Technical Assessment

World Vision

One of the biggest European company focused on e-commerce would like to better understand their customers purchase profile and be more successful on upcoming marketing campaigns that include but no limited to market basket analysis. A dataset representing a sample of their B2B transactions was shared with you (BI Engineer) and it was requested end-to-end analytical solution that will support the decision making of SLT group.

Dataset

File Name: Dataset.xlsx

Attributes Information:

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.

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.

Main Tasks

The solution includes:

1) The dataset should be loaded on a RDBMS, you can use free versions available like MySQL.

2) The dataset has potentially missing or non-expected values based on the columns definition that will be presented later. It means that a data cleansing process should be applied first, and these purged data must be moved to a temporary data structure to be analyzed and manually fixed. This process will create a new and filtered dataset that should be loaded on the same RDBMS.

Note: The data purge process(es) and the populated data structure are part of the solution and should be part of the deliverable.

3) To better support data analysis, a dimensional data model (s) should be created and a set of ETL processes developed to feed this model.

Note: The data model diagram and the ETL scripts are part of the solution and should be part of the deliverable. The ETL scripts should be part of a job or similar functional object to support a one-shot load considering dependencies. It means that a "job call" can load the dimensional data model with no need to start ETLs one by one. It must be considered in this job the possibility to perform a full data-load or incremental data-load as well.

4) The use of a BI Tool (Power BI, Tableau, ...) to build up reports and dashboards (Sales Book) is mandatory (you can use free versions available) to give a self-service experience to the final user. The expertise of the BI Engineer to design valuable data analysis is a key asset in this technical assessment considering all the attributes belonging to the dataset, but additionally the SLT group would like to see important charts like:

- What time do people often purchase online?
- How many items each customer buy?
- Top 10 best sellers' products?
- Average transaction value (total revenue / number of transactions) Year over year.

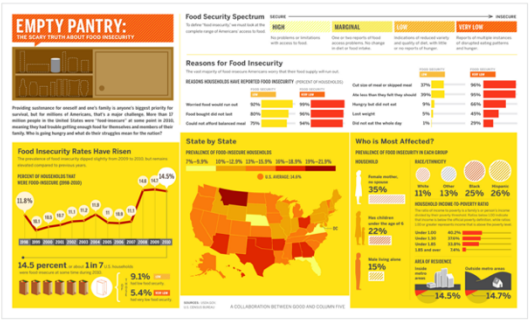
  Note: A high dollar amount could mean that shoppers are purchasing your more expensive products or they're buying larger quantities.
- Basket Analysis including average size of basket and the set of common products purchased.
- The frequency of cancelation (number of cancelled invoices) and average amount of cancelation. Is there any common product associated with cancelations?

Note: Histograms, Time-series (day, month, quarter, year), maps / heat-maps visualizations and the use dimensional filters are well appreciated technical features in analytical "books". Tip: Data Analysis is composed of a balanced set of reports and dashboards that creates a compelling Storytelling.

Examples of Expected Visualizations

Reports and Charts Examples (histograms, Bars, Maps, Heats, Timeseries ...)

Story Telling Approach



Timeline

- May.05 (noon): Send Technical assessment to the candidates (this document)
- May.06 (noon): The candidate can send a suggested hour for a call in case of doubts or questions.
- May.07 (afternoon): Call Marcos-Candidate (dismiss potential technical doubts).
- May.14: Due Date to deliver the solution (send the hyperlink)
- May.15: Solution Presentation (1 hour). To be scheduled.
- May.18: Candidates will have their solutions ranked, and results send to hiring manager.
- Final Result (Expected Date): Until the end of May the selected candidate will be notified.

Information

- Each candidate will have 1-hour call-conference to present a detailed solution (May.15).
- Each candidate should send a hyperlink where all deliverables can be accessed by WVC (May.14).
- Deliverables: Database Scripts, Reports/Dashboards (pdfs) and a readme file with instructions or orientation to the WVC-evaluator to support the solution understanding and evaluation.