



LEVELS H&H



TRAYECTORIA

ENGINEERING

Procesos de limpieza y filtrado, creación de Data Warehouse y automatización de procesos.

MACHINE LEARNING

Uso de las herramientas de AI en la plataforma GCP.
Determinar features y posibles modelos para procesar la información.

ANALYTICS

Primeras propuestas de visualización de los datos proporcionados.



ETL



01

02

03

EXTRACCIÓN

- Extracción de datasets de Google Maps.

TRANSFORMACIÓN

- Unificación de archivos.
- Filtro de restaurantes
- Desanidado de columnas.

CARGA

- Google BigQuery



TRANSFORMACIONES

UNIFICACIÓN DE ARCHIVOS

{j s o n}



FILTRADO DE LA DATA



DESANIDADO DE COLUMNAS

MISC:

Filtro por presencia.

Categories:

Generalización.

ENGINEERING

GOOGLE CLOUD PLATFORM

AUTOMATIZACIÓN Y CARGA INCREMENTAL



CLOUD STORAGE



COLUD
FUNCTIONS



BIG QUERY





CLOUD STORAGE

ALMACENAMIENTO

CARGA DE DATOS SIN PROCESAR

etl_metadata_sitios

Ubicación

us-central1 (Iowa)

Clase de almacenamiento

Standard

Acceso público

No público

Protección

Ninguna

OBJETOS

CONFIGURACIÓN

PERMISOS

PROTECCIÓN

CICLO DE VIDA

OBSERVABILIDAD

INFORMES DE INVENTARIO

Depósitos > etl_metadata_sitios

SUBIR ARCHIVOS

SUBIR CARPETA

CREAR CARPETA

TRANSFERIR LOS DATOS

ADMINISTRAR CONSERVACIONES

EDITAR LA RETENCIÓN

DESCARGAR

BORRAR

Filtrar solo por prefijo de nombre

Filtro Filtrar objetos y carpetas

Mostrar datos borrados

<input type="checkbox"/>	Nombre	Tamaño	Tipo	Fecha de creación	Clase de almacenamiento	Última modificación	Acceso público	Historial de ve	
<input type="checkbox"/>	1.json	244.3 MB	application/json	8 feb 2024 12:13:09	Standard	8 feb 2024 12:13:09	No público	—	
<input type="checkbox"/>	10.json	268.8 MB	application/json	8 feb 2024 12:21:07	Standard	8 feb 2024 12:21:07	No público	—	
<input type="checkbox"/>	11.json	274.2 MB	application/json	8 feb 2024 12:21:12	Standard	8 feb 2024 12:21:12	No público	—	
<input type="checkbox"/>	2.json	245.5 MB	application/json	8 feb 2024 12:13:17	Standard	8 feb 2024 12:13:17	No público	—	
<input type="checkbox"/>	3.json	247.3 MB	application/json	8 feb 2024 12:13:07	Standard	8 feb 2024 12:13:07	No público	—	
<input type="checkbox"/>	4.json	249.9 MB	application/json	8 feb 2024 12:20:00	Standard	8 feb 2024 12:20:00	No público	—	
<input type="checkbox"/>	5.json	252.1 MB	application/json	8 feb 2024 12:20:01	Standard	8 feb 2024 12:20:01	No público	—	
<input type="checkbox"/>	6.json	254.7 MB	application/json	8 feb 2024 12:14:05	Standard	8 feb 2024 12:14:05	No público	—	
<input type="checkbox"/>	7.json	260.3 MB	application/json	8 feb 2024 12:14:40	Standard	8 feb 2024 12:14:40	No público	—	
<input type="checkbox"/>	8.json	271 MB	application/json	8 feb 2024 12:15:28	Standard	8 feb 2024 12:15:28	No público	—	
<input type="checkbox"/>	9.json	264.2 MB	application/json	8 feb 2024 12:22:15	Standard	8 feb 2024 12:22:15	No público	—	

Filas por página:

