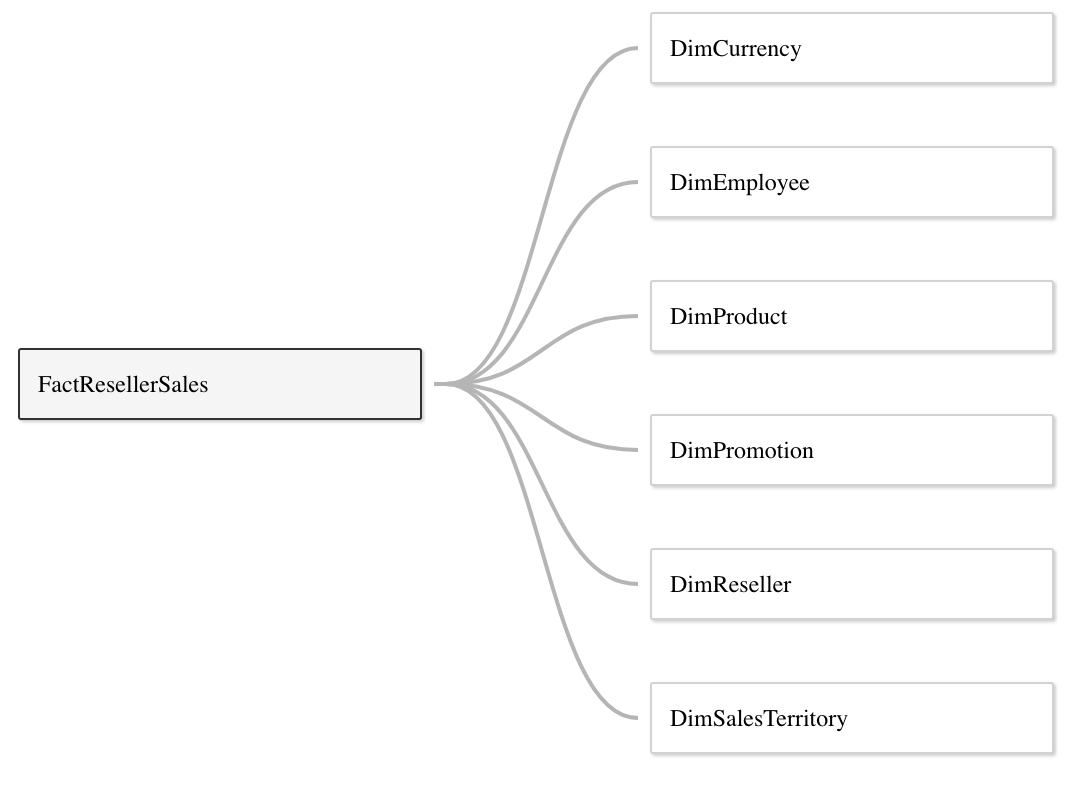
Assignment 2: Tableau Analysis

Name: Megh Dave

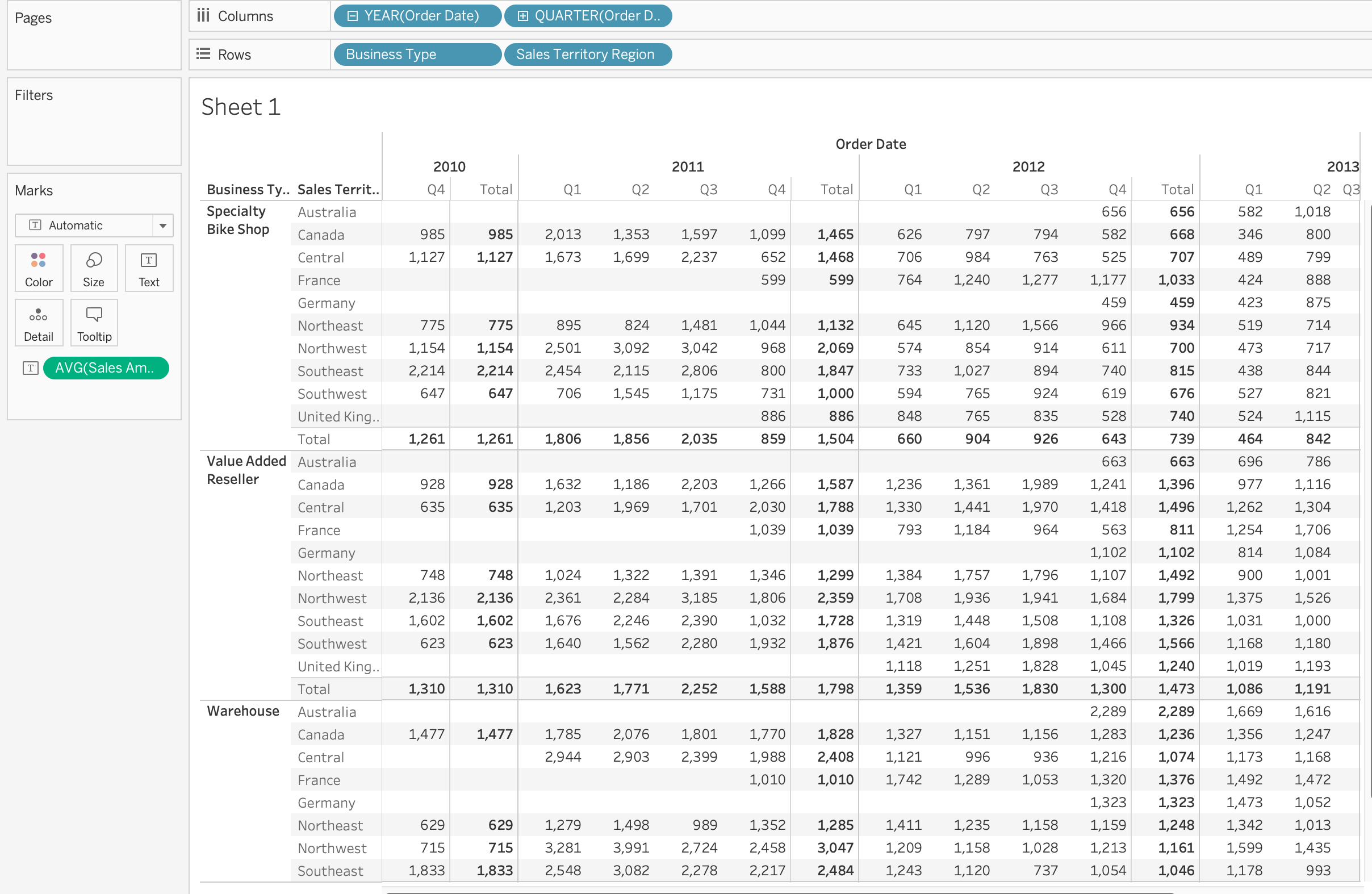
This assignment adopt the sample Adventure Works Data Warehouse, connecting the tables FactResellerSales, DimCurrency, DimEmployee, DimProduct, DimPromotion, DimSeller, DimSalesTerritory.



