Practical 1

Aim-: Import the legacy data from different sources such as (EXCEL,SQLSERVER,ORACLE) and load in the target system

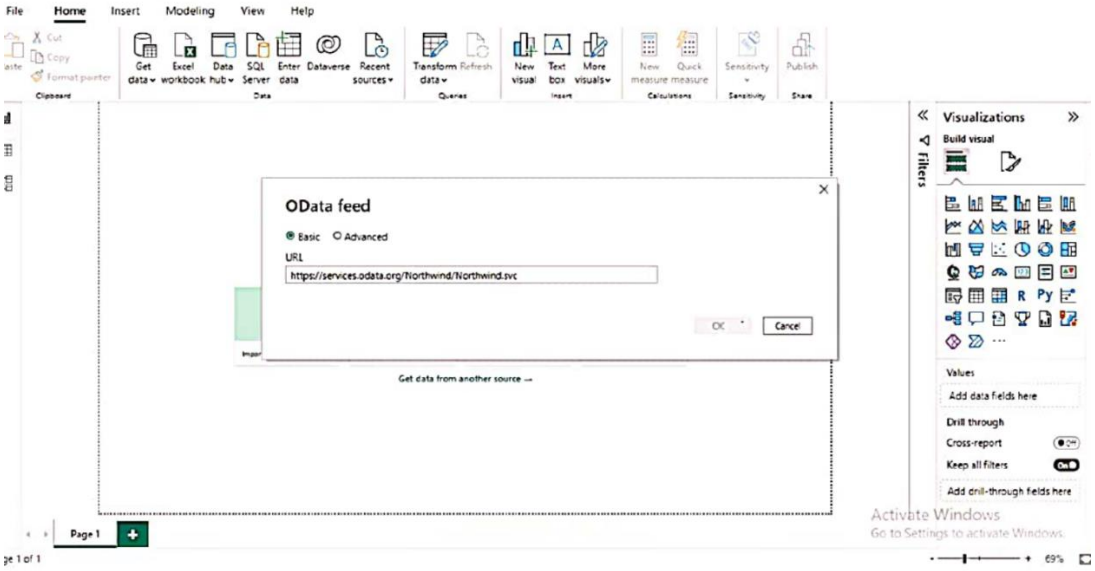
Definition of legacy data-:

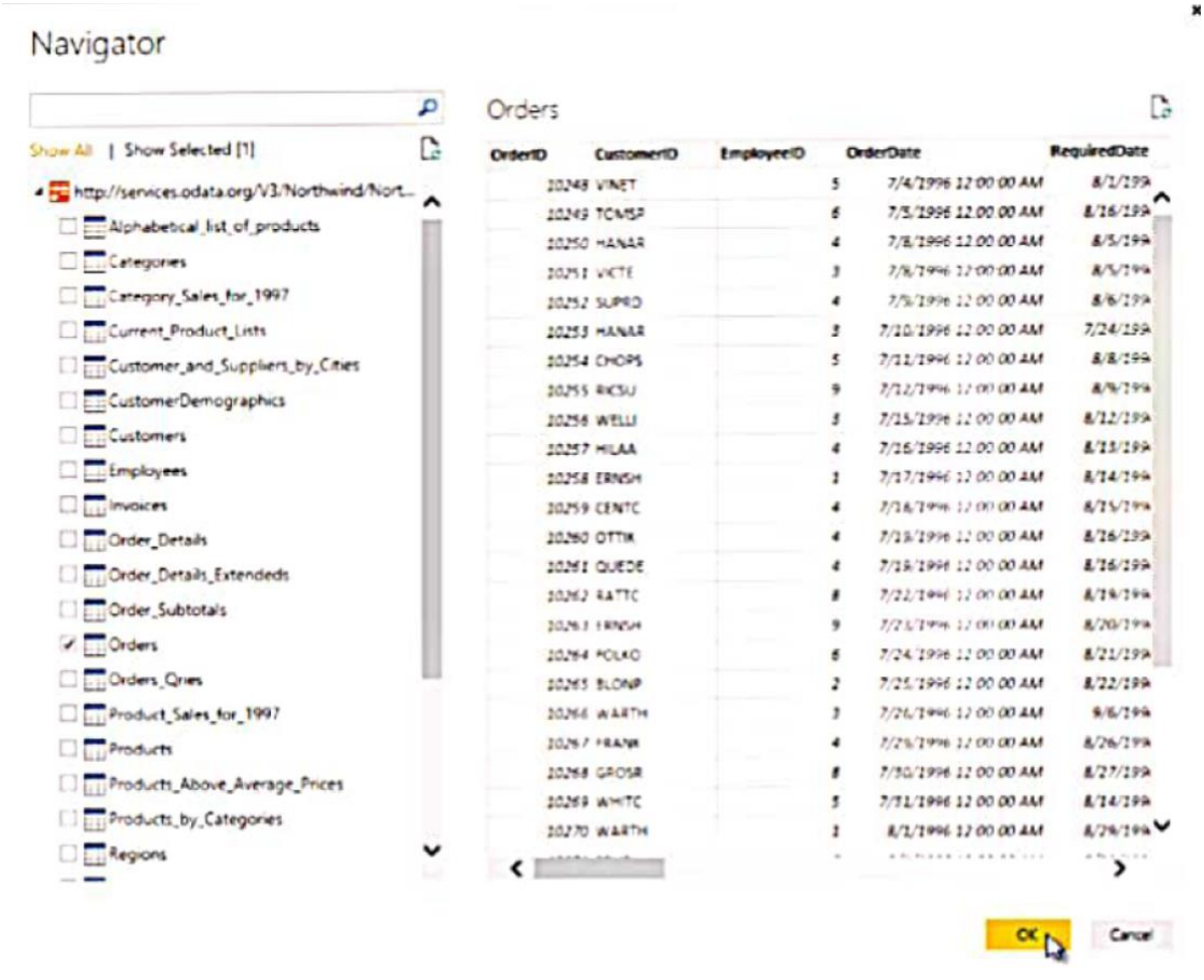
1) The Legacy data is an electronic data that is only understood by only compatible with or only resides on hardware or software that has become obsolete.
2) This means that legacy data includes everything that modern technology is not able to open.

Dataset Used-: Northwind Database

1) The Northwind database is a simple database used to demonstrate the features of some of it products.
2) The database contains the sales data for Northwind Traders, a fictious Speciality foods export-import company.

Output-:


File
Home
Insert
Modeling
View
Help
OData feed
Basic
Advanced
URL
https://services.odata.org/Northwind/Northwind.svc
OK
Cancel
Get data from another source
Visualizations
Build visual
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1
Activate Windows
Go to Settings to activate Windows.


Navigator
Show All | Show Selected [1]
http://services.odata.org/V3/Northwind/Nort...
Alphabetical_list_of_products
Categories
Category_Sales_for_1997
Current_Product_Lists
Customer_and_Suppliers_by_Cities
CustomerDemographics
Customers
Employees
Invoices
Order_Details
Order_Details_Extendeds
Order_Subtotals
Orders
Orders_Qries
Product_Sales_for_1997
Products
Products_Above_Average_Prices
Products_by_Categories
Regions
Orders
OrderID
CustomerID
EmployeeID
OrderDate
RequiredDate
OK
Cancel

Practical 2

Aim-: Perform the Extraction Transformation and Loading (ETL) Process to Construct the database in the SQL Server

Definition of Extraction Transformation and Loading-:

1) In Computing, extract, transform, load is a 3 Phases process where data is extracted, transformed & loaded into an output data container.
2) It is a tool performed on a given set of data

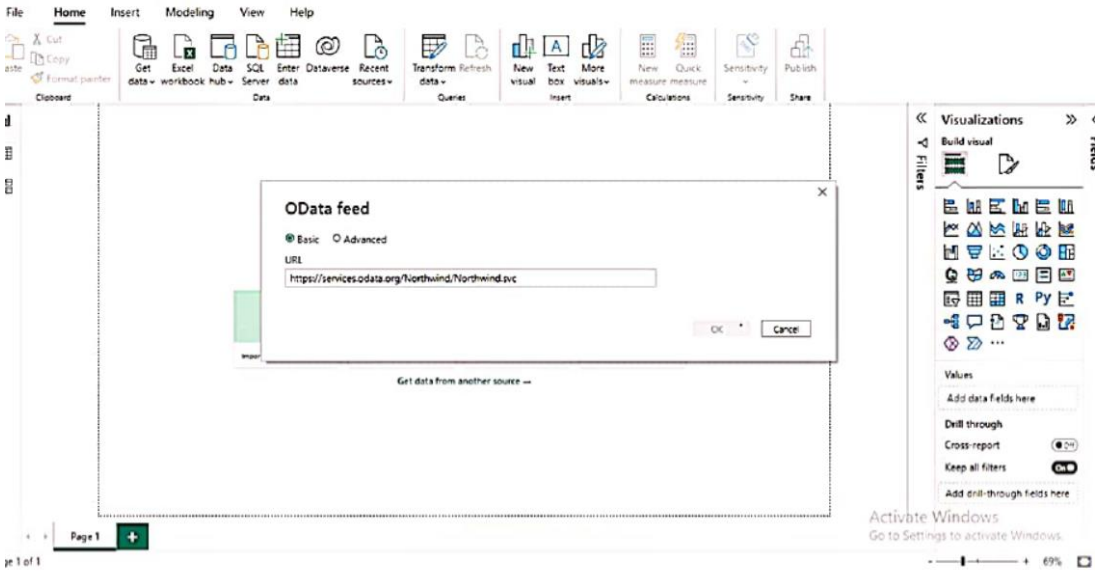
   in order to accomplish a specific business goal.
3) Power B.I. Extract Transform load (ETL)Dataflow is a cloud-based platform that users in data preparation.
4) Users can also utilize Power B.I. Dataflows to ingest, convert, and load data into Microsoft Dataverse environments, Power B.I. Workspace.
5) Without ET2 Tools, the analytical reports and dashboards look old because of outdated data.
6) ETL helps update them so your reports are current.
7) ETL includes the following processes :-
   The data is extracted from a data source, then transformed, validated standardized, corrected

