01. Sales Analysis with Power BI

🕐 Created	@July 21, 2024 10:17 AM
☰ Tags	Dashboard Power BI SQL Visualization
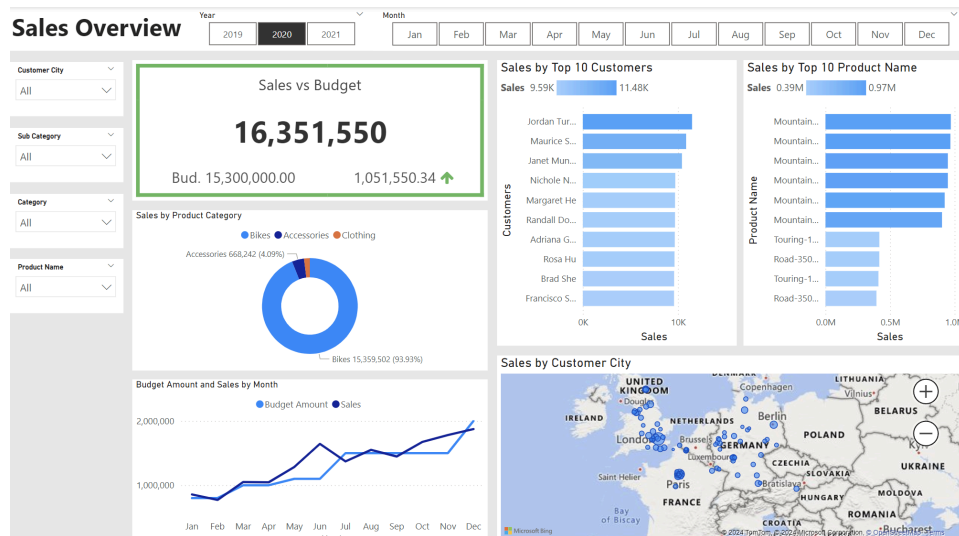
Dashboard Public Link: https://app.powerbi.com/groups/me/reports/99693684-59e2-4ffd-a011-bb44f0e6fec7?ctid=485d0c2a-b3bc-407c-98fb-825408258656&pbi_source=linkShare

Overview

Background

In this scenario, I am a Data Analyst at a certain company. One day, my Sales Manager, Steven, emailed me a request:

▼ Read Steve's Email Here

Hi Rania!
I hope you are doing well. We need to
improve our internet sales reports and want to move from static reports to visual dashboards.

Essentially, we want to focus it on

how much we have sold of what products, to which clients and how it has been over time.

Seeing as
each sales person works on different products and customers it would be beneficial to be able to filter them also.

We
measure our numbers against budget so I added that in a spreadsheet so we can compare our values against performance.

The budget is for 2021 and we usually look
2 years back in time when we do analysis of sales.
Let me know if you need anything else!

Regards,

Steven

Business Demand Overview

From Steven's email, we can infer the objective of our analysis through identifying the business demand points:

1. Reporter: Sales Manager, Steven

2. Value of Change: Visual dashboards and improved sales reporting

3. Necessary Systems: Power BI, Customer Relationship Management (CRM)

4. Other Relevant Information: Budgets have been delivered in Excel for 2021

Now we focus on how this request can benefit our clients with identifying **User Stories:**

User Stories

No	Role	Request	User Value	Acceptance Criteria
1	Sales Manager	A dashboard overview of	Follow up which customers and	Operational Power BI dashboard with

No	Role	Request	User Value	Acceptance Criteria
		internet sales	products sales the best and sales overtime against budget	graphs and KPIs to visualize growth of sales and its relationship with budget
2	Sales Representative	A detailed overview of internet sales per customers and per products	Follow up customers that buys or sells most and follow up on products that sells most	Operational Power BI dashboard that can filter data for each customer and product

In **conclusion**, we need an interactive dashboard that can:

1. Visualize relationship between sales against budget on 2019—2021

2. Analyze customers and products based on sales

3. Provide a tool that can visualize sales growth

Data Understanding

Information on sales, customers, and products are provided by AdventureWorks sample databases - SQL Server | Microsoft Learn that was created by Microsoft for learning purposes. For this project, we use AdventureWorksDW2019.bak database from the provided link as well as the Excel given by our Sales Manager that provides information on the 2021 budget.

