

Introducción a Big Data y la ciencia de datos.

1.1 Big Data: Datos estructurados y datos no estructurados

1.2 Data Analysis, Data Analytics, Data Engineer, Data Scientist

1.3 De la hoja de cálculo al trabajo en la nube

1.4 El arte de la visualización de los datos: Flourish – Gapminder - Gephi: Estadísticas animadas

1.5 Casos prácticos

Big Data

El Big Data es un término que describe el gran volumen de datos, tanto estructurados como no estructurados, que se generan y recopilan a diario. Estos datos pueden provenir de una variedad de fuentes, como redes sociales, sensores, dispositivos móviles, transacciones comerciales y registros gubernamentales.

Algunos ejemplos de cómo se utiliza el Big Data en la práctica incluyen:

- **Marketing:** Las empresas utilizan el Big Data para segmentar a los clientes, personalizar las ofertas y medir el rendimiento de las campañas de marketing.
- **Ventas:** Las empresas utilizan el Big Data para predecir la demanda de productos, optimizar las rutas de entrega y mejorar la atención al cliente.
- **Fabricación:** Las empresas utilizan el Big Data para mejorar la eficiencia de la producción, reducir los costes y prevenir los defectos.
- **Salud:** Las empresas utilizan el Big Data para desarrollar nuevos tratamientos, mejorar la atención al paciente y prevenir las enfermedades.
- **Gobierno:** Los gobiernos utilizan el Big Data para mejorar la seguridad pública, la planificación urbana y la prestación de servicios públicos.



Las 6V's del Big Data (Grandes volúmenes de datos)

- **Volumen:** Se refiere a la cantidad masiva de datos generados, recopilados y almacenados. Con Big Data, las organizaciones manejan conjuntos de datos que pueden variar desde gigabytes hasta terabytes o incluso petabytes.
- **Variedad:** Indica la diversidad de tipos de datos. Big Data no se trata solo de datos estructurados (como bases de datos tradicionales), sino también de datos no estructurados y semiestructurados, como texto, imágenes, videos, sonidos, registros de redes sociales, etc.
- **Velocidad:** Hace referencia a la velocidad a la que se generan, procesan y deben utilizarse los datos. En algunos casos, como en aplicaciones de transmisión en vivo, la velocidad de procesamiento de datos es crítica.
- **Veracidad:** Se relaciona con la calidad y confiabilidad de los datos. En un entorno de Big Data, los datos pueden provenir de diversas fuentes con diferentes grados de precisión y confiabilidad, por lo que es esencial garantizar la veracidad de los datos.
- **Valor:** Es la capacidad de convertir los datos en información valiosa y, finalmente, en conocimiento para la toma de decisiones. El objetivo último de trabajar con Big Data es obtener información significativa que agregue valor a una organización.
- **Viabilidad:** Algunas fuentes añaden la "viabilidad" como una sexta V, refiriéndose a la necesidad de evaluar la viabilidad económica y técnica de manejar y procesar grandes volúmenes de datos. Esto incluye considerar la infraestructura necesaria, los costos asociados y la capacidad para gestionar eficientemente los recursos.

Que es la ciencia de datos

La ciencia de datos es un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas científicos para extraer conocimiento y penetrar en la comprensión de fenómenos complejos. Combina elementos de estadísticas, matemáticas y ciencias de la computación para analizar y comprender datos.

El proceso típico de ciencia de datos implica la recopilación, limpieza, análisis y visualización de datos para obtener información significativa y tomar decisiones informadas.

- **Recopilación de datos:** Implica la adquisición de datos de diversas fuentes, que pueden ser estructurados (como bases de datos) o no estructurados (como texto o imágenes).
- **Preprocesamiento de datos:** Incluye la limpieza y transformación de datos para eliminar errores, valores atípicos o información innecesaria, y para preparar los datos para su análisis.
- **Análisis de datos:** Aplica técnicas estadísticas, algoritmos de aprendizaje automático y otros métodos para descubrir patrones, relaciones y tendencias en los datos.



Machine Learning: Es una rama de la ciencia de datos que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender de los datos y realizar tareas específicas sin ser programadas explícitamente.