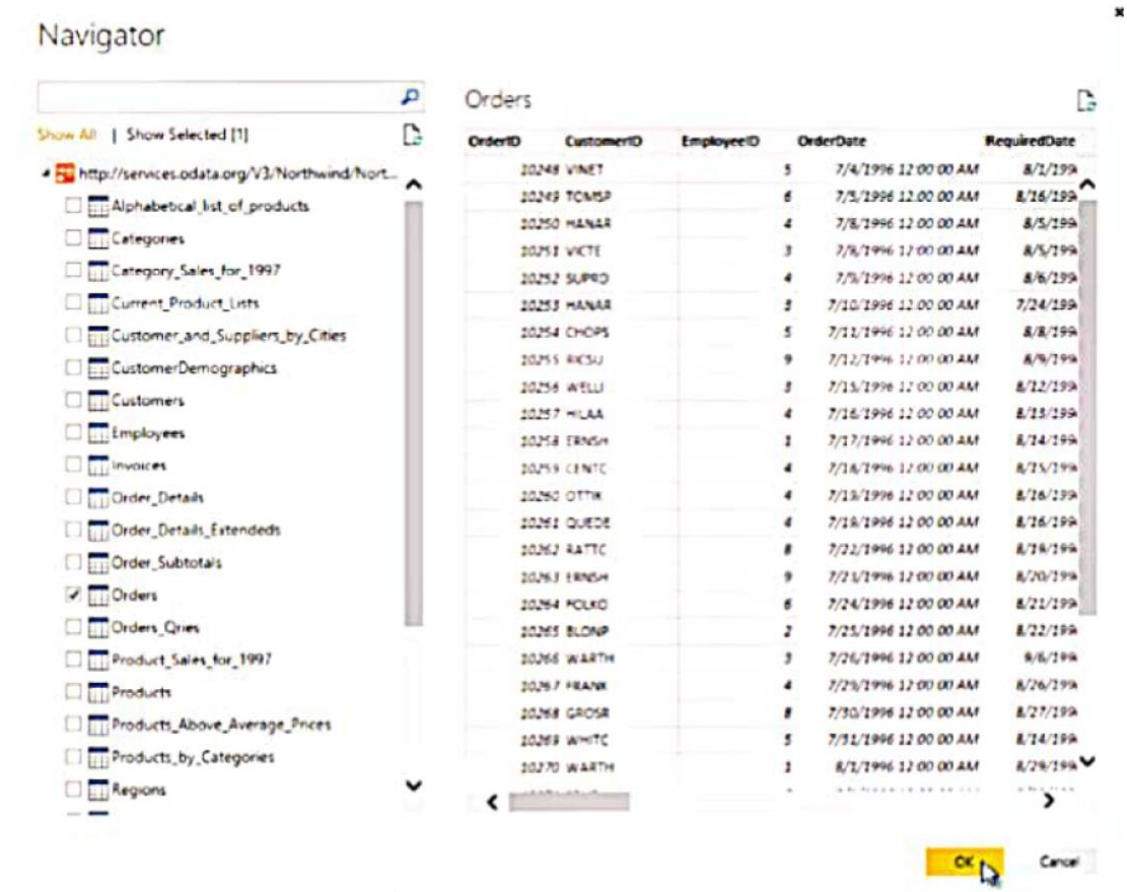
loaded into a-quality checked and ultimately data repository, such as data mart or data warehouse, where it is streamlined for analysis and reporting.

Output-:

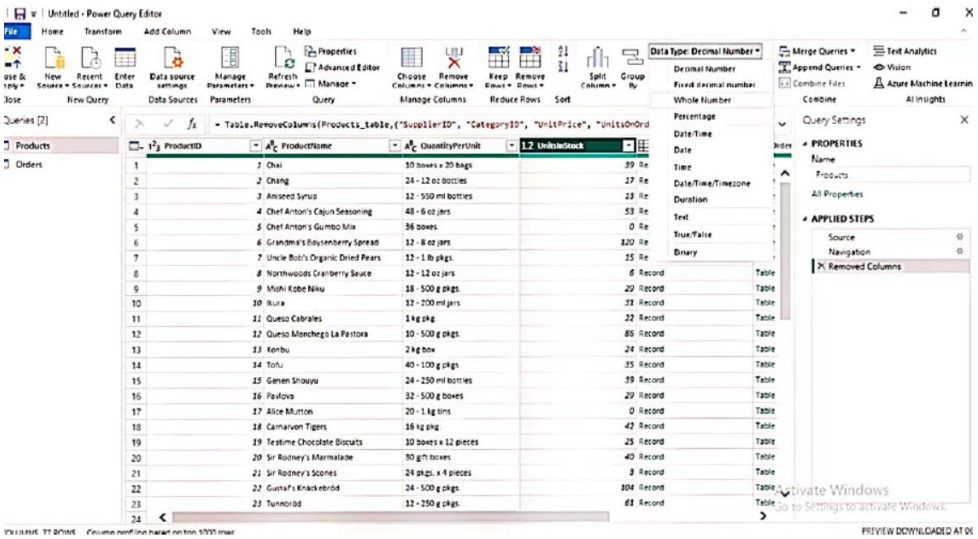
Extract-:

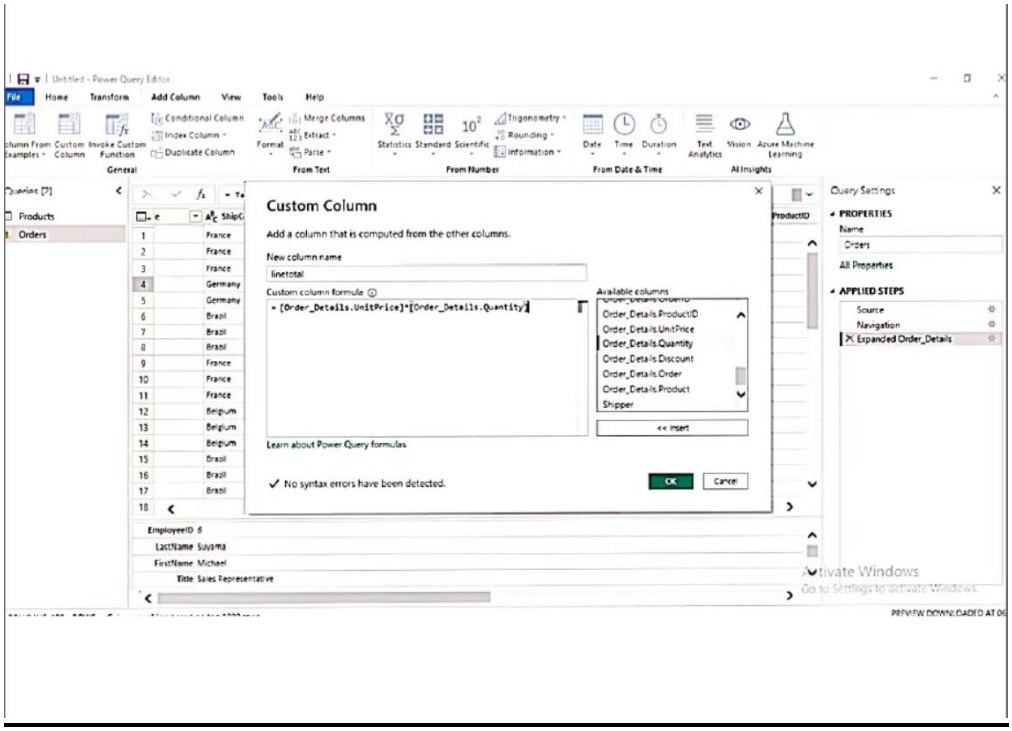
Extract




Navigator
Show All | Show Selected [1]
http://services.odata.org/V3/Northwind/Nort...
Alphabetical_list_of_products
Categories
Category_Sales_for_1997
Current_Product_Lists
Customer_and_Suppliers_by_Cities
CustomerDemographics
Customers
Employees
Invoices
Order_Details
Order_Details_Extendeds
Order_Subtotals
Orders
Orders_Qries
Product_Sales_for_1997
Products
Products_Above_Average_Prices
Products_by_Categories
Regions
Orders
OrderID
CustomerID
EmployeeID
OrderDate
RequiredDate
OK
Cancel

