Option #1: Agriculture - BI Use Case: Farm to Market

Scott Miner

Colorado State University – Global Campus

**Option #1: Part A: Agriculture - BI Use Case: Farm to Market**

The assignment asks the student to design a single-solution BI framework for a large farming corporation to streamline its ordering and delivery processes. The system will combine historical data from multiple sources and make use of Radio Frequency Identification (RFID) technology in real-time to predict product ordering, optimize product prices, and modify delivery options for customers. Additionally, the framework incorporates a sensor network and a cloud-based data warehouse (DW) comprising two layers, a server layer and a client layer, that store data, and present data to users, respectively.

**Benefits of Consolidating Historical Data**

Woodard (2016) describes historical data in the agricultural context as data pertinent to weather, climate, land use, and water resources, among others. The integration of this data is important because it leads to high costs and time savings. Current data collection tasks may be resource-intensive, non-transparent, and error-prone if done manually. Additionally, without live data, models become stale and unusable. A modern cloud-based data management system that incorporates real-time data processing using RFID and sensor network technology can accomplish in minutes what took a legacy system days. Therefore, we need to determine ways to increase logistic efficiency, shorten transportation times, and minimize losses. One such method involves using RFID technology.

**Real-Time Capabilities of RFID Technology**

We can use RFID technology to track product information along the food supply chain as agriculture products pass from growers to retailers. A passive tag replaces a bar code in the supply chain and identifies a product at the item level (Visich et al., 2007). Passive tags exchange data wirelessly between readers, which we can implement at key control points in the supply chain, including dock doors and facility entrances (Chow et al., 2006). No line of sight is required to read the tags, and we can implement them in harsh conditions, including environments with moisture, dust, and dirt. We can read hundreds of tags simultaneously, program tags easily, and store additional information beyond identifiers, including environmental conditions, historical move data, and future destination data (Visich et al., 2007). Figure 1 illustrates product tracking with RFID technology in the supply chain.

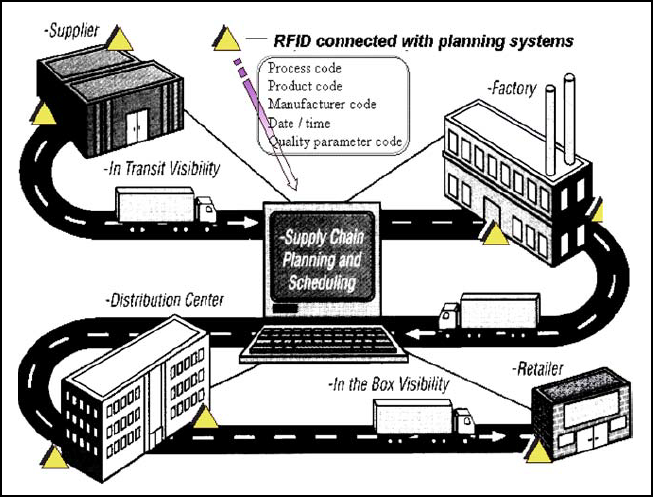
How does it work? The EPC is embedded in a tag microchip and is attached to an item, case, or pallet. As a tag comes within range of an RFID reader, the reader sends out an electromagnetic wave that couples with the antenna on the RFID tag, which it uses to power the circuit of its microchip. The RFID tag, in return, sends out an electromagnetic wave to the reader per its programmed code, and the reader decodes this information and decides to either store the information, act upon it, or transmit it to a receiving computer. This architecture makes use of Savant, software developed at MIT that, when receiving information from an RFID tag, uses the EPC to contact the Object Naming Service (ONS), which locates the server containing information on the scanned item. The Savant software receives this information and communicates it to the corresponding cloud-based DW. From this, we can determine this architecture comprises three main parts: (a) the RFID tag, (b) the RFID reader, and (c) the back-end database (Visich et al., 2007).

Figure 1. Product Tracking with RFID. From "Dynamic planning with a wireless product identification technology in food supply chains. International Journal of Advanced Manufacturing Technology," by Li et al., 2006.

