

Introduction :

Myself and Raj have worked in various technologies starting from Sybase through Ingress database. We have also worked in various other databases like Oracle, MS SQL Server etc.

Now we are into data Analysis. When we consider data analysis to derive an output, then we go by

- a. Planning , Organizing and documenting our research – this basically involves research and obtain (O)
- b. Cleaning Data – (S) – Basically, clean the data to fit our form and needed for our analysis
- c. Explore and Analyze (E)
- d. Model it (M) - we will model the database through diagram or manual or Erwin diagram etc. we can also model primary keys, foreign keys if we are using relational database. We can use excel or SQL query to model it more appropriately.
- e. iNterpret and depict and present (N)

Case Study/Project

In this case: Academy Award Best Picture. (Best Picture and Film Editing Award) –

Inference:

- a) Best Picture and Film Editing Nomination are richly correlated.
- b) Since 1981, Not a single film has won Best Picture without being nominated for Film Editing Award.
- c) About 2/3 of movie nominated for Film Editing , has gone to win Best Picture Award.

How to go About ?

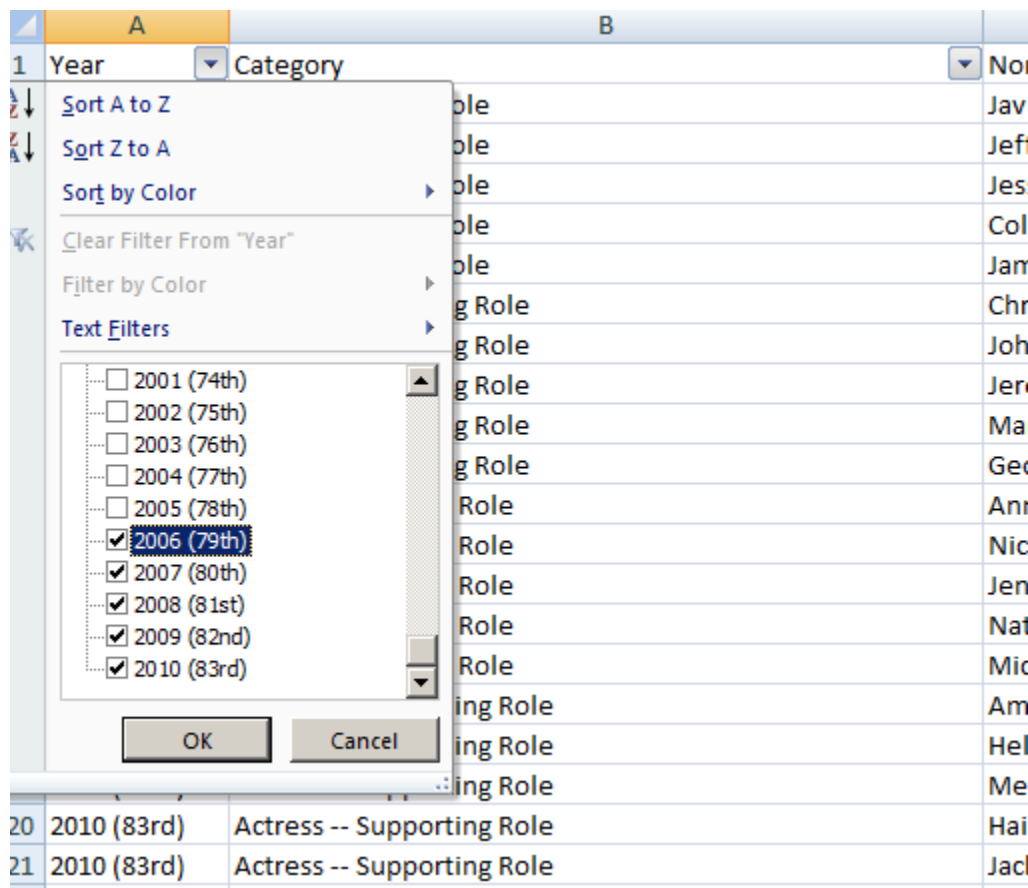
Please use data to determine whether Best Film Editing is the best predictor of Best Picture.

1. Identify categories of Academy Award Nominations –
 - a. current categories table
 - b. discontinued categories table
 - c. Proposed categories
2. Film Database year wise – We have got it upto 83rd Academy Awards
 - a. There are several ways to get it. if we have time, we can enter the data from the wiki pedia or go to http://awardsdatabase.oscars.org/ampas_awards/BasicSearchInput.jsp
 - b. Or we can download from available sources as excel sheet, database, data sets etc.
 - c. Identify the data and normalize it.

