BISM7233 Data analytics for Business
Assignment Team 58
Word Count: 3757

Data analytics for Business

Case Study Report

Student

44884343 Shu-Hui Yang

45193837 Sheng-Yi Huang

45348011 Xinyao Zheng

44643054 Xuan Xie

45513323 TszHo Chan

Instructor

Dr Ida Asadi Someh

Content

I. Executive Summary

In the competitive market, the essential goal for a sustainable business is maintaining competitive advantages. According to the analysis, WorldView Bikes (WV) company cannot persist the increasing profit in 2014, which presents the negative impact on future performance. To help WV's business expansion, we will further give few suggests to the decision maker.

WV should pay more attention on the key customers, provide more complete services to them and improve the loyalty of key customers. In addition, as the most profitable product is from bicycles, WV should promote different types of bicycles to the most related sales areas and the most profitable sales territories according to the function of bicycles; that is, mountain, road and touring bikes and according to the most profitable period every year. Therefore, WV can drastically increase the market share and stand out from competitors. Eventually, after rejecting the least important vendors, for making sure the quality of products, WV not only maintains existing customers but attract more potential customers as well.

Business analytics (BA) provides these suggestions to help business improve current situations and analyze future predictions. By utilizing BA, our company will take proactive actions to earn more profits and achieve higher market share compared with our competitors.

II. Business Analytics Framework

With the advancement in technology, enterprises can make good use of external and internal data to analyze the questions they are concerned with, such as the best-selling product, the most valuable customer and the most profitable sales area. Business analytics refers to businesses utilize statistical analysis and historical data to provide predictive analysis, solve businesses' problems and help managers make decisions. Enterprises value business analytics much because they not only can take preemptive actions before sales dropping happen but can seize opportunities to maximize the market share as well, which greatly improve business processes and help businesses sustain competitive advantages in such competing industry.

There is no denying that business analytics plays a pivotal role in enterprises. We will then illustrate five major benefits that result from business analytics and businesses

can take these advantages by implementing business analytics.

To begin with, business analytics provides quicker and smarter decision-making to enterprises, which complied with business rules and strategies. Enterprises have to make sure the high-quality data; that is, the accuracy, creditability and timeliness of the data and leverage statistical and quantitative analysis to achieve business ultimate goals, such as maximizing the profitability and minimizing the cost. Secondly, business analytics help enterprises improve current situations and business processes. For example, organizations can monitor and check the schedule that they are designed at the beginning of the year. If the actual figure is way behind that they expected, they could change business processes to catch up the goals through business analytics immediately. Thirdly, owing to the rapid change in the industry, customers' preferences will change all the time and companies should provide a variety of services and products to satisfy needs from different kinds of clients. As a result, businesses need to have an ability to discover unpredictable and unprecedented situations so that they can come up with innovative, creative and effective solutions to deal with problems and business analytics can indeed accomplish these tasks. Fourthly, under the huge competitive pressure, enterprises need to work more efficiently through business analytics, as business analytics focuses on the real-time data analysis, provide predictive reports to companies and assist organizations to set up proactive and anticipating actions. Finally, business analytics enables enterprises to utilize business insights through data visualization, give suggestions and help decision-making. Visualization presents comprehensive, meaningful and useful data through charts and graphs and allows business to obtain insights in a clear, readable and understandable way.

Nowadays, businesses utilize the technology and complex techniques to leverage the value of the data and practice business strategies. For example, in order to meet customers' expectation, enterprises need to come up with the tactic which considers different fronts, such as vendors, products and services. Therefore, a framework of business analytics, which consists of data, strategies, business processes and business performance management allows the company to derive data in a structured way. For different projects and goals, enterprises should define and construct various specific business analytic framework. Business analytic framework provides a step-by-step process and design appropriate relationships to implement and practice business

analytics.