1. Cross Table Calculation and Aggregation:

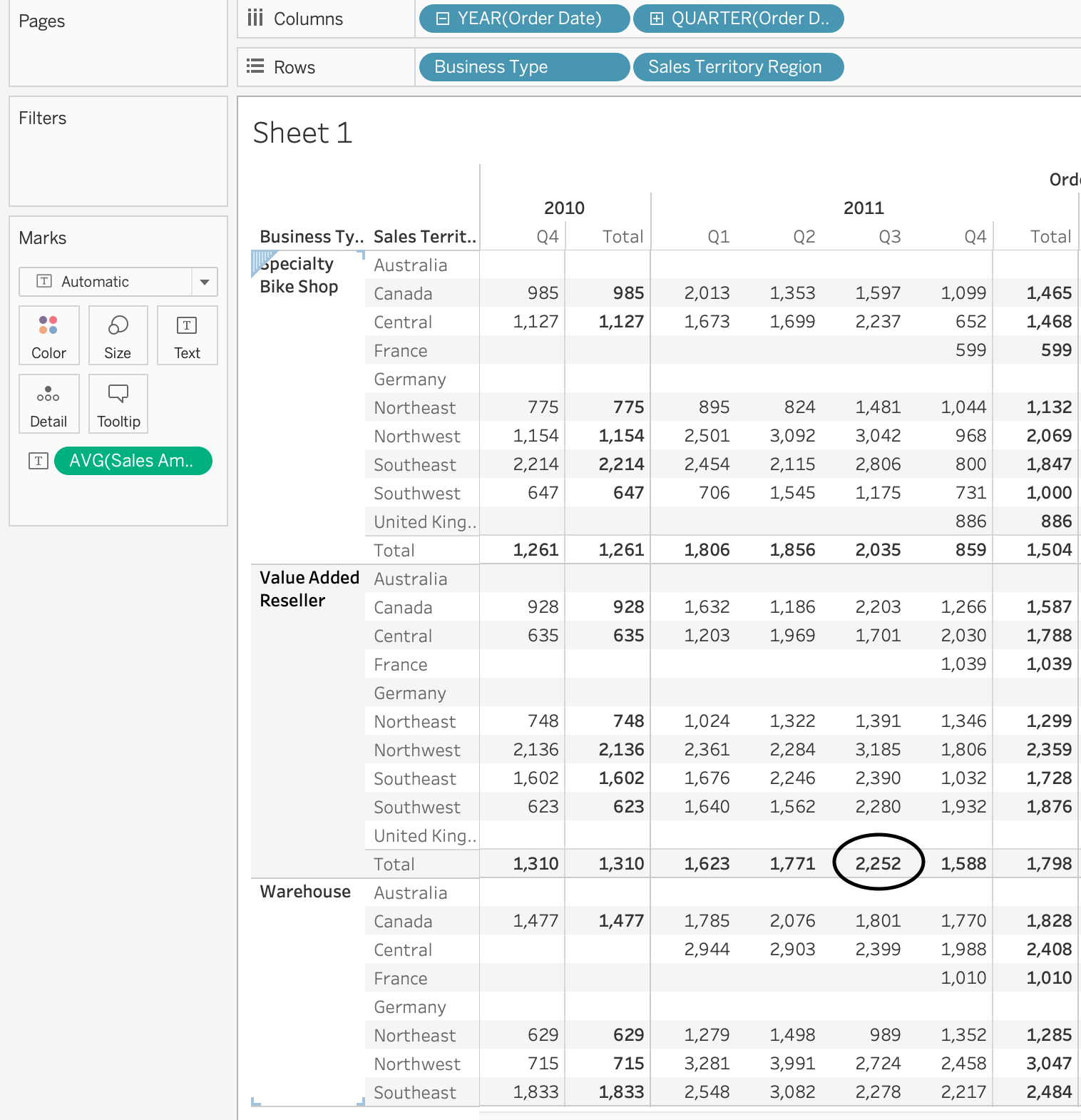
(1). Create and display a cross table with ***“Business Type”*** firstly (from DimReseller) and ***“Sales Territory Region”*** secondly (from DimSalesTerritory) as rows, with ***Year of “Order Date”*** and ***Quarter of “Order Date”*** (discrete types) from FactResellerSales as columns; The value of the cells in this cross table is the ***average of “Sales Amount”*** from FactResellerSales;

Besides the values in the cells, generate and display ***row grand totals***, ***column grand totals*** and ***all subtotals*** in the same cross table;

