Data Science and Machine Learning in Python

Assignment 2 – Market Basket Case

Stephan Weyers

Fachhochschule Dortmund
University of Applied Sciences and Arts

Assignment 2 – General Remarks

Provided files

- W02_Task.pdf (this document, contains the case description)
- W02_data_groceries.csv (the data)
- W02.ipynb (contains some code as starting point for analysis)
- W02_Results.xlsx (template for your evaluation of collaboration in your team)
- W02_Peer Review.xlsx (template for your feedback of other team‘s work)

Instructions

- Work together with the team mates who have been assigned to you (during lecture or via email).
- Read the case description on the following pages
- Solve the problem and make a recommendation to the store manager
- The end product should be one ipynb-file with your Python code and one pdf-presentation. Submit these files in ILIAS (one version per team, only one submission required. If two or more team members upload the presentation, the most recent version counts). Do not put your names in the presentation to enable anonymous peer feedback
- Evaluate the contributions of your own team members in the Excel file
- Submit the populated xlsx file in ILIAS (one version for each individual student)
- After all teams submitted their results, each individual students has to review 2 other team distributions, so that each team gets at least 6 student feedbacks
- By default the grading will be the median assessment of the 6 peer reviews weighted by the contributions of the team mates. The lecturer will (selectively) check the grading. In addition, teams can ask for revision, if they are not satisfied with their grades.

Due dates

- May 10th (23:59 German time) for submitting your solution
- On May 12th the files will be shared for peer review
- May 24th (23:59 German time) for providing your peer reviews

Market Basket Case Description

The manager of a grocery store asked for your help regarding the shelf layout of the shop. Until recently they had about 200 SKUs (stock keeping units: unique item numbers), but the headquarter of the grocery chain advised them to keep only 105 of them and introduce 64 new SKUs. The store manager is in charge of where to place those items. You find the layout of the shop on the next page. The 105 existing items were distributed evenly across the 7 shelves. In general everything can be changed, but the store manager suggests to keep those 105 items at their current position, unless there are very strong reasons for an alternative. Otherwise the customers could be even more confused than they will be anyway due to the change.

You have been given sales data from a different shop that made the transition last year. The layout of this shop is somewhat different, but you can get from the data, which items were purchased together and which not.

