

Data Mining & OLAP Reporting

Phase: Data Modelling & Frontend Design Week 49/2007

Christoph Johann Szczecina

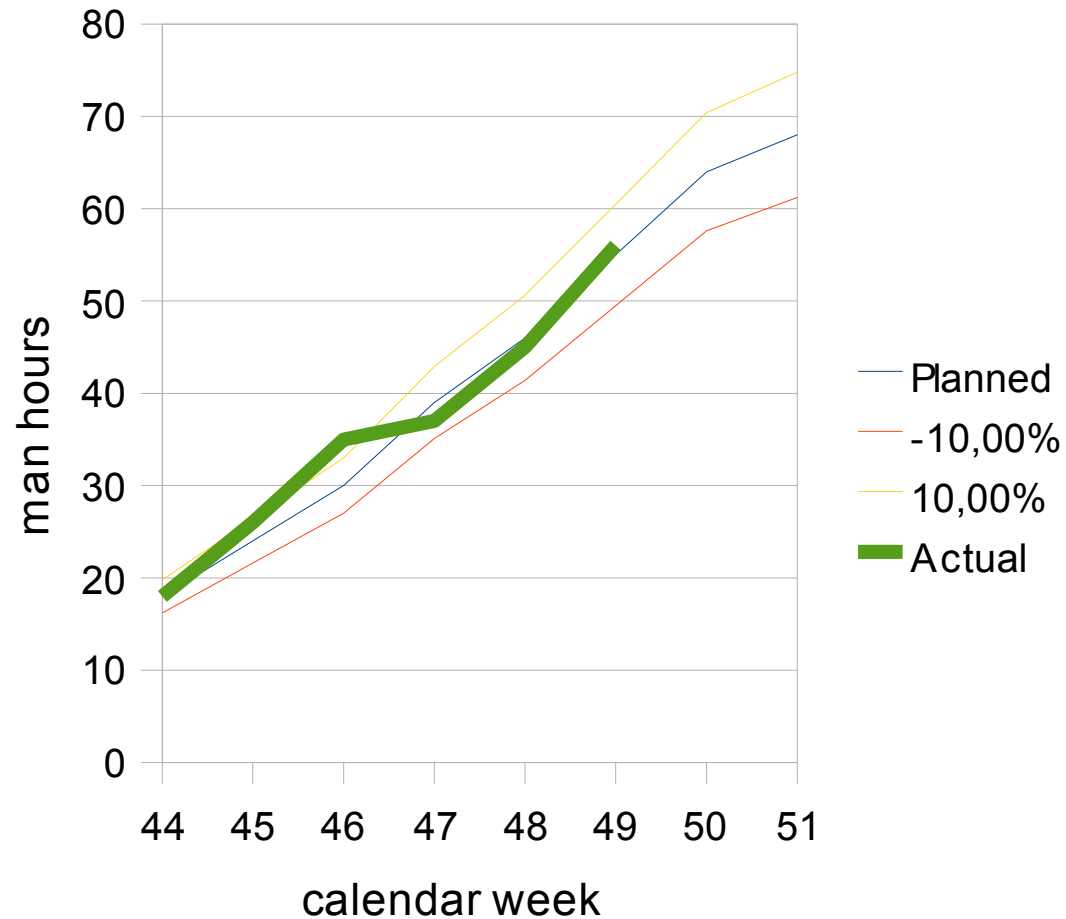
Udi Ibarretxe

Andoni Albizu





