Capstone 2 Project Proposal:

I found a data set that includes online sales transactions for over a year for an online company. I would like to consider using clustering to find any patterns of products purchased and when, or by whom.

My client would be the company's marketing department. This data and analysis would potentially allow them to market different products better based on the propensity of customers to purchase, whether that is by strategically changing the price or offering sales to repeat customers.

I found the excel file via the UCI archive. It is available publicly on their website as a free download: http://archive.ics.uci.edu/ml/datasets/Online+Retail

I propose to use unsupervised learning with clustering to find patterns of purchases/clients.

Deliverables would be code and a slide deck.

Info about the data from my initial ideas document:
- Has time-series and better labeled data than previous sales data
- Abstract: "This is a transnational data set which contains all the transactions occurring between 01/12/2010 and 09/12/2011 for a UK-based and registered non-store online retail"
- Link: http://archive.ics.uci.edu/ml/datasets/Online+Retail
  - Also a link to transaction data for 2 years
    http://archive.ics.uci.edu/ml/datasets/Online+Retail+II
- Attribute info:
  - 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.