



# CAPSTONE PROJECT

# ACTIONABLE

- ❖ Data Exploration and Summarization
- ❖ Data Cleaning and Statistical Analysis
- ❖ Graphical Analysis and Insights using Excel
- ❖ Database Setup and SQL Operation
- ❖ Running Queries in SQL
- ❖ Data Import to Power BI from SQL Database
- ❖ Interactive Dashboard Creation in Power BI
- ❖ Conclusion

01

# Data Exploration And Summarization



# Key Insights On the Data

- In the data set there are **16 Columns** which contains different variables like Segments, Country, Product, Discount Band, Units Sold, Manufacturing Price, Sales Price, Gross Sale, Discounts, Sales, COGS, Profit, Date, Month Number, Month Name and Year.
- These Columns contains different **types of columns**. They are as follows;
  - 05 categorical column
  - 10 numeric column &
  - 01 date wise column
- In the data set there are 700 rows which contains different values like numerical, Textual and non-numerical values. Specially, Discount Column contains 53 non-numerical data.

# VARIABLE OVERVIEW

**01**

## **Categorical Variables:**

Segment: 5 unique values  
Country: 5 unique values  
Product: 6 unique values  
Discount Band: 4 unique values

**02**

## **Numerical Variables:**

Units Sold: 510  
Manufacturing Price: 6  
Sale Price: 7  
Gross Sales: 550  
Discounts: 515  
Sales: 559  
COGS: 545  
Profit: 557

**03**

## **Date wise Variables:**

Date: 16 unique values  
Month Number: 12 unique values  
Month Name: 12 unique values  
Year: 1 unique value



02

## Data Cleaning and Statistical Analysis

# Data cleaning Steps

- **Structured the columns according to their nature of the values:** It stored before in appropriate format. Such as columns which represents currency was stored in custom format, due to that calculation could be difficult during analysis so I change it as currency format.
- **Removed extra spaces within the values and texts.**
- **Modified column into standardised way:** As most of the columns are not stored as their natural position like adjusted text value in left side of the column and numbers in right side of the column.
- **Replaced the non-numerical value \$ - with \$0 through customising method which was earlier stored in Discount column.**

# Statistical Analysis

STATISTICAL ANALYSIS IN EXCEL						
Statistical Values		Sales	COGS	Profit	Discount	Units Sold
Total		118726350.26	101832648.00	16893702.26	9205248.24	1125806.00
Mean		169609.07	145475.21	24133.86	13150.35	1608.29
Median		35540.20	22506.25	9242.20	2585.25	1542.50
Mode		32670.00	17430.00	0.00	0.00	727.00
Standard Deviance		236726.35	203865.51	42760.63	22962.93	867.43
Variance		56039363321.24	41561144584.94	1828471184.09	527296097.92	752431.09
Maximum Value		1159200.00	950625.00	262200.00	149677.50	4492.50
Minimum Value		1655.08	918.00	-40617.50	0.00	200.00
2nd Largest Value		1038082.50	948375.00	247500.00	125820.00	4251.00
Range		1157544.92	949707.00	302817.50	149677.50	4292.50
First Quartile		15928.00	7490.00	2805.96	800.32	905.00
Third Quartile		261077.50	245607.50	22662.00	15956.34	2229.13
Skewness		1.69	1.55	2.71	2.68	0.44
Kurtosis		2.19	1.61	8.68	7.91	-0.32