**Price adjustments and Delivery Modifications in Real-Time**

Li et al. (2006) relay how with RFID technology, we can identify product value in real-time, as well as estimate product prices at various stages in the supply chain. The main factor that affects food value is temperature. Therefore, at a minimum, our RFID microchip needs to contain information on identification, packaging, temperature, and the times and dates of entering and exiting processes. By using RFID readers at key control points along the supply chain, we can achieve real-time visibility of the product in transit from the supplier to the distribution center, as well as from the distribution center to the retailer. As control systems automatically read information from RFID tags, the system dynamically plans for the best processing routes based on information from the current products, as well as information upstream, from the next process in the supply chain. In this way, we can modify delivery methods for customers as needed, based on the perishability and value of the product in the supply chain. The main factors used to determine the value of a product along the supply chain are (a) demands, (b) orders, (c) product value deterioration rates, (d) time spent at each stage in the supply chain, and (e) shipping costs. The main factor affecting the quality and shelf life (SL) of a perishable food product is the temperature at harvest fields, processing workshops, warehouses, and retail stores. Therefore, we can monitor these data in our BI solution via RFID technology to calculate value reduction over time and adjust product prices per market and environmental changes (Li et al., 2006).

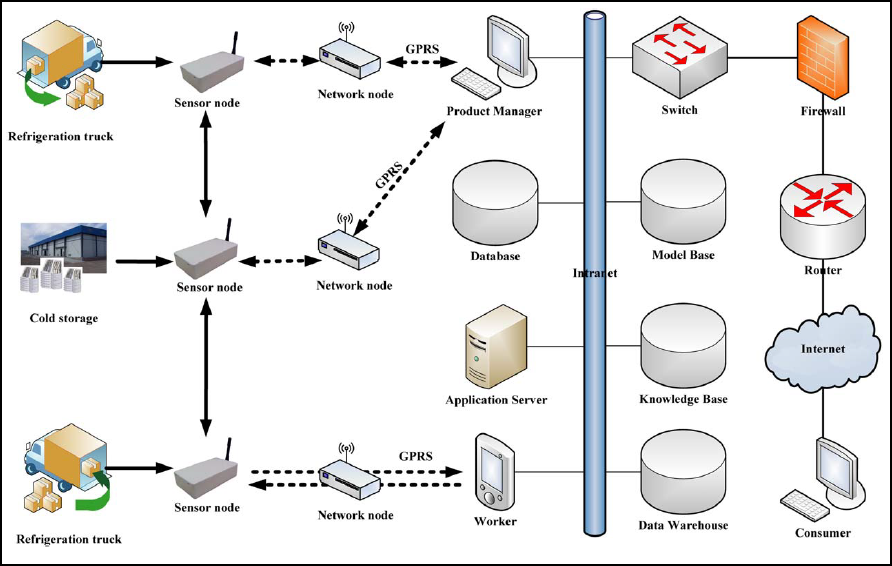
La Scalia et al. (2017) developed one such solution to monitor strawberry transportation from the “farm to the fork.” Among other factors found to influence product value, including light, moisture, oxygen levels, and microbiological growth, La Scalia et al. agree temperature is the factor known to most influence the SL of a fresh agricultural product. La Scalia et al. developed a smart logistic unit (SLU) to monitor these parameters in real-time along the food supply chain. The SLU contains probes, a global positioning system (GPS), and 3G modules to communicate product data to a cloud-based DW, accessible over the internet. The cloud-based platform allows for centralized data storage and real-time data access. The architecture of this framework comprises two layers consisting of sensor nodes, network nodes, and the cloud-based DW. The sensor nodes receive real-time temperature information from the trucks and send the data to the network coordinator, which aggregates the data and sends it to the remote DW via General Packet Radio Service (GPRS). The cloud-based DW consists of two layers, the server layer that connects users to the sensors and the client layer, that allows for the user-friendly presentation of data and allows customers to place orders over the web (Xiao et al., 2016). Figure 2 displays a proposed architecture of this solution.

Figure 2. Architecture diagram for BI solution framework. From "Developing an Intelligent Traceability System for Aquatic Products in Cold Chain Logistics Integrated WSN with SPC," by Xiao et al., 2016.

**Conclusion**

A cloud-based DW allows for a single solution BI system to incorporate both order and delivery data. Customers can log in via a web-based platform to place an order at any time of the day. As soon as a customer places an order, the order information, including quantity, cost, and shipping details, are updated in the cloud-based DW. Executives, analysts, and other key stakeholders have access to this data to adjust product prices when necessary and modify delivery methods based on market and environmental changes. Analysts can use the decision support system to predict which factors play the most significant role in product orders. In this way, order predictions can consider real-time data obtained from the RFID system, as well as historical data in the DW that contains information on past orders, past deliveries, environmental factors, and weather conditions. The BI solution framework proposed in this paper facilitates these requirements.

