

INT217
INTRODUCTION TO DATA MANAGEMENT
PROJECT REPORT

(Project Semester August-December 2021)

Analysis of US Regional Sales Data

Submitted by
Ayush Kumar Gupta
11903677

B. tech (CSE) RKM010A03
Course Code INT 217

Under the Guidance of
Harneet Kaur : 19371, Assistant Professor
Discipline of CSE/IT
Lovely School of Computer Science and Engineering
Lovely Professional University, Phagwara



L OVELY
P ROFESSIONAL
U NIVERSITY

CERTIFICATE

This is to certify that **Ayush Kumar Gupta** bearing Registration no. **11903677** has completed **INT 217** project titled, **Analysis of US Regional Sales Data** under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of

Lovely Professional University

Phagwara, Punjab.

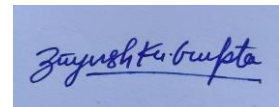
Date: 30/12/2021

DECLARATION

I, **Ayush Kumar Gupta** student of **Introduction to Data management** under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Date: 30/12/2021

Signature:



Registration No, 11903677

Ayush Kumar Gupta

Introduction:

Microsoft Excel is a general-purpose electronic spreadsheet used to organize, calculate, and analyze data, developed by Microsoft. The task you can complete with Excel ranges from preparing a simple family budget, preparing a purchase order, create an elaborate 3-D chart, or managing a complex accounting ledger for a medium size business.

Features in MS Excel:**Home:**

Comprises options like font size, font styles, font colour, background colour, alignment, formatting options and styles, insertion and deletion of cells and editing options

Insert:

Comprises options like table format and style, inserting images and figures, adding graphs, charts and sparklines, header and footer option, equation and symbols

Page Layout:

Themes, orientation and page setup options are available under the page layout option

Formulas:

Since tables with a large amount of data can be created in MS excel, under this feature, you can add formulas to your table and get quicker solutions

Data:

Adding external data (from the web), filtering options and data tools are available under this category

Review:

Proofreading can be done for an excel sheet (like spell check) in the review category and a reader can add comments in this part

View:

Different views in which we want the spreadsheet to be displayed can be edited here. Options to zoom in and out and pane arrangement are available under this category

Benefits of Using MS Excel

MS Excel is widely used for various purposes because the data is easy to save, and information can be added and removed without any discomfort and less hard work.

Given below are a few important benefits of using MS Excel:

Easy to Store Data: Since there is no limit to the amount of information that can be saved in a spreadsheet, MS Excel is widely used to save data or to analyse data. Filtering information in Excel is easy and convenient.

Easy to Recover Data: If the information is written on a piece of paper, finding it may take longer, however, this is not the case with excel spreadsheets. Finding and recovering data is easy.

Application of Mathematical Formulas: Doing calculations has become easier and less time-taking with the formulas option in MS excel

More Secure: These spreadsheets can be password secured in a laptop or personal computer and the probability of losing them is way lesser in comparison to data written in registers or piece of paper.

Data at One Place: Earlier, data was to be kept in different files and registers when the paperwork was done. Now, this has become convenient as more than one worksheet can be added in a single MS Excel file.

Neater and Clearer Visibility of Information: When the data is saved in the form of a table, analysing it becomes easier. Thus, information in a spreadsheet is more readable and understandable.

PivotTables

A PivotTable is an interactive way to quickly summarize large amounts of data. You can use a PivotTable to analyze numerical data in detail, and answer unanticipated questions about your data. A PivotTable is especially designed for:

1. Querying large amounts of data in many user-friendly ways.
2. Sub-totaling and aggregating numeric data, summarizing data by categories and subcategories, and creating custom calculations and formulas.
3. Expanding and collapsing levels of data to focus your results, and drilling down to details from the summary data for areas of interest to you.
4. Moving rows to columns or columns to rows (or "pivoting") to see different summaries of the source data.
5. Filtering, sorting, grouping, and conditionally formatting the most useful and interesting subset of data enabling you to focus on just the information you want.

6. Presenting concise, attractive, and annotated online or printed reports.

PivotCharts

PivotCharts provide graphical representations of the data in their associated PivotTables. PivotCharts are also interactive. When you create a PivotChart, the PivotChart Filter Pane appears. You can use this filter pane to sort and filter the PivotChart's underlying data. Changes that you make to the layout and data in an associated PivotTable are immediately reflected in the layout and data in the PivotChart.

