

Herramientas y Roles de Trabajo (Data Science)

Flujo de Trabajo en Ciencia de Datos

Fases, Roles & Oportunidades Laborales

- Ingeniero/arquitecto de Datos

Origen de datos

Este es la conexión entre **dispositivo y creación de DB** trabajan en la creación de API's, SQL y NoSQL

Maximo ambito o rol principal en **bases de datos**

El ingeniero tiene mayor conocimientos en ingeniería

- Analista

Analista de Datos

Apartir de la información que creo el **ingeniero de datos** y extraerla, va a "extraer" la información importante, creación de dashboards o gráficas, para posterior automatizar

Excel y SQL

Este tiene vista al presente de la empresa

- Data Scientist

Científico de Datos, esta combinación entre extraer, sabe "limpiar" la información y aparte predecir esta, el analista tiene vista a futuro de la empresa

R y python, Tambien SQL

El analista y el cientifico de datos tienen mayor conocimientos en matematica

- Data Translator (Gefe o Dirige los anteriores)

Es muy bueno en alguno de los roles de anterioridad, o una persona que su fuerte son los negocios pero que tambien sabe de datos pero no necesariamente una persona tecnica, comunica e interpreta los datos a otras personas de otras areas

Herramientas para cada etapa

Herramientas a lo largo de un analisis y características

- Rol del analista e Ingeniero (Extraccion de Informacion con SQL)

Personas que crean las bases de datos, Lenguaje SQL

- Extraccion de informacion
- Ayuda a encontrar la informacion para un estudio
- Para posterior entregar un **cuadro de control/grafico**

- Rol de cientifico de datos (Analisis y visualizacion)

Estas herramientas nos ayudan a realizar graficos (R y Python)

- R Es mas orientado al analisis estadisticos
- Simplifica el trabajo
- Analisis descriptivo y exploratorio

- **Analisis descriptivo** describir o interpretar la informacion que tenemos
- **Analisis exploratorio** Sirve para predecir a un futuro
- Packages: ggplot2 y dplyr
- Python mas orientado a Ingenieria, base ingenieria
- Este es mas para personas que saben programar, sintaxis de lenguaje facil
- Analisis descriptivo y exploratorio
 - **Analisis descriptivo** describir o interpretar la informacion que tenemos
 - **Analisis exploratorio** Sirve para predecir a un futuro
- Librerias: Pandas y Numpy

Que es y como usar una base de daots relacional con SQL

Este nos ayuda para aprender que son las bases de datos DB

Las bases de datos/tablas al momento de ya estar almacenadas tienen dos partes **columnas y Filas**

Columna = Van para abajo

Fila = Van para el lado

Comandos

A momento de nosotros queremos seleccionar una **columna** usamos

```
SELECT fecha
SELECT fecha, cliente
```

Indicar donde esta almacenada la informacion

```
FROM /* Tabla donde se almacena la informacion */
/* Ejemplo */
SELECT fecha, client FROM ventas

/* Condiciones para tener mejor la informacion */
WHERE /* Especificar condicion, por ejmplo quiere solo ventas del
2010 */

WHERE fecha = 2010

GROUP BY /* Agrupar, condicion agrupada */

/* Saber la informacion de venta de cada cliente */
GROUP BY cliente /* Agrupa por columnas */
ORDER BY /* Ordenar resultados, ventas de enero a diciembre */

/* OPERADORES LOGICOS */

AND /* Une varias condiciones, para que esta condicion sea
cumplida tiene que hacer las dos condiciones */
WHERE fecha = 2020 AND cliente = Jose

OR /* Este sirve para que una u otra condicion sea cumplida */
/* Podemos decir que queremos las ventas del 2010 o 2011 */
WHERE fecha = 2010 OR fecha = 2011

NOT /* NO incluyas.. */
NOT IN 2020 /* No incluir ventas del 2020 */
```

```
/* Funciones agregadas */  
/* Estos son para no tener que hacer sumas o promedios en  
grandes escalas */  
  
AVG /* Sacar un promedio */  
COUNT /* Hacer un recuento de x columna o tabla */  
DISTINCT /* Lleva a usuarios unicos */  
COUNT + DISTINCT  
  
SUM /* REALIZAR una suma de alguna columna o de un año */  
MAX /* Valores maximos o mejores o destacados */  
MIN /* Lo contrario, para ver el valor minimo o peor */
```

Como estructurar "queries" SQL

Extraccion de informacion de una tabla SQL

Ejemplo de SQL

Objetivo

Saber CUÁNTAS bocinas hemos vendido por más de 600 MXN desde 2019

Tabla en Excel

VENTAS_2020

Día	Mes	Año	Producto	Id	Valor
1	2	1998	Bocina	24	\$528
12	4	2004	Auriculares	31	\$240
14	8	2016	Auriculares	14	\$315
16	10	2019	Bocina	200	\$1,050
21	12	2020	Bocina	304	\$680

Comprensión

TABLA: VENTAS_2020

COLUMNAS

Día (1-31)

Mes (1-12)

Año (1990-2020)

Producto: bocinas y auriculares

Valor (MXN)

Código SQL

```
SELECT COUNT (DISTINCT id)
FROM VENTAS_2020
WHERE Producto = 'Bocina'
AND Valor > 600
AND Año >= 2019
```

Resultado

2

1. Comprobamos si la informacion de las tablas es real o si es veridica
2. Verificar informacion como creacion/fundacion de la empresa

Explicacion del Codigo SQL

```
/* La primera linea nos dice que seleccionamos la columna ID*/
SELECT COUNT (DISTINCT id)
/* FROM hace referencia el nombre de una tabla */
FROM VENTAS_2020
/* Con WHERE especificamos donde esta el producto que queremos y
comenzamos con algunas condicionales como la fila especifica */
WHERE Producto = "bocina"
/* Comenzamos con las condicionales, que el valor en pesos sea
mayor a 600 */
AND Valor > 600
/* El año de venta sea IGUAL o MAYOR a 2019 */
AND Año >= 2019
```

```
/* Esto da como resultado R// 2 */
```

Conflictos y retos actuales sobre la etica en relevancia a los datos

Reflexion sobre informacion sensible y discernir entre informacion confidencial y no confidencial

El peligro de usar aplicaciones como **FaceApp** nosotros brindamos datos de vectores de nuestra cara, ubicacion de puntos clave en la cara o incluso el clonar nuestra cara o copiarla

A pequeñas escalas no suena tan peligroso pero el algoritmo o su forma de funcionar es gracias a otras personas como/similares a nosotros tambien subieron fotos de sus caras y estas las almaceno y guardo

Con esto saldrán las siguientes preguntas

- Me gusta brindar informacion personal a este tipo de empresas?
- Me gusta sentirme vulnerable o inferior al saber que alguien externo podria tener acceso a estos datos?