Assignment 2
Predict future sales

Wojtek Kowalczyk
`wojtek@liacs.nl`

01-11-2022

Description

In this assignment you are supposed to take part in a Kaggle competition: Predict future sales. You will need to use various time series forecasting algorithms and data visualization techniques to accurately predict sales of a software company.

Your tasks are as follows:

1. Register your team on the Kaggle platform.

2. Download the data and study the problem description.

3. Perform exploratory data analysis: generate a number of plots, collect simple statistics (Correlations? Outliers? Missing values? Errors?), get some intuitions about what is in the data.

4. If needed, clean/transform/sample your data, so you can apply several existing forecasting algorithms to the data.

5. Additionally, over 1-2 weeks you will receive another task(s) related to this challenge - evaluation on an alternative error measure (e.g., instead of using MSE use an asymmetric measure that punishes "under-predictions" (empty shelves) more than "over-predictions"), create a short "management summary for business people" (e.g., in a form of a PowerPoint presentation).

6. Write a report, documenting your work and findings in detail.

Deliverables

Your submission on Brightspace should consist of (not zipped):

1. Report in *.pdf* format (max 7 pages)

2. Jupyter notebook containing your exploratory data analysis and modelling

You must also submit your results on Kaggle (you need to include the name of your team in the report, we will check the leaderboards)

Deadline: 9th December 2022, 23:30

Suggested material

As a starting point for generating predictions it is strongly recommended that you study:

- Time series forecasting basics (Kaggle)

- Methods for time series forecasting