

PRACTICAL 7

AIM:- Perform regression on the data set using R programming.

Solution:

Theory:

Regression :-

- Regression refers to a type of supervised machine learning technique that is used to predict any continuous-valued attribute.
- Regression involves the technique of fitting a straight line or a curve on numerous data points. It happens in such a way that the distance between the data points and curve comes out to be the lowest.

Regression Analysis in R:-

- Regression analysis is a group of statistical processes used in R programming and statistics to determine the relationship between dataset variables. Generally, regression analysis is used to determine the relationship between the dependent and independent variables of the dataset.
- Regression analysis helps to understand how dependent variables change when one of the independent variables is changes and other independent variables are kept constant. This helps in building a regression model and further, helps in forecasting the values with respect to a change in one of the independent variables.
- On the basis of types of dependent variables, a number of independent variables, and the shape of the regression line, there are 4 types of regression analysis techniques i.e., Linear Regression, Logistic Regression, Multinomial Logistic Regression and Ordinal Logistic Regression.

Code :

```
data <- read.csv("D:/Hemil.csv")
View(data)
nrow(data)
ncol(data)
X <- data$weight
Y <- data$height
relation <- lm(Y~X)
plot(X,Y)
plot(X,Y,col = "blue",main = "X & Y Regression",cex = 1,pch = 16,xlab = "X",ylab = "Y")
abline(relation)
summary(relation)
```

OUTPUT:

