



Assignment for BI Developer Intern

21.2.2023

Solution zip-file and running the programs generating the solution

All files needed for producing assignment solutions with Python are packed in a zip-file. Extracting this zip-file creates a folder with one program for each of the assignment tasks. It is assumed that the working directory is set to be the same directory as the program files are located in. Zip-file includes the data for the assignment in a separate folder. Programs will produce html and svg files as output and those will be generated to subfolder "output".

Connection to a Postgres database is configured in `wolt_config.py` and it should be updated first.

Visualizations are made with Plotly and output is opened in browser and also saved as vector graphics in svg file format. For handling svg-files kaleido package needs to be installed (`pip install -U kaleido`).

Reading assignment data from the csv-files to Postgres database

I used Python for creating Postgres tables and reading assignment data to the database tables. Data is read as is, except for the empty strings in csv-files which were converted to Postgres Null values. After tables are populated Postgres indexes are created to speed up SQL queries.

Source Code file: `wolt_read_csv_to_pg.py`

Config file: `wolt_config.py` (database connection and folders)

Task 1

Profitability of each purchase using SQL

The program for task 1 reads the SQL query from a text-file, prints the query text to console and runs the query. Query results are read directly to a Pandas dataframe. This dataframe is saved as Pickle-file if it has not been created before to speed up re-runs with same SQL-query and data in Postgres when development of the Python part of the solution to the assignment was made.

In the next step the program generates html-files with CSS styling to show the first and last 20 rows of the table with 216288 rows. These html-files will be generated in the output folder defined in `wolt_config.py` -file.

Source Code file: `wolt_task_1.py`

SQL file: `task1_final.sql`

First 20 rows from the table (task_1_first_20_rows.html):

	purchase_id	country	date_received	number_of_items	sales_per_purchase_ex_vat_eur	margin_per_purchase_ex_vat_eur	margin_perc
0	5d5d0dd546	HRV	2022-12-07	78	1350.61	1214.09	89.9
1	de7423e2dc	HRV	2022-12-29	43	772.58	694.46	89.9
2	c8fcd2cfaf	HRV	2022-12-21	12	619.10	559.04	90.3
3	8a0f9ca69a	HRV	2022-12-15	24	602.27	550.16	91.3
4	dc470e56cb	HRV	2022-12-22	55	588.39	529.17	89.9
5	c5d21e95fc	HRV	2022-11-08	44	564.81	513.32	90.9
6	b4c0e5e97a	HRV	2022-12-10	36	537.41	486.14	90.5
7	60eb2aab07	HRV	2022-12-16	28	474.98	427.49	90.0
8	3084993d8f	HRV	2022-12-23	18	463.62	423.64	91.4
9	8a75804d91	HRV	2022-12-17	36	459.74	418.17	91.0
10	450bcdcd39	HRV	2022-12-18	30	455.77	411.47	90.3
11	db6a9e7980	HRV	2022-12-15	32	428.14	385.82	90.1
12	271385c981	HRV	2022-12-23	7	419.95	377.61	89.9
13	873f81f841	HRV	2022-12-05	28	414.49	374.35	90.3
14	12a05f2fa2	HRV	2022-12-04	31	413.65	371.54	89.8
15	9163dc581f	HRV	2022-12-14	22	410.69	369.97	90.1
16	c5135fbb18	HRV	2022-12-01	10	404.74	367.30	90.7
17	d2efc4a249	HRV	2022-12-10	24	406.63	366.21	90.1
18	9e50322296	HRV	2022-12-02	27	405.72	364.73	89.9
19	b2ffcee675	HRV	2022-11-10	18	397.46	357.19	89.9
20	6a02ee5a35	HRV	2022-12-28	11	400.81	356.38	88.9

Last 20 rows from the table (task_1_last_20_rows.html):

