DATA QUERYING PROJECT

TEAM-3 MEMBERS:

BURCU MADEN
RASHA BURHANULDEEN
KEVAL LATHIYA

12-05-2022

Table of Contents

Business Context

Adventure Works is a global firm specializing in multi-facets goods such as manufacturing and selling bikes, components, clothing, and accessories to 19820 customers in 10 different territories, including the US, Europe, and the Pacific. The company employs more than 1000 employees, most of which are sales contacts who work in the stores and count to 753 employees.

Adventure Works sells 504 unique products in 4 categories and 37 subcategories online and in 701 retail stores. The main market of Adventure Works is the US which has five territories where the company makes up more than 60% of sales on an annual basis. Although the company has many retail stores, it earned 27% of its revenue from online shopping. In 2011-2013, Adventure Works grew rapidly and increased its revenue nearly four times; however, in 2014, the company lost more than 50% of its revenue, which was approximately 20 million Dollars.

Target Question

In this study, we will attempt to formulate a cohesive analysis of the impacts that cause the decrease and increase in sales within the firm's US territories, given that the US market is the firm's primary selling market. The change in the revenue per territory will be the essential part of this study to help the company better understand its market and the effects that different factors could have on its overall sales.

We are planning to analyze the product portfolio of Adventure Works to determine the highest and lowest-performing products so that the company can make a robust decision of whether to withdraw certain products or increase the manufacturing of others. Moreover, we will review the revenue changes between seasons and highlight any special offer products and whether they were a driving force for the sales or not.

"What is the impact of online sales, number of stores, and product types in terms of category/subcategory on Adventure Works' revenue in different US territories?"

Level of Observation

We mainly focus on the territories within the US and how their total sales and revenues change depending on the variables applied to them. Therefore, we are planning to territory-level observation in the research.

Variables

Table 1: Variables

#	Variable Name	Variable Description	Sample Data Values	Data Type	Source of Data	Variable Role (Predictor / Predicted)
1	% of the company revenue in each territory	Measure the sales activities of the company in each territory	**16.85%** revenue for territory **4**	decimal	Sales.Sales order detail	Predicted
2	% of online sales	measure the impact of the online sales in total revenue in the territories.	**30.75%** of revenue coming from online shopping for territory **4**	decimal	Sales.sales order header	predictor
3	% of the sales of each product / product category / subcategory	Measure the impact of the products to the company revenue Help find the most – least popular products in each territory to withdraw or continue to sell	4.68% of territory 4 revenue comes from product 782 (productid)	decimal	Production.product + sales.Sales order detail	predictor
4	% sales in the different seasons	Measure the impact of seasonality on total sales in each territory For ex: territory 2 reach its highest sales in fall season with x% average revenue in 2014	Territory **4's 26.39%** of sales is in fall season	decimal	sales order header	predictor
5	% of average sales in product category in the different seasons	Measure the impact of seasonality on product sales in each territory For ex: the consumers in territory 2 highly buy bikes in fall semester	Sales of bike were **50%** in average in spring season for territory **4**	decimal	sales order header + production.product	predictor
6	Number of stores	compare the difference in the number of stores between the territories and measure its effect on the company revenue	With **115** stores, territory **4** has a greater number of stores	smallint	sales store	predictor

Details about Source of Data

- 1. (% of the company revenue in each territory) sales order detail table includes unit price and quantity information. To analyze each territory, we merged sales stores, customer, sales order header and sales order detail tables
- (% of the sales of each product / product category / subcategory) We merged product, sales order detail and sales order header tables to analyze sales of per product in each territory by using inner join function. For further analysis, we will use inner join function to combine product category and product subcategory tables.
- 5. (% of sales in product category in the different seasons) We merged product, sales order detail and sales order header tables to analyze sales of per product in each territory by using inner join function in different seasons. For further analysis, we will use inner join function to combine product category and product subcategory tables.
- 6. (Number of stores) For number of stores in each territory, we merged store, customer, territory data tables.

Descriptive Statistics

Revenue is a predicted (dependent variable) in this study. To conduct analysis, revenue data was normalized. We analyzed each US territory revenue per year. We had 20 different observations for revenue including 5 territory and 4 years.