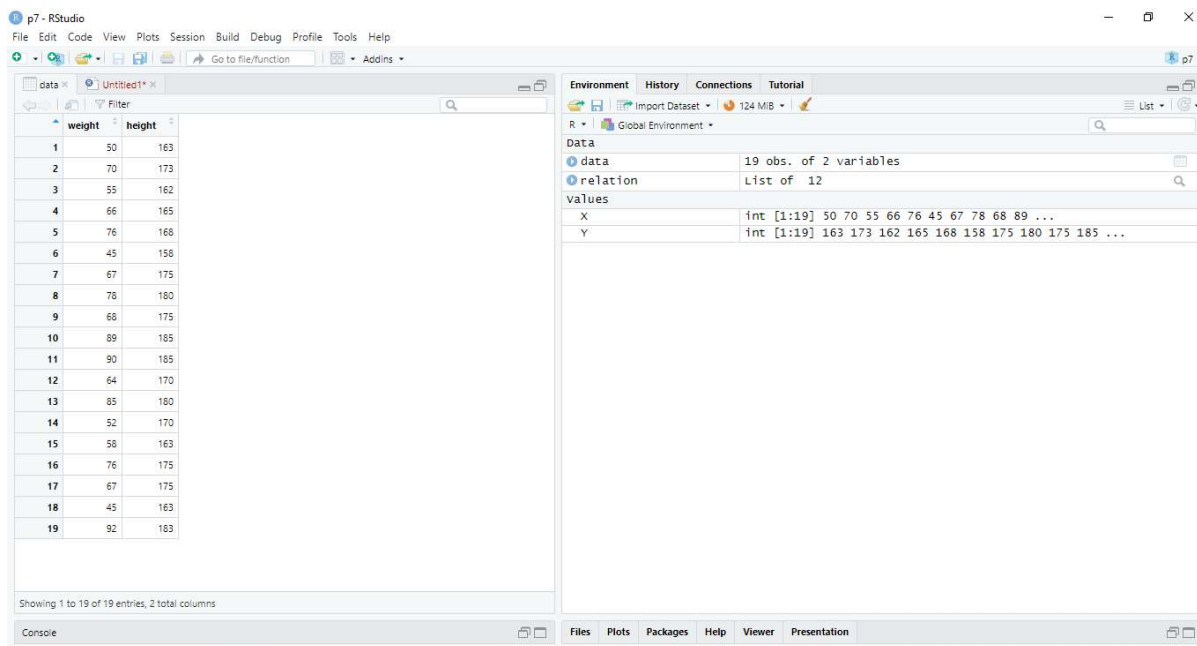


Figure 1 Display the relationship between dataset variables

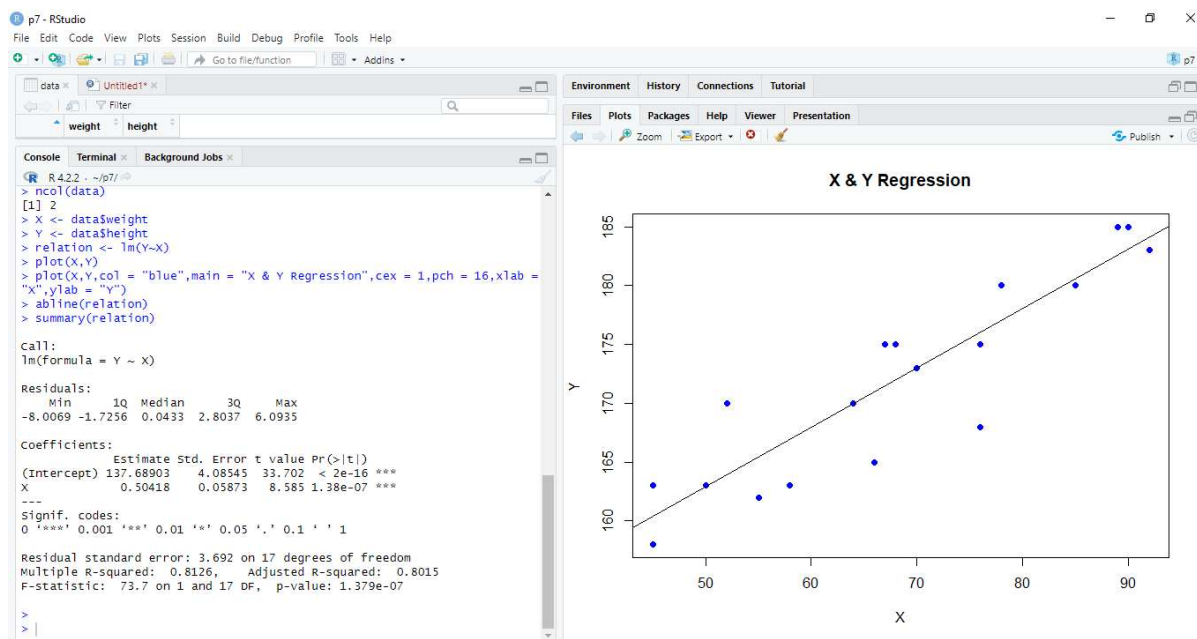


Figure 2 X&Y Regression Plots Diagram

PRACTICAL-1

AIM:- Design and Create cube by identifying measures and dimensions for Star Schema, Snowflake schema and fact Constellation Schema.

Solution:

Software Used: Analysis services- SQL Server-2008

Theory:

1) Creating a New Analysis Services Project

To create a new Analysis Services project, use the New Project dialog box in BIDSTo create a new Analysis Services project, follow these steps:

1. Select Microsoft SQL Server 2008 ⇒ SQL Server Business Intelligence. Development Studio from the Programs menu to launch Business Intelligence Development Studio.
2. Select File ⇒ New Project. ⇒
3. In the New Project dialog box, select the Business Intelligence Projects project type.
4. Select the Analysis Services Project template.
5. Name the new project AdventureWorksCube1 and select a convenient location to save it.
6. Click OK to create the new project.

(1) Star Schema:

- The star schema architecture is the simplest data warehouse schema.
- It is called a star schema because the diagram resembles a star, with points radiating from a center. The center of the star consists of fact table and the points of the star are the dimension tables. Usually the fact tables in a star schema are in third normal form(3NF) whereas dimensional tables are de-normalized.
- Despite the fact that the star schema is the simplest architecture, it is most commonly used nowadays and is recommended by Oracle.

Fact Tables:-

- A fact table typically has two types of columns: foreign keys to dimension tables and measures those that contain numeric facts.
- A fact table can contain fact's data on detail or aggregated level.

