**USER MANUAL**

**Technologies Used**

* Front-End: Visual Studio C# .NET with ASP .NET.
* Back-End: MS SQL SERVER 2012.
* Pre-processing: MACRO ENABLED Microsoft excel.
* Documentation : Latex and Microsoft Word.

**Detailed Implementation**

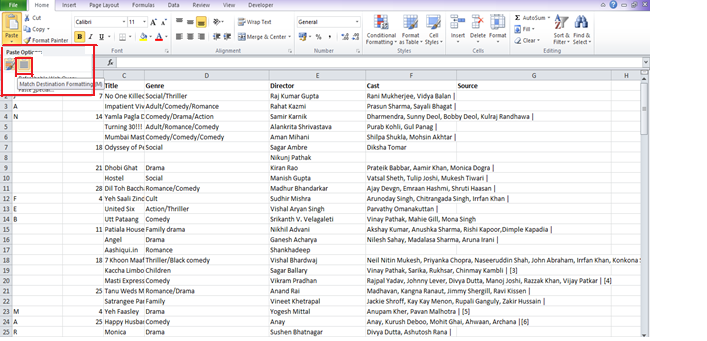
1. **DATA COLLECTION:**

* Data required for our project is collected from websites such as Wikipedia, box office.com, Bollywood hungama and koimoi.com.
* Some links from where the data is collected are as follows:
* <https://en.wikipedia.org/wiki/List_of_Bollywood_films_of_2015>
* <http://www.koimoi.com/box-office-bollywood-films-of-2013/>
* <http://boxofficeindia.com/Collections/net_box_office/delhi_up/#.Vg5fpS6qqkp>
* <http://www.koimoi.com/box-office-india-bollywood-100-crore-club/>

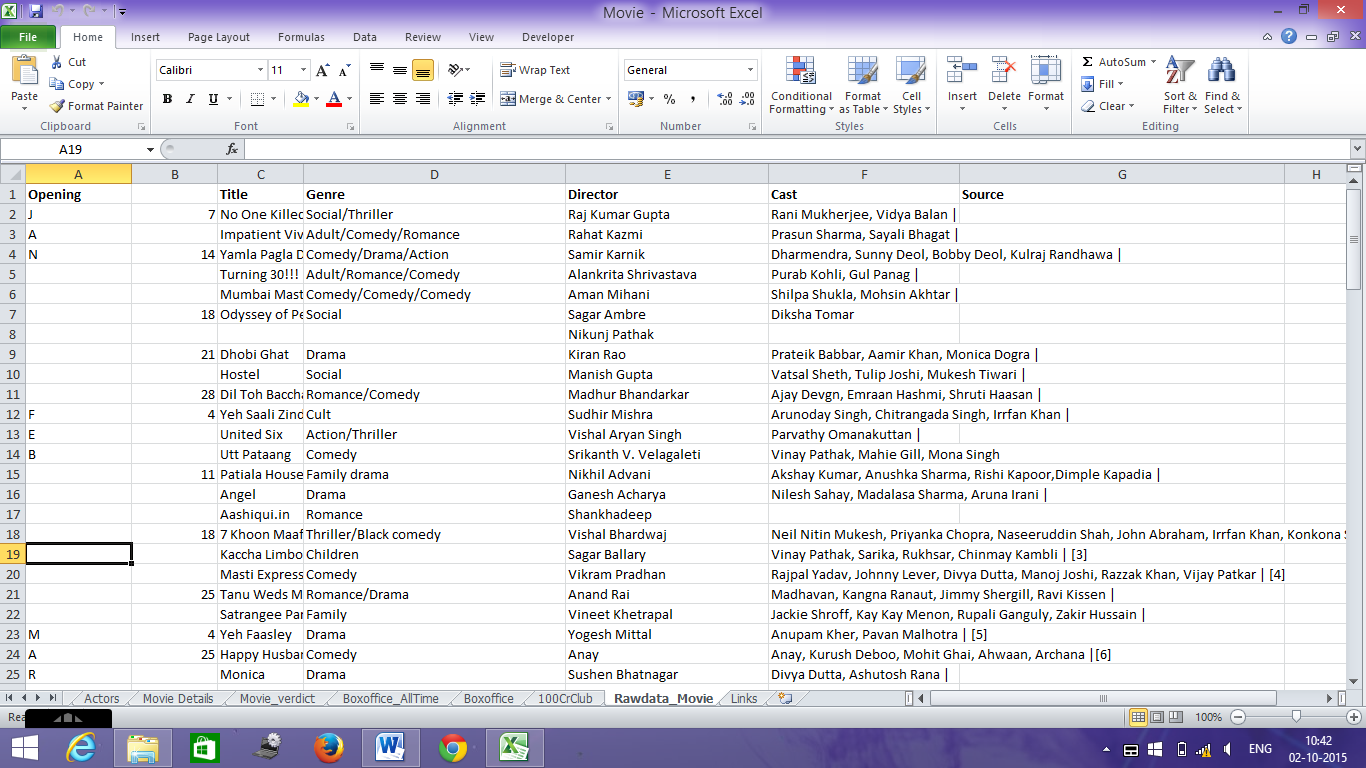
1. The example of data available on https://en.wikipedia.org/wiki/List\_of\_Bollywood\_films\_of\_2013 is as follows:



1. After selecting, the data it is pasted in excel sheet as follows:



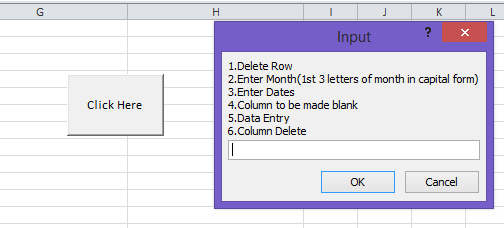
RAW DATA:



