

Università degli Studi di Firenze

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA MAGISTRALE IN INFORMATICA (CLASSE LM-18)

ANNO ACCADEMICO 2012/2013

Progetto per il corso di Data Warehousing

Autori:

Tommaso PAPINI, 5537529

Riccardo BERNINI, 5435313



dicembre 2016

Indice

Introduzione	1
1 Modellazione concettuale	3
2 Progettazione logica	5
2.1 Schema di fatto	5
2.2 Schema a stella	6
2.3 Flat Table	6
3 Elaborazione dei dataset (ETL)	9
3.1 Trasformazione	10
3.2 Job	13
4 Costruzione e alimentazione del Data Warehouse	15
5 Alcune analisi	19
5.1 Totale Costi	19
5.1.1 Confronto: Totale Costo 2012/2013	20
5.1.2 Confronto: Totale Pagamenti/Crediti 2012/2013	20
5.1.3 Confronto: Totale Fatture/Note di Credito 2012/2013	21
5.2 TOP 10 Pagamenti per Aziende	22
5.2.1 TOP 10 Pagamenti per Aziende 2012	22
5.2.2 TOP 10 Pagamenti per Aziende 2013	24
5.2.3 Confronto: TOP 10 Pagamenti per Aziende 2012/2013	27
5.3 TOP 10 Pagamenti per Direzioni	28
5.3.1 TOP 10 Pagamenti per Direzioni 2012	29
5.3.2 TOP 10 Pagamenti per Direzioni 2013	30
5.3.3 Confronto: TOP 10 Pagamenti per Direzioni 2012/2013	31
5.4 BOTTOM 10 Pagamenti per Direzioni	32
5.4.1 BOTTOM 10 Pagamenti per Direzioni 2012	32
5.4.2 BOTTOM 10 Pagamenti per Direzioni 2013	33

5.4.3	Confronto: BOTTOM 10 Pagamenti per Direzioni 2012/2013 . . .	34
5.5	TOP 10 Accoppiate Direzione-Azienda	36
5.6	TOP 10 Anni di Pagamento per Aziende	38
5.7	Analisi per tipo di Atto di Impegno	39
5.7.1	<i>Consigli Comunali</i> negli anni	39
5.7.2	Confronto: CC-DD nel 2009-2013	40
Conclusioni		43
	Alcune considerazioni	43
	Altri progetti	44
Bibliografia		49

Elenco delle figure

1.1	Schema-concettuale.	3
1.2	Schema riconciliato.	4
2.1	Fact-Schema.	5
2.2	Star-Schema.	6
2.3	<i>Flat Table</i> finale.	7
3.1	Trasformazione Kettle, prima parte.	10
3.2	Trasformazione Kettle, seconda parte.	11
3.3	Trasformazione Kettle, terza parte.	11
3.4	Trasformazione Kettle, quarta parte.	11
3.5	Trasformazione Kettle, quinta parte.	12
3.6	Trasformazione Kettle, sesta parte.	12
3.7	<i>Job</i> di trasformazione.	13
3.8	Impostazioni di scheduling del job.	14
4.1	Schema <i>MOLAP</i>	16
5.1	Totale Costo 2012/2013, grafico.	20
5.2	Totale Costo 2012/2013, tabella.	20
5.3	Totale Pagamenti/Crediti 2012/2013, grafico.	21
5.4	Totale Pagamenti/Crediti 2012/2013, tabella.	21
5.5	Totale Fatture/Note di Credito 2012/2013, grafico.	22
5.6	Totale Fatture/Note di Credito 2012/2013, tabella.	22
5.7	TOP 10 Pagamenti per Aziende 2012, grafico.	23
5.8	TOP 10 Pagamenti per Aziende 2012, tabella.	23
5.9	TOP 10 Pagamenti per Aziende 2012, tabella disaggregata.	24
5.10	TOP 10 Pagamenti per Aziende 2013, grafico.	25
5.11	TOP 10 Pagamenti per Aziende 2013, tabella.	25
5.12	TOP 10 Pagamenti per Aziende 2013, tabella disaggregata.	26
5.13	TOP 10 Pagamenti per Aziende 2012/2013, grafico.	27

