**Report: Retail Sales & Customer Insights**

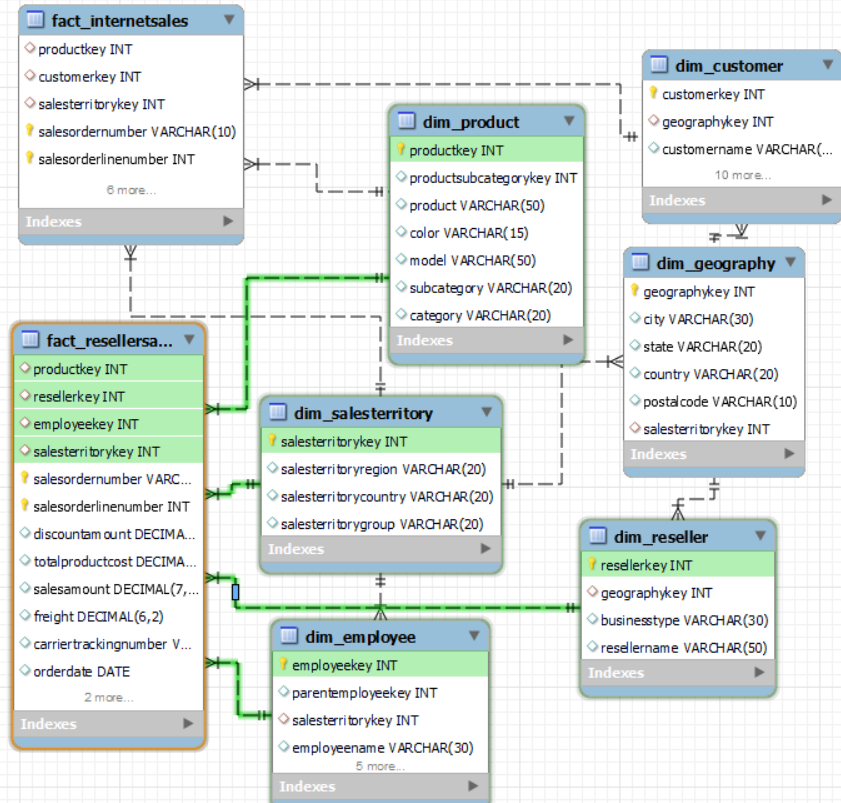
**Problem Statement**

Adventure Works is a bike manufacturer also dealing in components, accessories and clothing related to biking. It generates business through online (retail) and reseller (B2B) sales channels. Currently, the company is struggling with competitive pressures in online sales in some geographies, where it is losing out to the competition, and certain sections of the customer base. Brian Welcker, VP - Sales, needs to tackle these issues but with a data-driven approach. However, he is experiencing challenges in extracting actionable insights from the company’s disparate data sources. The company uses multiple systems for inventory management, customer relationship management (CRM), and point-of-sale (POS) transactions. However, these systems are not integrated, resulting in fragmented data that is difficult to analyze.

