

Data Analytics

Descripción del Evento

Este evento está diseñado para presentar a los asistentes el mundo del análisis de datos. Los asistentes aprenderán sobre los principales conceptos y técnicas de análisis de datos, así como cómo estos pueden ser aplicados en el contexto de negocio. Para ello, realizaremos un análisis de un conjunto de datos y construiremos un dashboard que nos permita mostrar los principales resultados.

Agenda del Evento

- Presentación Hackio
- Presentación de los Programas de Data
- Introducción al análisis de datos
- Recopilación y limpieza de datos
- Análisis estadístico
- Visualización de datos
- Interpretación de datos
- Preguntas y respuestas

Objetivos del Evento

- Proporcionar una introducción sólida al análisis de datos
 - Enseñar a los asistentes cómo recopilar, limpiar, analizar e interpretar datos
 - Demostrar cómo el análisis de datos puede aplicarse en un contexto empresarial
 - Inspirar a los asistentes a explorar más a fondo el mundo del análisis de datos
-

Introducción al análisis de datos

El análisis de datos es una práctica que consiste en la inspección, limpieza, transformación y modelado de datos con el objetivo de descubrir información útil, llegar a conclusiones y apoyar la toma de decisiones. En un mundo cada vez más digital, el análisis de datos se ha convertido en una parte crucial de muchos ámbitos, desde el empresarial hasta el académico, pasando por el gubernamental.

El análisis de datos puede involucrar una variedad de técnicas y métodos, desde estadísticas simples y gráficos de datos hasta algoritmos de aprendizaje automático más complejos. Sin embargo, independientemente del método utilizado, el objetivo final es el mismo: extraer significado de los datos para ayudar a tomar decisiones basadas en evidencia.

Una de las claves para un buen análisis de datos es la calidad de los datos en sí. Los datos deben ser precisos, completos y relevantes para la pregunta o problema que se está abordando. La limpieza y preparación de los datos es un paso crucial en este proceso, ya que incluso los datos más precisos pueden ser inútiles si están desordenados o mal formateados.

Una vez que los datos están limpios y preparados, pueden ser analizados para descubrir patrones, tendencias y relaciones. Esto a menudo implica el uso de técnicas estadísticas para resumir los datos y extraer información útil. Sin embargo, el análisis de datos no se trata solo de números y estadísticas. También se trata de interpretar los resultados y comunicarlos de manera efectiva a otros.

Un analista de datos es un profesional que recopila, procesa y realiza análisis estadísticos de datos. Su principal objetivo es ayudar a las empresas a tomar decisiones de negocio basándose en la información obtenida de los datos. Para ello, utiliza diversas herramientas y métodos para interpretar y analizar los datos, y luego presenta los resultados en un formato comprensible para los tomadores de decisiones. Las responsabilidades de un analista de datos pueden variar dependiendo de la industria y la empresa, pero generalmente incluyen la definición de problemas a resolver, la recopilación y limpieza de datos, el análisis de los datos y la interpretación y presentación de los resultados.

El flujo de trabajo de un analista de datos generalmente incluye los siguientes pasos:

1. **Definición del problema:** Identificar la pregunta o problema que se debe resolver.

2. **Recopilación de datos:** Recoger los datos necesarios para resolver el problema. Esto puede implicar extraer datos de bases de datos internas, recoger datos de fuentes externas, o ambas.
3. **Limpieza de datos:** Preparar los datos para el análisis. Esto puede implicar eliminar datos duplicados, llenar valores faltantes, corregir errores, etc.
4. **Análisis de datos:** Utilizar técnicas estadísticas y matemáticas para analizar los datos y extraer información relevante.
5. **Interpretación de los resultados:** Traducir los resultados del análisis en términos que sean comprensibles y útiles para los tomadores de decisiones.
6. **Presentación de los resultados:** Crear informes y visualizaciones de los resultados del análisis para presentarlos a los tomadores de decisiones.
7. **Aplicación de los resultados:** Utilizar los resultados del análisis para tomar decisiones y guiar estrategias.

