A

**Agenda**: Lista de citas programadas

**Agregación de datos:** Proceso de recolectar datos de múltiples fuentes y combinarlos en una sola colección resumida

**Agregación:** Proceso de recolectar o juntar muchas partes separadas en un todo

**Alcance del trabajo (SOW):** Esquema acordado de las tareas a realizar durante un proyecto

**Algoritmo:** Proceso o conjunto de reglas a seguir para realizar una tarea específica

**Análisis computacional de datos:** La ciencia de los datos

**Análisis de datos:** Recopilación, transformación y organización de los datos para sacar conclusiones, hacer predicciones e impulsar una toma de decisiones fundamentada

**Análisis de déficits:** Método para examinar y evaluar el estado actual de un proceso con el fin de identificar las oportunidades de mejora en el futuro

**Análisis:** Proceso usado para dar sentido a los datos recopilados

**Analista de datos:** Persona que recopila, transforma y organiza los datos para sacar conclusiones, hacer predicciones e impulsar la toma de decisiones fundamentada

**Anidado:** Código que realiza una función particular y está contenido dentro de código que realiza una función más amplia

**Anonimización de datos:** Proceso de protección de los datos privados o confidenciales de las personas mediante la eliminación de información que pueda asociarse a ellas

**Anotación:** Texto que explica brevemente los datos o ayuda al público a concentrarse en un aspecto concreto de los datos en una visualización

**Apertura:** Aspecto de la ética de datos que promueve el acceso libre, el uso y el uso compartido de los datos

**Árbol de decisiones**: Herramienta que ayuda a los analistas a tomar decisiones sobre características esenciales de una visualización

**Archivo CSV (valores separados por coma):** Archivo de texto delimitado que utiliza una coma para separar valores

**Archivo de audio:** Almacenamiento en audio digitalizado generalmente en MP3, AAC u otro formato comprimido

**Archivo de registro:** Archivo generado por computadora que registra eventos de sistemas operativos y otros programas de software

**Archivo de vídeo:** Conjunto de imágenes, archivos de audio y otros datos generalmente codificados en un formato comprimido como por ejemplo MP4, MV4, MOV, AVI o FLV

**Argumento (R):** Información necesaria para la ejecución de una función en R

**Asignación de alias:** Nombrar temporalmente una tabla o columna en una consulta para que sea más fácil de leer y escribir

**Atributo**: Característica o cualidad de los datos que se usa para etiquetar una columna en una tabla

**Atributos preatencionales:** Elementos de una visualización de datos que las personas reconocen automáticamente sin hacer un esfuerzo consciente

**AVERAGE:** Función de una hoja de cálculo que muestra el resultado de un promedio de los valores de un rango seleccionado

**AVERAGEIF:** Función de una hoja de cálculo que devuelve el promedio de todos los valores de las celdas de un determinado rango que cumplen una condición especificada

B

**Base de datos normalizada:** Base de datos en la que solo se almacenan datos relacionados en cada tabla

**Base de datos relacional:** Base de datos que contiene una serie de tablas que se pueden conectar para formar relaciones

**Base de datos:** Recopilación de datos almacenados en un sistema informático

**Biblioteca:** Directorio que contiene todos los paquetes instalados de un analista de datos

**Bordes**: Líneas que se pueden agregar alrededor de dos o más celdas en una hoja de cálculo

**Buscar y reemplazar:** Herramienta que encuentra un término de búsqueda específico y lo reemplaza por otro término

C

**C#:** Lenguaje de programación orientado al objeto usado para crear juegos y aplicaciones móviles en la plataforma de desarrollo de código abierto .NET

**C++:** Extensión del lenguaje de programación C que se usa para crear juegos de consola, como los juegos para Xbox

**Cadena de texto:** Grupo de caracteres en una celda, mayormente compuesto por letras

**Cálculo:** Rama de la matemática que implica el estudio de las tasas de cambio y los cambios entre valores que están relacionados por una función

**Campo calculado:** Campo nuevo dentro de una tabla dinámica que realiza ciertos cálculos en función de los valores de otros campos

**Campo:** Información de una fila o columna de una hoja de cálculo; en una tabla de datos, suele ser una columna de la tabla

**Canal:** Aspecto visual o variable que representa características de los datos en una visualización

**Canalización (R):** Herramienta en R para expresar una secuencia de varias operaciones, representadas por “%>%”

**Captación de interés:** Captar la atención de alguien y mantener su interés durante una presentación de datos

**CASE:** Instrucción de SQL que muestra resultados de registros que cumplen con las condiciones al incluir una instrucción si/entonces en una consulta

**Caso práctico:** Una manera común en que los empleadores evalúan destrezas profesionales y obtienen información sobre cómo un candidato aborda desafíos comunes relacionados con los datos

**CAST:** Función de SQL que convierte los datos de un tipo en otro

**Causa raíz:** Razón por la que ocurre un problema

**Causalidad:** Cuando una acción lleva directamente a un resultado, como una relación de causa-efecto

**Ciclo de vida de los datos:** Secuencia de etapas por las que pasan los datos, que incluye planificar, capturar, gestionar, analizar, archivar y destruir

**Ciencia de datos:** Campo de estudio que utiliza datos sin procesar para crear nuevas formas de modelar y entender lo desconocido

**Clasificación:** Sistema para ubicar valores de un conjunto de datos en una escala de logro o estado

**Clave externa:** Campo en una tabla de una base de datos que es una clave primaria en otra tabla (Ver clave primaria)

**Clave primaria:** Identificador en una base de datos que hace referencia a una columna en la que cada valor es único (Ver clave externa)

