



Golf Champions Data Mart

CSC 177 - Team #3

Nicholas Hanks

Dinesh Binjola



Agenda



- Introduction
- Background & Scope
- Implementation
 - Data cleaning & preprocessing
 - Data mart design
- Demo
- Lessons Learned
- References
- Questions?



Introduction



- The objective: produce a comprehensive data mart to query golf tournament champions
- Data was available online, but in multiple disparate data sets/pages on the website



Background & Scope



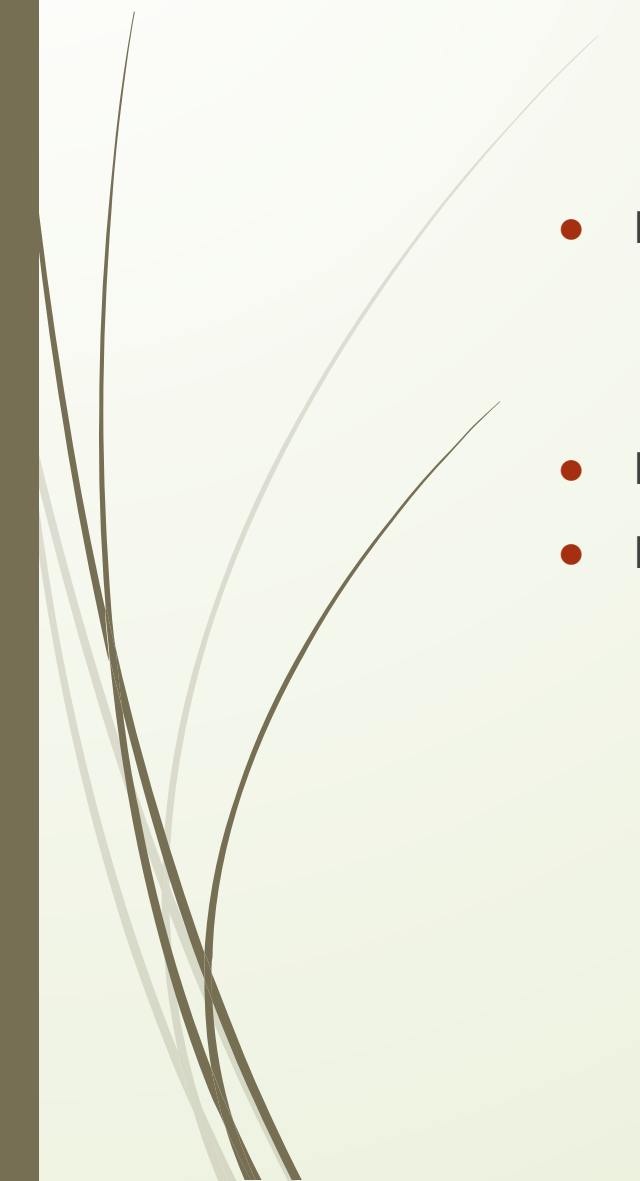
- Data source: www.databasegolf.com
- Data ranged from 1860 – 2015
- Sampled data from four tournaments: PGA Championship, Augusta Masters, US Open, & Open Championship (British Open)
- Data available on winning player & tournament
- Data changed in .csv format