Visualización de datos: Utiliza gráficos y visualizaciones para representar de manera efectiva la información descubierta durante el análisis, facilitando la interpretación y la toma de decisiones.

Big Data: Se refiere al manejo y análisis de conjuntos de datos extremadamente grandes y complejos que no pueden ser procesados con herramientas tradicionales de bases de datos.

Inteligencia Artificial (IA): A menudo se entrelaza con la ciencia de datos, ya que la IA utiliza técnicas de aprendizaje automático para permitir a las máquinas realizar tareas que normalmente requieren inteligencia humana.

Relación de la ciencia de datos con big data

Proceso de Datos.

Big Data: Se refiere a conjuntos de datos extremadamente grandes que superan la capacidad de las herramientas de procesamiento de datos convencionales. Esto puede incluir datos estructurados, no estructurados y semiestructurados. Las tecnologías de big data como Hadoop y Spark se utilizan para gestionar, almacenar y procesar estos volúmenes masivos de datos.

Ciencia de Datos: Implica el uso de técnicas avanzadas de análisis de datos, estadísticas y aprendizaje automático para extraer conocimientos y patrones significativos de los datos. La ciencia de datos puede aprovechar las capacidades de procesamiento de big data para analizar grandes cantidades de información y descubrir relaciones complejas.



Herramientas y Tecnologías.

Big Data: Implica el uso de tecnologías específicas para gestionar y procesar grandes cantidades de datos. Esto puede incluir sistemas de almacenamiento distribuido, bases de datos NoSQL, y marcos de procesamiento paralelo como Apache Hadoop y Apache Spark.

Ciencia de Datos: Utiliza herramientas y técnicas para analizar datos y extraer información. Esto puede incluir lenguajes de programación como Python y R, bibliotecas como Pandas y scikit-learn, y entornos de desarrollo integrados (IDE) como Jupyter.

Escalabilidad.

Big Data: Se centra en la escalabilidad para manejar volúmenes masivos de datos y asegurar un rendimiento eficiente incluso cuando se trabaja con petabytes de información.

Ciencia de Datos: Puede aprovechar la escalabilidad proporcionada por las tecnologías de big data para analizar conjuntos de datos a gran escala y realizar modelos más complejos y precisos.

Extracción de Valor.

Big Data: Proporciona la infraestructura necesaria para almacenar y procesar grandes cantidades de datos, pero no necesariamente realiza análisis detallados o modelado predictivo por sí mismo.

Ciencia de Datos: Se centra en extraer información significativa de los datos, identificando patrones, construyendo modelos predictivos y proporcionando conocimientos accionables para la toma de decisiones.

Nota: el big data proporciona la infraestructura para gestionar grandes volúmenes de datos, mientras que la ciencia de datos se centra en extraer conocimientos valiosos de esos datos. Juntas, estas disciplinas permiten abordar problemas complejos y tomar decisiones informadas basadas en datos a gran escala.



Datos estructurados

En el contexto de Big Data, los datos estructurados se refieren a la información organizada de manera formal y predefinida, que se presenta en un formato tabular con filas y columnas. Estos datos están altamente organizados y son fáciles de almacenar, procesar y analizar. A menudo se representan en bases de datos relacionales y se pueden manipular utilizando consultas SQL.

Las características clave de los datos estructurados incluyen:

Formato Tabular: Los datos están organizados en tablas con filas y columnas, donde cada columna tiene un nombre y un tipo de dato específico.

Esquema Definido: Existe un esquema predefinido que describe la estructura de los datos, especificando qué tipo de datos se almacenan en cada columna.

Facilidad de Consulta: Se pueden realizar consultas mediante lenguajes como SQL para extraer información específica de la base de datos.

Datos no estructurados

Los datos no estructurados en el contexto del big data se refieren a información que no sigue un formato predefinido o no está organizada de manera tabular. A diferencia de los datos estructurados, que se almacenan en bases de datos relacionales y se presentan en filas y columnas, los datos no estructurados no tienen un formato uniforme y pueden incluir diversos tipos de información, como texto sin formato, imágenes, audio, video, mensajes de redes sociales, correos electrónicos, entre otros.

