mujeres.*    **Imagen:** 228131\_i406  **! INSERTAR GRÁFICO DINÁMICO - histograma\_hombre\_vrs\_mujeres.html**  **Se Encuentra en la carpeta CF04\_228131 - GRAFICOS DINAMICOS HMTL** | |
| **Gráfico circular:**  muestra en forma de círculo generalmente en porcentaje, indica la proporción que aparece la cifra respecto al total  **Figura 2** *Muestra un gráfico Circular género Vs. monto del préstamo.*    **Imagen:** 228131\_i407  **Figura 3** *Muestra un gráfico Circular el porcentaje por estratos los cuales reciben el monto del préstamo.*    **Imagen:** 228131\_i408  **! INSERTAR GRÁFICOS DINÁMICOS 2 Y 3**  **Se Encuentra en la carpeta CF04\_228131 - GRÁFICOS DINÁMICOS HMTL** | |
| **Gráfico de cajas y bigotes:**  Estos en forma de caja o rectángulo se emplean para apreciar los cuartiles y el valor mínimo a la izquierda, valor máximo a la derecha con líneas verticales.  Q1: cuartil 1, Q2: cuartil 2(mediana), Q3: cuartil 3, valor mínimo, valor máximo  Debajo del Cuartil 1(Q1) está el 25% de los datos, debajo del cuartil 2(Q2) es la mediana o dato central se ubica el 50% de los datos, debajo del Q3 se obtiene el 75% de los datos.  En el rango intercuartil = Q3-Q1 está el 50% de los datos  Por debajo del valor mínimo y por encima del máximo están los valores atípicos o en inglés *outliers*  **Figura 4** *Muestra un gráfico de cajas y bigotes para la edad de las personas que reciben préstamo*    **Imagen:** 228131\_i409  **! INSERTAR GRÁFICOS DINÁMICOS 4 Y 5**  **INSERTAR AQUÍ INFOGRAFÍA INTERACTIVA # 1 – PUNTO CALIENTE**    **Imagen:** 228131\_i410  **INSERTAR AQUÍ INFOGRAFÍA INTERACTIVA # 2 – PUNTO CALIENTE** | |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva 1 - Punto caliente | |
| **Texto introductorio** | Gráfico de cajas y bigotes para la edad de las personas que reciben un préstamo. | |
| **Imagen**    **Imagen:** 228131\_i411 | | |
| **Código de la imagen** | 228131\_i409 | |
| **Punto caliente 1** | Muestra el número 18. La edad mínima de las personas que reciben préstamo. | P1 en amarillos |
| **Punto caliente 2** | Cuartil 1 (Q1 = 31) debajo de ese valor está el 25% de la edad de las personas que reciben préstamo. | P2 en amarillo |
| **Punto caliente 3** | Cuartil 2 (Q2 = 45). La mediana debajo de ese valor está el 50% de la edad de las personas que reciben préstamo. | P3 en amarillo |
| **Punto caliente 4** | Cuartil 3 (Q3 = 57). Debajo de ese valor está el 75% de edad de las personas que reciben préstamo. | P4 en amarillo |
| **Punto caliente 5** | Muestra el número 84. La edad máxima de las personas que reciben préstamo. | P5 en amarillo |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Infografía interactiva 2 Punto caliente | |
| **Texto introductorio** | gráfico de cajas y bigotes estrato vs edad de las personas que reciben crédito. | |
| **Imagen:** 228131\_i412 | | |
| **Código de la imagen** | 228131\_i410 | |
| **Punto modal 1** | **Estrato 0**, 50% tiene la Edad entre 32 y 50.75 años. | Punto 1 en amarillo |
| **Punto modal 2** | **Estrato 1**, 50% tiene la Edad entre 27 y 51 años. | Punto 2 en amarillo |
| **Punto modal 3** | **Estrato 2**, 50% tiene la Edad entre 30 y 54 años. | Punto 3 en amarillo |
| **Punto modal 4** | **Estrato 3**, 50% tiene la Edad entre 35 y 63 años. | Punto 4 en amarillo |
| **Punto modal 5** | **Estrato 4**, 50% tiene la Edad entre 28 y 54 años. | Punto 5 en amarillo |
| **Punto modal 6** | **Estrato 5**, 50% tiene la Edad entre 27 y 46.5 años. | Punto 6 en amarillo |
| **Punto modal 7** | **Estrato 6**, \*Solo 2 personas Edad 24 y 49 años respectivamente. | Punto 7 en amarillo |