Dario Casas

Borja Uriarte

Budget: Man-Hours

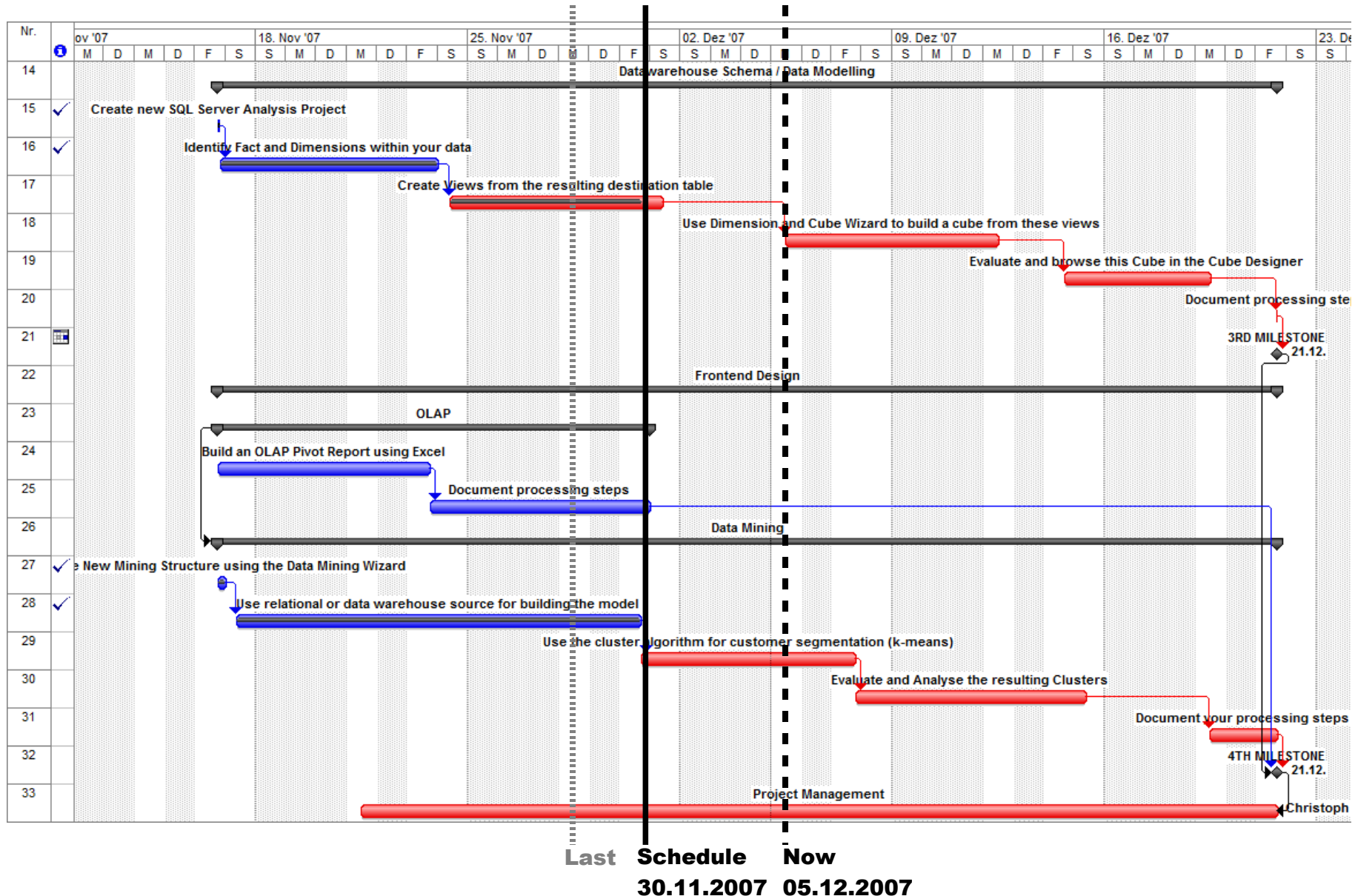


Since Last Report: Planned vs. Actual

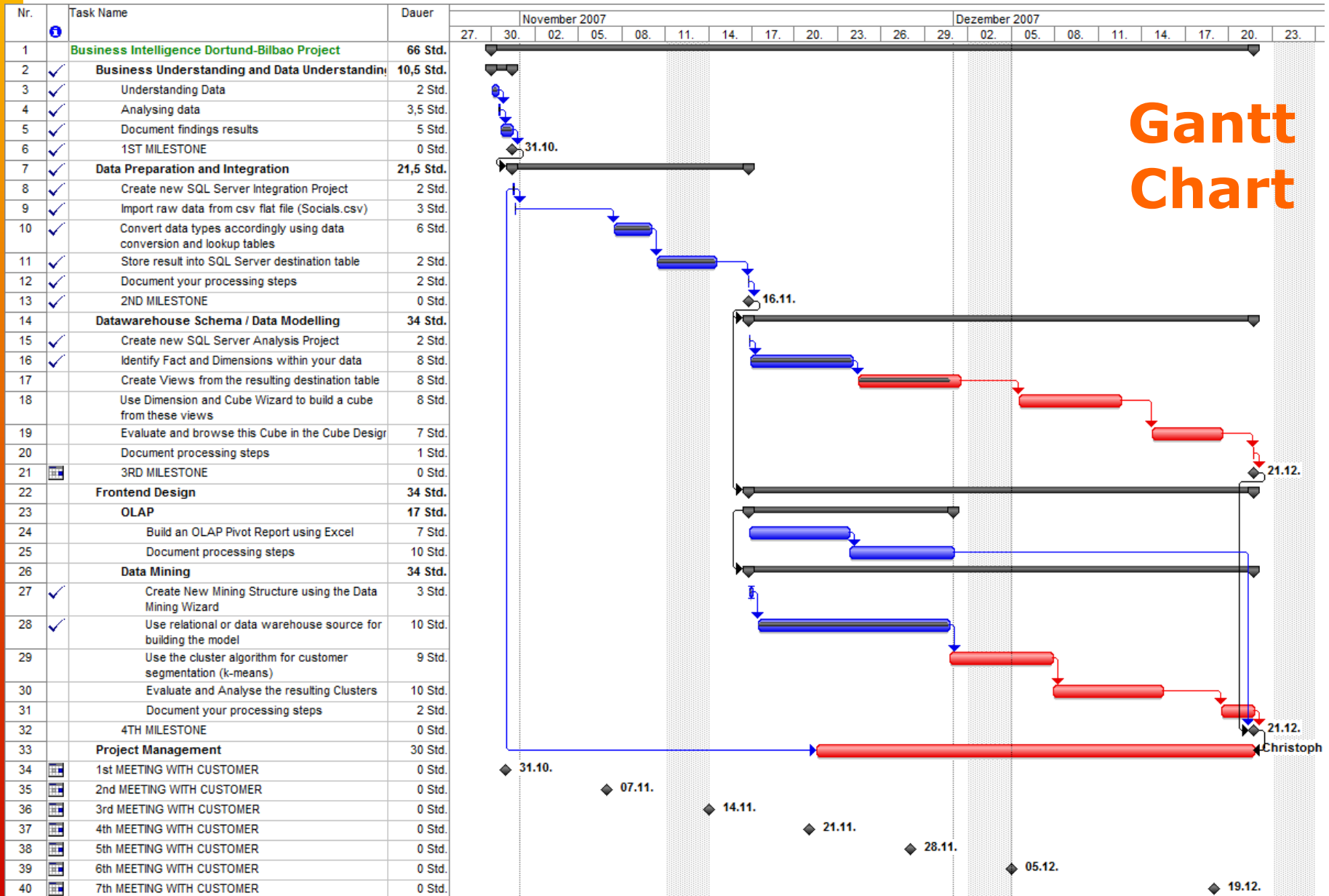
- [16] Identify Fact and Dimensions within data 
- [17] Create Views from the resulting destination table 
- [24] Build an OLAP Pivot Report using Excel 
- [28] Use relational or data warehouse source for building the model 