Now, the WorldView Bikes (WV) company takes expanding their business into consideration. To consider the expanding project deliberately, thoroughly and perfectly, the company assigned our team to do business analytics and help the CEO make a decent decision. Consequently, we will further extract, transform and load data through Pentaho to analyze five essential questions and provide insights to the decision maker. The analysis of five essential questions is shown as follows.

1. Key Customers

The company needs to identify the most profitable customers as key customers, so WV can provide more relevant products and more complete services to key customers.

2. Profitable products

The company should understand which products are the most popular and profitable, so WV can expand the business with the related products.

3. Profitable sales territories

Identifying the most profitable sales territory enables the business to analyze specific products in these territories. For example, in territories containing mountain areas, mountain-related products may be the best sales and the most valuable so that these areas can be the most profitable territories.

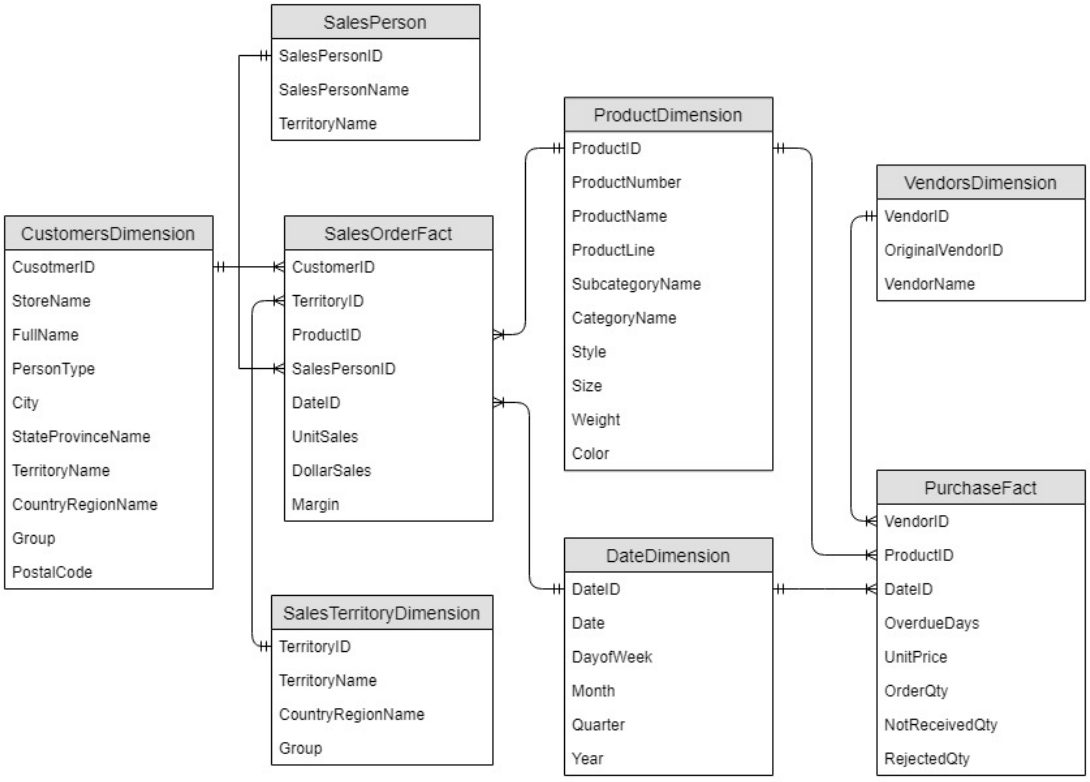
4. Profitable time periods

Company can identify the most popular time periods during a whole year to set up strategies and make a bulk of profits during these periods every year.

5. Vendors for replacement

The company should change suppliers to those who have more competitive purchase prices; therefore, the company can minimize the cost. Furthermore, vendors whose behaviors violate our business strategies will be rejected by our company. In other words, we consider vendors who provide poor quality products and have bad service quality to be the least important suppliers.

