Nayan Shivhare

Data Warehousing and Business Intelligence

What is a data warehouse?

A data warehouse is the central repository of Information for your organization. Data warehouse helps organizations to support and manage business intelligent (BI) activities. Data flows into the Data warehouse from the organization's operational system.

Goals of Datawarehouse and Business Intelligence:

You can readily understand goals by Walking through the halls of the organization and listening to business management people. Some of the important goals are-

1. Easy accessibility of data that has been collected from various operational systems.
2. Information in the Datawarehouse must be understandable and readable.
3. Dataware houses must be secure and trustworthy.
4. Break down data in every way possible for example- sales on weekends, sales on weekdays, or maybe you want to break sales by every Saturday.
5. To make an informative decision Business people need to get at data easily.
6. For making more fact-based decisions.

"MOST DATA WAREHOUSE IS A CONSTANT ONGOING PROCESS - THE MAIN USERS OF DATA WAREHOUSE WILL BE BUSINESS PEOPLE."

Datawarehouse Nuts and Bolts:

Now, let's understand Dataware housing system goals. And, how data is being presented in Datawarehouse.

The best way to present analytical data is achieved using “Dimensional modeling” which is a widely accepted technique. Dimensional models implemented in Relational database systems are referred to as star schemas. And, star schemas are made up of Fact and Dimension Tables. And multi-dimensional database environments are referred to as online analytical processing(OLAP). OLAP cubes rely on indexing and other techniques to form relationships between facts and dimensions.

Fact Tables:- In fact tables, data should be presented at the lowest level possible to describe business processes or events. Events are facts, and looking at the fact table you can see the story of when, how, where, and who did it. Facts are semi-additive or non-additive.

Dimension Tables:- Facts tables have two or more foreign keys which are used to connect to dimension tables. Dimension tables contain textual context associated with business events that have been recorded in facts tables. Dimension tables will have fewer rows than the facts tables but almost way more attributes. And, finding the most possible attributes is key to Datawarehouse reporting as dimensional attributes are what are usually used for queries.

Data Warehouse Architecture:

1. Operational Source Systems:

   These are the systems that provide data for the Datawarehouse. These systems capture the day-to-day transaction data of an organization. And, Datawarehouse designers have no control over these systems. Most of the time several systems can feed data into Datawarehouse. And, data coming from these systems oftentimes need to be adjusted or expended before being loaded into Datawarehouse.

2. ETL(Extract Transform Load) System:

   It is the area where data coming from different source systems will often have different formats and names. So, with the help of ETL, we adjust the data.

   a. Extract: It is a process of extracting necessary data coming from source systems. It takes attention to detail and must be completed by considering future goals. Extracting everything will increase the cost, and size and reduce performance so identifying and understanding necessary data would be very useful.
   b. Transform: Data is transformed before it is loaded into the Data warehouse. Transformations can take various forms depending on business requirements. But the most basic level transformation includes correcting misspellings, parsing data

into standard formats, changing variable names, etc. During transformation, we also need to consider the way to deal with data duplications.

c. Load: In this stage, data is introduced or loaded into DW systems. The key task of linking facts and dimension tables must be created.

3. Data presentation Area:

   After data is loaded in the ETL process it remains in the presentation area. This all BI users to see or get access to the Data warehouse. Everything else is behind the scenes for the BI users. At this stage, data is been loaded into the dimensional format and can be accessed for queries and reports.

4. Business Intelligence Applications

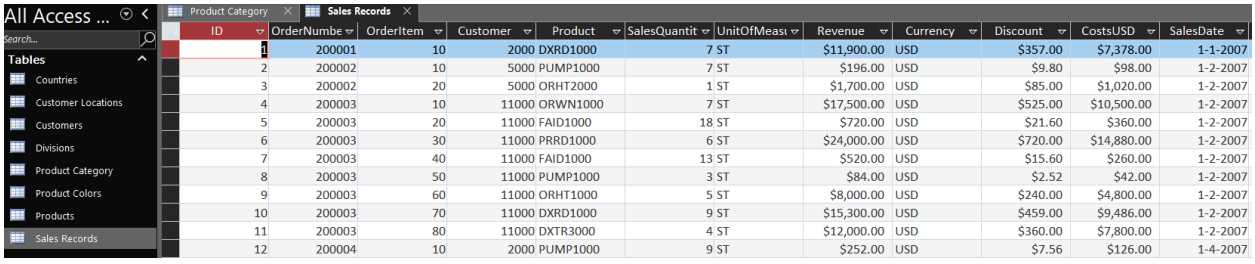
   It is an application that is used by Data Warehouse end-user to present the Data Warehouse data easily and understandably. It can be from simple query writing applications to complex data visualization tools.

