



Project Objective

Hi, I am Asif Khan. I am a data analyst. In this project I will work on sales related data of a UK based business that sells gift items in both local and foreign markets through their online store. Many of their customers are wholesalers who re-sell these items to the local retailers. The objective of this project is to get insight about:

- The top-selling products in local (UK) market
- The largest foreign markets for their products
- The top-selling products in the foreign markets
- Sales pattern across the largest foreign markets

This is a snapshot of the data in CSV format (First 5 rows):

	A	B	C	D	E	F	G	H
1	InvoiceNo	StockCode	Description	Quantity	InvoiceDate	UnitPrice	CustomerID	Country
2	536365	85123A	WHITE HANGING HEART T-LIGHT HOLDER	6	12/1/2010 8:26	2.55	17850	United Kingdom
3	536365	71053	WHITE METAL LANTERN	6	12/1/2010 8:26	3.39	17850	United Kingdom
4	536365	84406B	CREAM CUPID HEARTS COAT HANGER	8	12/1/2010 8:26	2.75	17850	United Kingdom
5	536365	84029G	KNITTED UNION FLAG HOT WATER BOTTLE	6	12/1/2010 8:26	3.39	17850	United Kingdom
6	536365	84029E	RED WOOLLY HOTTIE WHITE HEART.	6	12/1/2010 8:26	3.39	17850	United Kingdom

Part 1: Data Cleaning in SQL

The following steps were taken to make sure the data was more usable to provide us with accurate insights and info graphics:

1. Deleting duplicate transactions.

Removing duplicate transaction is an essential part of data cleaning. Transactions that look exactly alike and happened at the same time and same day are most likely duplicate, and I had to remove them from the data.

```
WITH cte AS (  
    SELECT *,  
           ROW_NUMBER() over (  
             partition by  
               InvoiceNo,  
               Description,  
               Quantity,  
               InvoiceDate,  
               UnitPrice  
             order by  
               InvoiceNo) row_num  
    from  
        dbo.Online_Retail  
    )  
DELETE from cte  
where row_num > 1;
```

%
Messages
(5349 rows affected)

2. Checking for any missing values.

Any row where the unit price is missing is useless for further analysis. So I have tracked and deleted them as part of the data cleaning process.

```
Select *  
From dbo.Online_Retail  
Where (UnitPrice is null or UnitPrice = '0')
```

100 %

	Description	Quantity	UnitPrice	Country
1		8	0	United Kingdom
2	label mix up	-59	0	United Kingdom
3	samples/damages	-20	0	United Kingdom
4		-16	0	United Kingdom
5		1	0	United Kingdom
6		5	0	United Kingdom
7		-22	0	United Kingdom

```

Delete
From dbo.Online_Retail
Where (UnitPrice is null or UnitPrice = '0')

```

Furthermore, since some of the figures in 'Quantity' column had negative values, I had to convert them into absolute (positive) figures.

```

update dbo.Online_Retail
set Quantity = ABS (Quantity)
where Quantity < 0

```

3. Splitting Date and Time of Sale into Different Columns.

The 'InvoiceDate' column includes the date and time of all the transactions. This is not very usable as it will be hard to put such a format into further analysis. Therefore, I have decided to slit them into two separate columns.

```

alter table dbo.Online_Retail
add DateOfSale date;

update dbo.Online_Retail
set DateOfSale = SUBSTRING(InvoiceDate, 1, CHARINDEX(' ', InvoiceDate))

alter table dbo.Online_Retail
add TimeOfSale time (7);

update dbo.Online_Retail
set TimeOfSale = SUBSTRING(InvoiceDate, 8, CHARINDEX(' ', InvoiceDate))

```

4. Removing Unused Columns.

```

Alter table dbo.Online_Retail
drop column InvoiceNo, StockCode, InvoiceDate, CustomerID