OPTION 1

(Option is to use the excel sheet, then write a small program and slice it to normalized tables to derive the data)

3. The data that we got from Internet is attached as excel sheet. Out of this, we considered loading the data for 2 years. Just for analysis purpose for two years. But concept can be extended further.



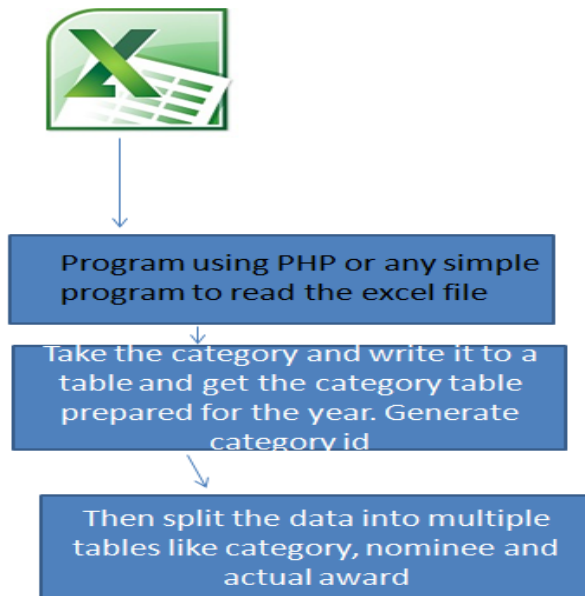
Year	Category	Nominee Name
2001 (74th)		
2002 (75th)		
2003 (76th)		
2004 (77th)		
2005 (78th)		
2006 (79th)		
2007 (80th)		
2008 (81st)		
2009 (82nd)		
2010 (83rd)		
2010 (83rd)	Actress -- Supporting Role	Hai
2010 (83rd)	Actress -- Supporting Role	Jack

4. Cleaned the data as below

The following categorical data , Nominee has not been properly entered or excel sheet does not have proper information.

Scientific and Technical (Scientific and Engineering Award)
 Scientific and Technical (Scientific and Engineering Award)
 Scientific and Technical (Scientific and Engineering Award)
 Scientific and Technical (Scientific and Engineering Award)

5. Main idea is to write a program to read the excel sheet and algorithm is as below



6. Once we get it to the normalized table, then it is very easy to find the “Film Editing Award” vs “Best Picture Award”

OPTION 2

(This is easiest way to find out from the excel sheet without have to load anything in PostgreSQL or R as below)

1. Open a excel sheet
2. Then filter it for a year, for “Best Picture” and Film Editing.

Year	Category	Nominee	Additional Info	Won ?
2010 (83rd)	Film Editing	Black Swan	Andrew Weisblum	NO
2010 (83rd)	Film Editing	The Fighter	Pamela Martin	NO
2010 (83rd)	Film Editing	The King's Speech	Tariq Anwar	NO
2010 (83rd)	Film Editing	127 Hours	Jon Harris	NO
2010 (83rd)	Film Editing	The Social Network	Angus Wall and Kirk Baxter Mike Medavoy, Brian Oliver and Scott	YES
2010 (83rd)	Best Picture	Black Swan	Franklin, Producers	NO
2010 (83rd)	Best Picture	The Fighter	David Hoberman, Todd Lieberman and	NO