5.14	TOP 10 Pagamenti per Aziende 2012/2013, tabella.	27
5.15	TOP 10 Pagamenti per Aziende 2012/2013, tabella disaggregata.	28
5.16	TOP 10 Pagamenti per Direzioni 2012, grafico.	29
5.17	TOP 10 Pagamenti per Direzioni 2012, tabella.	29
5.18	TOP 10 Pagamenti per Direzioni 2013, grafico.	30
5.19	TOP 10 Pagamenti per Direzioni 2013, tabella.	30
5.20	TOP 10 Pagamenti per Direzioni 2012/2013, grafico.	31
5.21	TOP 10 Pagamenti per Direzioni 2012/2013, tabella.	31
5.22	BOTTOM 10 Pagamenti per Direzioni 2012, grafico.	32
5.23	BOTTOM 10 Pagamenti per Direzioni 2012, tabella.	33
5.24	BOTTOM 10 Pagamenti per Direzioni 2013, grafico.	33
5.25	BOTTOM 10 Pagamenti per Direzioni 2013, tabella.	34
5.26	BOTTOM 10 Pagamenti per Direzioni 2012/2013, grafico.	34
5.27	BOTTOM 10 Pagamenti per Direzioni 2012/2013, tabella.	35
5.28	TOP 10 Accoppiate Direzione-Azienda, grafico.	36
5.29	TOP 10 Accoppiate Direzione-Azienda, tabella.	36
5.30	TOP 10 Accoppiate Direzione-Azienda, tabella disaggregata.	37
5.31	TOP 10 Anni di Pagamento per Aziende, grafico.	38
5.32	TOP 10 Anni di Pagamento per Aziende, tabella.	38
5.33	TOP 10 Anni di Pagamento per Aziende, tabella disaggregata.	39
5.34	<i>Consigli Comunali</i> negli anni, grafico.	40
5.35	<i>Consigli Comunali</i> negli anni, tabella.	40
5.36	<i>Consigli Comunali e Decreti Dirigenziali</i> negli anni, grafico a linee.	41
5.37	<i>Consigli Comunali e Decreti Dirigenziali</i> negli anni, grafici a torta.	41
5.38	<i>Consigli Comunali e Decreti Dirigenziali</i> negli anni, tabella.	42
C.1	AQS: trasformazione <i>Kettle</i> , prima parte.	45
C.2	AQS: trasformazione <i>Kettle</i> , seconda parte.	45
C.3	AQS: trasformazione <i>Kettle</i> , terza parte.	45
C.4	AQS: trasformazione <i>Kettle</i> , quarta parte.	46
C.5	<i>Schema a stella</i> costruito sul file di log finale.	46
C.6	Durata media degli attacchi per protocollo.	47
C.7	Tempistiche dell'attacco <i>nmap</i> con opzione <i>-T</i>	47

Introduzione

Nel mondo delle basi di dati sono essenziali le infrastrutture tramite le quali si possono organizzare e memorizzare i dati, poiché a seconda del contesto di utilizzo può essere necessaria una maggiore efficienza, una maggiore ottimizzazione dello spazio o una maggiore flessibilità. In ogni caso il punto focale della ricerca in tale direzione rimane comunque l'oggetto di interesse, ovvero i dati stessi, e dove poterli reperire per poter popolare le strutture ottimizzate in base ai requisiti di utilizzo. Spesso capita che se un qualsiasi ente o azienda decide di orientare le proprie risorse nella creazione e nell'utilizzo di una base di dati, il reperimento degli stessi è a carico dell'azienda, come anche l'utilizzo e la fruizione: in generale, la proprietà del dato e tutti i diritti inerenti sono e rimangono dell'azienda che li ha generati o raccolti. Come è facile immaginare questo tipo di approccio può favorire l'azienda in termini di preferenza dell'utente in base alla qualità dei dati, ma limita enormemente il progresso ed il raffinamento dei dati, in quanto un'eventuale altra azienda (o un qualsiasi privato) che volesse ampliare o migliorare la qualità o il numero dei dati contenuti nella base di dati dovrebbe ripetere il passaggio (tipicamente oneroso) del reperimento delle informazioni, disperdendo tempo e risorse a spese del progresso innovativo.