# Key Points On Statistical Analysis Table

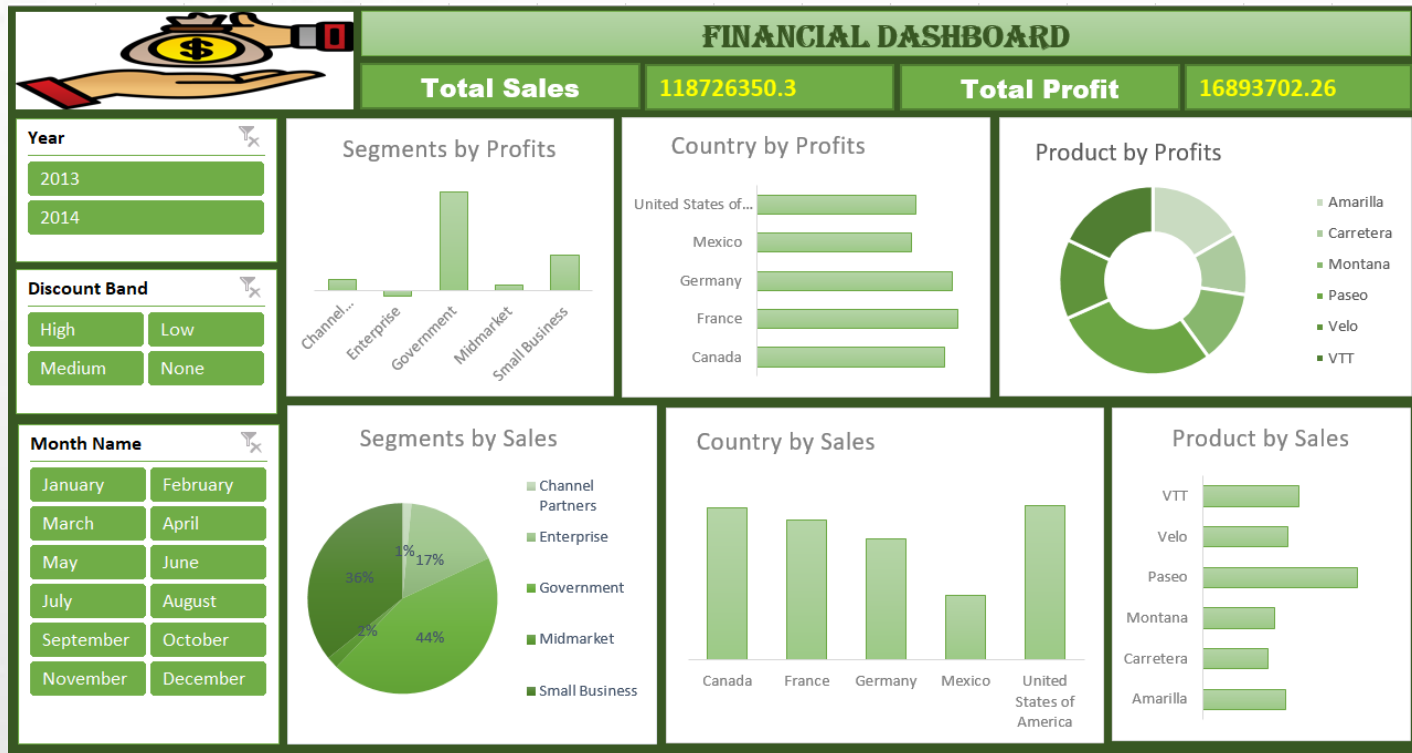
- In the table data signifies, For selling 1125806 units of Overall Products, The Manufacturing Companies issued 7.75% discounts and 16.58% profits generated on their respective total sales of \$118726350.26.
- By seeing the Statistical table we can observe difference mean, mode and median value. Which signifies that data is not normally distributed during the period of 2013-14.
- In The table data tells, profit is lied on 302817.50 range zone, By this we can consider overall companies are making profits while selling their products on market.