1. **Elaboración del informe estadístico.**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | | Pestañas o tabs Verticales |
| **Introducción** | | Es la forma de mostrar datos, gráficos, tablas, análisis, resultados y conclusiones el estudio o investigación realizada.  Este documento debe contener planteamiento del problema, objetivo general, objetivos específicos, metodología, análisis, resultados y por último conclusiones. |
| **Imagen:** 228131\_i413 | | |
| **Planteamiento del problema** | Se explica lo que se quiere resolver o investigar | |
| **Objetivo general** | Se centra en solucionar un problema o investigación de estudio | |
| **Objetivos específicos** | Son los pasos para obtener el objetivo general. | |
| **Metodología** | Se debe aclarar si es un informe estadístico descriptivo o inferencial, la población de estudio, el tamaño de la muestra. El tiempo de estudio y Software utilizado. | |
| **Análisis y resultados** | Presentación de gráficos, tablas, análisis con los respectivos hallazgos. Las tablas y gráficos deben estar enumerados y explicados. Señalados en el documento, ejemplo: ver gráfico 2 | |
| **Conclusiones** | Explicar de forma clara empleando los porcentajes y/o totales, además de hallazgos resultado de todo el análisis estadístico a la población objeto. | |

**Ejemplo de Informe estadístico:**

|  |  |
| --- | --- |
| **Tipo de recurso** | Acordeón tipo 1 |
| **Introducción** | En el recurso anterior se explicaron los elementos a tener presentes en un informe estadístico. seguidamente, se plantea un ejemplo donde se interpreta el paso a paso de un informe estadístico que servirá de guía a la hora de realizarlo. |
| **Imagen:** 228131\_i414 | |
| **Informe estadístico:**  Comparar los créditos de consumo otorgado en Medellín por la alcaldía desde el año 2017 hasta 2019 | |
| **Objetivo general:**  Cotejar los créditos de consumo otorgado en Medellín por la alcaldía desde año 2017 hasta 2019 | |
| **Objetivos específicos:**  Analizar los créditos de consumo otorgado en Medellín por la alcaldía desde año 2017 hasta 2019  Examinar los créditos de consumo otorgado en Medellín por la alcaldía desde año 2017 hasta 2019 por estrato, edad y género. | |
| **Metodología:**  La población estudio son personas que obtuvieron créditos de consumo otorgados en Medellín por la alcaldía desde año 2017 hasta 2019 (En este caso análisis de tipo estadístico descriptivo), utilizando medidas de tendencia central como: la media, moda y desviación. Gráficos Estadísticos: de cajas y bigotes, histograma, apilados. estáticos utilizando las librerías, *Matplotlib,*(2022), *Seaborn*,(2022), *Altair*,(2022), empleando Google Colab con el lenguaje *Python* y manejo de tablas de datos. | |
| **Resultados y análisis:**  Un total de 12152 personas recibieron créditos, de los cuales 8069 fueron mujeres y 4083 hombres. De estos corresponden a67.2% para el género femenino y 32.8% para el masculino. Monto prestado valor mínimo $50.000, máximo $450.000. Los estratos van del 0 al 6. Solo 2 personas en el estrato 6 recibieron crédito.  [Insertar aquí Rutas / Pasos. Verticales 1](#bookmark=id.1fob9te) | |
| **Conclusiones:**  Las edades de las personas oscilan entre 18 y 84 años.  Monto prestado valor mínimo $50.000 máximo $450.000  12152 personas recibieron créditos de consumo otorgado en Medellín por la alcaldía.  67.2% fueron para el género femenino y 32.8% para el masculino.  8069 eran mujeres y 4083 hombres.  8117 un 66.7% se beneficiaron con un crédito de $ 200.000.0  El 2018 fue el año donde se otorgaron más créditos, un total de 8765.  Los estratos de los beneficiarios van del 0 al 6 en este último solo 2 personas recibieron el crédito con edad de 24 años y 49 años respectivamente. | |