According to descriptive statistics, the mean of revenue is 0.29 and the median is approximately 0.25. The standard deviation indicates how the data is close to the mean which is 0.26 in our analysis. The data appear to be skewed to the right (positive), which explains why the mean is greater than the median. Revenue has positive kurtosis with 1.89 which indicates that the distribution has heavier tails than the normal distribution.

Table 2: Descriptive Statistics for Revenue

Revenue	
Mean	0.294009921
Standard Error	0.058881314
Median	0.246343165
Mode	-
Standard Deviation	0.263325243
Sample Variance	0.069340183
Kurtosis	1.890077523
Skewness	1.434610707
Count	20

We analyzed seasonality by looking each year's revenue considering seasons. To follow consumers' demand and trend of the year appropriately, we analyzed winter season by looking previous year December data and following year January and February data. (For example, for year 2012 Winter season, we used December 2011, January 2012 and February 2012.)

According to analysis, spring season showed highest mean. It shows that consumers bought more bicycle or related equipment in spring season. All standard deviations were quite similar between 0.03 – 0.04. While fall season and summer had negative (left) skewness, spring and winter season had positive (right skewness). All seasons showed negative kurtosis which indicated that the distribution has lighter tails than the normal distribution.

Table 3: Descriptive Statistics for Seasonality

Descriptive Statistics	Fall Season	Spring Season	Summer Season	Winter Season
Mean	0.246166935	0.291381399	0.247437074	0.195167219
Standard Error	0.040001467	0.04849727	0.03410752	0.032774942
Median	0.245078414	0.249888268	0.292537438	0.189529529
Mode	0	0	0	0
Standard Deviation	0.178891999	0.216886385	0.152533466	0.146573995
Sample Variance	0.032002347	0.047039704	0.023266458	0.021483936
Kurtosis	-0.905137042	-0.81618286	-0.743891189	-1.146529058
Skewness	-0.007762746	0.61978738	-0.938371749	0.030724635
Range	0.546054946	0.619802907	0.416615624	0.421744598
Minimum	0	0.019225399	0	0
Maximum	0.546054946	0.639028306	0.416615624	0.421744598
Count	20	20	20	20

To understand the impact of online sales and number of stores in revenue, we analyzed the % of online sales and number of stores in each territory.

Online sales had approximately 0.099 mean and 0.03 standard deviation while number of stores had 0.05 mean and 0.004 standard deviation. Both variables have positive skewness and negative kurtosis.

Table 4: Descriptive Statistics for Online Sales & Number of Stores

Descriptive Statistics	Online Sales	Number of Stores
Mean	0.099145	0.05
Standard Error	0.029763532	0.00438057
Median	0.00205	0.046755725
Mode	0	0.044847328
Standard Deviation	0.13310656	0.019590505
Sample Variance	0.017717356	0.000383788
Kurtosis	-0.350474934	-0.017662583
Skewness	0.965364677	0.661501799
Range	0.4032	0.073473282
Minimum	0	0.020992366
Maximum	0.4032	0.094465649
Count	20	20

To analyze the product's impact on territory revenue, we analyzed highly preferred subcategory and category revenues.

Highly preferred subcategories were x and y in each year and each territory. However, data highlighted that company changed its product portfolio frequently. Each year highly preferred product was changed, and previous years' preferred products were out of list.

In terms of highly preferred subcategories, the mean was 0.48 and standard error was 0.03. It had positive skewed and negative kurtosis.

For highly preferred product, mean was 0.07 and standard error was 0.01. It had positive skewness and kurtosis.

Table 5: Descriptive Statistics for Highly Preferred Product Subcategory/Product

Descriptive Statistics	Highly Preferred Subcategory	Highly Preferred Product
Mean	0.481283565	0.074770022
Standard Error	0.034750553	0.010261018
Median	0.452784122	0.061849706
Mode	-	-
Standard Deviation	0.155409199	0.045888669
Sample Variance	0.024152019	0.00210577
Kurtosis	-0.060736655	1.104844043
Skewness	0.623073217	1.177430371
Range	0.561199285	0.167333686
Minimum	0.291875995	0.021674345
Maximum	0.85307528	0.18900803
Count	20	20

Regression Analysis

To understand the relationship between dependent and independent analysis we conducted regression analysis. Regression analysis helped us interpret which independent variable can affect mostly, which can be eliminated, how variables interact each other and how the unit change in the independent variable impact dependent variable.