**Clúster:** Conjunto de puntos de datos en una visualización de datos con valores similares

**COALESCE:** Función de SQL que arroja valores que no son nulos en una lista

**Codificación:** El proceso de escribir instrucciones para una computadora usando la sintaxis de un lenguaje de programación específico

**Código abierto:** Código que está libremente disponible y puede ser modificado y compartido por las personas que lo usan

**Coherencia:** Grado de repetibilidad de los datos desde diferentes puntos de entrada o recopilación

**Combinación de datos:** Método de Tableau que combina datos provenientes de múltiples fuentes de datos

**Compatibilidad:** Qué tan bien dos o más conjuntos de datos pueden trabajar juntos

**Composición de datos:** Proceso que combina las partes individuales en una visualización y las muestra juntas como un todo

**CONCAT:** Función de SQL que une cadenas y crea nuevas cadenas de texto que se pueden usar como claves únicas

**CONCATENATE:** Función de una hoja de cálculo que une dos o más cadenas de texto

**Conjunto de datos:** Recopilación de datos que pueden ser manipulados o analizados como una unidad

**Consentimiento:** Aspecto de la ética de datos que presupone el derecho de una persona a conocer cómo y por qué se utilizarán sus datos personales antes de aceptar proporcionarlos

**Consulta externa:** Instrucción de SQL que contiene una subconsulta

**Consulta interna:** Subconsulta de SQL que está dentro de otra instrucción de SQL

**Consulta:** Solicitud de datos o información de una base de datos

**Contexto:** Condición en la que algo existe o sucede

**Control de acceso:** Funciones como la protección de contraseñas, permisos de usuario y cifrado que se usan para proteger una hoja de cálculo

**Controlador de relleno:** Cuadro en la esquina inferior derecha de una celda seleccionada de una hoja de cálculo que se puede arrastrar a través de las celdas vecinas para seguir una instrucción

**Convenciones de nomenclatura:** Pautas uniformes para el nombre de un archivo que describen el contenido, la fecha de creación y la versión

**Conversión de tipos:** Convertir datos de un tipo en otro

**CONVERT:** Función de SQL que cambia la unidad de medida de un valor en los datos

**Cookie:** Pequeño archivo almacenado en una computadora que contiene información acerca de sus usuarios

**Correlación:** Medición del grado de cambio de dos variables entre sí

**COUNT DISTINCT:** Función de SQL que solo devuelve los valores distintos en un rango especificado

**COUNT**: Función de la hoja de cálculo que cuenta el número de celdas en un rango que cumplen con una condición especificada

**COUNTA:** Función de la hoja de cálculo que cuenta el número total de valores en un rango especificado que cumplen con un criterio especificado

**COUNTIF:** Función de la hoja de cálculo que devuelve el número de celdas dentro de un rango que coinciden con un valor especificado

**CRAN (Comprehensive R Archive Network) (R):** Archivo en línea con paquetes R, código fuente, manuales y documentación

**CREATE TABLE:** Cláusula de SQL que añade una tabla temporal a una base de datos que puede ser utilizada por varias personas

**CSS (Cascading Style Sheets):** Lenguaje de programación usado para el diseño de páginas web que controla los elementos gráficos y la presentación de la página

**Cuarteto de Anscombe:** Cuatro conjuntos de datos que tienen estadísticas de resumen casi idénticas, pero contienen valores graficados diferentes

D

**DATEDIF:** Función de la hoja de cálculo que calcula el número de días, meses o años entre dos fechas

**Datos abiertos:** Datos que están disponibles para el público

**Datos booleanos:** Tipo de datos con solo dos valores posibles, generalmente verdadero o falso

**Datos continuos:** Datos que se miden y que pueden tener casi cualquier valor numérico

**Datos cualitativos:** Medida subjetiva y explicativa de una cualidad o característica

**Datos cuantitativos:** Medida específica y objetiva, como un número, cantidad o rango

**Datos de primera fuente:** Datos recopilados por una persona o por un grupo por medio de sus propios recursos

**Datos de segunda fuente:** Datos recopilados por un grupo directamente de su audiencia y que luego se venden

**Datos de terceros:** Datos proporcionados por fuentes externas que no los recopilaron de forma directa

**Datos desactualizados:** Cualquier dato que haya sido reemplazado por información más nueva y más precisa

**Datos discretos:** Datos que se cuentan y tienen un número limitado de valores

**Datos duplicados:** Cualquier registro que inadvertidamente comparte datos con otro registro

**Datos en formato ancho:** Conjunto de datos en el que cada tema tiene una sola fila con varias columnas para retener los valores de los distintos atributos del tema

**Datos en formato largo:** Conjunto de datos en el que cada fila es un punto de tiempo por tema; por lo tanto, cada tema tiene datos en varias filas

