

**Group Members:** Harshal Patil

**Student Number:** 10374924

**Course Title:** MSc in Information Systems with Computing

**Lecturer Name:**  Dr. Shazia A Afzal

**Module/Subject Title:** B9IS106 Data and Data Analytics

**Assignment Title:** Design and implementation of a hybrid (Relational/XML) database system of your choice.

**Requirements for developing the data warehouse:**

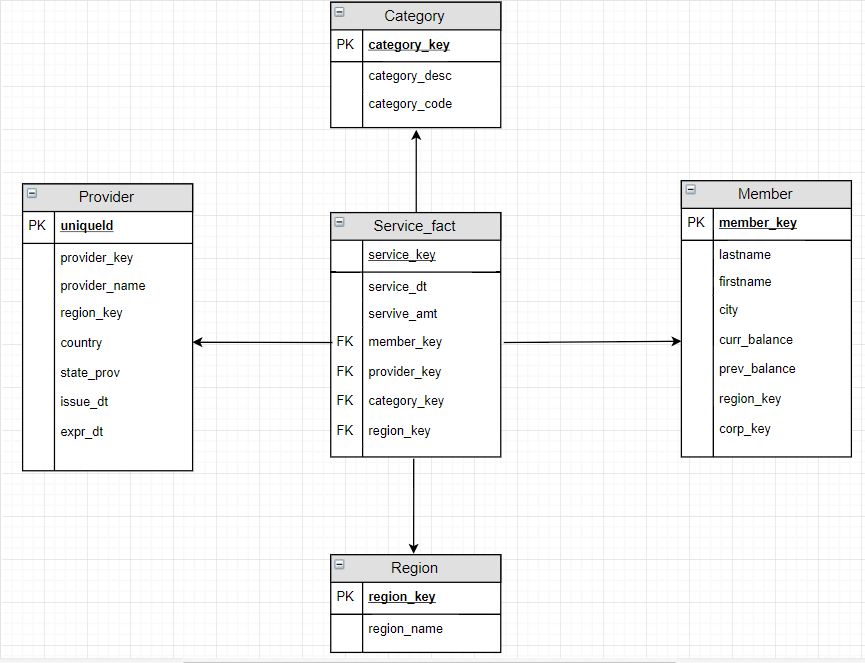
* The database that I have selected is Credit 2008 database from sqlskills in which there is charge for the service provided in each region according to the categories. There are members who uses the service that are provided by providers and are charged accordingly. I have used SQL Server 2014 Management studio to design and build data warehouse.
* I have downloaded .bak file database and restored in SQL server.

Link: <https://www.sqlskills.com/sql-server-resources/sql-server-demos/>

**Services Data Warehouse:**

* According to the source database, I have created a star schema including provider, category, member and region as dimension tables.
* Then I created a fact table with the measures service\_dt and service\_amount and foreign keys of all the dimension tables.

**Star Schema:**



**All the codes for implementing the Data Warehouse and for extraction:**

* **Creating Data Warehouse table**

use Book\_Sales

create table Authors(aut\_key varchar(20) not null Primary key, au\_id varchar(11), au\_lname varchar(40), au\_fname varchar(20), phone char(12), address varchar(40),

city varchar(20), state char(2), zip char(5), contract bit)

select \* from Authors

use Book\_Sales

create table Publishers(pub\_key varchar(20) Not null Primary Key, pub\_id char(4), pub\_name varchar(40), city varchar(20), state char(2), country varchar(30))

select \* from Publishers

use Book\_Sales

create table Stores(stor\_key varchar(20) Not null Primary Key, stor\_id char(4), stor\_name varchar(40), stor\_address varchar(40), city varchar(20),

state char(2), zip char(5), discounttype varchar(40), lowqty smallint, highqty smallint, discount decimal(4,2))

select \* from Stores

use Book\_Sales

create table Titles(title\_key varchar(20) not null Primary key, title\_id varchar(6), title varchar(80), type char(12), lorange int, hirange int, royalty int,

price money, pubdate datetime)