III. Design of a Dimensional Model



We choose 8,7,6,10,6,3,10 attributes for the Purchase Fact, Sales Order Fact, Date Dimension, Product Dimension, Sales Territory Dimension, Vendors Dimension, and Customers Dimension tables respectively; the reasons are as follows.

1. Purchase Fact Table

To identify WV's least important suppliers, we utilize these attributes:

VendorID, ProdcutID, DateID: Purchase Fact table needs the Vendor ID, ProdcutID, and DateID to link it with Vendors, Product, and Date dimension table because these enable the CEO to know the information about the vendors, products and all the different dates relevant to purchase orders.

OverdueDays: the overdue days can show how many days of delivery vendors are late and give the insight of vendors' delivery quality.

UnitPrice: the CEO can compare the unit price of the same product from different vendors to identify the suppliers who have competitive advantage in price.

OrderQty: the quantity of order shows how many products WV needs from different vendors. CEO can easily identify the least important suppliers by comparing the data with different vendors.

NotReceivedQty: the not received quantity allows CEO to identify the service quality of vendors by knowing how many products the vendors fail to deliver to WV.

RejectedQty: the rejected quantity can show the product quality of vendors by calculating how many products WV rejects and returns them to vendors.

2. Sales Order Fact Table

CustomerID, TerritoryID, Product ID, SalesPersonID and DateID: they are primary keys related to Customer Dimension, Sales Territory Dimension, Product Dimension, Sales Person Dimension and Date Dimension Table respectively. We can distinguish customers, identify where customers are from, understand what kinds of products were bought by customers and when customers placed orders.

Unitsales: It represents the product quantities we sold to customers. It is an important criterion to find out key customers.

Dollarsales: It represents line total, showing every product's total sales including discount. It is useful in calculating every customer's total purchase.

Margin: It represents how much profits the company generates every revenue. Margin is calculated in: Margin = (Line total - standard cost * Qty) / line total. It shows the sales situation of the company.

3. Date Dimension Table

DateID: it is the surrogate key of the date dimension for unique identifying all the dates and it is created by ourselves.

Date: we need the detailed dates to link all the other tables that have the relevant date information and trace the information through the date. It is also the finest level of hierarchy, the grain of the date dimension table.

DayofWeek: it is chosen because the CEO wants to know what the most profitable time periods are by Sundays. And it is also make it available to analyse data, such as sales, by drilling down through the hierarchy.

Month, Quarter, Year: they are chosen because the CEO would like to identify the key customers, the most profitable products, the most profitable territories, and the most profitable time periods are by month, quarter, season and year. They also comprise the hierarchy of the Date Dimension table.

4. Product Dimension Table

To identify the most profitable products of WV, we select these attributes:

ProductID: it is the surrogate key of the product dimension for unique identifying all the products and it is created by ourselves.

ProductNumber: it is a unique identifier number for a product.

ProductName, ProductLine, SubcategoryName, CategoryName, Style, Size, Weight, Color: these attributes allow the CEO to identify the characteristics of the most profitable products.

5. Sales Territory Dimension Table

To identify the most profitable sales territory, we select these attributes:

TerritoryID: it is a unique identifier for a sales territory.

TerritoryName, CountryRegionName, Group: it is clear to identify the most profitable territory's name, country region and geographic area.

6. Vendors Dimension Table

VendorID: It is the new primary key that we used in Pentaho to identify different vendors in sequence.

OriginalVendorID: It is the old primary key to identify different vendors.

VendorName: Name helps analysts to show vendors clearly in the final report. It also helps stakeholders easy to understand.

7. Customers Dimension Table

CustomerID: It is the primary key to identify different customers.

StoreName, Full name: these attributes represent store customers and private customers separately, which enables decision makers to identify specific store customers and private customers easily in the report.

PersonType: There are two types of customers, store and private customers.

City, State, Country, Territory, Group, Postcode: They are geographical information of customers so that we can analyse our key customers' areas.

