**Takeaways:-** Read json, Json\_tuple, to\_json, explode, collect\_list, concat\_ws

**Problem Statement:**

You are getting transactional data for your ecart business in json format and you have to transform data as per the given instructions

You can run the python script and generate the dummy data

Sample data:-

[

[

{

"transaction\_id": "T001",

"customer\_id": "C006",

"items": [

{

"item\_id": "I010",

"product\_name": "Printer",

"quantity": 1,

"price": 1369.88

},

{

"item\_id": "I007",

"product\_name": "Monitor",

"quantity": 1,

"price": 845.42

}

],

"transaction\_time": "2024-07-26T13:16:23Z",

"delivery\_info": {

"status": "Shipped",

"delivery\_time": "2024-07-31T13:53:23Z",

"address": {

"street": "532 Pine St",

"city": "Star City",

"state": "TX",

"zip": "32658"

}

}

},

{

"transaction\_id": "T002",

"customer\_id": "C009",

"items": [

{

"item\_id": "I003",

"product\_name": "Mouse",

"quantity": 3,

"price": 1101.22

}

],

"transaction\_time": "2024-07-26T12:06:23Z",

"delivery\_info": {

"status": "Pending",

"delivery\_time": "2024-07-30T13:53:23Z",

"address": {

"street": "924 Elm St",

"city": "Central City",

"state": "TX",

"zip": "19249"

}

}

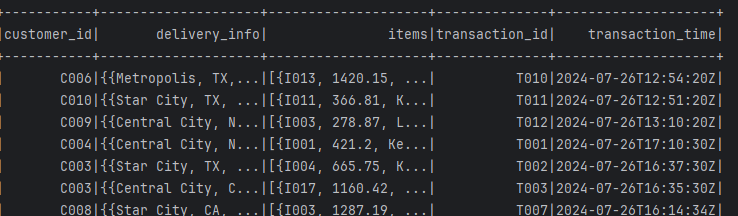
},

]

**Step 1**

Read the json file

Output should be in this format

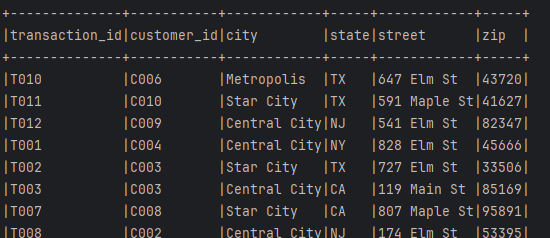


**Step 2**

The delivery\_info column contains city, state, street and zipcode.

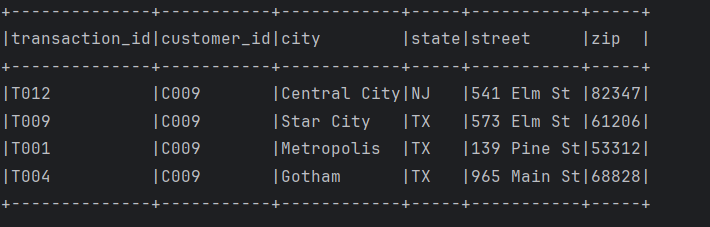
Fetch those details from the column, alonwith transaction\_id and customer\_id

Output format



**Step 3**

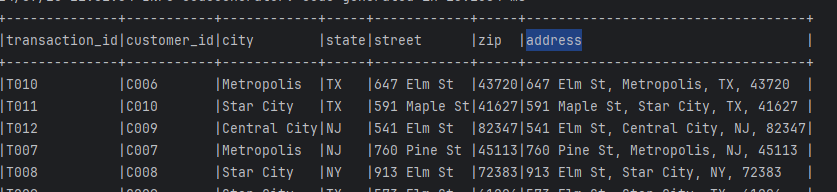
Filter customer\_ids with multiple addresses like in following eg customer\_id C009



**Step 4**

Create one address column in the format street,city,state,zip

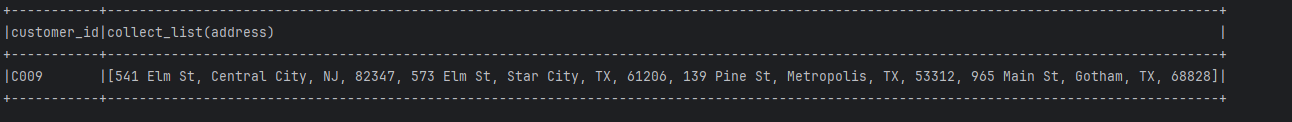
Output format



**Step 5**

Create a separate customerAddress Dataframe with all address corresponding to that customer\_id

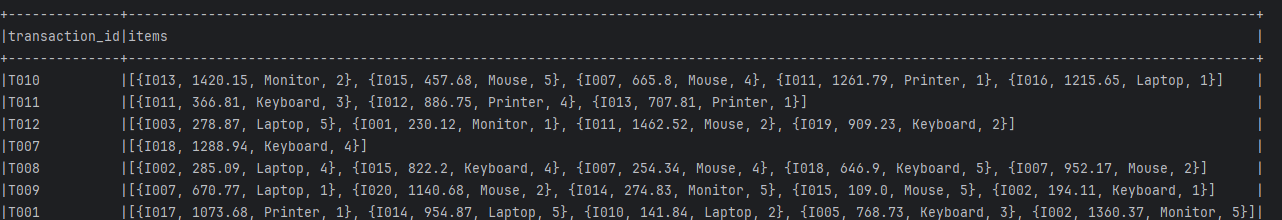
Output format



**Step 6**

Fetch item details corresponding to each transaction ID

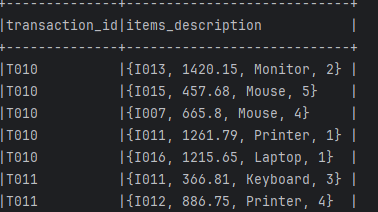
Output format



**Step 7**

Denormalize it with just 1 set of item information in one row

Output format



**Step 8**

Split item\_Description column as item\_id, price, product\_name, quantity

Output format