Differences between PivotCharts and standard charts:

If you are familiar with standard charts, you will find that most operations are the same in PivotCharts. However, there are some differences:

Row/Column orientation: Unlike a standard chart, you cannot switch the row/column orientation of a PivotChart by using the **Select Data Source** dialog box. Instead, you can pivot the Row and Column labels of the associated PivotTable to achieve the same effect.

Chart types: You can change a PivotChart to any chart type except an xy (scatter), stock, or bubble chart.

Source data: Standard charts are linked directly to worksheet cells, while PivotCharts are based on their associated PivotTable's data source. Unlike a standard chart, you cannot change the chart data range in a PivotChart's **Select Data Source** dialog box.

Formatting: Most formatting—including chart elements that you add, layout, and style—is preserved when you refresh a PivotChart. However, trendlines, data labels, error bars, and other changes to data sets are not preserved. Standard charts do not lose this formatting once it is applied. Although you cannot directly resize the data labels in a PivotChart, you can increase the text font size to effectively resize the labels.

Introduction:

Fictitious sales data for a certain company across the US regions. I have taken the data from the internet and started analysis on my own. Under the dedicative guidance of our teacher, we learned a lot and now it is the time to show the hard work we have done till now.

Objectives/Scope of Analysis:

- Top 10 Selling Products
- Best Selling Team
- Top 3 Customers/Buyers
- Max sale by Mode
- Year wise Sales Analysis

Source of the Dataset:

I searched over the internet and found a popular sales data of the US based company.

The source link is shown below:

<https://data.world/dataman-udit/us-regional-sales-data>

ETL Process:

As the name suggested ETL stands for Extract, Transform and Load. Just like the name applies ETL tool Extracts data from the source, and Transforms the data while in transit and then it Loads the data in to Specified Data destination

Extract: Here, firstly I downloaded the data from the website and stored it into the folder. Then I went to the data tab in ribbons and selected Get Data option. I used “get from the folder” option and then selected the folder. It extracted the data from the file. It opens in Power Query editor.

Transform: After Extraction I selected transform into Power Query tab. There I merged my some of columns, cleared the unwanted data and got the new refreshed data in a single table.

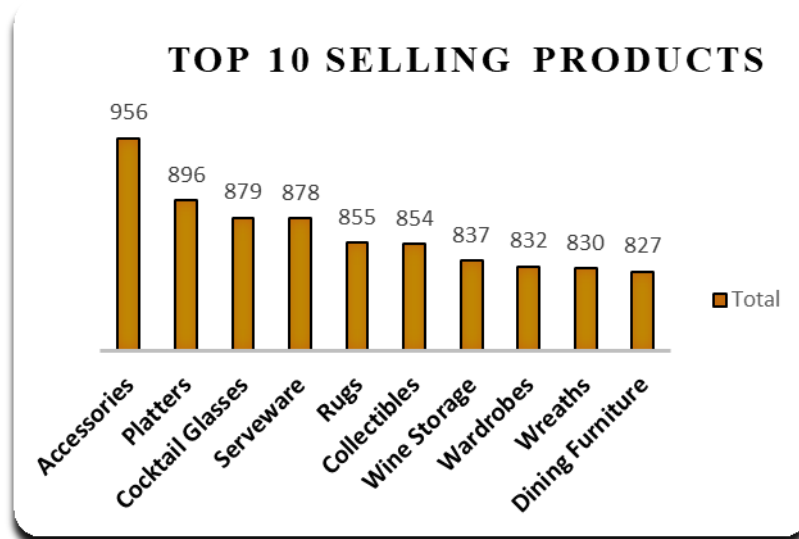
Load: After the transformation phase I loaded the data into my new excel sheet.

Now my excel Sheet is ready to create Pivot tables and Pivot Charts.