8. Salesperson Dimension Table

To identify the most profitable salesperson and the territory he or she is in charge of, we select these attributes:

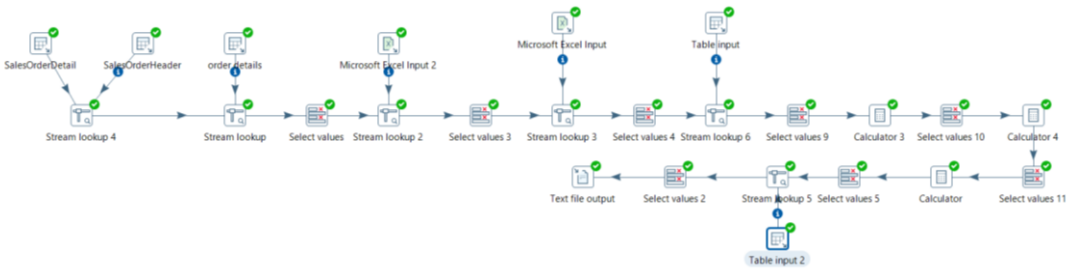
SalesPersonID: it is a unique identifier for a sales person.

SalesPersonName, SalesTerritoryName: it is clear to identify the most profitable territory that the specific salesperson is in charge of.

IV. Implementation of the Dimensional Model

1. Data Integration

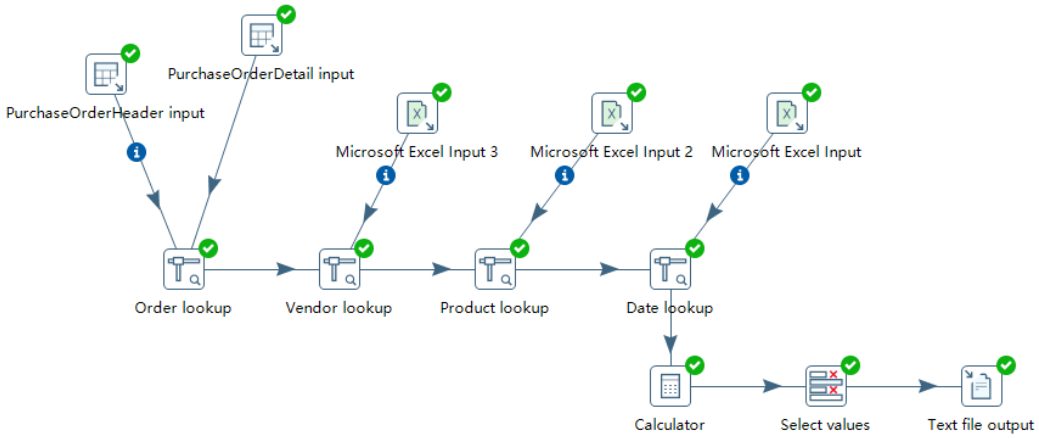
a. Sales Order Fact Table



First, use Stream Lookup function to join SalesOrder Detail table and SalesOrder Header table together with foreign key SalesOrderNumber, get the fields CustomerID, OrderDate, SalesOrderNumber, TerritoryID, SalesPersonID and name the new table as Stream Lookup 4. Second, use Stream Lookup function to join SalesOrder Detail table and Stream Lookup 4 table together with foreign key SalesOrderNumber, get the fields OrderQty, ProductNumber, SalesOrderNumber, SpecialOfferID, UnitPrice, LineTotal. Third, use Select Value function to extract fields OrderQty, UnitPrice, LineTotal, ProductNumber, CustomerID, OrderDate, TerritoryID, SalesPersonID. Fourth, use Microsoft Excel Input function to add Date Dimension sheet. Use Stream Lookup function to join Date Dimension table together. Fifth, select values OrderQty, UnitPrice, LineTotal, ProductNumber, CustomerID, DateID, TerritoryID, SalesPersonID. Rename OrderQty as UnitSales, Line Total as DollarSales. Sixth, input excel file