50

1 - 11 de 11

<

>





COLUD
FUNCTIONS

AUTOMATIZACIÓN



CODIGO OPTIMIZADO

```
@functions_framework.cloud_event
def etl_metadata(cloud_event):
    # Obtener información del evento
    data = cloud_event.data
    bucket_name = data['bucket']
    file_name = data['name']

    # Inicializar el cliente de Storage
    storage_client = storage.Client()
    bucket = storage_client.bucket(bucket_name)
    blob = bucket.blob(file_name)

    # Verificar si el objeto es un archivo JSON
    if blob.name.endswith(".json"):
        # Descargar el archivo JSON como objeto de bytes
        blob_data = blob.download_as_string()
        # Leer datos JSON desde el objeto de bytes
        df = pd.read_json(io.BytesIO(blob_data), lines=True)

        # Eliminar columnas innecesarias
        df = df.drop(columns=['description', 'num_of_reviews', 'hours', 'state'],
                        inplace=True)

        # Modificar columnas a tipo string
        df['category'] = df['category'].astype('string')

        # Filtramos los registros que contienen la categoría "restaurant":
        df = df[df['category'].str.contains("restaurant", case=False)].reset_index(drop=True)
```

CARGA INCREMENTAL

MARCA DE TIEMPO	RESUMEN
2024-02-09 07:07:52.400 ART	POST 200 656 B 22.8 s APIs-Google; (+https://developers.g... https://functi
2024-02-09 07:07:53.957 ART	POST 200 656 B 26.2 s APIs-Google; (+https://developers.g... https://functi
2024-02-09 07:07:56.131 ART	POST 200 656 B 21.1 s APIs-Google; (+https://developers.g... https://functi
2024-02-09 07:08:15.190 ART	Datos concatenados y cargados en BigQuery: restaurants.metadata_restaurants
2024-02-09 07:08:17.188 ART	Datos concatenados y cargados en BigQuery: restaurants.metadata_restaurants
2024-02-09 07:08:20.162 ART	Datos concatenados y cargados en BigQuery: restaurants.metadata_restaurants
2024-02-09 09:52:38.494 ART	POST 200 714 B 29.7 s APIs-Google; (+https://developers.g... https://functi
2024-02-09 09:52:40.732 ART	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on
2024-02-09 09:52:49.674 ART	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on
2024-02-09 09:52:52.670 ART	Default STARTUP TCP probe succeeded after 1 attempt for container "worker" on
2024-02-09 09:53:08.395 ART	Datos concatenados y cargados en BigQuery: restaurants.metadata_restaurants