Select \* from Titles

create table Fact\_Sales(ord\_num varchar(20), ord\_date datetime, qty smallint, payterms varchar(12),

stor\_key varchar(20), Foreign key (stor\_key) References Stores(stor\_key),

pub\_key varchar(20), Foreign key (pub\_key) References Publishers(pub\_key),

aut\_key varchar(20), Foreign Key (aut\_key) References Authors(aut\_key),

title\_key varchar(20), Foreign Key (title\_key) References Titles(title\_key)

select \* from Fact\_Sales

* **Inserting data into tables:**

Service Table:

USE [Services\_DW]

GO

INSERT INTO [dbo].[Category]

([category\_key]

,[category\_desc]

,[category\_code])

SELECT [category\_no]

,[category\_desc]

,[category\_code]

FROM [Credit].[dbo].[category]

Member Table:

USE [Services\_DW]

GO

INSERT INTO [dbo].[Member]

([member\_key]

,[lastname]

,[firstname]

,[city]

,[curr\_balance]

,[prev\_balance]

,[region\_key]

,[corp\_key])

SELECT [member\_no]

,[lastname]

,[firstname]

,[city]

,[curr\_balance]

,[prev\_balance]

,[region\_no]

,[corp\_no]

FROM [Credit].[dbo].[member]

GO

Provider table:

USE [Services\_DW]

GO

INSERT INTO [dbo].[Provider]

([provider\_key]

,[provider\_name]

,[region\_key]

,[country]

,[state\_prov]

,[issue\_dt]

,[expr\_dt])

SELECT [provider\_no]

,[provider\_name]

,[region\_no]

,[country]

,[state\_prov]

,[issue\_dt]

,[expr\_dt]

FROM [Credit].[dbo].[provider]

Region table:

USE [Services\_DW]

GO

INSERT INTO [dbo].[Region]

([region\_key]

,[region\_name])

SELECT [region\_no]

,[region\_name]

FROM [Credit].[dbo].[region]

Fact table:

USE [Services\_DW]

GO

INSERT INTO [dbo].[Service\_fact]

([service\_key]

,[service\_dt]

,[servive\_amt]

,[member\_key]

,[provider\_key]

,[category\_key]

,[region\_key])

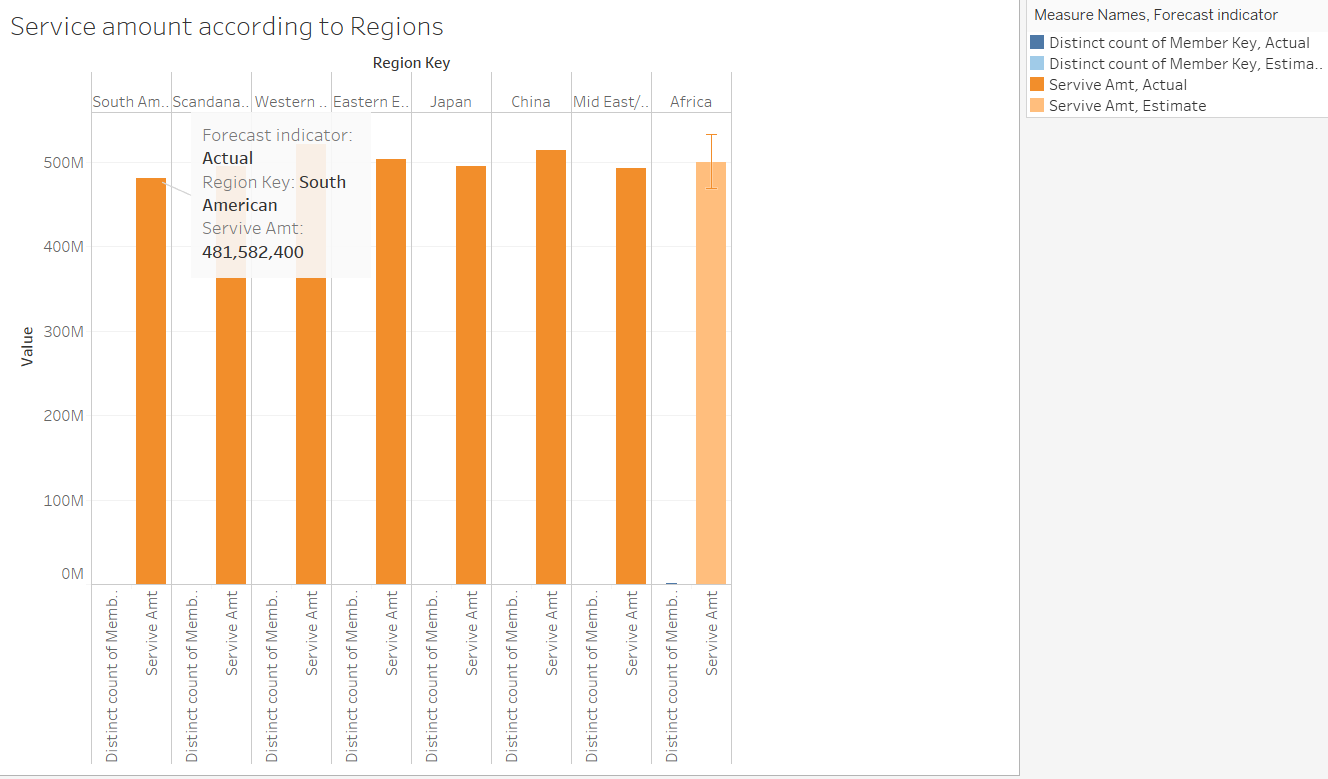
SELECT charge\_no, charge\_dt, charge\_amt, C.member\_no, provider\_no, category\_no, Credit.dbo.region.region\_no