Regression Statistics, ANOVA table and Coefficients table were analyzed in detail. In model summary the value of adjusted R square was interpreted to show whether adding additional predictors improve a regression model or not by considering the values of substantial (0.75), moderate (0.5) and weak (0.25). In the ANOVA table, the statistically significance of the independent variables are analyzed with p-value < 0.05.

The coefficient table provided coefficient (β) values and p-values (significance) for each variable. The p-value tests the null hypothesis whether the variable has a significant effect on the dependent variable or not. For statistically significant variables, coefficient (β) is analyzed that shows what is the impact on dependent variable if there is one unit change in independent variables.

We run regression analysis two times. In second regression analysis we eliminated the independent variable with highest p-value.

Table 6: 1st Regression Analysis Results

Regression Statistics	
Multiple R	0.98879575
R Square	0.97771703
Adjusted R Square	0.95766236
Standard Error	0.0541821
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	1.28810648	0.14312294	48.75257679	4.39725E-07
Residual	10	0.029357	0.0029357		
Total	19	1.31746349			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3.7478666	1.330028668	-2.8178841	0.018226733	-6.711355198	-0.784378101	-6.711355198	-0.784378101
territory	0.00648089	0.018838546	0.34402299	0.737953298	-0.035494004	0.04845579	-0.035494004	0.04845579
fall season revenue %	3.60828384	1.460114847	2.4712329	0.033035054	0.354945222	6.861622461	0.354945222	6.861622461
spring season revenue %	3.03910979	1.231773418	2.46726366	0.033260313	0.294547584	5.783672004	0.294547584	5.783672004
summer season revenue %	2.24076358	1.111583083	2.01583095	0.071475953	-0.235997873	4.717525038	-0.235997873	4.717525038
winter season revenue %	3.23900066	1.41037983	2.29654494	0.044515263	0.096478566	6.381522759	0.096478566	6.381522759
Online sales %	-0.5748585	0.203846901	-2.82005	0.01815915	-1.029057654	-0.120659254	-1.029057654	-0.120659254
Number of Stores %	6.60929134	3.083501237	2.14343723	0.057699616	-0.261177562	13.47976025	-0.261177562	13.47976025
subcategory revenue %	0.91634701	0.296485925	3.09069311	0.01143046	0.255735198	1.576958814	0.255735198	1.576958814
highly preferred product revenue %	4.64846648	1.329558618	3.49624786	0.00576252	1.686025265	7.61090769	1.686025265	7.61090769

According to regression statistics table, adjusted R square was 95.8% which means that 95.8% of the variance in the dependent variable (Revenue) can be explained by the independent variables (Territory, Seasonality, Online Sales, Number of Stores, Highly Preferred Subcategory and Product).

The ANOVA table shows that there were statistically significant differences between the means of factors (independent variables) and the constant having a significance value.

Setting up the test, the null hypothesis implies that the variables presented have no significance determination on the revenue. The alternative hypothesis, on the other hand, implies that the variables presented do have a significant impact on the revenue. The result of regression showed that territory, revenue percentage in summer season and number of stores with higher than 0.05 p-value did not significant impact on revenue. On the other hand, revenue percentage in fall, spring and winter seasons, online sales and highly preferred subcategory and product had significant impact on the dependent value (revenue) with less than 0.05 p-value. Since the p-values for these variables are less than the significance level, the null hypotheses were rejected.

Territory, as having the highest p-value, was eliminated from the study and regression analysis repeated.