**Datos en tiempo real:** Datos que se actualizan automáticamente

**Datos estáticos:** Datos que no cambian una vez que se registraron

**Datos estructurados:** Datos organizados en un cierto formato, por ejemplo, filas y columnas

**Datos externos:** Datos que se alojan y generan fuera de una organización

**Datos incoherentes:** Datos que usan diferentes formatos para representar lo mismo

**Datos incompletos:** Datos que carecen de campos importantes

**Datos incorrectos/inexactos:** Datos que son completos pero inexactos

**Datos internos:** Datos alojados en los sistemas propios de una empresa

**Datos limpios:** Datos que están completos y son correctos y pertinentes para el problema que se está resolviendo

**Datos no estructurados:** Datos que no se organizan de forma que sea fácil identificarlos

**Datos nominales:** Tipo de datos cualitativos que se categorizan sin un orden establecido

**Datos ordenados (R):** Forma de estandarizar la organización de los datos en R

**Datos ordinales:** Datos cualitativos con un orden o escala establecidos

**Datos sucios:** Datos que están incompletos o son incorrectos o irrelevantes para el problema a resolver

**Datos:** Recopilación de hechos

**Delimitador:** Carácter que indica el principio o el fin de un elemento de datos

***Design Thinking*:** Proceso que se usa para resolver problemas complejos de una manera centrada en el usuario

**Destrezas analíticas:** Cualidades y características asociadas al uso de hechos para resolver problemas

**Destrezas transferibles:** Habilidades y cualidades que se pueden transferir de un trabajo o un sector a otro

**Diagrama de caja:** Visualización de datos que muestra la distribución de valores a lo largo de un eje x

**Diagrama de dispersión:** Visualización de datos que representa relaciones entre las diferentes variables con puntos de datos individuales sin una línea de conexión

**Diagrama de Gantt:** Visualización de datos que muestra la duración de los eventos o actividades en una línea de tiempo

**Diseño de datos:** Cómo se organiza la información

**DISTINCT:** Palabra clave que se agrega a una instrucción SELECT en SQL para recuperar solamente entradas no duplicadas

**Dominio del problema:** Área de análisis que abarca cada actividad que afecta a un problema o se ve afectada por él

**dplyr (R):** Paquete de R en Tidyverse que ofrece un conjunto de funciones uniforme para completar tareas comunes de manipulación de datos

**DROP TABLE:** Cláusula de SQL que elimina una tabla temporal de una base de datos

E

**Ecosistema de datos:** Los distintos elementos que interactúan entre sí para producir, gestionar, almacenar, organizar, analizar y compartir datos

**Ecuación:** Cálculo que implica suma, resta, multiplicación o división (también se denomina expresión matemática)

**Eje X:** Línea horizontal de un gráfico que suele situarse en la parte inferior y, a menudo, se usa para representar escalas de tiempo y categorías puntuales

**Eje Y:** Línea vertical de un gráfico que suele situarse a la izquierda y, a menudo, se usa para representar frecuencias y otras variables numéricas

**Elemento de datos:** Información en un conjunto de datos

**Encabezado:** Primera fila en una hoja de cálculo que hace referencia al tipo de datos en cada columna

**Énfasis:** Principio de diseño que consiste en organizar los elementos visuales para centrar la atención del público en la información importante de una visualización de datos

**Equidad:** Cualidad del análisis de datos que no genera sesgos ni los reafirma

**Equilibrio:** Principio de diseño que consiste en crear un atractivo estético y claridad en una visualización de datos mediante la distribución uniforme de los elementos visuales

**Especialista en almacenamiento de datos:** Profesional que desarrolla procesos y procedimientos para almacenar y organizar datos efectivamente

**Esquema:** Forma de describir cómo se organiza algo, por ejemplo, los datos

**Estadísticas:** El estudio de cómo recopilar, analizar, resumir y presentar datos

**Estética (R):** Propiedad visual de un objeto en un diagrama

**Estrategia de datos:** Gestión de las personas, los procesos y las herramientas que se usan en el análisis de datos

**Estructura de los datos:** Formato para organizar y almacenar datos

**Ética de los datos:** Normas justificadas respecto de lo que está bien y lo que está mal a la hora de recopilar, compartir y usar datos

**Ética:** Normas justificadas respecto de lo que está bien y lo que está mal. Por lo general, presuponen lo que deben hacer los seres humanos, usualmente en términos de derechos, obligaciones, beneficios para la sociedad, equidad o virtudes específicas

**Etiqueta:** Texto en una visualización que identifica un valor o describe una escala

**Etiquetas y anotaciones (R):** Grupo de funciones de R para personalizar un diagrama

**Exactitud:** Grado de conformidad de los datos con respecto a la entidad real que se mide o describe

**Exhaustividad:** Grado en que los datos contienen todas las medidas o componentes deseados

**Expresión matemática:** Cálculo que implica suma, resta, multiplicación o división (también se denomina ecuación)

**Expresión regular (RegEx):** Regla que establece que los valores de una tabla deben coincidir con un patrón prescrito

F

**Facetas (R):** Serie de funciones que divide datos en subconjuntos en una matriz de paneles

**Factor (R):** Objeto que almacena datos de categoría en el que los valores de datos están limitados y generalmente se basan en un grupo finito, como país o año

**Filtrado:** Proceso que muestra solo los datos que cumplen con un criterio específico mientras oculta el resto

**Filtro de panel:** Herramienta que muestra solo los datos que cumplen con un criterio específico y oculta el resto

**Flotante:** Número que contiene un decimal

**Formato condicional:** Herramienta de la hoja de cálculo que cambia la forma en que aparecen las celdas cuando los valores cumplen con una condición específica

**Formato visual:** Apariencia de la visualización de datos que les aporta una estructura y un atractivo estético

**Fórmula:** Conjunto de instrucciones que se utilizan para realizar un cálculo utilizando los datos de una hoja de cálculo

**Foto digital:** Imagen electrónica o computarizada generalmente en formato BMP o JPG

**Framework:** Contexto que necesita una presentación para establecer conexiones lógicas entre la tarea empresarial y las métricas