Dimension Tables:-

- A dimension is a structure usually composed of one or more hierarchies that categorizes data. If a dimension hasn't got a hierarchies and levels it is called flat dimension or list.
- The primary keys of each of the dimension tables are part of the composite primary key of the fact table. Dimensional attributes help to describe the dimensional value.
- They are normally descriptive, textual values. Dimension tables are generally small in size then fact table.

The main characteristics of star schema:

- Simple structure -> easy to understand schema
- Great query effectiveness -> small number of tables to join
- Relatively long time of loading data into dimension tables -> de-normalization, redundancy data caused that size of the table could be large.
- The most commonly used in the data warehouse implementations -> widely supported by a large number of business intelligence tools.

(1) Snowflake Schema:

- A snowflake schema is a logical arrangement of tables in a multidimensional database such that the entity relationship diagram resembles a snowflake shape.
- The snowflake schema is represented by centralized fact tables which are connected to multiple dimensions.
- Star and snowflake schemas are most commonly found in dimensional data warehouses and data marts where speed of data retrieval is more important than the efficiency of data manipulations. As such, the tables in these schemas are not normalized much, and are frequently designed at a level of normalization short of third normal form.
- Normalization splits up data to avoid redundancy (duplication) by moving commonly repeating groups of data into new tables. Normalization therefore tends to increase the number of tables that need to be joined in order to perform a given query, but reduces the space required to hold the data and the number of places where it needs to be updated if the data changes

(2) Fact constellation Schema:

- Fact constellation is a measure of online analytical processing, which is a collection of multiple fact tables sharing dimension tables, viewed as a collection of stars.
- This is an improvement over Star schema.
- For each star schema it is possible to construct fact constellation schema (for example by splitting the original star schema into more star schemes each of them describes facts on another level of dimension hierarchies).
- The fact constellation architecture contains multiple fact tables that share many dimension tables.
- Moreover, dimension tables are still large. **tep 1:** shows the Solution Explorer window of the new project, ready to be populated with objects and Defining a data source.