Filter Database

The AdventureWorksDW2019 database consists of a version table, a log table, **16 dimension tables**, and **12 fact tables**. Let's analyze which dataset its column that will be the best fit for our analysis:

No	Request	Dataset	Necessary Columns
1	"Improve **Internet Sales**"	dbo.FactInternetSales	- SalesAmount: provides information on amount of customer's sales -

No	Request	Dataset	Necessary Columns
			CustomerKey: customer identifier - OrderDateKey: date of order identifier
2	"…how much we have sold of **what products**…"	dbo.DimProduct	- ProductKey: product identifier - EnglishProductName: name of product
		dbo.DimProductCategory	- ProductCategoryKey: category identifier - EnglishProductCategoryName: category name
		dbo.DimProductSubcategory	- ProductSubcategoryKey: subcategory identifier - EnglishProductSubcategoryName: subcategory name
3	"…how much we have sold…, to **which clients**…"	dbo.DimCustomer	- CustomerKey: customer identifier - FirstName: customer's first name - LastName: customer's last name - Gender: customer's gender - DateFirstPurchase: date of customer's first purchase - GeographyKey: customer's geography key
		dbo.DimGeography	- GeographyKey: customer's geography key - City: city in which customer purchased in

No	Request	Dataset	Necessary Columns
4	"…how much we have sold…**over time**"	dbo.Calendar	- DateKey: date identifier - EnglishMonthName: month name - MonthNumberOfYear: month number - CalendarYear: year

> 💡 Notice that there are identical columns among the datasets. This will give us a clue on the relationships in our database.

The **FACT_Budget** dataset, which was given by the Sales Manager, consists of 18 rows and 2 columns. The first columns represents the date in which the budget was given and the second column is the amount of budget. On the date column, each data is the 1st day of the month, with the months ranging from January 2020 to June 2021. Because there are no missing values and the datatype is already suitable, this dataset is already clean.

Data Cleaning, Data Transformation, and SQL Queries

As mentioned before, we don't need every column on each dataset to fulfill the request. There are also necessary columns that has a significant amount of missing values. Hence, cleaning dataset off of missing values and unnecessary columns are needed.

Few of the mentioned datasets at the previous sections only consists of two necessary columns. To be more efficient, we need to do data transformation (e.g. join) on these dataset to their related dataset. This way, we reduce the amount of tables needed for analysis.

1. **DIM_Customer**

   ▼ View SQL Code Here

   ```
   --Cleansed Customer Table--
   SELECT
   ```

```
      c.customerkey AS CustomerKey,
      c.firstname AS [First Name],
      c.lastname AS [Last Name],

      --concat first name and last name to form full name in new column
      c.firstname + ' ' + lastname AS [Full Name],

      --change gender written category type
      CASE c.gender WHEN 'M' THEN 'Male' WHEN 'F' THEN 'Female' END /

      c.datefirstpurchase AS DateFirstPurchase,

      -- Joined in Customer City from Geography Table
      g.city AS [Customer City]

    FROM
      dbo.dimcustomer AS c
      -- left join ~ keep whole left table (c) and join specified columns from ri
      LEFT JOIN dbo.DimGeography AS g ON g.GeographyKey = c.Geograph

      -- Ordered List by Customerkey in Ascending Order
      ORDER BY
         CustomerKey ASC
```

▼ View CSV Result Here

DIM_Customer.csv

▼ Key points on process

- Combined information from two columns (FirstName and LastName) to one column (as FullName). This procedure uses a concatenation method.

- Used SQL Case expression to edit the gender categories on the Gender column

- Used SQL Left Join clause to merge dbo.DimGeography on CustomerKey with dbo.Customer to get customer's city

2. **DIM_Calendar**

▼ View SQL Code Here

```sql
-- Cleansed DIM_DateTable --
SELECT
  [DateKey],
  [FullDateAlternateKey] AS Date,
  [EnglishDayNameOfWeek] AS Day,
  [WeekNumberOfYear] AS WeekNr,
  [EnglishMonthName] AS Month,

    -- take first three letters of EnglishMonthName and input it to a new col
  LEFT([EnglishMonthName], 3) AS MonthShort,

  [MonthNumberOfYear] AS MonthNo,
  [CalendarQuarter]AS Quarter,
  [CalendarYear] AS Year

FROM
  [AdventureWorksDW2019].[dbo].[DimDate]

--Extract only data from 2019-2021:
WHERE
  CalendarYear >= 2019 AND CalendarYear <= 2021
```

▼ View CSV Result Here

DIM_Calendar.csv

▼ Key points on the process

- Took first 3 letters of each data in EnglishMonthName column to create a new column that consists of first three letters of each month.