IS-607-Project 3 Academy Award

			Mark Wahlberg, Producers Emma Thomas and Christopher Nolan, Producers	NO
2010 (83rd)	Best Picture	Inception		
		The Kids		
2010 (83rd)	Best Picture	Are All Right	Gary Gilbert, Jeffrey Levy-Hinte and Celine Rattray, Producers	NO
2010 (83rd)	Best Picture	The King's Speech	Iain Canning, Emile Sherman and Gareth Unwin, Producers	YES
2010 (83rd)	Best Picture	127 Hours	Christian Colson, Danny Boyle and John Smithson, Producers	NO
2010 (83rd)	Best Picture	The Social Network	Scott Rudin, Dana Brunetti, Michael De Luca and Ceán Chaffin, Producers	NO
2010 (83rd)	Best Picture	Toy Story 3	Darla K. Anderson, Producer	NO
2010 (83rd)	Best Picture	True Grit	Scott Rudin, Ethan Coen and Joel Coen, Producers	NO
2010 (83rd)	Best Picture	Winter's Bone	Anne Rosellini and Alix Madigan-Yorkin, Producers	NO
2009 (82nd)	Film Editing	Avatar	Stephen Rivkin, John Refoua and James Cameron	NO
2009 (82nd)	Film Editing	District 9	Julian Clarke	NO
2009 (82nd)	Film Editing	The Hurt Locker	Bob Murawski and Chris Innis	YES
2009 (82nd)	Film Editing	Inglourious Basterds	Sally Menke	NO
		Precious: Based on the Novel 'Push' by Sapphire		
2009 (82nd)	Film Editing		Joe Klotz	NO
2009 (82nd)	Best Picture	Avatar	James Cameron and Jon Landau, Producers	NO
2009 (82nd)	Best Picture	The Blind Side	Gil Netter, Andrew A. Kosove and Broderick Johnson, Producers	NO
2009 (82nd)	Best Picture	District 9	Peter Jackson and Carolynne Cunningham, Producers	NO
2009 (82nd)	Best Picture	An Education	Finola Dwyer and Amanda Posey, Producers	NO
2009 (82nd)	Best Picture	The Hurt Locker	Kathryn Bigelow, Mark Boal, Nicolas Chartier and Greg Shapiro, Producers	YES
2009 (82nd)	Best Picture	Inglourious Basterds	Lawrence Bender, Producer	NO
		Precious: Based on the Novel 'Push' by Sapphire		
2009 (82nd)	Best Picture		Lee Daniels, Sarah Siegel-Magness and Gary Magness, Producers	NO

IS-607-Project 3 Academy Award

2009 (82nd)	Best Picture	A Serious Man	Joel Coen and Ethan Coen, Producers	NO
2009 (82nd)	Best Picture	Up	Jonas Rivera, Producer	NO
2009 (82nd)	Best Picture	Up in the Air	Daniel Dubiecki, Ivan Reitman and Jason Reitman, Producers	NO
2008 (81st)	Film Editing	The Curious Case of Benjamin Button	Kirk Baxter and Angus Wall	NO
2008 (81st)	Film Editing	The Dark Knight	Lee Smith	NO
2008 (81st)	Film Editing	Frost/Nixon	Mike Hill and Dan Hanley	NO
2008 (81st)	Film Editing	Milk	Elliot Graham	NO
2008 (81st)	Film Editing	Slumdog Millionaire	Chris Dickens	YES
2008 (81st)	Best Picture	The Curious Case of Benjamin Button	Kathleen Kennedy, Frank Marshall and Ceán Chaffin, Producers	NO
2008 (81st)	Best Picture	Frost/Nixon	Brian Grazer, Ron Howard and Eric Fellner, Producers	NO
2008 (81st)	Best Picture	Milk	Dan Jinks and Bruce Cohen, Producers	NO
2008 (81st)	Best Picture	The Reader	Anthony Minghella, Sydney Pollack, Donna Gigliotti and Redmond Morris, Producers	NO
2008 (81st)	Best Picture	Slumdog Millionaire	Christian Colson, Producer	YES
2007 (80th)	Film Editing	The Bourne Ultimatum	Christopher Rouse	YES
2007 (80th)	Film Editing	The Diving Bell and the Butterfly	Juliette Welfling	NO
2007 (80th)	Film Editing	Into the Wild	Jay Cassidy	NO
2007 (80th)	Film Editing	No Country for Old Men	Roderick Jaynes	NO
2007 (80th)	Film Editing	There Will Be Blood	Dylan Tichenor	NO
2007 (80th)	Best Picture	Atonement	Tim Bevan, Eric Fellner and Paul Webster, Producers	NO
2007 (80th)	Best Picture	Juno	Lianne Halfon, Mason Novick and	NO