Product Dimension and join it with Select Value 3 as Stream Lookup 3. Seventh, Select Value 4 select CustomerID, TerritoryID, ProdcutID, SalesPersonID, DateID, UnitSales, DollarSales, ProductNumber. Eighth, Product table input from database, stream lookup 6 joins product table and select value 4 together. Then select value 9 selects CustomerID, TerritoryID, ProdcutID, SalesPersonID, DateID, UnitSales, DollarSales, StandardCost. Tenth, use Calculator to calculate Margin = (DollarSales - standard cost * UnitSales) / DollarSales. Select values 5 selects CustomerID, TerritoryID, ProdcutID, SalesPersonID, DateID, UnitSales, DollarSales, Margin. Lastly, use text file output function to output csv. SalesOrder Fact Table.

b. Purchase Fact Table



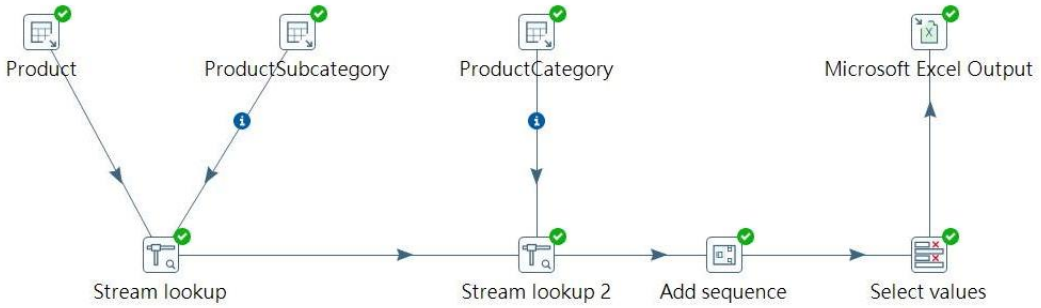
Firstly, we input all the data in PurchaseOrderHeader and PurchaseOrderDetail and change the attribute name of VendorID into OriginalVendorID in PurchaseOrderHeader. In order lookup step, we connect the two tables that are input in the first step with the same attribute PurchaseOrderID, adds the OriginalVendorID and OrderDate in PurchaseOrderHeader to all the attributes in PurchaseOrderDetail table. Then, we input the Vendors dimension table and look up and add the VendorID in Vendors dimension table to the table by the OriginalVendorID respectively. In the next Microsoft Excel Input 2 and Product lookup step, we will input the Product dimension table and look up and add the ProductID in Product dimension table to the table by the ProductNumber respectively. After that, input the Date dimension table (Microsoft Excel Input). In the data lookup step, we try to look up and add the DateID in Date dimension table to the table by the OrderDate in the table and Date in Date dimension table, and then calculate the OverdueDays by less the DueDate from the DeliveryDate in the table and NotReceivedQty by less the ReceivedQty from the OrderQty in the

table. And we also set the value type of these two new attributes into Integer. Finally, we delete all the other attributes except VendorID, ProdcutID, DateID, OverdueDays, UnitPrice, OrderQty, NotReceivedQty, and RejectedQty and sorts them in a proper order. We also set the value type for all the eight attributes in Select values and output all the transformed data into a csv excel table file.

c. Date Dimension Table

Firstly, we input all the data in Dates database table and changes the attribute name of DateID into Date in DateDimension Table input step. Secondly, we calculate the DayofWeek, Month, Quarter, and Year of each date input in the first step. And sets the value type of the DayofWeek and Month into String for the preparation of next Mapper steps. Next, we map the value of week and month (number) to text (Sun, Jan) respectively. Then, in add sequence step, we add the surrogate key, primary key – DateID, for Date dimension table to unique identify every date. After that, we sort all the attributes and set the value type for each attribute. Eventually, we output the all the transformed data into an excel table file.