Progress Overview

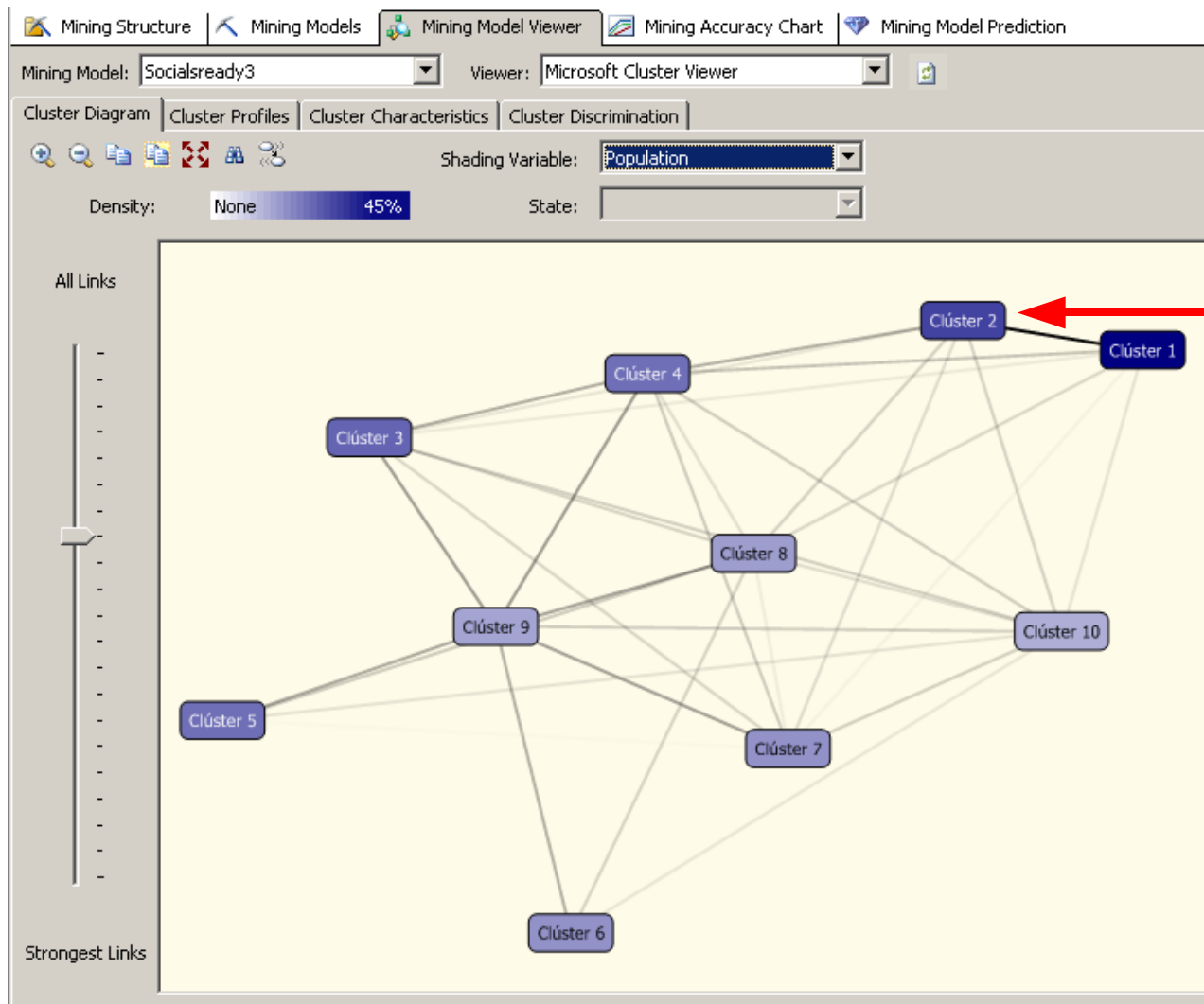
Nr.	Task Name	Dauer	Anfang	Ende	% Abgeschlossen
1	Business Intelligence Dortmund-Bilbao Project	66 Std.	Di 30.10.07	Fr 21.12.07	37%
2	Business Understanding and Data Understanding	10,5 Std.	Di 30.10.07	Mi 31.10.07	100%
3	Understanding Data	2 Std.	Di 30.10.07	Di 30.10.07	100%
4	Analysing data	3,5 Std.	Di 30.10.07	Di 30.10.07	100%
5	Document findings results	5 Std.	Di 30.10.07	Mi 31.10.07	100%
6	1ST MILESTONE	0 Std.	Mi 31.10.07	Mi 31.10.07	100%
7	Data Preparation and Integration	21,5 Std.	Mi 31.10.07	Fr 16.11.07	100%
8	Create new SQL Server Integration Project	2 Std.	Mi 31.10.07	Mi 31.10.07	100%
9	Import raw data from csv flat file (Socials.csv)	3 Std.	Mi 31.10.07	Mi 31.10.07	100%
10	Convert data types accordingly using data conversion and lookup tables	6 Std.	Mi 07.11.07	Fr 09.11.07	100%
11	Store result into SQL Server destination table	2 Std.	Sa 10.11.07	Mi 14.11.07	100%
12	Document your processing steps	2 Std.	Fr 16.11.07	Fr 16.11.07	100%
13	2ND MILESTONE	0 Std.	Fr 16.11.07	Fr 16.11.07	100%
14	Datawarehouse Schema / Data Modelling	34 Std.	Fr 16.11.07	Fr 21.12.07	41%
15	Create new SQL Server Analysis Project	2 Std.	Fr 16.11.07	Fr 16.11.07	100%