FROM Credit.dbo.charge as C INNER JOIN

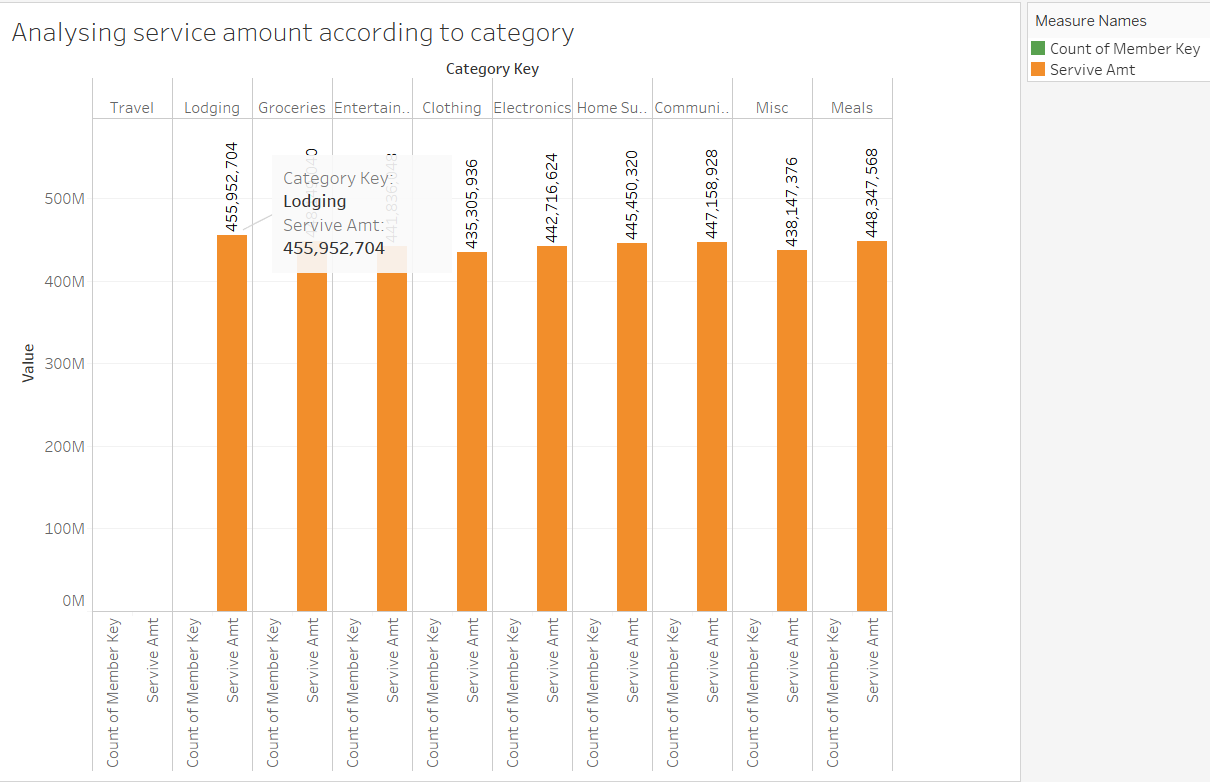
Credit.dbo.member ON C.member\_no = Credit.dbo.member.member\_no INNER JOIN

