Coding challenge for Babbel Strategic Data Analyst

How to submit

Please perform your analyses in a Jupyter notebook, using Python 3 and a recent version of Pandas. For charting you can use any visualisation library you like, including Pandas built-in charting, but please use Pandas for analysis.

Use text cells in your notebook to explain what you are doing or any assumptions you make. If you find a descriptions ambiguous (which we tried our best to avoid), make a sane assumption and document it.

Assume your target audience for charts is the sales department, and not analysts or data scientists. **Make sure to use appropriate chart types to be as clear as possible!** Good charts have titles, axis titles, etc! Whenever we ask you to create a chart, feel free to create more than one if (and only if) you believe this makes more sense.

We do care about clarity and legibility of the Python code that generates the analysis! Use comments if you feel something needs to be explained.

Make sure you solve all five tasks.

Please read the task descriptions carefully. They are dense but complete.

None of the tasks requires you to write a lot of code if you think about the data carefully and know the right tools in Pandas. Familiarity with pivot tables in Pandas helps a lot.

Don't hesitate to get in touch if you have questions.

Please submit two files:

1. An HTML export of you Pandas notebook (File → Download as → HTML). Make sure we see your **code and the charts** you created. **We will look primarily at this file, so make sure we see all you want us to see!**
2. The original Jupyter notebook.

The tasks

We have a record of sales of our candy wholesale company (`sales.csv`), with columns `customer_id, date, market` and `flavor`. Each row indicates that a customer has bought a container of candy of a certain flavor on a certain day. Our candy is very popular, so you can only buy one container per order. Each customer has an ID and belongs to a certain market region (column `market`).

We also have a price table (`prices.xlsx`) with prices per flavor and market. Unfortunately, the business gave us this as an Excel file.

Both data sets are clean and contain no errors or missing values.

Please perform the following analyses:

1 – Warm-up exercise

The business wants to know the revenue ***over time*** for each market and each flavor. Prepare a set of meaningful and readable plots.

2 – Customer lifetime revenue (CLR)

The business wants to understand how much money we make per customer in each market on average.

Create a plot that breaks down the average revenue per customer for each country **and** shows which flavor contributes how much to that revenue.

For example:
In market FRA we make on average 10 € per customer. Of the 10 €, 2 € come from Apple, 3 € come from Banana, etc.

3 – Flavor preferences in first vs following purchase

The business suspects that customers' first purchases follow different patterns than following purchases.

Create plots that show for each country the customers' flavor preferences for their first purchase versus their preferences for following purchases.

- Note that we are not interested in absolute sales or revenue, but in preference of flavor by country!
- We want to see "first" versus "following" purchase, and not "first" versus "second", "third", etc! “Following” means any purchase that is not the first one.

4 – First versus following in CLR

Create a plot like the one in task 2 showing average revenue by customer. This time, break it down by market (as in task 2) and by the contribution of the first purchase versus the contribution of all following purchases to the average revenue.

Example:

A customer has the following purchases:

- Jan-01-2018 – Apple – 5 €
- Jan-12-2018 – Banana – 3 €
- Feb-07-2018 – Coffee – 4 €

His first purchase contributes 5 €, his following purchases contribute 7 € to what we’ve earned with this customer.

- Here, we are not interested in which products customers actually bought in the first and following purchases! Only the “first vs following” split matters.
- For a customer with only one purchase, the contribution of the following purchases is simply 0 €.

Make sure you choose an appropriate visualisation.

5 – Transitions

We want to analyse how people switch between flavors. Visualise the probability of flavor transitions, i.e. of a customer buying flavor B after buying flavor A for all pairs of flavors. If a customer’s last purchase has occurred more than 6 months before the latest date in the dataset, we consider him as “churned”, and we record a transition from his last flavor to a special state called “churned”.

- A customer who bought “Apple, Banana, Apple, Doughnut” contributes 3 transitions: Apple → Banana,

Banana → Apple,
Apple → Doughnut

- If he bought Doughnut more than 6 months before the latest date in the dataset, there is an additional transition: Doughnut → churned
- A customer with a single purchase X contributes a transition “X → churned” if he bought X more than 6 months before the latest date in the dataset.

We look at this globally, so you don’t need to break down by market. Again, choose an appropriate visualisation.