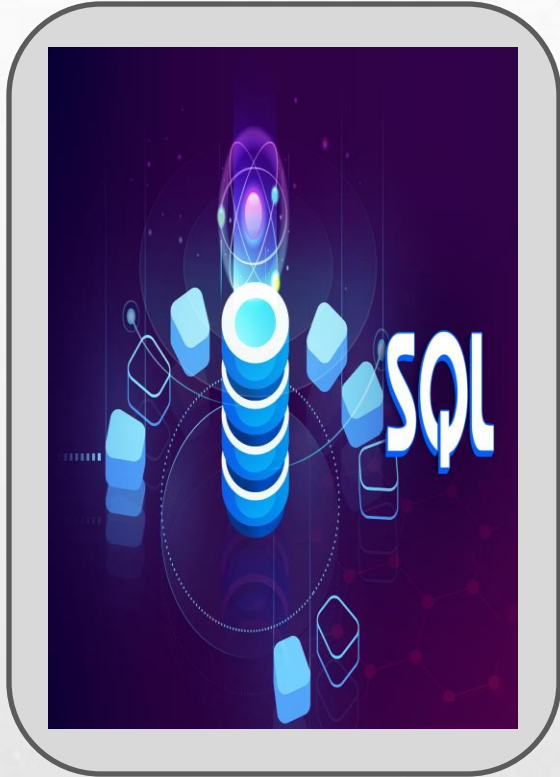


03

## Graphical Analysis using Excel

# Graphical Analysis





# 04

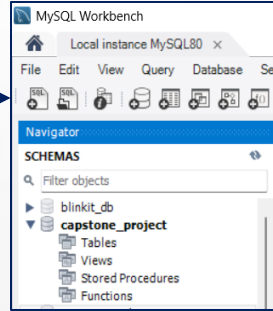
## Data Base Setup And SQL Operation

# Data Base Setup And SQL Operation

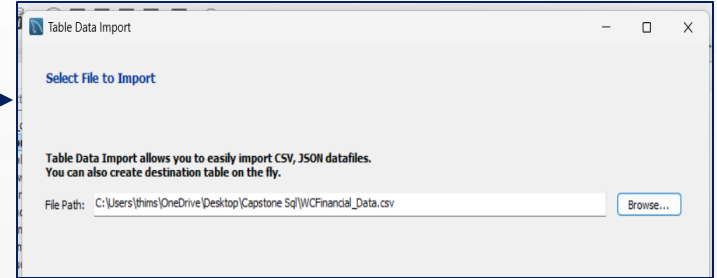
## Data Base Creation

```
1 • create database capstone_project;  
2 • use capstone_project;
```

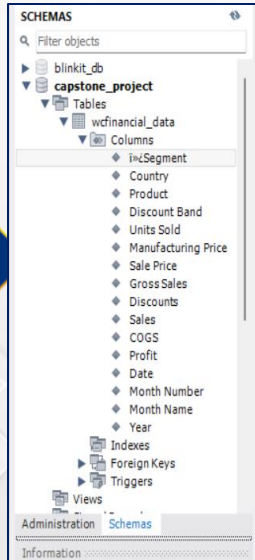
## Data Base Created



## Importing Data Set



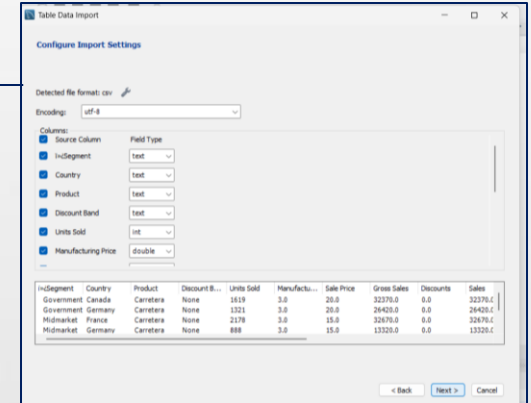
## Data Set Loaded



## Data Set Loading Process



## Data Type Addition



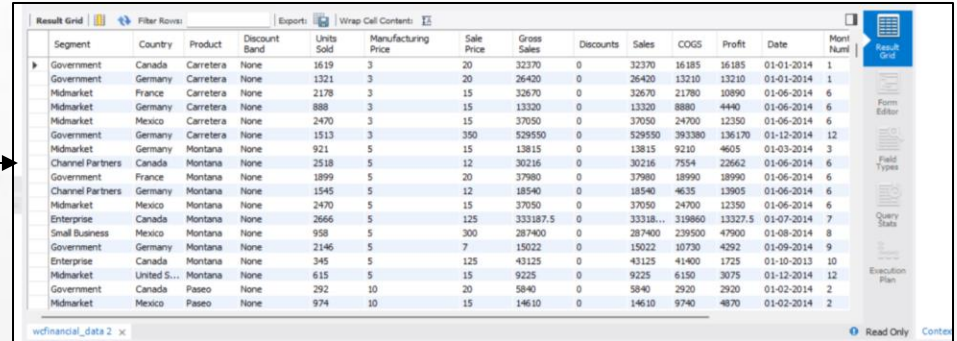
## Showing in Data Base

# Running Queries In SQL

## SQL Queries

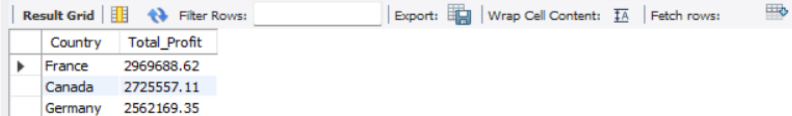
```
9 • select * from wcfinancial_data;
```

