f33

HubSpot and SOW Data Cleaning: Analysis and Suggestions on the HubSpot Data

Prepared by Youssef Franci

Summary

In our quest for impeccable data integrity in our HubSpot system, I executed an exhaustive data enhancement project. This document details the measures undertaken and offers tailored recommendations to elevate our ongoing data management standards.

Actions Overview

1. **Data Extraction**: Initiated by forging a link with Google Drive to capture key datasets. The primary focus was the HubSpot dataset, complemented by the DealID Sheet.
2. **Data Analysis**: Identified discrepancies within the "SOW ID" and "SOW #" columns of the main dataset. All inconsistencies were systematically recorded.
3. **Data Archiving**: Preserved the identified inconsistencies in an SQLite database, serving as a pivotal resource for current analytics and future evaluations.
4. **Data Cleanup & Integration**: Streamlined the DealID Sheet by removing duplicates. Subsequently, merged it with the primary dataset, achieving a unified and de-duplicated data repository. I further enhanced the data by normalizing text entries and filling in any gaps.
5. **Discrepancy Resolution**: Acknowledging the "SOW #" column's accuracy, I amended the "SOW ID" column to address irregularities. Entries that did not conform to the desired format were highlighted for potential follow-up.

Strategic Recommendations

a. **HubSpot Data Refresh**: Align HubSpot with the newly formulated CSV file, which will be regarded as the definitive source. This alignment guarantees the most recent and precise data representation.

b. **Uniformity in Data**: Address the disparities pinpointed in the "SOW ID" and "SOW #" columns. Using the polished data, we can make requisite modifications to achieve data consistency.

c. **Routine Data Maintenance**: Implement an automation system for monthly data housekeeping. This proactive measure will nip potential inconsistencies in the bud, maintaining data precision.

d. **Duplicate Management**: It's crucial to promptly identify and flag dataset duplicates. Post identification, a strategy can be devised—whether it's merging, deletion, or further assessment.

Appendix: Updated and Cleaned SOWs

▮▮▮		▮▮
▮	▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮	▮▮
▮▮		
▮▮	▮▮▮▮▮	▮▮

███████		
████	████████████████████████████	██████
███████		
████	██████████████████████████████	██████
███████		
████	██	██████
███████	███	
████	██████████████████	██████
███████	██████████████████████████████████████	
████	█████████████████████	██████
███████	██	
████	██████████████	██████
███████		
████	███████████████████████	██████
███████		
████	████████████████████████	██████
███████		
████	████████████████████████	██████
███████		
████	██████████████████████████	██████
███████		
████	██████████████████████████████	██████
███████		
████	██████████████████████████	██████
███████		
████	██████████████████████████████	██████
███████		
████	███████████████████████████████	██████
███████		
████	███████████████████████	██████

███████ ████	██	████████
███████ ████	██████████████████████	████████
███████ ████	████████████████████████████	████████
███████ ████	████████████████████████████	████████
███████ ████	██████████████████████	████████
███████ ████	████████████████	████████
███████ ████	████████████████████████	████████
███████ ████	██	████████
███████ ████	████████████████████████████████████	████████
███████ ████	████████████████████████████████████	████████
███████ ████	████████████████████████████████████	████████
███████ ████	████████████████████████████	████████
███████ ████	██████████████████████████████████	████████
███████ ████	████████████████	████████
███████ ████	██████████████████████	████████

■■■■		■■■
■■■	■■■■■■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■	
■■■■		■■■
■■■	■■■■■■■■■■■■■■	
■■■■	■■■■■■■■■■■■■	
■■■	■■■	■■■
■■■■		
■■■	■■■■■■■■■	■■■

███████		
████	██████████████████████████████████████	████████
███████		
████	████████████████████████████████████	████████
███████		
█████	████████████████████████████	████████
███████		
█████	██████████████████████████████	████████
███████		
█████	████████████████████████████	████████
███████		
█████	██████████████████████████████████████	████████
███████		
█████	██████████████████████	████████
███████		
█████	████████████████████████████████	████████
████████		
█████	██	████████
███████		
█████	████████████████████████████████	████████
███████		
█████	████████████████████████████████	████████
███████		
█████	██████████████████████████████████	████████
███████		
█████	████████████████████████████████████	████████
███████		
█████	██████████████████████	████████
███████		
█████	██████████████████████████████████████	████████

▮▮▮▮		▮▮▮▮
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	
▮▮▮▮		▮▮▮▮
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮
▮▮▮▮		
▮▮	▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		▮▮▮▮
▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮▮
▮▮▮▮		▮▮▮▮
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮

▬▬▬		▬▬▬
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬▬	▬
▬▬▬		▬▬▬
▬▬	▬▬▬▬▬▬▬	▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		▬▬▬
▬▬	▬▬▬▬▬▬▬▬▬	▬
▬▬▬		
▬▬	▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬

▇	▇	
▇	▇▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇	▇
▇		
▇	▇▇▇▇▇▇	▇

██████		
██	████████████████████████	████
██████		
██	█████████████████████████	████
██████		
██	█████████████████████	████
██████		
██	███████████████████████████	████
██████		
██	████████████████████████	████
██████		
██	███████████████████	████
██████		
██	█████████████████████████	████
██████		
██	████████████████████████████	████
██████		
██	██████████████████████████████████	████
██████		
██	██████████████████████	████
██████		
██	█████████████████████	████
██████		
██	███████████████████████	████
██████		
██	██	████
██████		
██	███████████████████████	████
██████		
██	███████████████████████	████

▬▬▬		▬▬▬
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬
▬▬▬		▬▬▬
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬	▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬▬▬▬	▬▬▬
▬▬▬		
▬▬	▬▬▬▬▬▬▬	▬▬▬

▆▆▆▆		▆▆▆▆
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆▆	▆▆▆▆
▆▆▆▆		
	▆▆ ▆▆▆▆▆▆▆▆▆▆	▆▆▆▆

███████			
	███	████████████████████████████	██████
███████			
	███	█████████████████	██████
███████			
	███	████████████████	██████
██████			
	███	███████████████████	██████
██████			
	███	████████████	██████
██████			
	███	██████████████████████	██████
██████			
	███	█████████████████████████	██████
██████			
	███	████████████████	██████
██████			
	███	████████████████	██████
██████			
	███	████████████████████	██████
██████			
	███	███████████████	██████
██████			
	███	█████████████	██████
██████			
	███	██████████████	██████
██████			
	███	██████████████	██████
██████			
	███	████████████	██████

▮▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮	▮▮▮
▮▮▮		
	▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮

▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮▮	▮▮▮
▮▮▮		
▮▮	▮▮▮▮▮▮▮▮▮▮▮	▮▮▮