Building Dataware House:

For Building we have taken SAP Relational Database for the Bike sales.

Database from Operational Source System which contains different data and information.

SAP Database



ID	OrderNumbe	OrderItem	Customer	Product	SalesQuantit	UnitOfMeasu	Revenue	Currency	Discount	CostsUSD	SalesDate
1	200001	10	2000	DXRD1000	7	ST	$11,900.00	USD	$357.00	$7,378.00	1-1-2007
2	200002	10	5000	PUMP1000	7	ST	$196.00	USD	$9.80	$98.00	1-2-2007
3	200002	20	5000	ORHT2000	1	ST	$1,700.00	USD	$85.00	$1,020.00	1-2-2007
4	200003	10	11000	ORWN1000	7	ST	$17,500.00	USD	$525.00	$10,500.00	1-2-2007
5	200003	20	11000	FAID1000	18	ST	$720.00	USD	$21.60	$360.00	1-2-2007
6	200003	30	11000	PRRD1000	6	ST	$24,000.00	USD	$720.00	$14,880.00	1-2-2007
7	200003	40	11000	FAID1000	13	ST	$520.00	USD	$15.60	$260.00	1-2-2007
8	200003	50	11000	PUMP1000	3	ST	$84.00	USD	$2.52	$42.00	1-2-2007
9	200003	60	11000	ORHT1000	5	ST	$8,000.00	USD	$240.00	$4,800.00	1-2-2007
10	200003	70	11000	DXRD1000	9	ST	$15,300.00	USD	$459.00	$9,486.00	1-2-2007
11	200003	80	11000	DXTR3000	4	ST	$12,000.00	USD	$360.00	$7,800.00	1-2-2007
12	200004	10	2000	PUMP1000	9	ST	$252.00	USD	$7.56	$126.00	1-4-2007

Raw Data screenshot.

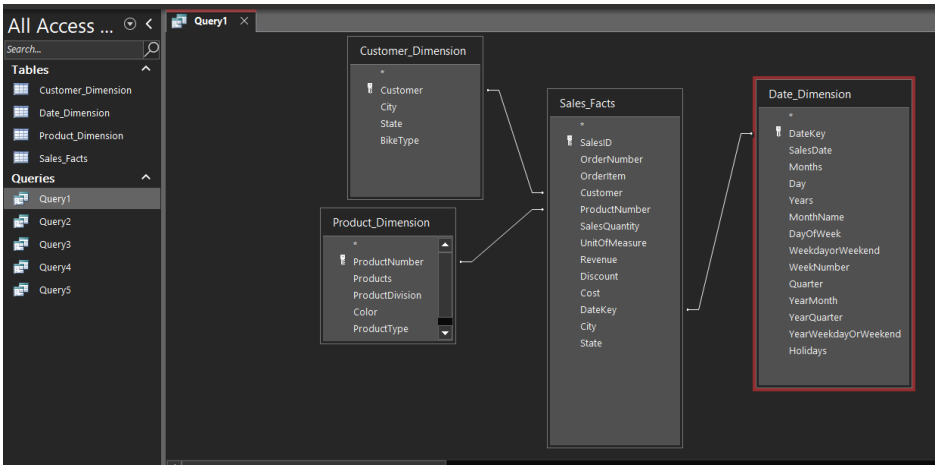
ETL(Extract, Transform, Load) System:

Extracting data- Extracted data from a relational database.

Transform - Formed relationship, cleaned the data, and transformed data into dimensional modeling.

For Dimensional modeling, we need a Sales facts table and a dimension table. Based on the database I have created one sales facts table, Date dimension table, customer dimension table, and Product Dimension table to give end users a better way to form relationships and filter queries based on the business requirement.

Load- After transformation, we loaded data into MS Access and formed relationships between tables for query and getting helpful insight from our transformed data.



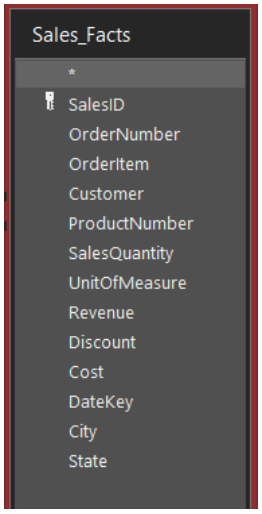
Screenshot of Data after being transformed.

