

PROJECT REPORT
ON
SALES DATA ANALYSIS

BY
SANTHOSH S

1. Introduction

Background

In today's fast-paced and competitive business environment, having a deep understanding of sales performance is crucial. Businesses generate vast amounts of sales data daily from various channels such as physical stores, online platforms, and third-party vendors. This data, if analyzed effectively, can provide valuable insights into customer preferences, sales trends, and market dynamics. Leveraging these insights can help businesses optimize their strategies, enhance customer satisfaction, and drive revenue growth.

Sales data analysis involves examining historical sales data to uncover patterns, trends, and anomalies. By applying analytical techniques, businesses can identify their best-selling products, peak sales periods, and the impact of different marketing campaigns. This analysis is instrumental in making informed decisions about inventory management, pricing strategies, and promotional activities.

Aim

The primary aim of this project is to conduct a comprehensive analysis of sales data to identify significant patterns and trends. Specifically, the project seeks to:

- Calculate key revenue metrics such as total sales.
- Analyze sales trends over different time periods.
- Identify best-selling products.
- Create visualizations to effectively communicate the findings.

These objectives are geared towards providing actionable insights that can inform business decisions and strategies aimed at increasing sales and profitability.

Purpose of the System

The purpose of this sales data analysis system is to enable businesses to transform raw sales data into meaningful insights. The system is designed to:

- **Collect and Store Sales Data:** Integrate and organize sales data from various sources into a central database.
- **Analyze Sales Performance:** Use statistical and computational techniques to evaluate sales data.
- **Visualize Data:** Create intuitive and interactive visualizations that make it easier to understand complex data sets.
- **Support Decision-Making:** Provide a solid foundation of data-driven insights that help businesses make strategic decisions regarding inventory, marketing, and sales tactics.

2. Analysis

Hardware and Software Requirements

➤ Hardware System Configuration

- processor - Pentium – IV
- RAM - 4 GB (min)
- Hard Disk - 20 GB

➤ Software System Configuration

- Operating System : Windows 7 or 8
- Software : MySQL Workbench

Feasibility Study

Technical Feasibility

The project requires knowledge of MySQL for database management and PowerBI for visualization. Both technologies are widely used and have ample documentation and community support.

Operational Feasibility

Businesses can easily integrate this system into their existing workflows. The use of common software tools like MySQL and PowerBI ensures compatibility with most business environments.

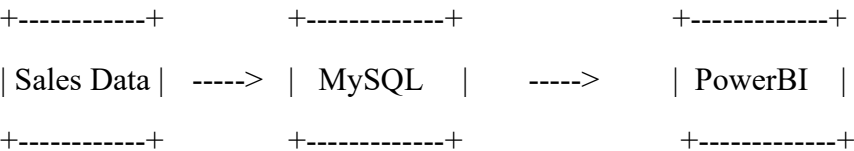
Economic Feasibility

The cost involved includes the subscription for PowerBI (if using the Pro version) and potential costs for MySQL hosting if cloud services are used. These costs are minimal compared to the potential benefits gained from data-driven decision-making.

3. Design

Block Diagram

The block diagram below illustrates the flow of the data cleansing process:



- 1. **Sales Data:** Raw sales data collected from various sources.
- 2. **MySQL:** Data storage and querying.
- 3. **PowerBI:** Data visualization and reporting.

4. Methodology

Methodology

The following steps outline the methodology used for the sales data analysis process:

4.1 Data Collection and Import

1. **Data Import:** Load the raw sales dataset into a MySQL database.

4.2 Data Storage

2. **Store Cleaned Data:** Store the prepared data in a structured format within the MySQL database. **Calculate Key Metrics:** Use SQL queries to calculate key revenue metrics:
3. **Total Sales:** Calculate the sum of the sales amount to determine the overall revenue generated.
4. **Sales Trends:** Group data by time periods (e.g., daily, monthly, quarterly) to observe trends and patterns over time.
5. **Best-Selling Products:** Identify top products by sales volume to understand which items contribute most to revenue.

4.3 Data Visualization

6. **Import Data into PowerBI:** Connect PowerBI to the MySQL database and import the cleaned and prepared sales data.
7. **Create Visualizations:** Design various charts and graphs to represent total sales, sales trends, and best-selling products. Use visual tools such as bar charts, line graphs, and pie charts to make the data easily interpretable.

5. CODING AND IMPLEMENTATION

6.