|  |
| --- |
| Cuadro de texto |
| Para afianzar los conocimientos en gráficos estáticos y dinámicos. Utilizando librerías como *Plotly, matplotlib*, *Seaborn*, Altair. Usando Google Colab.  Por favor observar el siguiente video donde se trabaja con datos reales. Sobre créditos de consumo otorgados en Medellín. Obtenidos de la página de datos abiertos de la Capital de Antioquía.  **VER ANEXO: 228131\_v02 - gráficos**  228131\_i432  SE ENCUENTRA EN LA CARPETA CF04\_228131 - VIDEO GRÁFICOS |

|  |
| --- |
| Cuadro de texto |
| Para comprender fácilmente el video, específicamente el manejo de Google Colab y el documento j*upyter*, *Notebook* con extensión ipynb, se recomienda descargar los 3 archivos adjuntos.   |  |  |  |  | | --- | --- | --- | --- | | **Tema** | **Tipo de archivo** | **Nombre de archivo** | **Enlace Web** | | Manejo de Google Colab | .PDF | Manejo de Google Colab.pdf |  | | Créditos de consumo | .CSV | creditos\_de\_consumo\_otorgados.csv |  | | Gráficos ploty | .IPYNB | gráficos\_Plotly\_video.ipynb |  |   SE ENCUENTRAN EN LA CARPETA CF04\_228131-Descargar Material |

[Viene del recurso anterior **resultados y análisis:** Rutas / Pasos. Verticales 1](#bookmark=id.30j0zll)

|  |  |
| --- | --- |
| **Tipo de recurso** | Rutas / Pasos. Verticales 1 |
| **Introducción** | A continuación, se muestra el análisis de ejemplo de las tablas y gráficos. |
| Imagen general que ilustre el tema    **Imagen:** 228131\_i415 | |
| **Botón 1** | Información sobre los años y número de beneficiarios  **Tabla 6** *Años vs beneficiarios del crédito*   |  |  | | --- | --- | | Años de los beneficiarios del crédito | números de beneficiarios del crédito | | 2017 | 1921 | | 2018 | 8765 | | 2019 | 1466 |   Nota: muestra los años y el número de beneficiarios del crédito de consumo. |
| **Botón 2** | **Tabla 7** *Estrato y el rango de edades que tienen el 50% de los créditos*   |  |  | | --- | --- | | Estrato | El 50% de las edades están entre: | | 0 | 32 y 50.75 años | | 1 | 27 y 51 años | | 2 | 30 y 54 años | | 3 | 35 y 63 años | | 4 | 28 y 54 años | | 5 | 27 y 46.5 años | | 6 | \*Solo 2 personas 24 y 49 años |   Nota: Indica el estrato y el intervalo de edades correspondiente al 50% de las personas que recibieron crédito de consumo. Solo 2 personas recibieron crédito en el estrato 6\* |
| **Botón 3** | **Tabla 8** *monto prestado y total de beneficiarios*   |  |  | | --- | --- | | Monto prestado | Beneficiarios en forma descendente | | $ 200.000.0 | 8117 | | $ 300.000.0 | 2115 | | $ 200.012.0 | 1549 | | $ 450.000.0 | 184 | | $ 400.000.0 | 136 | | $ 250.000.0 | 16 | | $ 100.000.0 | 10 | | $ 150.000.0 | 8 | | $ 50.000.0 | 7 | | $ 350.000.0 | 6 | | $ 199.975.0 | 2 | | $ 199.606.0 | 1 | | TOTAL | 12152 |   *Nota:* representa el monto prestado y el total de beneficiarios en forma descendente. |
| **Botón 4** | La distribución de las edades se observar en el histograma de frecuencias (**ver gráfico 6)**  **Figura 5** *Histograma de frecuencias*    *Nota:* muestra el histograma de frecuencia para la edad. Realizado con *matplotlib*.  **Imagen:** 228131\_i416 |
| **Botón 5** | La relación entre estrato y edad se puede observar en el gráfico de cajas y bigotes  **Figura 6** *Cajas y bigotes*    *Nota:* expone el gráfico de cajas y bigotes del estrato vs la edad. Realizado con *matplotlib.*  **Imagen:** 228131\_i417 |
| **Botón 6** | **Figura 7** *La relación entre el estrato, monto prestado y edad*    *Nota:* muestra el estrato, monto prestado y la barra de colores referente a la edad. Realizado con librería *altair****.***  **Imagen:** 228131\_i418 |
| **Botón 7** | El diagrama o gráfico de dispersión sirve para saber el comportamiento de dos variables entre sí, en este caso la edad y el monto prestado**.**  **Figura 8** *Diagrama de dispersión edad vs monto prestado*    *Nota:* diagrama de dispersión edad vs monto prestado.  **Imagen:** 228131\_i419  !ATENCIÓN! INSERTAR GRÁFICO 9 diagrama\_de\_dispersion\_edad\_vr\_monto\_prestado.html  **Se Encuentra en la carpeta CF04\_228131 - GRÁFICOS DINÁMICOS HMTL** |