IS-607-Project 3 Academy Award

2007 (80th)	Best Picture	Michael Clayton	Russell Smith, Producers Sydney Pollack, Jennifer Fox and Kerry Orent, Producers	NO
2007 (80th)	Best Picture	No Country for Old Men	Scott Rudin, Ethan Coen and Joel Coen, Producers	YES
2007 (80th)	Best Picture	There Will Be Blood	JoAnne Sellar, Paul Thomas Anderson and Daniel Lupi, Producers	NO
2006 (79th)	Film Editing	Babel	Stephen Mirrione and Douglas Crise	NO
2006 (79th)	Film Editing	Blood Diamond	Steven Rosenblum	NO
2006 (79th)	Film Editing	Children of Men	Alex Rodríguez and Alfonso Cuarón	NO
2006 (79th)	Film Editing	The Departed	Thelma Schoonmaker Clare Douglas, Christopher Rouse and Richard Pearson	YES
2006 (79th)	Film Editing	United 93	Alejandro González Iñárritu, Jon Kilik and Steve Golin, Producers	NO
2006 (79th)	Best Picture	Babel		NO
2006 (79th)	Best Picture	The Departed	Graham King, Producer	YES
2006 (79th)	Best Picture	Letters from Iwo Jima	Clint Eastwood, Steven Spielberg and Robert Lorenz, Producers	NO
2006 (79th)	Best Picture	Little Miss Sunshine	David T. Friendly, Peter Saraf and Marc Turtletaub, Producers	NO
2006 (79th)	Best Picture	The Queen	Andy Harries, Christine Langan and Tracey Seaward, Producers	NO

The above excel sheet , 2007 record is not proving our data analysis between Best Picture Vs Film Editing. However, 2010,2009 and 2008 is depicting the results as to any nominated File Editing , will also win Best Picture.

OPTION 3

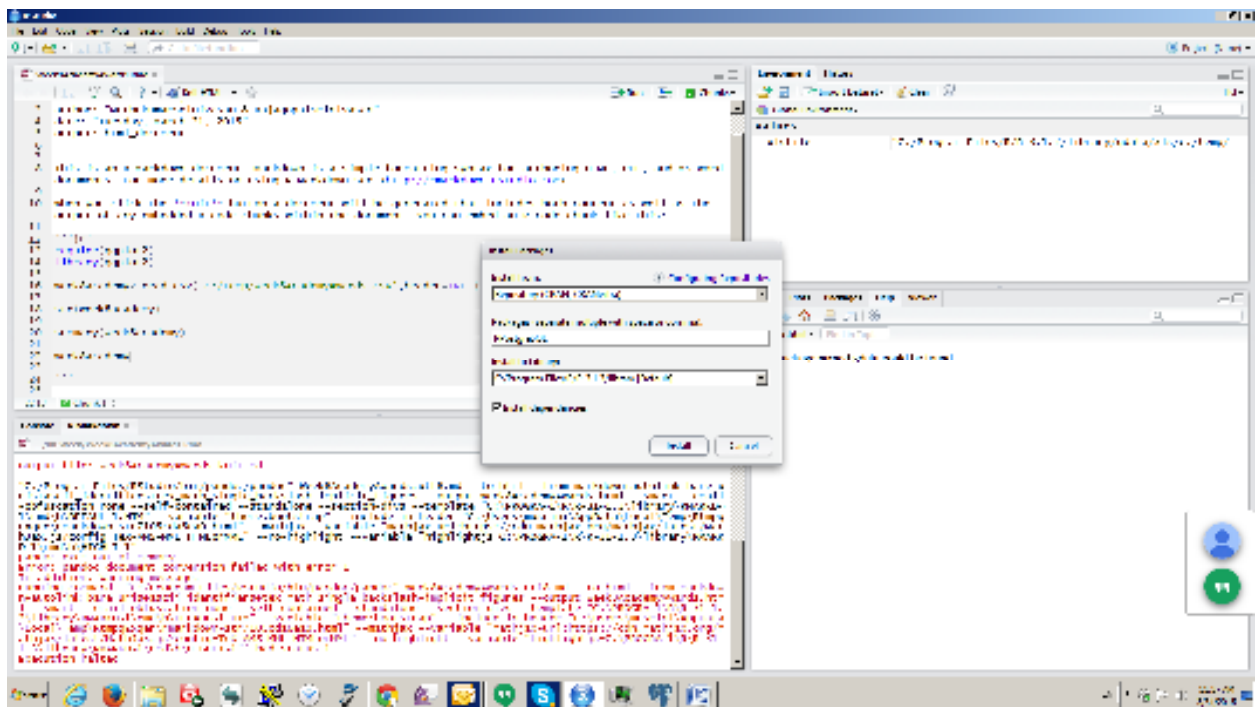
This is through R Studio Analysis

- a. We can load the csv file and analyze.