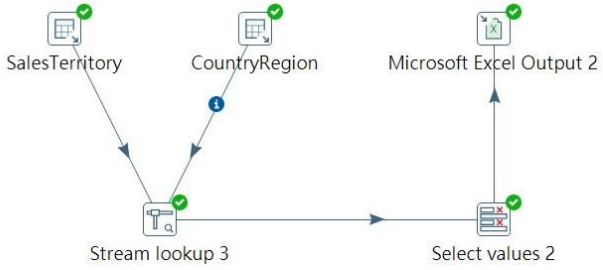
d. Product Dimension Table



Firstly, we select the attributes of Name as ProductName, ProductNumber, Color, Size, Weight, ProductLine, Style and ProductSubcategoryID from Product table. In ProductSubCategory table, we also select ProductSubcategoryID, ProductCategoryID and rename the Name in ProductSubCategory table as SubcategoryName. Then, using stream lookup to join these two tables together. Next, in ProductCategory table, we

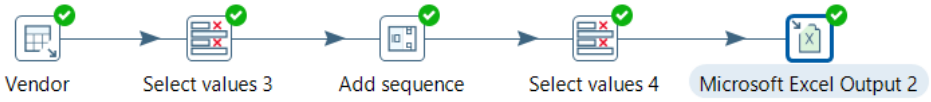
select ProductCategoryID and rename the Name in ProductCategory table as CategoryName, also use stream lookup 2 to join these three tables together. Furthermore, add a surrogate key which is ProductID to this dimension. Finally, we select all attributes to Product Dimension table and output all the transformed data.

e. Sales Territory Dimension Table



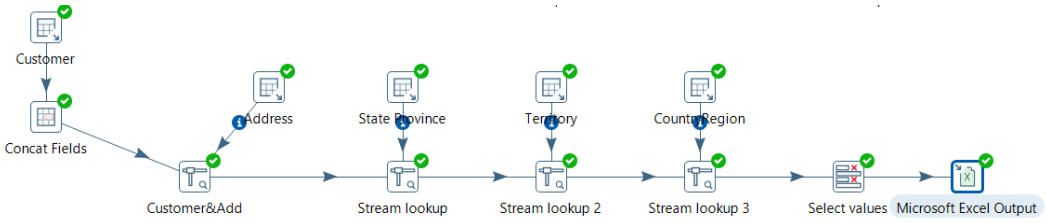
In SalesTerritory table, we select the attributes of TerritoryID, CountryRegionCode and Group and rename the Name in SalesTerritory table as TerritoryName. Also, choose CountryRegionCode, and rename the Name in CountryRegion table as CountryRegionName. Then, use stream lookup 3 to join the two tables together. Then, we choose TerritoryID, TerritoryName, CountryRegionName and Group to Sales Territory Dimension table and output all the transformed data.

f. Vendors Dimension Table



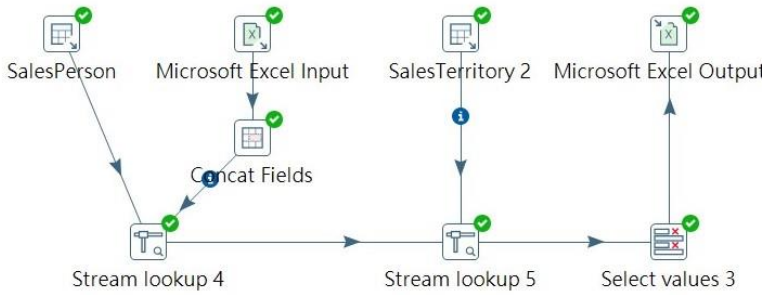
First, all fields are selected from Vendor table in the database provided, and use Select Value function to extract original VendorID and VendorName. Second, use Add Sequence function to generate new VendorID. Third, use Select Value function to extract new VendorID, original VendorID and VendorName. Last, output Vendor Dimension Table with Microsoft Excel Output function.

