RDD Drill 1

Tom Bohbot

May 2021

1 Introduction

Questions 1, 7, and 8 do not display photos in the report as no output is necessary for these questions. Additionally, the code is not included in the report for the other questions as the code can be found in GitHub.

2 1) Ingest the file

No output necessary. Please see code, and you can unmute the line that prints all the lines from file to verify visually. I muted this so that you are not spammed with a lot of output.

3 2) How many lines are in the file?

```
2) number of lines: 1000
```

4 3) How many unique customers?

```
3) number of unique customers: 100
```

5 4) Which customer made the most purchases?

```
4) ID of customer with most purchases is: 53
```

6   5) How many purchases did she make?

```
5) The previous customer made: 19 purchases
```

7   6) Print that set of purchases using the format below:

```
6) set of purchases that customer 53 made:
2015-03-30, 6:18 AM, 53, 42, 5, 2197.85
2015-03-30, 4:42 AM, 53, 44, 6, 9182.08
2015-03-30, 2:51 AM, 53, 59, 5, 3154.43
2015-03-30, 5:57 PM, 53, 31, 5, 6649.27
2015-03-30, 6:11 AM, 53, 33, 10, 2353.72
2015-03-30, 9:46 PM, 53, 93, 1, 2889.03
2015-03-30, 4:15 PM, 53, 72, 7, 9157.55
2015-03-30, 2:42 PM, 53, 94, 1, 921.65
2015-03-30, 8:30 AM, 53, 38, 5, 4000.92
2015-03-30, 6:06 AM, 53, 12, 6, 2174.02
2015-03-30, 3:44 AM, 53, 47, 1, 7556.32
2015-03-30, 10:25 AM, 53, 30, 2, 5107.0
2015-03-30, 1:48 AM, 53, 58, 4, 718.93
2015-03-30, 9:31 AM, 53, 18, 4, 8214.79
2015-03-30, 9:04 AM, 53, 68, 4, 9246.59
2015-03-30, 1:51 AM, 53, 40, 1, 4095.5
2015-03-30, 1:53 PM, 53, 85, 9, 1630.24
2015-03-30, 6:51 PM, 53, 100, 1, 1694.52
2015-03-30, 7:39 PM, 53, 100, 8, 7885.35
```

8   7) Apply a 5 percent discount to all purchases of item 25, whenever a given purchase involves more than one of these items.

No output necessary.

9   8) Create a set of "complimentary txs" in which the company gives a free item to certain customers. Specifically: every customer who bought five or more items with id 81 are given a free item 70. Each "complimentary tx" corresponds to a new transaction, and the set of these transactions should be added to the existing set of trans- actions.

No output necessary.

10   9) How many transactions are now in the data-set?

```
9) amount of transactions currently in dataset: 1006
```

11   10)Which customer spent the most money, and how much money did they spend?

```
10a) ID of customer who spent most money: 76
10b) Customer 76 spent 100049.0 dollars.
```