Dimension Table components:

Sales Fact Table:

A fact table is presenting it to the lowest grain possible. We have selected SalesID as our primary key.

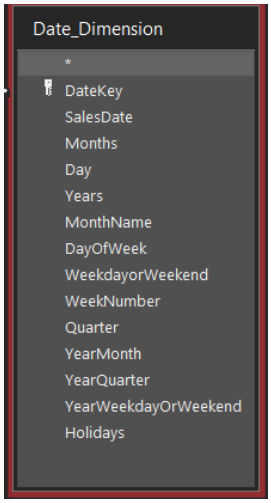
SalesID	OrderNumber	OrderItem	Customer	ProductNumber	SalesQuantity	UnitOfMeasure	Revenue	Discount	Cost	DateKey	City	State
1	200001	10	2000	DXRD1000	7	ST	11,900.00	357.00	7,378.00	01012007	New York City	New York State
2	200002	10	5000	PUMP1000	7	ST	196.00	9.80	98.00	01022007	Boston	Massachusetts
3	200002	20	5000	ORHT2000	1	ST	1,700.00	85.00	1,020.00	01022007	Boston	Massachusetts
4	200003	10	11000	ORWN1000	7	ST	17,500.00	525.00	10,500.00	01022007	Washington DC	District of Columbia
5	200003	20	11000	FAID1000	18	ST	720.00	21.60	360.00	01022007	Washington DC	District of Columbia
6	200003	30	11000	PRRD1000	6	ST	24,000.00	720.00	14,880.00	01022007	Washington DC	District of Columbia
7	200003	40	11000	FAID1000	13	ST	520.00	15.60	260.00	01022007	Washington DC	District of Columbia
8	200003	50	11000	PUMP1000	3	ST	84.00	2.52	42.00	01022007	Washington DC	District of Columbia
9	200003	60	11000	ORHT1000	5	ST	8,000.00	240.00	4,800.00	01022007	Washington DC	District of Columbia
10	200003	70	11000	DXRD1000	9	ST	15,300.00	459.00	9,486.00	01022007	Washington DC	District of Columbia
11	200003	80	11000	DXTR3000	4	ST	12,000.00	360.00	7,800.00	01022007	Washington DC	District of Columbia
12	200004	10	2000	PUMP1000	9	ST	252.00	7.56	126.00	01042007	New York City	New York State
13	200004	20	2000	DXTR3000	2	ST	6,000.00	180.00	3,900.00	01042007	New York City	New York State
14	200005	10	5000	DXTR3000	6	ST	18,000.00	900.00	11,700.00	01042007	Boston	Massachusetts
15	200005	20	5000	RKIT1000	8	ST	256.00	12.80	128.00	01042007	Boston	Massachusetts
16	200005	30	5000	CAGE1000	8	ST	144.00	7.20	72.00	01042007	Boston	Massachusetts
17	200005	40	5000	PRRD2000	3	ST	12,600.00	630.00	7,812.00	01042007	Boston	Massachusetts
18	200005	50	5000	PRTR3000	8	ST	25,600.00	1,280.00	16,640.00	01042007	Boston	Massachusetts
19	200005	60	5000	PUMP1000	9	ST	252.00	12.60	126.00	01042007	Boston	Massachusetts
20	200005	70	5000	OHMT1000	2	ST	100.00	5.00	50.00	01042007	Boston	Massachusetts



Date Dimension:

We have broken down dates between 2009 to 2017 in every possible way including holidays.

