Project Title: Clean & Analyze NYC Taxi Trip Data

Context:

You work as a data engineer for a company analyzing NYC transportation data. You're tasked with building a CLI tool that can ingest, clean, and summarize raw taxi trip data stored in CSV and JSON formats.

You can simulate rows with common issues:

- Missing values (`null`, empty strings)
- Inconsistent data (e.g., `payment_type` as "credit card", "Credit Card", "CREDIT CARD")
- Invalid entries (e.g., negative `trip_distance` or `fare_amount`)

Suggested Cleaning Tasks

- Drop or fill missing `passenger_count`
- Convert all `payment_type` to lowercase and standardize
- Filter out rows where `fare_amount` or `trip_distance` <= 0
- Parse date strings into datetime objects

Expected Summary Stats Output

```
$ python datacleaner.py --input nyc_taxi_sample.csv --summary

Summary Stats:
--------------
Total rows: 5
Valid rows after cleaning: 3

Numeric Columns:
- trip_distance: mean=4.3, min=3.5, max=5.2
- fare_amount: mean=15.2, min=12.5, max=18.0

Categorical Columns:
- payment_type:
  - credit card: 2
  - cash: 1
```