Per questo motivo nasce l'approccio *Open Data*, che ha il proposito di rendere disponibili dati "pubblici" senza introdurre restrizioni sull'accesso o sull'utilizzo degli stessi. La migliore definizione di *dato aperto* risulta essere:

*Un contenuto o un dato si definisce **aperto** se chiunque è in grado di utilizzarlo, riutilizzarlo e ridistribuirlo, soggetto, al massimo, alla richiesta di attribuzione e condivisione allo stesso modo.*

Da qualche anno il Comune di Firenze ha adottato l'approccio *open data* rendendo disponibile sul sito del Comune una collezione di 431 *dataset* (ad oggi, 14 dicembre 2016) che possono essere consultati ed utilizzati liberamente da qualsiasi utente vi acceda. Allo scopo di incentivare il contributo degli utenti alle informazioni il Comune ha reso disponibile una grande varietà di *dataset* che solo in minima parte risultano essere completi da un punto di vista quantitativo.

Il proposito di questo elaborato consiste nell'organizzazione dei *dataset* relativi al bilancio comunale degli anni 2012 e 2013 in un unico modello multidimensionale, consentendo quindi tutta una serie di analisi su di essi, come ad esempio il calcolo delle spese totali compiute nei due anni, il calcolo delle spese compiute dai vari settori comunali e come queste si siano modificate nei due anni, la possibilità di ricavare quali sono i settori più o meno costosi o di ricavare a quali aziende sono stati assegnati i lavori più lunghi. Questi sono soltanto alcuni esempi delle analisi condotte, come ultimo passo di questo progetto, all'interno del Capitolo 5.

Tra i due *dataset* vi è una stretta relazione temporale (il secondo segue il primo) ed hanno la stessa struttura. Occorre tuttavia alterare in parte la loro struttura includendo nel *dataset* finale ulteriori dati ed informazioni derivate; alla fase preparatoria segue una fase di modellazione concettuale e l'organizzazione dei dataset in un unico modello di Data Warehouse multidimensionale; infine la struttura potrà essere alimentata e successivamente interrogata tramite una serie di *query* (richieste) che risulterebbero, in linea di massima, piuttosto "scomode" (in termini di efficienza computazionale) sui *dataset* originali.

Oltre ai classici strumenti di modellazione concettuale viene sperimentato l'approccio al problema tramite il framework di *Business Intelligence* **Pentaho** che, oltre alle funzionalità di *Data Mining*, *analisi OLAP* ed utilizzo di metadati per la gestione dei *Big Data*, consente anche di alimentare agilmente la struttura modellata fornendo un tool di *Data Integration*.

Modellazione concettuale

A partire dalla pseudo-narrazione derivante dalla descrizione, dagli obiettivi di questo progetto e dall’analisi dei *dataset* utilizzati si può ottenere uno schema concettuale che, attraverso una serie di entità e le relazioni tra esse, possa modellare dettagliatamente la realtà che ci proponiamo di analizzare.

Gli obiettivi da tener presente nella costruzione dello schema concettuale sono: da una parte la necessità di soddisfare i requisiti utente; dall’altra la necessità di consistenza dello schema concettuale con gli schemi delle sorgenti operazionali.

L’approccio metodologico alla progettazione concettuale utilizzato è l’approccio guidato dai dati: lo schema concettuale viene cioè definito in funzione della struttura delle sorgenti stesse, evitando così il complesso compito di stabilire il legame con esse a posteriori.