1. Create Database

create database sales;

use sales;

2. Import .CSV file

select * from orders;

Row ID	Order ID	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	City	State	Country	Postal Code	Market	Region	Product ID	Category
32298	CA-2012-124891	31-07-2012	Same Day	RH-19495	Rick Hansen	Consumer	New York City	New York	United States	10024	US	East	TEC-AC-10003033	Technology
26341	IN-2013-77878	07-02-2013	Second Class	JR-16210	Justin Ritter	Corporate	Wollongong	New South Wales	Australia		APAC	Oceania	FUR-CH-10003950	Furniture
25330	IN-2013-71249	18-10-2013	First Class	CR-12730	Craig Reiter	Consumer	Brisbane	Queensland	Australia		APAC	Oceania	TEC-PH-10004664	Technology
13524	ES-2013-1579342	30-01-2013	First Class	KM-16375	Katherine Murray	Home Office	Berlin	Berlin	Germany		EU	Central	TEC-PH-10004583	Technology
47221	SG-2013-4320	06-11-2013	Same Day	RH-9495	Rick Hansen	Consumer	Dakar	Dakar	Senegal		Africa	Africa	TEC-SHA-10000501	Technology
22732	IN-2013-42360	01-07-2013	Second Class	JM-15655	Jim Mithum	Corporate	Sydney	New South Wales	Australia		APAC	Oceania	TEC-PH-10000030	Technology
30570	IN-2011-81826	09-11-2011	First Class	TS-21340	Toby Swindell	Consumer	Porirua	Wellington	New Zealand		APAC	Oceania	FUR-CH-10004050	Furniture
31192	IN-2012-86369	18-04-2012	Standard Class	MB-18085	Mick Brown	Consumer	Hamilton	Waikato	New Zealand		APAC	Oceania	FUR-TA-10002958	Furniture
40155	CA-2014-135909	21-10-2014	Standard Class	JW-15220	Jane Waco	Corporate	Sacramento	California	United States	95823	US	West	OFF-BI-10003527	Office Supplies
40936	CA-2012-116638	31-01-2012	Second Class	JH-15985	Joseph Holt	Consumer	Concord	North Carolina	United States	28027	US	South	FUR-TA-10000198	Furniture
34577	CA-2011-102988	09-04-2011	Second Class	GM-14695	Greg Maxwell	Corporate	Alexandria	Virginia	United States	22304	US	South	OFF-SU-10002881	Office Supplies

3. SQL queries to calculate Total Sales:

select Category,

sum(Quantity*Sales) as total_sales from orders group by Category;

Category	total_sales
Technology	23045350.70148007
Furniture	20049448.02810021
Office Supplies	18608455.020500634

4. Group data by time periods

select

year(order_date) as year,

sum(case when month(order_date)=1 then Sales else 0 end) as jan_sales,

sum(case when month(order_date)=2 then Sales else 0 end) as feb_sales,

sum(case when month(order_date)=3 then Sales else 0 end) as mar_sales,

sum(case when month(order_date)=4 then sales else 0 end) as apr_sales,

```

sum(case when month(order_date)=5 then sales else 0 end) as may_sales,
sum(case when month(order_date)=6 then sales else 0 end) as jun_sales,
sum(case when month(order_date)=7 then sales else 0 end) as jul_sales,
sum(case when month(order_date)=8 then sales else 0 end) as aug_sales,
sum(case when month(order_date)=8 then sales else 0 end) as sep_sales,
sum(case when month(order_date)=10 then sales else 0 end) as oct_sales,
sum(case when month(order_date)=11 then sales else 0 end) as nov_sales,
sum(case when month(order_date)=12 then Sales else 0 end) as dec_sales,
sum(sales) as Total_sales_year from orders
where year(order_date) in (2011,2012,2013,2014)
group by year(order_date)
order by year(order_date);

```

year	jan_sales	feb_sales	mar_sales	apr_sales	may_sales	jun_sales	jul_sales
2011	98822.94285999997	90752.0809800001	145564.55736	116759.98017999993	146734.21610000005	215121.24822000007	115064.89912000007
2012	135691.76024	100483.07798	162971.81116000027	159497.54351999983	207559.17523999997	256038.09042000017	144995.98112000007
2013	199076.00737999988	166967.25040000002	197710.70412000033	177658.41884000011	259968.15670000005	395803.26990000013	229774.99700000002
2014	241157.4676600001	184798.54456000018	262967.6436200003	242270.49830000012	288235.28813999996	401074.89909999934	257931.71248000022

jul_sales	aug_sales	sep_sales	oct_sales	nov_sales	dec_sales	Total_sales_year
115064.89912000007	207275.53521999973	207275.53521999973	199012.81204000043	298165.06152000016	333759.95059999999	2256967.1955399867
144995.98112000007	302625.53837999999	302625.53837999999	252917.08220000012	322238.886900000057	337815.92460000008	2671332.421399961
229774.99700000002	326387.81736000004	326387.81736000004	291837.35688000056	373065.6561000012	404497.04202000003	3397806.112379975
257931.71248000022	456142.46236000065	456142.46236000065	422520.4291600003	554214.583	500541.99048000004	4291645.176559992

