Shelf Engine Data Challenge

Overview

Thank you for your interest in Shelf Engine. This Data Science challenge aims to measure your problem solving and machine learning skills. Please code in Python. Expect us to run your code. If you need any non-standard libraries for your analysis, please point to public repositories or include them along with your analysis. Please refer to the **Output Files Expected** section for details on submission materials.

Data Overview

The attached dataset shows a public, transnational data set on transactions from 12/01/2010 and 12/09/2011 for a UK-based and registered non-store online retail company. Transaction level records such as these are similar to the data we work with at Shelf.

Columns:

InvoiceNo: Invoice number. Nominal, a 6-digit integral number uniquely assigned to each transaction. If this code starts with letter 'c', it indicates a cancellation.

StockCode: Product (item) code. Nominal, a 5-digit integral number uniquely assigned to each distinct product.

Description: Product (item) name. Nominal.

Quantity: The quantities of each product (item) per transaction. Numeric.

InvoiceDate: Invice Date and time. Numeric, the day and time when each transaction was generated.

UnitPrice: Unit price. Numeric, Product price per unit in sterling.

CustomerID: Customer number. Nominal, a 5-digit integral number uniquely assigned to each customer.

Country: Country name. Nominal, the name of the country where each customer resides.

Source: http://archive.ics.uci.edu/ml/machine-learning-databases/00352/

Data Challenge

The store is interested in maintaining the right inventory given its historical sales. Build a model that predicts sales quantities for the 7 days from 11/27/2011 - 12/3/2011 (Sun - Sat). To help the company prepare for the worst case, please focus your predictions and the evaluation of your model on only the top three selling items.

- Include any exploratory data analysis and evaluations that support your findings.
- Please give us your estimated number of sales for these three items, broken out by country.
- Please briefly discuss why you chose your approach, its strengths and drawbacks, and how you might approach this problem if you were to research it further.

We define top selling items as items which had the greatest total sales over this week across all countries.

Output Files Expected:

1. {Your Name}***_report.pdf***: A short memo on your analysis of the above questions.
2. {Your Name}***_code***: A set of executable scripts (or a Jupyter notebook) on a local machine
3. {Your Name}***_result.csv***: The sales quantity forecast of the items described above

Evaluation Criteria:

We present this data challenge to you to help us better understand your approach to Data Science problems like those we face at Shelf Engine. We are most interested in your reasoning, thought process, and coding proficiency, which will be calibrated by your level of experience.

We also value your communication skills in writing which would show us your ability to convey your thought process to others in the team.

Enjoy the challenge and happy modeling!

— Shelf Engine Data Science team