- Filtered dataset to only include data from 2019—2021.

3. **DIM_Products**

▼ <u>View SQL Code Here</u>

```sql
SELECT
  p.[ProductKey],
  p.[ProductAlternateKey] AS ProductItemCode,
  p.[EnglishProductName] AS [Product Name],

    -- joined from Sub Category Table
  ps.EnglishProductSubcategoryName AS [Sub Category],

    --joined in from Category Table
  pc.EnglishProductCategoryName AS [Product Category],

  p.[Color] AS [Product Color],
  p.[Size] AS [Product Size],
  p.[ProductLine] AS [Product Line],
  p.[ModelName] AS [Product Model Name],
  p.[EnglishDescription] AS [Product Description],

    --replace missing value with "Outdated" in the Status column
  ISNULL (p.Status, 'Outdated') AS [Product Status]

FROM
  dbo.DimProduct as p
    --left join for DimProductSubcategory and DimProduct Category Tables
  LEFT JOIN dbo.DimProductSubcategory AS ps ON ps.ProductSubcateg
  LEFT JOIN dbo.DimProductCategory AS pc ON ps.ProductSubcategoryK
--ascending order on product key
```

```
ORDER BY
  p.ProductKey asc
```

▼ View CSV Result Here

DIM_Products.csv

▼ Key points on the process

- Replaced missing values on column with a non-null value

- Used SQL Left Join clause to merge dbo.DimProductCategory and dbo.DimProductSubcategory with dbo.DimProduct

4. **FACT_InternetSales**

▼ View SQL Code Here

```
SELECT
  [ProductKey],
  [OrderDateKey],
  [DueDateKey],
  [ShipDateKey],
  [CustomerKey],
  [SalesOrderNumber],
  [SalesAmount]
FROM
  [AdventureWorksDW2019].[dbo].[FactInternetSales]
WHERE
  --only includes data from 2019-2021
  --4 is because the OrderDateKey data represents year on the first 4 cha
  LEFT
    (OrderDateKey, 4) >= 2019
--ascending order on OrderDateKey
```

```
ORDER BY
  OrderDateKey ASC
```

▼ View CSV Result Here

FACT_InternetSales.csv

▼ Key points on the process

- Extracted data from 2019-2021 based on OrderDateKey

All the necessary information from AdventureWorksDW2019 database has been cleaned and transformed to four datasets.

Data Modelling

Now, I need to identify all the possible relationships between dataset. Here is what I discovered:

- DIM_Products can be connected to FACT_InternetSales because the ProductKey column, that is present on both datasets, represents a product identifier.

- DIM_Calendar can be connected to FACT_InternetSales because the DateKey and OrderDateKey both represents a certain date and time in the same format.

- DIM_Calendar can be connected to FACT_Budget because the Date column is present on both datasets and each represent a date and time in the same format.

- DIM_Customers can be connected to FACT_InternetSales because the CustomerKey column, that is present on both datasets, represents a customer identifier.

- Key Measurements is a new table built out of the other tables that was created to store the measures created out of the other columns (this will be explained in later analysis)