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Slider Presentación | |
| **Introducción** | En Google *Colab* se tienen máquinas virtuales con tarjetas gráficas. Permite ejecutar el código en el lenguaje *Python* dentro de la nube con el formato de *jupyter* *notebook* extensión *ipynb*.  A continuación, se presenta el paso a paso que explica el funcionamiento y configuración de la herramienta. | |
| *jupyter* *notebook* se puede utilizar localmente o en la nube con Google Colab. Para usarlo localmente debe tenerlo instalado en su computador.  Dentro de Google *Colab* se está en un entorno interactivo denominado notebook de *Colab,* que permite escribir y ejecutar código celda por celda o todo de una vez, se obtiene en la pestaña código.  Escribe en un lenguaje etiquetado especial llamado *Markdown*, se obtiene en la pestaña de nombre texto. | | **Figura 9** *Pantalla de inicio de Google Colab.*  Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico  Descripción generada automáticamente  **Imagen:** 228131\_i432 |
| **Instalar una librería o paquete**  Ejemplo:  pip install pandas  Instalaría la librería de nombre pandas si ya está instalada lo dirá.  **Importar librería con un alias(as)**  Ejemplo  import pandas as pd  cargar la librería pandas con el alias o abreviación pd  Ejemplo:  import math  Cargar la librería de nombre *math* referente a matemática.  print(dir(math))  Muestra las funciones que están dentro de la librería *math*  Pow(2,3) una de las funciones que están en la librería *math*, es la función potencia en ingles Power, pow(2,3) significa que eleva el número 2 a la 3 da como resultado 8. | | **Figura 10**  *Importar la librería math y utiliza la función potencia: pow()*  Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico  Descripción generada automáticamente  **Imagen:** 228131\_i433 |
| ***DataFrame***  Es una estructura de datos con dos dimensiones (filas y columnas), se puede guardar referencias de distintos tipos (caracteres, enteros, valores de punto flotante, etc.) en columnas. Es similar a una hoja de cálculo o una tabla de SQL.  Un archivo CSV (valores separados por comas) es un documento de texto con un formato específico que permite guardar los datos en una configuración de tabla estructurada.  Un archivo CSV puede tener diferentes separadores diferentes de la coma (,) como:  Punto y coma (;), barra espaciadora (/), guión del medio (-), guión de piso(\_), etc. | | **Figura 11** *Muestra la lectura y contenido del data frame utilizando el alias pd.read\_csv()*  Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico  Descripción generada automáticamente  **Imagen:** 228131\_i434 |
| **Lectura de archivos CSV**  Para leer en formato CSV se utiliza el comando pd.read\_csv('nombre del archivo.csv'), pd es la abreviación de la librería pandas.  En la parte nombre del archivo se debe escribir la ruta o dirección donde está el archivo al final de esa ruta el nombre con extensión.csv  Se puede utilizar el comando pwd para saber la ruta donde se está ubicado.  Para cambiar el directorio o ruta se utiliza el comando cd  Ejemplo: cd /content/sample\_data/  Ejemplo: leer un archivo csv, utilizando el alias de pandas pd, nombrar el data\_frame y visualizarlo con el nombre que se le coloco en este caso data\_frame  import pandas as pd  data\_frame= pd.read\_csv('california\_housing\_test.csv')  data\_frame | | Se muestra la misma imagen de la diapositiva anterior |
| **Obtener determinado número de filas**  Para obtener las primeras 5 filas de la data frame se escribe:  data\_frame.head(5)  Para obtener las últimas 5 filas del data frame  data\_frame.tail() | | **Figura 12** *Contenido de las primeras y últimas 5 filas del data frame.*  Tabla  Descripción generada automáticamente  *Nota:* muestra el contenido de las primeras y últimas 5 filas del data frame.  **Imagen:** 228131\_i435 |
| **Visualizar dimensión**  Para ver la dimensión (largo y ancho), filas y columnas también el nombre de las columnas, en este caso 3000 filas 9 columnas se utilizan los comandos:  print("Dimensión: ", data\_frame.shape)  print("Columnas: ", data\_frame.columns)  Shape: (3000, 9) | | **Figura 13**  *Muestra la dimensión y el nombre de las columnas del data frame*  Interfaz de usuario gráfica, Texto, Correo electrónico  Descripción generada automáticamente  *Nota:* muestra la dimensión y el nombre de las columnas del data frame  **Imagen:** 228131\_i436 |
| Información sobre los estadísticos descriptivos de la columna con nombre 'total\_rooms', se escribe el nombre entre comillas sencillas o dobles.  data\_frame['total\_rooms'].describe()  También se puede: data\_frame["total\_rooms"].describe()  Redondear la salida de los valores a 2 decimales. | | **Figura 14** *Información sobre los estadísticos descriptivos de la columna con nombre 'total\_rooms'.*  Interfaz de usuario gráfica, Texto, Aplicación  Descripción generada automáticamente  *Nota:* Información sobre los estadísticos descriptivos de la columna con nombre *'total\_rooms'*.  **Imagen:** 228131\_i437 |