Table 7: 2st Regression Analysis Results

Regression Statistics	
Multiple R	0.988662382
R Square	0.977453306
Adjusted R Square	0.96105571
Standard Error	0.051965413
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	1.287759039	0.16096988	59.60955055	5.70313E-08
Residual	11	0.029704446	0.002700404		
Total	19	1.317463485			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3.388049444	0.787999245	-4.299559251	0.001256887	-5.122424089	-1.6536748	-5.122424089	-1.6536748
fall season revenue %	3.25592947	0.998057346	3.262266926	0.007568439	1.059220062	5.452638878	1.059220062	5.452638878
spring season revenue %	2.72496784	0.792880492	3.436795162	0.005556008	0.979849643	4.470086037	0.979849643	4.470086037
summer season revenue %	1.94468525	0.674726417	2.882183359	0.014912924	0.45962242	3.42974808	0.45962242	3.42974808
winter season revenue %	2.850872089	0.811733152	3.512080395	0.004865758	1.064259468	4.63748471	1.064259468	4.63748471
Online sales %	-0.56097625	0.19163828	-2.927266154	0.013757522	-0.982769259	-0.13918324	-0.982769259	-0.13918324
Number of Stores %	7.496429784	1.621452203	4.623281382	0.000736424	3.927637547	11.06522202	3.927637547	11.06522202
subcategory revenue %	0.851507791	0.219506868	3.879185191	0.002566966	0.368376432	1.33463915	0.368376432	1.33463915
highly preferred product revenue %	4.279095145	0.752094434	5.689571618	0.000140338	2.623746457	5.934443832	2.623746457	5.934443832

In second regression analysis, adjusted R square increased to 96.1% which means that 95.8% of the variance in the dependent variable (Revenue) can be explained by the independent variables (Territory, Seasonality, Online Sales, Number of Stores, Highly Preferred Subcategory and Product).

The ANOVA table shows that there were statistically significant differences between the means of factors (independent variables) and the constant having a significance value.

All hypotheses had less than 0.05 p-value which means that seasonality (revenue percentage in fall, spring, summer, winter seasons) online sales and highly preferred subcategory and product had significant impact on the dependent value (revenue) with less than 0.05 p-value. Since the p-values for these variables are less than the significance level, the null hypotheses were rejected.

Conclusion

In this study, we analyzed Adventure Works sales data in detail by using SQL and excel to understand the decline in the revenue and its reason. Although company showed incremental growth until 2014**,** in year 2014 company lost 50% of its revenue. To understand the downgrade of the company, we needed to understand important factors in revenue and determined independent variables accordingly online sales, number of stores, product type, seasonality, and shipping time on revenue for US territories. We conducted our analysis in territory level with the reason that company has many territories, and each territory might have different consumer and sales structures.

With the help of SQL, we analyzed dependent and independent variables for each territory and prepared a data set. Independent variable of shipping time was excluded from the study that all territories have same average shipping time which should not be the main factor of difference revenues in each territory and decline.

To understand the mean level, standard deviation and skewness of variables, descriptive analyses were conducted, and results were discussed in detail in ‘Descriptive Statistics’ part.

To understand the impact of independent variables on dependent variable, regression analysis was conducted. Regression analysis results showed that all independent variables have significant impact on dependent variable of each territory revenue.

Appendices

1. SQL Queries

- **The Query below gives output for revenue generated by each territories in each year and also shows the territory revenue as the percentage of total**

```
select year (S.ModifiedDate) as year,S.TerritoryID,
      sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as 'US revenue',
      sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount))/
(select sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as revenue
            from Sales.SalesOrderDetail
            JOIN sales.salesorderheader
       ON   Sales.SalesOrderDetail.salesorderid=sales.salesorderheader.salesorderid
where sales.salesorderheader.TerritoryID < 6 AND
YEAR(Sales.SalesOrderDetail.ModifiedDate) = YEAR(S.ModifiedDate))*100 as 'Percentage of
US revenue'
from Sales.SalesOrderDetail
JOIN sales.salesorderheader S
ON Sales.SalesOrderDetail.salesorderid = S.salesorderid
where S.TerritoryID < 6
group by year (S.ModifiedDate), S.TerritoryID
order by  year (S.ModifiedDate), 'Percentage of US revenue' DESC
```

- **The query to generate total US sales along with how much of it was contributed by online sale is as below:**

```
select year (S.ModifiedDate) as year,S.TerritoryID,
       sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as 'US revenue',
      (select sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount))
            from Sales.SalesOrderDetail
            JOIN sales.salesorderheader
            ON Sales.SalesOrderDetail.salesorderid =
            sales.salesorderheader.salesorderid
            where sales.salesorderheader.TerritoryID < 6
                  AND sales.salesorderheader.OnlineOrderFlag = 1
                  AND YEAR(sales.salesorderheader.ModifiedDate) = YEAR(S.ModifiedDate)
                  AND sales.salesorderheader.TerritoryID = S.TerritoryID)/
      (sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)))*100 AS 'Online sale'
from Sales.SalesOrderDetail
JOIN sales.salesorderheader S ON Sales.SalesOrderDetail.salesorderid = S.salesorderid
where S.TerritoryID < 6
group by year (S.ModifiedDate), S.TerritoryID
order by  year (S.ModifiedDate),'Online sale' DESC
```