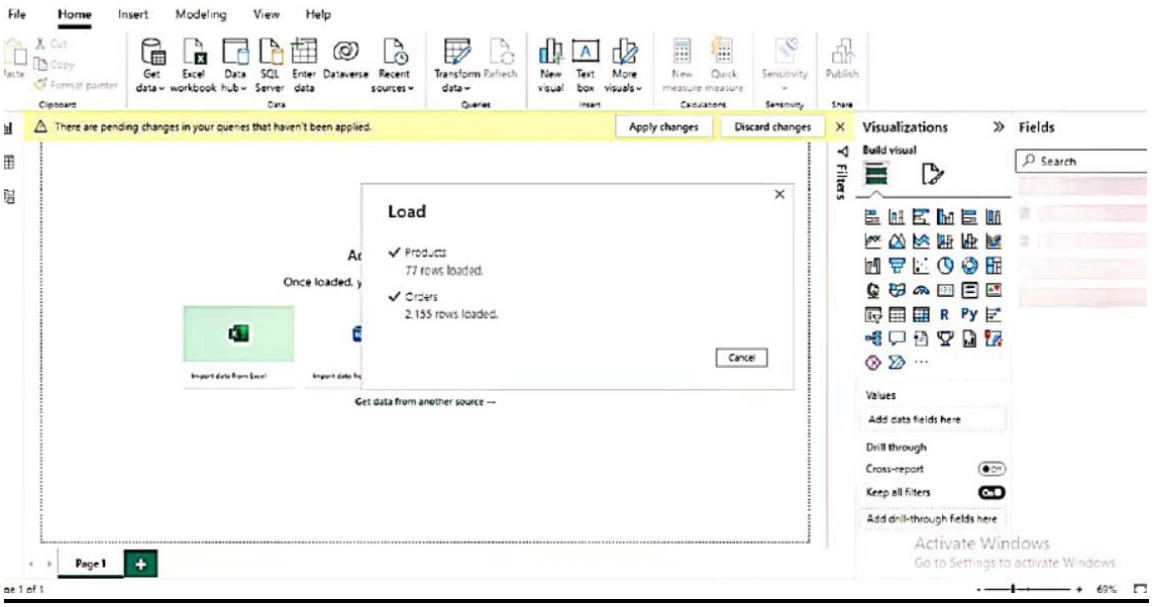
Transform-:


Untitled - Power Query Editor
Home
Transform
Add Column
View
Tools
Help
Queries [2]
Products
Orders
Query Settings
PROPERTIES
Name
Products
All Properties
APPLIED STEPS
Source
Navigation
Removed Columns
Activate Windows


Custom Column
Add a column that is computed from the other columns.
New column name
linetotal
Custom column formula
= [Order_Details.UnitPrice]*[Order_Details.Quantity]
Available columns
Order_Details.ProductID
Order_Details.UnitPrice
Order_Details.Quantity
Order_Details.Discount
Order_Details.Order
Order_Details.Product
Shipper
<< Insert
Learn about Power Query formulas
No syntax errors have been detected.
OK
Cancel
Query Settings
PROPERTIES
Name
Orders
All Properties
APPLIED STEPS
Source
Navigation
Expanded Order_Details

Load-:

Load


There are pending changes in your queries that haven't been applied.
Apply changes
Discard changes
Load
Products
77 rows loaded.
Orders
2,155 rows loaded.
Cancel
Get data from another source
Visualizations
Build visual
Fields
Search
Values
Add data fields here
Drill through
Cross-report
Keep all filters
Add drill-through fields here
Page 1

Practical 3

Aim-: Perform Data visualization from ETL (Using Power B.I)

Definition of Data Visualization-:

1) Data visualisation helps you turn all the granular data into easily understood, visually compelling and useful business information.

2) By using Power B.I. as a tool for Data visualization it helps us to see our KPIs more clearly, it unifies the data.

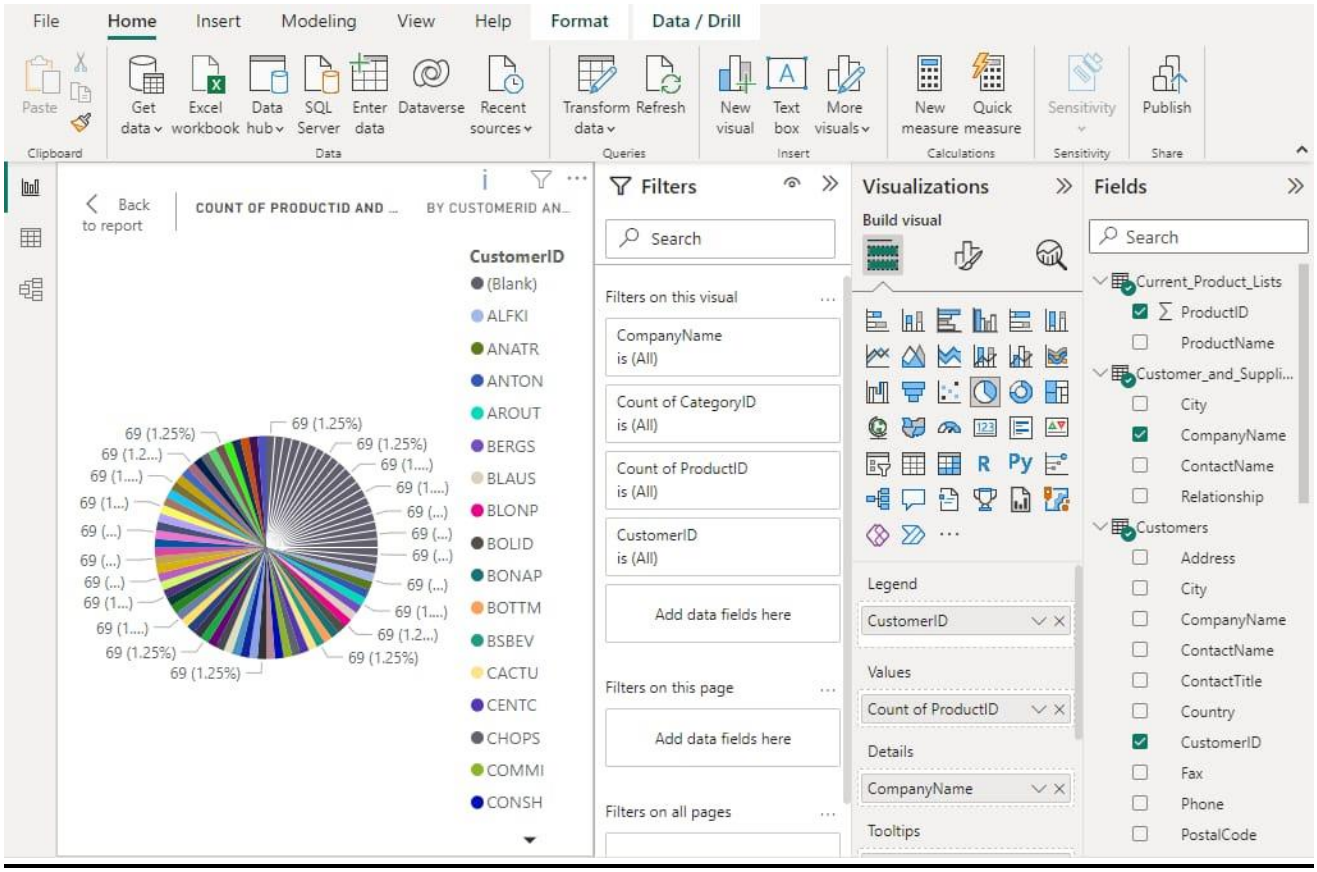
3) Data visualization brings data to life, making you the master storyteller of the insights hidden within your data.

(4) Data Visualization helps users develop powerful business insight quickly and effectively.

Definition of ETL Process-:

1) Tableau Prep tool is a tool which excellent for Data Visualization and has some ETL capabilities in the Tableau.
2) ETL Tool is anacronym for Extract, Transform and load
3) In ETL Tool the data is collected from different heterogenous sources and transformed into information and loaded into data warehouses.