Implementation



Cleaning & Preprocessing

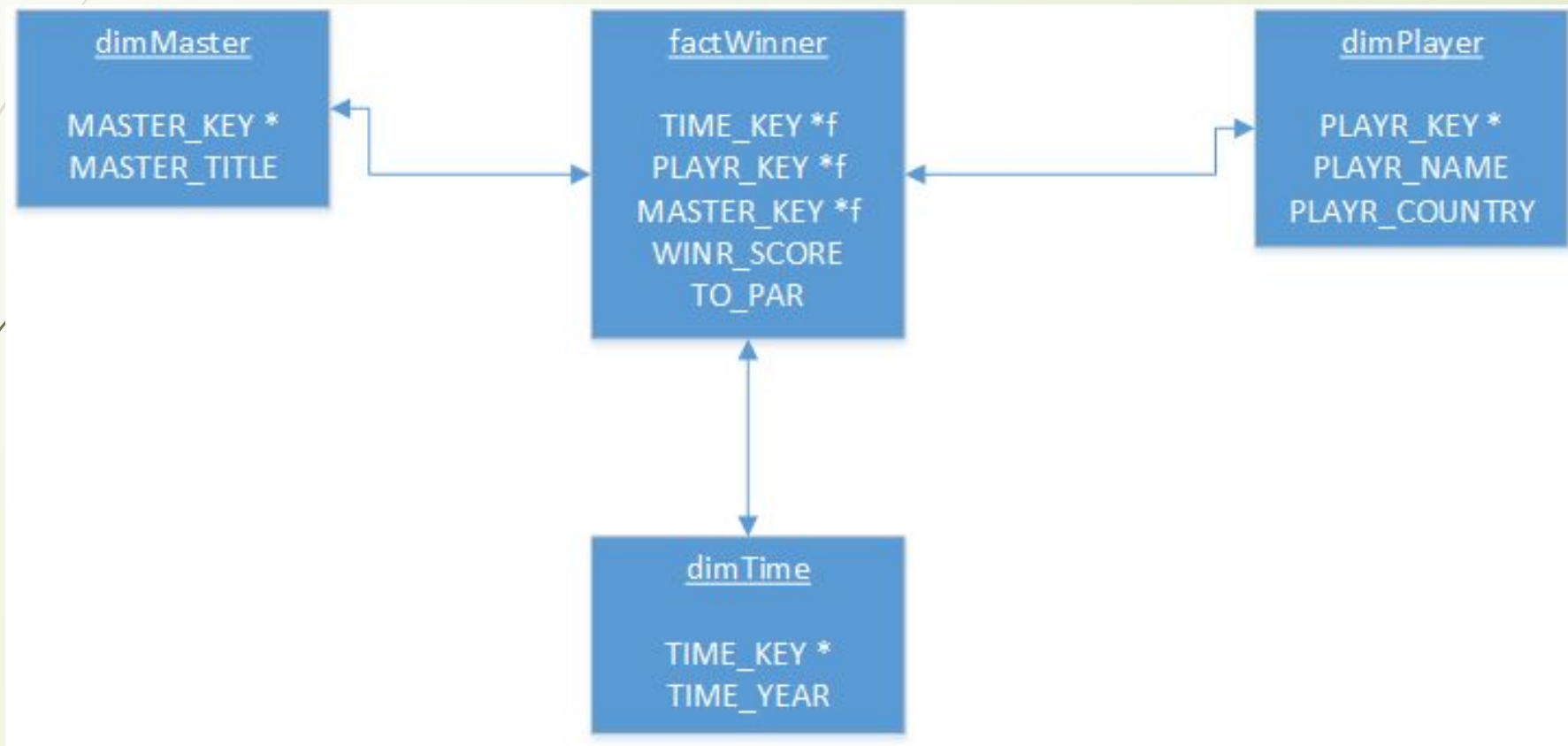
- Data loaded into Excel & scanned for errors
 - Removed non-required attributes
 - Noise detected in player names & scores
 - Manually corrected errors in records
 - Prepared to load into SQL Server
- 

	A	B	C	D	E	F	G	H
1	1	1934	United States	Horton Smith	284	-4	agusta_masters	
2	2	1935	United States	Gene Sarazen	282	-6	agusta_masters	
3	3	1936	United States	Horton Smith	285	-3	agusta_masters	
4	4	1937	United States	Byron Nelson	283	-5	agusta_masters	
5	5	1938	United States	Henry Picard	285	-3	agusta_masters	
6	6	1939	United States	Ralph Guldahl	279	-9	agusta_masters	
7	7	1940	United States	Jimmy Demaret	280	-8	agusta_masters	
8	8	1941	United States	Craig Wood	280	-8	agusta_masters	
9	9	1942	United States	Byron	280	-8	agusta_masters	
10	10	1946	United States	Herman Keiser	282	-6	agusta_masters	
11	11	1947	United States	Jimmy Demaret	281	-7	agusta_masters	
12	12	1948	United States	Claude Harmon	279	-9	agusta_masters	
13	13	1949	United States	Sam Snead	282	-6	agusta_masters	
14	14	1950	United States	Jimmy Demaret	283	-5	agusta_masters	
15	15	1951	United States	Ben Hogan	280	-8	agusta_masters	
16	16	1952	United States	Sam Snead	286	-2	agusta_masters	
17	17	1953	United States	Ben Hogan	274	-14	agusta_masters	
18	18	1954	United States	Sam Snead	289	1	agusta_masters	
19	19	1955	United States	Cary Middlecoff	279	-9	agusta_masters	
20	20	1956	United States	Jack Burke, Jr.	289	1	agusta_masters	
21	21	1957	United States	Doug Ford	283	-5	agusta_masters	
22	22	1958	United States	Arnold Palmer	284	-4	agusta_masters	
23	23	1959	United States	Art Wall, Jr.	284	-4	agusta_masters	
24	24	1960	United States	Arnold Palmer	282	-6	agusta_masters	
25	25	1961	South Africa	Gary Player	280	-8	agusta_masters	
26	26	1962	United States	Arnold Palmer	280	-8	agusta_masters	
27	27	1963	United States	Jack Nicklaus	286	-2	agusta_masters	
28	28	1964	United States	Arnold Palmer	276	-12	agusta_masters	
29	29	1965	United States	Jack Nicklaus	271	-17	agusta_masters	
30	30	1966	United States	Jack Nicklaus	288	E	agusta_masters	
		final						

