

Analysis and forecast of daily revenue and receipts in restaurants

Data Science Lab Project Report

Francesco Paolo Luigi Caruso [906416], Alberto Porrini [826306], Roberto Ferrari [852220], Andrea Muscio [889901]

Informations on codes and datasets provided

In the "CODES AND DATASETS.zip" file, you can find:

- R code used for the forecasting part, named "FORECAST MODELS.R".
- Python code for the preprocessing and exploratory analysis part, named "DATA PREPROCESSING - EXPLORATORY ANALYSIS". As specified later, it is recommended to use the code directly on Colab, accessible at the link: https://colab.research.google.com/drive/1fHx7VPkf3tAbkMIN_x6b6K5HzStNtOvD (which is what was actually used).
- In the "STARTING DATASETS" subfolder:
 - The original dataset named "ristoranti.csv".
 - 2 datasets created with pivot tables on Excel, corresponding to revenue ("ristoranti_fatturato.csv") and receipts ("ristoranti_scontrini.csv").
- In the "ORGANIZED DATASETS" subfolder contains datasets where the 2018 monthly part as well as the last rows (which had null values) have been removed. In particular in this folder there are:
 - 6x2 datasets created with Python code, each containing data for a specific restaurant. One dataset contains revenue data, and the other contains receipts data. For example, for restaurant R000, the revenue dataset is named "fat_r000.csv," and the receipts dataset is named "scon_r000.csv." For other restaurants, you can simply change the number in the dataset name accordingly.
 - 2 datasets containing all the restaurants together, one for revenue ("fatturato.csv") and another for receipts ("scontrini.csv")

The code used for preprocessing and exploratory analysis is available in the Python notebook. However, it was created and used on Google Colab. You can view and run it directly at the following link: https://colab.research.google.com/drive/1fHx7VPkf3tAbkMIN_x6b6K5HzStNtOvD . To use it directly on Google Colab, you can place your datasets in your Google Drive and modify the path indicated in `pd.read_csv` command. Alternatively, to use it on Jupyter, you should remove the line `drive.mount('/content/drive')` and provide the dataset path directly in `pd.read_csv` command.

Additionally, note that part of the code is used to print (always on Google Drive) the datasets that are in the "ORGANIZED DATASETS" subfolder.

The code is divided in sections for revenue and receipts and each of them is divided in three subsections: time series plots, boxplots with pre and post COVID-19 data and boxplots with post COVID-19 data.

For the forecast models, the R code is provided as an attachment. The code provided is specific to a single restaurant (specifically R000). You can change the restaurant by modifying the .csv files to be loaded and also replacing "R00_" with the corresponding number in the code lines marked with "*****CHANGE RESTAURANT NUMBER HERE*****".

The code is divided into two parts: one for the analysis with pre and post COVID-19 data and one for the analysis with only post COVID-19 data. Each part is further divided into REVENUE and RECEIPTS, and within those, SARIMA and TBATS models are used separately.