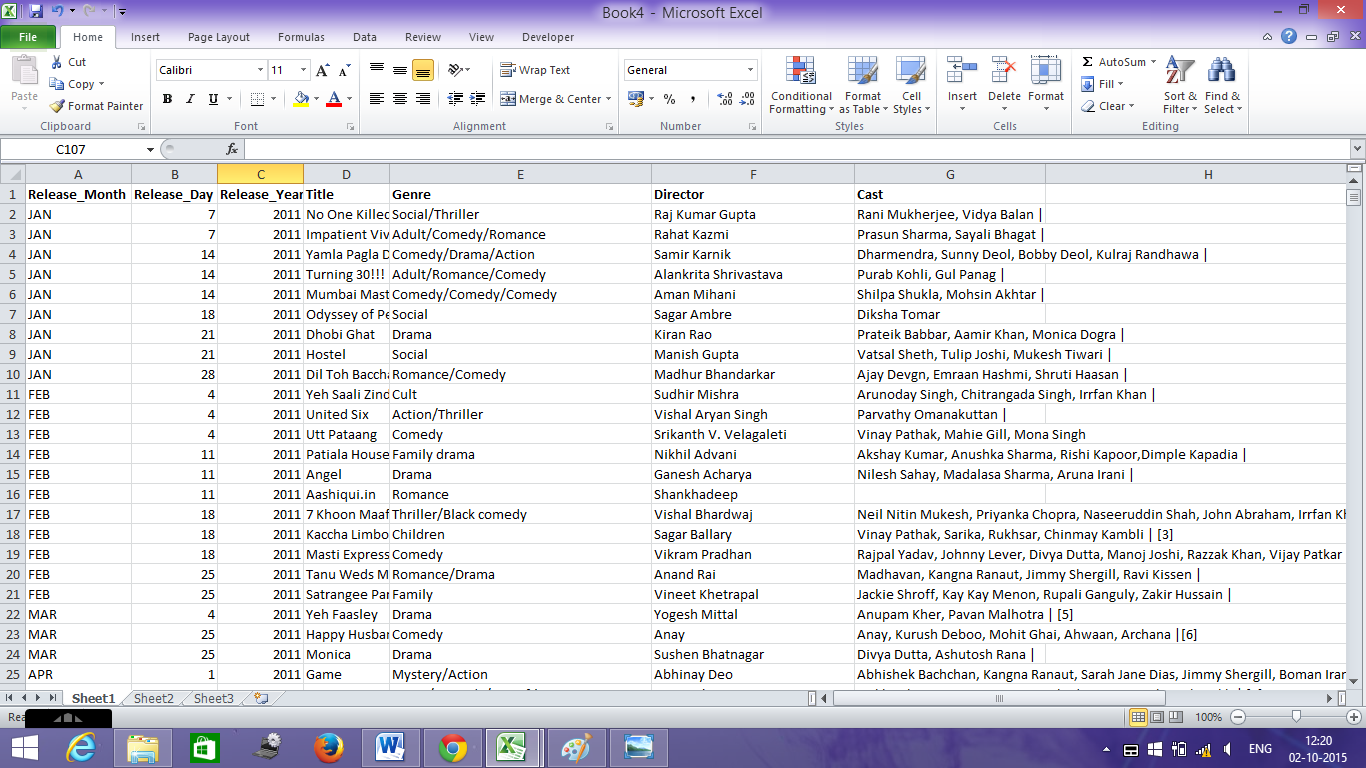
1. **DATA PRE-PROCESSING:**

Data collected is pre-processed using macro enabled Excel Workbooks.We are doing so, because websites keep changing their format of webpages without informing any of its users, so complete dependency of our data on any website will result in affecting our database.

Some options provided by macro is as follows:



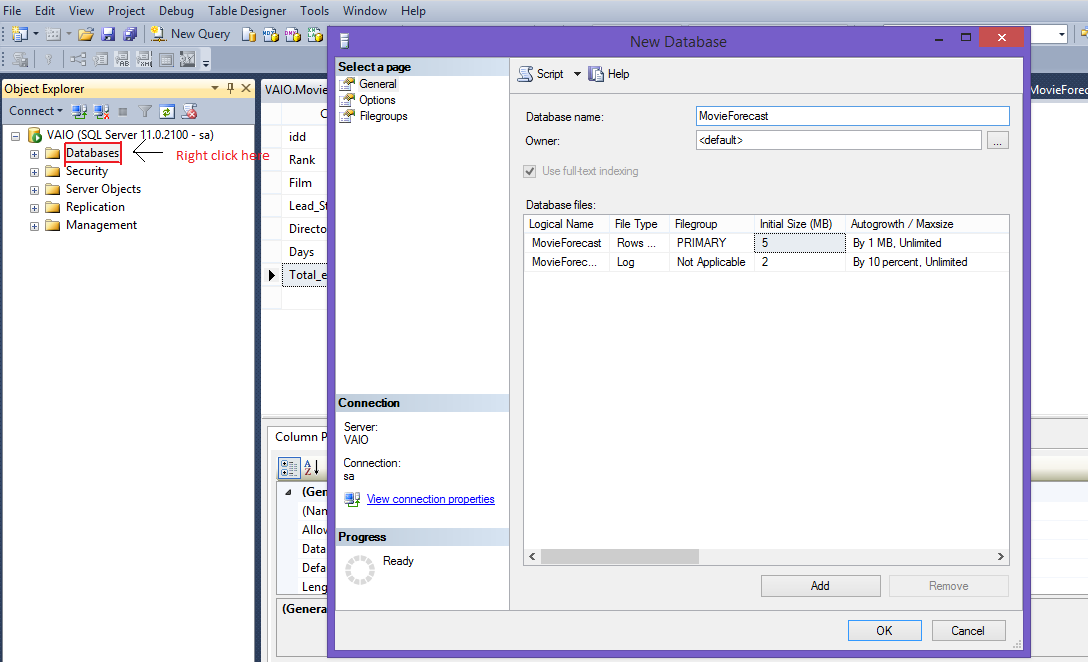
Pre-processed data is shown below:



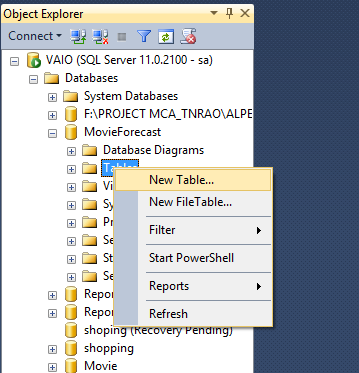
1. **Designing Database:**
2. **For our database we will be considering the following tables:**
   1. **Movie Data**: It consists of the following columns.