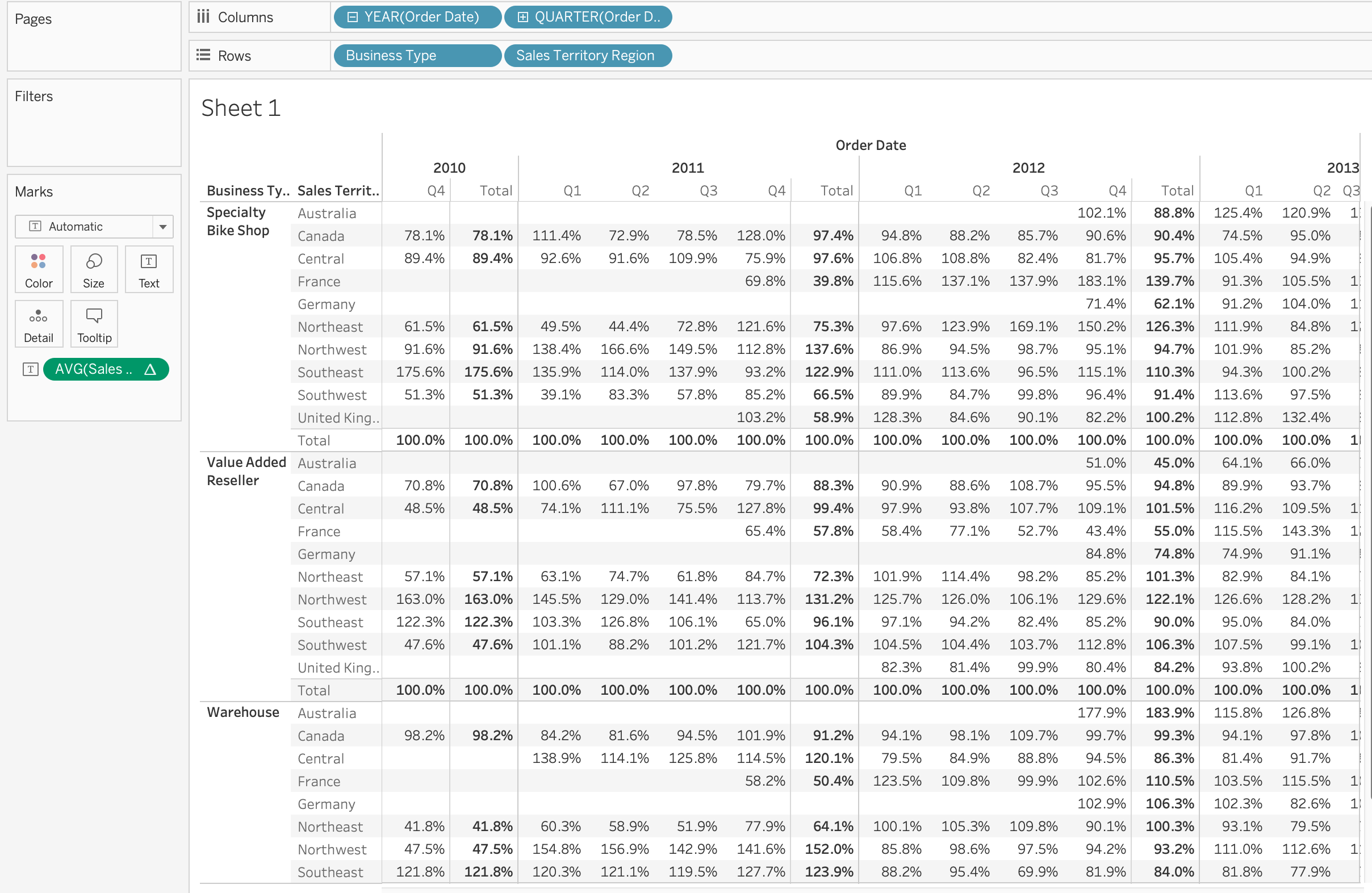


(2). What is the subtotal of sales amount from ***Value Added Reselle***r in ***2011 Q3***?

2252



(3). Based on the prior cross table, generate and display ***quick table calculation*** with ***percentage of total***, using ***pane (down)***;



(4). Based on the cross table from (3), Why do we have ***100%*** as the subtotal value at each pane level?

Pane down creates separate panes vertically, with each pane displaying a subset of data based on a specific category or criteria. We have 100% as the subtotal at each pane level because the data in each pane is displayed as a percentage of the total data in the preceding pane. It represents the percentage of the total data for all categories that are attributed to that category.