	purchase_id	country	date_received	number_of_items	sales_per_purchase_ex_vat_eur	margin_per_purchase_ex_vat_eur	margin_perc
216268	2cf6d606d0	FIN	2022-12-29	25	43.41	-14.55	-33.5
216269	28b6bbdcc3	FIN	2022-12-24	23	59.45	-15.33	-25.8
216270	e308dec23f	FIN	2022-12-24	22	46.65	-15.44	-33.1
216271	9602a99710	HUN	2022-10-28	11	32.17	-15.81	-49.1
216272	2ba5fa0dbc	FIN	2022-12-25	11	22.68	-15.81	-69.7
216273	eafea1b2b2	GRC	2022-08-23	4	24.12	-16.64	-69.0
216274	8a79e7d030	MLT	2022-09-10	18	62.30	-16.68	-26.8
216275	0be936cf36	HUN	2022-11-15	8	28.88	-16.98	-58.8
216276	42816b92d0	FIN	2022-12-24	15	30.75	-17.78	-57.8
216277	03182c6e9d	GRC	2022-07-18	5	26.57	-17.88	-67.3
216278	524abd524d	FIN	2022-12-28	18	47.74	-18.34	-38.4
216279	9e8c67d697	FIN	2022-12-24	14	28.20	-18.62	-66.0
216280	322b142bd3	NOR	2022-12-29	19	29.32	-20.65	-70.4
216281	6720695183	MLT	2022-10-09	7	50.22	-21.24	-42.3
216282	71bc43cd09	HUN	2022-11-17	12	43.32	-25.48	-58.8
216283	7db4c899e4	FIN	2022-12-26	20	43.20	-26.41	-61.1
216284	c0afa8a722	FIN	2022-12-26	25	42.88	-27.69	-64.6
216285	5b51e3d50a	NOR	2022-12-27	32	42.36	-32.48	-76.7
216286	617b023eca	HUN	2022-10-11	20	72.20	-42.46	-58.8
216287	9236c8d40d	SVK	2022-12-26	37	7.03	-42.56	-605.1

Task 2

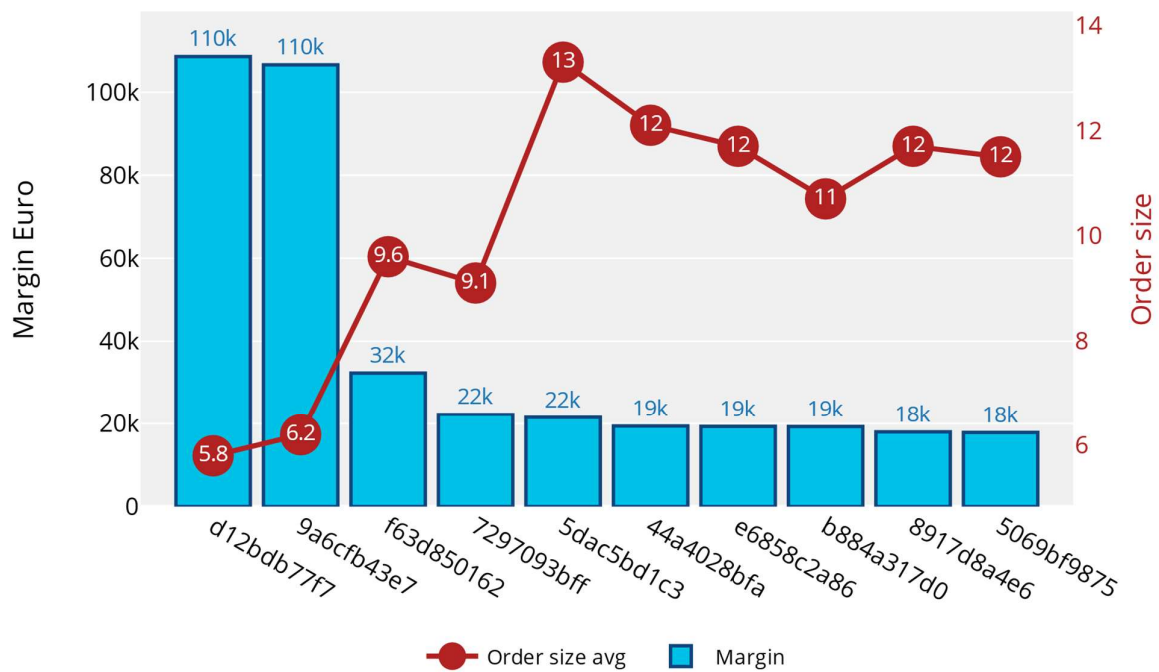
Top 10 venues with highest margin and their average order size

Task is solved similarly like task 1 with Python, Postgres and Pandas. For visualization Plotly is used.