16	Identify Fact and Dimensions within your data	8 Std.	Fr 16.11.07	Fr 23.11.07	100%
17	Create Views from the resulting destination table	8 Std.	Sa 24.11.07	Sa 01.12.07	50%
18	Use Dimension and Cube Wizard to build a cube from these views	8 Std.	Mi 05.12.07	Mi 12.12.07	0%
19	Evaluate and browse this Cube in the Cube Design	7 Std.	Fr 14.12.07	Mi 19.12.07	0%
20	Document processing steps	1 Std.	Fr 21.12.07	Fr 21.12.07	0%
21	3RD MILESTONE	0 Std.	Fr 21.12.07	Fr 21.12.07	0%
22	Frontend Design	34 Std.	Fr 16.11.07	Fr 21.12.07	25%
23	OLAP	17 Std.	Fr 16.11.07	Fr 30.11.07	0%
24	Build an OLAP Pivot Report using Excel	7 Std.	Fr 16.11.07	Fr 23.11.07	0%
25	Document processing steps	10 Std.	Fr 23.11.07	Fr 30.11.07	0%
26	Data Mining	34 Std.	Fr 16.11.07	Fr 21.12.07	38%
27	Create New Mining Structure using the Data Mining Wizard	3 Std.	Fr 16.11.07	Fr 16.11.07	100%
28	Use relational or data warehouse source for building the model	10 Std.	Sa 17.11.07	Fr 30.11.07	100%
29	Use the cluster algorithm for customer segmentation (k-means)	9 Std.	Fr 30.11.07	Fr 07.12.07	0%
30	Evaluate and Analyse the resulting Clusters	10 Std.	Fr 07.12.07	Sa 15.12.07	0%
31	Document your processing steps	2 Std.	Mi 19.12.07	Fr 21.12.07	0%
32	4TH MILESTONE	0 Std.	Fr 21.12.07	Fr 21.12.07	0%