```

5. Adding Additional Column.

Since the figures for revenue would be extremely important for our analysis, it was important to create a column to store that data.

```

alter table dbo.Online_Retail
add SalesRevenue int;

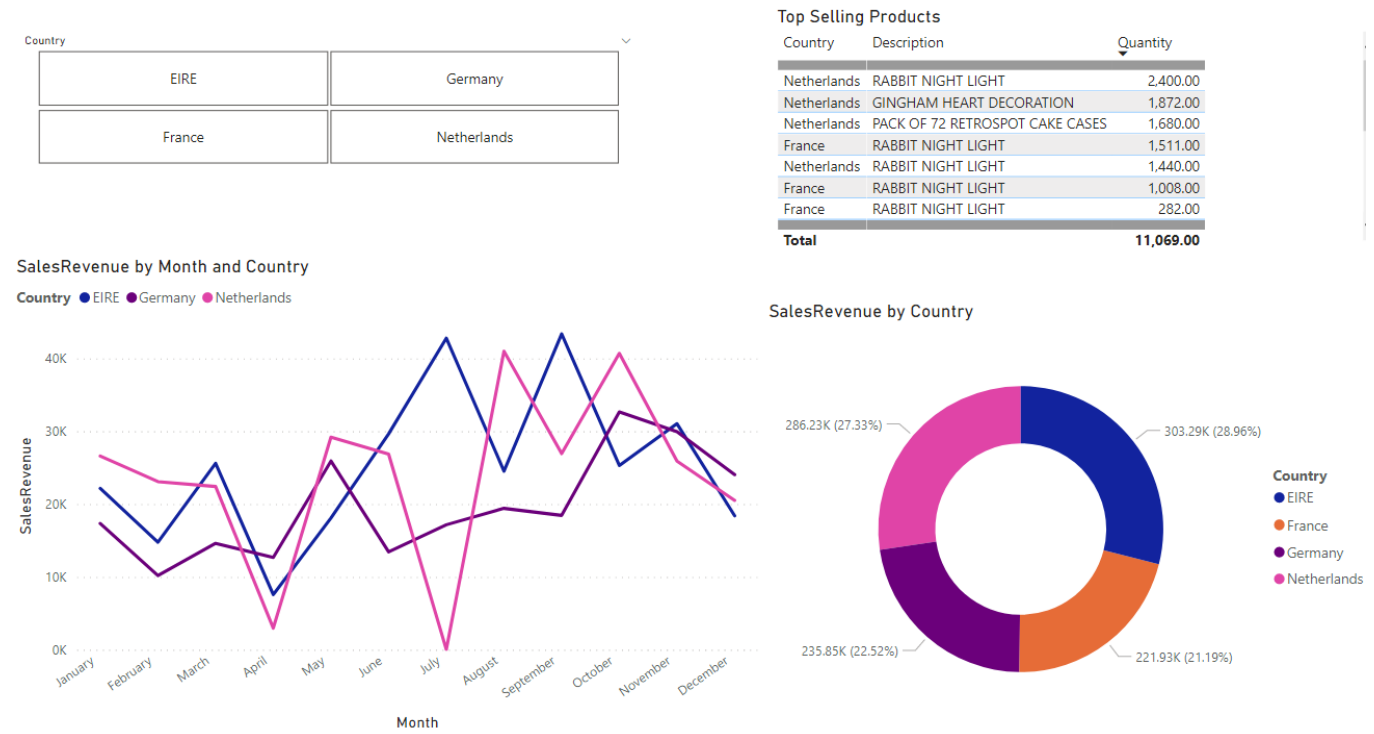
update dbo.Online_Retail
set SalesRevenue = Quantity * UnitPrice