5. Identify Best-Selling Products

```

select Category, total_sales_amount, RANK() over(order by total_sales_amount DESC) AS
sales_rank

```

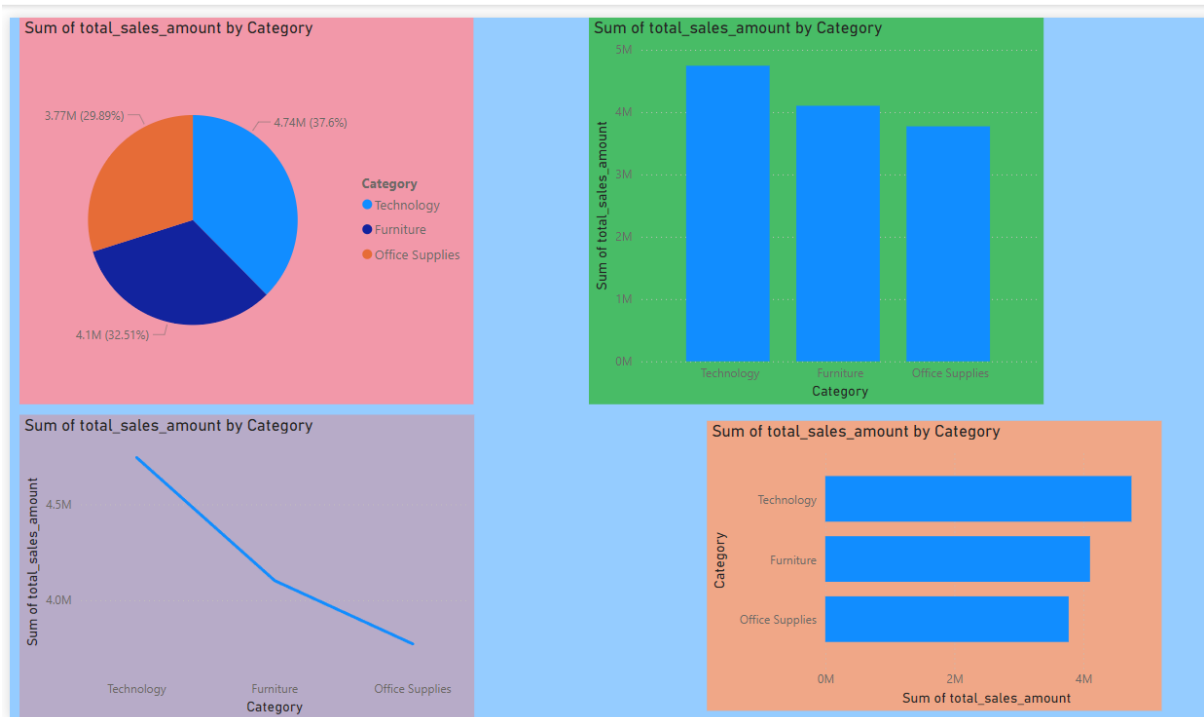
```

from ( select category, sum(sales) as total_sales_amount from orders group by Category) as
sales_summary order by total_sales_amount DESC ;

```

Category	total_sales_amount	sales_rank
Technology	4744303.53208002	1
Furniture	4101921.252299971	2
Office Supplies	3771526.1214999114	3

6. Visualisation using Power BI



5. Discussion

Results and Observations

- **Total Sales:** The system provides a clear calculation of total sales.
- **Sales Trends:** Trends over time help in identifying peak sales periods and seasonal effects.
- **Best-Selling Products:** Highlighting top products aids in inventory and marketing strategy planning.

Challenges and Solutions

- **Data Quality:** Ensuring data accuracy was challenging. Implemented thorough data cleaning processes.
- **Integration:** Combining MySQL and PowerBI required ensuring compatible data formats and structures. Used consistent data schemas and data transformation tools.

Future Enhancements

- **Predictive Analytics:** Incorporate machine learning models to forecast future sales.
- **Real-Time Data Analysis:** Implement real-time data processing for up-to-the-minute insights.
- **Enhanced Visualizations:** Utilize advanced PowerBI features for more interactive and detailed visualizations.