Credit.dbo.region ON Credit.dbo.region.region\_no = Credit.dbo.region.region\_no

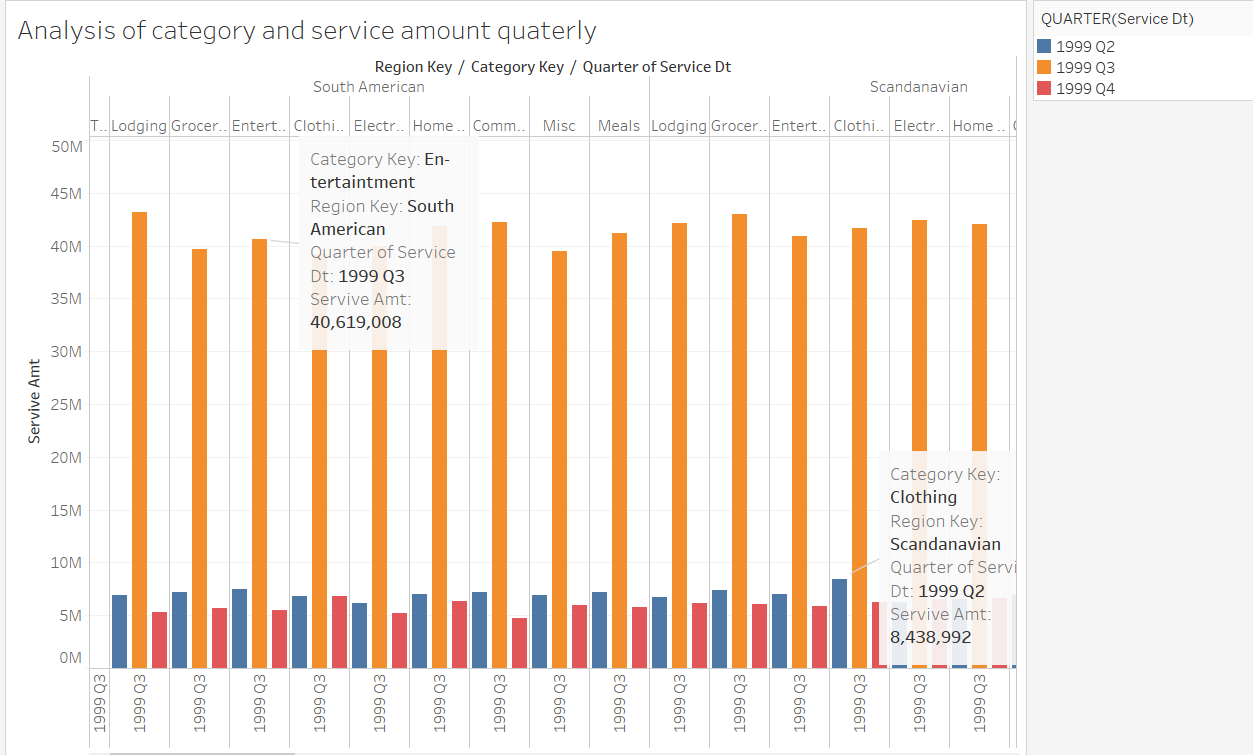
* **Tableau screenshot:**



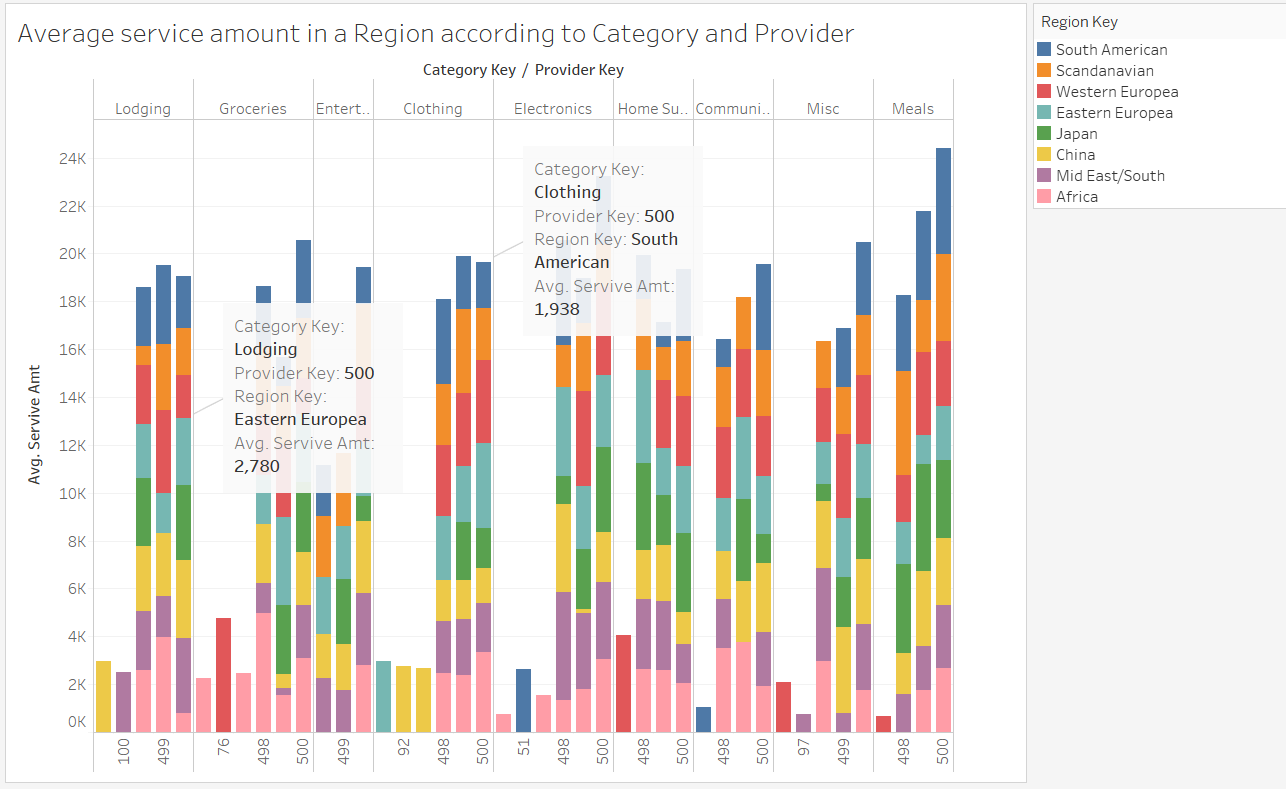
* Here I have analysed service amount in a particular region and have forecasted as well.



* In the above screenshot, I have analysed service amount according to category.



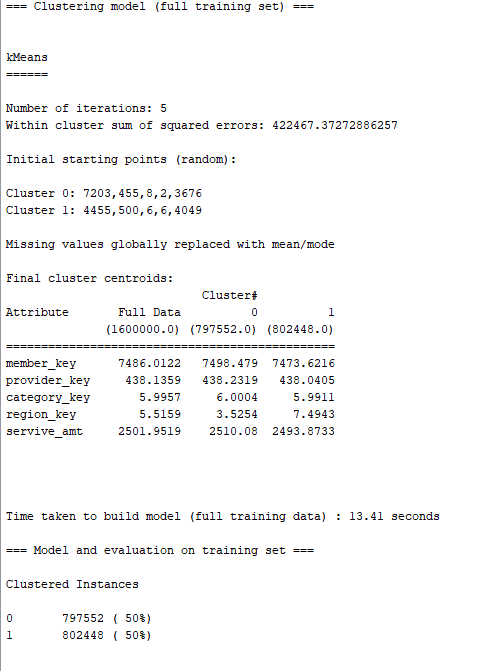
* Here there is analysis of category and service amount quarterly.