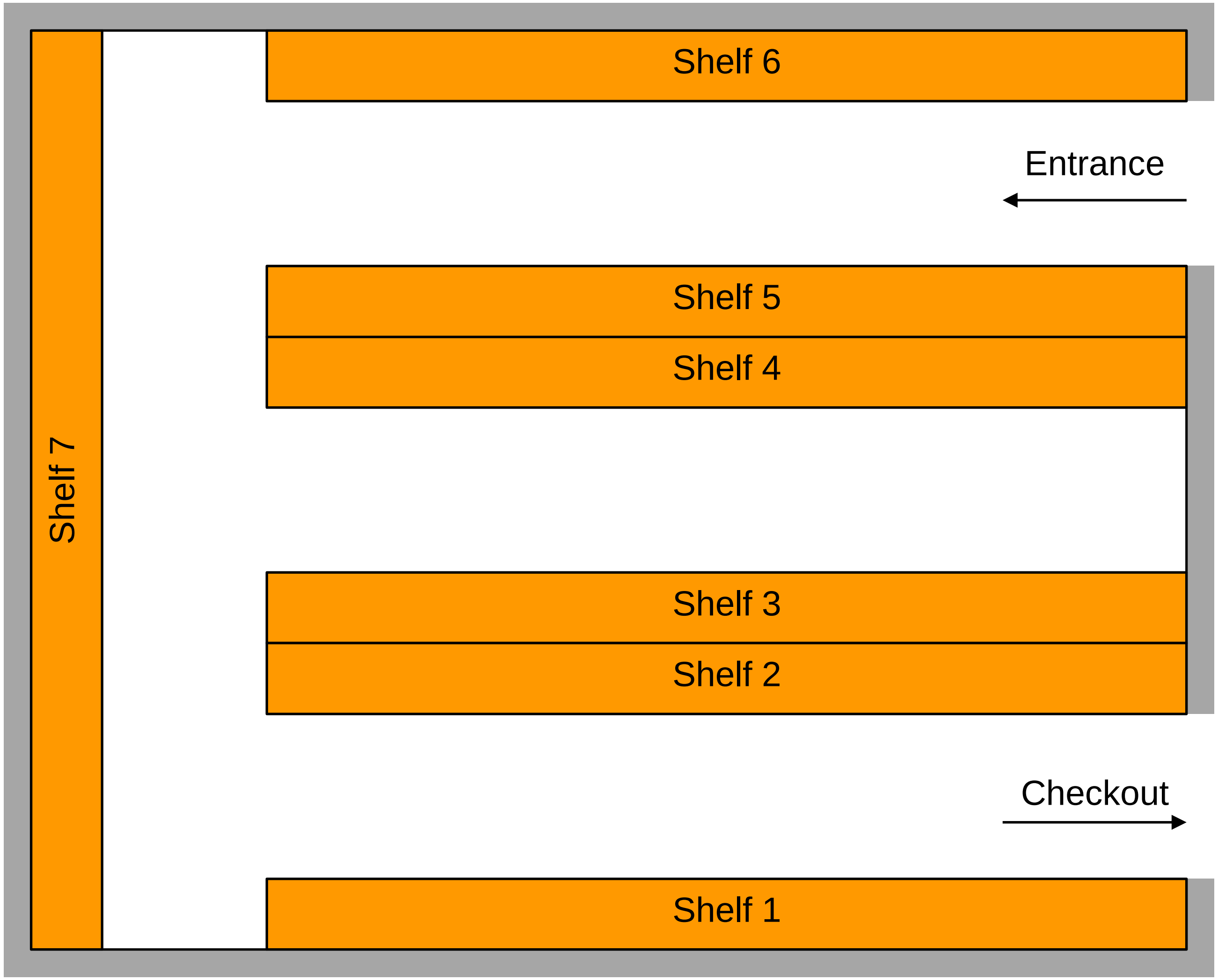
Your task is to perform smart analysis with the data using Python code and come up with a good recommendation for the store manager.

This is an open exercise. Take into account any retail knowledge that you have. Think out of the box and be creative. But be fact-based and explain as much as possible with the data that you received. Don‘t forget to use your common sense and check the real life applicability of your solutions!