```

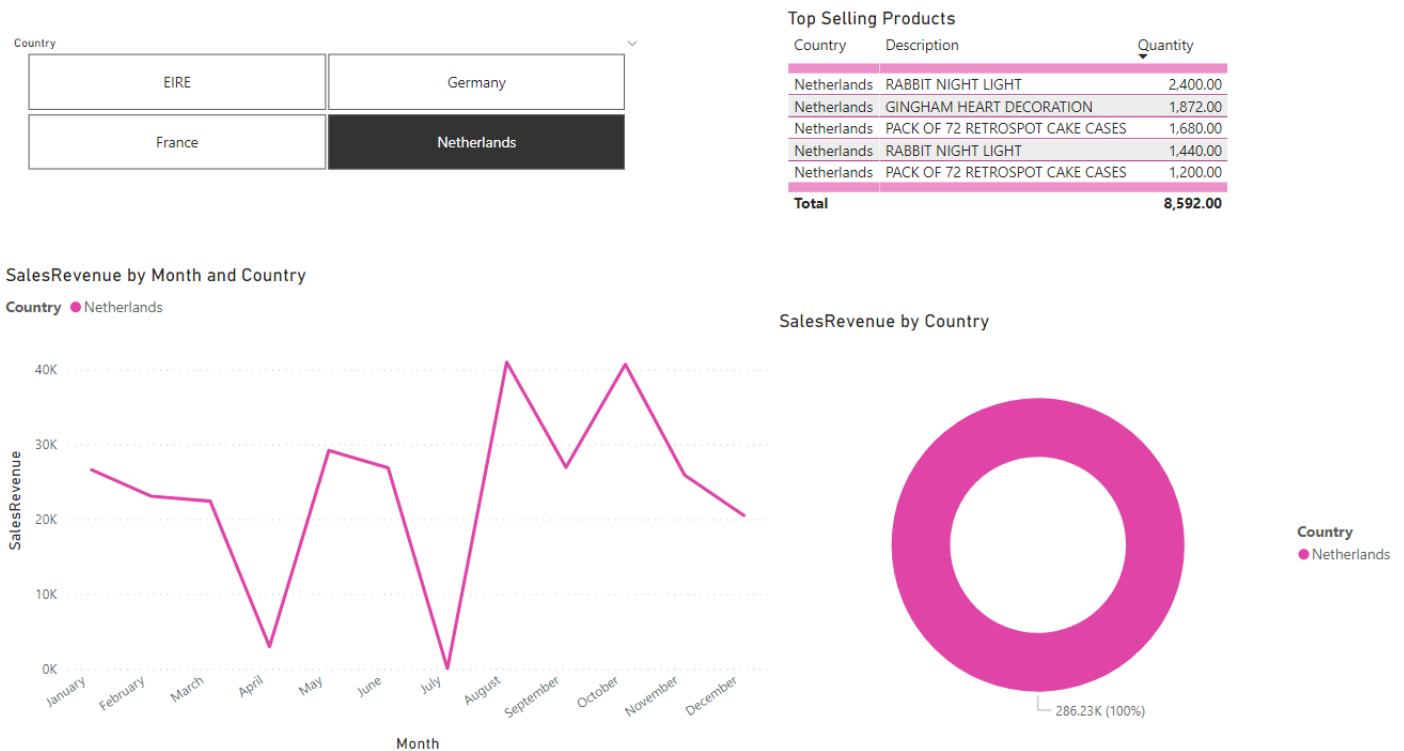
Part 2: Data Visualization in Microsoft Power BI

1. Creating the Dashboard for Foreign Markets

After importing the SQL database in Microsoft power BI, I have created a simple dashboard and added a slicer that will enable me to analyze the biggest foreign markets for the company. For example, in the 1st snapshot on the next page, clicking on Netherlands on the slicer displays the top selling products, overall sales pattern and volume inside the Netherlands market:



Clicking on any country on the top left slicer will show the country specific information in my analysis.



Now, let us have a closer look at the different parts of the dashboard:

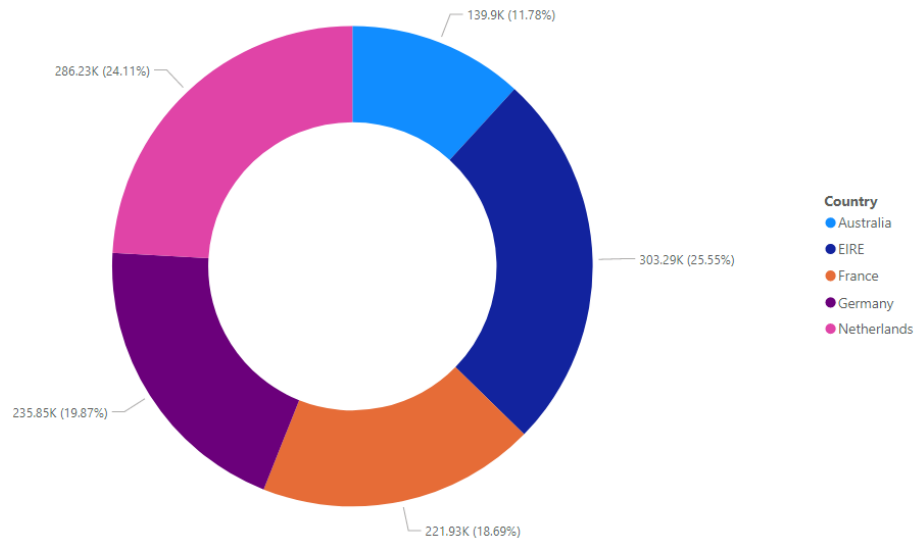
2. Top 5 Selling Products in the Local (UK) Market:

Country	Description	Quantity
United Kingdom	"PAPER CRAFT , LITTLE BIRDIE"	161,990.00
United Kingdom	MEDIUM CERAMIC TOP STORAGE JAR	148,430.00
United Kingdom	POPCORN HOLDER	12,327.00
United Kingdom	ROTATING SILVER ANGELS T-LIGHT HLDR	11,668.00
United Kingdom	WORLD WAR 2 GLIDERS ASSTD DESIGNS	9,663.00
Total		344,078.00

3. Largest 5 Foreign Markets

As shown in the doughnut chart below, Ireland (EIRE), Netherland and Germany are the company's largest foreign markets. This is based on the percentage of total sales revenue the company earned in all foreign markets.

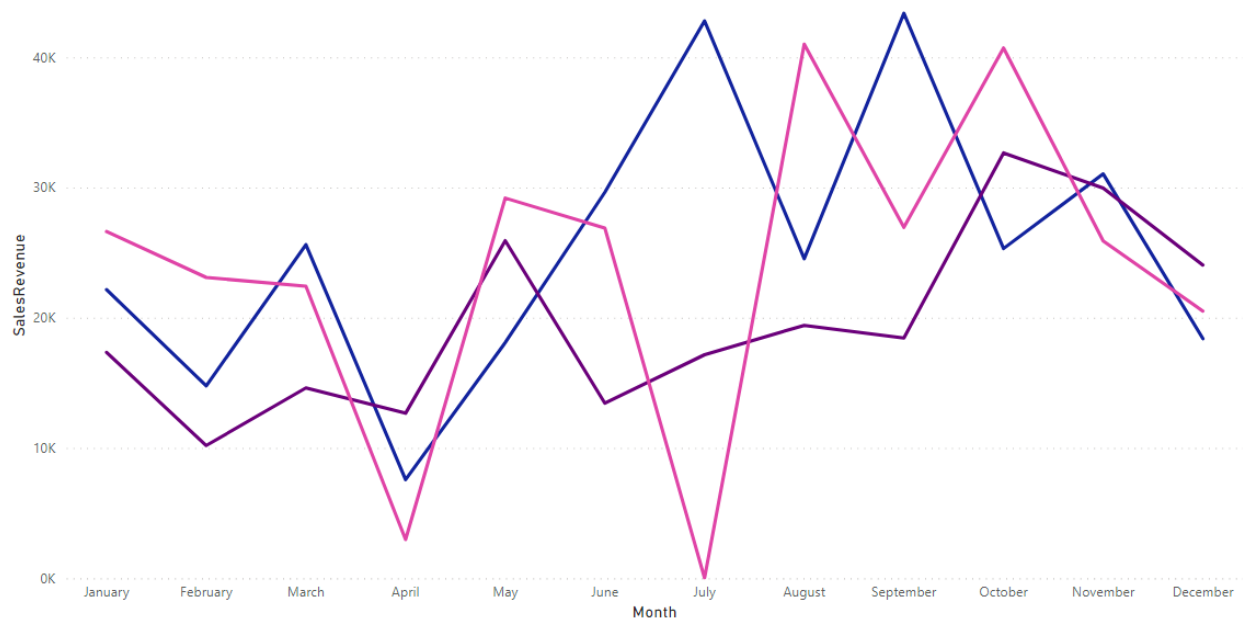
SalesRevenue by Country



4. Sales Pattern in the Top 3 Foreign Markets.

SalesRevenue by Month and Country

Country ● EIRE ● Germany ● Netherlands



The trends above show some interesting insights into the sales pattern of the company's largest three markets outside UK and demands further probe. For instance, why do the sales in Ireland boom in the month of July whereas there are almost no sales in Netherlands?