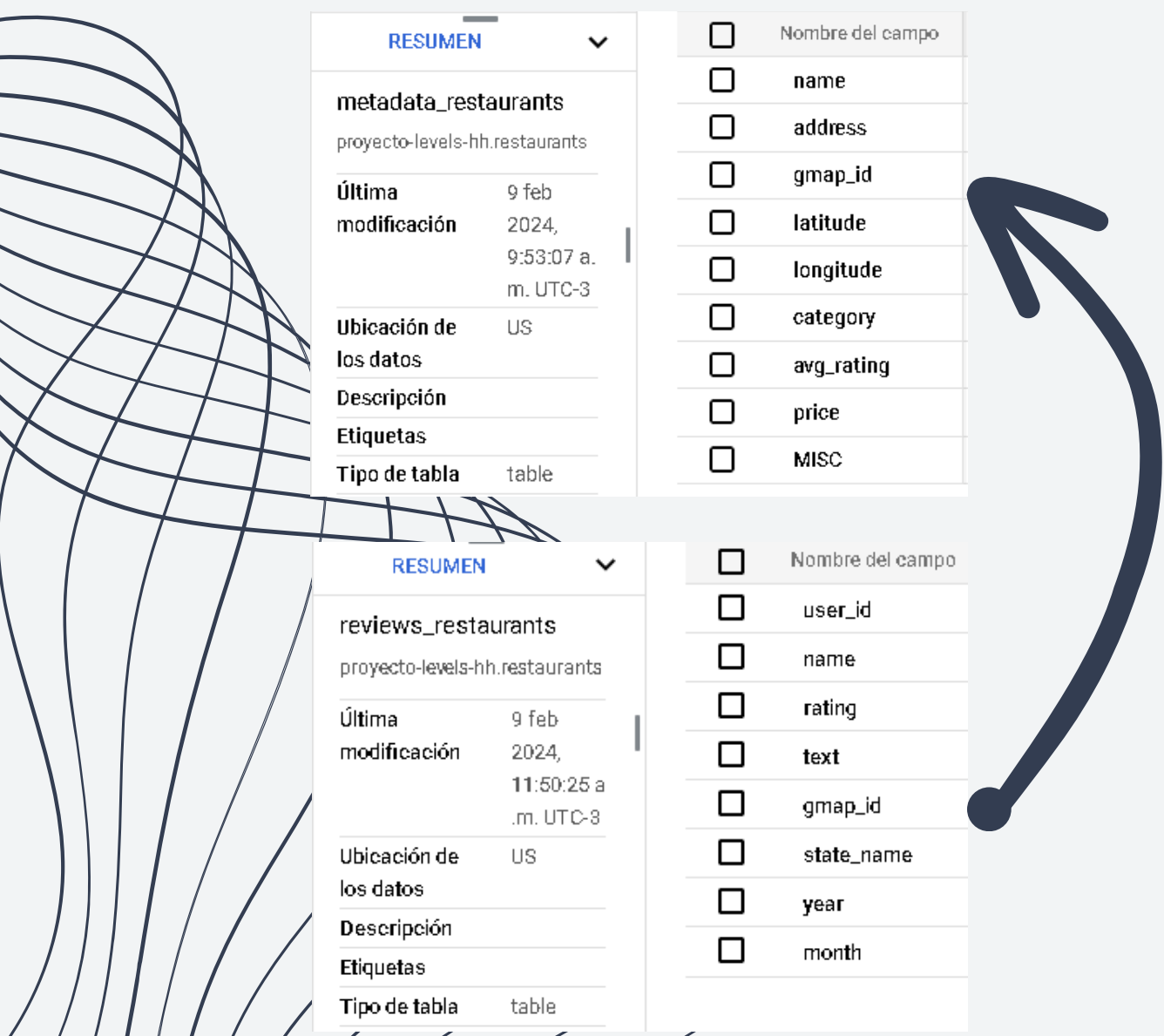
BIG QUERY

BASE DE DATOS



MODELO ENTIDAD-RELACIÓN

MUESTRA DE DATOS DE RESEÑAS



reviews_resta...						
<div>CONSULTA COMPARTIR COPIAR INSTANTÁNEA BORRAR EXPORTAR</div>						
ESQUEMA		DETALLES	PREVIEW	LINAJE	PERFIL DE DATOS	CALIDAD DE LOS DATOS
Fila	user_id	name	rating	text	gmap_id	state_name
1	1.03899886...	Brad G	5	For those of you seeking out fuel that's not E10, as of 1/3/10 this place has it.	0x88626f53d1eaa553:0x9daf...	Alabama
2	1.18341119...	Matt Spivey	4	null	0x88621354b99f3e39:0xbd5d...	Alabama
3	1.05359917...	Brant Powell	5	null	0x88626b57fd53d359:0xce05...	Alabama
4	1.04280646...	Joao Natali	2	null	0x889a4d74b15d2553:0x4cb...	Alabama
5	1.16677325...	carole clark	5	I was so nervous and asked him to just talk while he worked and he talked about a book he was reading and it made the time go faster and I	0x888e81fe47ebcbeb:0x826b...	Alabama