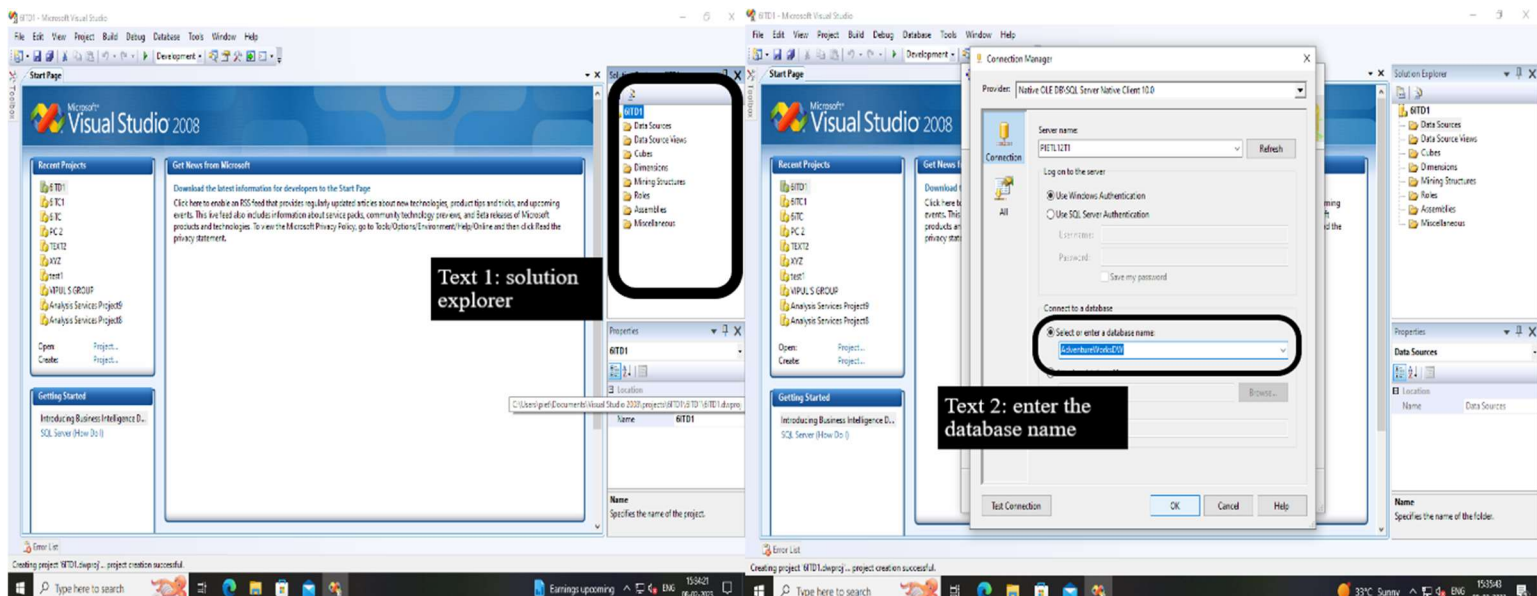


Figure 1 New Analysis Services project& Setting up a connection

step 2: create data source view Wizard select the tables and show the analysis service.

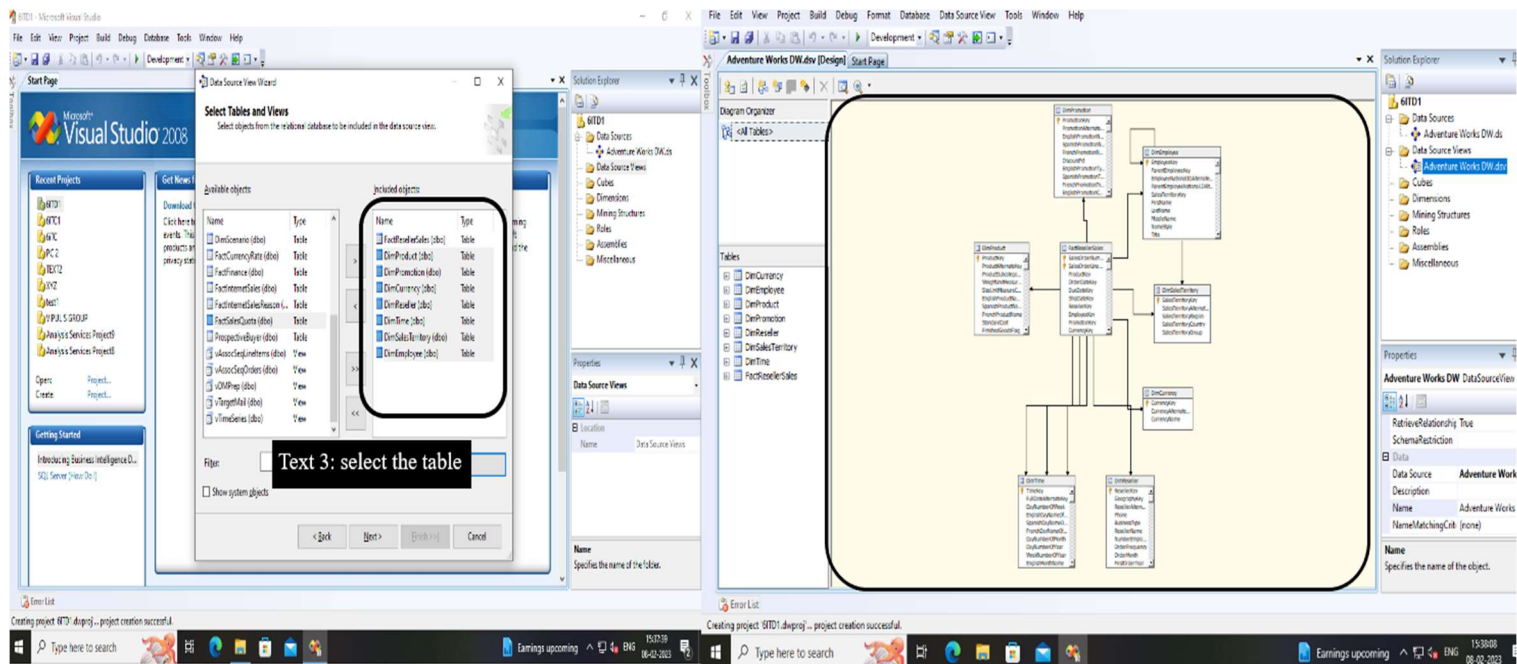


Figure 2 Selecting tables for the data source view & The Sales data source view

step 3: Deploying and Processing a Cube.

