**Computer Vision Scientist Take-Home Exercise**

## Data

We have a **picture of a receipt** and an **OCR engine** that very accurately converts everything on the receipt into text, and for each text entity gives us its 4-point bounding polygon. So for each of the recognized text entities it will output something like:

“text”: “Murphy”

“polygon”: [x1, y1, x2, y2, x3, y3, x4, y4]

## Task

Create a model that, once trained, would

1. **Classify** **text on the receipt** based on several categories that include, among others:

store\_name, store\_address, store\_city, store\_state, store\_zip, product\_quantity, product\_description, product\_price, receipt\_subtotal

(Anything not belonging to these is classified as “text”)

These categories are marked with their own colors in the example illustration.

1. **Semantically group** entries in the products table together as **purchased items.**

E.g.:

“Items”: [

{

“product\_quantity”: 1

“product\_description”: “305s KING FF BPACK”

“product\_price”: 4.69

}

…

]

## Deliverable

Description of the model in the format you are most comfortable with:

* diagrams
* documents
* slides

Got some code? It’s a bonus.

Text, letter

Description automatically generated