**Síntesis**

|  |  |
| --- | --- |
| **Tipo de recurso** | Síntesis |
| Nombre del mapa: estadística descriptiva, tablas, gráficos e informe estadístico. | |
| **Introducción** |  |
| **Figura 15**  *Síntesis del componente formativo*    Imagen: 228131\_i420  Nota. Muestra la síntesis del componente 4 realizado con CmapTool ver 6.04.  Enlace Web: <https://cmapscloud.ihmc.us/viewer/cmap/1YFM4NMZG-1S976TR-KR> | |

**Actividad interactiva**

|  |  |  |
| --- | --- | --- |
| **Tipo de recurso** | Actividad didáctica. Verdadero y falso | |
| Durante el transcurso de este componente formativo se afianzó el concepto de estadística descriptiva, disciplina responsable de recopilar, almacenar, organizar, tabular o graficar y calcular los parámetros básicos de los conjuntos de datos. Por ello, el propósito de la siguiente actividad es poner a prueba el nivel de conocimientos y aprendizajes obtenidos. | | Colegas analizando tablas de crecimiento  Imagen: 228131\_i421 |
| 1. El cuartil 2 (Q2) es el promedio: | | Calcular primer cuartil | CÃ³mo se calcula el Q1 | Calculadora EstadÃ­stica  Imagen: 228131\_i422 |
| Verdadero | | **Falso (correcto)** |
| Retroalimentación: Los cuartiles dividen los datos en 4 partes iguales. El cuartil 2 (Q2) es la mediana o dato central. Debajo del el cuartil 2 (Q2) está el 50% de los datos. | | |
| 1. Debajo del cuartil 1(Q1) está el 20% del dato | | Calcular primer cuartil | CÃ³mo se calcula el Q1 | Calculadora EstadÃ­stica  Imagen: 228131\_i423 |
| Verdadero | | **Falso (correcto)** |
| Retroalimentación: Los cuartiles dividen los datos en 4 partes iguales. Debajo del cuartil 1(Q1) está el 25% de los datos experimento: | | |
| 1. Debajo del cuartil 1(Q3) está el 70% de los datos: | | Calcular tercer cuartil | CÃ³mo se calcula el Q3 | EstadÃ­stica  Imagen: 228131\_i424 |
| **Verdadero** | | **Falso (correcto)** |
| Retroalimentación: Los cuartiles dividen los datos en 4 partes iguales. Debajo del cuartil 1(Q3) está el 75% de los datos | | |
| 1. Las tablas se enumeran sólo con números arábigos:1,2,3,4,5 | | plantilla de infografÃ­a moderna, opciones, pasos, filas, columna, grÃ¡fico / fondo de tabla infogrÃ¡fica  Imagen: 228131\_i425 |
| Verdadero | | **Falso (correcto)** |
| Retroalimentación: Las tablas se pueden numerar con números arábigos:1,2,3,4,5, … y con números romanos: I, II, III, IV, V, VI, VII, … | | |
| 1. En el gráfico de cajas y bigotes se obtienen sólo los cuartiles. | | Diagrama de grÃ¡fico de cajas y bigotes  Imagen: 228131\_i426 |
| Verdadero | | **Falso (correcto)** |
| Retroalimentación: en el gráfico de cajas y bigotes se obtienen Q1: cuartil 1, Q2: cuartil 2(mediana), Q3: cuartil 3, valor mínimo, valor máximo | | |