- **The query for subcategory wise revenue generated by each territory in each year:**

```
select year(sales.SalesOrderHeader.orderdate) year,
       sales.SalesOrderHeader.territoryid  territory,
       sum(d.OrderQty*d.UnitPrice-(d.OrderQty*d.UnitPriceDiscount)) subcategory_revenue,
      (Production.ProductSubcategory.Name) subcat,
      (Production.ProductSubcategory.ProductSubcategoryID) subcat_id
from sales.SalesOrderHeader join Sales.SalesOrderDetail d
on sales.SalesOrderHeader.SalesOrderID = d.SalesOrderID
join Production.Product
```

```
on d.ProductID = Production.Product.ProductID
join Production.ProductSubcategory
on Production.Product.ProductSubcategoryID =
Production.ProductSubcategory.ProductSubcategoryID
where sales.SalesOrderHeader.TerritoryID < 6
group by year(sales.SalesOrderHeader.orderdate),sales.SalesOrderHeader.territoryid,
       (Production.ProductSubcategory.Name),
       (Production.ProductSubcategory.ProductSubcategoryID)
order by year(sales.SalesOrderHeader.orderdate), sales.SalesOrderHeader.territoryid,
        subcategory_revenue desc
```

- **The query for revenue generated in each season of the each year:**

```
select o.TerritoryID,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)
       -((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2011-06-01 00:00:00.000'
                     and '2011-09-01  00:00:00.000' and
                     o.TerritoryID = sales.SalesOrderHeader.TerritoryID) summer_2011,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)
       -((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2012-06-01 00:00:00.000'
                     and '2012-08-31  00:00:00.000' and
                      o.TerritoryID = sales.SalesOrderHeader.TerritoryID) summer_2012,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
       ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2013-06-01 00:00:00.000'
                     and '2013-08-31  00:00:00.000' and o.TerritoryID =
                     sales.SalesOrderHeader.TerritoryID) summer_2013,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
       ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2014-06-01 00:00:00.000'
                     and '2014-08-31  00:00:00.000' and o.TerritoryID =
                     sales.SalesOrderHeader.TerritoryID) summer_2014,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
       ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2010-12-01 00:00:00.000'
                     and '2011-02-28  00:00:00.000' and o.TerritoryID =
                     sales.SalesOrderHeader.TerritoryID) winter_2011,
       (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
       ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
              from sales.SalesOrderHeader join sales.SalesOrderDetail
              on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
              where sales.salesorderheader.OrderDate between '2011-12-01 00:00:00.000'
```

```
            and '2012-02-28 00:00:00.000' and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) winter_2012,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2012-12-01 00:00:00.000'
            and '2013-02-28 00:00:00.000' and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) winter_2013,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2013-12-01 00:00:00.000'
            and '2014-02-28  00:00:00.000' and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) winter_2014,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2011-03-01 00:00:00.000'
            and '2011-05-31  00:00:00.000' and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) spring_2011,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2012-03-01 00:00:00.000'
            and '2012-05-31  00:00:00.000'  and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) spring_2012,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2013-03-01 00:00:00.000'
            and '2013-05-31  00:00:00.000'  and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) spring_2013,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2014-03-01 00:00:00.000'
            and '2014-05-31  00:00:00.000'  and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) spring_2014,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2011-09-01 00:00:00.000'
            and '2011-11-30  00:00:00.000'  and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) fall_2011,
(select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
      from sales.SalesOrderHeader join sales.SalesOrderDetail
      on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
      where sales.salesorderheader.OrderDate between '2012-09-01 00:00:00.000'
            and '2012-11-30  00:00:00.000'  and o.TerritoryID =
            sales.SalesOrderHeader.TerritoryID) fall_2012,
```

```
        (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
        ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
               from sales.SalesOrderHeader join sales.SalesOrderDetail
               on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
               where sales.salesorderheader.OrderDate between '2013-09-01 00:00:00.000'
                      and '2013-11-30  00:00:00.000'  and o.TerritoryID =
                      sales.SalesOrderHeader.TerritoryID) fall_2013,
        (select sum((sales.SalesOrderDetail.OrderQty)*(sales.SalesOrderDetail.UnitPrice)-
        ((sales.SalesOrderDetail.OrderQty)*sales.SalesOrderDetail.UnitPriceDiscount))
               from sales.SalesOrderHeader join sales.SalesOrderDetail
               on sales.salesorderheader.SalesOrderID =sales.SalesOrderDetail.SalesOrderID
               where sales.salesorderheader.OrderDate between '2014-09-01 00:00:00.000'
                      and '2014-11-30  00:00:00.000'  and o.TerritoryID =
                      sales.SalesOrderHeader.TerritoryID) fall_2014
from sales.SalesOrderHeader o join sales.SalesOrderDetail
on o.SalesOrderID = sales.SalesOrderDetail.SalesOrderID
where o.TerritoryID<6
group by o.TerritoryID
```

