

Daisy Intelligence Hackathon

Graham Hoyes,
Daniil Orekov,
Danial Hasan

Preprocessing Flyers

- Notice that Product, Price, and Discount data are in **Red**, **Black**, and **Grey** only
- Removed irrelevant colors from flyer images to improve OCR performance



Google Vision OCR

- Used Google Cloud Vision OCR
 - Powerful
 - Reliable
 - Affordable
- Detects text in images and returns:
 - Text symbols
 - Bounding boxes for blocks, paragraphs, words, etc.



Detecting Text: Products

- Exploit the color theme to identify different text types, e.g.:
 - **Product Name: Black**
 - **Price: Red**
- Use bounding box height to identify different font sizes
- Use *fuzzywuzzy* library for fuzzy string matching to products



Use **height** and **color** to identify

Detecting Text: Price and Discounts

- Exploit the color theme to identify different text types, e.g.:
 - **Product Name: Black**
 - **Price: Red**
- Use **regex** expressions for string type matching for \$, /, %, **SAVE**, etc.
- Assign Price blocks to closest Product block by vertical distance

Price blocks close to
their respective product



Use **color and regex**
to locate and assign
price, unit, etc.

Thank You!