

Description of supporting materials

by Bilgin Sherifov

Note: Unfortunately, of all the work I've done at OTB, I can share with you only these very basic things.

1. **cv_and_job_description:** it includes my CV and an informal description of the some of the works I've done at On The Beach.
2. **marketing_reports:** these are an example of stand-alone application that runs daily and sends emails to various stakeholder with various plots of daily and hourly OTB performance metrics on Google Paid Search. This was writhe in 2017. I've also deployed an application for ordering of items on the search results when OTB has no prior information about the visitor. This second application is much more complex, but the style and structuring is similar to the one here.

3. **item_feature_visualisation:** these are interactive visualizations of items vs. features data sets. In these examples hotels are items, and airports or departure dates are features. You can run each by going to

Direct Line\supporting material\item_feature_visualisation
and from there in the command prompt running this command:

```
python -m http.server --cgi 8000
```

Then on your browser go to:

<http://localhost:8000/>

Navigate either to

\departure_months__hotels

or to

\ airports__hotels

Then click on the “*Choose File*” button and load the JSON document.

Size of the hotel circles represents how many times that hotel was booked, while size of the airport or departure_month circles represents how many bookings were from those airports or for those months. The more links there are between a hotel and an airport (or departure month), the shorter and the thicker the link between them. Some circles have no links, because they were removed to make the visualization lighter.

This is a nice visualization of how feature and item cluster. It uses simulations of physical electrostatic and gravitational forces to achieve this.

4. **econ_simulation_hidden_state_space:** This (*econ_torch.ipynb*) is a simulation of a non-linear dynamical system, where some of the variables are never observed directly.

The goal is to come up with methods to estimate these hidden time-varying variables. This is meant to simulate the impact TV ads have on sales figures, in the presence of other macroeconomic or business drivers. The goal is to measure both the short-term and the long-term impact.

I'm trying various time-series and other methods and comparing how well are they able to capture the true impact.

There is also a PowerPoint presentation that is based on these results

(MancML__Bilgin_Sherifov__2019_03_14.pptx) that I had presented at MancML event.

5. **econ_regression_time_varying:** this is just one of many stages in a major project I have developed about assessing impact of TV ads on sales figures. The final system consists of various models, one of which is presented here. The bulk of the work though, is about getting the right inputs, in the right format, cleansing them, correcting for known abnormalities etc.
6. **customer_value_challenge:** this includes a notebook and its PDF version of a submission of mine to a data challenge by another company that I have applied for.
The challenge is to investigate the dataset, performing a simple exploratory analysis, and then a more focused and advanced analysis. Some of the things they asked were:
 - Can you identify any relevant groups of customers? From your perspective, what are the most relevant variables that characterize those groups?
 - Do you notice any patterns in the data?
 - From your perspective, which are the most valuable customers?
7. **telemetric_challenge:** this includes a notebook and its PDF version of a submission of mine to a data challenge by another company that I have applied for.
The challenge is to identify accidents and their severity, given GPS and Accelerometer time series from various vehicles.