Intern Coding Test - Data Science

The following exercises are intended to help us get a feel for your data domain expertise. Please post Code Solutions to a public Github repository. Commit directly to the master branch. Commit as often as you like but at least once per exercise. Please don't squash commits or rewrite history.

The exercise will be given to the candidate couple of days before we discuss solutions. At a high level, below is the break-up of the exercise

---

Data

Below represents <u>sample</u> coupon data for a retailer. Each column (tokenized value) is extracted from **OfferDetails**

Product - Can be a single value or an array

FaceValue - The savings on the offer , generally a number

UOM - Unit of measure, represents the measuring unit for the product

Product	FaceValue	UOM	OfferDetails
Downy Liquid Fabric Conditioner Downy Odor Protect Downy Wrinkle Guard Downy Nature Blends Bounce/Downy Wrinkle Guard In Wash Scent Boosters Downy Unstopables, Fresh Protect, Odor Protect, and Infusions Downy Libre Enjuague, Gain Fireworks	$2.00	72 ld 48 oz 40 oz 67 oz 130 ct 80 ct 8.6 oz	Save $2.00 ONE Downy Liquid Fabric Conditioner 72 ld or larger (includes Downy Odor Protect 48 oz or larger OR Downy Wrinkle Guard 40 oz or larger OR Downy Nature Blends 67 oz or larger) OR Bounce/Downy Sheets 130 ct or larger (includes Bounce/Downy Wrinkle Guard 80 ct or larger) OR In Wash Scent Boosters 8.6 oz or larger (includes Downy Unstopables, Fresh Protect, Odor Protect, and Infusions) (excludes Downy Libre Enjuague, Gain Fireworks, and trial /travel size).
Tide PODS Tide Power PODS Tide Liquid/Powder Laundry Detergent, Tide Simply, Tide Simply PODS	$2.00		Save $2.00 ONE Tide PODS OR Tide Power PODS (excludes Tide Liquid/Powder Laundry Detergent, Tide Simply, Tide Simply PODS, and Tide PODS 9 ct and below, and trial/travel size).
Tide Laundry Detergent Tide Purclean Laundry Detergent, Studio by Tide Laundry Detergent, Tide PODS, Tide Rescue, Tide Simply Laundry Detergent, Tide Simply PODS, Tide Detergent	$2.00	10 oz	Save $2.00 ONE Tide Laundry Detergent (excludes Tide Purclean Laundry Detergent, Studio by Tide Laundry Detergent, Tide PODS, Tide Rescue, Tide Simply Laundry Detergent, Tide Simply PODS, Tide Detergent 10 oz and trial/travel size).
General Mills cereal	$.5	8.9 - 15.2 OZ	SAVE 50¢ ON TWO when you buy TWO BOXES (8.9 - 15.2 OZ. EXCLUDES LARGE SIZE) any flavor/variety General Mills cereal listed: *
Pepperidge Farm® Farmhouse buns, bread, or cookies	$3.00		$3.00 OFF when you purchase any THREE (3) Pepperidge Farm® Farmhouse buns, bread, or cookies

Exercise 1

- Write a function that takes **OfferDetails** as input, and returns **FaceValue** as output
- Run the function against all rows in the attached data and determine function accuracy (Higher the better)

coupons_ner.csv

Exercise 2

- Write a function that takes **OfferDetails** as input, and returns **Product** as output (Product can be single or array)
- Run the function against all rows in the attached data and determine function accuracy (Higher the better)

Hint

- For Exercise 1, regex function should do (ensure that all edge cases are taken care of)
- For Exercise 2, build a corpus of products (built manually) and go from there. Mention the order of complexity (in any measures Big O, memory, cpu etc.)
- For Exercise 2, our expectation is that you implement NER model

Tech Stack

Your choice really. There are many out there NLTK, Spacy, Pytorch, TensorFlow etc.

Duration/delivery

Revert back with github/gitlab/bitbucket link by pushing all of the code into the repo. And add **clear readme.md** setup instructions and any other instructions within 3 days.