Project Proposal: supermarket receipt visualizer

Input: scanned receipt from supermarket

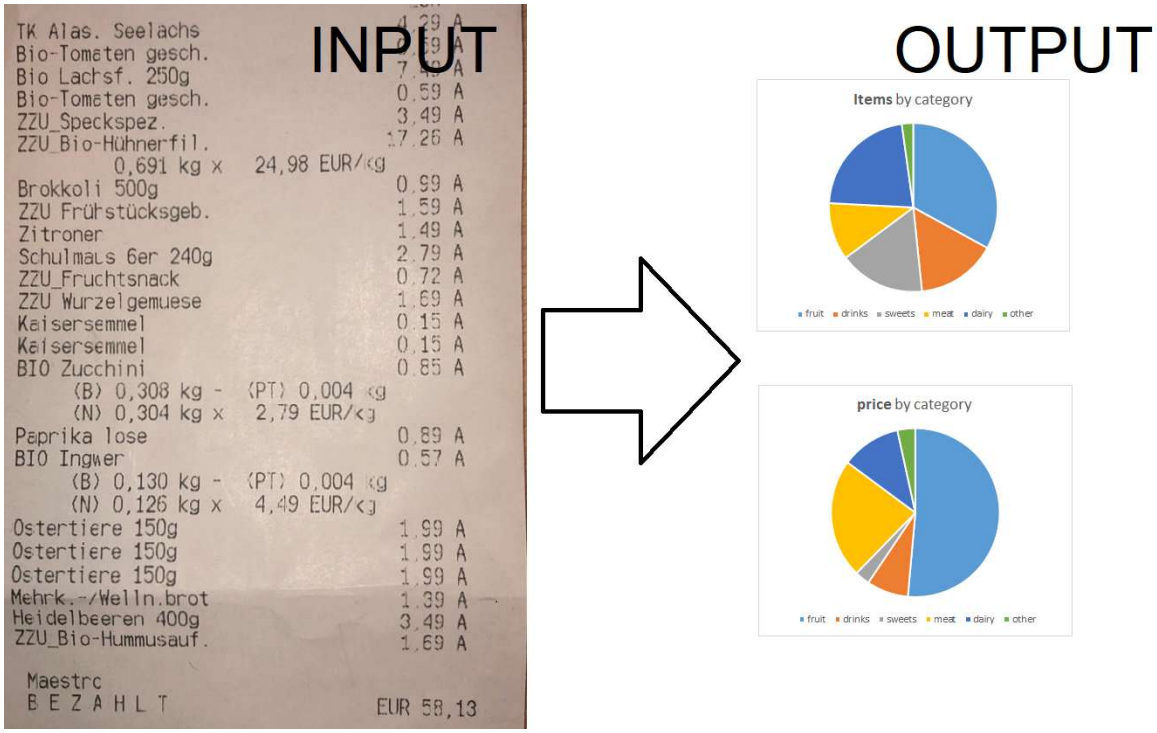
Output: pie chart to see distribution of items categories (fruit/veggy, dairy products, sweets, drinks, etc)

Database in the background will be fed to collect, item, category and price.

With NLP I am planning to identify the categories.

With pytesseract I will extract text from receipt image

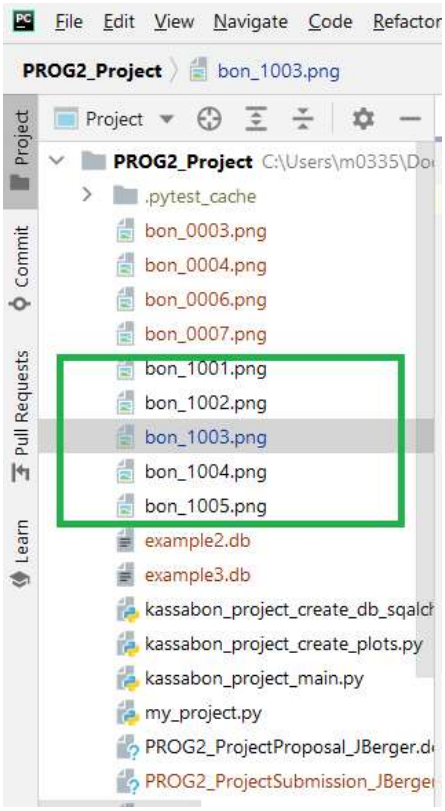
With matplotlib I will do visualisation



I deem this project useful to have an overview on shopping behaviour in terms of preferences on certain categories. We can assume a relationship between shopping behaviour and consummation. I hope to create awareness on shopping and hence consumption behaviour to support creating a healthier lifestyle by adapting groceries.