Analysis on the Data Set:

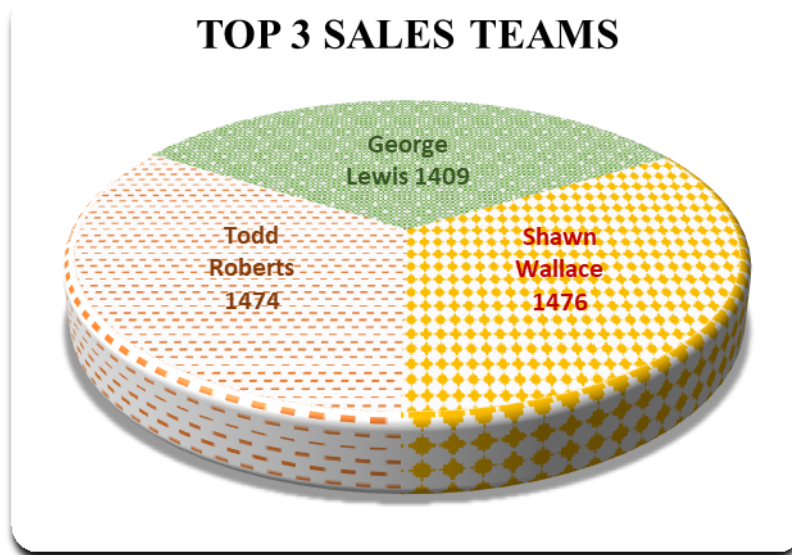
1. **Introduction:** Top 10 Selling Products

2. **General Description:** There were many products available in the data set but I just wanted to showcase the top 10 Selling Products on the basis of Order quantity.
3. **Requirements:** A pivot table with product name as a row and Order Quantity as the values. In the options of value, I opted the Sum of the Order Quantity function, and the I selected the row column in the pivot table and from the value filters selected Top 10. Then selected the Order Quantity and used sorting type Highest to Lowest. Then I selected the option to create Pivot Chart and selected the Clustered Columns chart. In the last pasted it into the dashboard worksheet.
4. **Analysis Results:** I found the different kind of accessories were most selling and Dining Furniture was the lowest selling product.
5. **Visualization:**

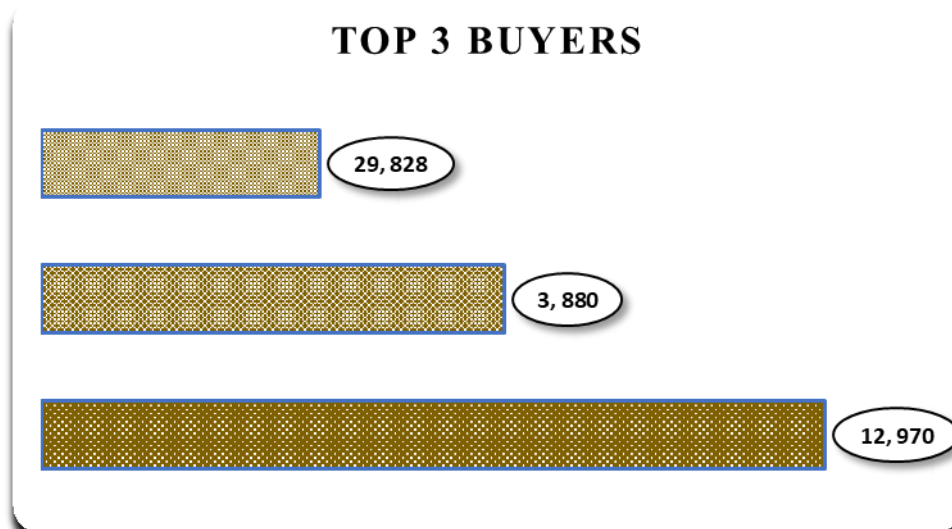


1. **Introduction:** Best Selling Team
2. **General Description:** There were 28 Teams available in the data set but I just wanted to showcase the top 3 Selling Selling on the basis of Order quantity.
3. **Requirements:** A pivot table with Sales team name as a row and Order Quantity as the values. In the options of value, I opted the Sum of the Order Quantity function, and the I selected the row column in the pivot table and from the value filters selected Top 3. Then selected the Order Quantity and used sorting type Highest to Lowest. Then I selected the option to create Pivot Chart and selected the Pie chart. In the last pasted it into the dashboard worksheet.
4. **Analysis Results:** I found that Shawn's team was the best among those top 3 teams

