Data Science case study

"We use Machine Learning and Data Science to make sport accessible to the many."

This case study is to be performed using your prefered ML & DS tools and libraries. We accept the following submission file types: .ipynb, .py, .html.

1. The data

train.csv.gz & *test.csv.gz* files contain weekly store-department turnover data (*test.csv.gz* does not contain the turnover feature).

For confidentiality reasons the turnover values are rescaled.

bu_feat.csv.gz file contains some useful store information.

2. Preliminary questions & data exploration

a. Which department made the highest turnover in 2016?
b. What are the top 5 week numbers (1 to 53) for department 88 in 2015 in terms of turnover over all stores?
c. What was the top performer store in 2014?
d. Based on sales can you guess what kind of sport represents departement 73?
e. Based on sales can you guess what kind of sport represents departement 117?
f. What other insights can you draw from the data? Provide plots and figures if needed.

3. Multi-step forecasting

In stores many decisions are made by managers at the department level.
In order to help store managers in making mid-term decisions driven by economic data, we want to forecast the turnover for the next 8 weeks at store-department level.

a. Build an estimator able to predict the turnover of *test.csv.gz* data.
b. Which metric would you use to evaluate your predictions and why?
c. Build an end-to-end pipeline from data processing to results exposition.