We have done this type of analysis in the past. Loaded CSV has got Year, nominee, did they won etc. based on this, we can find out Best Picture Vs Film Editing each year and provide the analysis.

- b. Or we can connect to PostgreSQL and derive the table to analyze.

Another way is to connect R Studio to PostgreSQL and provide the analysis. For this, we need to install the RPostgreSQL package as below



RPostgreSQL Package was installed

```
> install.packages("RPostgreSQL")
trying URL 'http://cran.rstudio.com/bin/windows/contrib/3.1/RPostgreSQL_0.4.zip'
Content type 'application/zip' length 434912 bytes (424 KB)
opened URL
downloaded 424 KB
```

package 'RPostgreSQL' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
C:\Users\musrini\AppData\Local\Temp\RtmpmsLMS\downloaded_packages

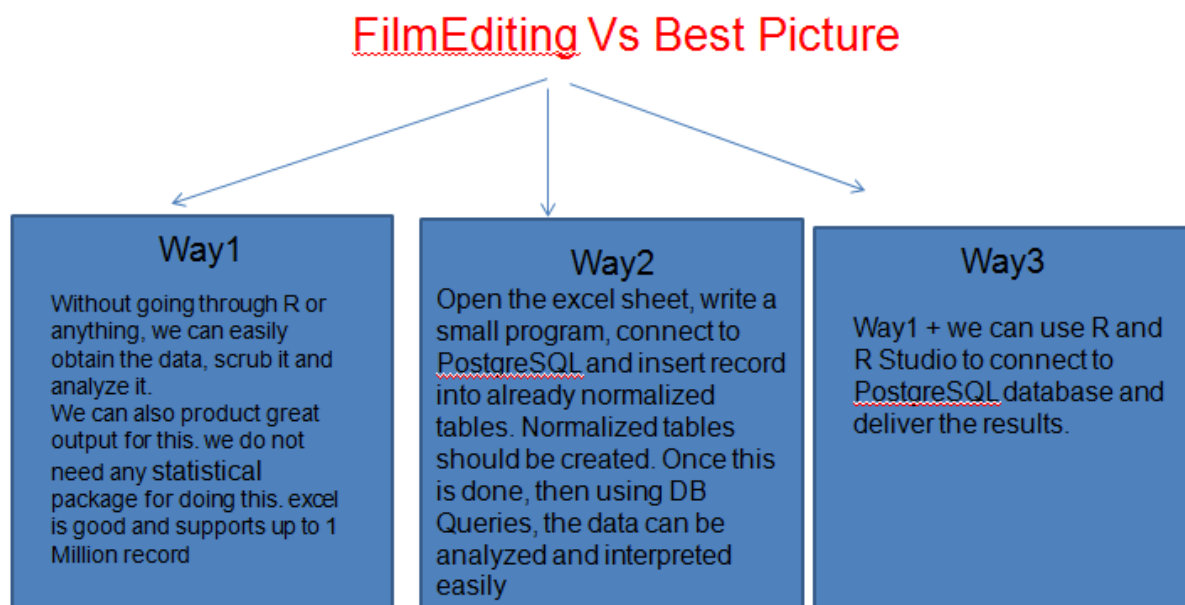
>

The mark down files are given. Through R Studio, we can connect to the PostgreSQL and get the data and interpret.

Disclaimer

We have got data in excel sheet. We could not write the program to put it in the database or slide it out. So we have manually entered and. We have designed the overall flow and we have also seen it working with few data in the R Studio.

CONCLUSION



Using one of the above, we can conclude that

- a) Best Picture and Film Editing Nomination are richly correlated.
- b) Since 1981, Not a single film has won Best Picture without being nominated for Film Editing Award.
- c) About 2/3 of movie nominated for Film Editing , has gone to win Best Picture Award.

Reference Points

Muthukumar Srinivasan & Rajagopal Srinivasan

IS-607-Project 3 Academy Award

<https://medium.com/silk-stories/why-oscar-is-a-male-a-data-driven-analysis-of-gender-representation-in-87-years-of-academy-awards-837ddd812d27>

http://en.wikipedia.org/wiki/Academy_Awards#Awards_of_Merit_categories

Academy award database

<https://www.aggdata.com/awards/oscar> - Not in Normalized form

<http://www.ya-shin.com/awards/awards.html#oscars> – Not in normalized form

<http://blog.silk.co/post/78199230161/how-we-made-the-oscar-database> - could not obtain because it was too much for us for this project