Recopilación y Limpieza de Datos

Recopilación de Datos

La recopilación de datos es el primer paso en cualquier proyecto de análisis de datos. Este proceso implica identificar y recoger información relevante para el problema o pregunta que se está tratando de resolver. Los datos pueden provenir de una variedad de fuentes, incluyendo bases de datos internas, encuestas, entrevistas, registros públicos y más.

Es importante asegurarse de que los datos que se están recogiendo son relevantes para la pregunta de investigación y que se recojan de manera sistemática y consistente. Esto incluye definir claramente qué tipo de datos se necesita, cómo se va a recoger y cómo se va a almacenar.

Limpieza de Datos

Una vez que los datos han sido recogidos, el siguiente paso es prepararlos para el análisis. Este proceso, conocido como limpieza de datos, implica revisar y modificar los datos para asegurarse de que son precisos, completos y adecuados para el análisis.

La limpieza de datos puede implicar una variedad de tareas, incluyendo:

- **Lidiar con datos faltantes:** Esto puede implicar imputar valores faltantes basándose en otros datos disponibles, eliminar registros con datos faltantes o usar métodos estadísticos para estimar valores faltantes.
- **Corregir errores en los datos:** Esto puede implicar corregir errores tipográficos, estandarizar formatos de datos o validar la precisión de los datos contra fuentes confiables.
- **Eliminar duplicados:** Esto implica identificar y eliminar registros duplicados en los datos.
- **Transformar los datos:** Esto puede implicar convertir los datos a diferentes formatos o unidades, crear nuevas variables a partir de variables existentes o recodificar variables para el análisis.

La limpieza de datos es un paso crucial en el análisis de datos, ya que los datos de mala calidad pueden llevar a resultados incorrectos o engañosos. Por lo tanto, es importante dedicar tiempo y esfuerzo para asegurarse de que los datos estén limpios y listos para el análisis.

Análisis estadístico para análisis de datos

El análisis estadístico es una parte fundamental de la ciencia de datos y es esencial para entender mejor los datos que estás trabajando. A través de las estadísticas, se pueden obtener conclusiones significativas a partir de los datos y tomar decisiones informadas.

Conceptos básicos de estadísticas

- **Población y muestra:** Una población es el conjunto total de observaciones que nos interesan. Una muestra es un subconjunto de esa población. En general, es práctico trabajar con muestras cuando la población es demasiado grande.
- **Variables:** Las variables son las características que se miden. Pueden ser cualitativas (categóricas) o cuantitativas (numéricas).
- **Medidas de tendencia central:** El promedio, la mediana y la moda son todas medidas de tendencia central que indican el valor "típico" en un conjunto de datos.

- **Medidas de dispersión:** La varianza, la desviación estándar y el rango son medidas de dispersión que indican cuánto varían los datos alrededor de la media.

Técnicas de análisis estadístico

- **Análisis descriptivo:** Este es el primer paso en cualquier análisis de datos. Incluye la generación de estadísticas descriptivas (como la media, la mediana y la moda), la creación de tablas de frecuencia y la visualización de los datos mediante gráficos y diagramas.
- **Pruebas de hipótesis:** Las pruebas de hipótesis son útiles para comparar dos o más grupos y determinar si las diferencias observadas en los datos son significativas o se deben simplemente al azar.
- **Correlación y regresión:** La correlación mide la relación entre dos variables. La regresión, por otro lado, permite predecir una variable a partir de otra.
- **Análisis multivariante:** Este tipo de análisis estadístico se utiliza cuando se trabaja con tres o más variables. Permite entender las relaciones entre varias variables a la vez.

Al final de esta lección, deberías tener una comprensión sólida de los conceptos básicos de las estadísticas y cómo se aplican al análisis de datos. Recuerda, la clave para un buen análisis de datos es entender tus datos y las técnicas estadísticas que estás utilizando.

