Explanation and Documentation

1. Data Preparation
    a. NULL descriptions were filled with 'MISSING'
    b. Sales quantity cannot be negative or zero, so related transactions were filtered out
    c. Unit price cannot be negative or zero, so related transactions were filtered out
    d. NULL customer ids were filled with 'MISSINGCUST'
    e. Months were selected as the basic temporal unit of evaluation. A new column called 'Time Period' was created. December 2010 was set as Time Period 1, and so on, and November 2011 was set as Time Period 12
    f. There was only partial data for December 2011, so that was filtered out
    g. A new column called 'Revenue' was created as the product of Quantity and Unit Price
    h. Data was collapsed into 2 different data sets:
        i. Customer-centric, and
        ii. Stock-centric
    i. In Customer-centric data, each row had a unique combination of Customer ID and Time Period. Quantity, Price, and Revenue were aggregated to get the average Quantity, average Price, and average Revenue per Customer ID per Time period. Quantity and Revenue were also aggregated to get the total Quantity and total Revenue per Customer ID per Time period
    j. In Stock-centric data, each row had a unique combination of Stock Code and Time Period. Quantity, Price, and Revenue were aggregated to get the average Quantity, average Price, and average Revenue per Stock Code per Time period. Quantity and Revenue were also aggregated to get the total Quantity and total Revenue per Stock Code per Time period

2. Machine Learning models were fit to both sets of data
    a. In Customer-centric data, customer ID, country, and Time Period were used to predict the average Quantity, average Price, average Revenue, total Quantity, and total Revenue per Customer ID per Time period for all customer IDs for the following time periods: 13 (December 2011), 14 (January 2012), 15 (February 2012)
    b. In Stock-centric data, stock code, Time Period, and each word from the Description column (count-vectorized) were used to predict the average Quantity, average Price, average Revenue, total Quantity, and total Revenue per Stock Code per Time period for all Stock Codes for the following time periods: 13 (December 2011), 14 (January 2012), 15 (February 2012)

3. Visualization and Analysis
    a. Customer Behavior:
        i. Time Period slicer on the top-left to select the temporal range of interest. Time periods 1 through 12 are actual data, whereas time periods 13-15 are predicted data

ii. The pie chart below the slicer shows total revenue generated by all customers in different countries for the time period selected in the slicer
iii. To drill down further, the bar chart below the pie chart shows the top customers in terms of total revenue generation during the time period selected in the slicer (excluding the predicted revenue for last 3 months)
iv. Line graph on the top-right shows how total revenue generation has progressed through the time periods (excluding predicted ones) for each customer
v. Table below the line graph filters out the customer for whom we have all 15 data points, 12 for the actual time periods and 3 for the predicted time periods
vi. Bar graph to the right of the table shows the minimum total revenue possible for the predicted time periods 13, 14, 15. Country-wise split is shown. It is minimum expected because the data is drawn only from customers who have been engaging in a transaction in each time period

b. Stock Performance
i. Time Period slicer on the top-left to select the temporal range of interest. Time periods 1 through 12 are actual data, whereas time periods 13-15 are predicted data
ii. The bar chart below the slicer shows the top total revenue generating stocks/products for the selected time slice (excluding the predicted revenue for last 3 months)
iii. Since Stock Code is not completely descriptive, the table below the bar chart shows exactly the same insights as the bar chart, but just adds the description as a column
iv. Line graph on the bottom-left shows how total revenue generation has progressed through the time periods (excluding predicted ones) for each product/description
v. The world map on the top right shows the top 10 revenue generating products/descriptions along with the consuming country for the first 12 time periods for which actual data is available
vi. The tree map on the bottom-right shows total revenue prediction for time periods 13, 14, 15. The prediction for each time period is split by the stock code and does not include stock codes expected to generate less than 5000 in total revenue