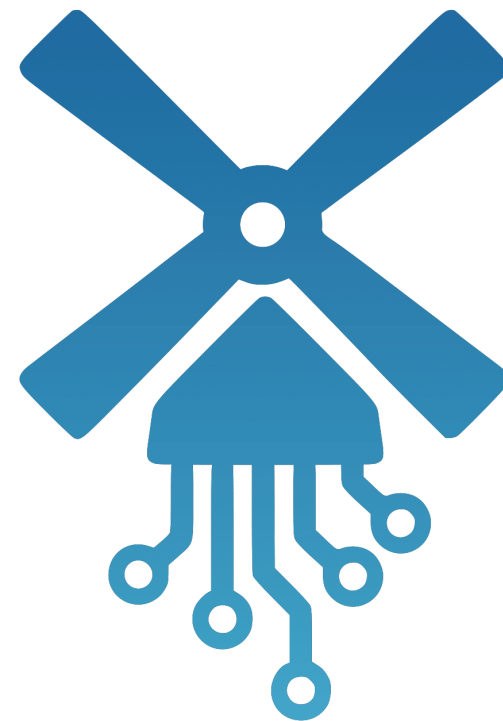


# Datos Abiertos

La información del mundo, abierta para ti.  
Atrévete a descubrir ideas que pueden  
cambiar tu forma de ver las cosas.



Luis Nuñez  
SAS Sr. Systems Engineer



26 de Febrero 2026, 11:30h  
Albacete



# 50

Years of  
BUSINESS  
ANALYTICS

# #1

Market Leader in  
**AI & ADVANCED  
ANALYTICS**



# 3000+

Education Institution  
Partners



SAS is deemed the

# World's Most Innovative Company

(Fast Company  
March 2021)

## 8 YEARS RUNNING



Only vendor named leader  
in Gartner's Magic Quadrant  
for Data Science and Machine Learning  
Platforms.