Visualización de Datos para Analistas de Datos

La visualización de datos es una parte esencial del análisis de datos. Permite presentar los datos de manera gráfica para facilitar su comprensión y extracción de conclusiones. Además, puede destacar tendencias y patrones que podrían pasar desapercibidos en los datos brutos.

La visualización de datos es el proceso de convertir datos en gráficos, gráficos y otros elementos visuales. El propósito principal de la visualización de datos es facilitar la comprensión de los datos, permitiendo a los usuarios ver los datos de manera más intuitiva y clara.

¿Por qué es Importante la Visualización de Datos?

- **Comprensión rápida de la información:** Un gráfico puede transmitir información de manera eficiente y rápida, permitiendo a los usuarios entender los datos en un corto período de tiempo.
- **Identificación de patrones y tendencias:** Los gráficos y diagramas pueden ayudar a identificar patrones, tendencias y correlaciones en los datos que de otro modo podrían pasar desapercibidas.
- **Toma de decisiones basada en datos:** La visualización de datos permite a los tomadores de decisiones ver y entender los datos de manera más efectiva, facilitando la toma de decisiones basada en datos.

Técnicas de Visualización de Datos

- **Gráficos de barras:** Son útiles para comparar datos categóricos o para mostrar cambios en los datos a lo largo del tiempo.
- **Gráficos de líneas:** Son ideales para mostrar tendencias a lo largo del tiempo.
- **Histogramas:** Son útiles para mostrar la distribución de los datos.
- **Diagramas de dispersión:** Son ideales para mostrar la relación entre dos variables numéricas.
- **Mapas de calor:** Son útiles para visualizar matrices de datos y encontrar patrones de correlación.

Un buen recurso para decidir que tipo de gráfica podemos usar estas páginas webs como recursos:

- [Data to Viz](#)
- [Data Viz](#)

Herramientas de Visualización de Datos

- **Excel:** Ofrece una variedad de opciones de gráficos y es una buena opción para visualizaciones simples.
- **Tableau:** Es una herramienta de visualización de datos muy potente que permite crear gráficos interactivos y dashboards.

- **Python (matplotlib, seaborn, plotly):** Estas bibliotecas de Python ofrecen amplias posibilidades para la visualización de datos.

Interpretación de Datos y Storytelling

Interpretación de Datos

La interpretación de datos es un paso crucial en el proceso de análisis de datos. Una vez que hemos recopilado y analizado nuestros datos, necesitamos interpretar los resultados para entender qué nos están diciendo.

Interpretar los datos implica examinar los resultados de nuestro análisis y tratar de entender qué significan en el contexto de nuestra pregunta de investigación o problema empresarial. Esto puede involucrar identificar patrones o tendencias, hacer comparaciones, o sacar conclusiones sobre las relaciones entre variables.

Aunque las técnicas específicas de interpretación de datos pueden variar dependiendo del tipo de análisis que se esté realizando, hay algunas estrategias generales que pueden ser útiles:

- **Relacionar los resultados con la pregunta de investigación o problema empresarial:** ¿Cómo responden los resultados a la pregunta que estábamos tratando de responder? ¿Ofrecen alguna solución al problema que estábamos tratando de resolver?
- **Considerar el contexto:** ¿Cómo se relacionan los resultados con lo que ya sabemos sobre el tema? ¿Hay factores externos que podrían estar influyendo en los resultados?
- **Examinar los datos desde diferentes ángulos:** ¿Cambian los resultados si miramos los datos de una manera diferente? ¿Hay subgrupos o segmentos de datos que muestran patrones diferentes?
- **Buscar explicaciones para los resultados inesperados:** Si los resultados son sorprendentes o no son lo que esperábamos, ¿podemos pensar en alguna razón por la que esto podría ser así? ¿Necesitamos recopilar más datos o realizar un análisis adicional?

Storytelling con Datos

Una vez que hemos interpretado nuestros datos, el siguiente paso es comunicar nuestros hallazgos a otros. Una forma efectiva de hacerlo es a través del storytelling con datos.

