

Análisis exploratorio y visualización de datos 01

...

20 de Noviembre, 2021



Francisco J. Palm R.
Científico de Datos
@mapologo

Contenido

Conociendo tus datos

1. Definición del problema
2. Metadatos
3. Carga de Datos
4. Datos ordenados
5. Formatos y codificación
6. Tipos de datos

Estadísticas Descriptivas

1. Estadísticas univariantes
2. Estadísticas multivariantes

Limpieza de datos

1. Valores faltantes
2. Valores atípicos

Conociendo tus datos

Los datos que utilicemos (o que nos sean entregados) en la solución de un problema pueden darnos una visión más clara del problema que vamos a abordar (y pueden mostrar otros problemas).

¡Los datos acompañan la narrativa del problema y también ayudan a construir una narrativa de la solución!

Definición del problema:

Es, básicamente, una observación sobre los componentes de una situación que depende de quien observa, cómo observa y desde donde lo hace.



Preguntas clave sobre el problema

Al buscar la solución a un problema, cambiamos repetidamente de punto de vista y de la forma en que miramos el problema.

Pólya nos dice que hay cuatro fases al abordar un problema:

- **Entender** el problema.
 - **Planificar** el abordaje.
 - **Ejecutar** el plan diseñado.
 - **Revisar** la solución obtenida.
-

[illegible]

Metadatos

Son datos que *hablan* de otros datos.

- Son *altamente estructurados* y describen características de otros datos: contenido, calidad, información y otros atributos.
 - Tienen *diferencias entre sí*, según las reglas establecidas para determinar la estructura interna de los esquemas de los datos que describen.
 - Pueden *clasificarse* según distintos criterios: contenido, variabilidad o función.
-

Ventajas

- Facilitan búsqueda y análisis.
- Permiten estandarizar.
- Contribuyen a la integración de estructuras de datos.
- Protege los datos críticos.
- Soporta el desarrollo ágil de proyectos.
- Facilita la gobernanza de datos..



Intermedio de 10 mins

Carga de datos

Formatos y Codificación

Tipos de datos



Source: glazestock.com
Illustrated by Rudityas

Estadísticas Descriptivas

Estadística Descriptiva

Es una de las dos ramas de la estadística que, como disciplina ayuda a recoger, almacenar, ordenar, visualizar y calcular parámetros básicos sobre un conjunto de datos.

Parámetros de estadística descriptiva

- Medidas de tendencia central
- Medidas de dispersión
- Medidas de posición
- Frecuencia absoluta
- Frecuencia relativa
- Frecuencia acumulada

Tablas y Gráficos en estadística descriptiva

- Histograma.
- Gráfico de barras.
- Gráfico de sectores.
- Tablas de probabilidad.
- Tablas bidimensionales.
- Gráfico de cajas.

Estadísticas Univariantes

Estas estadísticas se enfocan en el análisis de una única característica o cualidad observada.

Cada característica (variable) analizada presenta k modalidades, exhaustivas y mutuamente excluyentes.

Las variables pueden ser

- Cualitativas (ordinales/nominales)
- Cuantitativas (pocos/muchos valores)

Estadísticas Multivariantes

Como rama de la estadística, busca observar y analizar de forma simultánea más de una variable respuesta.

También se le conoce con el nombre de análisis multivariante y comprende varios métodos:

- Métodos de dependencia (regresión, correlación, multivarianza, regresión logística ...)
 - Métodos de interdependencia (cluster, análisis discriminante ...)
 - Métodos estructurales (ecuaciones estructurales ...)
-

Limpieza de Datos



Limpieza de datos

Al recibir o capturar los datos, pueden no estar dispuestos de forma idónea para su análisis, manipulación o visualización.

Por eso la limpieza de datos es una **tarea clave** en la solución de cualquier problema.

Los datos limpios son aquellos que han sido depurados y formateados según un conjunto de reglas establecidas por quien los analiza o por las necesidades de análisis del problema planteado.

¡Los datos son limpiados para mejorar su análisis e interpretación!

Limpieza de datos: tres tareas clave.

1) Inspección de los datos

Busca detectar datos incompletos, duplicados, incorrectos o inconsistentes.

2) Depuración de los datos

Es útil para ajustar o remover las anomalías detectadas en la inspección.

3) Verificación de los datos

Tarea clave para validar que los datos sean correctos una vez hayan sido depurados.

Valores Faltantes

Valores Atípicos

Ahora, una breve pregunta ...

Opción 1: Utiliza este enlace para responder una pregunta

<https://www.menti.com/gdvwzpd1vq>

Opción 2: entra en menti.com y utiliza el siguiente código:

3214 6021

Opción 3: Escanea este Qr



Saludos

Francisco J. Palm R.
Científico de Datos
francisco.p@reworth.co
@mapologo

