ZOMATO DATASET - COMPLETE DATA DICTIONARY

Field Name	Data Type	Description	Range/Values	Unique Values	Null Count	ETL Transformations	Sample Value
ID	String/ Object	Unique hexadecimal transaction ID	Hexadecimal format	45,546	0	None	`0xcdcd`
Delivery_person _ID	String/ Object	Delivery personnel identifier	[CITY][RES?][D EL##] pattern	1,320	0	None	`DEHRES1 7DEL01`
Delivery_person _Age	Float6 4	Age of delivery person	20.0 - 50.0	21	0	Missing → Median (29.6), Invalid ages removed	29.6
Delivery_person _Ratings	Float6 4	Customer ratings (1-6 scale)	2.5 - 6.0	27	0	Missing → Median (4.7)	4.2
Restaurant_latit ude	Float6 4	Restaurant latitude	-30.91 to 30.91	654	0	None	30.32796 8
Restaurant_long itude	Float6 4	Restaurant longitude	-88.37 to 88.43	514	0	None	78.04610 6
Delivery_locatio n_latitude	Float6 4	Delivery destination latitude	0.01 to 31.05	4,373	0	None	30.39796 8
Delivery_locatio n_longitude	Float6 4	Delivery destination longitude	0.01 to 88.56	4,373	0	None	78.11610 6

Field	Type	Description	Format/Values			Transformation	Example
Order_Date	String/ Object	Date order placed	DD-MM-YYYY format	44	0	None	`12-02-2022`
Time_Orderd	Time/ Object	Time order placed	00:00:00 - 23:59:59	160	5,761	Converted from mixed formats	`21:55:00`
Time_Order_pic ked	Time/ Object	Time order picked up	00:00:00 - 23:59:59	174	5,004	Converted from mixed formats	`22:10:00`
Weather_conditi ons	String/ Object	Weather during delivery	Fog, Sunny, Stormy, Sandstorms, Cloudy, Windy, Not Specified	7	0	Missing → "Not Specified", Title case	`Fog`
Road_traffic_de nsity	String/ Object	Traffic conditions	Jam, High, Medium, Low, Not Specified	5	0	Missing → "Not Specified", Title case	`Jam`
Vehicle_conditi on	Integer	Vehicle condition (0-3 scale)	0, 1, 2, 3	4	0	None	2
Type_of_order	String/ Object	Food order category	Snack, Meal, Drinks, Buffet	4	0	None	`Snack`
Type_of_vehicle	String/ Object	Vehicle type used	motorcycle, scooter, bicycle, [others]	4	0	None	`motorcy cle`
multiple_deliveri es	Integer	Simultaneo us deliveries	0, 1, 2, 3	4	0	Converted to int, NaN → 0	3
Festival	String/ Object	Festival day indicator	Yes, No, Not Specified	3	0	Missing → "Not Specified", Title case	`No`

| City | String/Object | City classification | Metropolitian, Urban, Semi-Urban, Not Specified | 4 | 0 | Missing → "Not Specified", Title case | `Metropolitian` |
| Time_taken (min) | Float64 | Delivery time in minutes | 10.0 - 49.0 | 40 | 0 | Outliers capped at 99th percentile | 26.3 |

KEY CHANGES FROM ORIGINAL TO CLEANED DATA

Aspect	Original Dataset	Cleaned Dataset	ETL Action
Total Records	45,584	45,546	38 invalid age records removed
Delivery_person_Age nulls	1,854	0	Filled with median (29.6)
Delivery_person_Ratings nulls	1,908	0	Filled with median (4.7)
multiple_deliveries nulls	993	0	Filled with 0
Weather_conditions nulls	616	0	Filled with "Not Specified"
Road_traffic_density nulls	601	0	Filled with "Not Specified"

City nulls	1,200	0	Filled with "Not Specified"
Festival nulls	228	0	Filled with "Not Specified"
multiple_deliveries type	Float64	Integer	Converted to integer
Time_taken range	10-54	10-49	Outliers capped at 99th percentile
Time format	Mixed formats	Standardized time	Converted to HH:MM:SS

ETL TRANSFORMATIONS APPLIED

Step	Fields Affected	Transformation	Impact
1	multiple_deliveries	Converted to numeric, NaN → 0, integer type	Standardized to integer
2	Delivery_person_Age, Delivery_person_Ratings	Missing → Median values	0 nulls
3	Delivery_person_Age	Remove invalid ages (<18, >65)	38 records removed

4	Time_taken (min)	Cap outliers at 99th percentile	Reduced skew
5	Time_Ordered, Time_Order_picked	Convert mixed formats to time objects	Standardized time data
6	Weather_conditions, Road_traffic_density, City, Festival	Missing → "Not Specified", Title case	0 nulls, consistent text