**FROM:** Sección de una consulta que indica de qué tabla(s) extraer los datos

**Fuente de datos correctos:** Fuente de datos que es confiable, original, integral, actual y citada (ROCCC)

**Fuente de datos erróneos:** Fuente de datos que no es confiable, original, integral, actual ni citada (ROCCC)

**Función (R):** Contenido de código reutilizable para realizar tareas específicas en R

**Función anidada:** Función que está contenida completamente dentro de otra función

**Función matemática**: Función que se utiliza como parte de una fórmula matemática

**Función:** Comando preestablecido que realiza automáticamente un proceso o tarea específico utilizando los datos de una hoja de cálculo

**Fusión de datos:** Proceso de combinar dos o más conjuntos de datos en un único conjunto de datos

**Fusión:** Acuerdo que une dos organizaciones en una organización nueva

**FWF (archivo de ancho fijo):** Archivo de texto con un formato específico que permite guardar datos textuales de manera organizada

G

**Geolocalización:** Ubicación geográfica de una persona o dispositivo mediante información digital

**Geom (R):** Objeto geométrico usado para representar datos

**ggplot2 (R):** Paquete de R en Tidyverse que crea diversas visualizaciones de datos aplicando diferentes propiedades visuales a las variables de datos en R

**Gobierno de datos:** Proceso para garantizar la gestión formal de los recursos de datos de una empresa

**Gráfico circular:** Visualización de datos que usa segmentos de un círculo para representar las proporciones de cada categoría de datos en comparación con el total

**Gráfico combinado:** Visualización de datos que combina más de un tipo de visualización

**Gráfico de anillos:** Visualización de datos en donde los segmentos de un anillo representan valores de datos que se suman en un conjunto

**Gráfico de áreas:** Visualización de datos que utiliza puntos de datos individuales para una variable cambiante conectados por una línea continua con un área de relleno debajo

**Gráfico de barras:** Visualización de datos que usa el tamaño para contrastar y comparar dos o más valores

**Gráfico de burbujas agrupadas:** Visualización de datos que muestra los datos en círculos agrupados

**Gráfico de burbujas**: Visualización de datos que muestra puntos de datos individuales como burbujas, y compara valores numéricos por su tamaño relativo

**Gráfico de calibre:** Visualización de datos que muestra un único resultado en un rango de valores progresivo

**Gráfico de columnas:** Visualización de datos que usa puntos de datos individuales para una variable cambiante, representada mediante columnas verticales

**Gráfico de distribución:** Visualización de datos que muestra la frecuencia de diversos resultados en una muestra

**Gráfico de líneas:** Visualización de datos que usa una o más líneas para mostrar cambios o variaciones en los datos a lo largo del tiempo

**Gráfico de viñetas:** Visualización de datos que muestra los datos como un gráfico de barras horizontal que se mueve hacia un valor deseado

**Gráfico dinámico:** Gráfico creado a partir de los campos en una tabla dinámica

**Gráfico:** Representación gráfica de los datos en una hoja de cálculo

**GROUP BY:** Cláusula de SQL que agrupa las filas que tienen los mismos valores de una tabla en filas de resumen

**Guiones bajos**: Líneas utilizadas para subrayar palabras y conectar caracteres de texto

H

**Habilidades interpersonales:** Rasgos y comportamientos no técnicos que se relacionan con la manera en que las personas trabajan

**Hacer contactos:** Construir relaciones con otros tanto en persona como en línea

**HAVING:** Cláusula de SQL que añade un filtro a una consulta en lugar de la tabla subyacente que solo se puede usar con funciones de agregado

**head() (R):** Función en R que devuelve una vista previa de los nombres de las columnas y las primeras filas de un conjunto de datos

**Hipótesis:** Teoría que se intenta corroborar o refutar con datos

**Histograma:** Visualización de datos que muestra con qué frecuencia los valores de los datos se encuadran en ciertos rangos

**Historia:** Narrativa de una presentación de datos que la hace significativa e interesante

**Hoja de cálculo:** Hoja de cálculo digital

**HTML5:** Lenguaje de programación que brinda una estructura para páginas web y conecta a plataformas de hosting

I

**IDE (entorno de desarrollo integrado):** Aplicación de software que junta todas las herramientas que un analista de datos puede querer usar en un solo lugar

**Informe:** Recopilación estática de datos que se entrega periódicamente a los interesados

**Ingeniero de datos:** Profesional que transforma los datos en un formato útil para su análisis y les da una estructura confiable

**Ingresos:** Cantidad total de ingresos generados por la venta de mercaderías o servicios

**INNER JOIN:** Función de SQL que devuelve registros con valores coincidentes en ambas tablas

**Instrucción condicional:** Declaración de que si una determinada condición es verdadera, entonces debe producirse un determinado evento

**Integridad de datos:** Exactitud, exhaustividad, coherencia y confiabilidad de los datos a lo largo de su ciclo de vida

**Interesados:** Personas que invierten tiempo y recursos en un proyecto y se interesan por su resultado

**Interoperabilidad de los datos:** Capacidad de integrar datos de varias fuentes y un factor clave que conduzca al uso satisfactorio de los datos abiertos entre las empresas y los gobiernos

**Intervalo de confianza:** Rango de valores que transmite qué probabilidad hay de que una estimación estadística refleje la población

J

**Java:** Lenguaje de programación que se usa ampliamente para crear aplicaciones web de empresas que pueden ejecutarse en múltiples clientes