ANALYTICS



ANALISIS DE RATINGS

PANEL DE CONTROL

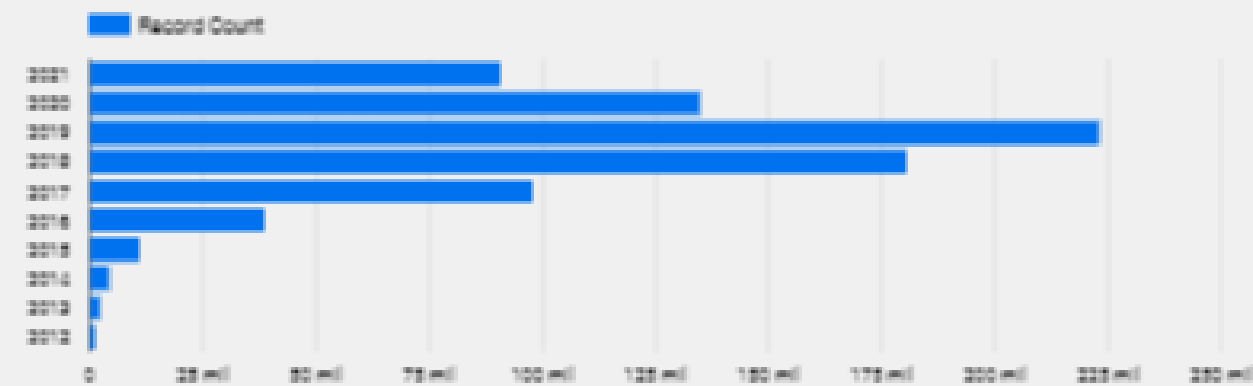
year *

<input checked="" type="checkbox"/> month	year
<input checked="" type="checkbox"/> marzo	1289
<input checked="" type="checkbox"/> enero	1183
<input checked="" type="checkbox"/> febrero	1184
<input checked="" type="checkbox"/> junio	1174
<input checked="" type="checkbox"/> mayo	1183
<input checked="" type="checkbox"/> julio	1182
<input checked="" type="checkbox"/> agosto	1123

state_name *

name *

TOP 10 AÑOS CON MENOS RATING

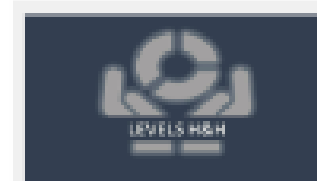


TOP 10 AÑOS CON MEJOR RATING



Fecha de la última actualización: 9/2/2024 7:42:33

ANALYTICS



ANALISIS DE RESTAURANTES

PANEL DE CONTROL

year

☒ month

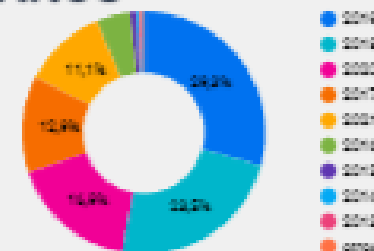
year

<input checked="" type="checkbox"/> julio	11/22
<input checked="" type="checkbox"/> agosto	11/22
<input checked="" type="checkbox"/> abril	10/22
<input checked="" type="checkbox"/> diciembre	10/22
<input checked="" type="checkbox"/> septiembre	10/19
<input checked="" type="checkbox"/> octubre	10/12
<input checked="" type="checkbox"/> noviembre	988897722