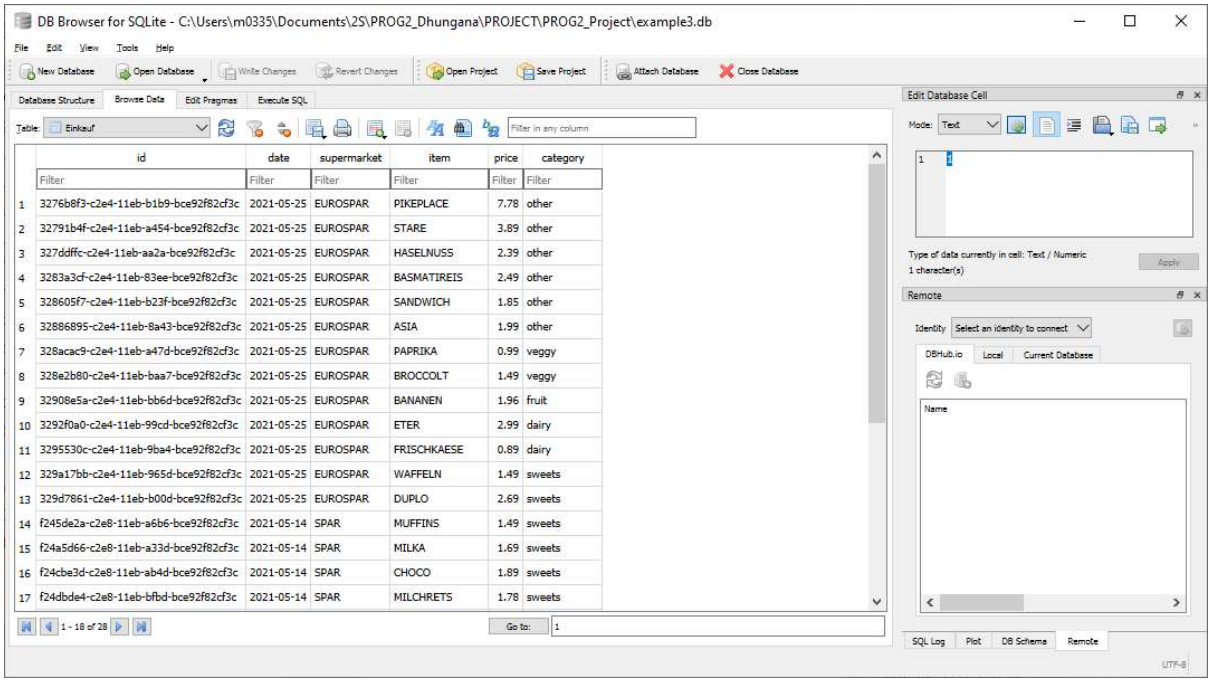
Project Delivery: supermarket receipt visualizer

Input: scanned receipt from supermarkets (bon_1001.png, bon_1002.png, etc)

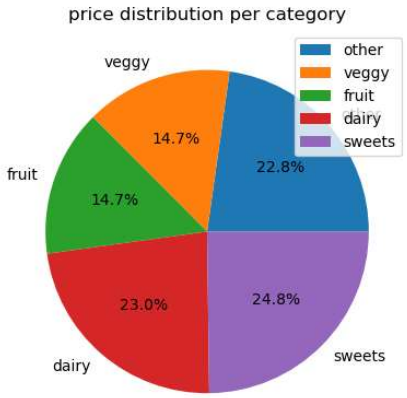


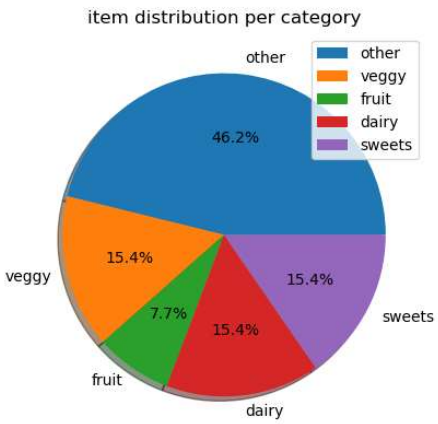
Files:

kassabon_project_create_db_sqlalchemy.py creates database entries



kassabon_project_create_plots.py generates 2 pie plots. The first plot illustrates the distribution of categories over items. The second plot shows the distribution if categories on the prices.





kassabon_project_main.py running all functions like:

1. select_image()
2. extract_text_from_image()
3. define_categories()
4. extract_date()
5. extract_supermarket()
6. confirm_prices()
7. assign_cat_to_item()
8. count_items_per_cat()
9. price_per_category()

my_project.py runs the program addressing the other modules mentioned earlier

my_project.py | kassabon_project_main.py

```python
import kassabon_project_main as kassa
from nltk.tokenize import word_tokenize
from kassabon_project_create_db_sqalchemy import *
from kassabon_project_create_plots import *


text = kassa.extract_text_from_image(kassa.select_image())
words = word_tokenize(text)

#print(text)
#print(words)
#exit()

#creates a dictionary with categories as keys and list of tuples as values, where each tuple has item and price
dict1 = kassa.define_categories()

# words getting distributed into the dict1 for each category
kassa.assign_cat_to_item(dict1,words,text)

#supermarket = kassa.extract_supermarket(words)
#date = kassa.extract_date(text)

#populate database with data from dictionary
enter_data(dict1,kassa.extract_date(text),kassa.extract_supermarket(words))
print(dict1)
#create pie plots
create_plot_item_pie(kassa.count_items_per_cat(dict1),dict1.keys())
create_plot_price_pie(kassa.price_per_category(dict1),dict1.keys())
```

Test_kassabon_project_main.py to test the functions:

1. extract_date()
2. extract_supermarket()
3. confirm_prices()
4. assign_cat_to_item()
5. count_items_per_cat()
6. price_per_category()