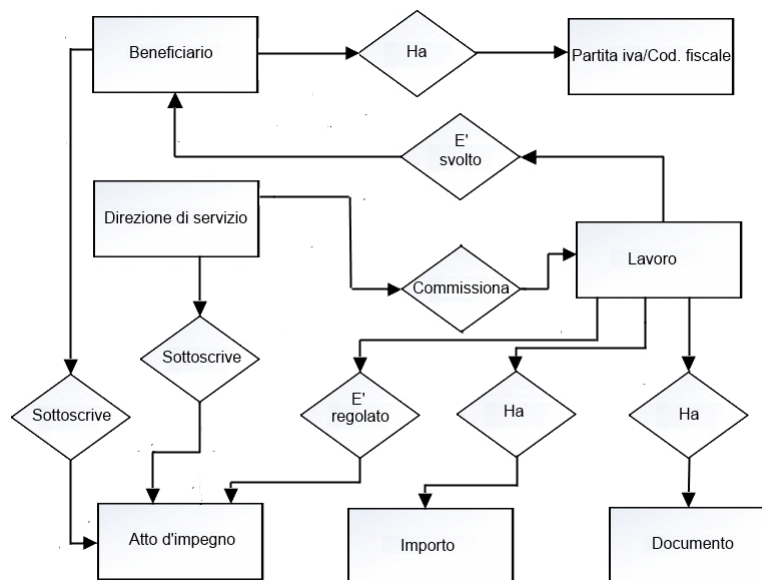


Figura 1.1: Schema-concettuale.

Una modellazione di questo tipo fornisce inoltre le basi per la costruzione di un'ontologia e una conseguente interrogazione dei dati da un punto di vista semantico.

Il passaggio successivo consiste nell'attribuzione di un **URI** (*Uniform Resource Identifier*) alle entità coinvolte nello schema, in modo tale da creare una correlazione tra i riferimenti del modello e la posizione reale dei dati.

Lo schema concettuale presentato può essere successivamente tradotto in uno schema entità-relazione più appropriato, che prende il nome di *Dimensional Fact Model*. Questo modello concettuale è specificatamente concepito per fungere da supporto alla progettazione di un *Data Warehouse* e può essere considerato come una specializzazione del modello multidimensionale.

Come si può osservare nella Figura 1.2, la rappresentazione generata dal DFM consiste in un insieme di schemi di fatto: gli elementi di base modellati sono i fatti stessi, le misure e le dimensioni.

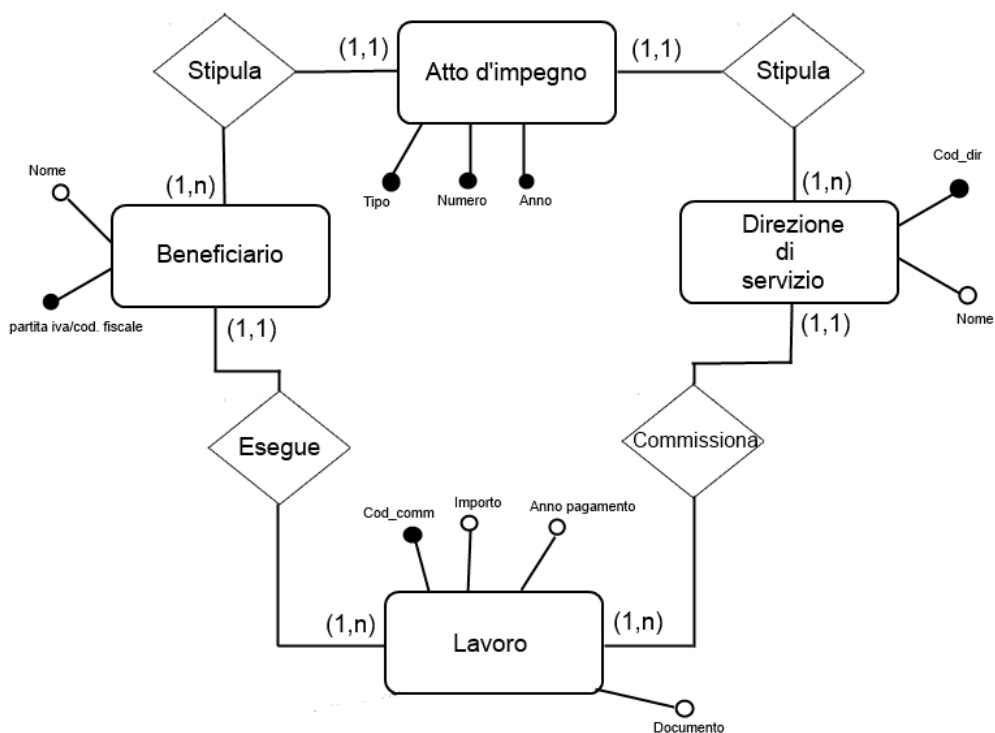


Figura 1.2: Schema riconciliato.