## SQL Out Comes



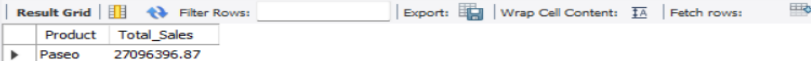
Segment	Country	Product	Discount Band	Units Sold	Manufacturing Price	Sale Price	Gross Sales	Discounts	Sales	COGS	Profit	Date	Month Num
Government	Canada	Carretera	None	1619	3	20	32370	0	32370	16185	16185	01-01-2014	1
Government	Germany	Carretera	None	1321	3	20	26420	0	26420	13210	13210	01-01-2014	1
Midmarket	France	Carretera	None	2178	3	15	32670	0	32670	21780	10890	01-06-2014	6
Midmarket	Germany	Carretera	None	888	3	15	13320	0	13320	8880	4440	01-06-2014	6
Midmarket	Mexico	Carretera	None	2470	3	15	37050	0	37050	24700	12350	01-06-2014	6
Government	Germany	Carretera	None	1513	3	350	529550	0	529550	393380	136170	01-12-2014	12
Midmarket	Germany	Montana	None	921	5	15	13815	0	13815	9210	4605	01-03-2014	3
Channel Partners	Canada	Montana	None	2518	5	12	30216	0	30216	7554	22662	01-06-2014	6
Government	France	Montana	None	1899	5	20	37980	0	37980	18990	18990	01-06-2014	6
Channel Partners	Germany	Montana	None	1545	5	12	18540	0	18540	4635	13905	01-06-2014	6
Midmarket	Mexico	Montana	None	2470	5	15	37050	0	37050	24700	12350	01-06-2014	6
Enterprise	Canada	Montana	None	2666	5	125	333187.5	0	333187.5	319860	13327.5	01-07-2014	7
Small Business	Mexico	Montana	None	958	5	300	287400	0	287400	239500	47900	01-08-2014	8
Government	Germany	Montana	None	2146	5	7	15022	0	15022	10730	4292	01-09-2014	9
Enterprise	Canada	Montana	None	345	5	125	43125	0	43125	41400	1725	01-10-2013	10
Midmarket	United S...	Montana	None	615	5	15	9225	0	9225	6150	3075	01-12-2014	12
Government	Canada	Paseo	None	292	10	20	5840	0	5840	2920	2920	01-02-2014	2
Midmarket	Mexico	Paseo	None	974	10	15	14610	0	14610	9740	4870	01-02-2014	2

```
11 -- Top 3 Profit earning countries in 2014  
12 • select Country, round(sum(Profit),2) as Total_Profit  
13 from wcfinancial_data where Year = '2014' group by  
14 Country order by Total_Profit Desc limit 3;
```



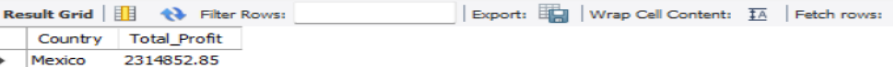
Country	Total_Profit
France	2969688.62
Canada	2725557.11
Germany	2562169.35

```
23 -- Which brand is most sold in the year 2014?  
24 • select Product, round(sum(Sales),2) as Total_Sales  
25 from wcfinancial_data where Year = '2014' group by  
26 Product order by Total_Sales desc limit 1;
```



Product	Total_Sales
Paseo	27096396.87

```
35 -- Which country has the earned the lowest profit in 2014?  
36 • select Country, round(sum(Profit),2) as Total_Profit  
37 from wcfinancial_data where Year = '2014' group by  
38 Country order by Total_Profit Asc limit 1;
```



Country	Total_Profit
Mexico	2314852.85



05

## Data Import To Power Bi From SQL Data Base

# Data Import to Power Bi From SQL Data Base

**Get Data**

Search

Database

- All
- File
- Database
- Microsoft Fabric
- Power Platform
- Azure
- Online Services
- Other

Database

- SQL Server database
- Access database
- SQL Server Analysis Services database
- Oracle database
- IBM Db2 database
- IBM Informix database (Beta)
- IBM Netezza
- MySQL database
- BusinessESM database
- Import data from a MySQL database
- Teradata database
- SAP HANA database
- SAP Business Warehouse Application Server
- SAP Business Warehouse Message Server
- Amazon Redshift
- Impala