| 1. El rango intercuartil se obtiene restando Q3 y Q1 | | Diagrama de grÃ¡fico de cajas y bigotes  Imagen: 228131\_i427 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación: El rango intercuartil = Q3-Q1 está el 50% de los datos | | |
| 1. Un informe estadístico debe tener metodología | | Mujer empresaria que trabaja en el informe financiero de operaciones empresariales. Hoja de cÃ¡lculo del informe contable.  Imagen: 228131\_i428 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación: Uno de los elementos que compone un informe estadístico es la metodología. En ella se debe explicar si es un documento estadístico descriptivo o inferencial, la población de estudio, el tamaño de la muestra. El tiempo de estudio y software utilizado. | | |
| 1. Los gráficos dinámicos se pueden hacer con la librería *Plotly* | | Barra de estadÃ­sticas, dona y grÃ¡fico de lÃ­neas  Imagen: 228131\_i429 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación: *Plotly* es una librería que permite hacer gráficos dinámicos o interactivos. Utilizando el lenguaje *Python* | | |
| 1. La librería *matplotlib* es para hacer gráficos estáticos | | ColecciÃ³n de grÃ¡ficas de barras infogrÃ¡ficas  Imagen: 228131\_i430 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación: gráficos estáticos se pueden hacer utilizando las librerías *matplotlib, seaborn, altair* | | |
| 1. En Google Colab se puede trabajar el lenguaje Python | | Como empezar a analizar datos con Python usando Google Colab | by Gustavo  Juantorena | Medium  Imagen: 228131\_i431 |
| **Verdadero (correcto)** | | Falso |
| Retroalimentación: en Google Colab se puede escribir y ejecutar celda por celda código en el lenguaje Python. | | |

**Material complementario**

|  |  |  |  |
| --- | --- | --- | --- |
| Tipo de recurso | Material complementario | | |
| Tema | Referencia APA del material | Tipo | Enlace |
| Estadística descriptiva. | Aguilar, J. (2021), *Estadística Descriptiva regresión y probabilidades con aplicaciones* (1ma ed), Bogotá, Ediciones de la U. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=15156&pg=202> |
| Python practico | Moreno, A., & Córcoles, S. (2020),*Python práctico* (1ma ed) Bogotá: Ediciones de la U. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=10285&pg=1> |
| Jupyter for Data Science | TOOMEY, D. (2017) *Your One-stop Guide to Building an Efficient Data Science Pipeline Using Jupyter. Birmingham: Packt Publishing.* |  | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1637911&lang=es&site=ehost-live> |
| Learning Jupyter 5 | TOOMEY, D. (2018). *Explore Interactive Computing Using Python. 2nd* Edition. |  | <https://search-ebscohost-com.bdigital.sena.edu.co/login.aspx?direct=true&db=e000xww&AN=1881491&lang=es&site=ehost-live> |
| Estadística | Triola, M. (2018). *Estadística* (Vol. 12 edición). Mexico: Pearson. | Libro digital | <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=7386&pg=149> |