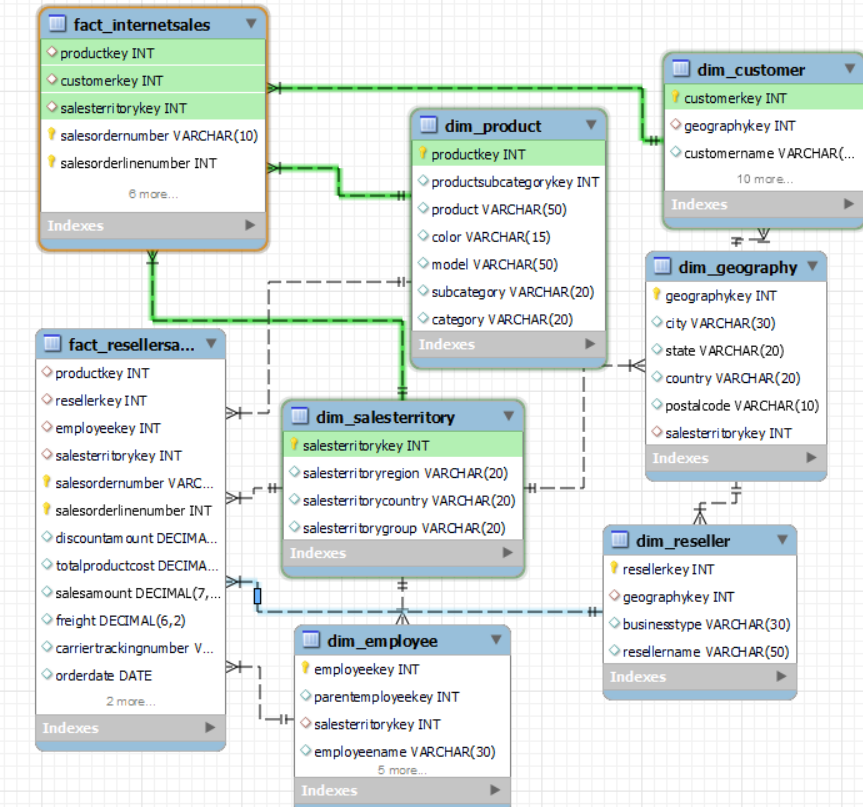
**Workflow:**

**Database Creation**:

* Created a database ‘retail1’ in MYSQL Workbench.
* **Star Schema Design –** Build a warehouse schema to optimize queries for inventory and supply chain management.
* I have created tables in the database using ‘star schema’ design.
* I created total 8 tables – 6 dimension tables and 2 fact table.
* dim\_product
* dim\_salesterritory
* dim\_geography
* dim\_employee
* dim\_customer
* dim\_reseller
* fact\_internetsales
* fact\_resellersales
* are all the tables.

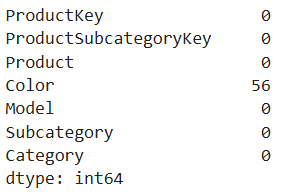


------------------------------------------------------------------------------------------------



**Importing data into Tables using Python**:

* Now I have connected the python and database tables using mysql connector library.
* The cursor is used to execute SQL queries.
* **product** data is residing in sheet ‘DimProduct’ of ‘DimTables.xlsx’.
* loading data into product DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There are null values in color column and duplicate values.

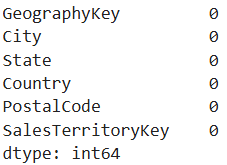
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_product table.
* Checked the dim\_product table in the MySQL Workbench all the records have been inserted.



* Findings: All the 397 records are imported.
* **SalesTerritory** data is residing in sheet ‘DimSalesTerritory’ of ‘DimTables.xlsx’.
* loading data into SalesTerritory DataFrame.
* Checked for null values and duplicates in the DataFrame.
* Findings: one row is null but cannot exclude it as other tables have data related to this row. No duplicate records.
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_ salesterritory table.
* Checked the dim\_ salesterritory table in the MySQL Workbench all the records have been inserted.



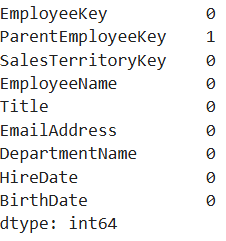
* Findings: All the 11 records are imported.
* **Geography** data is residing in sheet ‘DimGeography’ of ‘DimTables.xlsx’.
* loading data into Geography DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There are no null and duplicate values.

* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_ geography table.
* Checked the dim\_ geography table in the MySQL Workbench all the records have been inserted.



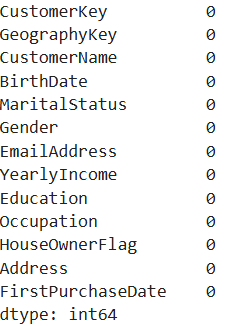
* Findings: All the 655 records are imported.
* **Employee** data is residing in sheet ‘Dim Employee’ of ‘DimTables.xlsx’.
* loading data into Employee DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There is a null value in ParentEmployeeKey column and no duplicate values.

* Filled the null value as zero in ParentEmployeeKey column and converted data type to int.
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_employee table.
* Checked the dim\_ employee table in the MySQL Workbench all the records have been inserted.