	A	B	C	D	E	F
1	Year	Country	Name	Score	Par	Title
31	1966	United States	Jack Nicklaus	288	E	agusta_masters
102	1916	United States	Chick Evans*	286	Ø2	us_open
121	1938	United States	Ralph Guldani	284	E	us_open
132	1953	United States	Ben Hogan	283	Ø5	us_open
150	1971	United States	Lee Trevino	280	E	us_open
158	1979	United States	Hale Irwin	284	E	us_open
174	1995	United States	Corey Pavin	280	E	us_open
177	1998	United States	Lee Janzen	280	E	us_open
184	2005	New Zealand	Michael Campbell	280	E	us_open
189	2010	N. Ireland	Graeme McDowell	284	E	us_open
271	1948	England	Henry Cotton	284	E	open_champ
272	1949	South Africa	Bobby Locke	283	Ø5	open_champ
275	1952	South Africa	Bobby Locke	287	Ø1	open_champ
282	1959	South Africa	Gary Player	284	E	open_champ
309	1986	Australia	Greg Norman	280	E	open_champ
321	1998	United States	Mark O'Meara	280	E	open_champ
386	1966	United States	Al Geiberger	280	E	pga
436						

Data Mart Design - Star Schema



Data Loading in SQL Server

SQL Server Import and Export Wizard

Choose a Data Source
Select the source from which to copy data.

Data source: Flat File Source

Select a file and specify the file properties and the file format.

File name: C:\Users\Nick\Documents\Factscsv.csv Browse...

Locale: English (United States) ☐ Unicode

Code page: 1252 (ANSI - Latin I)


Format: Delimited

Text qualifier: <none>

Header row delimiter: [CR][LF]

Header rows to skip: 0


☒ Column names in the first data row

 Columns are not defined for this connection manager.












Help < Back Next > Finish >> Cancel

SQL Server Import and Export Wizard

The execution was successful

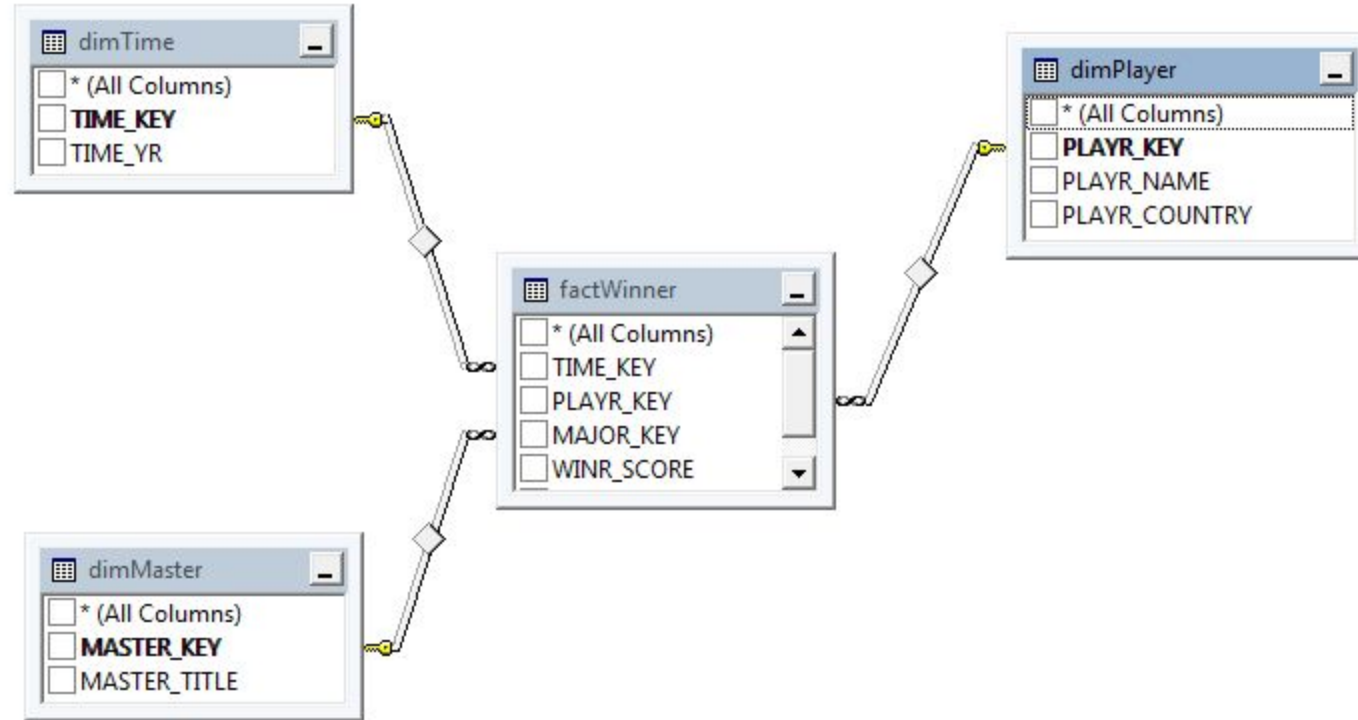
 **Success** 11 Total 0 Error
11 Success 0 Warning