**GLOSARIO**

|  |  |
| --- | --- |
| **Tipo de recurso** | Glosario |
| Censo | Un conjunto de actividades destinadas a recoger, recopilar, el padrón o lista de la población de una nación en un momento determinado, la información demográfica social y económica, para luego evaluar, analizar y publicar o divulgar. |
| Continua | Es aquella que puede tomar un número infinito de valores entre dos valores distintos de una característica. La altura de las 5 personas: 1.85, 1.89, 1.93, 1.59, 1.65. En la práctica se mide la altura con dos decimales, pero también se puede dar con tres decimales o más. |
| Cualitativa | Son aquellas en la que los resultados posibles no son valores numéricos. Por ejemplo: color del pelo, tipo de ropa preferida, lugar de veraneo, etc. |
| Cuantitativa | Se trata de características cuyas variaciones pueden ser expresadas de forma numérica. Por ejemplo: edad, peso, estatura, número de hijos, años de servicio, etc. |
| Cuartiles | Son 4 dividen un conjunto de datos ordenados en cuatro partes iguales. Cuartil 1 (Q1) 25%, la mediana es el Cuartil 2 (Q2) 50%,Cuartil 3 (Q3) 75% Cuartil 4 (Q4) 100%.Se trabaja principalmente con los Q1,Q2,Q3 |
| Dato atípico | Es un valor mayor a 1.5 veces el rango intercuartil. |
| Desviación estándar | Mide la dispersión de una distribución de datos. Entre más dispersa está una distribución de referencias, más grande es su desviación estándar. |
| Discreta | Es cuando no se puede tomar ningún valor entre dos consecutivos, y que es continua al momento de tomar cualquier valor dentro de un intervalo. Ejemplos de variable discreta: número de empleados de una fábrica; número de hijos; número de cuentas en un banco. |
| Escala cronológica | Hace referencia a los sucesos, es la ubicación de éstos en una línea temporal. |
| Escala nominal | Es la gama de medición en la cual los números sirven como “etiquetas” solamente para identificar o clasificar un objeto. normalmente trata sólo con variables no numéricas (no cuantitativas). |
| Escala ordinal | Es la que permite evaluar la actitud de un encuestado con respecto a un tema utilizando un conjunto de respuestas ordenadas. Por ejemplo: “Muy satisfecho”, “Satisfecho”, “Insatisfecho” y “Muy insatisfecho”. |
| Estadísticos | Hace referencia a cualquier función real medible de la muestra en una variable aleatoria. |
| Media | Resultado que se obtiene al dividir la suma de varias cantidades por el número de sumandos. |
| Muestreo | Selección de un conjunto de personas o cosas que se consideran representativas del grupo al que pertenecen, con la finalidad de estudiar o determinar las características del grupo. |
| Rango intercuartil | Es una medida de dispersión. Se obtiene restando el cuartil 3 (Q3) menos el cuartil 1(Q1); Q3-Q1 en ese rango está el 50% de los datos. |

**Referentes bibliográficos**

|  |  |
| --- | --- |
| **Tipo de recurso** | Bibliografía |
| Lind, D., & Marchal, W. (2019), *Estadística Aplicada a los negocios y la economía.* <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9246&pg=315>. | |
| Metada alcaldía de Medellín, (2022), *Créditos de consumo otorgados*. <https://acortar.link/zGU2gK> | |
| Martínez, C. (2019), *Estadística y muestreo*, <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=9278&pg=387>.  No puedo comprobar la fuente me pide usuario y contraseña, página del SENA | |
| Martínez, C. (2019), *Estadística básica aplicada*. <https://www-ebooks7-24-com.bdigital.sena.edu.co/?il=8975&pg=82> | |
| Matplotlib. (2022), *Visualización con Python*. <https://matplotlib.org/> | |
| Plotly(2022), *Plotly Biblioteca de gráficos de código abierto para Python*. <https://plotly.com/python/> | |
| Seaborn. (2022). visualización de datos estadísticos. <https://seaborn.pydata.org/> | |
| Vega, A (2022). *Visualización declarativa en Python*. <https://altair-viz.github.io/> | |
| Vladimirovna, O. & Gutiérrez, E. (2016). *Estadística inferencial 1 para ingeniería y ciencias*.  <https://elibro-net.bdigital.sena.edu.co/es/ereader/senavirtual/40474?page=69> | |