IE 575: Predictive Analytics
Final Project Description (20% Weight)
Due date: First Day of Exam Week

The intent of this project is to make you familiarize yourself with a data set and use the techniques learnt in this class to undertake extensive data analytics. Though we have given broader guidelines, please do use your imagination and creativity to blow us off. So this is an open ended project. We will use the "online retail data set", which is sent along with this project description; this data relates to a UK based non-store online retail agency. This data which is from UCI Machine Learning Repository is described below:

This data represents online transactions. A transaction in this data set is of the form (*InvoiceNo, StockCode, Description, Quantity, InvoiceDate, UnitPrice, CustomerID, Country*).

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: Invoice 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.

Our objective in this project is to analyze the data to help the organization to improve their sale (total revenue). In an undergraduate course taught by us at PSU UP campus, we asked the students to analyze retail data (similar but not exact). A sample project report is sent to you along with this for your reference.

Though one can do several things with this data, we feel you should have at least the following accomplished in your project:

1. Show an understanding of the data through descriptive statistics and visualization
2. Use clustering to find interesting groups
3. Use at least 3 methods to predict sales
4. Compare the performance of these methods
5. Recommend strategies for the organization
6. Produce a report

Note: As the number of items are more than 1500, through your initial analysis narrow down your prediction analytics to the top 10 best sellers.

Deliverables: Project Report, A 20 minute recoded presentation, and ppt slides. Please submit your code. We are not providing a report template- use your imagination and show some insights.

Reference:

Daqing Chen, Sai Liang Sain, and Kun Guo, Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining, Journal of Database Marketing and Customer Strategy Management, Vol. 19, No. 3, pp. 197â€“208, 2012

(This link may not work directly. Please copy this and use PSU account to go to Google Scholar and paste it in scholar search)