Output-:


File
Home
Insert
Modeling
View
Help
Format
Data / Drill
Paste
Get data
Excel workbook
Data hub
SQL Server
Enter data
Dataverse
Recent sources
Transform data
Refresh
New visual
Text box
More visuals
New measure
Quick measure
Sensitivity
Publish
Clipboard
Data
Queries
Insert
Calculations
Sensitivity
Share
Back to report
COUNT OF PRODUCTID AND ...
BY CUSTOMERID AN...
CustomerID
(Blank)
ALFKI
ANATR
ANTON
AROUT
BERGS
BLAUS
BLONP
BOLID
BONAP
BOTTM
BSBEV
CACTU
CENTC
CHOPS
COMMI
CONSH
69 (1.25%)
69 (1.2...)
69 (1....)
69 (1...)
69 (...)
Filters
Search
Filters on this visual
CompanyName is (All)
Count of CategoryID is (All)
Count of ProductID is (All)
CustomerID is (All)
Add data fields here
Filters on this page
Add data fields here
Filters on all pages
Visualizations
Build visual
Legend
CustomerID
Values
Count of ProductID
Details
CompanyName
Tooltips
Fields
Search
Current_Product_Lists
ProductID
ProductName
Customer_and_Suppli...
City
CompanyName
ContactName
Relationship
Customers
Address
City
CompanyName
ContactName
ContactTitle
Country
CustomerID
Fax
Phone
PostalCode

Practical 4

Aim-:

a) Import the data warehouse in Microsoft Excel and create the Pivot table and Pivot Chart.

b) Import the cube in Microsoft Excel and create the Pivot Chart to perform data analysis.

Definition of Pivot Table-:

1) Pivot is table is a simple tool used to create a Summarized report from a large set of databases.
2) Pivot Table is an interactive way to quickly Summarize large amounts of data.
3) Pivot tables in Power BI are used to turn rows into columns and unpivot columns into rows.

Definition of Pivot Charts-:

1) Pivot Charts are an in-built helps you summarize sale program tool that selected rows and spreadsheet columns of data in a spreadsheet.
2) It gives you the big picture of your raw data using various types of graphs and layouts.

Definition of Data Analysis-:

1) Data analysis is a transforming, and monitoring process of inspecting, valuable insights.
2) Data insights helps in making the required decision for the growth of the business & Company.
3) It involves subject data to obtain precise informed decisions or expand knowledge which can help businesses make on various subjects.
4) Data analysis has multiple facets and approaches.

Definition of Cube-:

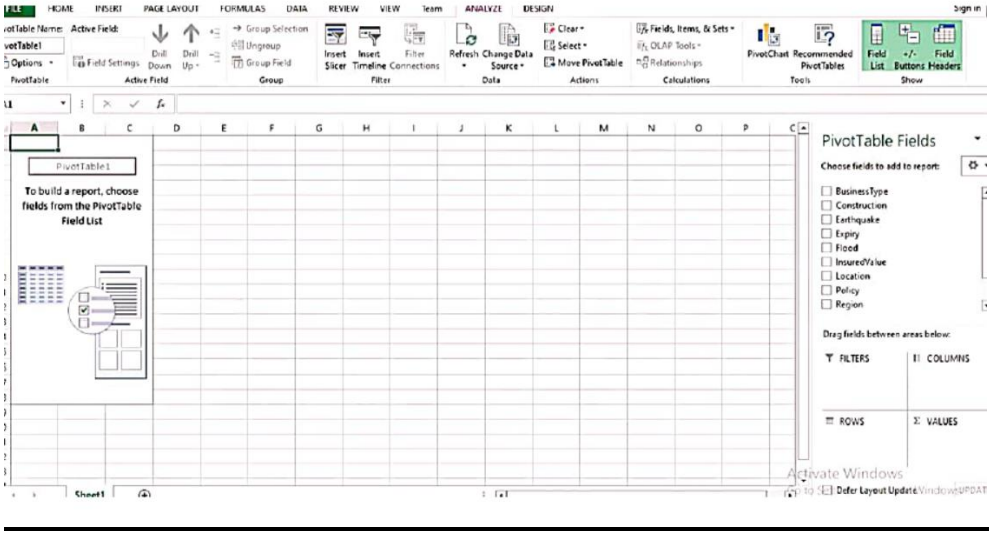
a) ***ROLAP-:*** It stands for Relational Online Analytical Processing and stores data in column, Rows.

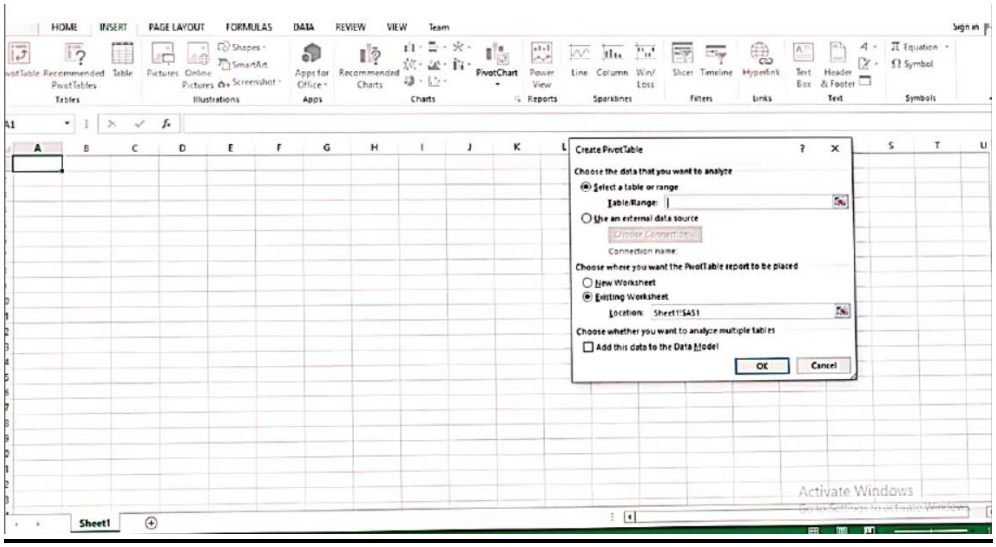
b) ***MOLAP-:*** It stands for Multidimensional online Analytical Processing and accesses data through various combinations.

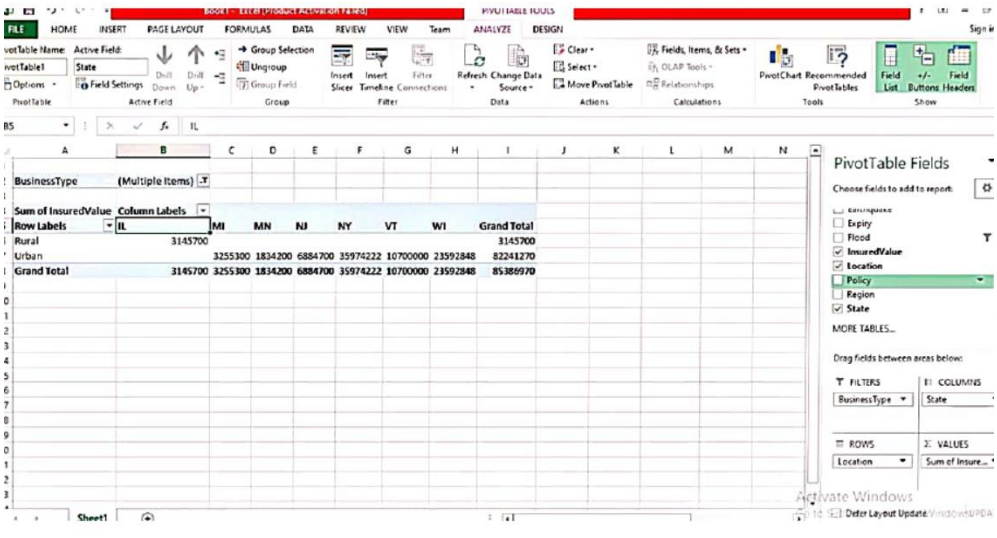
c) ***HOLAP-:*** It for Hybrid Online Analytical & is combination of both.

