

Final Assignment Report

Dataset Description:

The dataset is named “Online Retail II” and is publicly available at [UC Irvine Machine Learning Repository](#). The dataset consists of transactions that occurred for a UK-based and registered online retail store between the dates 01/12/2009 and 09/12/2011. This online store sells unique giftware. They sell goods to wholesalers, so most of the transactions have a quantity of more than 1. The dataset contains 1,067,371 instances. The below table describes each column.

Column Name	Data Type	Description
Invoice	Integer	A 6-digit identifier for every order made by the customer
StockCode	String	A 5-digit product code unique for every product
Quantity	Integer	Number of items purchased for each transaction
InvoiceDate	Datetime	Time at which the order was placed
Price	Float	Price of a single product in sterling
Description	String	Details of the product
Customer ID	Integer	A 5-digit identifier of the customer
Country	String	Country of purchase

Additional columns have been **derived** to better work with the data for the Analysis.

Column Name	Data Type	Description
Month	Integer	The month of the purchase
Year	Integer	The year of the purchase
WeekDay	String	Describes the day of the week
Month_year	Date	Provides both month and date together
Hour	Integer	The hour of the purchase
Revenue	Float	Revenue made on the purchase

Dashboard Title:

Sales and Revenue Analysis of an Online Gift Store given the data between the years 2009-2011.

The dashboard aims to analyze the sales and revenue of the online store. Some of the targeted **objectives** of the dashboard are:

1. A visual of sales expansion into different countries in the world, which originated in the United Kingdom.

This is presented through a **Geographical graph** in the dashboard colored based on the revenue contributed by the country. It can be drilled over three different years and visualizes the expansion to different countries in the ascending order of year selection.

2. Visualize an hour-wise invoice graph to understand the customer's buying patterns in a day.

This is presented by a **tree map** which highlights in descending order of size the highest invoice receiving hour to lowest invoice receiving hour. This can be further drilled into by choosing a specific day of the week.

3. Identifying the pattern in monthly revenue changes over the year.

This is presented by a **line chart** plotting month vs. revenue in sterling for the years 2010 and 2011. 2009 was not considered for this because it has only one month's data.

4. To identify Loyal Customers.

This is presented through a horizontal **stacked bar chart**. Here a consideration has been made that the customer's loyalty is greater if they continue the purchases for the upcoming years. So, a calculation showed that 683 Customer IDs had been purchased over three years, making these Customer IDs the most loyal customers.

5. Visually identify the busiest day in the week in terms of invoices made.

A **bar plot** presents this with weekday vs. count of invoices. This plot is re-arranged to visually identify the most busy day and the least.

Steps towards Development:

1. Data collection: The above-mentioned data in the description has been downloaded.
2. Data Exploration and Transformation: I have performed a few transformations to the data to have the above-mentioned derived columns in the datasets in Jupiter notebook.
3. Once the data is ready, it is exported to **Tableau** a separate selection of graphs is built into each sheet. While building these graphs, the following standard practices are followed.
 - a. Font type: Tahoma
 - b. Font Size: Sheet title: 10, Sheet sub-titles: 8, Axis text font: 6
 - c. Sheets are named with a relevant name to understand individually
4. Once the sheets are built, a dashboard is created, and all the charts in each sheet are moved to it in floating mode. Floating mode enables easy movement of all the chart parts when arranging the dashboard.
5. Located the graphs in the background sheets for the best visual access, keeping the aspect ratios of each chart intact and following the key principle specified.
6. Dashboard title was added with the same font and font size of 16.

Key Principles Employed:

Having Interactive charts: All the graphs built are interactive in nature with a tooltip enabled. Apart from that, some important interactions added are:

- Panning feature in the Geographical chart enables a clear vision of all the countries in the world
- At the same-time geographical chart have the filtering enabled on the year
- Hour-wise invoice graph as a weekday selection interaction enabled as well

Consistent Color-palette: Stuck to using a color palette that is compatible with the color-blind audience. The palette ranges from warm oranges to cool blues having good contrast visually. All the charts have been aligned to use the same colors. Focused on having darker colors to emphasize the point to be made in each chart.

Having multiple levels of Titles or descriptions: Ensured to put a hierarchical variation in the titles to enable the focus and variation of the information being put forth. Also, I used a concise vocabulary to describe each chart.

Visually serial processing and concise layout: To employ good visual experience and understanding of the information showcasing, arranged all the graphs in a planned fashion. Having no unnecessary overlaps, clear fonts and charts, access to the charts, and not giving out too much information to clutter the space.

Gestalt Principle of Continuity, Proximity, and Similarity: Enabled the placement of similar parameter graphs next to each other for easy comparison and relational availability. Put revenue and invoice-related graphs next to each other. Continuity is practiced visually in all charts by arranging the graphs in ascending and descending order. This also helps in creating a focus on something which is either small or large.

Gestalt Principle of Enclosure: When using multiple graphs in a dashboard, they can look extremely cumbersome if clear boundaries are made to focus on those specific graphs. I have used a gray background for the title of each chart and, by arrangement, created an enclosed space for each graph.

Use of Visual Encoding: Darker shade of blue is used consistently for all the graphs where the outcome is to be recognized. So, when someone looks at the dashboard, the dark blue is aligned with an important outcome.