Certified Connectors | Template Apps

Connect Cancel

**MySQL database**

Server

localhost:3306

Database

Capstone\_Project

> Advanced options

OK Cancel

Get data from another source →

**Navigator**

Display Options

- localhost:3306: Capstone\_Project [1]
- capstone\_project.wcfinancial\_data

capstone\_project.wcfinancial\_data

Segment	Country	Product	Discount Band	Un
Government	Canada	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	France	Carretera	None	
Midmarket	Germany	Carretera	None	
Midmarket	Mexico	Carretera	None	
Government	Germany	Carretera	None	
Midmarket	Germany	Montana	None	
Channel Partners	Canada	Montana	None	
Government	France	Montana	None	
Channel Partners	Germany	Montana	None	
Midmarket	Mexico	Montana	None	
Enterprise	Canada	Montana	None	
Small Business	Mexico	Montana	None	
Government	Germany	Montana	None	
Enterprise	Canada	Montana	None	
Midmarket	United States of America	Montana	None	
Government	Canada	Paseo	None	
Midmarket	Mexico	Paseo	None	
Channel Partners	Canada	Paseo	None	
Government	Germany	Paseo	None	
Channel Partners	Germany	Paseo	None	
Government	Mexico	Paseo	None	
Midmarket	France	Paseo	None	

