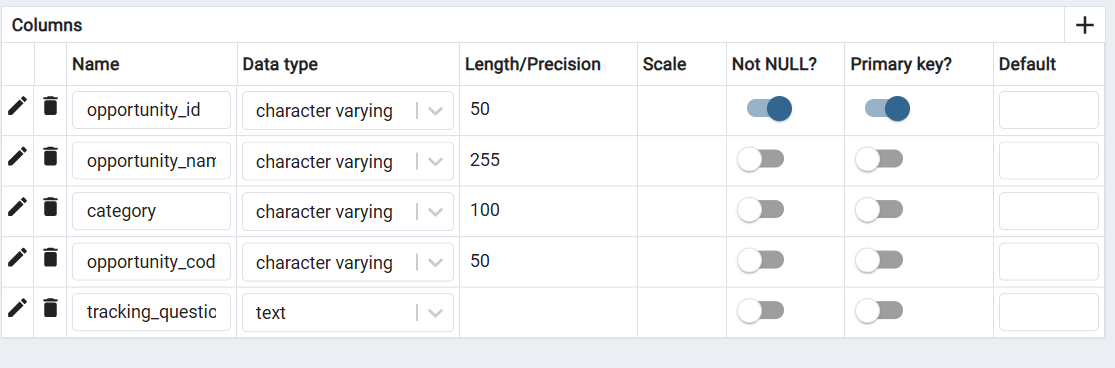
**EDA OF OPPORTUNITY DATASET**

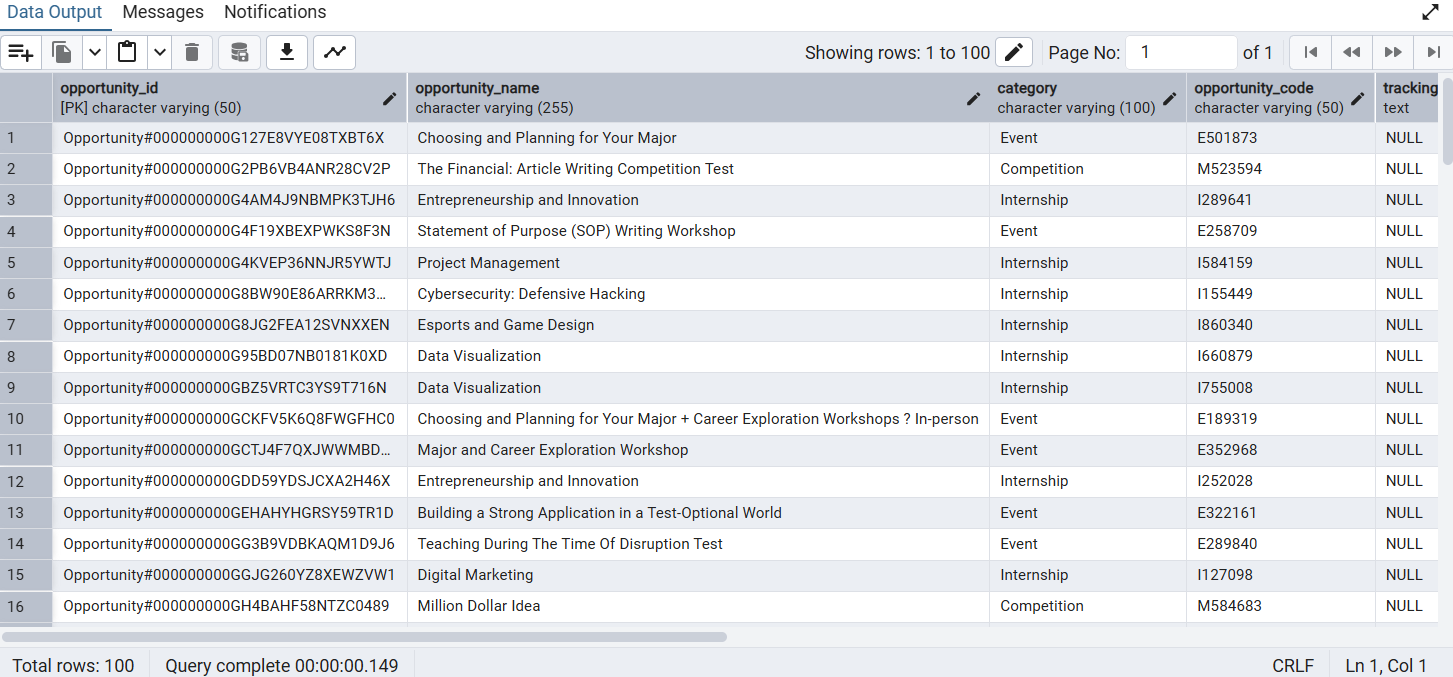
**Understand the Dataset**

#### Step 1: Understand the Dataset

* View the first few rows to get an overview of the data.
* Check column names, data types, and the total number of records.

Research how to inspect a table’s structure and data in PostgreSQL.

****



Identify Missing and Duplicate Values Detect missing data to determine if any cleaning is needed. Identify duplicate records that may need removal. Look up methods to check for missing values and duplicates in PostgreSQL.

## **1. Dataset Structure**

* **Rows**: 187
* **Columns**: 5 (all text/object type)
* **No purely numeric columns**
* **Memory usage**: ~7.4 KB

**Columns**:  
opportunity\_id, opportunity\_name, category, opportunity\_code, tracking\_questions

## **2. Missing Values**

| **Column** | **Missing Count** | **Missing %** |
| --- | --- | --- |
| tracking\_questions | 69 | 36.9% |
| Others | 0 | 0% |

## **3. Duplicates**

* **Duplicate rows**: **0**

## **4. Top Categories**

| **Column** | **Most Frequent Value** | **Frequency** |
| --- | --- | --- |
| opportunity\_id | Opportunity#000000000G127E8VYE08TXBT6X | 1 |
| opportunity\_name | Project Management | 4 |
| category | Internship | 43 |
| opportunity\_code | E501873 | 1 |
| tracking\_questions | {"is\_required\_for\_badge\_award":"true","code":"QBWTGGT","question":"All deliverables..."} | 1 |

## **5. Correlations**

* No numeric columns → no correlation analysis possible.

## **6. Outliers**

* No numeric columns → no outlier detection possible.

## **Key Insights**

1. **Data type issue**: All columns are stored as object; there’s no numeric or date field.
2. **Missing data**: Only tracking\_questions has missing values (~37% missing).
3. **Categorical distribution**:
   * category has 7 unique values, dominated by **Internship** (43 occurrences).
   * opportunity\_name has many unique values (170/187 rows are unique).
4. **Duplicates**: No duplicate records found.
5. **EDA limitation**: No numeric analysis possible without transforming data.

#### Visualize Data Trends

* Use graphs like histograms, line charts, and box plots to identify patterns.
* Analyze relationships between different variables.

 Find out how to create basic visualizations for data analysis.