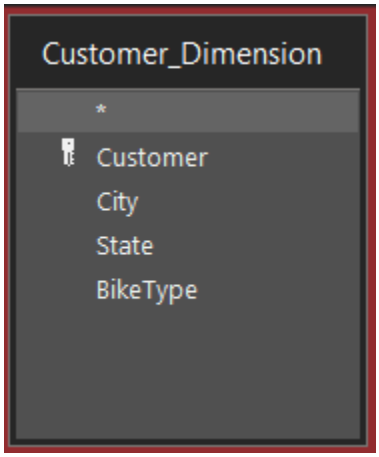
DateKey	SalesDate	Months	Day	Years	MonthName	DayOfWeek	WeekdayorWeekend	WeekNumber	Quarter	YearMonth	YearQuarter	YearWeekdayOrWeekend	Holidays
01012007	01-01-2007	1	1	2007	January	Monday	Weekday	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	Its new year
01022007	01-02-2007	1	2	2007	January	Tuesday	Weekday	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01032007	01-03-2007	1	3	2007	January	Wednesday	Weekday	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01042007	01-04-2007	1	4	2007	January	Thursday	Weekday	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01052007	01-05-2007	1	5	2007	January	Friday	Weekday	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01062007	01-06-2007	1	6	2007	January	Saturday	Weekend	1	Quarter 1	2007-January	2007-Quarter 1	2007-Weekend	No Holiday
01072007	01-07-2007	1	7	2007	January	Sunday	Weekend	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekend	No Holiday
01082007	01-08-2007	1	8	2007	January	Monday	Weekday	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01092007	01-09-2007	1	9	2007	January	Tuesday	Weekday	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01102007	01-10-2007	1	10	2007	January	Wednesday	Weekday	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01112007	01-11-2007	1	11	2007	January	Thursday	Weekday	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01122007	01-12-2007	1	12	2007	January	Friday	Weekday	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01132007	01-13-2007	1	13	2007	January	Saturday	Weekend	2	Quarter 1	2007-January	2007-Quarter 1	2007-Weekend	No Holiday
01142007	01-14-2007	1	14	2007	January	Sunday	Weekend	3	Quarter 1	2007-January	2007-Quarter 1	2007-Weekend	No Holiday
01152007	01-15-2007	1	15	2007	January	Monday	Weekday	3	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01162007	01-16-2007	1	16	2007	January	Tuesday	Weekday	3	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01172007	01-17-2007	1	17	2007	January	Wednesday	Weekday	3	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday
01182007	01-18-2007	1	18	2007	January	Thursday	Weekday	3	Quarter 1	2007-January	2007-Quarter 1	2007-Weekday	No Holiday



Customer Dimension:

We have included all textual context for the customer in the customer dimension. We have taken the customer ID and based on it we have included the city, state, and bike type they ordered.

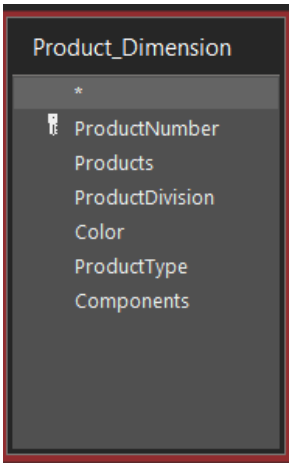
Customer	City	State	BikeType
1000	Denver	Colorado	Rocky Mountain Bikes
2000	New York City	New York State	Big Apple Bikes
3000	Philadelphia	Pennsylvania	Philly Bikes
4000	Atlanta	Georgia	Peachtree Bikes
5000	Boston	Massachusetts	Beantown Bikes
6000	Chicago	Illinois	Windy City Bikes
7000	Grand Rapids	Michigan	Furniture City Bikes
8000	Detroit	Michigan	Motown Bikes
9000	Irvine	California	SoCal Bikes
10000	Palo Alto	California	Silicon Valley Bikes
11000	Washington DC	District of Columbia	DC Bikes
12000	Seattle	Washington State	Northwest Bikes



Product DImension:

Product dimension consists of the product number, color, product type, and other relevant information.

ProductNumber	Products	ProductDivision	Color	ProductType	Components
BOTL1000	Water Bottle	Accessories	Green	Accessories	
CAGE1000	Water Bottle Cage	Accessories	Green	Accessories	
CITY1000	City Bike Max	Bicycles	Green	Trend Bikes	
DXRD1000	Deluxe Road Bike (Shimano)	Bicycles	Black	Road Bikes	Shimano Ultegra
DXRD2000	Deluxe Road Bike (SRAM)	Bicycles	Black	Road Bikes	SRAM Force
DXTR1000	Deluxe Touring Bike (black)	Bicycles	Black	Touring Bikes	Shimano Road
DXTR2000	Deluxe Touring Bike (silver)	Bicycles	Silver	Touring Bikes	Shimano Road
DXTR3000	Deluxe Touring Bike (red)	Bicycles	Red	Touring Bikes	Shimano Road
ELBK1000	E-Bike Tailwind	Bicycles	Red	Electric Bikes	
EPAD1000	Elbow Pads	Accessories	Red	Accessories	
FAID1000	First Aid Kit	Accessories	Red	Accessories	
FXGR1000	Fixed Gear Bike Plus	Bicycles	Red	Trend Bikes	
HVBD1000	Hoverboard	Bicycles	Red	Electric Bikes	
KPAD1000	Knee Pads	Accessories	Red	Accessories	
OHMT1000	Off Road Helmet	Accessories	Red	Accessories	
ORHT1000	Men's Off Road Bike Hard Tail (Shimano)	Bicycles	Red	Offroad Bikes	Shimano XT
ORHT2000	Men's Off Road Bike Hard Tail (SRAM)	Bicycles	Red	Offroad Bikes	SRAM X9
ORMN1000	Men's Off Road Bike Fully	Bicycles	Red	Offroad Bikes	SRAM XO
ORWN1000	Women's Off Road Bike Fully	Bicycles	Red	Offroad Bikes	SRAM XO