**Part B: Cleaning a Messy Data Set**

Part B of the assignment asks the student to clean a messy data set contained within the MESSYDATA.SAS7BDAT file. Figure 3 displays the SAS code used to clean the messy data set, which attaches labels to variable names to enhance the readability of the output. Figure 4 displays the results of printing five variables from the first ten rows of the data set with the label names displayed. Figure 5 shows a screenshot of the Output Data, the tab that contains the SAS table with all variables present. Notice on the left side of the display, we can see the label names of all 15 variables, along with a symbol representing the variable type (i.e., character or numeric data). Elliott and Woodward (2016) identify several other ways to clean messy data sets. Popular options include fixing case problems, allowing categories, and deleting unneeded lines. Commonly, we need to standardize cases among variables. The case of the gender code might vary between “m” and “M,” and we need to standardize all the values accordingly. In such scenarios, we can use the UPCASE() and LOWCASE() functions in SAS. To verify whether all values in a category are allowable, we can compare them to pre-coded values. For instance, if a category allows nothing but the values “YES” or “NO.” we compare all values to these values, and if the data do not equal any allowed values, we set the value to missing. Lastly, we can delete irrelevant records. If a line is missing a subject ID, or some other element we deem important for analysis, we can delete the record. Thus, we are provided with multiple ways to clean data sets, including fixing labels and renaming variables, fixing capitalization issues, allowing categories, and deleting unneeded records.

