**Business Requirement - Complete the series of task below to provide useful reports**

**to aid decision making.**

**Getting your data into MS Excel**

* Create a new MS Excel workbook, name the workbook Superstores DataSet
* In the workbook, using an existing or newly inserted worksheet, rename the worksheet (Sales­\_Transaction\_Data)
* In the same workbook, using another existing or newly inserted worksheet, rename the worksheet (Returned\_Items)
* In the same workbook, using another existing or newly inserted worksheet, rename the worksheet (Regional\_Managers)
* Using the appropriate method, import the sales transaction.txt data (can be found in the assignment folder) into the Sales transaction worksheet.
* Using the appropriate method, import the Returned Items.txt data (can be found in the assignment folder) into the Returned Items worksheet.
* Using the appropriate method, import the Regional Managers.txt data (can be found in the assignment folder) into the Regional Managers worksheet.

**Data Exploration & Understanding**

* Investigate and explore each of the fields(aka Variables, columns)
* How many records (rows) of transactions are there?
* How many variables are in each dataset (Data in the worksheets)? Sales transaction-184655,
* Check the formats (number format in Excel) of each field and ensure fields aren’t assigned the wrong number format. From the field name, you should be able to attribute the right number format in excel to the field e.g. Text values in a Date field.
* Are there any missing values in any of the fields? If so, how do you deal with the missing values(Research this on google – How to deal with missing values in data analysis)
* If any, what is the percentage of missing values across each of the fields?
* Are there any spurious values or evidence of noise? These could include symbols, wrong data value in a wrong data field e.g. text/strings in a numeric field or date in a numeric field.



* How many unique categories are there in each field? (Get this from Pivot)
* Make notes of all identified issues with the data good or bad.

**Data Preparation – making data robust and fit for purpose**

Note: To derive a new field means insert/create a new column in excel, functions and formulas must be used to implement the data preparation and show all formulas and functions used in the spreadsheet

1. Data cleansing – Remove all spurious and unclean values from the data.
2. Using appropriate functions or methods – derive a new field called (RealOrderDate) from Order Date. Your new field should have this date format - 13/10/2010
3. Missing Data – Input missing values using an appropriate method and explain why you have chosen the method (Tip: At least 2-3 variables in the data contain missing values). Apply business sense to your method of choice, could filling the missing values mislead the business?
4. Data Formats – Re-assign the appropriate format to fields with incorrect formats (Tip: Order quantity)
5. Derive(create) new variables(fields) from existing variables to enhance the data
6. Using the (RealOrderDate) field and appropriate function in excel, derive 5 new columns – (OrderDay), (OrderMonth), (OrderYear), (Month-Year) and (DayOfMonth). This will enable you to slice the data by day, month and year when you start your analysis.
7. Using8r the Order Priority field and appropriate functions in excel, derive a new variable/column (PriorityCode). Making Critical - 5, High - 4, Medium - 3, Low - 2 and Not Specified - 1.
8. Derive a new field (DaysToShip) by calculating the Number of Days between the newly derived column RealOrderDate and Ship Date.
9. Using the appropriate function, derive a new field by combining the first name and last name into one column (FullName).
10. Using the birthdate field, derive the customer age field (Age)
11. Group customers into age band (AgeBand) using the newly derived Age column.
12. Using all the transactional fields provided, derive the Cost per Unit field.
13. Looking up data – show all formula/functions used.
    1. Using the Returned Items data and appropriate look up functions/formula, derive a new field in Sales Transaction Data by returning the corresponding Order ID Status to show the orders that were returned in the Sales Transaction Data.
    2. Following the same procedure as above (E.a.) and using the Regional Managers dataset, derive a new field (RegionalManager) in Sales Transaction Data by returning the corresponding regional managers’ names against the right region.

**Analysis**

Recommendation - Your analysis should be carried out using pivot tables and other advanced excel functions. Mastering pivot tables will enhance your analysis skills, it is the main analytical/aggregation tool in excel. Use as many pivot tables as possible and ensure you retain each pivot tables in your summaries sheet.

Create a new worksheet and name it Dashboard, this is where all charts created will be formatted and designed for presentation purposes.

1. Create/insert a new worksheet, name it summaries. This sheet should contain all the rough work/pivot tables you will use to create your reports.
2. Using pivot table, summarise the data and display the Month-Year and yearly profit/revenue report for Sales transaction. Plot this on an appropriate chart (Tip – check recommended charts).
3. Which product category, region, Province and Customer Segment is most profitable? (Tip: Use the pivot table to get the information and create a table).
4. What is the Minimum, Maximum, Median, Mode and Mean (average) days to ship from the order date to ship date? (Tip: Use DaysToShip column).
5. How many Critical priority orders took longer to ship than the mean (average) days to ship (from question 5)? Which region/manager defaults the most based on this key performance indicator?
6. Plot the chart of revenue based on the DayOfMonth. Which day do we perform well the most?
7. Which Ship mode cost the most and by how much? What is the average shipping cost across all sales transactions?
8. Which age group of customers are driving profit, what is the commercial value (Revenue and Profit worth) of each age group?
9. The business is looking to understand the drivers of returned orders. Which Product Category, Product Sub-Category, Product, Region, Manager is this predominant to? How much revenue was lost due to returned products? Make some recommendation on how we can prevent this and what would be the commercial value after 6months if percentage of returns went down by 50%.
10. The business intends to promote one of the regional managers to become Head of Sales, based on performance analysis of the sales transaction data who would you recommend and why?
11. Format and design all charts in the dashboard.

**Presentation and Recommendation**

1. Insert relevant slicers and timelines as filters to aid proper analysis.
2. Report connections of all slicers and timelines.
3. Cut and paste necessary derived tables from summaries to dashboard

**Hints**

* Use all necessary formatting tools
* Remove all outlines from charts

**Outputs**

You are required to send your MS Excel Workbook (showing all rough work).