El storytelling con datos implica usar técnicas narrativas para presentar nuestros resultados de una manera que sea interesante y fácil de entender. En lugar de simplemente presentar una serie de hechos o cifras, intentamos contar una historia con nuestros datos que capte la atención de nuestra audiencia y la ayude a entender lo que los datos significan.

Aquí hay algunas estrategias que pueden ser útiles para el storytelling con datos:

- **Comenzar con un gancho:** Inicia tu historia con una afirmación o pregunta interesante que capte la atención de tu audiencia y la haga querer saber más.
- **Usar visualizaciones de datos:** Los gráficos y diagramas pueden ser una forma efectiva de mostrar tus resultados y hacer que tu historia sea más atractiva.
- **Hacer que tu historia sea relevante para tu audiencia:** Intenta conectar tus resultados con algo que sea importante o significativo para tu audiencia.
- **Usar el lenguaje sencillo y claro:** Evita el jargon técnico y explica los conceptos complejos en términos que tu audiencia pueda entender.
- **Terminar con un llamado a la acción:** ¿Qué quieres que haga tu audiencia con la información que has presentado? ¿Hay algún paso que puedan tomar o alguna decisión que puedan hacer basada en tus resultados?

Recordar que el objetivo del storytelling con datos no es solo informar, sino también inspirar y motivar a la acción. Al contar una historia efectiva con nuestros datos, podemos ayudar a nuestra audiencia a entender no solo lo que los datos dicen, sino también lo que significan y por qué son importantes.

Recursos

Para la sesión usaremos un conjunto de datos que presentan un registro detallado sobre la logística de distribución de productos desde un origen específico (generalmente un almacén) hasta diversos destinos (tiendas en diferentes ubicaciones). Cada fila del excel representa una transacción de distribución de un producto particular. Las columnas que tenemos son:

1. **Id.:** Cada registro tiene un identificador único que permite diferenciarlo de los demás. Esto es crucial para llevar un seguimiento preciso y evitar duplicidades.
2. **Cantidad:** Esta columna indica la cantidad de unidades del producto distribuido en cada transacción. Es importante para medir el volumen de distribución y gestionar inventarios.
3. **Fecha_Entrega:** La fecha específica en la que se realizó la entrega del producto. Este dato es esencial para el seguimiento de tiempos de entrega y la planificación logística.
4. **Año:** Año en el que se realizó la entrega. Aunque se puede derivar de la columna "Fecha_Entrega", tenerlo como una columna separada facilita análisis por año.
5. **Estado_Revisión:** Indica el estado de revisión de la transacción, por ejemplo, si está "Finalizado". Ayuda a entender el progreso y la finalización de los registros de distribución.
6. **Origen:** Especifica el punto de origen desde donde se envían los productos, usualmente un almacén. Esta información es útil para la gestión de inventarios y optimización de rutas de distribución.
7. **Id.Destino:** Un identificador único para cada destino, que ayuda a identificar a dónde se envían los productos.
8. **Descripción_Destino:** Descripción más detallada del destino, generalmente una ciudad o una región. Ayuda a comprender la distribución geográfica de los envíos.
9. **Id.Tienda:** Identificador específico de la tienda de destino. Este detalle permite rastrear las entregas hasta un punto de venta específico.
10. **COD. Producto2:** Código del producto que está siendo distribuido. Este identificador único facilita la gestión y seguimiento de inventarios de diferentes productos.
11. **Descripción_Producto:** Proporciona una descripción del producto (por ejemplo, tijeras, ligas). Es útil para entender qué productos están siendo distribuidos sin necesidad de consultar códigos adicionales.
12. **COD.Vehículo:** Código del vehículo utilizado para realizar la entrega. Es importante para la logística y la planificación de rutas, así como para el mantenimiento de los vehículos.

13. **Costos:** Los costos asociados con la distribución de cada transacción. Esto puede incluir costos de transporte, manejo, etc., y es crucial para análisis financieros y de eficiencia.

[Panel de Distribución de Productos.xlsx](#)

El dashboard creado durante la sesión lo encontrareis en [este](#) link