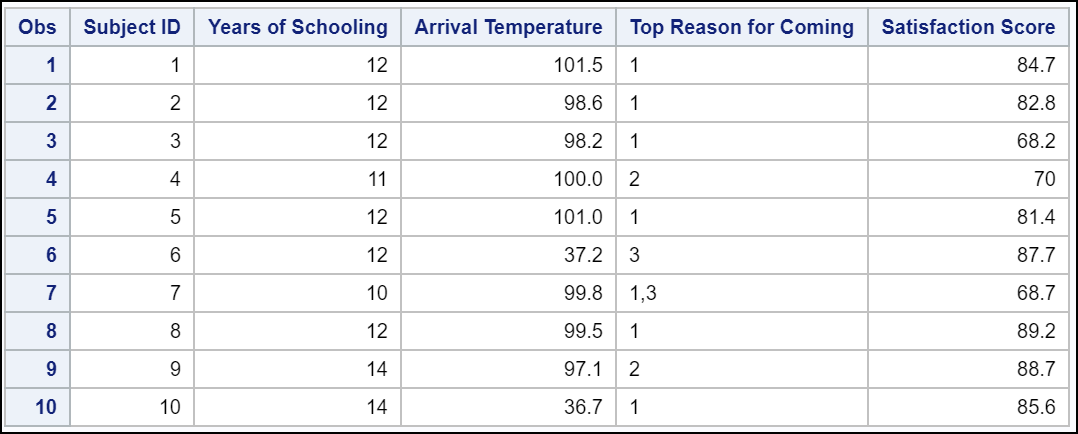


Figure 4. Output of first 10 rows with labels

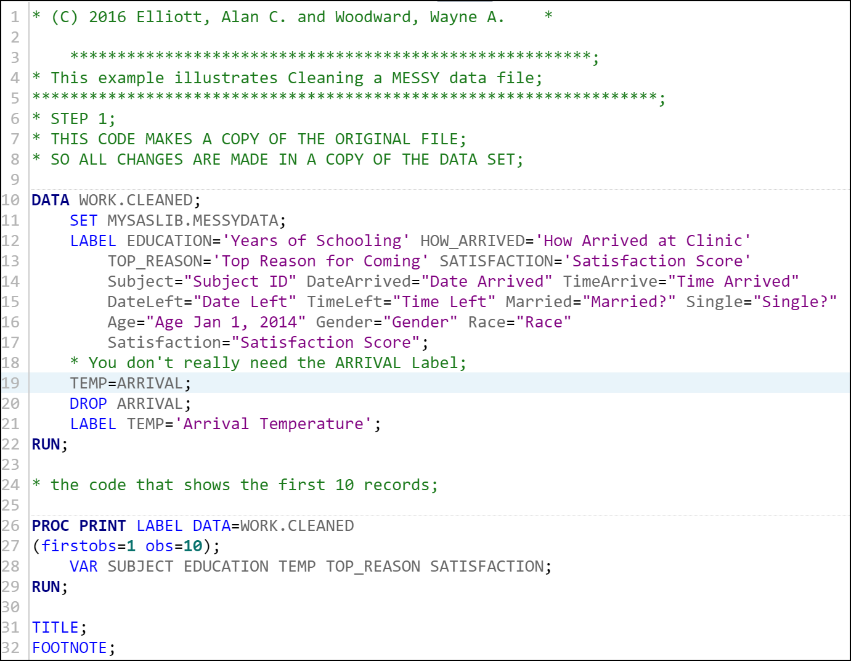


Figure 3. SAS code to clean a messy data set

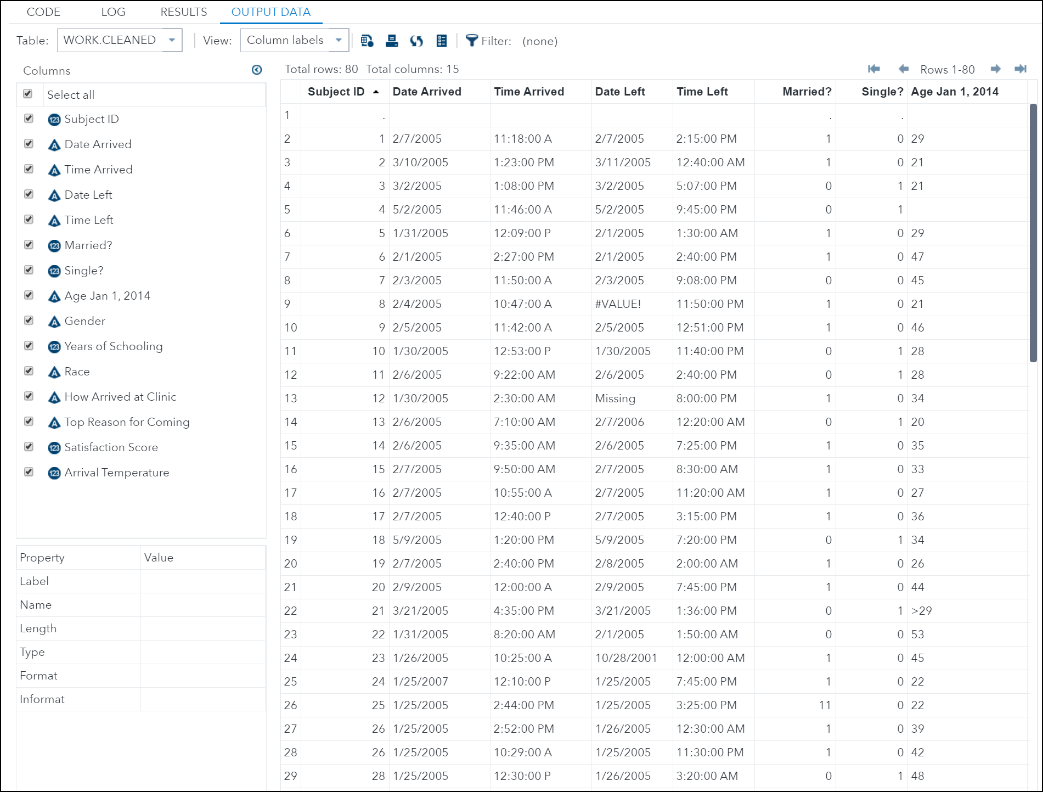


Figure 5. Output Data Tab contains the SAS data table with all variables present. Notice the label names on the left-hand side of the image.

References

Elliott, A. C., & Woodward, W. A. (2016). SAS Essentials: Mastering SAS for data analytics (Second edition). John Wiley and Sons, Inc.

La Scalia, G., Nasca, A., Corona, O., Settanni, L., & Micale, R. (2017). An Innovative Shelf Life Model Based on Smart Logistic Unit for an Efficient Management of the Perishable Food Supply Chain. Journal of Food Process Engineering, 40(1), n/a-N.PAG. <https://doi.org/10.1111/jfpe.12311>

Li, D., Kehoe, D., & Drake, P. (2006). Dynamic planning with a wireless product identification technology in food supply chains. International Journal of Advanced Manufacturing Technology, 30(9/10), 938–944. <https://doi.org/10.1007/s00170-005-0066-1>

Visich, J. K., Suhong Li, & Khumawala, B. M. (2007). Enhancing Product Recovery Value in Closed-loop Supply Chains with RFID. Journal of Managerial Issues, 19(3), 436–452.

Woodard, J. D. (2016). Data Science and Management for Large Scale Empirical Applications in Agricultural and Applied Economics Research. Applied Economic Perspectives & Policy, 38(3), 373–388. <https://doi.org/10.1093/aepp/ppw009>

Xiao, X., Fu, Z., Zhang, Y., Peng, Z., & Zhang, X. (2016). Developing an Intelligent Traceability System for Aquatic Products in Cold Chain Logistics Integrated WSN with SPC. Journal of Food Processing & Preservation, 40(6), 1448–1458. https://doi.org/10.1111/jfpp.12730