Progettazione logica

La progettazione logica permette di determinare lo schema logico del DW. Mentre durante la fase di progettazione concettuale viene ritagliata la porzione del dominio applicativo che dovrà essere considerato, con la fase di progettazione logica si determinano le strutture dati che andranno a rappresentare il DW stesso.

In questo scenario, l'obiettivo primario è quindi di massimizzare la velocità di reperimento dei dati.

2.1 Schema di fatto

La prima fase della progettazione logica consiste nel dare luogo allo schema che rappresenta l'albero degli attributi. Nello schema seguente (Figura 2.1) la parte centrale esprime il “fatto” (*fact*, in inglese), ovvero ciò a cui è volta la costruzione del modello. I campi del *fatto* rappresentano le misure a cui si fa riferimento per correlare le dimensioni dello schema, che sono ottenute dalla disaggregazione iniziale dei *dataset*.

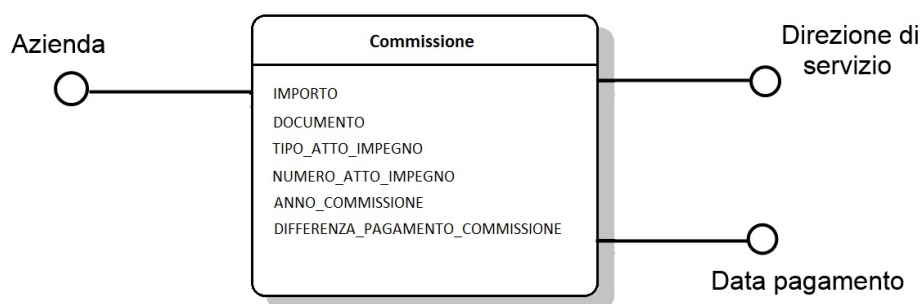


Figura 2.1: Fact-Schema.

Uno schema di fatto può essere modellato in ambito relazionale mediante uno schema a stella in cui la cosiddetta *fact table* (*tabella di fatto*) contiene tutte le misure e gli attributi descrittivi direttamente collegati al fatto e nella quale, per ogni gerarchia presente, viene creata una *dimension table* (*tabella dimensionale*) che ne contiene tutti gli attributi.

2.2 Schema a stella

Lo schema a stella rappresenta il rapporto tra le dimensioni del modello e la *fact table* (posta al centro della stella), che andrà a contenere le chiavi inter-dimensionali (indicate nella Figura 2.2 con un simbolo) le quali costituiscono i punti di contatto tra le dimensioni considerate. Questo schema a stella permette efficacemente di rappresentare un cubo dimensionale, incentrato sul fatto di interesse del processo decisionale o di analisi. Lo schema ed il cubo rappresentano quindi un insieme di eventi, descritti quantitativamente da misure numeriche. Ogni asse del cubo (ogni dimensione) rappresenta una possibile dimensione di analisi e ciascuna dimensione può essere vista a più livelli di dettaglio, individuati dagli attributi strutturati in gerarchie.