The end product should be a presentation and a file with Python code

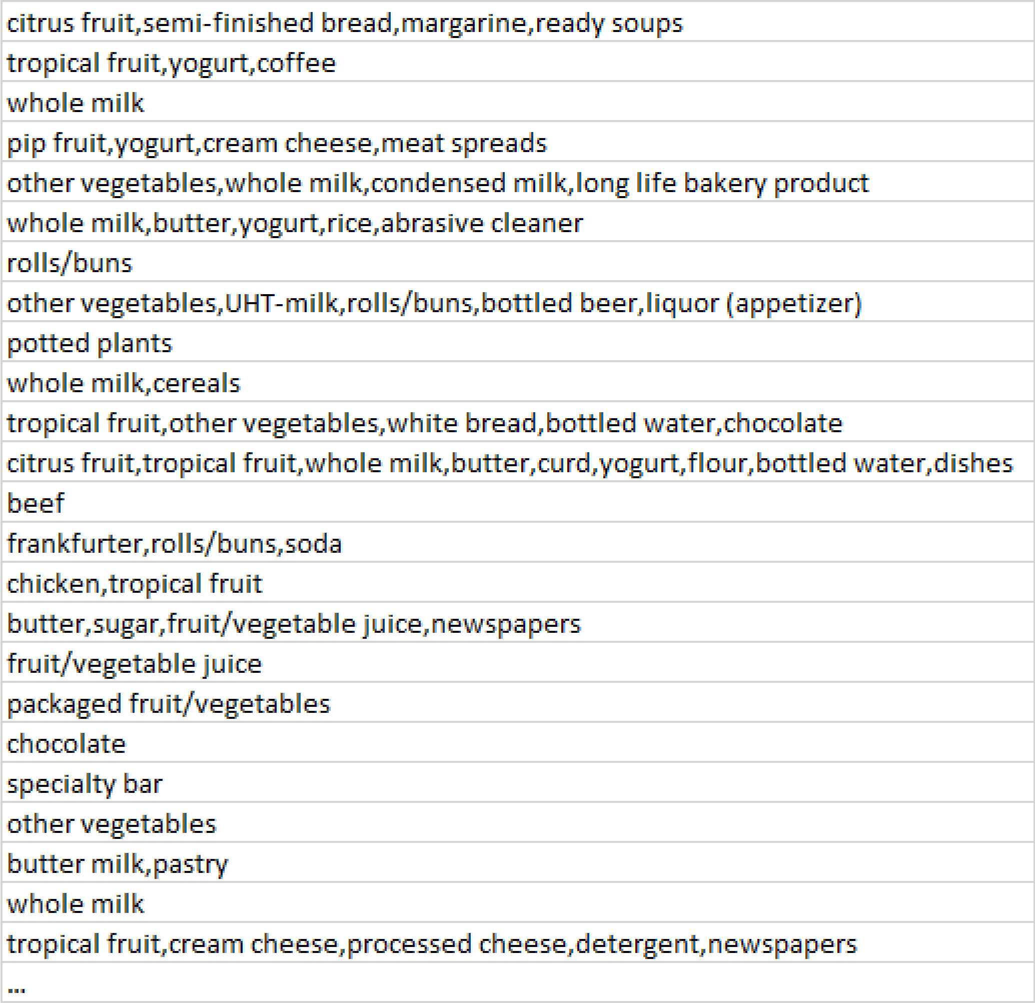
- The key stakeholder and recipient of the presentation is the store manager. Focus on the main insights and the business impact. It should look professional. Explain your findings with arguments and facts
- The code should be understandable. Delete all superfluous parts. Try to make the code as short as possible but still easy to read. Provide meaningful comments. Think of the business analytics team of the headquarter as target group for the code.

Current Store Layout



The data set

- 9835 rows
- Each row is a transaction (customer basket)
- The items purchased in each row are separated by commas
- 169 unique items



citrus fruit,semi-finished bread,margarine,ready soups
tropical fruit,yogurt,coffee
whole milk
pip fruit,yogurt,cream cheese,meat spreads
other vegetables,whole milk,condensed milk,long life bakery product
whole milk,butter,yogurt,rice,abrasive cleaner
rolls/buns
other vegetables,UHT-milk,rolls/buns,bottled beer,liquor (appetizer)
potted plants
whole milk,cereals
tropical fruit,other vegetables,white bread,bottled water,chocolate
citrus fruit,tropical fruit,whole milk,butter,curd,yogurt,flour,bottled water,dishes
beef
frankfurter,rolls/buns,soda
chicken,tropical fruit
butter,sugar,fruit/vegetable juice,newspapers
fruit/vegetable juice
packaged fruit/vegetables
chocolate
specialty bar
other vegetables
butter milk,pastry
whole milk
tropical fruit,cream cheese,processed cheese,detergent,newspapers
...