* Release\_Month: month in which the movie was released.
* Release\_Day: day on which the movie was released.
* Release\_Year: year in which the movie was released.
* Film: name of the movie.
* Genre:Defines the category of movie.
* Director: name of the director
* Cast: actor and actresses in the movie
* Upload\_Date: Date on which the data was uploaded using the interface which we will create.
  1. **Movie Verdict**: It consists of the following columns.
* Rank: rank of the movie
* Film: name of the movie.
* Release\_Date: date on which the movie was released
* Verdict: It describes the movie as:
* Super Hit
* Hit
* Plus
* Average
* Losing
* Flop
* Year: year in which the movie was released.
  1. **Stars Data**: It consists of the following columns.
* Cast\_Name: name of the actor or actress.
* Rank: his/her rank in the given year
* Year: year of which the data is recorded
* Cast\_type: whether the cast is an actor or actress.
  1. **Box-office Data**: It consists of the following columns.
* Rank: rank of the movie
* Film: name of the movie.
* Release\_date: date on which the movie was released
* Nett\_gross: the total amount of business done.
* Region: regions in which the movie did good business.
* Remark: remark for the movie in the specified region.
  1. **Box-office All Time**: It consists of the following columns.
* Rank: rank of the movie
* Film: name of the movie.
* Year: year of which the data is recorded
* Nett\_Gross: the total amount of business done.
* Gross: amount of business done after cutting some expenditures.
* Distributor\_Share: the share given to distributors.
* Verdict: Describes the verdict generated for movie like hit,super hit etc.
* Region: regions in which the movie did good business.
  1. **100cr Club**: It consists of the following columns.
* Rank: Movie rank
* Film: name of the movie
* Lead\_Stars: Lead actor and actress of movie.
* Directors: name of the director
* Days: Days been taken to reach the 100cr club.
* Total\_earn: total amount earned.
  1. **Box-office Gross: It consists of the following columns**.
* Rank: Movie rank
* Film: name of the movie.
* Release\_date: Movie release date
* Nett\_gross:Total gross earned by the movie
  1. **New releases: It consists of the following columns.**
* Release\_date: Movie release date
* Movie\_name: name of the movie.
* Genre:genre of the movie
* Director:name of the director
* Cast:star cast

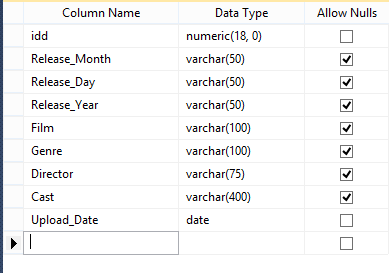
1. **Creating Database In SQL Server**
2. Open SQL server management studio and enter the required credentials.
3. In the object explorer window on left right click on Database and create a new database with name ‘MovieForecast’ then click OK. The database will be created in the object explorer under Database.



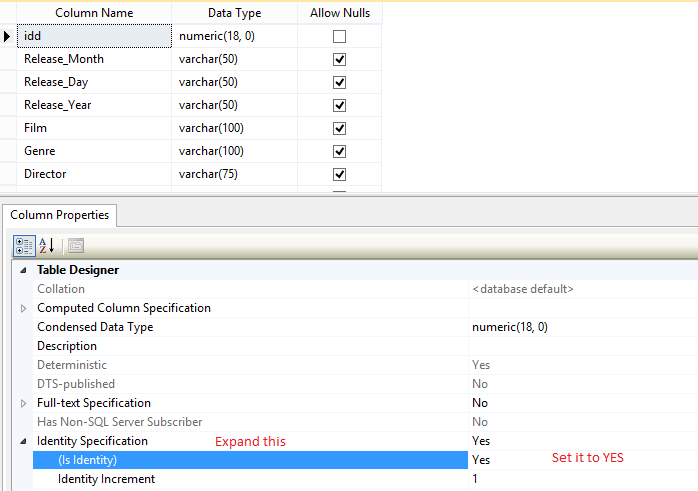
1. To create table expand your database created and in that right click on Table and select New Table to create a new table .



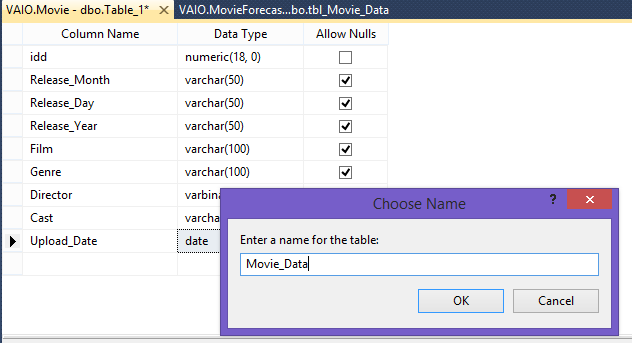
1. Enter the column name and its data type as follows:



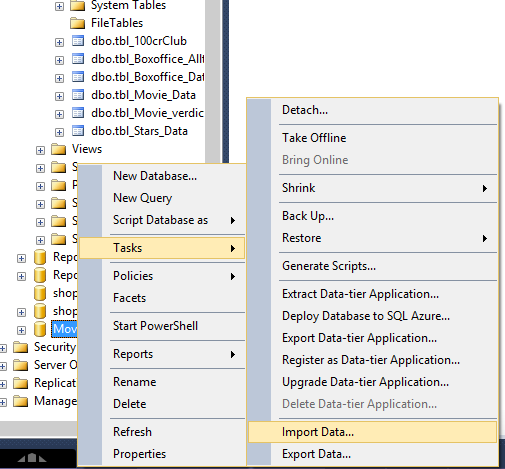
1. For the idd column to auto- increment by 1 do the following changes:



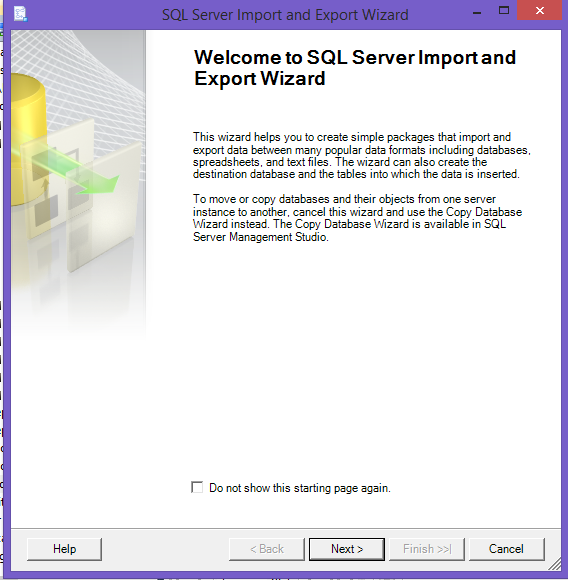
1. Set the name of the table



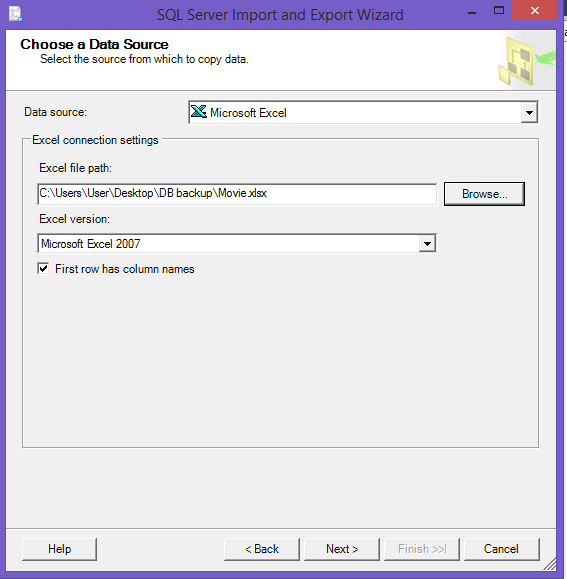
1. The table will be created in the database.
2. Follow the above steps to create all the tables.
3. **UPLOADING DATA IN DATABASE**
4. Right click on Database ‘Movie Forecast’->Task->Import data



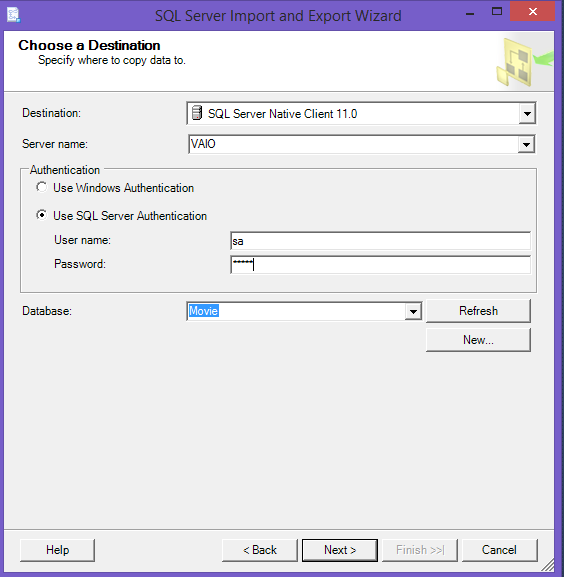
1. Click Next

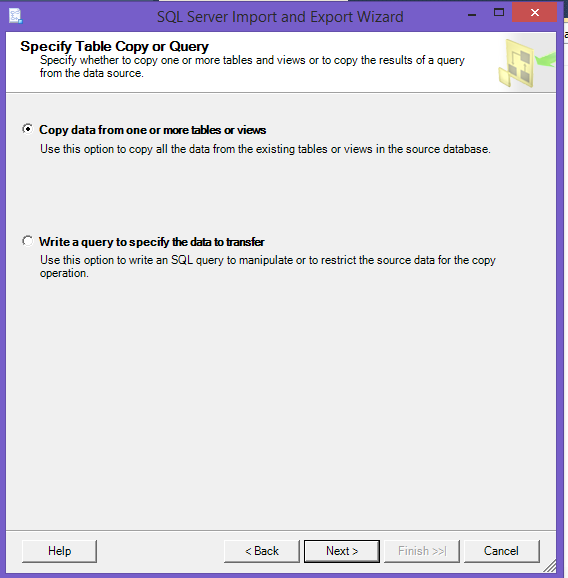


1. Enter the data source, the path of excel file and its version then click Next.

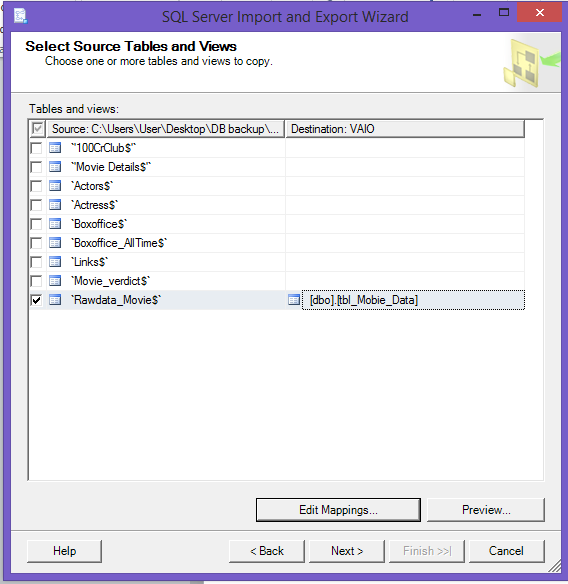


1. Enter the credentials required

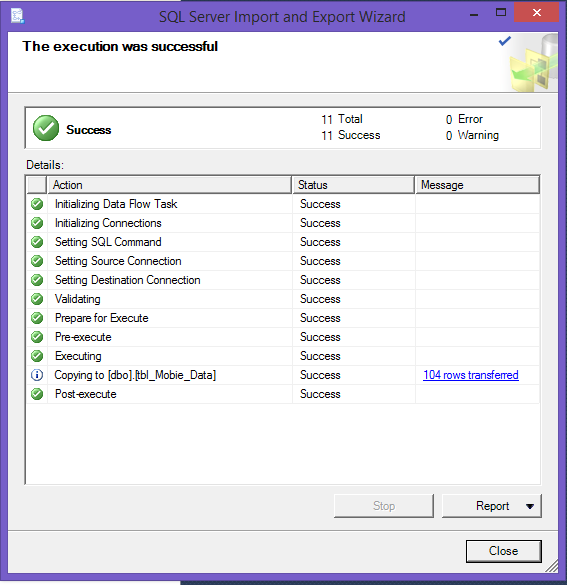


1. Click Next****
2. Select the correct sheet of excel in which your data is present and the table in which you want to load the data.