Algunos ejemplos de datos no estructurados incluyen:

- Texto sin formato: Documentos, informes, blogs, artículos, correos electrónicos, entre otros.
- Imágenes: Fotografías, gráficos, mapas, escaneos, entre otros.
- Audio: Archivos de sonido, grabaciones, llamadas telefónicas, entre otros.
- Video: Grabaciones de video, transmisiones en vivo, clips, entre otros.
- Datos de redes sociales: Mensajes, comentarios, tweets, publicaciones, entre otros.



Tipos de formatos de datos estructurados en big data

- CSV (Comma-Separated Values): Este formato utiliza comas para separar los valores en cada fila, y las filas representan registros. Es simple y fácil de entender, pero puede no ser eficiente para conjuntos de datos muy grandes.
- JSON (JavaScript Object Notation): Es un formato de intercambio de datos ligero y fácil de leer que utiliza pares de clave-valor. Es comúnmente utilizado en aplicaciones web y es fácilmente parseable por muchas tecnologías.
- XML (eXtensible Markup Language): Similar a JSON, XML es un formato de marcado que utiliza etiquetas para estructurar la información. Es más verboso que JSON y ha sido ampliamente utilizado en la web y en integraciones de sistemas.
- Avro: Es un formato de serialización binaria desarrollado dentro del proyecto Apache Hadoop. Avro es compacto, eficiente y permite la evolución de esquemas, lo que lo hace adecuado para sistemas Big Data.
- Parquet: Un formato de almacenamiento columnar que está diseñado para ser eficiente en términos de almacenamiento y procesamiento. Parquet es ampliamente utilizado en entornos de Big Data, especialmente con herramientas como Apache Spark y Apache Hive.

Tipos de formatos de datos estructurados en big data

- ORC (Optimized Row Columnar): Similar a Parquet, es otro formato de almacenamiento columnar que se utiliza para mejorar el rendimiento en entornos Big Data, especialmente con tecnologías como Apache Hive.
- Apache Arrow: Es una plataforma de procesamiento de datos en memoria que también define un formato de intercambio de datos columnar. Arrow se utiliza para facilitar la transferencia eficiente de datos entre diferentes sistemas.
- HDF5 (Hierarchical Data Format version 5): Es un formato de archivo y conjunto de herramientas para gestionar datos complejos. Aunque no es exclusivo de Big Data, se utiliza en este contexto para almacenar grandes conjuntos de datos científicos y de otro tipo.
- Protocol Buffers (protobuf): Un formato de serialización binaria desarrollado por Google. Es eficiente en términos de tamaño y se utiliza en diversas aplicaciones, incluidas aquellas en el ámbito de Big Data.



Tipos de formatos de datos “no estructurados” en big data

- Texto sin formato (Plain Text): Datos en forma de texto simple, como documentos, correos electrónicos, tweets o cualquier otro contenido de texto sin una estructura rígida.
- Datos de registro (Log Files): Archivos que registran eventos y actividades de sistemas, aplicaciones o servicios. Estos archivos pueden contener información diversa y suelen seguir un formato de texto.
- Imágenes: Datos visuales en forma de imágenes, fotos o gráficos. Estos datos pueden almacenarse en formatos como JPEG, PNG, GIF, entre otros.
- Audio y Video: Datos en forma de archivos de audio o video. Los formatos comunes incluyen MP3, WAV para audio, y MP4, AVI para video.
- Documentos Binarios: Archivos binarios que no siguen una estructura de texto, como documentos de Ofimática, archivos PDF, entre otros.
- Datos de Redes Sociales: Información recopilada de plataformas de redes sociales, que pueden incluir texto, imágenes, videos, enlaces y otros tipos de contenido.