Output-:

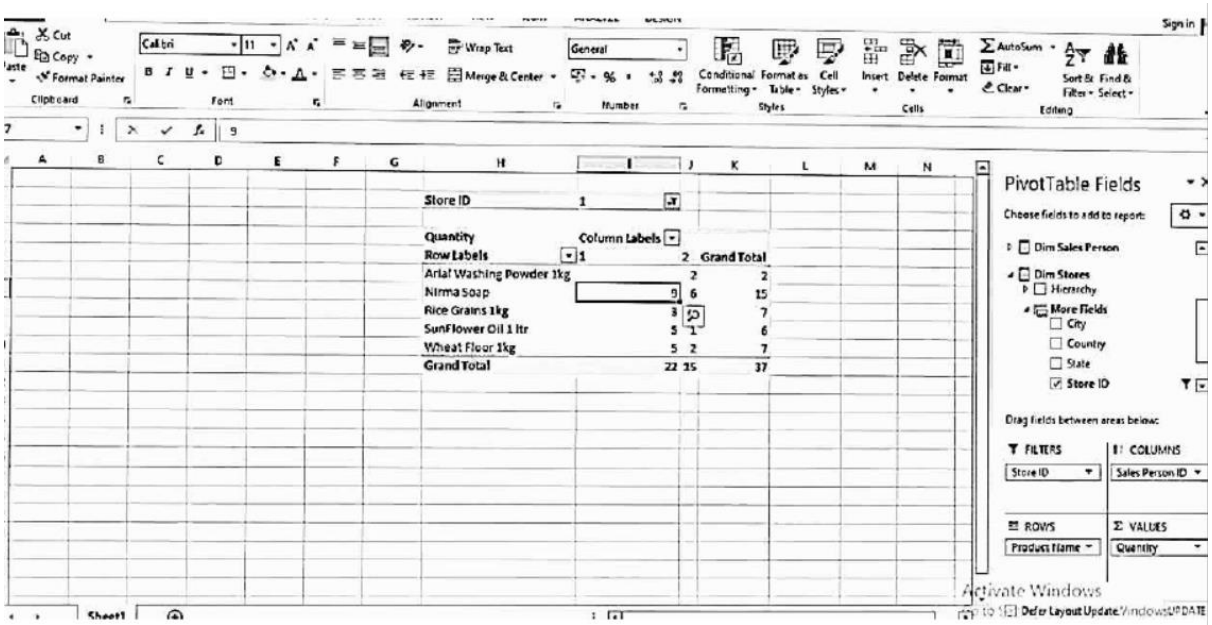
4a)


PivotTable Fields
Choose fields to add to report:
BusinessType
Construction
Earthquake
Expiry
Flood
InsuredValue
Location
Policy
Region
Drag fields between areas below:
FILTERS
COLUMNS
ROWS
VALUES
To build a report, choose fields from the PivotTable Field List
Sheet1

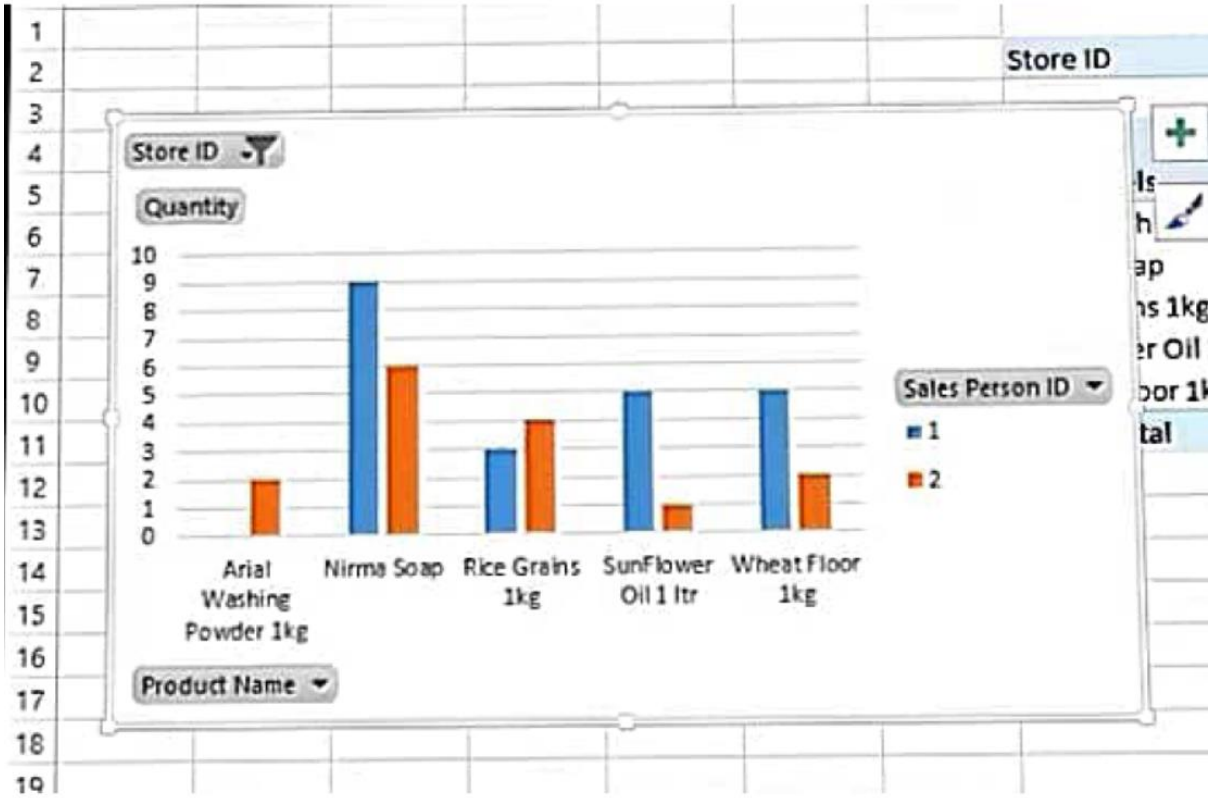

Create PivotTable
Choose the data that you want to analyze
Select a table or range
Table/Range:
Use an external data source
Choose Connection...
Connection name:
Choose where you want the PivotTable report to be placed
New Worksheet
Existing Worksheet
Location: Sheet1!A1
Choose whether you want to analyze multiple tables
Add this data to the Data Model
OK
Cancel
Sheet1


BusinessType (Multiple Items)
Sum of InsuredValue Column Labels
Row Labels IL MI MN NJ NY VT WI Grand Total
Rural 3145700 3145700
Urban 3255300 1834200 6884700 35974222 10700000 23592848 82241270
Grand Total 3145700 3255300 1834200 6884700 35974222 10700000 23592848 85386970
PivotTable Fields
Choose fields to add to report:
Expiry
Flood
InsuredValue
Location
Policy
Region
State
MORE TABLES...
Drag fields between areas below:
FILTERS
COLUMNS
BusinessType
State
ROWS
VALUES
Location
Sum of Insure...
Sheet1

4b)



Store ID	1		
Quantity	Column Labels		
Row Labels	1	2	Grand Total
Arial Washing Powder 1kg		2	2
Nirma Soap	9	6	15
Rice Grains 1kg	3	4	7
SunFlower Oil 1 ltr	5	1	6
Wheat Floor 1kg	5	2	7
Grand Total	22	15	37



Practical 5

Aim-: Apply the what-if Analysis for data visualization. Design and generate necessary reports based on the data warehouse data.

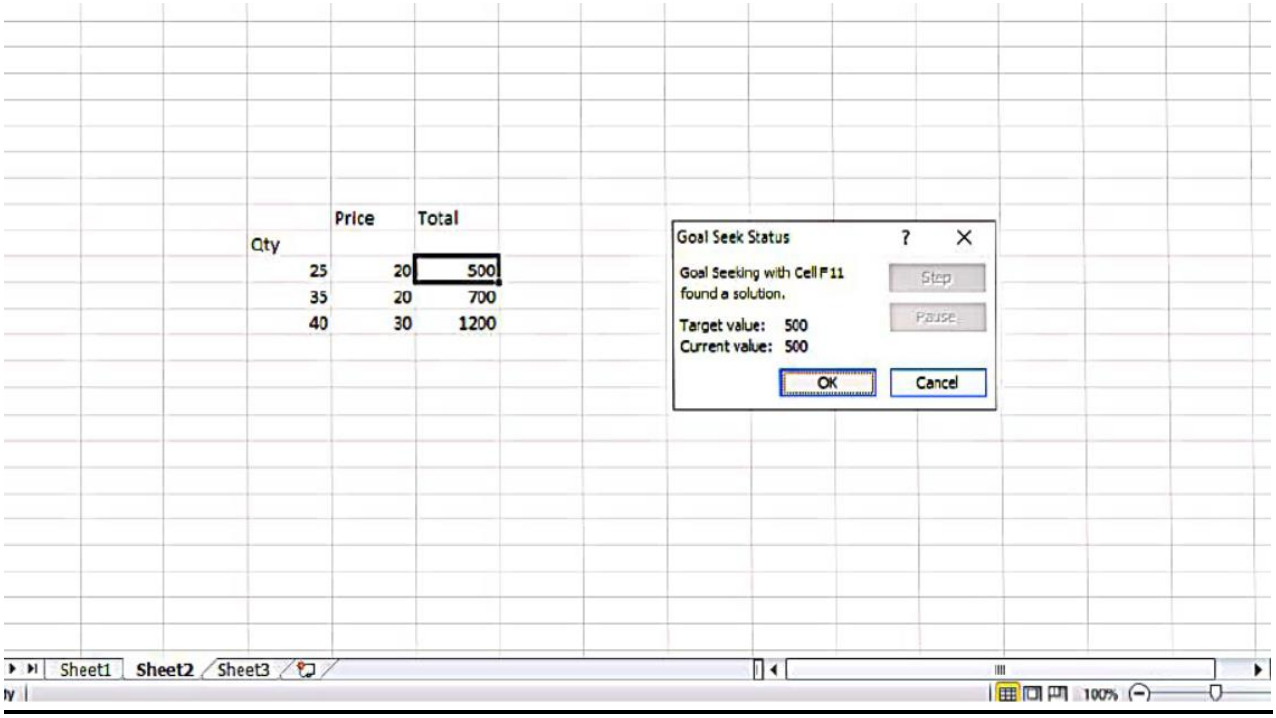
Definition of What-if Analysis:

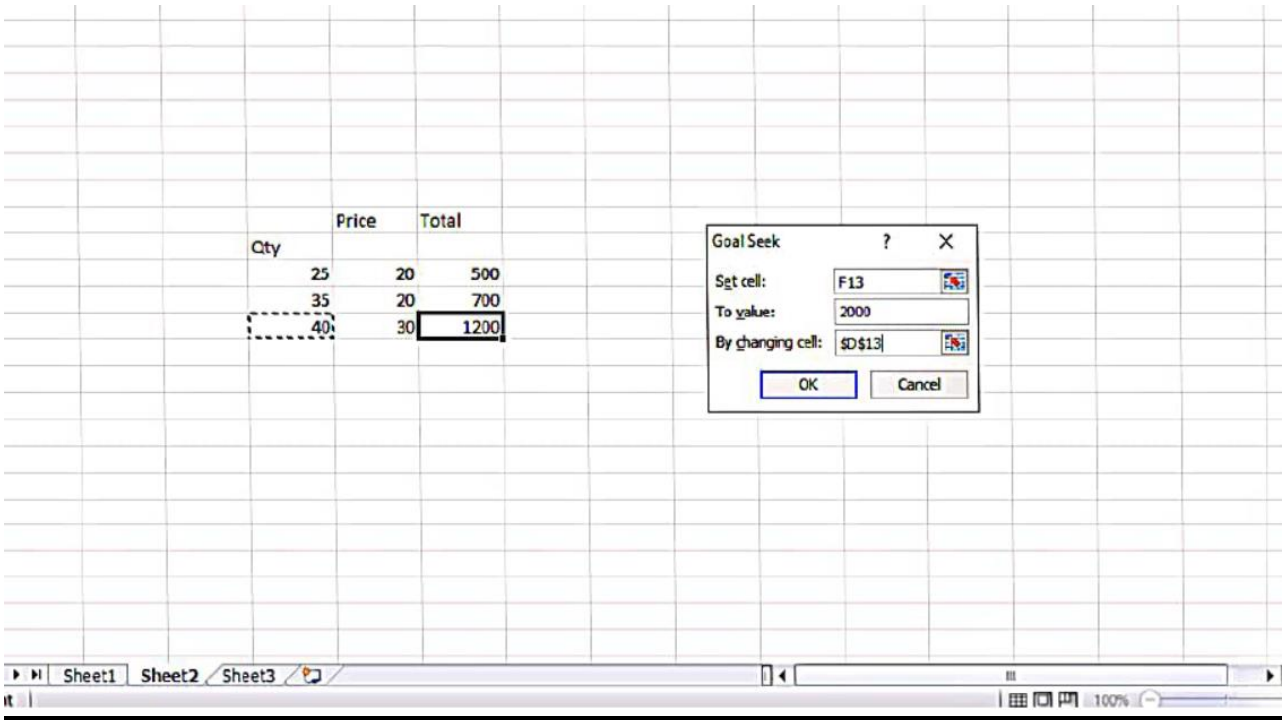
1) What-if analysis is often used to compare different scenarios and their potential outcomes based on changing conditions.
2) For using this feature, we can conditions. create What-it parameter and interact with the variable as a slicer.
3) Based on the input in the slicer, we can visualize and quantify changes in report data.

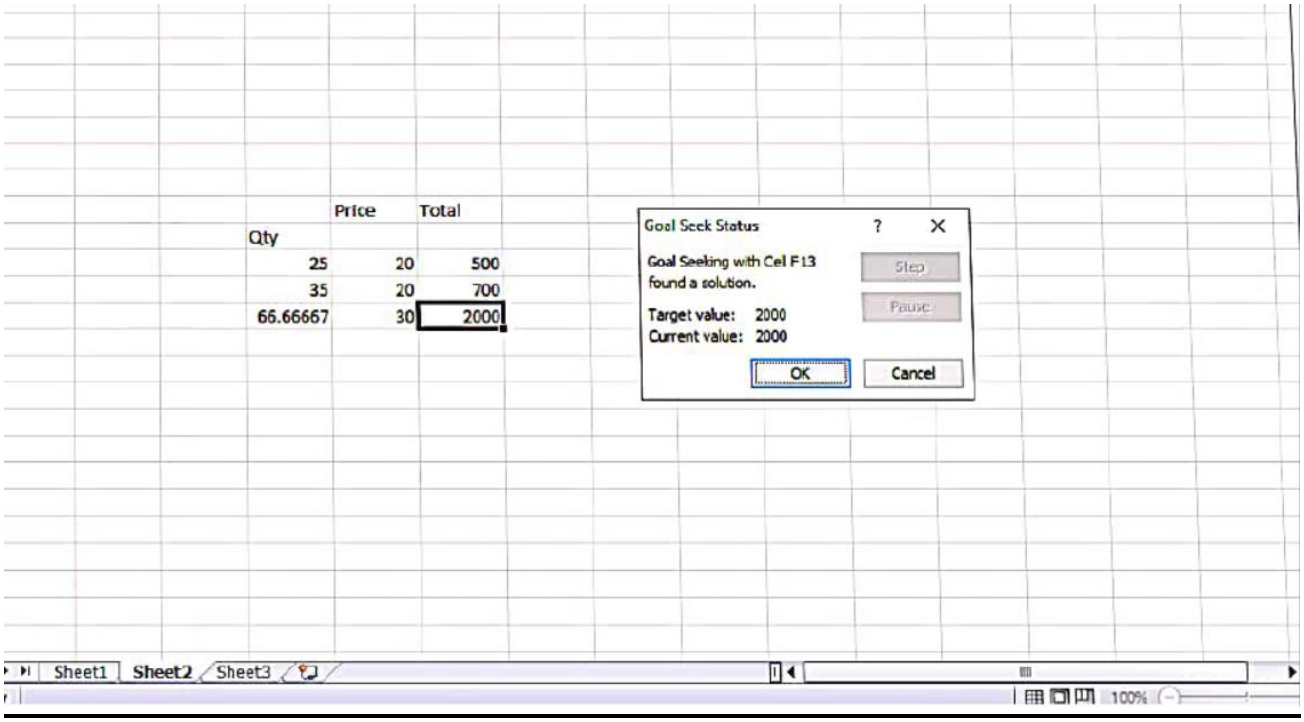
Definitions of Data Visualization-:

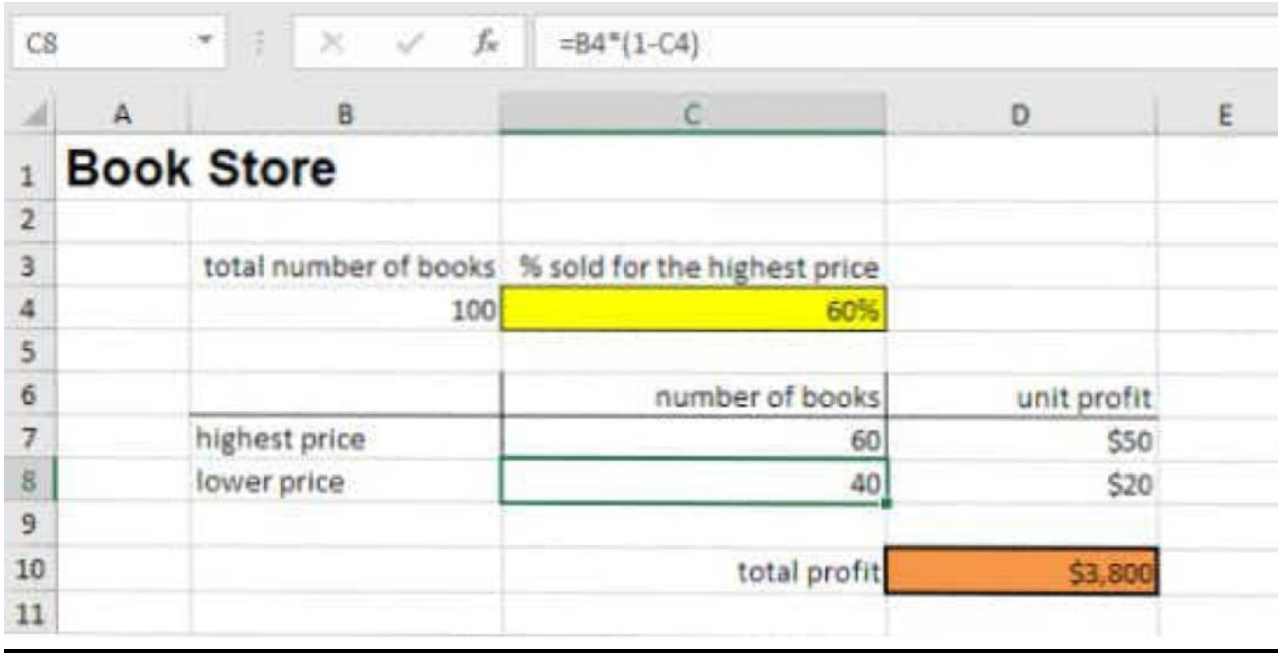
1) It is the representation of data through use of common graphics, such as charts, plots, infographic and even animation.
2) These visual displays of information communicate complex data relationships and data-driven to insight in a way easy to understand.
3) Visualizations are used to effectively present data and are the basic are the basic building your of any Business intelligence tool.