**JOIN:** Función de SQL que se usa para combinar filas de dos o más tablas basadas en una columna relacionada

L

**LEFT JOIN:** Función de SQL que devuelve todos los registros de la tabla izquierda y solo los registros coincidentes de la tabla derecha

**LEFT:** Función que devuelve un número establecido de caracteres a la izquierda de una cadena de texto

**LEN:** Función que indica la longitud de una cadena de texto al contar el número de caracteres que contiene

**Lenguaje de consulta estructurado:** Lenguaje de programación informática usado para comunicarse con una base de datos

**Lenguaje de consulta:** Lenguaje de programación informática usado para comunicarse con una base de datos

**Lenguaje de programación:** Sistema de palabras y símbolos usados para escribir instrucciones que las computadoras siguen

**Leyenda:** Herramienta que identifica el significado de diversos elementos en una visualización de datos

**LIMIT:** Cláusula de SQL que especifica el número máximo de registros devueltos en una consulta

**Limitaciones de datos:** Criterio que determina si un dato está limpio y es válido

**Línea de suavizado (R):** Línea en una visualización de datos que usa suavizado para representar una tendencia

**Lista:** Vector con elementos que pueden ser de cualquier tipo

**Longitud de campo:** Herramienta para determinar cuántos caracteres pueden incluirse en el campo de una hoja de cálculo

**Longitud:** Número de caracteres en una cadena de texto

M

**Macrodatos:** Conjuntos de datos grandes y complejos que, generalmente, se recopilan durante largos períodos y que permiten que los analistas de datos aborden los problemas comerciales de gran alcance

**Manipulación de datos:** Proceso para cambiar los datos, de manera que estén más organizados y sean más fáciles de leer

**Mapa coroplético:** Visualización de datos que colorea las áreas en un mapa según las mediciones o las dimensiones

**Mapa de densidad:** Visualización de datos que representa concentraciones, con un color que representa el número o la frecuencia de los puntos de datos en un área determinada de un mapa

**Mapa de símbolos:** Visualización de datos que muestra una marca sobre una determinada longitud y latitud

**Mapa térmico:** Visualización de datos que usa el contraste del color para comparar categorías en un conjunto de datos

**Mapa:** Visualización de datos que organiza los datos geográficamente

**Mapeo (R):** Proceso de emparejar una variable específica de un conjunto de datos con una estética específica

**Mapeo de datos:** Proceso de hacer coincidir campos entre una fuente de datos y otra

**Marca:** Objeto visual en una visualización de datos, como un punto, una línea o una forma

**Marco de datos:** Conjunto de columnas que contienen datos, que es similar a una hoja de cálculo o una tabla de SQL

**Margen de beneficio:** Porcentaje que indica cuántos centavos de ganancia se generaron por cada dólar de venta

**Margen de error:** Cantidad máxima que se espera que los resultados de la muestra difieran de los de la población real

**MATCH:** Función de la hoja de cálculo que se usa para ubicar la posición de un valor de búsqueda específico

**Matriz:** Conjunto bidimensional de elementos de datos con filas y columnas

**MAX:** Función de la hoja de cálculo que muestra el resultado del valor numérico más alto de un rango de celdas

**MAXIFS:** Función de la hoja de cálculo que devuelve el valor máximo de un determinado rango que cumple una condición especificada

**Mentalidad técnica:** Capacidad de dividir las cosas en pasos o piezas más pequeñas y trabajar con ellas de forma ordenada y lógica

**Mentor:** Persona que comparte su conocimiento, sus habilidades y su experiencia para ayudar a otras personas a crecer, tanto en el campo profesional como en el personal

**Metadato administrativo:** Metadato que indica el origen técnico de un recurso digital

**Metadatos descriptivos:** Metadatos que describen datos y que se pueden utilizar para identificarlos más adelante

**Metadatos estructurales:** Metadatos que indican cómo se organizan ciertos datos y si forman parte de una recopilación de datos o de varias

**Metadatos:** Datos sobre los datos

**Método McCandless:** Método que se utiliza para presentar visualizaciones de datos y que pasa de la información general a la específica

**Metodología SMART:** Herramienta para determinar la eficacia de una pregunta basándose en si es específica, medible, orientada a la acción, relevante y con plazos determinados

**Métrica de negocios:** Criterio de medición que se utiliza para resolver una tarea empresarial

**Métrica:** Tipo único y cuantificable de datos que pueden utilizarse para medición

**Microdatos:** Puntos de datos pequeños, específicos, que generalmente involucran un breve período y que son útiles para tomar decisiones diarias

**MID:** Función que extrae un segmento desde el medio de una cadena de texto

**MIN**: Función de la hoja de cálculo que muestra el resultado del valor numérico más bajo de un rango de celdas

**MINIFS:** Función de la hoja de cálculo que devuelve el valor mínimo de un rango determinado que cumple una condición especificada

**Modelo de datos:** Herramienta para organizar los elementos de los datos y la forma en que se relacionan entre ellos

**Modelo mental:** Proceso de pensamiento del analista de datos y enfoque a un problema

**Módulo:** Operador (%) que devuelve el resto cuando se divide un número por otro

**Movimiento:** Principio de diseño que consiste en organizar los elementos visuales para guiar la atención del público de una parte de una visualización de datos a otra

**Muestra:** En el análisis computacional de datos, segmento de una población que la representa en su totalidad

**Muestreo aleatorio:** Forma de seleccionar una muestra de una población para que todos los tipos posibles de la muestra tengan la misma oportunidad de ser elegidos

