Data Dictionary

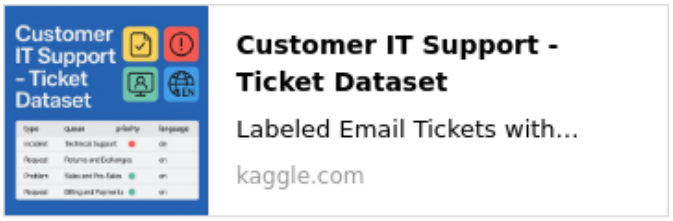
AI-Driven IT Helpdesk & Incident Resolution Analytics

Course: IT300 – Business Intelligence
Institution: Tunis Business School
Student: Mohamed Amin Mnari
Dataset Name: clean_it_helpdesk_tickets.csv
Project Phase: Week 1–2 – Data Preparation & ETL
Date: December 2025

1. Purpose of the Data Dictionary

This data dictionary provides a detailed description of the cleaned dataset used in the *AI-Driven IT Helpdesk & Incident Resolution Analytics* project. Its purpose is to ensure a clear and consistent understanding of each data field, support correct KPI calculation, facilitate dimensional modeling, and enable accurate dashboard development within the Business Intelligence solution.

2. Dataset Overview

Attribute	Description
Data Source	Kaggle – Multilingual Customer Support Tickets 
Dataset Type	Structured, transactional
Number of Records	Approximately 28,500
Number of Fields	12
Granularity	One row per support ticket
Intended Use	BI analysis, KPI computation, and dashboarding

3. Field-Level Data Dictionary

Field Name	Data Type	Description	Example Value	BI Usage
subject	Text	Short description of the reported issue	password reset request	Ticket categorization
type	Categorical	Type of ticket (Incident, Problem, Request)	Incident	Filtering and trend analysis
queue	Categorical	Support queue assigned to the ticket	IT Support	Workload distribution
priority	Categorical	Urgency level of the ticket	High	SLA and priority analysis
language	Categorical	Language of the ticket	EN	Multilingual and AI performance analysis
all_tags	Text	Combined list of issue tags	login, access	Issue grouping
created_date	Date	Date the ticket was created	2024-03-15	Time-series analysis
resolved_date	Date	Date the ticket was resolved	2024-03-16	Resolution metrics
resolution_time_hours	Numeric	Time taken to resolve the ticket (in hours)	12	Performance KPIs
sla_breached	Binary (0/1)	Indicates whether SLA was breached	0	SLA compliance metrics
ai_resolved	Binary (0/1)	Indicates if ticket	1	AI

		was resolved by AI		effectiveness analysis
ai_confidence_score	Numeric	Confidence score of AI resolution	0.87	AI quality assessment

4. Engineered Fields Explanation

The following attributes were not present in the original raw dataset and were generated during the ETL process to support Business Intelligence analysis and simulate a realistic AI-enabled IT helpdesk environment.

Engineered Field	Purpose
created_date	Enables time-based analysis and trend reporting
resolved_date	Required to calculate resolution duration
resolution_time_hours	Measures operational efficiency
sla_breached	Supports SLA compliance monitoring
ai_resolved	Simulates AI-assisted ticket resolution
ai_confidence_score	Represents AI decision reliability

Note: Engineered AI-related fields are simulated for analytical purposes and do not represent outputs from a real production AI system.

5. Data Quality Rules

The following data quality rules were applied during the ETL process:

Rule	Description
Uniqueness	**Each row represents a unique support ticket**
Mandatory fields	**type, priority, queue, created_date must not be null**
Date consistency	**resolved_date is always greater than or equal to created_date**
Resolution time	**resolution_time_hours is strictly positive**
SLA logic	**SLA thresholds are defined based on ticket priority**

6. Mapping to Dimensional Model

The cleaned dataset supports a star schema design for BI analysis. Key fields are mapped as follows:

Field	Target Table
created_date	DimDate
priority	DimPriority
queue	DimQueue
language	DimLanguage
type	DimTicketType
resolution_time_hours	FactTickets
sla_breached	FactTickets
ai_resolved	FactTickets