SAS recognized on  
People magazine's  
**COMPANIES THAT  
CARE**

OVER

# 12K

Employees  
worldwide



SAS recognized  
as leader in  
**more than 30**  
vendor ranking  
reports

# 96%

Net customer  
retention rate



Growing  
Partner Network  
of more than



# 2500

# 88

of the top

# 100

Companies on the GLOBAL  
LIST ARE SAS CUSTOMERS  
or their affiliates

# FORTUNE 500

# 80,000+

Customer sites in 150 countries



**\$3.2 billion**  
REVENUE

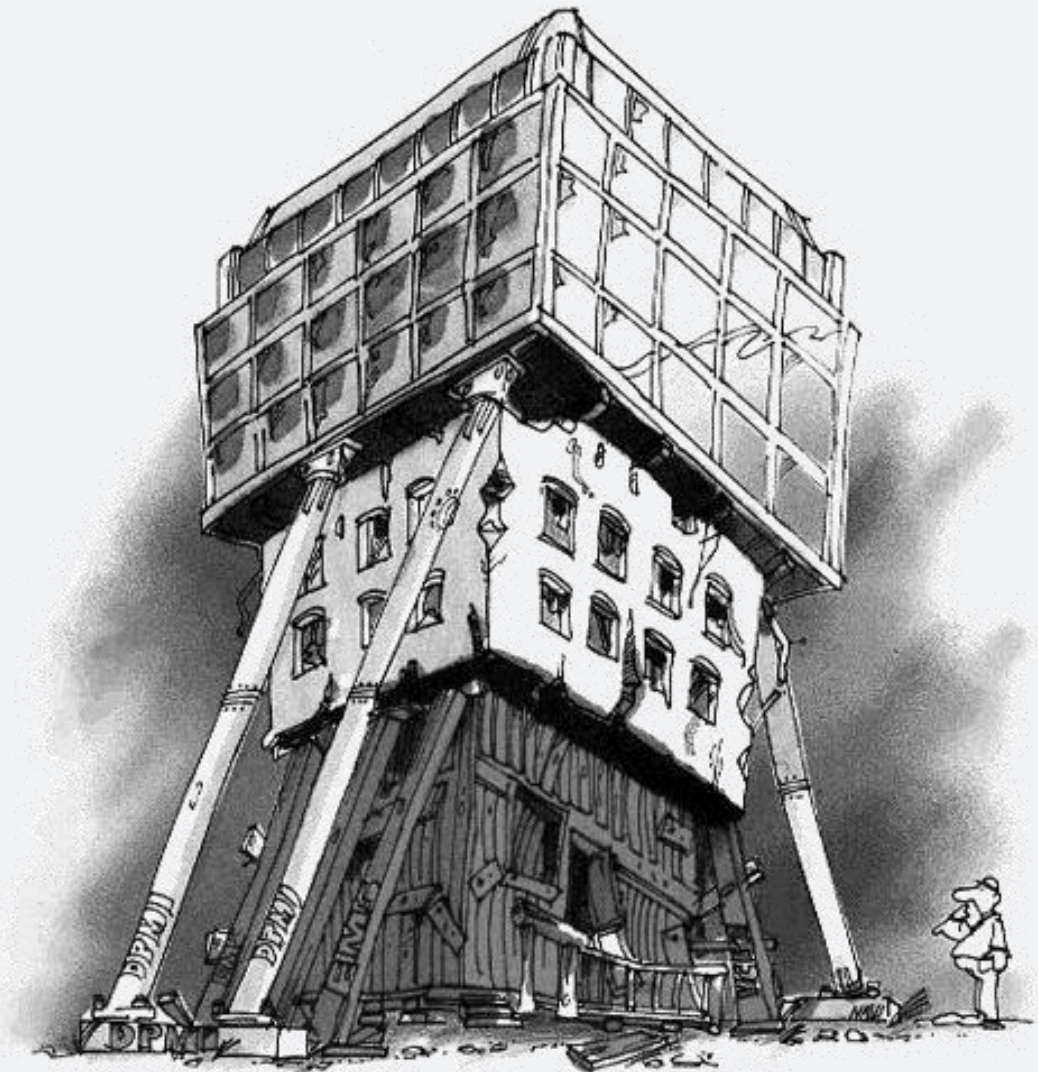


**¿Construiríais un cohete  
con piezas defectuosas?**

¿Construiríais un cohete con piezas defectuosas?



¿Viviríais en una casa sin unos buenos cimientos?





 Gemini

 Copilot

 perplexity

ANTHROPIC





# Garbage *in* Garbage **OUT**



We need more data!

We need more **GOOD** data!



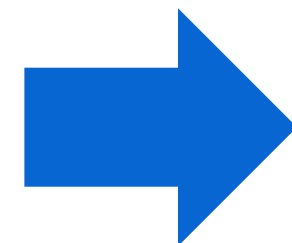
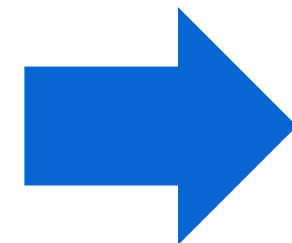
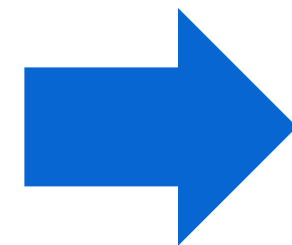
# Asegurar la Calidad de Datos

-  **Completeness**
-  **Uniqueness**
-  **Timeliness**
-  **Validity**
-  **Accuracy**
-  **Consistency**



