**Project Proposal**

**Dataset:** AdventureWorksDW2014

**Description of Dataset:** This is a sample data warehouse dataset which gives students a hands on various business scenario which includes Manufacturing, Sales, Purchasing, Product Management etc of a fictitious company. This dataset will help one to understand the basic scenarios of marketing and help one to analyze various ups and down in marketing which in turn is affected by region, product type, age group etc.

**FactInternetSales.csv**

|  |  |
| --- | --- |
| **Field** | Description |
| **ProductKey** | Primary key for **Product** rows. |
| **OrderDateKey** | YYYYMMDD Order date |
| **DueDateKey** | YYYYMMDD Due date |
| **ShipDateKey** | YYYYMMDD Ship date |
| **CustomerKey** | Foreign Key to Customer data |
| **PromotionKey** | Foreign key to promotion data |
| **CurrencyKey** | Foreign key to Currency data |
| **SalesOrderNumber** | Unique sales order identification number. |
| **SalesOrderLineNumber** | Customer purchase order number reference. |
| **RevisionNumber** | Incremental number to track changes to the sales order over time. |
| **OrderQuantity** | Quantity ordered per product. |
| **ExtendedAmount** |  |
| **UnitPriceDiscountPct** | Discount amount per unit |
| **DiscountAmount** | Discount amount |
| **TotalProductCost** | Total Standard cost of the product. |
| **SalesAmount** | Total sales price of product |
| **TaxAmt** | Tax amount. |
| **Freight** | Shipping cost. |
| **OrderDate** | Actual order date |
| **DueDate** | Actual Due Date |
| **ShipDate** | Actual Ship Date |

**Customer.csv**

|  |  |
| --- | --- |
| **Fields** | **Description** |
| **CustomerKey** | Primary key of Customer data |
| **GeographyKey** | Foreign key to Geography data |
| **CustomerAlternateKey** | Alternate key of customer table |
| **Title** | A courtesy title. For example, Mr. or Ms. |
| **CustomerFullName** | Firstname and last name of customer |
| **FirstName** | First name of the person. |
| **MiddleName** | Middle name or middle initial of the person. |
| **LastName** | Last name of the person. |
| **BirthDate** | Birthdate of customer |
| **MaritalStatus** | Married status M,S |
| **Suffix** | Surname suffix. For example, Sr. or Jr. |
| **Gender** | F,M |
| **EmailAddress** | Email address of person |
| **YearlyIncome** | Income of customer |
| **TotalChildren** | Total children of customer |
| **NumberChildrenAtHome** | Number of children at home |
| **Education** | Education of customer |
| **Occupation** | Current occupation of customer |
| **HouseOwnerFlag** | Customer is owner of house |
| **NumberCarsOwned** | Number of cars owned by customer |
| **Phone** | Contact detail of customer |
| **DateFirstPurchase** | Date at which customer bought first product |
| **CommuteDistance** | Distance of commutation |

**ProductHierarchy.csv**

|  |  |
| --- | --- |
| **Fields** | **Description** |
| **ProductKey** | Primary key for **Product** rows. |
| **ProductSubcategoryKey** | Foreign Key to ProductSubcategory |
| **ProductSubcategoryAlternateKey** | Alternate key for ProductSubcategory |
| **ProductCategoryKey** | Foreign Key to ProductCategory |
| **ProductCategoryAlternateKey** | Alternate key for ProductCategory |
| **ProductName** | Name of the product. |
| **ProductDescription** | Description of product |
| **ProductSubcategory** | Product Subcategory Name |
| **ProductCategory** | Product Category Name |
| **ProductModel** | Product Model Name |
| **ProductLine** | R = Road  M = Mountain  T = Touring  S = Standard |
| **WeightUnitMeasureCode** | Unit of measure for Weight column. |
| **SizeUnitMeasureCode** | Unit of measure for **Size** column. |
| **StandardCost** | Standard cost of the product. |
| **FinishedGoodsFlag** | 0 = Product is not a salable item.  1 = Product is salable. |
| **Color** | Product color. |
| **SafetyStockLevel** | Minimum inventory quantity. |
| **ReorderPoint** | Inventory level that triggers a purchase order or work order. |
| **ListPrice** | Selling price. |
| **Size** | Product size. |
| **SizeRange** |  |
| **Weight** | Product weight. |
| **DaysToManufacture** | Number of days required to manufacture the product. |
| **DealerPrice** |  |
| **Class** | H = High  M = Medium  L = Low |
| **Style** | W = Women's  M = Men's  U = Universal |
| **ProductStartDate** | Date the product was available for sale. |
| **ProductEndDate** | Date the product was no longer available for sale. |
| **ProductStatus** |  |

**GeoHierarchy.csv**

|  |  |
| --- | --- |
| **Fields** | **Description** |
| **GeographyKey** | Primary key |
| **City** | city |
| **StateProvinceCode** | State or province code. |
| **StateProvince** | State or province description. |
| **CountryRegionCode** | Standard code identifying countries and regions |
| **CountryRegion** | Country or region name. |
| **PostalCode** | Zipcode of given location |
| **SalesTerritoryKey** | Foreign key to Salesterritory |

**Analysis:**

Analysing TotalSales and TotalProfit for Customer segregated for each product using reduce side join and map reduce chaining for calculating total sales and total profit of products

Top 10 products by total sales using reduce side joining and top 10 filtering pattern

Analysing TotalSales and TotalProfit of Products sold segregated by countryregion using reduce side join for joining 3 files and seggerating data based on country by using binnig pattern

Total number of distinct customer in each country by using distinct and counter pattern

Forecasting the sales for next year based on algorithms.