Gantt Chart

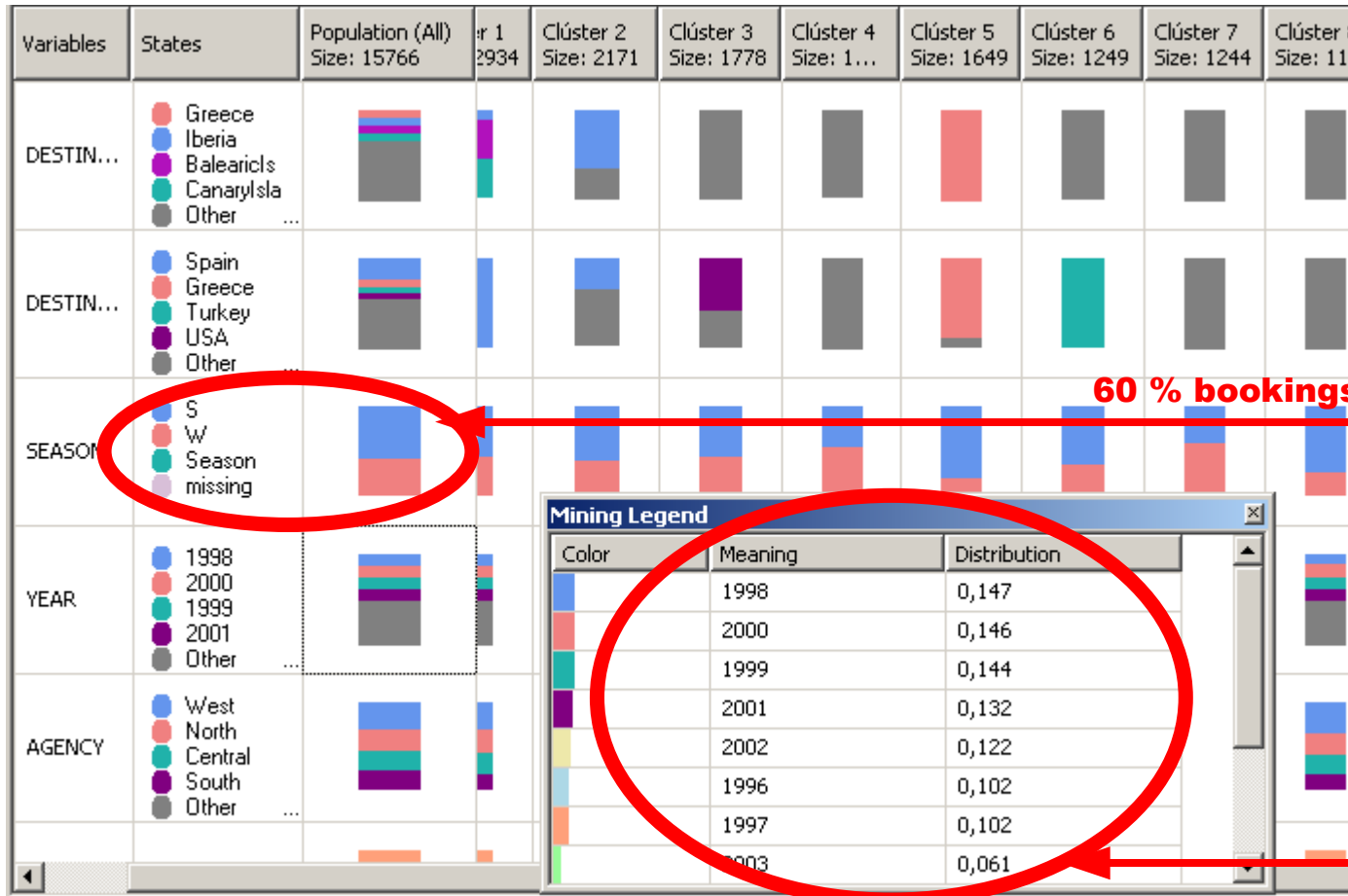


Data Mining: Cluster Diagram



**Strongest relation
within clusters 1 & 2**

Cluster Profiles: Quick Facts



60 % bookings for summer travels

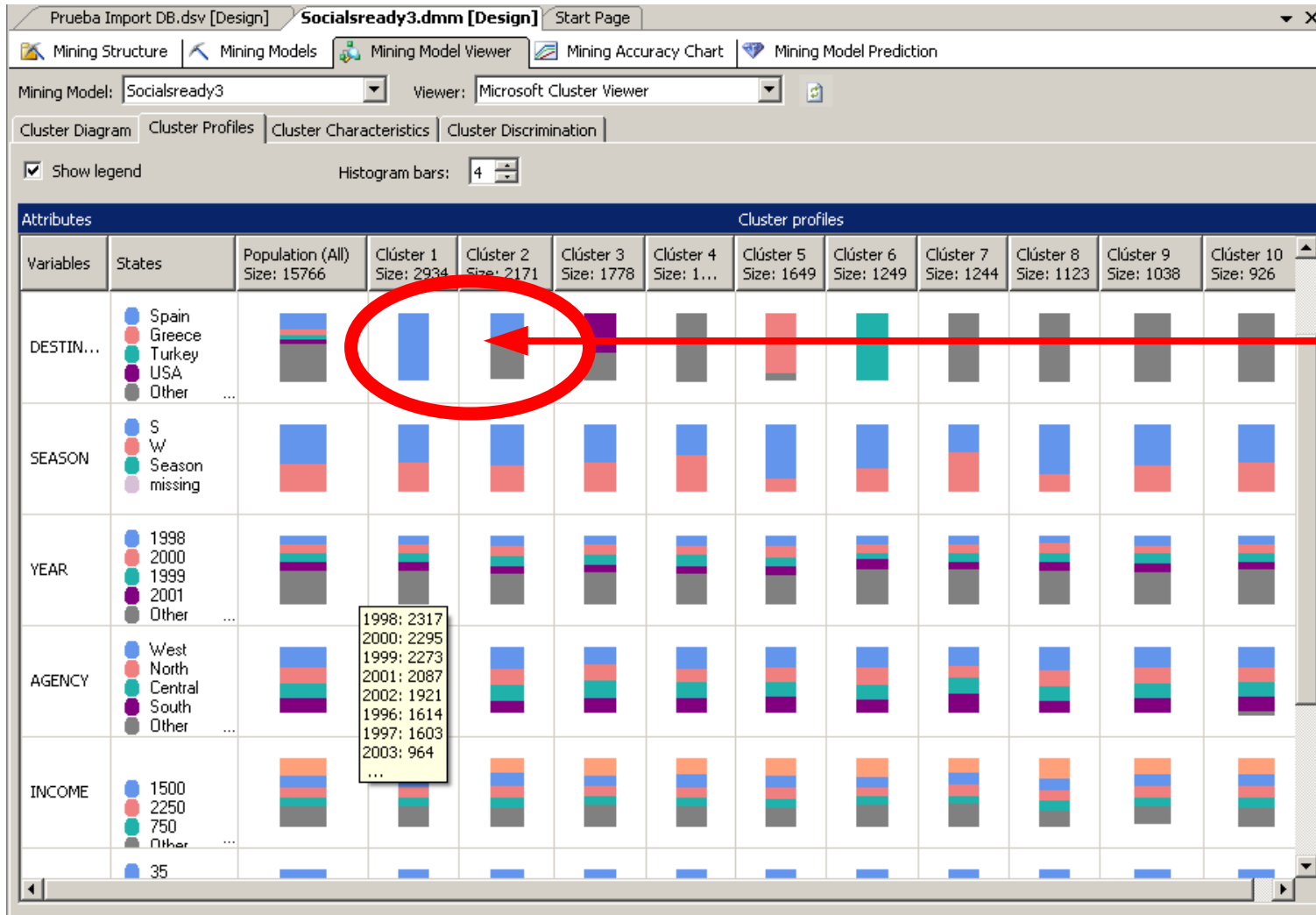
Decreasing amount
of YEAR data since
2001 -
should be evaluated
with OLAP
techniques

Cluster Profiles: Quick Facts

Distribution of destination regions within clusters:

- Cluster 1: Balearic and Canary Islands, and Iberia
- Cluster 2: Iberia and Tunisia
- Cluster 3: North America, Egypt
- Cluster 4: Germany, North Africa, Central America
- Cluster 5: Greece
- Cluster 6: Turkey
- Cluster 7: Asia
- Cluster 8: Italy
- Cluster 9: Dominican Republic, Central & North Europe
- Cluster 10: Central Africa, Great Britain