# ¿Cómo lo hacemos?

## Fuentes de Datos



## Trabajar con los Datos



Acceso a la información



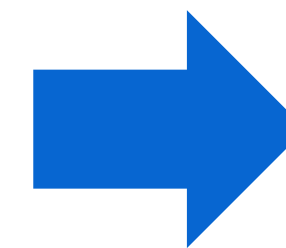
Análisis del nivel de calidad



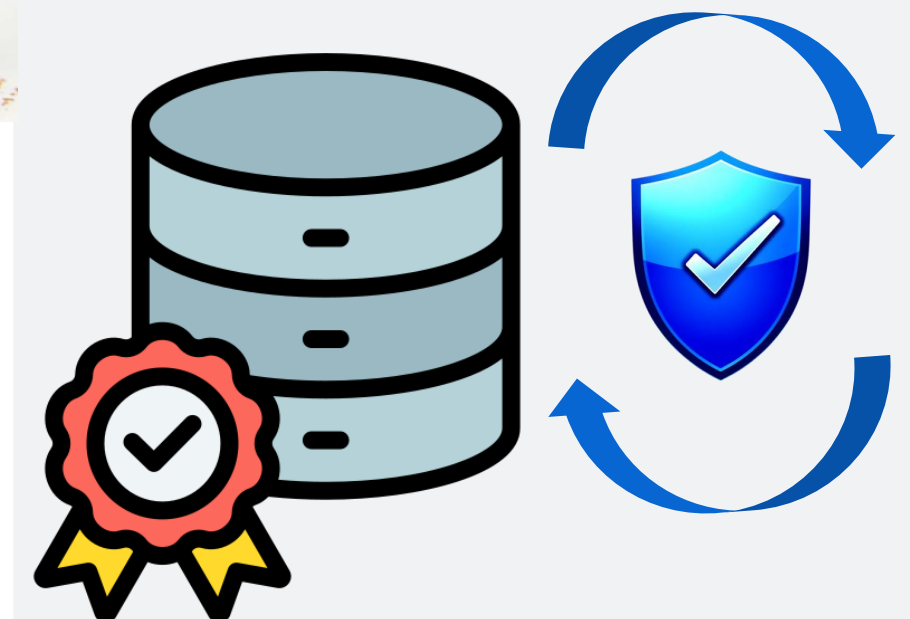
Limpiar la Información



Controlar el nivel de calidad



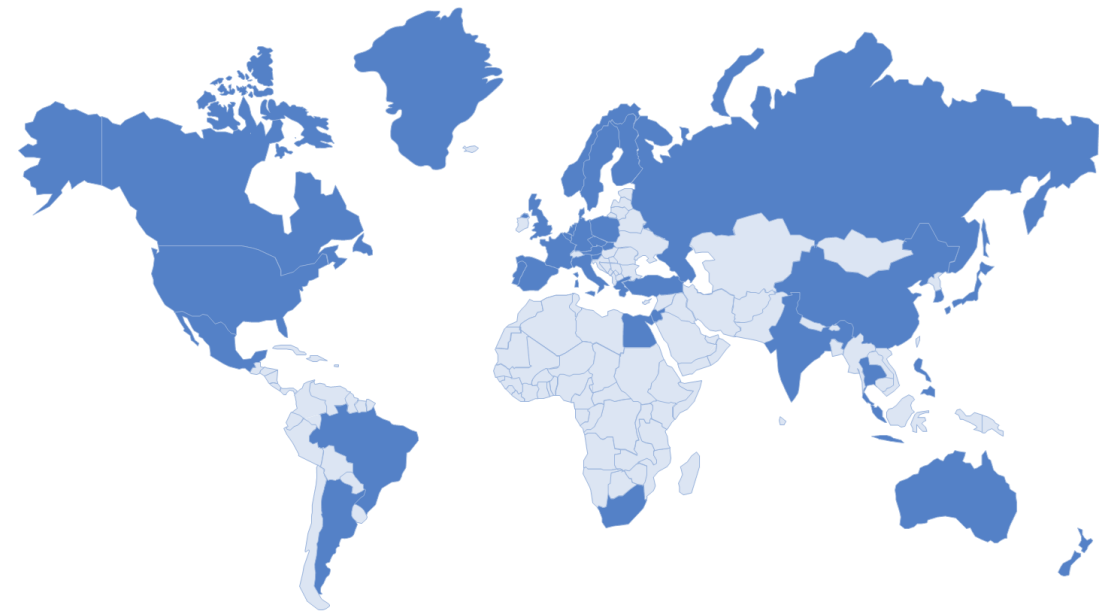
## Datos Limpios





**Ssas** viiya<sup>®</sup>

# Asegurar Data Quality para Geografías e Idiomas



**47 países**



**36 idiomas**

over 51 localizaciones, incluyendo diferentes sistemas de escritura como Árabe, Kanji, Griego, Hebreo, Chino ...