g. Customers Dimension Table



First, use Concat function to join customers' FirstName and LastName together, name the new column as FullName. Then, use Stream Lookup function to join Concat fields and Address table together with foreign key addressID and name the new table as Customer&Add. Second, select all fields of State Province Table and use Stream Lookup function to join it and Customer&Add with foreign key StateProvinceID, and name the new table as Stream Lookup. Third, select all fields of Territory Table and use Stream Lookup function to join it and Stream Lookup table with foreign key TerritoryID, and name the new table as Stream Lookup 2. Fourth, select all fields of Country Region Table and use Stream Lookup function to join it and Stream Lookup 2 table with foreign key CountryRegionCode, and name the new table as Stream Lookup 3. Fifth, use Select Values functions to extract and rename several columns as CustomerID, StoreName, FullName, PersonType, City, State, Country, Territory, Group, and Postcode. Lastly, output all the transformed data.

h. Salesperson Dimension Table



At first, we select SalesPersonID and TerritoryID from SalesPerson table, also input the attributes, SalesPersonID, Name and LastName from SalesPersonDetails table. Then, we combine Name with LastName as SalesPersonName attribute. Next, use stream lookup 4 to join these two tables together. Secondly, we choose TerritoryID and rename Name attribute as TerritoryName, then use stream lookup 5 to join the three tables together. Finally, we choose SalesPersonID, SalesPersonName and TerritoryName to Sales Person Dimension table and output all the transformed data.

2. Data Visualization

V. Generate Business Insights and Strategies

1. Who are the key customers?

We utilize sum of unit sales as total quantities and sum of dollar sales as total sales to measure the following questions.

In general, the total sales and total quantities from store customers is higher than that from individual customers; however, the average margin of individual customers is higher than store customers. Owing to having two different types of customers, we separate our customers into store and individual customers and analyze the key customer in each type by utilizing three criteria: total sales, which refers to our company's total revenue, total quantities, and margin.

In the key customer table, we use these three criteria to demonstrate our key customers. We can find out Jordan Turner as an individual customer has contributed highest total sales and total quantities to our company, especially in July 2013 and January 2014, and Kaitlyn Henderson has contributed the highest profit margin to our company in the individual type, especially in July 2013.

In customers of the store type, we value Brakes and Gears in the United State and Vigorous Exercise Company in Canada as key customers. Brakes and Gears provided the most total sales in September 2013 and Vigorous Exercise Company contributed the highest margin and total quantities to our company in October 2013.

2. Which products are the most profitable?

We also identify the most profitable product by analyzing total sales, total

quantities and margin of each product. In the product table, the most profitable products are Black Mountain-200 and Hitch Rack-4 Bike because they provide the highest total sales, total quantities and margin to our company.

We further analyze these two profitable products separately. During the period from 2011 to 2014, Mountain-200, Black had contributed 4.4M total sales to our company and the dashboard also shows that the average margin of this product is 22% and total quantities is 2.98k. Among the analyzed period, 2013 is the best-selling year. When we break down the period, it demonstrates that in quarter two we sold the product most compared with another three quarters. Also, the month with the highest total sales is March.

On the other hand, although Hitch Rack-4 Bike had provided only 237.1K total sales to our company, it still made great contributions on margin (47%) and total quantities (3.17K) during the period of our analysis. If we break down the time, we can find out we sold Hitch Rack-4 Bike most in July, quarter 3, 2013.

3. Which sales territories are the most profitable?

We utilize total sales, total quantities and margin to measure the most profitable sales territories. The most profitable territory is Southwest in the United State and the salesperson is Erfan Adam. This territory contributed 24.18M total sales, 59.1K total quantities and 27% margin to our company. In 2013, the total sales, total quantities and margin in Southwest is higher than that in other years. When we drill down the year 2013, we can find out the most profitable month is July in 2013.

