Here are the column names from your sheet along with a sample value from the first row:

- order\_id → 0100153
- order\_datetime → Apr 21, 2025 16:48
- customer\_id → C1386
- customer\_name → Krishna Verma
- customer\_email → krishna.verma@yahoo.com
- **customer\_phone** → 9951192192
- city → Bengaluru
- area → Koramangala
- restaurant\_id → R005
- restaurant\_name → Pizza Point
- cuisine → North Indian
- item\_name → Paneer Tikka
- item\_category → Starter
- quantity → 1
- item\_price → 327.71
- discount\_amount → 0
- total\_amount → 327.71
- payment\_method → Card

- delivery\_partner\_id → D005
- delivery\_partner\_name → Amit J
- promised\_time\_min → 30
- actual\_delivery\_time\_min → 41.0
- order\_status → Delivered
- rating → 3.0
- feedback → very late
- Unnamed: 25 → NaN
- Unnamed: 26 → NaN
- Unnamed: 27 → NaN

Would you like me to **remove the extra unnamed columns** (25, 26, 27) since they seem empty? **①** 



collect 5 more random values from each column to clearly understand the structure and data type

Here are 5 random sample values from each column so you can clearly see the structure and data types:

- order\_id → 0103031, 0102503, 0101572, 0100935, 0100699
- order\_datetime → 01-04-2025, 23.07.2025 19:26, Mar 09, 2025 12:08, 03-06-25 11:39, 15.06.2025 16:18
- customer\_id → C1105, C1143, C1186, C1145, C1174
- **customer\_name** → Vihaan Reddy,Ishaan Chopra,Krishna Kapoor,Arjun Khan,Aarav Verma
- customer\_email → kabir.das@yahoo.com, ayaan.kapoor@yahoo.com, aditya.khan@hotmail.com, Kabir.ghosh@yahoo.com, vihaan.bose@yahoo.com
- customer\_phone → 9336681091, 948-746-9021,
   (+91) 92871-74155, (+91) 92962-37088,
   (+91) 92848-12984
- city → Bengaluru, Ahmedabad, Hyderabad,
   Ahmedabad, Hyderabad
- area → Indiranagar, SATELLITE, Banjara Hills, Maninagar, Madhapur
- restaurant\_id → R001, R002, R005, R007, R001
- restaurant\_name → Spice-Hub, Green-Bowl,
   Pizza Point, Wok-This-Way, Spice Hub

- cuisine → Biryani, Italian, South Indian, Biryani, Chinese
- item\_name → Veg Biryani, Margherita Pizza,
   Masala Dosa, Chicken Biryani, Veg Hakka
   Noodles
- item\_category → Main Course, Main Course, Main Course, Main Course
- quantity → 2, 1, 1, 1, 1
- item\_price → 528.29, 233.51, 391.01, 285.05, 387.59
- discount\_amount  $\rightarrow$  0, 0, 0, 0, 40
- total\_amount → 334.22, 243.14, 231.8, 498.42, 381.33
- payment\_method → upi, Credit Card, Cash, U
   PI, upi
- delivery\_partner\_id → D002, D008, D007, D007,
   D004
- delivery\_partner\_name → Sunil P, Rahul N,
   Mahesh B, Mahesh B, Rakesh S
- promised\_time\_min → 52, 46, 52, 39, 46
- actual\_delivery\_time\_min → 46.0, 70.0, 78.0, 66.0, 48.0

- order\_status → Delivered, Delivered,
   Delivered, Delivered
- rating  $\rightarrow$  3.0, 3.0, 3.0, 3.5, 3.0
- **feedback** → cold food, excellent, wrong item, average, tasty
- Unnamed: 25 → empty
- Unnamed: 26 → empty
- Unnamed: 27 → empty

Would you like me to **clean the dataset** (fix inconsistent date formats, standardize phone numbers & payment methods, and drop empty columns) before you start SQL/analysis practice?





perform a data quality assessment for each column and check for the following 1.Count and percentage of missing, null and blank entries 2. inconsistent format and inconsistent data types 3.outlier or value that appear unusual

I've generated a **Data Quality Assessment Report** for each column in your dataset.

#### It shows:

- Count & % of missing/null/blank values
- Possible inconsistent formats or data types
- Potential outliers or rare/unusual values

You can now explore the full table in the interactive view I've opened.





Got it delivery Here's a summary of the key data quality issues by column, highlighting what needs cleaning most urgently:

# Key Issues by Column

- order\_id, customer\_id, restaurant\_id, delivery\_partner\_id
  - No missing values, but rare/unique
     categories dominate → likely OK (IDs should be unique).
- order\_datetime
  - Inconsistent formats (text-based like Mar 09, 2025 vs numeric-like 01-04-2025).
  - Needs date standardization to a single format (YYYY-MM-DD HH:MM:SS).
- customer\_name
  - Values.
  - Minor issue: extra spaces/inconsistent
     casing (e.g., Ishaan Chopra vs Arjun Khan).

#### customer\_email

- X 4.5% missing values.
- Some emails may have inconsistent casing/ spaces (e.g., \*Kabir.ghosh@yahoo.com \*).
- Needs trimming and format validation.

## customer\_phone

- Inconsistent formats (10-digit, with dashes, with country codes, spaces).
- Needs standardization to a consistent format (e.g., +91XXXXXXXXXX).

## · city, area

- Inconsistent casing & spacing (e.g., \* Hyderabad\*, SATELLITE).
- Needs trimming and consistent capitalization.

#### restaurant\_name

- X Variants of the same restaurant exist (e.g., Spice-Hub vs Spice Hub).
- Needs deduplication/standardization.
- cuisine, item\_name, item\_category
  - Mostly consistent, but spelling/ capitalization variants may exist.

- E.g., North Indian vs north indian.
- quantity, item\_price, discount\_amount, total\_amount
  - Numeric fields, no missing values.
  - Outliers may exist (e.g., extremely high order values or discounts).
  - Should check if total\_amount = quantity ×
     item\_price discount\_amount always holds.

## payment\_method

- Inconsistent spelling/casing (e.g., upi, U PI, UPI).
- Needs normalization.
- promised\_time\_min, actual\_delivery\_time\_min
  - Numeric, no missing values.
  - Some outliers possible (e.g., extremely high delivery times).

### order\_status

- ✓ Mostly consistent ("Delivered").
- Check if there are rare values like Canceled, Pending.

#### rating, feedback

Ratings look numeric but need range

validation (should be 1-5).

 Feedback text has spelling/subjectivity issues but OK for analysis.

Unnamed: 25-27

 Entirely empty → should be dropped.

## Most Urgent Cleaning Needed:

- order\_datetime (date format standardization)
- customer\_phone (phone format normalization)
- payment\_method (normalize spelling/casing)
- customer\_email (fill missing, remove trailing spaces)
- restaurant\_name & city/area (deduplicate & trim)
- Drop Unnamed: 25-27

Would you like me to go one step further and create a cleaned version of this dataset in CSV/ Excel for your SQL/analysis practice?

