Users' Analysis

Technical aspects

Python was used to manipulate the dataset, while the presentation was created in MS Power BI.

In order to run the analysis script, 2 files are needed: 'efood.py' and 'requirements.txt'. After, the code can be run using Python 3.10 and by opening the terminal and executing the following commands:

```
pip3.10 install -r requirements.txt

python3.10 efood.py
```

Explanation of processing

The analysis was divided into 3 steps: cleaning data, transforming data, and visualizing the results.

During the 1st step, 3 quality checks were made:

1. Existence of '*nulls'* across dataframe. In the case where '*nulls'* were existed, the respective records were removed.
2. Duplicates across all attributes were removed.
3. Duplicates on '*order_id*'. Since '*order_id*' is unique for each order, possible duplicates were deleted.
4. Negative values on '*amount'*. These records were excluded to keep only the valid orders.

Moving forward to 2nd step, 4 transformations were applied:

1. '*weekday'* attribute was created using '*order_timestamp*' so to find the exact days of week that an order was placed. The reason for this action, was to visualize later the behavior of users for each day of week.
2. '*order_timestamp*' wasn't able to be used in timeline chart. As a result, it was divided to keep only the '*date'* details.
3. '*amount'* was transformed to number with only 2 decimals.
4. '*paid_cash*' information was in the format of True/False. To be used later in the visualizations, these values were changed to the explicitly words of '*card*' and '*cash'*

Apart from these transformations, one more table was created. '*summarized_info'*, contains the number of orders, the total amount and the number of days that each user ordered for each cuisine.

After completing cleaning and transformations steps, the 2 tables (transformed '*order.csv*' and '*summarized_info.csv*') were loaded to MS Power BI. Moreover, one last measure created in the '*summarized_info*' table. Using DAX, users were divided into 6 buckets based on the number of days they ordered during January ('*Bucket_label_order_days_6_size_bin*').

So, to present the results of the analysis, 2 slides were created. Both the overall performance of these 4 cuisines in these smaller cities of Greece and the performance of 'Breakfast' explicitly were visualized. Finally, at the last slide the most important findings and insights were noted.