Data Dictionary: Anonymized Hubspot Tickets Dataset

Overview

This dataset contains anonymized implementation/support ticket records from Hubspot Service Hub. All personally identifiable information (PII) and sensitive business information has been removed or masked. This data describes the implementation projects, their status, milestones, and training completion.

Field Descriptions

Field Name	Data Type	Description	Values	Usage Notes
Ticket ID	Integer	Unique identifier for the ticket	Numeric ID	Primary key for identifying tickets
Create date	Date	Date when the ticket was created	YYYY-MM-DD format	Start of the implementation project
Close date	Date	Date when the ticket was closed	YYYY-MM-DD format	End of the implementation project
Last modified date	Date	Date of last update to the ticket	YYYY-MM-DD format	Used to track activity recency
Ticket status	String	Current status of the ticket	'New', 'In Progress', 'Waiting on contact', 'Waiting on us', 'Closed'	Core field for implementation status tracking
Pipeline	String	Pipeline the ticket belongs to	Pipeline name	Support Pipeline – For handling support tickets. Project Success Plan – To track activities during trials. Sym Creation – To monitor the Sym creation process.
Priority	String	Importance/urgency of the ticket	'Low', 'Medium', 'High', 'Critical'	Used for resource allocation decisions

Field Name	Data Type	Description	Values	Usage Notes
Type	String	Classification of the ticket type	'Implementation', 'Support', 'Training', etc.	Categorizes the nature of the ticket. Internal – For internal issues or requests. Customer Issue – For customer-related concerns.
Latest Milestone	String	Linked to the Project Success Plan pipeline. It reflects the most recent milestone achieved during a trial.	Milestone name	Tracks progress through implementation
Latest Milestone Update Date	Date	Date of the last milestone change	YYYY-MM-DD format	Important for tracking milestone velocity
Time to close (HH:mm:ss)	Time	Duration from creation to closure	Hours:minutes:seconds	Direct measure of implementation time
Response time (HH:mm:ss)	Time	Time to first response	Hours:minutes:seconds	Service level metric
1st Syms presented for review	Date	Date when first simulations were presented	YYYY-MM-DD format	Key implementation milestone
1st Syms approved for production	Date	Date when simulations were approved	YYYY-MM-DD format	Key implementation milestone
1st syms run in production	Date	Date when simulations went live	YYYY-MM-DD format	Key implementation milestone
Deployment plan approved	Date	Date of deployment plan approval	YYYY-MM-DD format	Important project planning milestone
Trial Required	String	Whether a trial was required	'Yes', 'No'	Indicates implementation approach
Trial Start Date	Date	Start date of the trial period	YYYY-MM-DD format	Beginning of evaluation period
Trial End Date	Date	End date of the trial period	YYYY-MM-DD format	End of evaluation period

Field Name	Data Type	Description	Values	Usage Notes
Training: Sym Building 101	String	Indicate whether a customer has completed specific training sessions. Linked to Project Success Plan	Date or empty	Training milestone
Training: Sym Building 201	String	Indicate whether a customer has completed specific training sessions. Linked to Project Success Plan	Date or empty	Training milestone
Training: General Overview	String	Indicate whether a customer has completed specific training sessions. Linked to Project Success Plan	Date or empty	Training milestone
Training: Reporting	String	Indicate whether a customer has completed specific training sessions. Linked to Project Success Plan	Date or empty	Training milestone
Training: Deployment/User Management	String	Indicate whether a customer has completed specific training sessions. Linked to Project Success Plan	Date or empty	Training milestone
Target Launch Date	Date	Planned go-live date	YYYY-MM-DD format	Important for measuring on-time delivery
Stage Date - Project Initiation	Date	Date entered initiation stage	YYYY-MM-DD format	Stage progression tracking
Stage Date - Project Launch	Date	Date entered launch stage	YYYY-MM-DD format	Stage progression tracking
Stage Date - Execution	Date	Date entered execution stage	YYYY-MM-DD format	Stage progression tracking

Field Name	Data Type	Description	Values	Usage Notes
Stage Date - Closure Phase	Date	Date entered closure stage	YYYY-MM-DD format	Stage progression tracking
Ticket name	String	Anonymized ticket name	Format: Ticket_[hash]	Masked to protect client identity
Associated Company	String	Anonymized company name	Format: Company_[hash]	Masked to protect client identity
Associated Company (Primary)	String	Anonymized primary company	Format: Company_[hash]	Masked to protect client identity
Associated Deal	String	Anonymized associated deal	Format: Deal_[hash]	Links to the originating sales deal

Derived/Calculated Fields

These fields are typically added during analysis and not present in the raw data:

Field Name	Data Type	Description	Calculation Method
Implementation_Duration_Days	Integer	Days from creation to closure	(Close date - Create date) in days
Days_To_First_Sym	Integer	Days to first simulation delivery	(1st Syms presented - Create date) in days
Time_To_Close_Hours	Float	Hours from creation to closure	Converted from Time to close (HH:mm:ss)
Implementation_Status	String	Simplified status categories	Mapped from Ticket status
Training_Completion_Count	Integer	Number of completed trainings	Count of completed training fields
Training_Completion_Pct	Float	Percentage of trainings completed	(Completed / Total possible) × 100
Create_Year	Integer	Year when ticket was created	Extracted from Create date
Create_Month	Integer	Month when ticket was created	Extracted from Create date
Create_YearMonth	String	Year-Month for time-based analysis	Format: YYYY-MM

Relationships

This dataset can be linked to other datasets using:

- Associated Company (Primary) → links to companies dataset
- Associated Deal → links to deals dataset
- Ticket ID → links to companies dataset via CompanyToTickets mapping

Usage Examples

Implementation Duration Analysis

```
# Calculate average implementation duration
avg_duration = df['Implementation_Duration_Days'].mean()

# Implementation duration by priority
duration_by_priority = df.groupby('Priority')
['Implementation_Duration_Days'].agg(['mean', 'median'])
```

Training Completion Analysis

```
# Calculate overall training completion rate
training_completion = df['Training_Completion_Pct'].mean()

# Count of tickets by completion level
completion_distribution = pd.cut(
    df['Training_Completion_Pct'],
    bins=[0, 25, 50, 75, 100],
    labels=['0-25%', '26-50%', '51-75%', '76-100%']
).value_counts()
```

Milestone Achievement Tracking

```
# Calculate days between key milestones
df['Days_Approval_to_Production'] = (
    pd.to_datetime(df['1st syms run in production']) -
    pd.to_datetime(df['1st Syms approved for production'])
).dt.days

# Percentage of implementations that achieved each milestone
milestone_achievement = {
    'Presented': df['1st Syms presented for review'].notna().mean() * 100,
    'Approved': df['1st Syms approved for production'].notna().mean() *
100,
    'Production': df['1st syms run in production'].notna().mean() * 100
}
```

Data Quality Notes

- Time-based fields may contain some inconsistencies due to manual data entry
- Training completion fields may be incomplete in some records
- Some milestone dates may be missing if the implementation is still in progress
- Not all tickets have associated deals, particularly for support-oriented tickets