Tipos de formatos de datos “no estructurados” en big data

- Datos de Sensores y Dispositivos IoT: Datos generados por sensores y dispositivos conectados a Internet de las Cosas (IoT). Estos datos pueden ser variados y no seguir una estructura específica.
- Datos Geoespaciales: Información relacionada con la ubicación, como coordenadas geográficas, mapas y datos de geolocalización.
- Secuencias de ADN: En el ámbito de la bioinformática, los datos genéticos pueden representarse en forma de secuencias de ADN, que son datos no estructurados y complejos.
- Flujos de Datos en Tiempo Real: Datos que llegan continuamente y en tiempo real, como transmisiones de eventos en redes sociales, registros de servidores en tiempo real, etc.
- Correo Electrónico: Los mensajes de correo electrónico, que pueden contener texto, archivos adjuntos y otros elementos multimedia.

Nota: La gestión de datos no estructurados en entornos de Big Data puede ser un desafío debido a su diversidad y complejidad. Herramientas y plataformas, como Apache Hadoop y Apache Spark, se utilizan para procesar y analizar datos no estructurados en conjunción con datos estructurados.



Introducción a Big Data y la ciencia de datos.

1.1 Big Data: Datos estructurados y datos no estructurados

1.2 Data Analysis, Data Analytics, Data Engineer, Data Scientist

1.3 De la hoja de cálculo al trabajo en la nube

1.4 El arte de la visualización de los datos: Flourish – Gapminder -
Gephi: Estadísticas animadas

1.5 Casos prácticos

El **Análisis de Datos** es una disciplina fundamental en el mundo actual, marcado por la explosión de la información digital. Se trata de un proceso sistemático que implica la inspección, transformación y modelado de conjuntos de datos con el propósito de descubrir patrones, tendencias y conocimientos que puedan ser utilizados para la toma de decisiones informadas. En esencia, es el arte y la ciencia de convertir datos crudos en información valiosa.

- Exploración Inicial: Los analistas revisan la estructura y la calidad de los datos, identificando posibles errores o inconsistencias.
- Entendimiento del Contexto: Comprender el contexto en el que se generaron los datos es crucial para una interpretación precisa.
- Limpieza de Datos: Eliminación de duplicados, manejo de valores atípicos y corrección de errores para garantizar la integridad de los datos.
- Transformación: Conversión de datos en un formato adecuado y creación de variables adicionales para facilitar el análisis.



- Identificación de Patrones: Búsqueda de relaciones y tendencias ocultas mediante herramientas estadísticas y visuales.
- Correlación y Causalidad: Evaluación de la relación entre variables y la identificación de factores que pueden influir en los resultados.
- Modelado Predictivo: Uso de algoritmos y modelos estadísticos para predecir futuros eventos o tendencias.
- Validación del Modelo: Evaluación del rendimiento del modelo en datos adicionales para garantizar su generalización.
- Recomendaciones y Estrategias: Utilización de resultados predictivos para sugerir acciones específicas que puedan mejorar los resultados.
- Optimización: Identificación de escenarios óptimos para maximizar o minimizar ciertos resultados.

Herramientas y Tecnologías:

El análisis de datos se beneficia enormemente de las herramientas y tecnologías modernas. Lenguajes de programación como Python y R, plataformas de Business Intelligence (BI) como Tableau, y frameworks de Machine Learning como TensorFlow son esenciales para llevar a cabo análisis sofisticados y automatizados.

Aplicaciones Prácticas:

El Análisis de Datos tiene aplicaciones en una variedad de campos, desde la toma de decisiones empresariales hasta la investigación científica. Desde la optimización de procesos internos en una empresa hasta la identificación de patrones de comportamiento del consumidor, el análisis de datos proporciona información valiosa que impulsa el progreso y la eficiencia.

Consideraciones Éticas:

La privacidad y la ética son consideraciones críticas en el análisis de datos. La manipulación inadecuada de información sensible puede tener consecuencias significativas. La transparencia en el proceso y la consideración de posibles sesgos son fundamentales para garantizar resultados justos y confiables.



La **Data Analytics** se refiere al proceso de examinar conjuntos de datos extensos con el objetivo de descubrir patrones, tendencias, correlaciones y conocimientos significativos. A través de diversas técnicas y herramientas, esta disciplina busca convertir datos en información valiosa, proporcionando una visión más clara y profunda del contexto en el que operan las organizaciones.