2. Discrete and Continuous Date Fields:

(1). Create and display a bar chart, with ***sum of “Sales amount”*** as rows, and **discrete** type ***month of “Order Date”*** from FactResellerSales table. Which month overall has the lowest ***sum of sales amount***?

Chart, bar chart

Description automatically generated

June has the overall lowest sum of sales amount

(2). Create and display a line chart, with ***sum of “Sales Amount”*** as rows, and **continuous** type ***month of “Order Date”*** from FactResellerSales table. Which month and year has the highest sum of sales amount?

January 2013 has the highest sum of sales amount

Chart, line chart

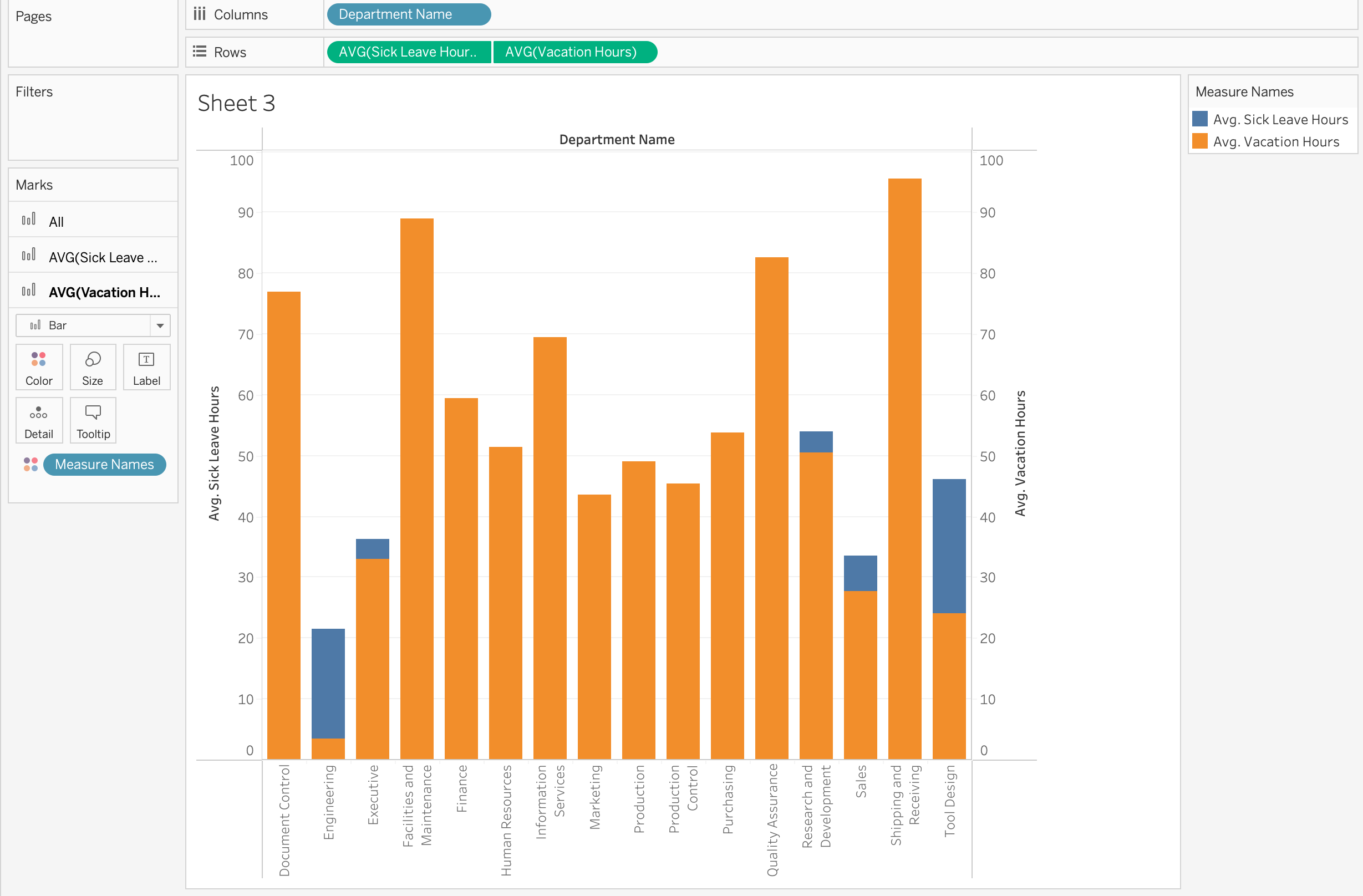
Description automatically generated

(3). When do we usually use continuous date field, and when do we use discrete date field?

We use a continuous date field to analyze data that occurs at regular intervals over time, and we use a discrete date field to group data by specific time periods or categories. Continuous date fields are useful for visualizing trends over time, while discrete date fields are useful for comparing data across different time periods or categories (like years, quarters, months, weeks and days)

3. Dual axis and Combined axis chart:

(1). Compare the ***average of “Sick Leave Hours”*** (from DimEmployee) and the ***average of “Vacation Hours”*** (from DimEmployee) using a dual axis chart, with ***“Department Name” (from DimEmployee)*** as columns; synchronize the two axis and display the chart;



(2). Based on the dual axis chart, which departments have higher average sick leave hours than average vacation hours?

Engineering, Executive, Research and Development, Sales, Tool Design are the departments that have higher average sick leave hours than average vacation hours.

(3). Generate and display a combined axis chart, with ***“Product Line”*** (from DimProduct) as columns, and ***average of “Dealer Price”***, ***average of “List Price”***, and ***average of “Standard Cost”*** (all from DimProduct) as rows;