4. Which time periods are the most profitable?

The most profitable time periods will be identified by the total sales, total quantities and margin which can directly represent which period has the most profit.

In year 2013, our company has the most total sales, total quantities and margin. When drilling down and analyze the most profitable month and quarter in year 2013, we can identify the most profitable period as follows.

The period with the most total sales is in June, Quarter two, 2013.

The period with the most total quantities is in July, Quarter three, 2013.

The period with the highest average margin is in November, Quarter four, 2013.

If we focus on Sunday, the period with most total sales and total quantities on Sunday is in June 2013, but the average margin is not the highest in this period.

5. Which vendors should be considered for replacement?

When it comes to identify the least important supplier, we come up with four criteria as follows.

Firstly, we consider Team Athletic Co., the vendor whose overdue days is the highest to be the least important supplier because this vendor has the lowest delivery quality. Secondly, we view Electronic Bike Repair & Supplies, the vendor who provides the least quantity of order to our company as the least significant supplier. Thirdly, SUPERSALES INC., the vendor who has most quantities that fail to hand over to our company, is a symbol of the least significant supplier; meanwhile, this supplier has the lowest service quality. Finally, we regard Integrated Sport Products, the vendor whose rejected quantities are larger than other vendors as the least important supplier because this vendor provides the poorest products to our company compared with other vendors.

Among these four criteria, we can obtain four different suppliers that represent the least important suppliers. However, if we would like to reject only one supplier, we should also consider our business strategy. The ultimate goals of our company are to stand out from competitors and keep sustainable competitive advantages; as a result, the product quality is our top priority. Due to having the lowest product quality, Integrated Sport Products will be rejected by our company in the future.

Appendix

1. **Data Dictionary**

   We submit the Data Dictionary in the attachment.

2. **Kettle transformations and Qlik Visualizations**

   We submit Kettle and Qlik files in the attachment.

3. **SQL**

   a. **Product Dimension**

```
SELECT
  Name as ProductName, ProductNumber, Color, StandardCost, "Size", Weight,
  ProductLine, Style, ProductSubcategoryID
FROM Product
SELECT
  ProductSubcategoryID, ProductCategoryID, Name as SubcategoryName
FROM ProductSubCategory
SELECT
  ProductCategoryID, Name as CategoryName
FROM ProductCategory
```

   b. **Sales Territory Dimension**

```
SELECT
  SalesPersonID, TerritoryID
FROM SalesPerson
SELECT
  TerritoryID, Name as TerritoryName, CountryRegionCode, "Group"
FROM SalesTerritory
SELECT
  CountryRegionCode, Name as CountryRegionName
FROM CountryRegion
SELECT
  StateProvinceID, TerritoryID
FROM StateProvince
SELECT
  AddressID, StateProvinceID, PostalCode
```

FROM Address

c. Date Dimension Table

SELECT

DateID as Date

FROM Dates

d. Purchase Fact Table

SELECT

PurchaseOrderID, RevisionNumber, Status, EmployeeID, VendorID as OriginalVendorID, ShipMethodID, OrderDate, ShipDate, SubTotal, TaxAmt, Freight, TotalDue, DeliveryDate

FROM PurchaseOrderHeader

4. Work Breakdown

Work	**Responsibility**	
	ID	**Name**
Executive Summary	44884343	Shu-Hui Yang
Business Analytics Framework	44884343	Shu-Hui Yang
Design of a Dimensional Model	45348011	Xinyao Zheng
	45193837	Sheng-Yi Huang
	44643054	Xuan Xie
	44884343	Shu-Hui Yang
	45513323	TszHo Chan
Implementation: Data Integration	45348011	Xinyao Zheng
	45193837	Sheng-Yi Huang
	44643054	Xuan Xie
Implementation: Data Visualization	45513323	TszHo Chan
Generate Business Insights and Strategies	45513323	TszHo Chan
	44884343	Shu-Hui Yang