La Data Analytics encuentra aplicaciones en diversos campos y sectores, desde el comercio minorista y la atención médica hasta la logística y la banca. Algunos ejemplos incluyen la personalización de experiencias de usuario en plataformas digitales, la optimización de cadenas de suministro y la identificación de patrones de enfermedades en poblaciones.

Aunque la analítica de datos ofrece enormes beneficios, no está exenta de desafíos. La privacidad y la seguridad de los datos, así como la interpretación adecuada de los resultados, son preocupaciones clave. La ética en el manejo de datos se vuelve esencial para garantizar un uso responsable y beneficioso.

El Rol de la Analítica de Datos en la Toma de Decisiones Estratégicas

La toma de decisiones estratégicas basada en datos se está convirtiendo en la norma en el mundo empresarial. La analítica de datos proporciona a los líderes empresariales la capacidad de tomar decisiones más informadas y precisas, alineando sus estrategias con las necesidades cambiantes del mercado. La capacidad de evaluar riesgos, identificar oportunidades y adaptarse ágilmente a las condiciones del entorno empresarial confiere una ventaja estratégica significativa.

Componentes

- **Recopilación de Datos:** El primer paso es recopilar datos relevantes de diversas fuentes, como bases de datos, redes sociales, sensores, y más.
- **Limpieza y Transformación de Datos:** Los datos a menudo vienen en formas desorganizadas y crudas. La limpieza y transformación son procesos esenciales para garantizar que los datos sean coherentes y útiles.
- **Análisis Descriptivo:** Aquí, se utilizan estadísticas y visualizaciones para resumir y describir características clave de los datos, como media, mediana, desviación estándar, histogramas, etc.
- **Análisis Exploratorio:** Se busca descubrir patrones y relaciones más profundas mediante técnicas visuales y estadísticas más avanzadas.
- **Modelado Predictivo:** Utilizando algoritmos de aprendizaje automático, se crean modelos que pueden predecir futuros eventos o resultados en función de datos históricos.
- **Toma de Decisiones Informada:** Los resultados del análisis de datos sirven como base para la toma de decisiones informadas, lo que puede impulsar la estrategia empresarial, la eficiencia operativa y la innovación.



Un **Ingeniero de Datos** es un profesional especializado en diseñar, desarrollar y mantener sistemas y arquitecturas que permiten la gestión efectiva de grandes volúmenes de datos. Este rol es esencial en el contexto actual, donde la cantidad de información generada por empresas, usuarios y dispositivos ha alcanzado proporciones colosales. La función del Ingeniero de Datos abarca diversas áreas cruciales para garantizar que los datos sean manejados de manera eficiente, segura y significativa para las organizaciones.

La ingeniería de datos también implica la transformación de datos brutos en información significativa. El ingeniero de datos se enfrenta al desafío de limpiar y procesar datos de manera que sean coherentes, precisos y útiles. Este proceso, conocido como ETL (extracción, transformación y carga), no solo requiere habilidades técnicas avanzadas, sino también un profundo conocimiento del dominio en el que opera la empresa. La capacidad para entender la naturaleza y el contexto de los datos es esencial para garantizar que la información resultante sea confiable y aplicable.

El Ingeniero de Datos es un arquitecto fundamental en la construcción y mantenimiento de la infraestructura que permite a las organizaciones aprovechar al máximo el potencial de sus datos. Su papel va más allá de lo técnico, involucrando una comprensión profunda de los procesos de negocio y la capacidad de traducir esos requisitos en soluciones tecnológicas eficientes.

Habilidades

- La ingeniería de datos también implica la transformación de datos brutos en información significativa. El ingeniero de datos se enfrenta al desafío de limpiar y procesar datos de manera que sean coherentes, precisos y útiles. Este proceso, conocido como ETL (extracción, transformación y carga), no solo requiere habilidades técnicas avanzadas, sino también un profundo conocimiento del dominio en el que opera la empresa. La capacidad para entender la naturaleza y el contexto de los datos es esencial para garantizar que la información resultante sea confiable y aplicable.
- Pipeline de datos, implementa procesos ETL (Extracción, Transformación y Carga) para garantizar que los datos se muevan de manera fluida desde diversas fuentes hasta su destino final. Esto implica la limpieza, transformación y procesamiento de datos para asegurar su calidad y coherencia.
- Habilidades avanzadas en programación, utilizando lenguajes como Python, Java, SQL, entre otros. Está familiarizado con el desarrollo de aplicaciones y scripts para automatizar procesos y manipular datos.
- Administrar bases de datos, asegurándose de que sean eficientes, seguras y estén bien estructuradas. Esto puede involucrar el uso de sistemas de gestión de bases de datos (DBMS) y la implementación de mejores prácticas en términos de rendimiento y seguridad.