* Findings: All the 21 records are imported.
* **Customer** data is residing in sheet ‘Dim Customer’ of ‘DimTables.xlsx’.
* loading data into Customer DataFrame.
* Checked for null values and duplicates in the DataFrame.

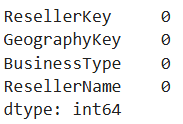
 Findings: There are no null and duplicate values.

* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_customer table.
* Checked the dim\_ customer table in the MySQL Workbench all the records have been inserted.



 Findings: All the 18484 records are imported.

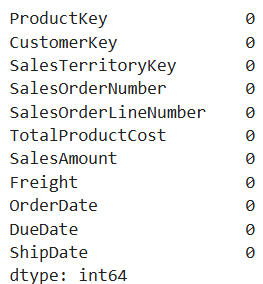
* **Reseller** data is residing in sheet ‘Dim Reseller’ of ‘DimTables.xlsx’.
* loading data into Reseller DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There are no null and duplicate values.

* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the dim\_reseller table.
* Checked the dim\_ reseller table in the MySQL Workbench all the records have been inserted.



* Findings: All the 701 records are imported.
* **Internet sales** data is residing in ‘FactInternetSales.xlsx’.
* loading data into InternetSales DataFrame.
* Drop DiscountAmount, CarrierTrackingNumber columns as both columns are empty.
* Checked for null values and duplicates in the DataFrame.

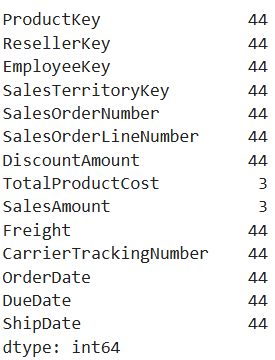
 Findings: There are no null and duplicate values.

* For TotalProductCost, Freight, SalesAmount columns rounded values to 2 decimal.
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the fact\_internetsales table.
* Checked the fact\_internetsales table in the MySQL Workbench all the records have been inserted.

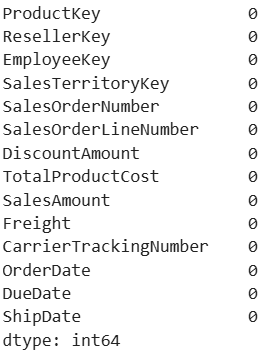


 Findings: All the 60398 records are imported.

* **Reseller Sales** data is residing in ‘Fact ResellerSales.xlsx’.
* loading data into ResellerSales DataFrame.
* Checked for null values and duplicates in the DataFrame.

 Findings: There are 44 null records.

* Drop the rows where SalesOrderNumber is null.
* After dropping there are no null values.

 Findings: There are no null and duplicate values.

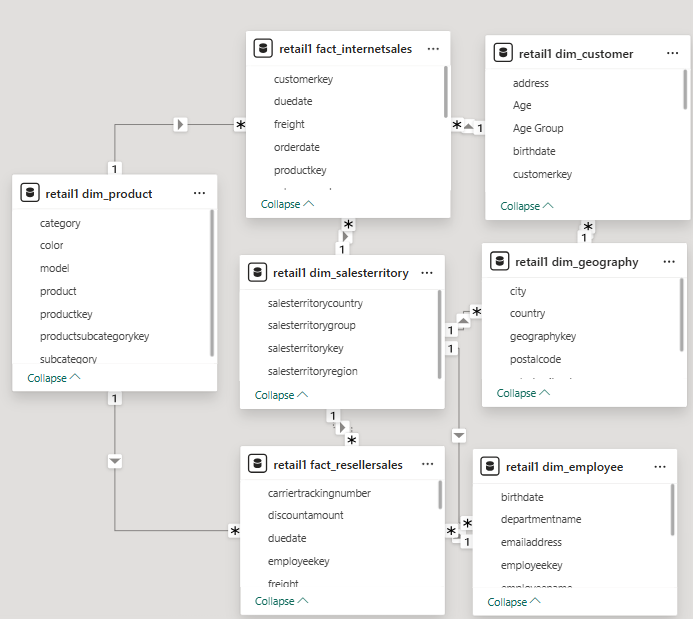
* For TotalProductCost, DiscountAmount, Freight, and SalesAmount columns rounded values to 2 decimal.
* For ProductKey, EmployeeKey, ResellerKey, SalesTerritoryKey, and SalesOrderLineNumber columns converted data type to int.
* Converted the DataFrame into a list of tuples for database insertion.
* Executed an SQL INSERT query to insert the data into the fact\_ resellersales table.
* Checked the fact\_ resellersales table in the MySQL Workbench all the records have been inserted.