- **The query below shows the number of stores in each territory in each year along with revenue**

```
select YEAR(SalesOrderHeader.OrderDate)  'Year',sales.salesorderheader.TerritoryID,
       COUNT(distinct(storeid)) 'Number of stores',
       sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as 'Revenue',
 (select sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as 'Revenue'
       from sales.salesorderheader s
       join Sales.Customer on s.CustomerID = Sales.Customer.CustomerID
       join Sales.SalesOrderDetail ON Sales.SalesOrderDetail.salesorderid =
              s.salesorderid
       where s.TerritoryID < 6 and OnlineOrderFlag =1 and s.TerritoryID =
              sales.salesorderheader.TerritoryID and
              YEAR(s.OrderDate) = year(sales.salesorderheader.OrderDate)
       group by YEAR(s.OrderDate),s.TerritoryID) 'online sales',
(select sum(OrderQty*UnitPrice-(OrderQty*UnitPriceDiscount)) as 'Revenue'
       from sales.salesorderheader s
       join Sales.Customer on s.CustomerID = Sales.Customer.CustomerID
       join Sales.SalesOrderDetail ON Sales.SalesOrderDetail.salesorderid =
              s.salesorderid
       where s.TerritoryID < 6 and OnlineOrderFlag =0 and s.TerritoryID =
              sales.salesorderheader.TerritoryID and
              YEAR(s.OrderDate) = year(sales.salesorderheader.OrderDate)
       group by YEAR(s.OrderDate),s.TerritoryID) 'offline sales'
from sales.salesorderheader
join Sales.Customer on Sales.SalesOrderHeader.CustomerID = Sales.Customer.CustomerID
join Sales.SalesOrderDetail ON Sales.SalesOrderDetail.salesorderid =
sales.salesorderheader.salesorderid
where sales.salesorderheader.TerritoryID < 6
group by YEAR(SalesOrderHeader.OrderDate),sales.salesorderheader.TerritoryID
order by YEAR(SalesOrderHeader.OrderDate),COUNT(distinct(storeid)) desc
```

2. Data Used in Analyses

Raw Data						
productID revenue	SubCategory ID	sub_category	subcategory revenue	Online sales %	Number of Stores	Whole US revenue by territory
235506.582	1	MountainBikes	1483257.9	20.91	32	2307294.96
75143.67	2	RoadBikes	534564.391	0	22	626632.143
92887.4468	1	MountainBikes	582426.001	0	26	1000918.84
257635.44	2	RoadBikes	1569093.43	27.98	47	2791392.96
184678.973	1	MountainBikes	1152949.64	0	34	1642280.52
205319.637	1	MountainBikes	2394026.46	18.62	62	4382547.85
116828.069	2	RoadBikes	1601488.83	0	37	2635333.93
82373.7463	2	RoadBikes	1523136.28	0	51	2734622.05
263344.043	2	RoadBikes	4461947.9	15.9	88	7718695.74
122037.998	1	MountainBikes	1328276.38	0.05	61	2772935.3
393786.247	1	MountainBikes	2320801.27	20.16	73	5983781.61
143298.914	2	RoadBikes	1185971.38	0.14	45	2780356.68
142112.117	2	RoadBikes	1250500.19	0.09	51	2984893.07
436097.275	2	RoadBikes	3540227.94	20.26	99	9261523.77
123197.198	2	RoadBikes	937819.921	0.09	63	2437471.52
182130.166	1	MountainBikes	1053470.95	32.7	61	3499010.6
63480.4668	2	RoadBikes	291059.582	0.27	37	920715.25
50009.106	2	RoadBikes	370393.254	0.01	36	1215255.84
214367.076	1	MountainBikes	1326325.89	40.32	76	4544141.73
66095.712	2	RoadBikes	321861.786	0.79	47	1067714.53