**Muestreo imparcial:** Cuando la muestra de la población que se está midiendo es representativa de la población como un todo

**mutate() (R):** Función de R que realiza cambios a un marco de datos separando y combinando columnas o creando nuevas variables

N

**Narrativa de datos:** Comunicar el significado de un conjunto de datos con elementos visuales y una historia personalizada según el público

**Narrativa:** (Ver Historia)

**Nivel de confianza:** Probabilidad de que el tamaño de una muestra refleje con precisión a la porción más grande de la población

**Notebook:** Entorno de programación interactivo y editable para generar informes de datos y mostrar destrezas en el uso de datos

**Nube:** Lugar para mantener los datos en línea, en vez de guardarlos en el disco duro de una computadora

**Nulo:** Indicación de que un valor no existe en un conjunto de datos

O

**Objetivo métrico:** Objetivo medible establecido por una empresa y evaluado mediante métricas

**Obligatorio:** Valor de datos que no puede quedar en blanco ni vacío

**Observación:** Atributos que describen los datos contenidos en la fila de una tabla

**Oficina del Censo de los Estados Unidos:** Agencia del Departamento de Comercio de los Estados Unidos que funciona como proveedor principal de la nación de datos de calidad sobre las personas y la economía

**Operador aritmético:** Operador usado para realizar operaciones matemáticas básicas como suma, resta, multiplicación y división

**Operador de asignación:** Operador usado para asignar valores a variables y vectores

**Operador lógico:** Operador que devuelve un tipo de datos lógico

**Operador relacional:** Operador usado para comparar valores, también conocido como comparador

**Operador:** Símbolo que designa la operación o cálculo a realizarse

**Orden de las operaciones:** Uso de paréntesis para agrupar los valores de la hoja de cálculo a fin de aclarar el orden en el que deben realizarse las operaciones

**Ordenación:** Proceso de organizar los datos en un orden significativo para que sean más fáciles de entender, analizar y visualizar

**Ordenar hoja:** Función del menú de la hoja de cálculo que ordena todos los datos por la clasificación de una columna específica ordenada y mantiene los datos juntos a través de las filas

**ORDER BY:** Cláusula de SQL que ordena los resultados devueltos en una consulta

**Organización Mundial de la Salud:** Organización cuya función principal es la de dirigir y coordinar la salud a nivel internacional dentro del sistema de las Naciones Unidas

**OUTER JOIN:** Función de SQL que combina RIGHT y LEFT JOIN para devolver todos los registros coincidentes en ambas tablas

P

**Paleta de color divergente:** Tema de color que muestra dos rangos de valores de datos que usan dos tonalidades diferentes, donde la intensidad del color representa la magnitud de los valores

**Panel:** Herramienta que monitorea los datos entrantes en vivo

**Paquete (R):** Una unidad de código R reproducible

**Patrocinador:** Profesional que se compromete a hacer progresar la carrera profesional de otra persona

**Patrón:** Principio de diseño que usa elementos visuales similares para demostrar tendencias y relaciones en una visualización de datos

**Pensamiento analítico:** Proceso de identificar y definir un problema para luego resolverlo mediante el uso de datos de manera organizada, paso a paso

**Pensamiento estructurado:** Proceso de reconocer el problema o la situación actuales, organizar la información disponible, revelar déficits y oportunidades e identificar opciones

**PHP (preprocesador de hipertexto):** Lenguaje de programación para el desarrollo de aplicaciones web

**Píxel:** En imágenes digitales, área pequeña de iluminación en una pantalla de visualización que, cuando se combina con otras áreas adyacentes, forma una imagen digital

**Población:** En el análisis computacional de datos, todos los valores posibles en un conjunto de datos

**Poder estadístico:** Probabilidad de que una prueba de importancia reconozca un efecto presente

**Poner de relieve:** Explorar los datos para identificar rápidamente la información más importante

**Portfolio:** Conjunto de materiales que puede compartirse con empleadores potenciales

**Pregunta con límite de tiempo:** Pregunta que especifica un plazo para ser analizada

**Pregunta específica:** Pregunta simple, significativa y enfocada en un solo tema o en algunas ideas estrechamente relacionadas entre sí

**Pregunta injusta:** Pregunta en la que se hacen suposiciones o que es difícil de responder honestamente

**Pregunta medible:** Pregunta cuyas respuestas se pueden cuantificar y evaluar

**Pregunta orientada a la acción:** Pregunta cuyas respuestas conducen al cambio

**Pregunta principal:** Pregunta que orienta a las personas hacia cierta respuesta

**Pregunta relevante:** Pregunta que tiene importancia para el problema que se debe resolver

**Privacidad de los datos:** Preservación de la información sobre los datos de una persona cada vez que ocurre una transacción de datos

**Proceso de análisis de datos:** Las seis fases de preguntar, preparar, procesar, analizar, compartir y actuar cuyo propósito es el de obtener conocimiento que propicie la toma de decisiones informada

**Proceso de validación de datos:** Proceso de comprobar y volver a comprobar la calidad de los datos para que sean completos, precisos, seguros y coherentes

**Programación informática:** El proceso de darle instrucciones a una computadora para que realice una acción o un conjunto de acciones

**Propiedad:** Aspecto de la ética de datos que presupone que cada persona es dueña de los datos sin procesar que proporciona y que tiene control primordial sobre su uso, procesamiento y uso compartido