- Implementar medidas de seguridad para proteger la integridad y confidencialidad de los datos. Se asegura de que se cumplan las regulaciones y normativas pertinentes en cuanto a privacidad y seguridad de la información.
- Optimizar el rendimiento de los sistemas de gestión de datos, asegurando tiempos de respuesta rápidos y eficiencia en el procesamiento de consultas.
- Estar dispuesto a aprender nuevas tecnologías y métodos para mantenerse actualizado.
- Trabajar en estrecha colaboración con científicos de datos, analistas de negocios y otros profesionales para comprender las necesidades de la organización y garantizar que los sistemas de datos sean compatibles con los objetivos empresariales.

El **científico de datos** es un profesional altamente capacitado y versátil cuya función principal es extraer conocimientos valiosos a partir de datos. Este experto multidisciplinario combina habilidades en programación, estadísticas y conocimientos de dominio específico para abordar problemas complejos y tomar decisiones informadas. La descripción de un científico de datos abarca una variedad de habilidades y responsabilidades que convergen para transformar datos brutos en información estratégica.

Un científico de datos debe poseer habilidades técnicas sólidas en programación y manipulación de datos. Lenguajes de programación como Python y R son fundamentales, junto con la capacidad de trabajar con bases de datos y herramientas especializadas. La experiencia en el manejo de conjuntos de datos masivos y técnicas de procesamiento distribuido también es esencial.

El conocimiento profundo de estadísticas y técnicas de modelado es una piedra angular en el arsenal de un científico de datos. La capacidad para aplicar algoritmos de aprendizaje automático, entender la validación de modelos y realizar análisis predictivos y descriptivos distingue a estos profesionales en su capacidad para convertir datos en información significativa.

La figura del científico de datos encarna la fusión de habilidades técnicas, conocimientos de dominio y una mentalidad analítica. Este profesional desempeña un papel crucial en la sociedad contemporánea, guiando a organizaciones y gobiernos a través del océano de datos hacia un futuro más informado y eficiente.



Habilidades

Conocimientos de Dominio: Comprender el contexto en el que se generan los datos es crucial. Ya sea en medicina, finanzas, marketing o cualquier otro campo, el científico de datos debe ser capaz de traducir problemas específicos del dominio en preguntas analíticas y resultados tangibles.

Comunicación Efectiva: La habilidad para comunicar hallazgos de manera clara y comprensible es un aspecto subestimado pero crítico del trabajo del científico de datos. Desde presentaciones a ejecutivos hasta la explicación de resultados a equipos no técnicos, la comunicación efectiva es clave para asegurar que los insights sean comprendidos y utilizados.

Exploración y Descubrimiento: El científico de datos es un explorador de datos por naturaleza. Debe ser curioso y capaz de identificar patrones no evidentes, formular preguntas y diseñar experimentos analíticos para descubrir relaciones valiosas entre variables.

Ética y Responsabilidad: El manejo ético de los datos es una preocupación central. El científico de datos debe estar comprometido con prácticas éticas, asegurando la privacidad y la seguridad de la información que maneja y considerando las implicaciones éticas de sus análisis.

Impacto en la Toma de Decisiones: La esencia del trabajo del científico de datos es influir en la toma de decisiones. Ya sea optimizando procesos empresariales, personalizando experiencias del cliente o mejorando la eficiencia operativa, el científico de datos juega un papel crucial en la toma de decisiones basada en datos.

Actividades y responsabilidades asociadas con Data Analysis, Data Analytics, Data Engineering y Data Science

Data Analysis

Análisis de Ventas:

- Utilizar datos de ventas para identificar patrones y tendencias.
- Crear informes para ayudar en la toma de decisiones estratégicas.

