#### **Problem Statement**

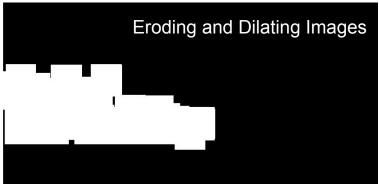




flyer_name	product_name	unit_promo_price	uom	least_unit_for_promo	save_per_unit	disco	unt	organic
week_1_page_1	Ground Sirloin or Round	3.99	lb	1	4.00		0.50	0
week_1_page_1	Blueberries	2.5	1 Pint	2	3.49	Y	0.58	0
wook 1 nogo 2	Marning Dounds	2 00	C Dools	1	0.50		Λ 11	0

## Data Processing: Feature Detection





### Cooked Tail-On Shrimp

31-40 Count

No Water Additives or Preservatives. 100% Traceable and Sustainably Sourced, Caught with Turtle-Safe Nets

Discount Taken at Register

#### Text Identification - Tesseract OCR



2/\$6 SAVE \$2.98 on 2

Blueberries

**Half Pint** 

A healthy snack—perfect for smoothies, added to yogurt, or make blueberry pancakes! High in Vitamin C and antioxidants.

Discount Taken at Register

HEALTHY REWARDS OFFER WITH CARD

2/°6 SAVE \$2.98 on 2

Blueberries
Half Pint
Ahealthy snack—perfect for
smoothies, added to yogurt, or
make blueberry pancakes! High
in Vitamin C and antioxidants.
Discount Taken at Register



## Sentimental Analysis - Smart Heuristics

A few heuristics were integrated to improve the results:

- 1) Correct Translations:
  - a) E.g. lb = ["ib", "lb", "Lb", "ib.", "lb.", "Lb."]
- 2) Problem Dependent Ordering:
  - a) E.g. Line appearing before "SAVE" line includes useful numeric attribute information
- 3) Use Cases:
  - a) E.g. Numeric values appearing in "SAVE" line usually indicate one of the following scenarios:
    - i) Uom and save per unit
    - ii) save\_per\_unit and least\_unit\_for\_promo
    - iii) discount

### Sentimental Analysis - Fuzzy

```
process.extract("Cooked Tail-On Shrimp", product, limit=10, scorer=fuzz.token_sort_ratio)
[(['Cooked Tail-On Shrimp'], 100),
(['Cooked Shrimp'], 76),
(['Key West Pink Shrimp'], 63),
 (['Jumbo Key West Pink Shrimp'], 60),
 (['Wild American Shrimp'], 59),
 (['Bacon Wrapped Shrimp'], 59),
 (['Coconut & Tornado Shrimp Rings'], 57),
 (['New Zealand Sharp Cheddar'], 57),
 (['East Coast White Shrimp'], 55),
 (['Wild White American Shrimp'], 55)]
```

# Solution

	Α	В	С	D	E	F	G	Н	1
1	flyer_name	product_na	unit_promo	uom	least_unit_	save_per_u	discount	organic	.0
2	week_10_page_1	Cooked Tai	9.99	lb	1	5		0	
3	week_10_page_1	Organic Gre	een Kale		2	0.5		1	
4	week_10_page_1	Whole Chic	0	lb	1	1.3		0	
5	week_10_page_1	NY Strip		lb	1	6		0	
6	week_10_page_1	Organic Re	d & Green S	lb	1	1		1	
7	week_10_page_1	Alaska Coh	6	lb	1			0	
8	week_10_page_1	Ground Ro	0	lb	1	3.5		0	
9	week_10_page_1	Blueberries		pint	2	1.49		0	
10	week_10_page_2	Brand Wide	e Sale		1		0.2	0	
11	week_10_page_2	Cleanses, F	ibre Suppor	t, & Skinny	( 1			1	
12	week_10_page_2	Clawhamm	949	pack	1	0.5		0	
13	week_10_page_2	Rustic Red	0	lb	1	5		0	
14	week_10_page_2	Pumpkin D	9.99	pack	1	0		0	
15	week_10_page_2	Maple Cara	0	inch	1	1		0	
16	week_10_page_2	Hummus		oz	2	1.99		1	