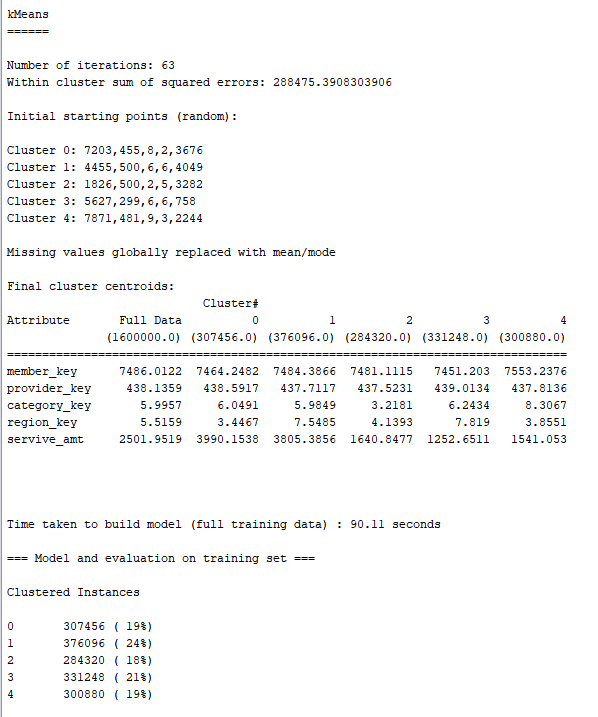
* I have analysed service amount in an region according to category and provider.

**Explanation of Data mining project:**

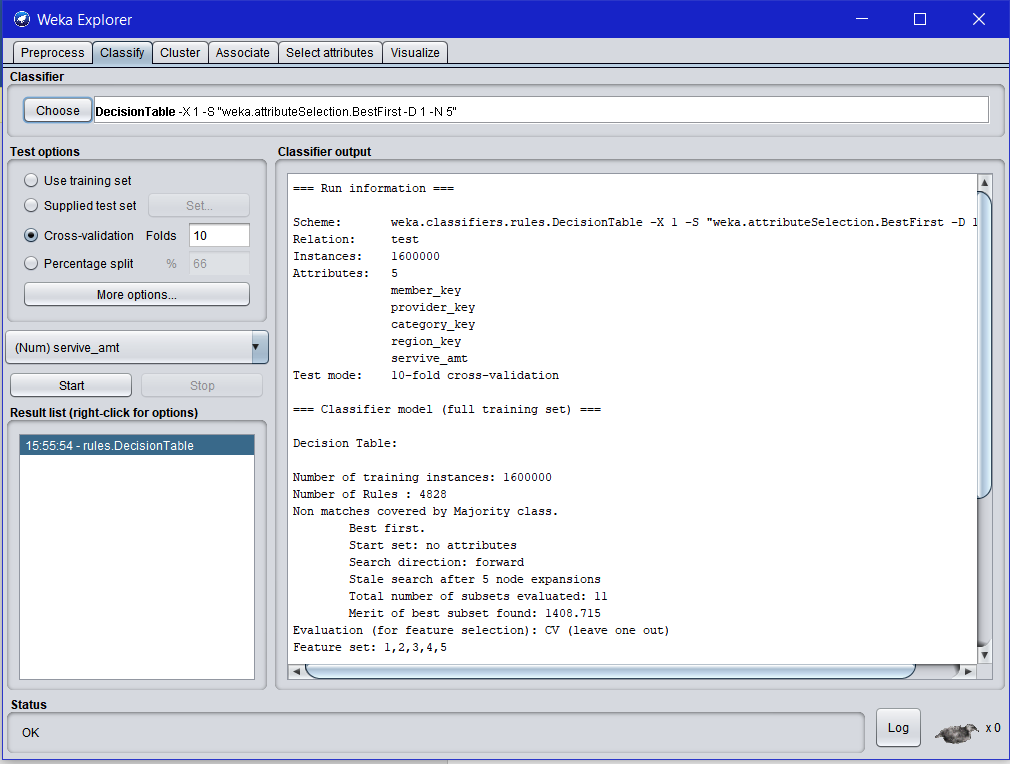
* kMeans



* Here I have selected 2 clusters and the clustered instances are 50%.



* In this I have created 5 clusters and got the above results. Creating 5 clusters is more useful to predict the outcome. Cluster 1 is most accurate among all the clusters.
* Decision Tree
* The main thing in decision tree is basically creating training data and then we can build predictive model which is mapped in the form of tree structure.



* Visualization of all the attributes