Cluster Profiles: Cluster 1 + 2



Preferred destination country: Spain

**Cluster 1:
Canary Islands,
Balears**

**Cluster 2:
Spain (Mainland),
Portugal and Tunesia**

Cluster Profiles: Cluster 1 + 2



Big percentage of missing values within these clusters! 28 %



Biggest age class 50, followed by 35 and 20 years old

Income between 750€ and 1500€

Until Next Report: Planned

- [17] Create Views from the resulting destination table
- [18] Use Dimension and Cube Wizard to build a cube from these views
- [24] Build an OLAP Pivot Report using Excel
- [29] Use the cluster algorithm for customer segmentation (k-means)

Until Next Report: Planned

Nr.		Task Name	Dauer	Anfang	Ende	% Abgeschlossen
14		Datawarehouse Schema / Data Modelling	34 Std.	Fr 16.11.07	Fr 21.12.07	41%
15	<input checked="" type="checkbox"/>	Create new SQL Server Analysis Project	2 Std.	Fr 16.11.07	Fr 16.11.07	100%
16	<input checked="" type="checkbox"/>	Identify Fact and Dimensions within your data	8 Std.	Fr 16.11.07	Fr 23.11.07	100%
17		Create Views from the resulting destination table	8 Std.	Sa 24.11.07	Sa 01.12.07	50%
18		Use Dimension and Cube Wizard to build a cube from these views	8 Std.	Mi 05.12.07	Mi 12.12.07	0%
19		Evaluate and browse this Cube in the Cube Design	7 Std.	Fr 14.12.07	Mi 19.12.07	0%
20		Document processing steps	1 Std.	Fr 21.12.07	Fr 21.12.07	0%
21		3RD MILESTONE	0 Std.	Fr 21.12.07	Fr 21.12.07	0%
22		Frontend Design	34 Std.	Fr 16.11.07	Fr 21.12.07	25%
23		OLAP	17 Std.	Fr 16.11.07	Fr 30.11.07	0%
24		Build an OLAP Pivot Report using Excel	7 Std.	Fr 16.11.07	Fr 23.11.07	0%
25		Document processing steps	10 Std.	Fr 23.11.07	Fr 30.11.07	0%
26		Data Mining	34 Std.	Fr 16.11.07	Fr 21.12.07	38%
27	<input checked="" type="checkbox"/>	Create New Mining Structure using the Data Mining Wizard	3 Std.	Fr 16.11.07	Fr 16.11.07	100%
28	<input checked="" type="checkbox"/>	Use relational or data warehouse source for building the model	10 Std.	Sa 17.11.07	Fr 30.11.07	100%
29		Use the cluster algorithm for customer segmentation (k-means)	9 Std.	Fr 30.11.07	Fr 07.12.07	0%
30		Evaluate and Analyse the resulting Clusters	10 Std.	Fr 07.12.07	Sa 15.12.07	0%
31		Document your processing steps	2 Std.	Mi 19.12.07	Fr 21.12.07	0%

Risks

- Bad quality of results
 - Description: Lack of verification methods within team
 - Action plan: Reviewing of each completed phase
 - Status: High
 - Responsible: project team

Risks

- Intercommunication within the project team
 - Description: Lack of communication
 - Action plan: Regular reporting E-Mails
 - Status: Average
 - Responsible: project leader
- Wrong time prediction
 - Description: Wrong scheduling, Spanish members only being able to contribute on weekends
 - Action plan: adding resources, project manager
 - Status: High
 - Responsible: whole team

QA Status

- Regular Reporting E-Mails
 - Keep team members on track
 - Double check of results
 - Allows additional improvements through members
- Constant updates to Time Plan
 - Flexible time allocation
- Backups
 - Comparison to former results
 - Security



... for your attention!

support@kyynel.biz