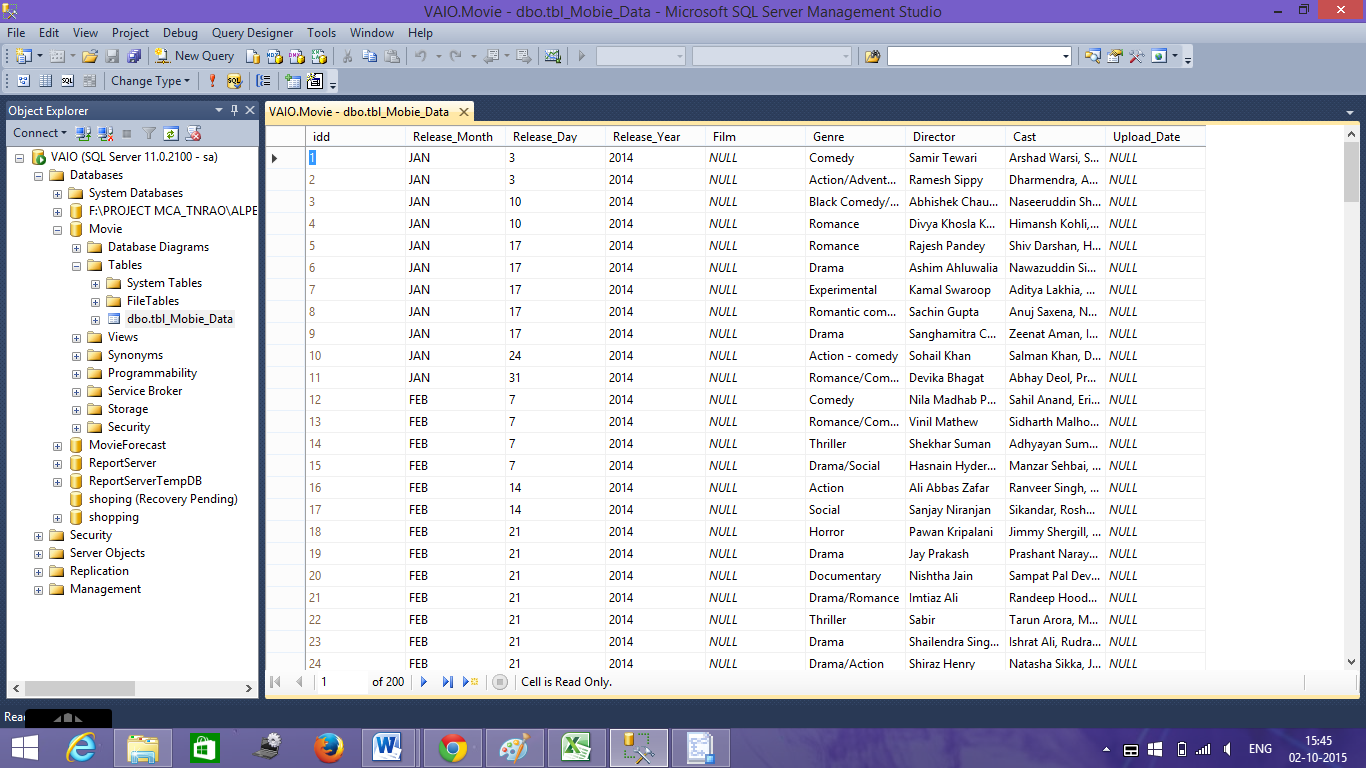
NOTE: The 1st line of excel sheet and the columns of the tables should be same.