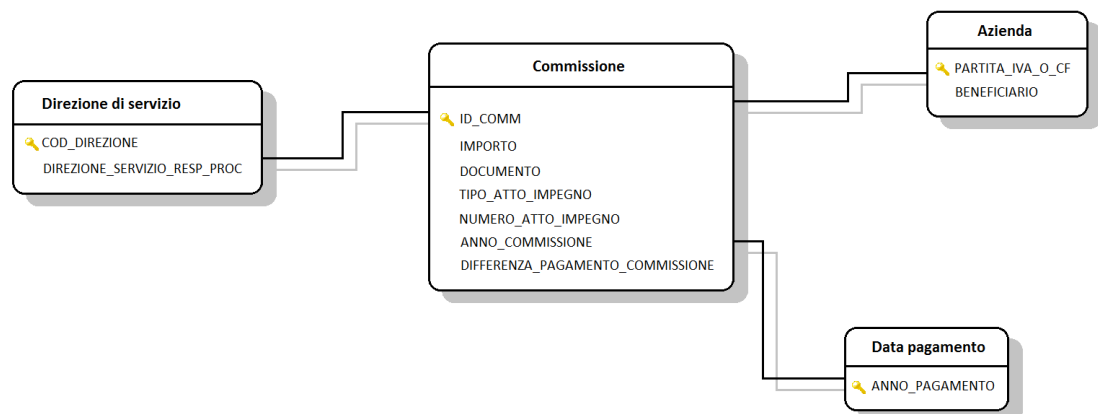


Figura 2.2: Star-Schema.

2.3 Flat Table

Fin'ora abbiamo parlato di modelli logici a supporto dei nostri dati, come schemi relazionali, concettuali e a stella. Facendo un passo avanti, entriamo quindi nell'ambito dell'implementazione effettiva di questi schemi logici, introducendo il concetto di *Flat Table*: una *Flat Table*, o *Flat-File Database* è un unico file di testo, contenente soltanto dati e delimitatori, che, privo di qualsiasi tipo di riferimento relazionale, rappresenta in concreto il nostro database.

Nel nostro caso, il file `.csv` ottenuto come risultato del processo di *Data Integration* sarà allora il supporto “fisico” per il nostro database. Questo file, denominato `Commissioni.csv`, sarà di fatto un unico, enorme file contenente sia i dati della *fact table* sia i dati di tutte le *dimension table* presenti nello schema di fatto. In questo modo viene eliminata ogni tipo di relazione (o se vogliamo, collegamento) tra i dati da analizzare, rendendo necessario più spazio per la memorizzazione ma migliorando notevolmente le tempistiche in fase di analisi.

In Figura 2.3 possiamo vedere le prime righe e colonne del file `Commissioni.csv` sopra citato. Per le trasformazioni di *Data Integration* che generano tale file si veda il Capitolo 3.