Queries Created:

1. I wanted to see the sum of revenue for each type of bike sold in Chicago with the no of quantities sold.

SumOfRevenue	Products	City	BikeType	SalesQuantit
1448601.8	Professional Road Bike (Shimano)	Chicago	Windy City Bikes	7
1346789.73	Professional Road Bike (Shimano)	Chicago	Windy City Bikes	6
1306461.25	Professional Road Bike (Shimano)	Chicago	Windy City Bikes	9
1001258.9	Professional Touring Bike (silver)	Chicago	Windy City Bikes	8
939916.4	Professional Road Bike (Shimano)	Chicago	Windy City Bikes	5
880472.65	Deluxe Touring Bike (silver)	Chicago	Windy City Bikes	9
873269.470000001	Deluxe Road Bike (Shimano)	Chicago	Windy City Bikes	9
832229.82	Professional Touring Bike (silver)	Chicago	Windy City Bikes	14
826438.93	Men's Off Road Bike Hard Tail (SRAM)	Chicago	Windy City Bikes	8
824817.68	Professional Touring Bike (silver)	Chicago	Windy City Bikes	7
796102.37	Deluxe Touring Bike (silver)	Chicago	Windy City Bikes	8
763693.94	Men's Off Road Bike Fully	Chicago	Windy City Bikes	9
745632.78	Deluxe Road Bike (Shimano)	Chicago	Windy City Bikes	8
740791.07	Deluxe Touring Bike (silver)	Chicago	Windy City Bikes	7
740349.26	Professional Touring Bike (silver)	Chicago	Windy City Bikes	10
738293.29	Men's Off Road Bike Hard Tail (SRAM)	Chicago	Windy City Bikes	9
715440.63	Men's Off Road Bike Hard Tail (SRAM)	Chicago	Windy City Bikes	7
713941.85	Professional Road Bike (Shimano)	Chicago	Windy City Bikes	8
61632364.56				

2. I wanted to see how many bicycles are sold in 2016, to different customers, with the color Black.

Query2

Customer	ProductDivisi	SumOfSalesQuantity	Years	Color
1000	Bicycles	412	2016	Black
11000	Bicycles	493	2016	Black
12000	Bicycles	347	2016	Black
2000	Bicycles	528	2016	Black
3000	Bicycles	467	2016	Black
4000	Bicycles	426	2016	Black
5000	Bicycles	985	2016	Black
6000	Bicycles	365	2016	Black
7000	Bicycles	183	2016	Black
8000	Bicycles	515	2016	Black
9000	Bicycles	416	2016	Black

3. I wanted to find out how much revenue the state of “District of Columbia” generates every year on the New Year Holiday from a different line of products that have been sold.