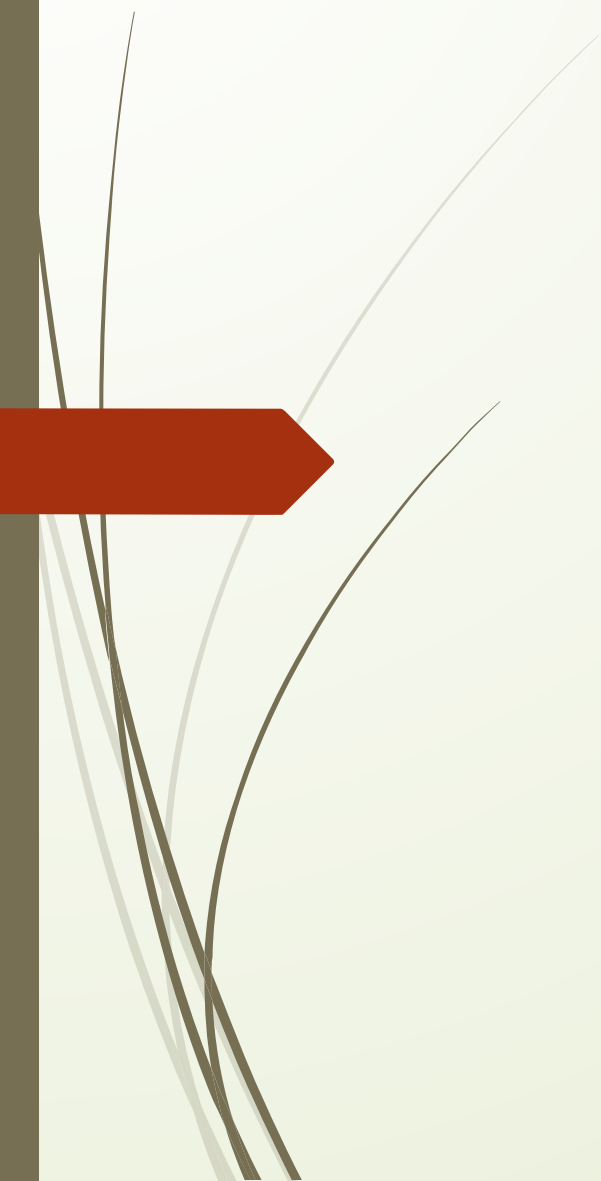
Details:

Action	Status	Message
 Initializing Data Flow Task	Success	
 Initializing Connections	Success	
 Setting SQL Command	Success	
 Setting Source Connection	Success	
 Setting Destination Connection	Success	
 Validating	Success	
 Prepare for Execute	Success	
 Pre-execute	Success	
 Executing	Success	
 Copying to [dbo].[Factscsv]	Success	434 rows transferred
 Post-execute	Success	

Stop Report


Close





Demo!

Fun Quiz

- 
- Which file format was imported in Weka?
 - .csv
 - .arff
 - .excel
 - .dat



Lessons Learned



- Data cleaning & preprocessing takes time
- It's important to start early
- Many tools to choose from
 - Excel for cleaning and preprocessing
 - MySQL, Sql Server & SS Management Studio for DB implementation
 - Learning PHP, MySQL, CSS, HTML, Javascript, and bootstrap for website implementation
 - Learning WEKA and Tableau
- Coordination & teamwork are key.
- Always Start Early. Don't keep it for the last minute.
- We believe that Tableau was the best of the tools among we used. We had outputs from WEKA and Tableau and outputs and analysis from Tableau were far better than WEKA.



References



- www.databasegolf.com
- https://en.wikipedia.org/wiki/Main_Page
- <http://www.masters.com/index.html>
- www.amatuergolf.com



Questions??