Encuestas y Retroalimentación del Cliente:

- Analizar datos de encuestas y comentarios de clientes para obtener información valiosa.
- Identificar áreas de mejora en productos o servicios.

Seguimiento del Rendimiento del Producto:

- Evaluar el rendimiento de productos utilizando datos de usuarios y métricas clave.
- Proporcionar recomendaciones para mejorar la experiencia del usuario.



Actividades y responsabilidades asociadas con Data Analysis, Data Analytics, Data Engineering y Data Science

Data Analytics

Segmentación de Clientes:

- Utilizar técnicas de análisis para identificar segmentos de clientes.
- Personalizar estrategias de marketing según los perfiles de los clientes.

Predicción de Ventas:

- Aplicar modelos predictivos para estimar las ventas futuras.
- Evaluar la eficacia de las estrategias de ventas actuales.

Análisis de Redes Sociales:

- Analizar datos de redes sociales para comprender la percepción del público.
- Identificar oportunidades de participación y mejorar la presencia en línea.

Actividades y responsabilidades asociadas con Data Analysis, Data Analytics, Data Engineering y Data Science

Data Engineering:

Diseño de Arquitectura de Datos:

- Desarrollar arquitecturas de datos escalables y eficientes.
- Garantizar la integridad y disponibilidad de los datos.

Desarrollo de Pipelines de Datos:

- Crear pipelines para la ingestión, transformación y carga (ETL) de datos.
- Automatizar procesos para mejorar la eficiencia operativa.

Gestión de Bases de Datos:

- Administrar bases de datos, asegurando un rendimiento óptimo.
- Implementar medidas de seguridad y privacidad de datos.



Actividades y responsabilidades asociadas con Data Analysis, Data Analytics, Data Engineering y Data Science

Data Scientist

Desarrollo de Modelos Predictivos:

- Utilizar algoritmos de aprendizaje automático para predecir resultados.
- Evaluar y ajustar modelos para mejorar la precisión.

Análisis de Texto y Minería de Datos:

- Aplicar técnicas de procesamiento de lenguaje natural para analizar texto.
- Extraer información valiosa de grandes conjuntos de datos no estructurados.

Optimización de Decisiones Empresariales:

- Utilizar análisis avanzado para tomar decisiones estratégicas.
- Identificar oportunidades para mejorar la eficiencia operativa.

Relación de Data Analysis, Data Analytics, Data Engineer, Data Scientist

Estrategia	Enfoque	Responsabilidades
Data Analysis (Análisis de Datos)	Se centra en examinar datos para identificar patrones, tendencias y obtener información significativa.	Utiliza herramientas estadísticas y de visualización para analizar datos y hacer interpretaciones. Generalmente trabaja con conjuntos de datos históricos para realizar análisis retrospectivos y proporcionar información para la toma de decisiones.
Data Analytics (Análisis de Datos)	Similar al Data Analysis, pero más amplio. Incluye la interpretación de datos, la identificación de tendencias y patrones, y la aplicación de modelos predictivos.	Implica el uso de técnicas más avanzadas, como el análisis predictivo y el análisis prescriptivo. Busca respuestas a preguntas específicas y utiliza datos para informar y respaldar la toma de decisiones.
Data Engineer (Ingeniero de Datos)	Se centra en el diseño y la construcción de sistemas y arquitecturas para la recopilación, almacenamiento y procesamiento eficiente de datos.	Desarrolla y mantiene la infraestructura de datos, garantiza la calidad y la integridad de los datos, y facilita la transferencia de datos entre sistemas. Trabaja en la creación y mantenimiento de pipelines de datos.
Data Scientist (Científico de Datos)	Combina habilidades de programación, estadísticas y conocimientos de dominio para extraer conocimientos y modelos predictivos a partir de grandes conjuntos de datos.	Desarrolla algoritmos y modelos predictivos, utiliza técnicas de aprendizaje automático, y presenta soluciones a problemas complejos. También interpreta los resultados y proporciona información accionable.