1. Then Click Next till you reach the below screen

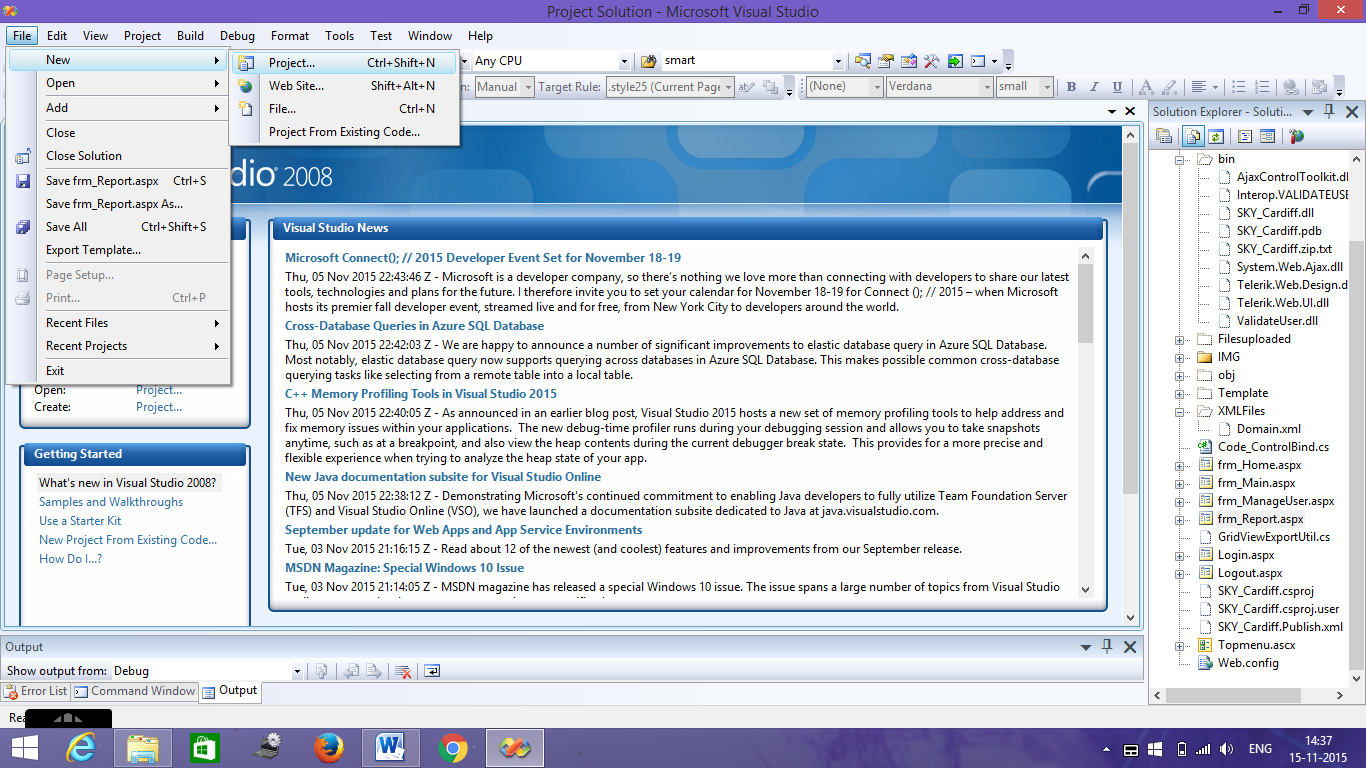


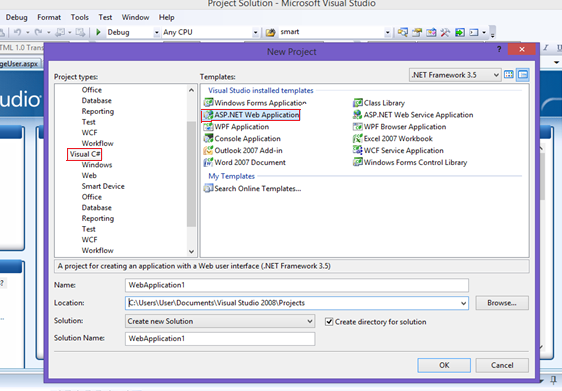
1. Final database is as follows



1. **GUI**
2. **DESIGNING INTERFACE**

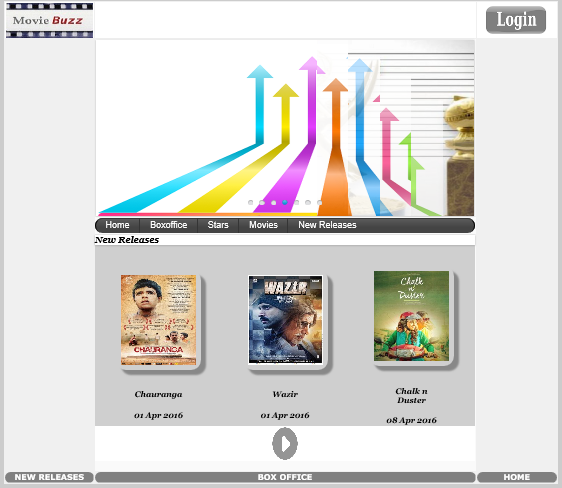
* Visual studio 2008 is used to create user interface.
* In visual studio an ASP.NET Web App project is created as shown below:



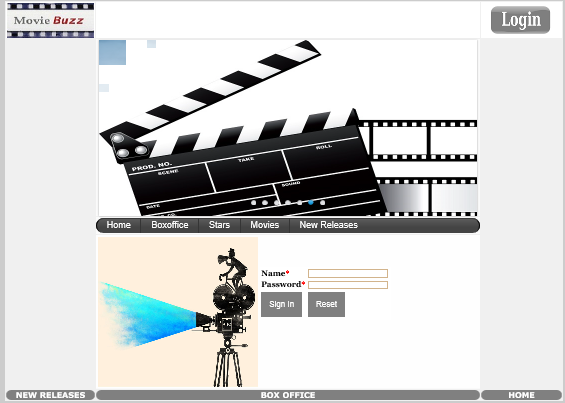


Enter the name of your project in the name column and location where your project should be saved

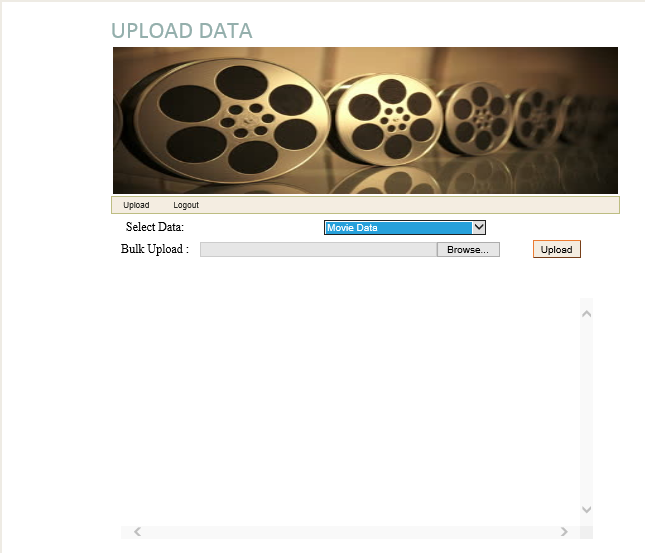
* User interface designed is as follows:
* Home Page:



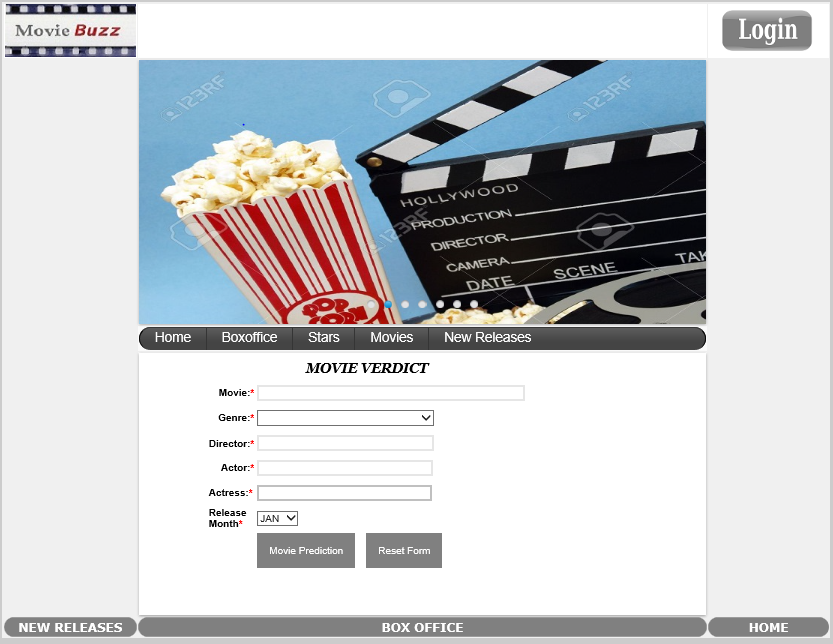
* Then login page is designed as shown below



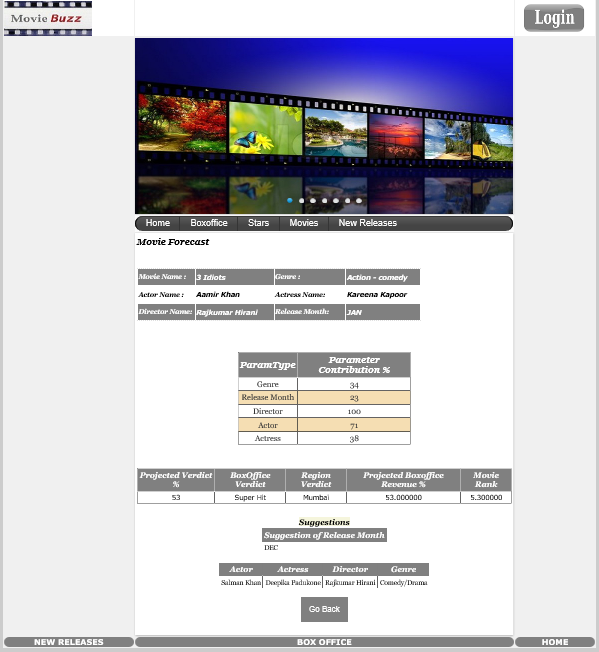
* After login the following page appears:



* To get the forecasted verdict user has to enter the details as mentioned in the form below



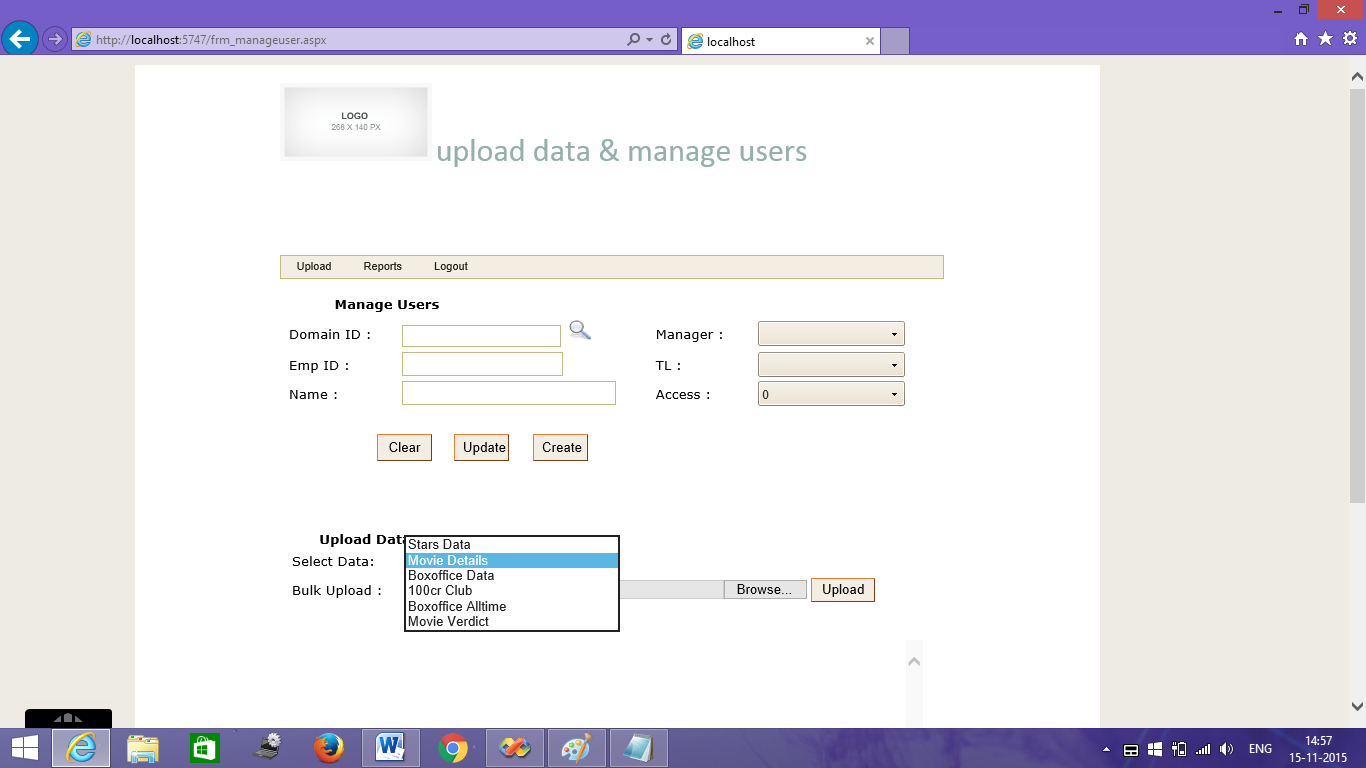
* After entering the details the output shown is as follows:



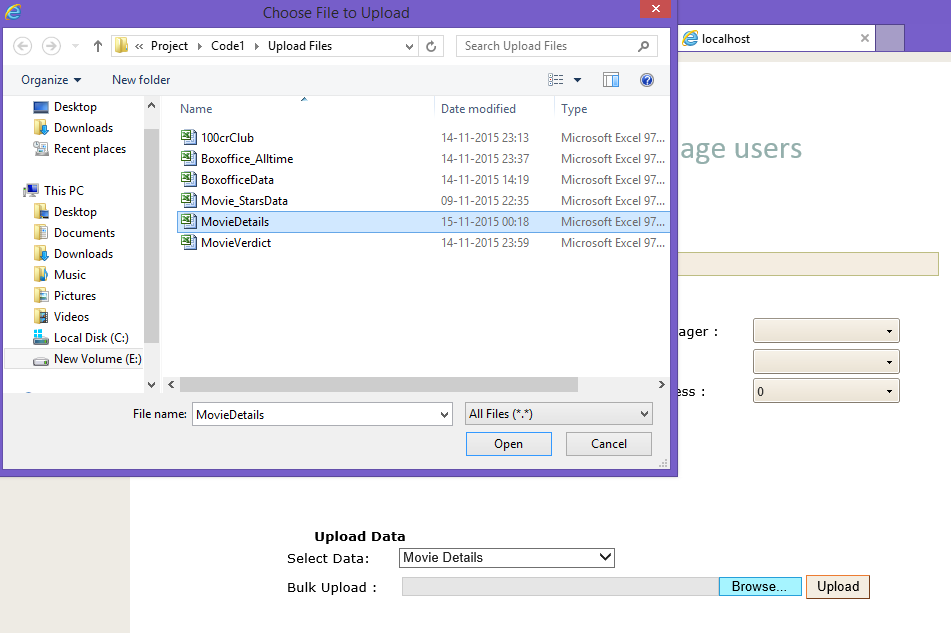
1. **UPLOADING DATA IN DATABASE**

Admin can upload data in database from front-end as follows

1. From the drop down menu select the table in which you want to upload the data.



1. Click on browse and select the path where your file is stored.



1. To upload data from front-end the admin has to enter the data in one of the 6 excel files available. Admin has to enter data in the format specified in excel file.

The files are available at following path:

**E:\Project\Code1\Upload Files**

1. After selecting the required file admin has to click on upload button for data to get uploaded in database.
2. **Algorithms Used:**
3. **Min-Max Normalization:** Normalization is done to scale the attribute data so that it fits in a specific range. There are many types of normalization techniques available like Z-score, Min-Max, etc. A simple Min-Max algorithm is used to normalize the data since using it we could normalize our values in the range of 0 to 1.

Y –the transformed normalized value.

X –the value which you want to normalize

If you want to convert the value in to range [0, 1] then A is 0 and B is 1.

1. **Regression algorithms:**
   * + 1. Linear regression algorithm - Linear regression involves a response variable y and a single predictor variable x. For linear regression value of y will be approximately calculated as follows:

Here A is the slope and B is the intercept.

Many prediction projects use Linear regression algorithm for predicting results .But since our verdict is depended on many variables we used Multiple – linear regression algorithm to forecast.

* + - 1. Multiple-Linear regression algorithm - It is used when there is 1 response variable y and many predictor variables for e.g. x1, x2, x3, x4... Here the function will look like

for i=1,2,…k.

where A0 is the intercept. A1, A2, …, An are partial regression coefficients and ℰ is the random error .

1. **CALCULATIONS OF OUTPUT:**
2. **Verdict Calculation**:

The user enters the following parameters:

* Genre
* Release month
* Actor
* Actress
* Director

Contribution of each parameter to calculate the verdict is as follows:

**For e.g. consider the Genre entered as ‘Drama’**

* Then the movies net gross is selected whose genre is as mentioned above.
* Data is normalised year wise.
* Net grosss is normalised by using the following formula

v'=((v-min)(newmax-newmin))/(max-min)+newmin

Where v=the gross which is to be normalised

max=max of that year

min=min of that year

new max=1

new min=0 and

v’=normalised value

After the data is normalised the contribution of genre for success of that movie is as follows:

* From the normalised data maximum value is selected
* Then each value is divided by maximum and then average of all the values obtained is calculated.

**In the same manner the contribution of other parameters is calculated.**

**Projected verdict= Average of (genre, actor, actress, release month, director)**

Verdict= If Projected verdict <35 Verdict=Flop

If Projected verdict >=35 and <43 then Verdict=Semi hit

If Projected verdict >=43 and <65 then Verdict=Super hit

If Projected verdict >=65 and <=100 then Verdict=Block Buster

1. **Region Where Movie will do good business calculation**

* Actor ,Actress and Genre Parameter entered above are considered here
* Net gross of movies is considered which consist of actor, actress or genre

i.e. parameter which has highest number of movies is considered .

* **The gross of the movies available is divided by the highest gross available in that table.**
* **After division the values are grouped region wise. Average of all the movies belonging to same region is done** .The region which has maximum average is considered as the region where the movie may do good business.

1. **Movie Rating**= The projected verdict calculated in output 1 **divided by** 10.
2. Suggestions for the movie.

* The releases month suggested is as follows:
  + - * + Actor, Actress and genre Parameter is considered.
        + Net gross of movies consisting of actor or actress or genre is considered.
        + Netgross is divided by maximum net gross available in that table.
        + Average of divided value is calculated, which is grouped by months.
        + The month which gives the highest value is suggested as the month of release to increase the overall performance of the movie.

1. Estimated gross prediction

It is calculated as follows

Suppose the parameters entered above contributed as

Genre = 32

Release month=42

Director=40

Actor=41

Actress=57

Estimated net gross=(Average(32,42,40,41,57)\*max(32,42,40,41,57))/100=23.94% the movie will contribute to the net gross og box office.