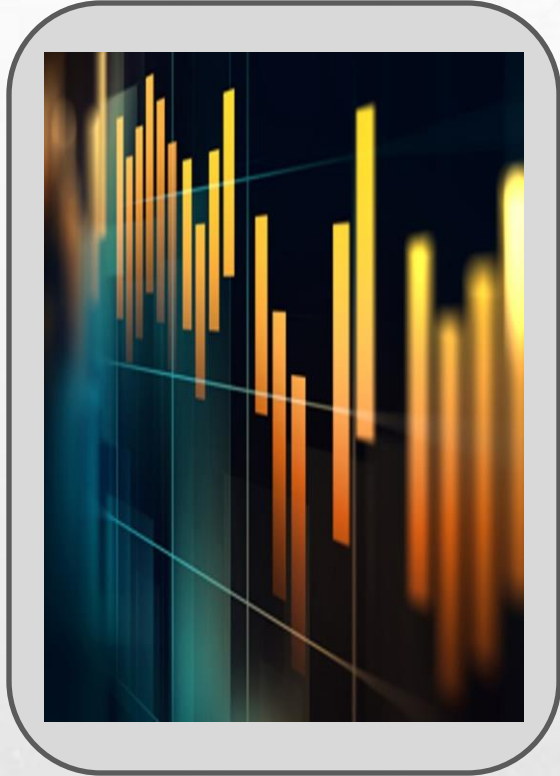
Select Related Tables

Load Transform Data Cancel



# Data Importing Steps

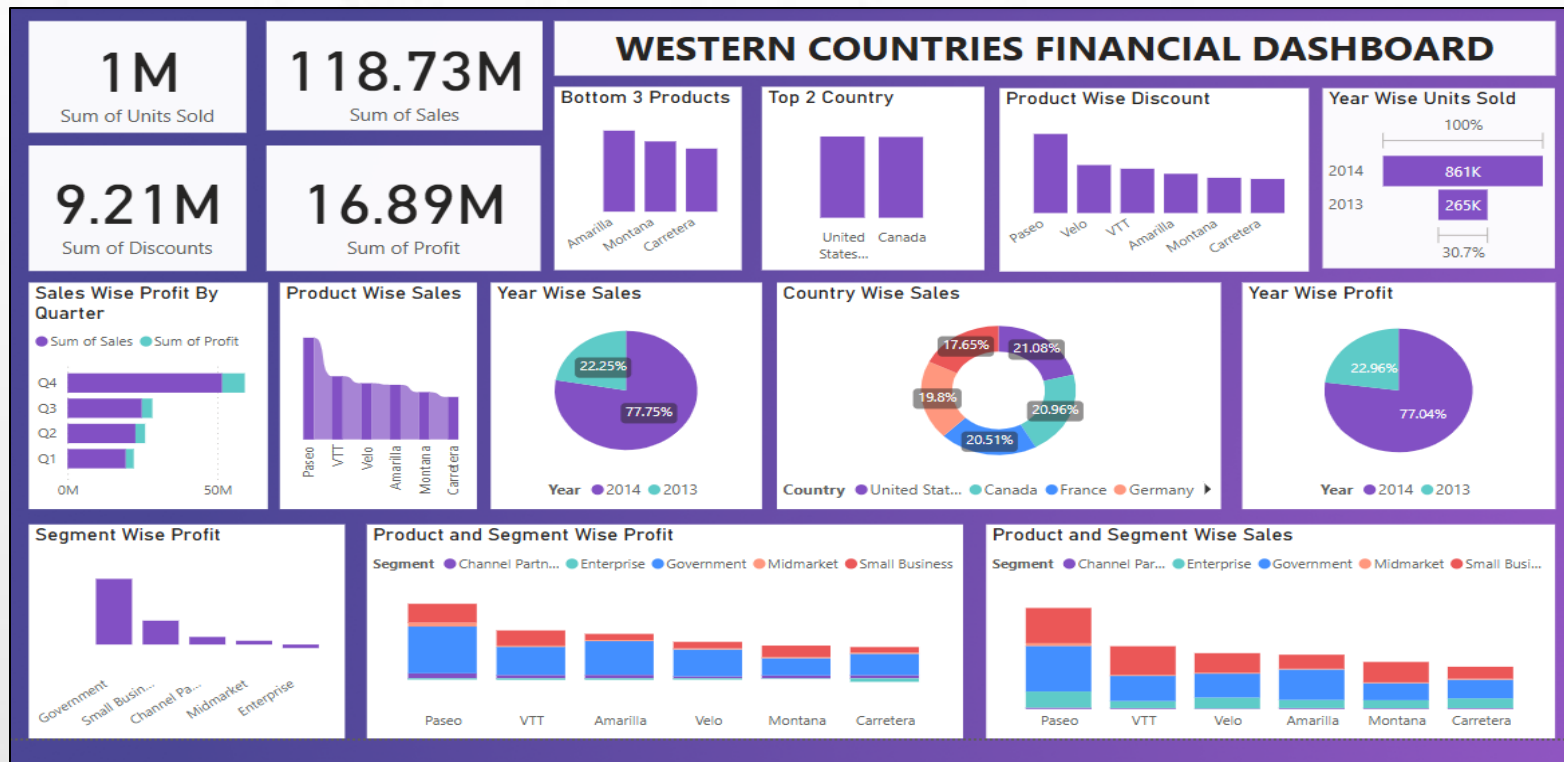
- Opened Power Bi and navigate to home tab
- Clicked on “Get Data” and selected “My SQL Data Base” from the list
- Entered the server address as “localhost:3306”
- Entered the data base name as “Capstone\_Project”
- Clicked OK button and chooses required table “wcfinancial\_data” for transfer file to power query editor page
- Made necessary changes for data visualisation purpose.
- Loaded the data into Power Bi dashboard and Performed the necessary visualisation process.



06

## Interactive Dashboard Creation in Power Bi

# Power Bi Dash Board





07

Conclusion

# Key Insights and Conclusion On The Western Countries Sales Data

- As appears in the Western Country Sales Data 78% of the sales occurs in year 2014 which indicates YOY products are gained significant growth on their sales.
- USA and Canada has emerged as top most product selling countries and Mexico has in the bottom line of the sales.
- 77% of the sales comes in the year 2014 and rest 23% in 2013.
- In year 2014 861k units are sold and in 2013 there are 265k units. Which describes demand for the products are tripled on Year on Year basis.
- In segment wise profit Government Segment has emerged has highest profit earning segment and Enterprise segment has in negative side of their earnings.
- In Product and Segment wise Profit Paseo has earned more profit when compared to its rivalries where as Montona and Carretera Products are least in the list.
- In Product and Segment wise sales Paseo gained the top position and Carretera in the bottom least side of the sales.
- Paseo issued most no of discounts to sell to sell their products and Carretera issued less no. of discounts which signifies customers are shown keen interest to buy discounted product.
- During Quarter-4 (October to December) there is a significant growth in sales when compare to Quarter-1 (January to March) which describes companies can invest more on Q4 to earn profit.

# THANKS!