Source Code file: wolt_task_2a.py

SQL file: task2-a_final.sql

Top 10 venues with highest margin and their average order size



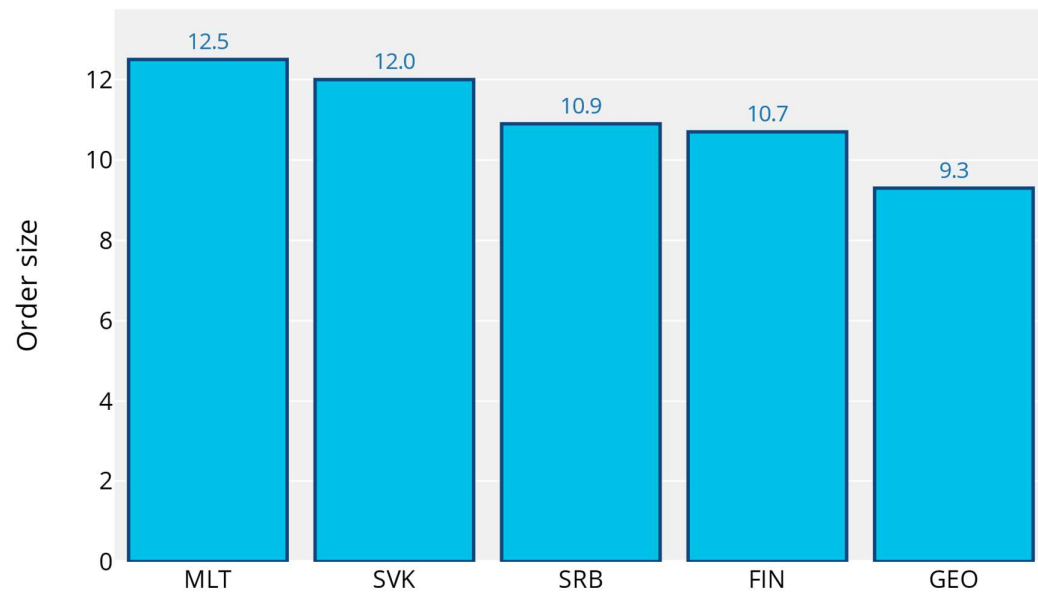
Top 5 countries by average order size

Task is solved similarly to previous task with Python, Postgres, Pandas and Plotly.

Source Code file: `wolt_task_2b_1_avg_order_size.py`

SQL file: `task2-b_avg_order_size_final.sql`

Top 5 countries by average order size

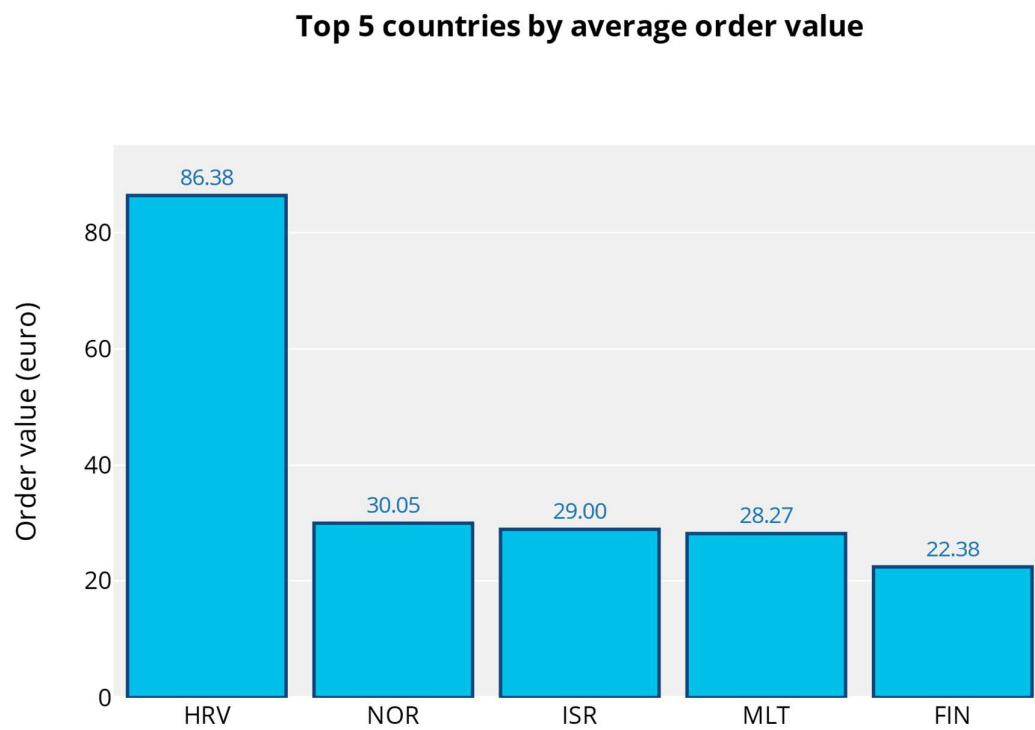


Top 5 countries by average order value

Task is solved similarly to previous tasks with Python, Postgres, Pandas and Plotly.