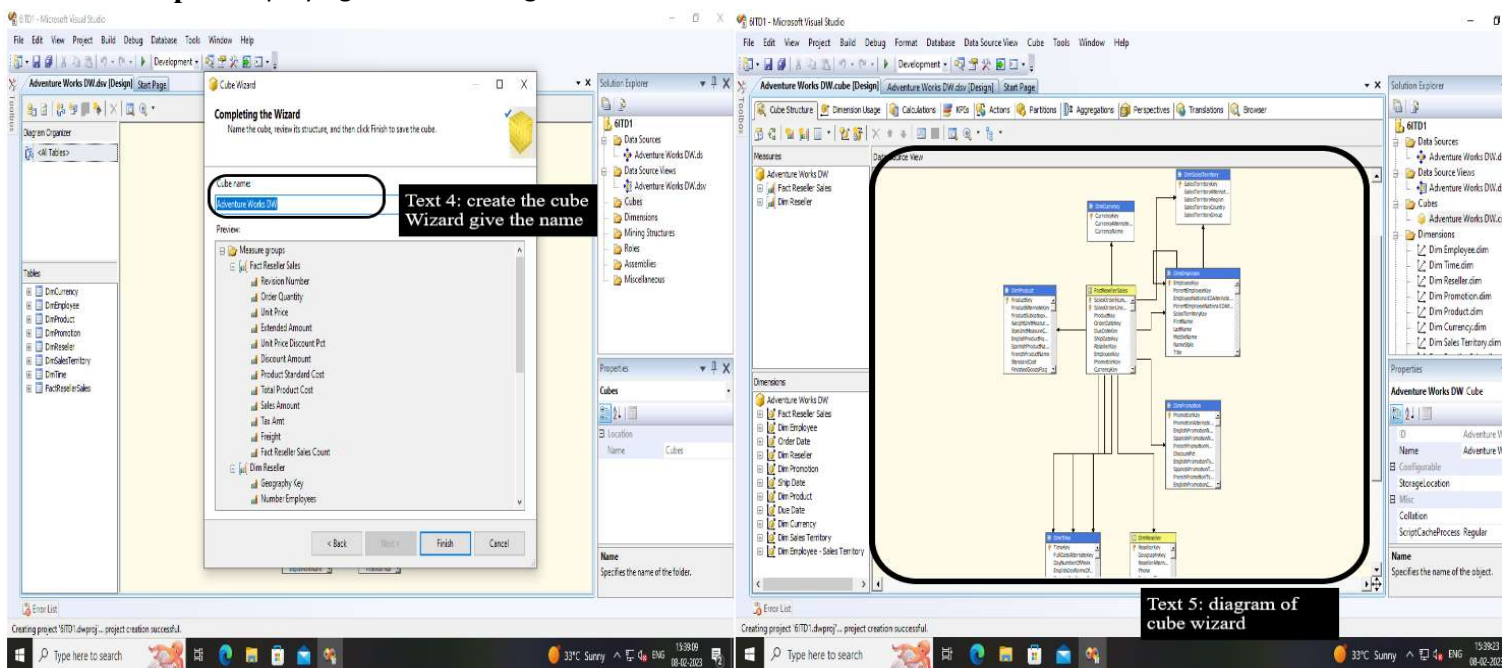


Figure 3 cube wizard diagram

step 4: Create Star Schema, Snowflake schema and fact Constellation Schema.



Figure 4 Diagram for schema