

# Grocery Receipt Scanner

Calvin Houser, Rebecca Duong, Rebecca Bui

## Receipt Data Set:

|                | Calvin H. | Rebecca D. | Celine | Rebecca B. | Person 1 | Person 2 |
|----------------|-----------|------------|--------|------------|----------|----------|
| SafeWay        | 17        | 10         | 0      | 2          | 4        | 7        |
| Trader Joe's   | 3         | 3          | 5      | 2          | 0        | 0        |
| Costco         | 3         | 0          | 0      | 1          | 3        | 0        |
| Grocery Outlet | 0         | 0          | 4      | 0          | 0        | 0        |
| New Leaf       | 5         | 0          | 0      | 0          | 1        | 0        |

# Tasks and Libraries

## Tasks:

|               |  |
|---------------|--|
| Parsing       | Ran Tesseract-OCR on scanned grocery receipts. Wrote parser functions to parse each line of each receipt.  |
| Wrangling     | A GUI was also created to manually correct the data that were not recognized or omitted by tesseract. Additional wrangling of the data with a fuzzy match done after the data is saved in a json format. |
| Visualization | Bokeh is used to create graphs or charts for our results.  |

## Libraries:

Tesseract   FuzzyWuzzy   Tkinter   Bokeh   Numpy   Pandas

# Questions

1. What do people spend money on the most?
2. Do our spending habits change over time?
3. How well does the system work?

# References

## Visualization References:

1. [https://bokeh.pydata.org/en/latest/docs/user\\_guide.html](https://bokeh.pydata.org/en/latest/docs/user_guide.html)
2. <https://towardsdatascience.com/data-visualization-with-bokeh-in-python-part-ii-interactions-a4cf994e2512>