**INTRODUCTION TO DATA MANAGEMENT**

**PROJECT REPORT**

(Project Semester August-December 2022)



***DATA ANALYSIS ON SALES***

Submitted by

Ede.Naveen Siva Sai Srinivas

Registration No. 12003093

Section: KM082

Course Code: INT217

Under the Guidance of

**Baljinder Kaur**

**UID: 27952**

**Discipline of CSE/IT**

**Lovely School of Computer Science and Engineering**

**Lovely Professional University, Phagwara**

1

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Contents** | | **Page no.** |
| 1. | Certificate | 3 |
| 2. | Declaration | 4 |
| 3. | Acknowledgement | 5 |
| 4. | Introduction | 6 |
| 5. | Objectives/Scope of the Analysis | 7 |
| 6. | ETL Process | 8-11 |
| 7. | Analysis on Dataset | 12-17 |
| 8. | Bibliography | 18 |

2

**CERTIFICATE**

This is to certify that Ede.Naveen Siva Sai Srinivas bearing

Registration no. **12003093** has completed **INT217** project titled, **“Data Analysis on Sales”** 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.

**Baljinder Kaur**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab.

Date: 10-11-2022

3

**DECLARATION**

I, Ede.Naveen Siva Sai Srinivas student of Computer Science and Engineering 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: 10-11-2022

Registration No. 12003093

Ede.Naveen Siva Sai Srinivas

4

**Acknowledgement**

I would like to express my special thanks of gratitude to my teacher Miss Baljinder Kaur who gave me the golden opportunity to do this wonderful project on the topic Data Analysis on Sales which also helped me in doing a lot of Research and I came to know about so many new things.

Secondly, I would also like to thank my friends who helped me a lot in finalizing this project within the limited time frame.

5

**Introduction**

Data analysis is a process of inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Data analysis has multiple facets and approaches, encompassing diverse techniques under a variety of names, while being used in different business, science, and social science domains.

People generally get attracted towards shopping so here I have taken a data set on sales of different categories of product. And with this data we will analyze many things like what category of people buy which product.

This thing will help us to analyze which products sales is more and what kind of product that company should produce more in order to again a good amount of profit.

.

**Objectives/Scope of the Analysis**

* Total sales according to region.
* Sales on first class.
* Sales on second class.
* Sales on standard class.
* Sales on same day.
* Progress sales of the year.
* Progress profit of the year.
* Progress quantity of the year.
* Progress discount the year.
* Analysis of sales based on shipping mode

**Source of dataset**

All the dataset used in this is collected from very well-known online website [www.kaggle.com](http://www.kaggle.com/)

**The dataset contains the following features**:

Region,Sales,Profit,Discount,ShippingMode,OrderId,CustomerID,Orderdate,Quantity etc.

## Step 1: Remove the blank cells from the dataset.

## 

For this, select the whole dataset. Go to Find and Select in the Home tab of excel. Select Go to Special from the drop-down menu and then tick the blank option. All the blank cells will be selected. Then go to Delete option in the home tab again and select Delete Rows from the drop-down menu.

This will remove any rows with blank cells.

## Step 2: Remove columns which are not properly defined or not crucial to our analysis.

For this we will columns which are redundant like the column with just the index numbers. For this we will select that particular column and then go to delete option in the home tag and then select Delete Columns from the drop-down menu.

Table

Description automatically generated

## Step 3: Giving proper column names.

The dataset does not have proper columns so our next step would be to giver proper column names to wherever required.

Table

Description automatically generated

## Step 4: Removing the Na values from the Shipping cost columns to zero.

Table

Description automatically generated

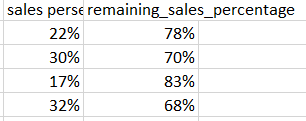
We will do this with the help of find and replace.

7

Chart, bar chart

Description automatically generated

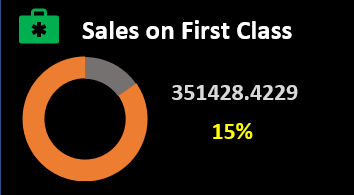
Calculating the total sales using the bar graph based on the regions I have observes the sales are high in the east and west are high compared to the central and south sales and the total sales.



Graphical user interface, text, application, chat or text message

Description automatically generated

The sales happen on the different shipping modes



Graph-based on one of the shipping modes (sales on first class) and calculating the total sales done on the first class and percentage on the sales done on the first class based on the total sales done on the shipping mode.

A picture containing graphical user interface

Description automatically generated

Graph-based on one of the shipping modes (sales on the standard class) and calculate the total sales done on the standard class and the percentage of the sales done on the first class based on the sales done on the shipping mode.

Chart

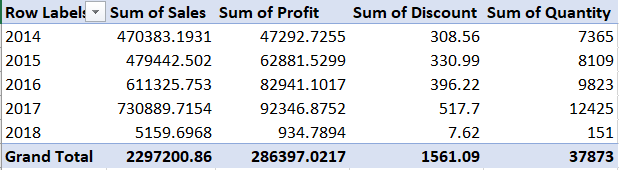
Description automatically generated with low confidence

Graph-based on one of the shipping modes (sales on the second class) and calculate the total sales done on the second class and the percentage of the sales done on the second class based on the sales done on the shipping mode.

A picture containing timeline

Description automatically generated

Graph-based on one of the shipping modes (sales on the same day) and calculate the total sales done on the sameday and the percentage of the sales done on the first class based on the sales done on the shipping mode.



The sales, profit, discount, and quantity happened based on the years .



Table with total sales ,profit ,discount, quantity.

A picture containing line chart

Description automatically generated

Graph based on the sales based on the years (I analyzed the sales is down at the certain time I have analyzed this is based on the decrease in the discount at the same time)

A green line on a black background

Description automatically generated with low confidence

Graph based on the profit based on the years (I analyzed the profit is down at the certain time I have analyzed this is based on the decrease in the discount at the same time and sales is also down at the same time)

Graphical user interface

Description automatically generated with medium confidence

Graph based on the discount based on the years (I analyzed the profit is down at a certain time and sales is also down at the same time because of the decrease of the discount)

A picture containing shape

Description automatically generated

Graph based on the quantity based on the years (I analyzed the profit is down at a certain time and sales is also down at the same time because of the decrease of the discount)

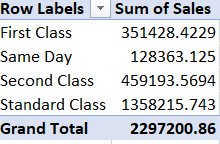


Table of sales based on the different shipping modes

Chart, pie chart

Description automatically generated

The total sales are done on the different shipping modes(the sales are done high on the standard class and low on the same day because there are less number of couriers to send on the same day and the second highest no of couriers is on the second class and the third highest couriers are on the first class)

Bibliography

* Dataset from [www.kaggle.com](http://www.kaggle.com)
* Information from Wikipedia
* Used MS-Excel 2016
* Used MS-Word 2016
* Information from Google.com
* Learnt some things from Tutorial Point YouTube channel.