**Proporción:** Principio de diseño que consiste en usar el tamaño relativo y la disposición de los elementos visuales para mostrar la información en una visualización de datos

**Pruebas A/B:** Proceso de probar dos variaciones de la misma página web para determinar qué página es más exitosa para atraer el tráfico de usuarios y generar ingresos

**Python:** Lenguaje de programación de uso general

Q

**Quitar duplicados:** Herramienta de la hoja de cálculo que busca y elimina automáticamente las entradas duplicadas de una hoja de cálculo

R

**R:** Lenguaje de programación que se usa para análisis estadísticos, visualización y análisis de datos

**Rango de datos:** Valores numéricos que se encuentran entre valores máximos y mínimos predefinidos

**Rango de ordenación:** Función del menú de la hoja de cálculo que ordena un rango especificado y preserva las celdas que no están incluidas en ese rango

**Rango:** Conjunto de dos o más celdas en una hoja de cálculo

**readr (R):** Paquete de R en Tidyverse usado para importar datos

**Redes sociales:** Sitios web y aplicaciones donde los usuarios crean y comparten contenido o interactúan entre sí

**Redundancia:** Cuando los mismos datos se almacenan en dos o más lugares

**Referencia absoluta:** Referencia dentro de una función que está bloqueada para que las filas y las columnas no cambien si se copia la función

**Referencia de celda:** Celda o rango de celdas en una hoja de cálculo que se usa generalmente en fórmulas y funciones

**Reformulación:** Proceso de replantear un problema o desafío, que se redirecciona luego hacia una posible resolución

**Registro de cambios:** Archivo que contiene una lista ordenada cronológicamente de las modificaciones realizadas en un proyecto

**Registro:** Conjunto de datos relacionados en una tabla de datos, generalmente sinónimo de fila

**Reglamento General de Protección de Datos de la Unión Europea (RGPD):** Organismo formulador de políticas en la Unión Europea, creado para ayudar a proteger a las personas y sus datos

**Relatividad:** Proceso de considerar las observaciones en relación con o en proporción a algo más

**Repetición:** Principio de diseño que consiste en repetir elementos visuales para demostrar el significado en una visualización de datos

**Replicación de datos:** Proceso de almacenamiento de datos en varias ubicaciones

**Repositorio de metadatos:** Base de datos creada para almacenar metadatos

**Resaltar tabla:** Visualización de datos que usa formato condicional y color en una tabla

**Retorno de la inversión (ROI):** Fórmula que utiliza las métricas de inversión y ganancias para evaluar el éxito de una inversión

**RIGHT JOIN:** Función de SQL que devuelve todos los registros de la tabla derecha y solo los registros coincidentes de la izquierda

**RIGHT:** Función que muestra un número establecido de caracteres a la derecha de una cadena de texto

**Ritmo:** Principio de diseño que consiste en crear movimiento y fluidez en una visualización de datos para captar la atención de una audiencia

**ROUND:** Función de SQL que devuelve un número redondeado hasta un número determinado de decimales.

**Ruby:** Lenguaje de programación orientado a objetos para el desarrollo de aplicaciones web

S

**Seguridad de los datos:** Emplear medidas de seguridad para proteger los datos contra el acceso no autorizado o contra la corrupción

**Selección:** Conjunto de valores en celdas de una hoja de cálculo

**SELECT INTO:** Cláusula de SQL que copia datos de una tabla a una tabla temporal sin añadir la nueva tabla a la base de datos

**SELECT:** Sección de una consulta que indica de qué columna(s) extraer los datos

**Sesgo de confirmación:** Tendencia de buscar o interpretar la información de manera que confirme creencias preexistentes

**Sesgo de interpretación:** Tendencia a interpretar situaciones ambiguas de manera positiva o negativa

**Sesgo de los datos:** Cuando una preferencia a favor o en contra de una persona, un grupo de personas o una cosa sesga sistemáticamente los resultados del análisis de datos en una cierta dirección

**Sesgo del investigador:** Tendencia de distintas personas a observar las cosas de forma diferente (Ver Sesgo del observador)

**Sesgo del muestreo:** Representar en mayor o en menor medida a ciertos miembros de una población debido a que se trabaja con una muestra que no representa a la población en su totalidad

**Sesgo del observador:** Tendencia de distintas personas a observar las cosas de forma diferente (también se denomina sesgo del investigador)

**Sesgo:** Preferencia consciente o subconsciente a favor o en contra de una persona, un grupo de personas o una cosa

**Significancia estadística:** Probabilidad de que los resultados de una muestra no se deban a una posibilidad aleatoria

**Sintaxis:** Estructura predeterminada de un lenguaje que incluye todas las palabras, los símbolos y la puntuación requeridos, así como su correcta ubicación

**SPLIT:** Función de la hoja de cálculo que divide el texto en base a un carácter específico y ubica cada fragmento en una nueva celda separada