Output-:


Price
Total
Qty
25 20 500
35 20 700
40 30 1200
Goal Seek Status
Goal Seeking with Cell F11 found a solution.
Target value: 500
Current value: 500
Step
Pause
OK
Cancel
Sheet1
Sheet2
Sheet3
100%


Price
Total
Qty
25 20 500
35 20 700
40 30 1200
Goal Seek
Set cell: F13
To value: 2000
By changing cell: D13
OK
Cancel
Sheet1
Sheet2
Sheet3
100%


Price
Total
Qty
25 20 500
35 20 700
66.66667 30 2000
Goal Seek Status
Goal Seeking with Cell F13 found a solution.
Target value: 2000
Current value: 2000
Step
Pause
OK
Cancel
Sheet1
Sheet2
Sheet3
100%


C8
=B4*(1-C4)
Book Store
total number of books
% sold for the highest price
100
60%
number of books
unit profit
highest price
60
$50
lower price
40
$20
total profit
$3,800

Practical 6

Aim-: Perform the data classification using classification algorithm.

Definition of data classification-:

1) Data classification is a method available in Power B.I. which allows users to tag dashboards. that alert consumers to sensitivity in their data.
2) Data classifications are enabled and configured at the tenant level, Once established, a visible tag will be present on the dashboards.
3) Data classification is not a data security implementation.
4) It is a tag for dashboards and can only be applied on the service, not on Power BI Desktop.

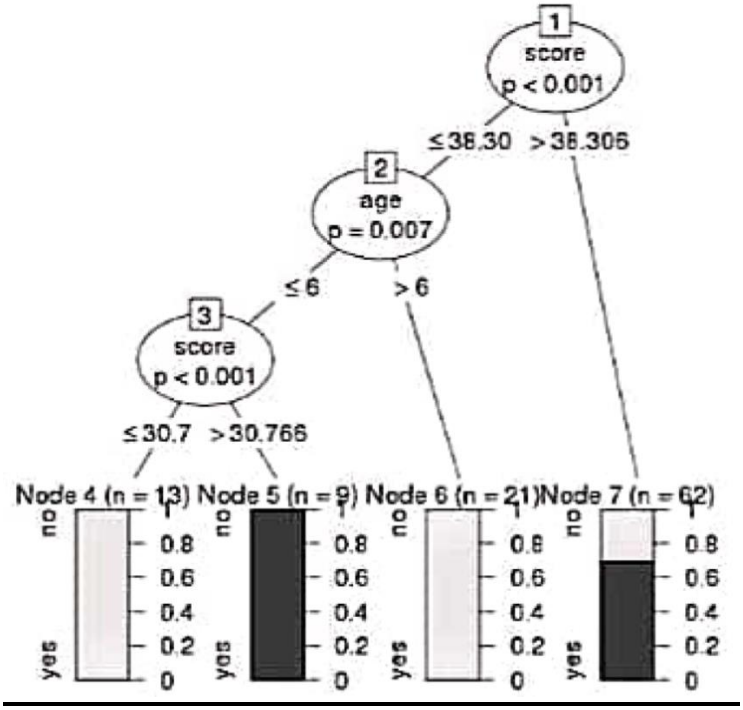
Definition of classification algorithm-:

1) The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training.

2) In Classification, a program learns from the given dataset or observations and then Classifies new observation into a no. of classes or groups.

Output-:

```
Loading required package: methods
Loading required package: grid
Loading required package: mvtnorm
Loading required package: modeltools
Loading required package: stats4
Loading required package: strucchange
Loading required package: zoo

Attaching package: 'zoo'

The following objects are masked from 'package:base':

as.Date, as.Date.numeric

Loading required package: sandwich
```




```
# Load the party package. It will automatically load other
# dependent packages.
library(party)

# Create the input data frame.
input.dat <- readingSkills[c(1:105),]

# Give the chart file a name.
png(file = "decision_tree.png")

# Create the tree.
  output.tree <- ctree(
  nativeSpeaker ~ age + shoeSize + score,
```

Practical 7

Aim-: Perform Data Analysis using Time series Analysis

Definition of Time Series Analysis:-

1) Time series is a series of data each data point is associated with a timestamp.
2) Simple example is the price of a stock in the stock market at different points of time on a given data.
3) The time data for the time series stored in an R object called time-series object.
4) R language uses many plot functions to create the time series dote manipulate data.
5) Following is the description of the parameters used-:

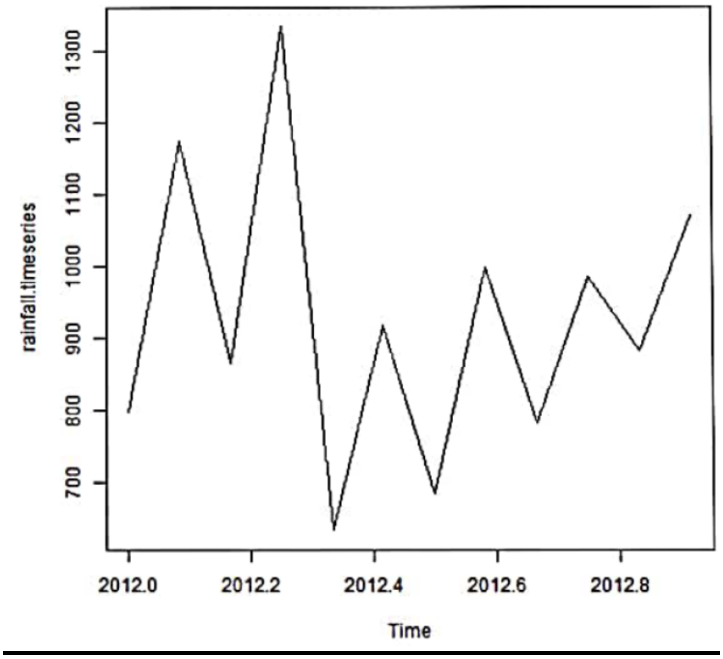
- Data is a vector containing the value used in timeseries
- Start specifies the start time for 1st observation in time series
- End specifies the end time for last observation in time series
- Frequency specifies the number of observation per unit.

6) The time series object is created by using ts() function.
   The basic syntax for ts() function in time series analysis.

[timeseries.object.name<-ts(data,start,end,frequency)]

Output-:

```
 Get the data points in form of a R vector.
ainfall <-
(799,1174.8,865.1,1334.6,635.4,918.5,685.5,998.6,784.2,985,882.8,1071)

 Convert it to a time series object.
ainfall.timeseries <- ts(rainfall,start = c(2012,1),frequency = 12)

 Print the timeseries data.
rint(rainfall.timeseries)

 Give the chart file a name.
ng(file = "rainfall.png")

 Plot a graph of the time series.
lot(rainfall.timeseries)

 Save the file.
ev.off()
```




```
Jan       Feb     Mar     Apr May Jun    Jul
          Aug     Sep

  2012  799.0  1174.8   865.1   1334.6
          Oct   635.4   918.5   685.5   998.6
                784.2
                  Nov     Dec
  2012  985.0   882.8  1071.0
```

Practical 8

Aim-: Perform the data clustering using clustering algorithm.