Chart, bar chart

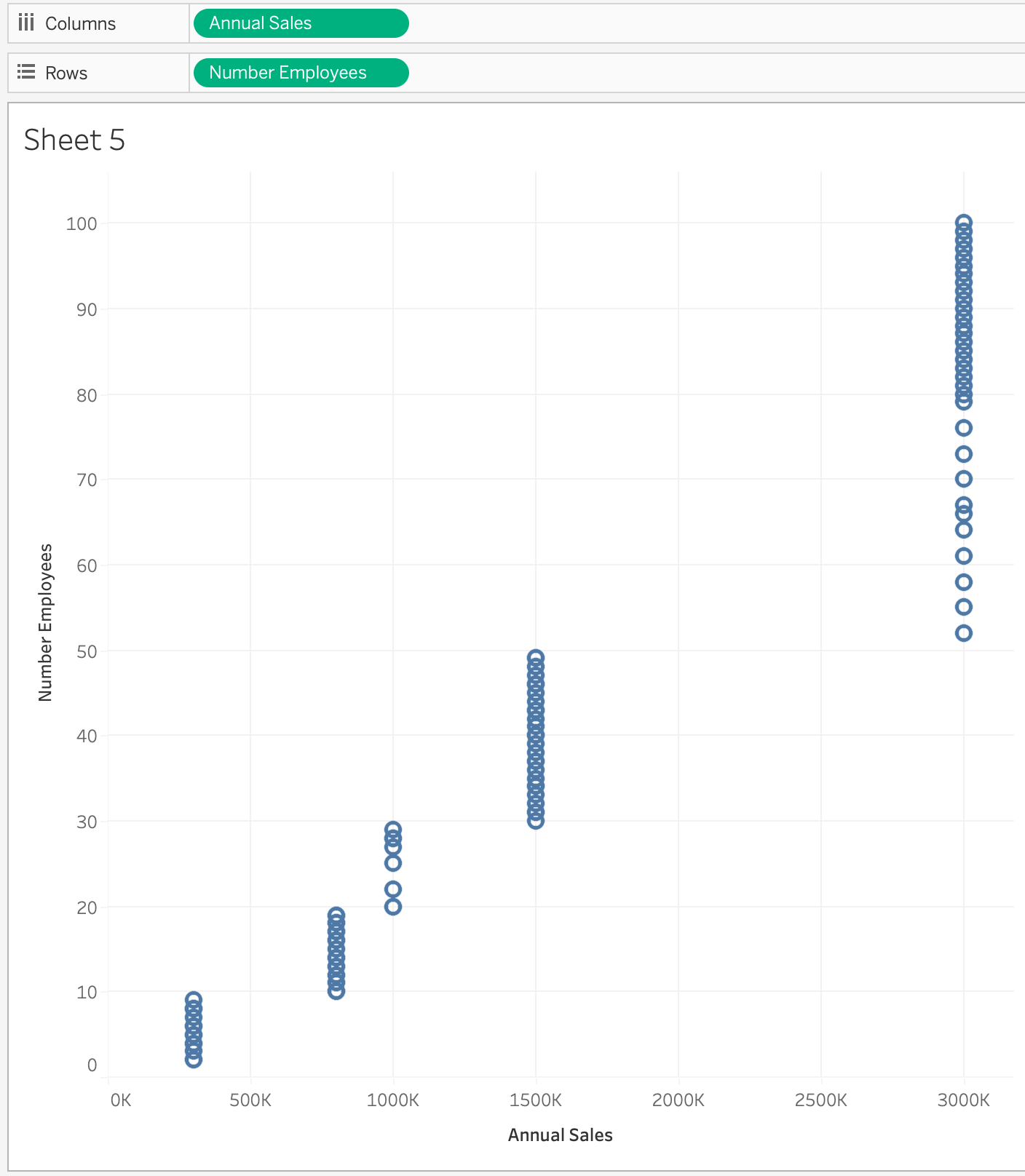
Description automatically generated

(4). Based on the prior combined axis chart, which product line has the highest average “List Price” and which product line has the lowest average “List Price”?

Product line M has the highest average “List Price” and product line S has the lowest average “List Price”

4. Analyze the correlation coefficient between two facts:

(1). Generate and display a scatter plot with facts ***“Annual Sales”*** and ***“Number Employees”*** from DimReseller; (hint: uncheck the aggregated measures)



(2). Generate and display ***a linear trend line*** on top of the scatter plot with facts “Annual Sales” and “Number Employees”, to enable the confidence interval of the linear line.

Chart, line chart

Description automatically generated

(3). Based on the linear trend line, what is the R square value? What is the p value? What is the direction of the correlation coefficient between Annual Sales” and “Number Employees”? Are the correlation coefficient significant?

R square value = 0.91797, p-value < 0.0001, and the direction of the correlation coefficient between Annual Sales and the Number of Employees is positive. The correlation coefficient is significant since p-value < 0.05 and they are strongly positively correlated as R square is closer to +1.

5. Create Calculated Fields:

(1). Create and display calculated fields:

a. ***Profit Margin*** = ((Unit Price – Product Standard Cost)\*Order Quantity)/Sales Amount;

b. ***Total Average Profit Margin*** = Average ((Unit Price – Product Standard Cost)\*Order Quantity)/ Average (Sales Amount);

Chart, waterfall chart

Description automatically generated

(2). Generate and display two column charts, with ***average of “Profit Margin”*** and “***Sales Territory Region”*** (from DimSalesTerritory) in one chart, and “Total Average Profit Margin” and “Sales Territory Region” in another chart; Why are these two charts look different?