Normalized Data										
year	revenue normalized	territory	fall season revenue %	spring season revenue %	summer season revenue %	winter season revenue %	Online sales %	Number of Stores %	subcategory revenue %	higly preferred product revenue %
2011	0.19463624	1	0.50673726	0.04217 3019	0.36360 4192	0	0.20 91	0.03053 4351	0.64285 578	0.10207 0427
2011	0	2	0.4562 3073	0.03446 7408	0.41661 5624	0	0	0.02099 2366	0.85307 528	0.03256 7865
2011	0.0433 4585	3	0.5460 5495	0.01922 5399	0.38564 4805	0	0	0.02480 916	0.58189 1338	0.04025 8159
2011	0.2506 9925	4	0.4845 9312	0.03609 5073	0.40337 0148	0	0.27 98	0.04484 7328	0.56211 8432	0.11166 125
2011	0.1176 2144	5	0.4262 3896	0.08473 8382	0.37243 5694	0	0	0.03244 2748	0.70204 1839	0.08004 1337
2012	0.4349 6964	1	0.2216 8853	0.27710 056	0.26613 6225	0.24653 2111	0.18 62	0.05916 0305	0.54626 3622	0.08898 7165
2012	0.2326 2617	2	0.3049 8576	0.22293 837	0.36077 0637	0.10898 7627	0	0.03530 5344	0.60769 8634	0.05063 4215
2012	0.2441 2465	3	0.2422 0546	0.28932 9488	0.28082 902	0.17035 8967	0	0.04866 4122	0.55698 2375	0.03570 1437
2012	0.8213 263	4	0.2479 5137	0.24878 261	0.33461 4094	0.17303 7223	0.15 9	0.08396 9466	0.57807 0189	0.11413 5404
2012	0.2485 6168	5	0.2182 6284	0.25099 3927	0.26900 6239	0.24194 0377	0.00 05	0.05820 6107	0.47901 4559	0.05289 224
2013	0.6204 0726	1	0.3233 7885	0.18996 0921	0.33001 0402	0.13813 8225	0.20 16	0.06965 6489	0.38784 8592	0.17067 0094
2013	0.2494 2114	2	0.1958 514	0.25659 5161	0.27224 169	0.27531 1752	0.00 14	0.04293 8931	0.42655 3684	0.06210 689
2013	0.2731 0834	3	0.2535 8132	0.24196 5657	0.30061 0215	0.20384 2807	0.00 09	0.04866 4122	0.41894 3045	0.06159 2523
2013	1	4	0.2621 8307	0.26279 0237	0.28446 4661	0.17521 6252	0.20 26	0.09446 5649	0.38225 1131	0.18900 803
2013	0.2097 1188	5	0.2333 9508	0.24642 862	0.30356 244	0.21661 3864	0.00 09	0.06011 4504	0.38475 113	0.05339 4646
2014	0.3326 4789	1	0	0.63462 3436	0.00275 8526	0.40395 995	0.32 7	0.05820 6107	0.30107 6811	0.07893 6664
2014	0.0340 5753	2	0	0.63542 2018	0	0.36457 7982	0.00 27	0.03530 5344	0.31612 3342	0.02751 294
2014	0.0681 6805	3	0	0.57825 5402	0	0.42174 4598	0.00 01	0.03435 1145	0.30478 6238	0.02167 4345
2014	0.4536 837	4	0	0.63902 8306	0.00196 0141	0.39990 336	0.40 32	0.07251 9084	0.29187 5995	0.09290 8397
2014	0.0510 814	5	0	0.63671 3981	0.00010 6733	0.36317 9287	0.00 79	0.04484 7328	0.30144 9289	0.02864 6408