**SQL:** (Ver Lenguaje de consulta estructurado)

**Suavizado (R):** Proceso usado para hacer que las visualizaciones de datos en R sean más claras y legibles

**Suavizado con GAM (modelo aditivo generalizado) (R):** Proceso para suavizar diagramas con un gran número de puntos

**Suavizado Loess (R):** Proceso usado para suavizar diagramas con menos de 1,000 puntos

**Subcadena:** Subconjunto de una cadena de texto

**Subconsulta:** Consulta de SQL anidada dentro de una consulta más grande

**SUBSTR:** Función de SQL que extrae una subcadena de una variable de cadenas

**Subtítulo:** Texto que complementa al título al indicar el contexto y la descripción

**SUM:** Función de una hoja de cálculo que suma los valores de un rango de celdas seleccionadas

**SUMIF:** Función de una hoja de cálculo que suma datos numéricos basados en una condición

**SUMPRODUCT:** Función que multiplica las matrices y muestra el resultado de la suma de esos productos

**Swift:** Lenguaje de programación para macOS, iOS, watchOS y tvOS

T

**Tabla de resumen:** Tabla que se usa para resumir información estadística sobre los datos

**Tabla dinámica:** Herramienta de resumen de datos que se utiliza para clasificar, reorganizar, agrupar, contar, totalizar o promediar datos

**Tabla temporal:** Tabla de una base de datos que se crea y existe temporalmente en el servidor de una base de datos

**Tableau:** Plataforma de inteligencia de negocios y análisis que ayuda a las personas a visualizar y comprender datos, y a tomar decisiones basadas en los datos

**Tarea empresarial:** Pregunta o problema que el análisis de datos resuelve para un negocio

**Tasa de respuesta estimada**: Número promedio de personas que suele completar una encuesta

**Tasa de rotación:** Ritmo en el que los empleados abandonan voluntariamente una empresa

**Testeo de hipótesis:** Proceso que se realiza para determinar si una encuesta o un experimento tiene resultados significativos

**Texto alternativo:** Texto que ofrece una alternativa al contenido no textual, como las imágenes y los vídeos

**Tibble (R):** Variación optimizada de marcos de datos

**tidyr (R):** Paquete de R en Tidyverse usado para limpieza de datos para generar datos ordenados

**Tidyverse (R):** Sistema de paquetes en R con una filosofía de diseño en común para la manipulación, exploración y visualización de datos

**Tipo de datos de cadena:** Secuencia de caracteres y puntuación que contiene información textual (también se denomina tipo de datos de texto)

**Tipo de datos de texto:** Secuencia de caracteres y puntuación que contiene información textual (también se denomina tipo de datos de cadena)

**Tipo de datos:** Atributo que describe cierto dato según sus valores, su lenguaje de programación o las operaciones que puede realizar

**Tipos de problemas:** Distintos problemas que encuentra el analista de datos; entre ellos, categorizar elementos, descubrir conexiones, hallar patrones, identificar temas, hacer predicciones y detectar algo inusual

**Título:** Texto en la parte superior de una visualización que comunica los datos que se están presentando

**Toma de decisiones basada en datos:** Uso de datos para guiar la estrategia empresarial

**Toma de decisiones inspirada en datos:** Explorar diferentes fuentes de datos para descubrir qué tienen en común

**Transacciones:** Aspecto de la ética de los datos que considera que las personas deben conocer las transacciones financieras resultantes del uso de sus datos personales y la dimensión de esas transacciones

**Transferencia de datos:** Proceso de copiar datos de un dispositivo de almacenamiento a la memoria de la computadora, o de una computadora a otra

**Transparencia de la transacción:** Aspecto de la ética de datos que presupone que se deben explicar todas las actividades de procesamiento de datos y los algoritmos a la persona que proporciona los datos y que también presupone que esta persona debe comprenderlos

**TRIM:** Función que quita los espacios al principio, al final o los repetidos en los datos

**TSV (archivo con valores separados por tabuladores):** Archivo de texto que almacena una tabla de datos separando columnas de datos con tabuladores

U

**Único:** Valor que no puede tener un duplicado

**Unidad:** Principio de diseño que consiste en usar elementos visuales que se complementan entre sí para crear un atractivo estético y claridad en la visualización de datos

V

**Validación de campos cruzados:** Proceso que garantiza el cumplimiento de ciertas condiciones para múltiples campos de datos

**Validación de datos:** Herramienta para corroborar la exactitud y la calidad de los datos

**Validez:** Grado de conformidad de los datos con respecto a las restricciones cuando se los ingresa, recopila o crea

**VALUE:** Función de la hoja de cálculo que convierte una cadena de texto que representa un número en un valor numérico

**Variable (R):** Representación de un valor en R que puede almacenarse para uso posterior

**Variedad:** Principio de diseño que consiste en usar diferentes tipos de elementos visuales en una visualización de datos para captar la atención del público

**Vector (R):** Grupo de elementos de datos del mismo tipo almacenados en una secuencia unidimensional en R

**Verificación:** Proceso que confirma que se ejecutó correctamente un esfuerzo de limpieza de datos y que los datos resultantes son precisos y confiables

**Vignette (R):** Documentación para un paquete en R que describe el problema que el paquete está diseñado para resolver, explica cómo pueden usarse sus funciones y enumera cualquier dependencia de otros paquetes

**Vista en círculo:** Visualización de datos que muestra la fortaleza comparativa en los datos

**Visualización de datos:** Representación gráfica de los datos

**Visualización estática:** Visualización de datos que no cambia a lo largo del tiempo a menos que se edite

**Visualización:** (Ver Visualización de datos)

**Visualizaciones dinámicas:** Visualizaciones de datos que son interactivas o cambian a lo largo del tiempo

**VLOOKUP:** Función de una hoja de cálculo que busca verticalmente cierto valor en una columna y arroja la información correspondiente

W

**WHERE:** Sección de una consulta que especifica los criterios que deben cumplir los datos solicitados

**WITH:** Cláusula de SQL que crea una tabla temporal que se puede consultar muchas veces