 Findings: All the 60855 records are imported.

**Power BI Reporting and Analytics**

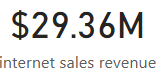
* The 8 Tables in MYSQL Workbench are loaded into power query for transformation.
* The data types are changed for ‘key’ columns and item number from number to text as there is no need of performing mathematical calculations.
* After transformation data is loaded into report.
* Goto Model view and check the relationships.



* I added title ‘**Retail Sales & Customer Insights Dashboard**’ using text box.
* Created measure ‘**total internet sales’**. Displayed using card visual under retail1 fact\_ internetsales table.

 Findings: total number of internet sales are 60 thousand.

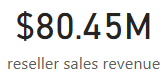
* Created measure ‘**internet sales revenue’**. Displayed using card visual under retail1 fact\_ internetsales table.

 Findings: total revenue generated from internet sales is 29.36 million.

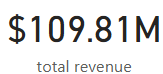
* Created measure ‘**total reseller sales’**. Displayed using card visual under retail1 fact\_ resellersales table.

 Findings: total number of reseller sales are 61 thousand.

* Created measure ‘**reseller sales’**. Displayed using card visual under retail1 fact\_ resellersales table.

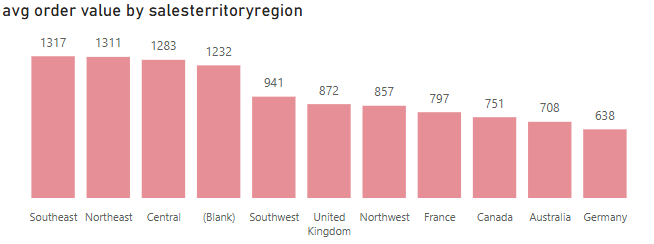
 Findings: total revenue generated from reseller sales is 80.45 million.

* Created measure ‘**Total Revenue’**. Displayed using card visual under retail1 fact\_ resellersales table.

 Findings: total revenue generated from sales is 108.81 million.

**Worst performing sales territories**

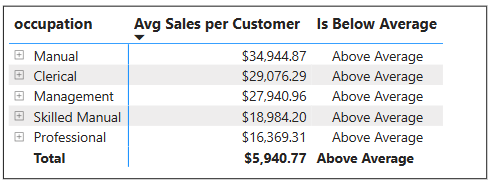
* Create a visual showing Total Sales and Avg Order Value by Sales Territory to help identify the where the company needs to improve penetration
* Created measure ‘**avg order value**’ under retail1 fact\_ resellersales table. Using total revenue/ (total internet sales + total reseller sales).
* plotted avg order value and salesterritoryregion using clustered column chart.

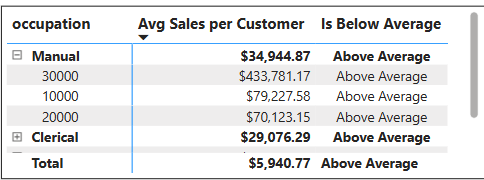


* Findings: South east region has highest avg order value and Germany has the lowest.

**Customer Segmentation based on Occupation and Annual Income**

* Show Avg Sales per Customer by Occupation and Annual Income of Customers so as to identify customer groups who are below average in Sales per Customer.
* Created measure ‘**total customers**’ under retail1 dim\_customer.
* Created measure ‘**Avg Sales per Customer’** under retail1 fact\_ resellersales table. Using total revenue/ total customers.
* Findings: avg sales per customer is 5940.77.
* Created measure ‘**Is Below Average**’ under retail1 dim\_customer table. Using avg sales per customer < 5940.77 "Below Average", "Above Average".
* Displayed using Matrix visual Occupation and Yearly Income in rows, Avg Sales per Customer and Is Below Average in values.