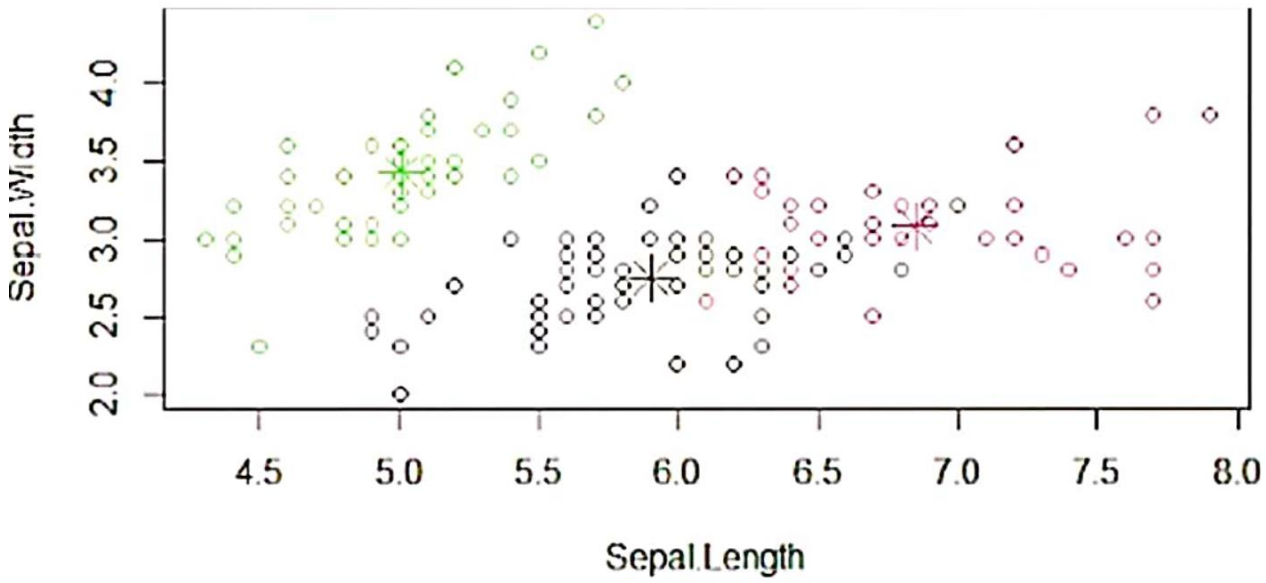
Definition of Data Clustering-:

1) Data clustering is an unsupervised machine learning algorithm that looks for patterns in data by dividing it into clusters.
2) These clusters are created such that the points are homogenous within the clusters and heterogenous across clusters.
3) It is commonly used in market segmentation, customer segmentation, image analysis, bioinformatics, data compression & computer graphics.

Definition of clustering algorithm-:

1) The Clustering algorithm provides two methods assigning for creating clusters and data points to the clusters.
2) The first, the K-means algorithm, is a hard clustering method. This means that data point can belong to only one clusters, and that single probability is calculated for the membership of each data point in that a cluster.

Output-:




```
#removing the species from data to cluster
newiris<-iris
newiris$Species<-NULL
#apply the k-means functions to newiris and store the result in kmeans.results
(kc<-kmeans(newiris,3))
print(kc)
#campare the clusters label
table(iris$Species,kc$cluster)
# plot the cluster and their centers
plot(newiris[c("Sepal.Length","Sepal.Width")],col=kc$cluster)
# plot the cluster center
points(kc$centers[,c("Sepal.Length","Sepal.Width")],col=1:3,pch=8,cex=2)
```

Practical 9

Aim-: Perform the linear regression on the given data warehouse.

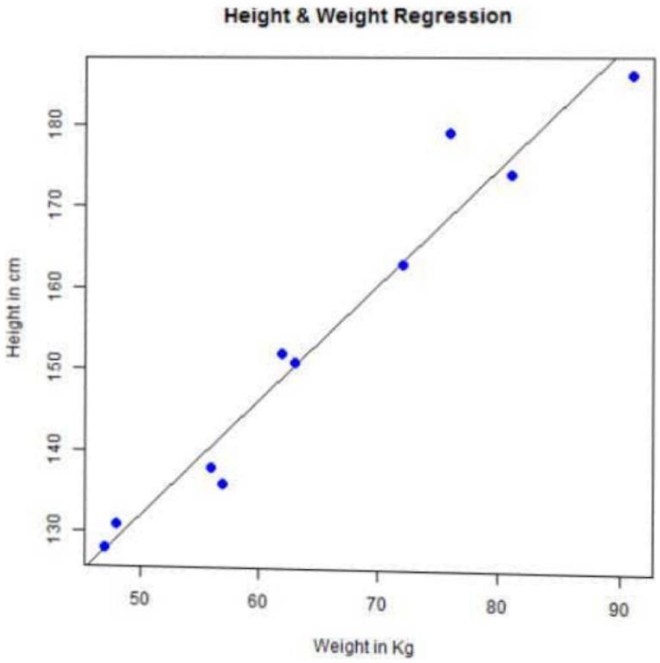
Definition of linear regression-:

1) linear regression is a statistical model apply to businesses to help forecast events based historical trend analysis.
2) Simple linear regression uses one variable, called the independent variable and other variable called as dependent variable. coordinate in a cartesian coordinate system and finds a linear function.

Definition of lm() function:-

1) The lm() function is used to fit linear models. to data frame in & R language.
2) It can be used to carry out regression, single stratum analysis of variance, and analysis of covariance to predict the value that is not in dataframe.

Output-:




```
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131
y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

# Apply the lm() function.
relation <- lm(y~x)
```

```
# The predictor vector.
x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131

# The resposne vector.
y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

# Apply the lm() function.
relation <- lm(y~x)

# Find weight of a person with height 170.
a <- data.frame(x = 170)
result <-  predict(relation,a)
print(result)
```

```
# Create the predictor and response variable.

x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131)

y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62, 48)

relation <- lm(y~x)


# Give the chart file a name.
png(file = "linearregression.png")


# Plot the chart.

plot(y,x,col = "blue",main = "Height & Weight Regression",
abline(lm(x~y)),cex = 1.3,pch = 16,xlab = "Weight in Kg",ylab = "Height in
cm")


# Save the file.
dev.off()
```

Practical 10

Aim-: Perform the logistic regression on the given data warehouse.

Definition of logistic regression:-

Logistic regression is a supervised machine learning algorithm that that accomplishes binary classification task by predicting the probability of an outcome, even binary for observation. The model delivers a or dichotomous outcome limited to two possible outcomes, yes/no, o/1, or true/false.

Dataset used:- mt cars

1) mtcars (motor trend car road sets)
2) There is a popular built-in data set in R called "mtcars" which is retrieved from the 1974 motor Trend Us Magazine.
3) The mtcars dataset is a built-in dataset in R. that contains measurements on 11 different attributes for 32 different cars.

GLM(Generalized linear Model):-

1) The glm (function in R can be used to generalized linear models.
2) This function is particularly useful for things for fitting logistic regression models, Poison regression models and other complex models.

Output-:

```
> input=mtcars[,c("am","cyl","hp","wt")]
> print(head(input))
                  am cyl  hp    wt
Mazda RX4          1   6 110 2.620
Mazda RX4 Wag      1   6 110 2.875
Datsun 710         1   4  93 2.320
Hornet 4 Drive     0   6 110 3.215
Hornet Sportabout  0   8 175 3.440
Valiant            0   6 105 3.460
> input=mtcars[,c("am","cyl","hp","wt")]
> am.data=glm(formula=am~cyl|hp|wt,data=input,family=binomial)
> summary(am.data)

Call:
glm(formula = am ~ cyl | hp | wt, family = binomial, data = input

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.021  -1.021  -1.021   1.342   1.342

Coefficients: (1 not defined because of singularities)
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)       -0.3795     0.3599  -1.054    0.292
cyl | hp | wtTRUE      NA         NA      NA       NA

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.23  on 31  degrees of freedom
Residual deviance: 43.23  on 31  degrees of freedom
AIC: 45.23

Number of Fisher Scoring iterations: 4
```

```
> mtcars
                   mpg cyl  disp  hp drat    wt  qsec vs am gear carb
Mazda RX4         21.0   6 160.0 110 3.90 2.620 16.46  0  1    4    4
Mazda RX4 Wag     21.0   6 160.0 110 3.90 2.875 17.02  0  1    4    4
Datsun 710        22.8   4 108.0  93 3.85 2.320 18.61  1  1    4    1
Hornet 4 Drive    21.4   6 258.0 110 3.08 3.215 19.44  1  0    3    1
Hornet Sportabout 18.7   8 360.0 175 3.15 3.440 17.02  0  0    3    2
Valiant           18.1   6 225.0 105 2.76 3.460 20.22  1  0    3    1
Duster 360        14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
Merc 240D         24.4   4 146.7  62 3.69 3.190 20.00  1  0    4    2
Merc 230          22.8   4 140.8  95 3.92 3.150 22.90  1  0    4    2
Merc 280          19.2   6 167.6 123 3.92 3.440 18.30  1  0    4    4
Merc 280C         17.8   6 167.6 123 3.92 3.440 18.90  1  0    4    4
Merc 450SE        16.4   8 275.8 180 3.07 4.070 17.40  0  0    3    3
Merc 450SL        17.3   8 275.8 180 3.07 3.730 17.60  0  0    3    3
Merc 450SLC       15.2   8 275.8 180 3.07 3.780 18.00  0  0    3    3
Cadillac Fleetwood 10.4  8 472.0 205 2.93 5.250 17.98  0  0    3    4
```