Data Modelling Diagram:
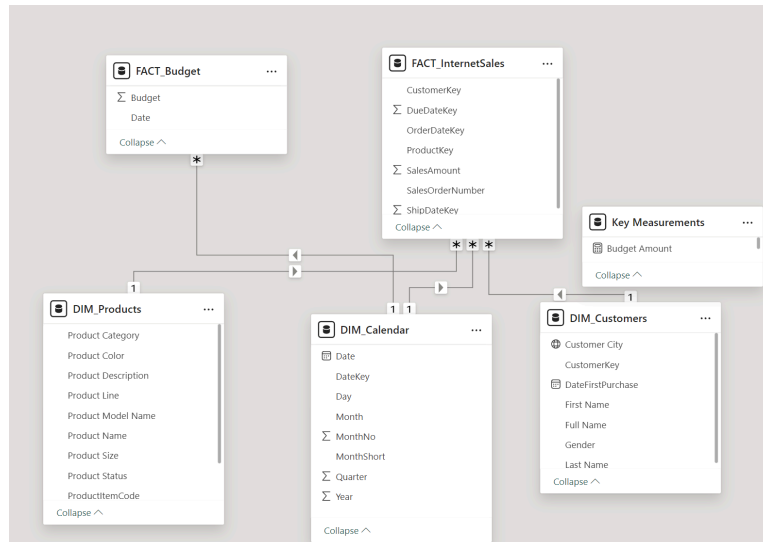
Power BI Reports and Dashboard

Here is my decision making process while creating the dashboard

1. **Slicers**

   a. Customer City: to show customer behavior based on customer's city

   b. Sub Category: to show product behavior based on its subcategory

   c. Category: to show product behavior based on its category

   d. Year: to show data based on year

   e. Month: to show data based on month

2. **KPI**

   a. Key Performance Indicator (KPI) is a tool that is used to showcase sales as the indicator, budget as the first comparison, and sales minus budget as the second comparison.

   b. When the sales minus budget shows a negative number, this shows underperformance. The KPI display's outline will turn red when this happen in the dashboard.

3. **Donut Chart**

   a. To visualize the sales components of each product category, donut chart is used. This way, client can easily see what product category sales the most overtime.

4. **Line Chart**

    a. To see the growth of sales throughout the year, we need a line chart that displays each month as the x-axis.

    b. To compare growth sales with budget throughout the year, we also need a line chart that represents budget growth throughout the year.

5. **Bar Chart**

    a. Sales by Top 10 Customers: shows the highest sales based on customer.

    b. Sales by Top 10 Product Names: shows the highest sales based on product.

6. **Map**

    a. Map is provided to show cities in which customers made their orders in.

7. **Customer Table**

    a. Customer table is provided to help sales representative see which customers made the highest sales throughout the year.

    b. The table also shows a deeper color on higher sales numbers.

Insight and Impact

From the dashboard, we can conclude that:

- Jordan Turner made the highest sales during the observed time period with a sales value of $11,484

- Bikes are the most sold product throughout 2019 to 2020, while Accessories are the most sold product in 2021

- KPI and line chart shows that the sales surpass the budget in most months, but there are still a pattern of inconsistency. There are 4 months in 2021 with sales shortfall, which are February, July, September, and December. This could happen because of seasonal factors. Since products sold are mostly bikes, this shows that products relies on the outdoor conditions of customer's area. Therefore, an alternative marketing scheme needs to be executed during these months that will benefit customers even in seasonal changes.

- Customer cities ranges throughout the whole world with most sales in Europe, South America, and Australia.

Challenges and Solutions

Challenges that I faced during the making of this dashboard and its solutions for future improvements:

- The request is made for 2019—2021. If the Sales Department wants a wider range, there needs to be adjustments on the SQL code. If the Sales Department wants to look at growth for every 2 years back, SQL code needs to be adjusted to include data in today's year minus 2. The Sales Department also needs to send a database on budget that updates overtime. Here is the code adjustment in FACT_InternetSales:

```
SELECT
  [ProductKey],
  [OrderDateKey],
  [DueDateKey],
  [ShipDateKey],
  [CustomerKey],
  [SalesOrderNumber],
  [SalesAmount]
FROM
  [AdventureWorksDW2019].[dbo].[FactInternetSales]
WHERE
  --only includes data from 2019-2021
  --4 is because the OrderDateKey data represents year on the first 4 characters
  LEFT
    (OrderDateKey, 4) >= YEAR(GETDATE())-2
--ascending order on OrderDateKey
ORDER BY
  OrderDateKey ASC
```

- The budget provided is only for January 2020 to June 2021, so KPI analysis can only be done fully on the year 2020. The Sales Department could provide a more complete budget data for better analysis.

Summary

The Internet Sales has a relatively positive trend. This shows that the Sales Department are already doing relatively efficient sales. However, there are still some inconsistency in the Sales vs Budget trend on four distinct months caused by a possible seasonal factor. The Sales Department could look more into the sales during those months.