SalesQuantity	Products	Holidays	State	Years	SumOfReven
1	Deluxe Road Bike (Shimano)	Its new year	District of Columbia	2012	1874.47
1	Professional Road Bike (SRAI	Its new year	District of Columbia	2012	4631.04
1	Water Bottle Cage	Its new year	District of Columbia	2015	20.43
12	Deluxe Touring Bike (black)	Its new year	District of Columbia	2012	39694.6
13	Deluxe Touring Bike (silver)	Its new year	District of Columbia	2015	44256.1
17	First Aid Kit	Its new year	District of Columbia	2011	731.46
2	Deluxe Touring Bike (silver)	Its new year	District of Columbia	2010	6347.78
2	Men's Off Road Bike Fully	Its new year	District of Columbia	2010	5078.22
2	Men's Off Road Bike Hard Ta	Its new year	District of Columbia	2015	3858.22
2	Professional Road Bike (Cam	Its new year	District of Columbia	2012	9923.65
2	T-shirt	Its new year	District of Columbia	2015	68.09
3	Deluxe Road Bike (Shimano)	Its new year	District of Columbia	2012	5623.4
3	Professional Road Bike (Cam	Its new year	District of Columbia	2012	14885.48
5	Men's Off Road Bike Fully	Its new year	District of Columbia	2011	12908.14
5	Professional Road Bike (SRAI	Its new year	District of Columbia	2011	22589.25
5	T-shirt	Its new year	District of Columbia	2011	161.35
5	Water Bottle Cage	Its new year	District of Columbia	2011	96.81
6	Elbow Pads	Its new year	District of Columbia	2010	476.08
6	Men's Off Road Bike Fully	Its new year	District of Columbia	2012	15877.84
6	Women's Off Road Bike Fully	Its new year	District of Columbia	2012	16539.42
7	Water Bottle Cage	Its new year	District of Columbia	2015	142.98
8	Men's Off Road Bike Hard Ta	Its new year	District of Columbia	2011	14629.23
9	Air Pump	Its new year	District of Columbia	2015	285.96

4. I wanted to find out Accessories and Bicycles sold in Quarter 1 of the year 2010, and revenue generated by each state.

Query4

ProductDivisi	Quarter	Years	State	SumOfSalesQuantity
Accessories	Quarter 1	2010	California	358
Bicycles	Quarter 1	2010	California	350
Bicycles	Quarter 1	2010	Colorado	68
Accessories	Quarter 1	2010	Colorado	33
Bicycles	Quarter 1	2010	District of Colur	68
Accessories	Quarter 1	2010	District of Colur	55
Bicycles	Quarter 1	2010	Georgia	174
Accessories	Quarter 1	2010	Georgia	133
Bicycles	Quarter 1	2010	Illinois	62
Accessories	Quarter 1	2010	Illinois	25
Bicycles	Quarter 1	2010	Massachusetts	332
Accessories	Quarter 1	2010	Massachusetts	253
Bicycles	Quarter 1	2010	Michigan	237
Accessories	Quarter 1	2010	Michigan	189
Bicycles	Quarter 1	2010	New York State	97
Accessories	Quarter 1	2010	New York State	44
Accessories	Quarter 1	2010	Pennsylvania	137
Bicycles	Quarter 1	2010	Pennsylvania	142
Bicycles	Quarter 1	2010	Washington Sta	140
Accessories	Quarter 1	2010	Washington Sta	90
Total				**2987**

5. Our organization planning to acquire a profitable and affordable “touring bikes” manufacturer company. But before we can make a decision we want to get factual numbers for our sales that come out from Touring Bikes sales in each state and overall. Specifically in the last two years 2018 and 2019.

ProductType	SumOfRevenue	State	Years
Touring Bikes	4183502.78	Massachusetts	2018
Touring Bikes	3071190.65	Colorado	2018
Touring Bikes	2765883.2	District of Columbia	2018
Touring Bikes	2025613.09	Georgia	2018
Touring Bikes	2477497.61	California	2018
Touring Bikes	2679844.28	Illinois	2018
Touring Bikes	3213000.64	Michigan	2018
Touring Bikes	2043250.11	New York State	2018
Touring Bikes	1720067.23	Pennsylvania	2018
Touring Bikes	1736750.68	Washington State	2018
Touring Bikes	3891471.12	Michigan	2019
Touring Bikes	4242782.7	Massachusetts	2019
Touring Bikes	2157082.74	Washington State	2019
Touring Bikes	3021475.08	New York State	2019
Touring Bikes	1878128.76	Georgia	2019
Touring Bikes	1844751.6	District of Columbia	2019
Touring Bikes	2263061.04	Pennsylvania	2019
Touring Bikes	2714260.38	Colorado	2019
Touring Bikes	2834125.08	California	2019
Touring Bikes	2319095.49	Illinois	2019
Total	**53082834.26**		

Conclusion:

Datawarehousing is a very powerful way to generate facts-based decisions for Businesses. But, the most important consideration when any enterprise or organization wants to create a data warehouse- is justification for the projects, preliminary scope (it will change based on need), and identification of resources. A successful Data Warehouse is very complex and needs everyone's involvement from business people, to technical people and Senior Business leads. Data warehouse is and will be an ongoing process based on business needs. Detailed planning and collecting business requirements are important before we can decide whether we need Data Warehouse or not.

Reference:

1. https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/