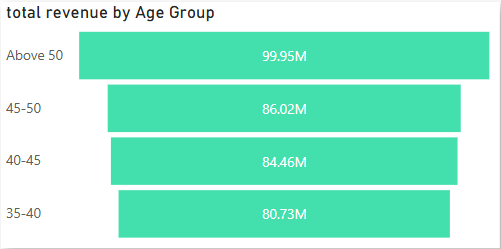




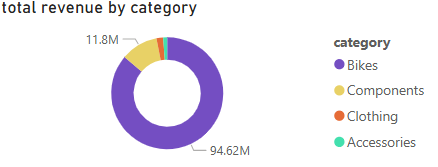
* Findings: there no occupation with avg sales per customer below average.

**Customer age-group to target for increasing sales**

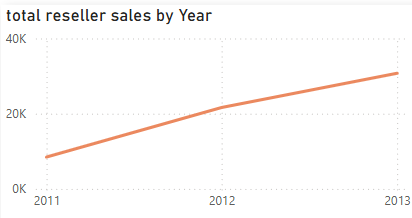
* Create a visual to help Brian identify an age-group (25-30, 30-35, and so on) under the age of 50 to target. Add the Age column in the data model, not in Power Query.
* Created calculated column ‘**Age**’ under retail1 dim\_customer table. By date difference of birth date and current date.
* Created calculated column ‘**Age Group**’ under retail1 dim\_customer table. Grouped customers into various ages.
* Total revenue by Age Group displayed using water fall chart visual.

 Findings: highest revenue generated is from Above 50 age group.

* **total revenue by category –** plotted total revenue and category using donut chart**.**

Findings: highest revenue is generated by Bikes category.

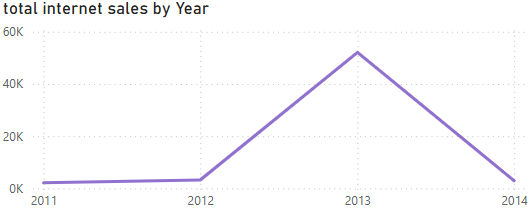
* **total reseller sales by Year –** plotted total reseller sales and Year using line chart**.**

Findings: highest total reseller sales are in 2013. total reseller sales has been increasing sales every year constantly.

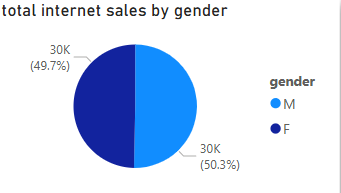
* Plotted sales territory region using map visual**.**

Findings: total there are 10 regions.

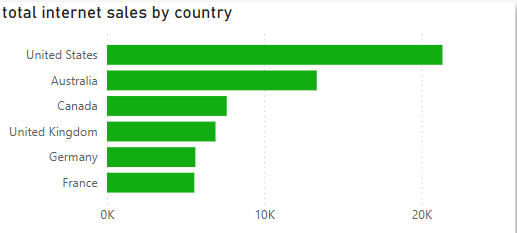
* **total internet sales by Year –** plotted total internet sales and Year using line chart**.**

Findings: highest total internet sales are in 2013. total reseller sales has been constantly changing every year.

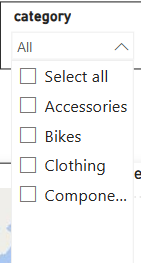
* **total internet sales by gender -** plotted total internet sales and gender using pie chart

Findings: customers of both the genders are equally contributing to internet sales.

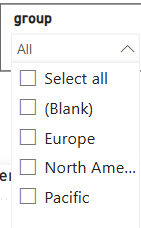
* **total internet sales by country -** plotted total internet sales and country using clustered bar chart**.**

Findings: highest total internet sales are in United States country.

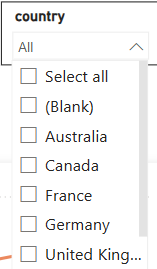
* Created slicer for category to filter all the visuals.



* Created slicer for group to filter all the visuals.

****

* Created slicer for country to filter all the visuals.



**Complete Report**