Existing Items With Given Positions

Shelf 1	Shelf 2	Shelf 3	Shelf 4	Shelf 5	Shelf 6	Shelf 7
shopping bags	soda	rolls/buns	napkins	chocolate	other vegetables	whole milk
newspapers	bottled water	sausage	hygiene articles	salty snack	root vegetables	yogurt
long life bakery product	bottled beer	pastry	softener	specialty chocolate	tropical fruit	butter
cling film/bags	canned beer	brown bread	cleaner	candy	citrus fruit	curd
flower (seeds)	fruit/vegetable juice	frankfurter	male cosmetics	specialty bar	pip fruit	frozen meals
pet care	red/blush wine	pork	abrasive cleaner	chewing gum	sauces	spread cheese
photo/film	white wine	beef	skin care	cake bar	spices	frozen dessert
candles	liquor	white bread	bathroom cleaner	chocolate marshmallow	ketchup	condensed milk
dog food	sparkling wine	waffles	decalcifier	popcorn	tea	specialty cheese
rice	rum	hamburger meat	hair spray	nuts/prunes	canned fruit	frozen potato products
instant coffee	brandy	zwieback	make up remover	artif. sweetener	potato products	finished products
kitchen towels	cocoa drinks	cereals	toilet cleaner	snack products	organic sausage	curd cheese
light bulbs	prosecco	syrup	baby cosmetics	cookware	ready soups	cream
preservation products	liqueur	fish	kitchen utensil	cooking chocolate	specialty vegetables	frozen fruits
sound storage medium	whisky	honey	baby food	pudding powder	salad dressing	frozen chicken

Association rules

$$\mathbf{N = \text{total number of transactions}}$$

$$\mathbf{Support}(X) = \frac{\mathbf{Frequency}(X)}{N}$$

$$\mathbf{Support}(X \rightarrow Y) = \frac{\mathbf{Frequency}(X\&Y)}{N}$$

$$\mathbf{Confidence}(X \rightarrow Y) = \frac{\mathbf{Frequency}(X\&Y)}{\mathbf{Frequency}(X)}$$

$$\mathbf{Lift}(X \rightarrow Y) = \frac{\mathbf{Support}(X \rightarrow Y)}{\mathbf{Support}(X) \cdot \mathbf{Support}(Y)}$$

$$\mathbf{Leverage}(X \rightarrow Y) = \mathbf{Support}(X \rightarrow Y) - \mathbf{Support}(X) \cdot \mathbf{Support}(Y)$$

In general, items X and Y with high

- $\mathbf{Support}(X \rightarrow Y)$
- $\mathbf{Confidence}(X \rightarrow Y)$
- $\mathbf{Lift}(X \rightarrow Y)$
- $\mathbf{Leverage}(X \rightarrow Y)$

should be placed closely together.

$\mathbf{Support}(X \rightarrow Y)$ measures the absolute appearance of X and Y together. Can be misleading, e.g. if X and Y are both fast moving items

$\mathbf{Confidence}$ overcomes this shortcoming and provides a more complete picture

$\mathbf{Lift}$ and $\mathbf{Leverage}$ compare the actual share of $X\&Y$ with the expected share.
$\mathbf{Lift} > 1$ or $\mathbf{Leverage} > 0$ means positive correlation.
$\mathbf{Lift} < 1$ or $\mathbf{Leverage} < 0$ means negative correlation.
It holds $-1 < \mathbf{Leverage} < 1$

Source: https://learn.datacamp.com/courses/market-basket-analysis-in-python

Grading criteria

Grading of own team task

- At least 6 students from other teams will assess your submissions anonymously (so please don't write your names in the pdf and ipynb file). You will see their aggregated feedback incl. the comments and suggestions.
- The median of their votes is the default grading.
- Stephan Weyers will selectively check the grading. Unsatisfied teams can ask for explicit checking their grades, but this could also lead to a downgrade!

The evaluation criteria are transparent in advance

For more details check file Assignment_02_Peer Review.xlsx

Python code evaluation criteria

- Structure of code
- Comments
- Code is working
- Complexity of code

Presentation evaluation criteria

- Business impact
- Design and formatting
- Communication
- Methods, analysis and reasoning

Grading of peer review process

- The peer review is part of the assignment grade.
- You have to provide marks for all criteria and give a decent amount of valuable constructive feedback. However, you don't have to comment each criterion or write novels. Just try to help the other team
- The deviation of your evaluation from the median assessment will be taken into account, so that you are incentivized to do give a fair mark that is just right.