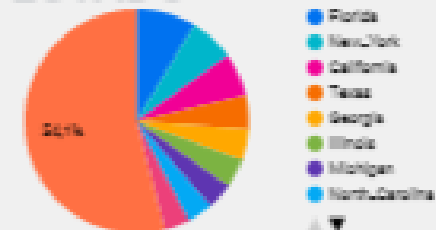
state_name

name

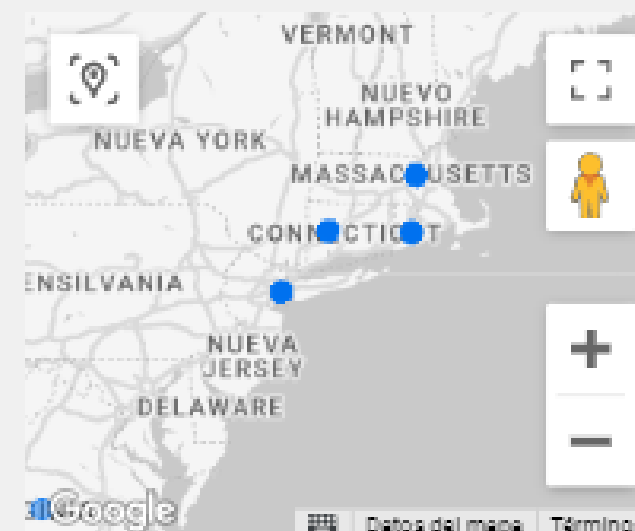
AÑOS



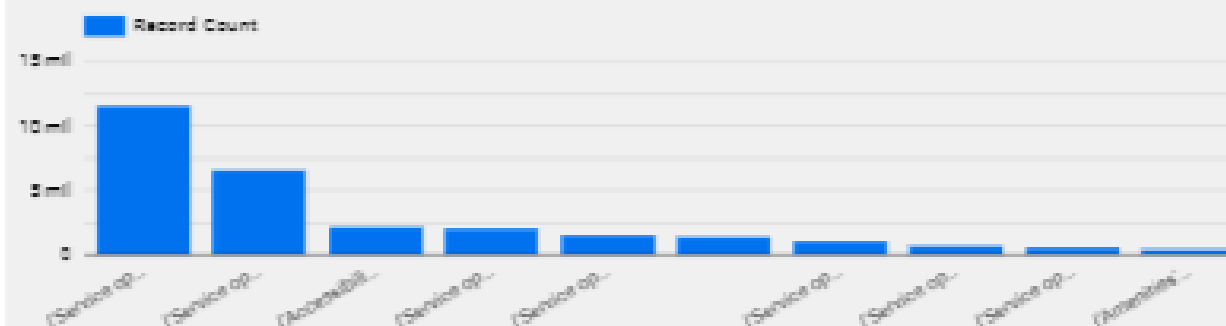
ESTADO



MAPA



SERVICIOS



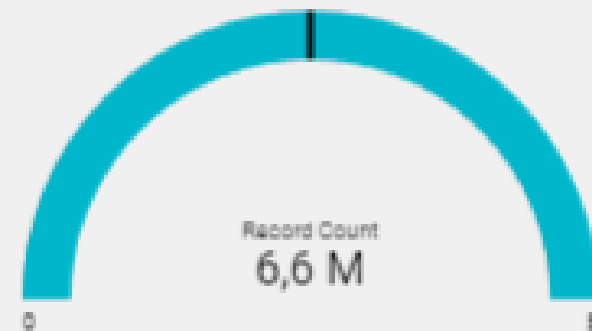
Fecha de la última actualización: 9/2/2024 7:42:33

ANALYTICS

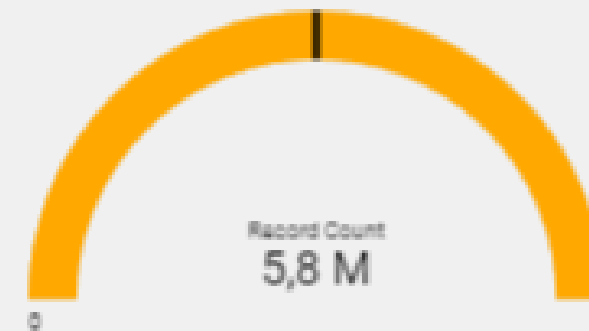


KPI'S

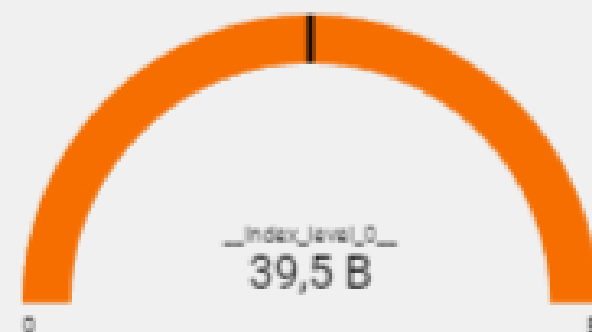
Tasa de satisfacción del cliente



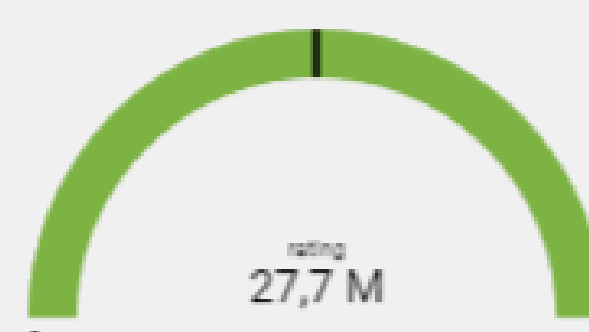
Promedio de rating



Nivel de clientela



NPS (Net Promoter Score)

