

# Daisy Intelligence Hackathon 2020

Solution by Team Artemis

David Weitzenfeld

Lorna Licollari

# Step 1 OCR

Images are analysed for text using **Google Cloud Vision API** (OCR).



Sales Valid Wed. September 6 – Tues. September 12

<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$3.99</b></p> <p>SAVE \$3 New Crop</p> <p><b>Organic Gala Apples</b> 3 lb. Bag Crisp, with a mildly sweet flavor and floral aroma. Discount Taken at Register</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>2/\$6</b></p> <p>SAVE \$2.98 on 2 <b>Blueberries</b> 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</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$3.99/lb.</b></p> <p>SAVE \$3.50/lb. Ground Fresh 3X Daily <b>USDA Choice or Better Ground Round</b> Never Administered Antibiotics or Growth Hormones, 100% Vegetarian Diet, 3rd Party Certified Humanely Handled, Genetically Certified Angus, Traceable from Farm to Fork. Discount Taken at Register</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$8.99/lb.</b></p> <p>SAVE \$6/lb. Fresh, Wild Caught <b>Alaska Coho Salmon Fillets</b> 100% Traceable and Sustainably Sourced, All Earth Fare Salmon is non-GMO, Superior Flavor, Great Source of Omega-3 Fatty Acids, Product of the USA (Alaska). Discount Taken at Register</p>
<p><b>ORGANIC</b></p> <p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$2.99/lb.</b></p> <p>SAVE \$1/lb. <b>Organic Red &amp; Green Seedless Grapes</b> "Known to boost an important heart antioxidant known as glutathione, grapes can protect against heart failure and high blood pressure." - Angela Hind, M.D. and Wellness Expert Discount Taken at Register</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$1.29/lb.</b></p> <p>SAVE \$1.30/lb. <b>Whole Chickens</b> Never Administered Antibiotics or Any Added Hormones, 100% Vegetarian Diet, 3rd Party Certified Humanely Raised Discount Taken at Register</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$9.99/lb.</b></p> <p>SAVE \$5/lb. <b>Cooked Tail-On Shrimp</b> 31-40 Count No Water Additives or Preservatives, 100% Traceable and Sustainably Sourced, Caught with Turtle-Safe Nets Discount Taken at Register</p>	<p><b>HEALTHY REWARDS OFFER WITH CARD</b></p> <p><b>\$7.99/lb.</b></p> <p>SAVE \$2/lb. <b>Napa Chicken Salad</b> From the Deli Signature item, made fresh in-house in small batches using American Humane Certified chicken that is never administered antibiotics or growth hormones. Discount Taken at Register</p>

**SIGN UP FOR EMAIL DEALS & MORE AT EARTHFARE.COM**

## Step 2 Clustering

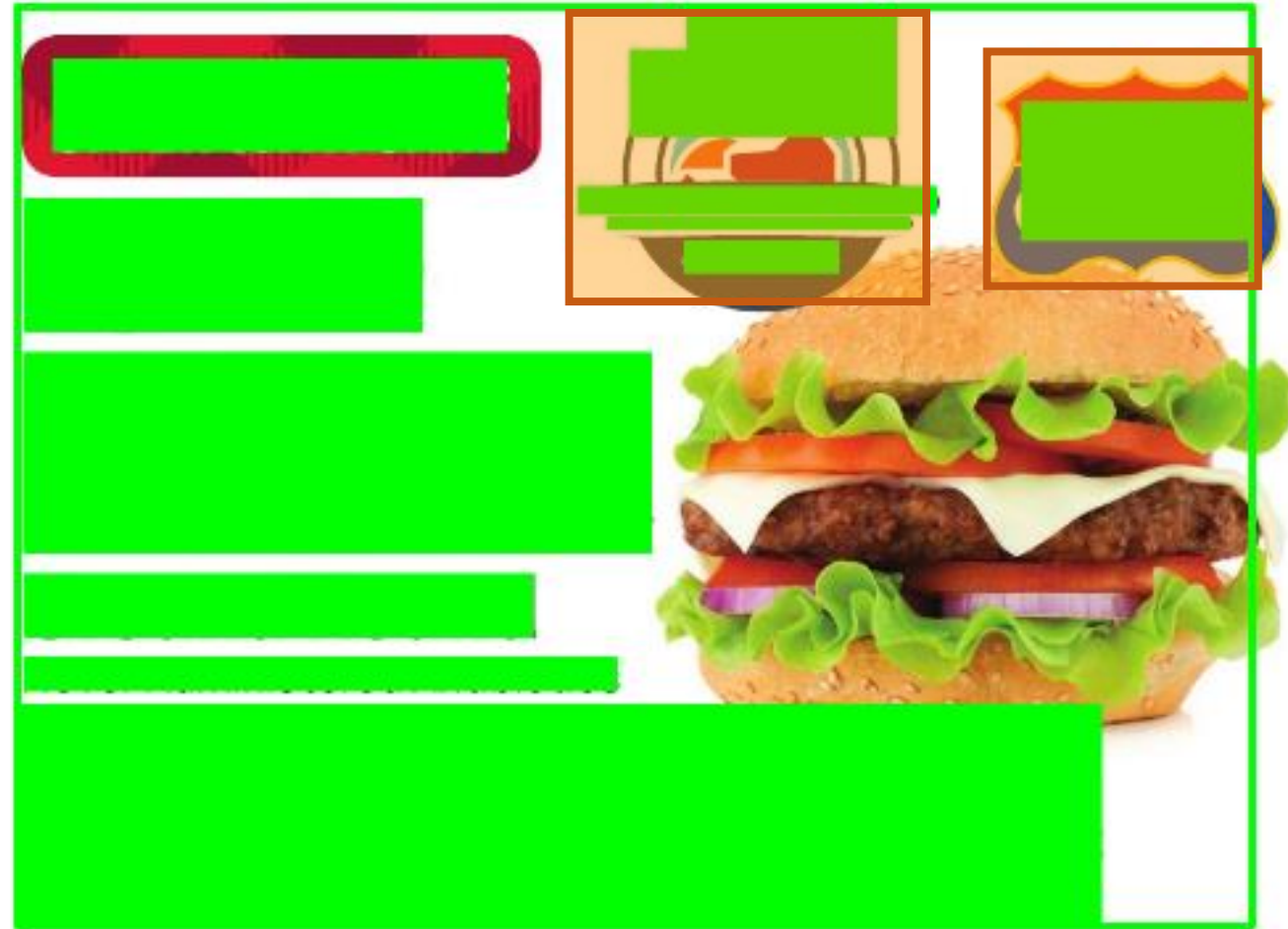
Text blocks are grouped using a custom clustering algorithm which makes use of the text alignment present in each ad-block.





## Step 3 Cluster Refinement

Clusters are refined by drawing an outer rectangle around each one and forcing big clusters to “eat” small ones that fall in their domain.



## Step 4 Information Parsing

Each cluster text is then analysed to **extract the required fields** based on **string matching** and **patterns observed** from the flyers.