&

Parseo

Estandarización

Matching

Resolución de Entidades

Identificación

Análisis de Género

Enriquecimiento



**115**

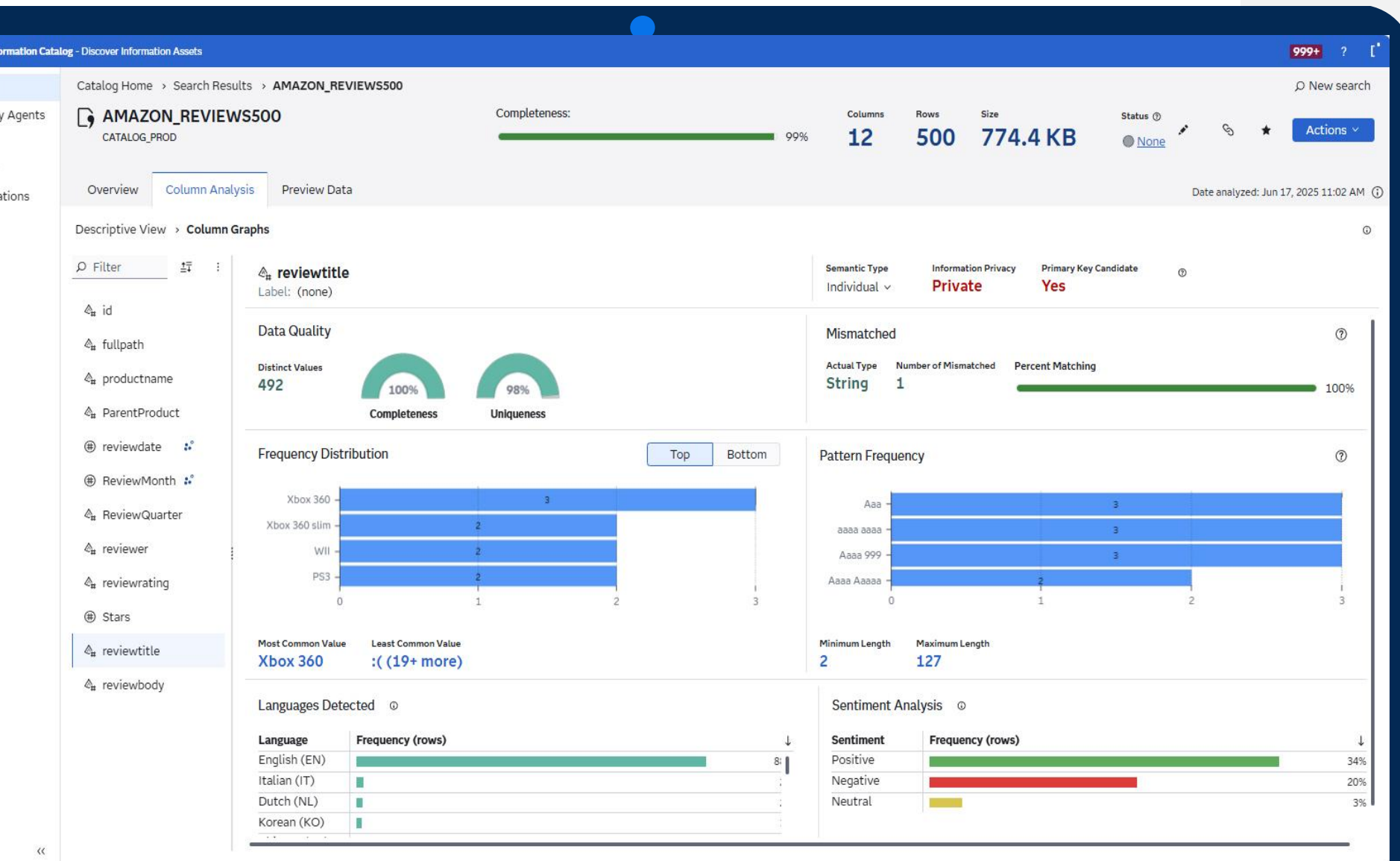
Tipos de Datos



**3rd Party  
& Product  
Data Types**

Tipos de Datos como Personas, Organizaciones, Proveedores, Ciudadanos, Pacientes, Vehículos, Productos, Tamaños, Colores ...

# Evaluar la Calidad de Nuestros Datos



- **Agentes Automáticos de Discovery de Datos** para evaluar la Calidad de la información
- **Identificar Datos Personales** a través de análisis semántico
- **Buscar Información de manera ágil** entre todos los datos en la plataforma

# Limpiar Datos

Estandarizar información

Data	Standard
SAS	SAS Institute, SA
Institute, S.A.	SAS Institute, SA
María del Carmen González	Sra. María del Carmen González
Sra. M <sup>a</sup> Carmen Gonzalez	Sra. María del Carmen González
Maricarmen Gonzalez	Sra. María del Carmen González
+34912007300 ext250	(+34) 912007300 ext 250
91-2007300 x250	(+34) 912007300 ext 250
C/ Torre Laguna 23 Primero A	Calle Torrelaguna, 23 1ºA
CL.Torrelaguna nº23 1º-letra A	Calle Torrelaguna, 23 1ºA

- **Unificar representaciones de datos** escribir los datos de la misma manera
- **Corregir información**, errores tipográficos, abreviaturas o derivados de nombres
- **Tratamiento de caracteres no deseados**