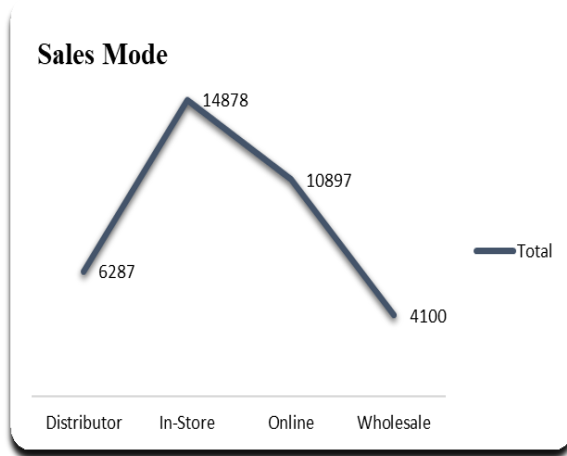
5. **Visualization:**



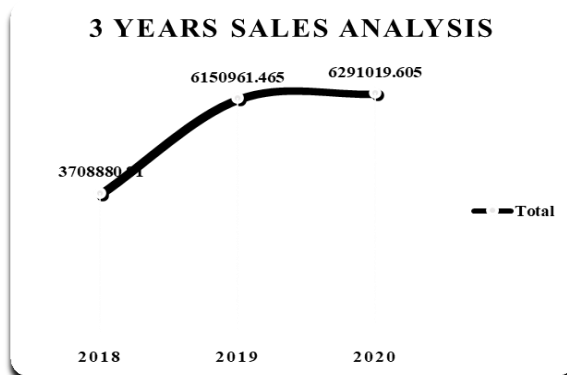
1. **Introduction:** Top 3 Customers/Buyers
2. **General Description:** There were 50 buyers available in the data set but I just wanted to showcase the top 3 Buyers on the basis of Order quantity.
3. **Requirements:** A pivot table with Customer ID as a row and Order Quantity as the values. In the options of value, I opted the Sum of the Order Quantity function, and the I selected the row column in the pivot table and from the value filters selected Top 3. Then selected the Order Quantity and used sorting type Highest to Lowest. Then I selected the option to create Pivot Chart and selected the Bar chart. In the last pasted it into the dashboard worksheet.
4. **Analysis Results:** I found that Shawn's team was the best among those top 3 teams
5. **Visualization:**



1. **Introduction:** Max sales by Modes
2. **General Description:** There were 4 modes of sales available in the data
3. **Requirements:** A pivot table with Sales Channel as a row and Order Quantity as the values. In the options of value, I opted the Sum of the Order Quantity function. Then I selected the option to create Pivot Chart and selected the Line chart. In the last pasted it into the dashboard worksheet.
4. **Analysis Results:** I found the Max sales was done by In-Store division.
5. **Visualization:**



1. **Introduction:** Yearly sales Analysis
2. **General Description:** The data was available for 3 years of sales
3. **Requirements:** A pivot table with Years as a row and Sum of the Sales Prices as the values. In the options of value, I opted the Sum of the Sales Price function. Then I selected the option to create Pivot Chart and selected the Line chart. In the last pasted it into the dashboard worksheet.
4. **Analysis Results:** I Plotted the Sales data graph
5. **Visualization:**



List of Sales analysis:

Top 3 Buyers:

Customer ID	Sum of Order Quantity
12	970
3	880
29	828

3 years sales Analysis:

Years	Sum of Selling Price
2018	3708880.81
2019	6150961.465
2020	6291019.605

Top 10 Selling products:

Products	Sum of Order Quantity
Accessories	956
Platters	896
Cocktail	
Glasses	879
Serve ware	878
Rugs	855
Collectibles	854
Wine Storage	837
Wardrobes	832
Wreaths	830
Dining	
Furniture	827

Top 3 selling Teams:

Sales Team Name	Sum of Order Quantity
Shawn Wallace	1476
Todd Roberts	1474
George Lewis	1409

Top Mode of Sales:

Mode of Sales	Sum of Order Quantity
Distributor	6287
In-Store	14878
Online	10897
Wholesale	4100

References:

Name of the Dataset: US Regional Sales Data

The Owner of the dataset: Udit Kumar Chatterjee

Website: Data.world.com

Link: <https://data.world/dataman-udit/us-regional-sales-data>

Bibliography:

1. Data.world.com
2. https://www.youtube.com/channel/UCfJT_eYDTmDE-ovKaxVE1ig by Kevin Stratvert
3. TutorialsPoint
4. Teachers Lectures videos and Notes