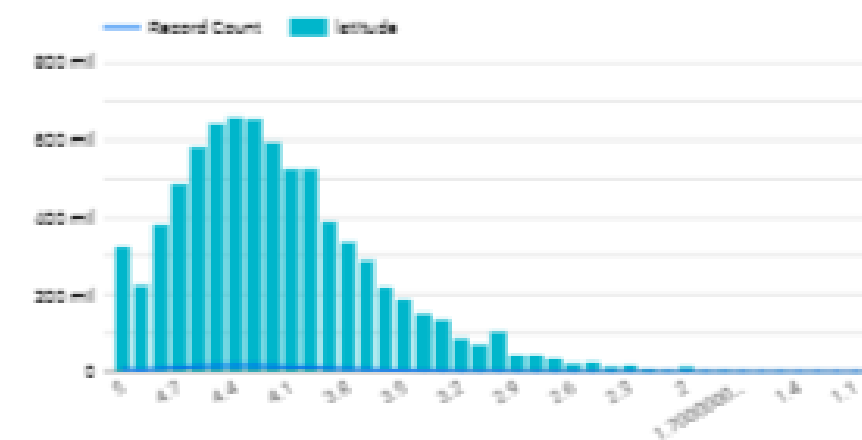
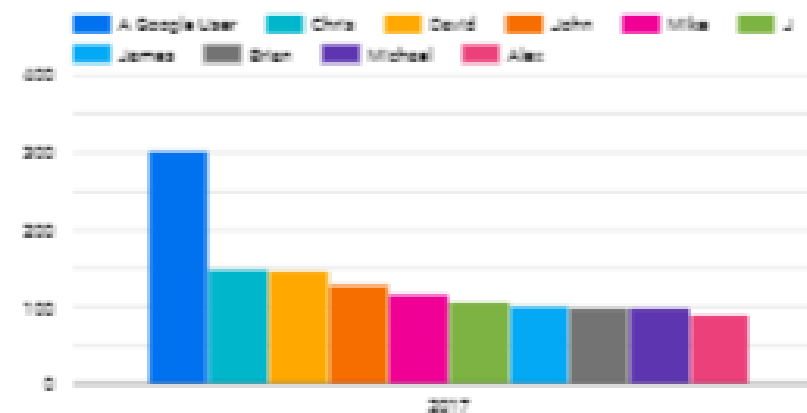
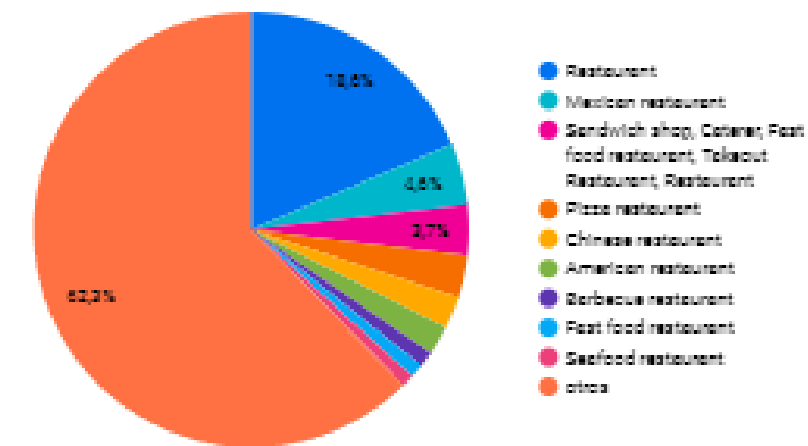
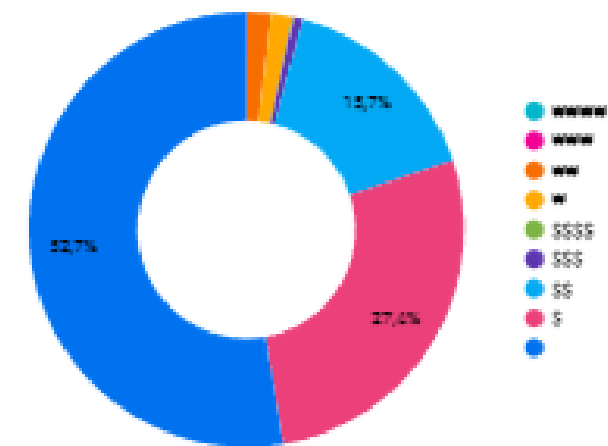


Fecha de la última actualización: 8/2/2024 16:23:56

ANALYTICS



MACHINE LEARNING



Fecha de la última actualización: 8/2/2024 15:33:52

MACHINE LEARNING



ANALISIS DE
SENTIMIENTO



ALGORITMO
PARA
PREDICCIÓN
DE SECTORES

Algoritmos para series
temporales

Algoritmos de IA para datos
tabulares

CIERRE Y CONCLUSIONES



GRACIAS POR SU ATENCIÓN

LEVELS H&H

Mariela Bracamonte

Rolando Stiwan

Matias Spak

Valentín Amat

Pedro Garcia