# Limpiar Datos

## Parsear información

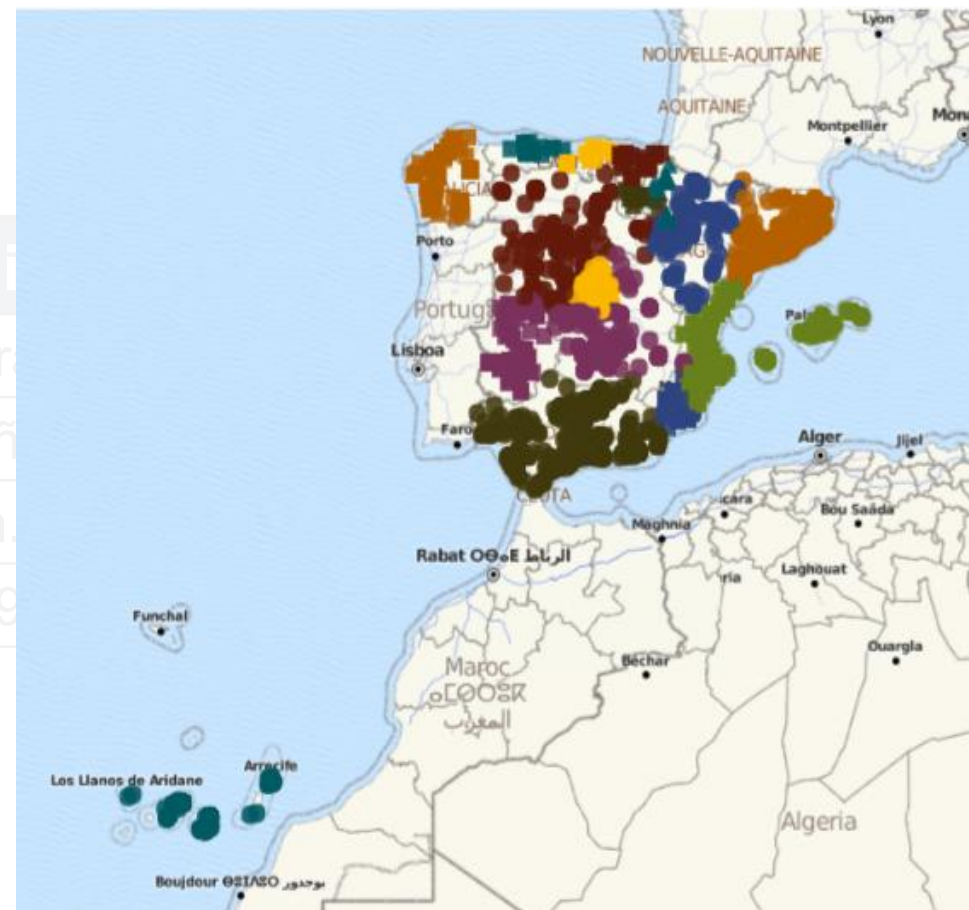
Avenida de Manoterias, 44 Madrid 28050  
Avda. Mañoterias núm. 44 28040  
Madrid, Av. Maanoterias 44



Tipo Vía	Nombre	Número	CP	Ciudad
Avenida	de Manoterias	44	28050	Madrid
Avda.	Mañoterias	núm. 44	28040	
Av.	Maanoterias	44		Madrid

- **Reconocer entidades dentro del texto**
- **Extraer componentes con significado**
- **Identificación de datos** como personas, organizaciones o datos personales

# Enriquecer Datos



- **Validar direcciones** contra un callejero
- **Enriquecer direcciones** con datos la información que falta
- **Geolocalización de los datos** para poder representarlos en un mapa

# Integración de Datos

CODPER	Given Name	Last Name
2399765	M. ARANZAZU	MACIÑEIRAS
8934412	DIEGO	LOPEZ DE AZCONA

ID	Name	Family Name
3345565	DIEGO	LOPEZ-AZCONA
2587650	ARANTXA	MACINERAS

- Identifica o elimina **duplicados** a través de lógica difusa (**Fuzzy Matching**)
- Selecciona el **mejor registro (Golden Record)** vinculando datos de diferentes fuentes
- Permite crear **unidades familiares** en los datos

# Cómo funcionan los códigos de Match

Se parsean los datos en sus componentes

Name	Name Matchcode @ 85 Sensitivity
Ms. Elizabeth Marie Lyte	W&~\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$M_~2\$\$\$\$\$\$\$\$\$\$\$\$
Beth M Lyte	W&~\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$M_~2\$\$\$\$\$\$\$\$\$\$\$\$
“ lizzie light ”	W&~\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$M_~2\$\$\$\$\$\$\$\$\$\$\$\$
Lyte, Liz M	W&~\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$M_~2\$\$\$\$\$\$\$\$\$\$\$\$
Elizabeth Lite	W&~\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$\$M_~2\$\$\$\$\$\$\$\$\$\$\$\$

Según la sensibilidad, se usan componentes para el matchcode



SAS®

Viya®



**Gracias!!!**

