

## Assignment

We have exported 2 datasets from our system, one contains orders from customers and another contains barcodes (with an order\_id if they are sold).

To print the Tickets vouchers we need a csv file with all the barcodes and orders\_ids per customer.

Write a program that reads these two files, orders.csv and barcodes.csv, and generates an output file that contains the following data:

```
customer_id, order_id1, [barcode1, barcode2, ...]  
customer_id, order_id2, [barcode1, barcode2, ...]
```

Bonus points:

- We want to have the top 5 customers that bought the most amount of tickets. The script should print (to stdout) the top 5 customers of the dataset. Each line should be in the following format:  
customer\_id, amount\_of\_tickets
- Print the amount of unused barcodes (barcodes left).
- Model how you would store this in a SQL database (e.g. UML, data model with relations and optionally indexes)

## Input files:

Two files in comma separated formatting.

### **orders.csv**

*order\_id, customer\_id*

This contains a list of orders. order\_id is unique.

### **barcodes.csv**

*barcode, order\_id*

The barcodes in our system. If a barcode has been sold, it's assigned to an order using order\_id, otherwise order\_id is empty.

## Validation:

Make sure the input is validated correctly:

- No duplicate barcodes
- No orders without barcodes

Items which failed the validation should be logged (e.g. stderr) and ignored for the output.

## Requirements:

- Write your solution in Python
- Deliver solution using git or a zip file.
- Runs out of the box and is production ready