Source Code file: `wolt_task_2b_2_avg_order_value.py`

SQL file: `task2-b_avg_order_value_final.sql`

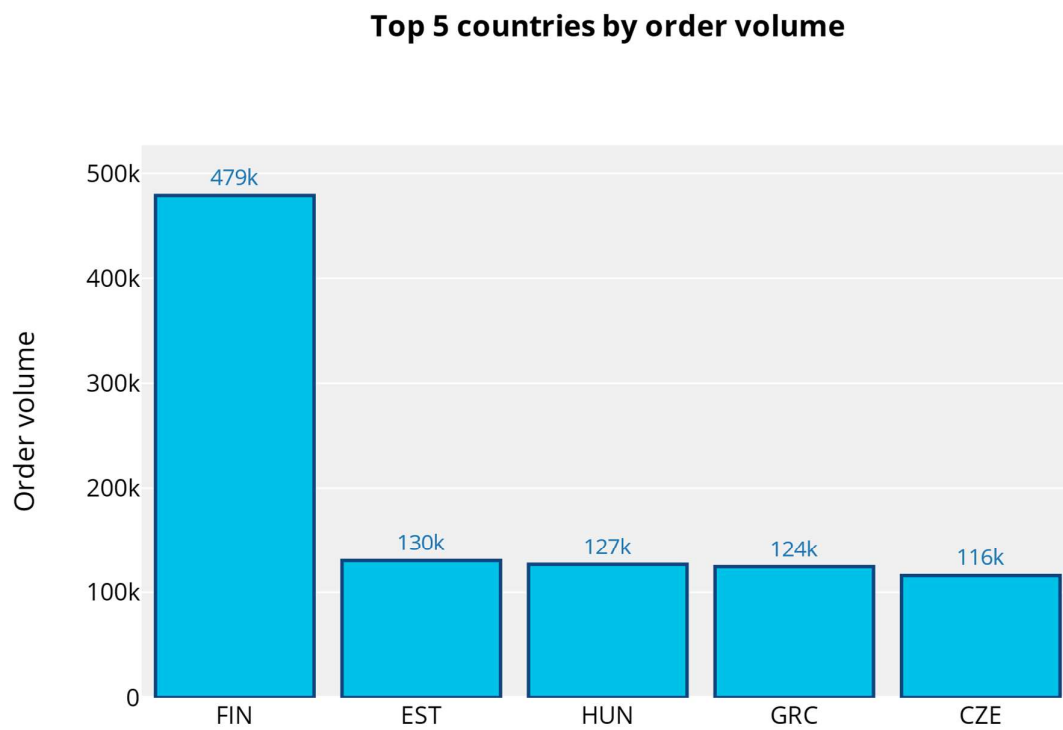


Top 5 countries by order volume

Task is solved similarly to previous tasks with Python, Postgres, Pandas and Plotly.

Source Code file: `wolt_task_2b_3_order_volume.py`

SQL file: `task2-b_order_volume_final.sql`



Monthly cumulative Woltwide margin

Task is solved similarly to previous tasks with Python, Postgres, Pandas and Plotly.

Source Code file: wolt_task_2c.py

SQL file: task2-c_final.sql



Additional questions about assignment

Assumptions about the data when calculating gross margin

Sales price of the item was converted to euro with an exchange rate calculated based on the items cost available in both the local currency and euro. Actual currency rate at the time of sale can be different than the rate used in calculation of gross margin. Error from using wrong exchange rate could be especially substantial in a country with very high inflation rate and in case the cost prices in item data are from long time ago before the time of purchase.

Above way of handling exchange rate also assumes that cost in local currency is in the same currency that is the currency the purchase was made in. This assumption might maybe break naturally in the case where multiple currencies are in use at a venue.

Problems in data and how they were solved

Some purchase items did not have a cost in local currency resulting division by zero error. This was solved by a CASE clause in the SQL query.

Some purchases had suspiciously high or low negative margins as can be seen from the output of task 1. This can point to mistakenly entered data into the system about the cost of the item or sales price. For solving this problem checks on price data in item table followed by a lookup for correct alternative venue-item-price could have been a solution but this was not implemented.

Additional data sources or business information to improve the solution

For calculation of profitability of the business one important data missing is the information on the individual customer making the purchase. This data is needed for calculating how frequently customers make purchases and makes it possible to calculate the customer acquisition costs in relation to frequency of purchase of a given customer or customer group.

Data on delivery cost of each purchase is another piece of information that would help to analyze profitability and get insight how to improve it.

There might also be some fixed cost per country, city and venue that should be taken into consideration when analyzing profitability and calculating metrics to help making business decisions that improve the bottom line of the business.

In addition to the above financial performance measurements there many important non-financial performance metrics that could fit in to set of most important metrics to follow. For example, customer satisfaction ratings and time from order to delivery.