Chart, bar chart

Description automatically generated

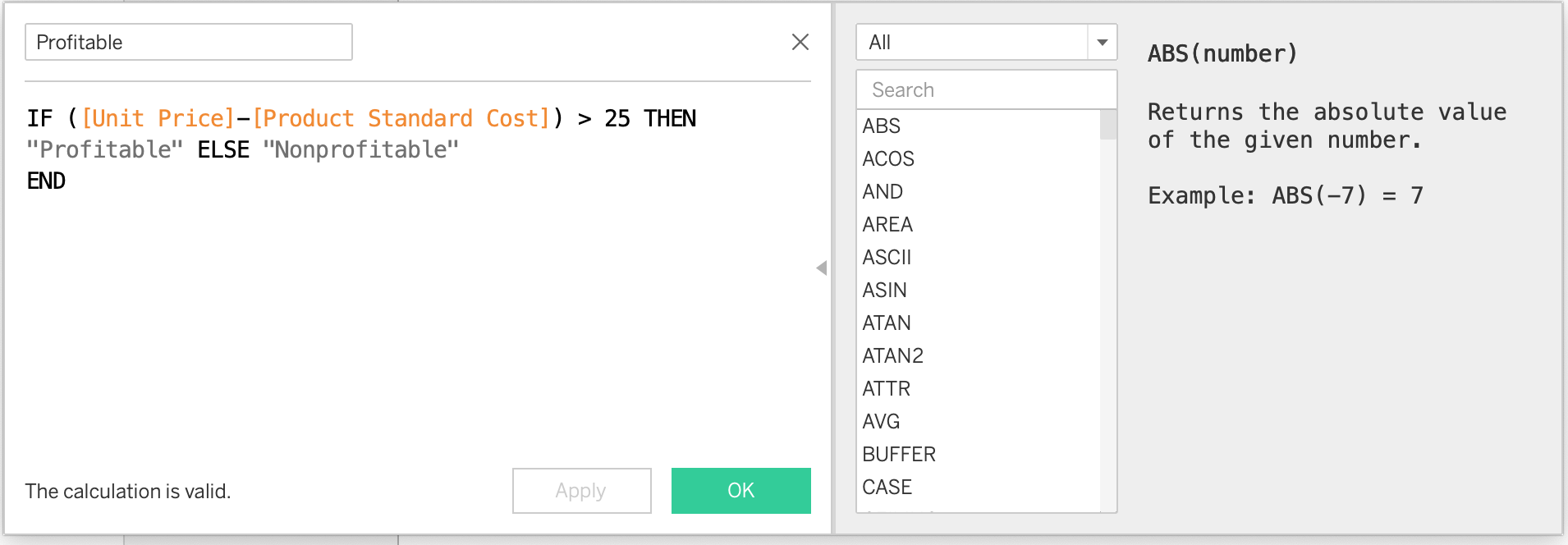
A picture containing timeline

Description automatically generated

The reason why the two charts look different is that the first one only shows the average profit margin, while the second one calculates the total average profit margin by dividing the average difference between the unit price multiplied by the order quantity and the average sales amount.

(3). Create another calculated field using IF THEN ELSE statement:

Create and display and new calculated field ***“Profitable”***: If “Unit Price” – “Product Standard Cost” is greater than 25, then we assign a value “Profitable”; otherwise, we assign the value “Nonprofitable”;



(4). Create and display a cross table, with the levels from ***“Profitable”*** as rows, and ***average of “Profit margin”*** as cell values in this cross table;

Graphical user interface, application

Description automatically generated

(5). What are the ***average profit margins*** for nonprofitable and profitable categories respectively?

Average profit margins for Nonprofitable category = 0.07130 and Profitable category = 0.12615

Graphical user interface, text, application

Description automatically generated

6. Forecasting and Clustering:

(1). Generate and display a forecasting line predicting sum of “Sales Amount” on the basis of month of “Order Date” (continuous) from FactResellerSales table;

Graphical user interface, chart

Description automatically generated

(2). Use default settings from forecast options, how many months’ sum of sales are predicted in the model?

14 months sum of sales are predicted in the model

(3). From the summary of the models, what are the quality metrics of the forecasting model?

Text, timeline

Description automatically generated

Graphical user interface, text, application

Description automatically generated

(4). Generate a scatter plot with “Sick Leave Hours” and “Base Rate” from Dimemployee; Cluster the plots in the scatter plot; Display the clusters;

Chart, scatter chart

Description automatically generated

(5). Use the default setting, how many clusters are generated?

2 clusters are generated using the default setting.

(6). Change the cluster number to 3, display the clustered scatter plot, how many items are in each cluster?

Chart, scatter chart

Description automatically generated

Cluster 1 – 95

Cluster 2 – 106

Cluster 3 – 95

A picture containing text

Description automatically generated