	A	B	C	D	E	F
1	ID_COMM	PARTITA_IVA_O_CF	BENEFICIARIO	COD_DIREZIONE	DIREZIONE_SERVIZIO_RESP_PROC	IMPORTO_DOC
2	1	5549710480	A PICCOLI PASSISOCIETA' COOPERATIVA SOCIALE A R.L.	65	SERVIZI ALL'INFANZIA	6340FAT
3	2	2249270485	A.A.P. DI VILLORESI ANDREA E FERRUZZI NEVA	48	SERV. GEST. E MANUT. (EX)	13FAT
4	3	2249270485	A.A.P. DI VILLORESI ANDREA E FERRUZZI NEVA	48	SERV. GEST. E MANUT. (EX)	16FAT
5	4	2249270485	A.A.P. DI VILLORESI ANDREA E FERRUZZI NEVA	48	SERV. GEST. E MANUT. (EX)	27FAT
6	5	1557920483	ABACO DI INNOCENTI E C S.R.L.	5	DIR. CORPO POLIZ. MUN.	405FAT
7	6	1783180480	ABATI LUIGI S.R.L.	40	SERV. SUPPORTO SCUOLA	10132FAT
8	7	1783180480	ABATI LUIGI S.R.L.	40	SERV. SUPPORTO SCUOLA	1117FAT
9	8	1783180480	ABATI LUIGI S.R.L.	40	SERV. SUPPORTO SCUOLA	6765FAT
10	9	1783180480	ABATI LUIGI S.R.L.	40	SERV. SUPPORTO SCUOLA	11731FAT
11	10	2616950248	ABITAT SRLSISTEMI INFORMATIVI TERRITORIALI	50	SERV. GEST. INFRASTR. TECNOL.	2487FAT
12	11	1883740647	ACCA SOFTWARE S.P.A.	7	DIR. NUOVE INFR. E MOBIL.	-18NOT.
13	12	1883740647	ACCA SOFTWARE S.P.A.	7	DIR. NUOVE INFR. E MOBIL.	620FAT
14	13	2179630484	ACCADEMIA EUROPEA DI FIRENZE	18	DIREZ. ISTRUZIONE	21106FAT
15	14	2179630484	ACCADEMIA EUROPEA DI FIRENZE	18	DIREZ. ISTRUZIONE	22481FAT
16	15	2179630484	ACCADEMIA EUROPEA DI FIRENZE	55	SERV. PROMOZ. ECON. TURISTIC	13370FAT
17	16	2179630484	ACCADEMIA EUROPEA DI FIRENZE	55	SERV. PROMOZ. ECON. TURISTIC	6685FAT
18	17	3931440485	ACCADEMIA MUSICALE DI FIRENZE	28	S. ATT. EDUC. E FORMAT.	600FAT
19	18	3931440485	ACCADEMIA MUSICALE DI FIRENZE	28	S. ATT. EDUC. E FORMAT.	8000FAT
20	19	2185260482	ACER GIARDINI	56	SERV. QUALITA' VERDE	89FAT
21	20	4566850287	ACERT DI PAOLO ZAMBUSI	5	DIR. CORPO POLIZ. MUN.	312FAT
22	21	3555130487	A.C.F. FIRENZE A.S.D.	68	SERVIZIO SPORT	445FAT
23	22	3555130487	A.C.F. FIRENZE A.S.D.	68	SERVIZIO SPORT	650FAT
24	23	3555130487	A.C.F. FIRENZE A.S.D.	68	SERVIZIO SPORT	650FAT
25	24	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	1041FAT
26	25	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	13163FAT
27	26	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	1322FAT
28	27	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	763FAT
29	28	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	2397FAT
30	29	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	704FAT
31	30	4094120484	ACS AUTOMATISMIDI TURCHI MONICA	48	SERV. GEST. E MANUT. (EX)	1098FAT
32	31	3358520967	ACTALIS S.P.A.	12	DIR. SIST. INFORMATIVI	529FAT
33	32	3358520967	ACTALIS S.P.A.	12	DIR. SIST. INFORMATIVI	6273FAT
34	33	1067020527	ADARTE ASSOCIAZIONE DANZA ARTETEATRO	28	S. ATT. EDUC. E FORMAT.	2649FAT
35	34	5439000489	Ad.D. Advertising DistributionADVERTISING DISTRIBUTION	43	SERV. BIBL. ARCH. EVENT	278FAT
36	35	890370372	ADS S.P.A.	12	DIR. SIST. INFORMATIVI	15125FAT
37	36	890370372	ADS S.P.A.	14	DIREZ. RISORSE UMANE	2197FAT
38	37	890370372	ADS S.P.A.	12	DIR. SIST. INFORMATIVI	90495FAT
39	38	890370372	ADS S.P.A.	14	DIREZ. RISORSE UMANE	700FAT
40	39	1080070158	AFFESSE MISURE SRI	5	DIR. CORPO POLIZ. MUN.	411FAT

Figura 2.3: *Flat Table* finale.

Capitolo 3

Elaborazione dei dataset (ETL)

Nello sviluppo di questo progetto sono stati usati i *dataset* contenenti il bilancio del 2012 e del 2013 del Comune di Firenze, reperibili sul sito degli *opendata* di Firenze¹.

Questi *dataset* contengono i dati principali delle fatture dei fornitori del Comune di Firenze per le quali, nel periodo che va dal 26 giugno 2012 (data di entrata in vigore di quanto disposto dall'art.18 del Decreto Sviluppo 2012) fino al 17 dicembre 2012, sono stati sia emessi che quietanzati i mandati di pagamento. Sempre con riferimento a tale articolo, sono stati selezionati gli elementi informativi pubblicati per ciascuna fattura.

Per quanto riguarda la struttura dei *dataset* essi sono strutturati allo stesso modo, con l'unica eccezione per il bilancio del 2012, che possiede un campo aggiuntivo identificante il numero di mandato (`NUMERO_MANDATO`).

Escluso l'eccezione appena introdotta la struttura generale è la seguente:

- `BENEFICIARIO`: Azienda o privato a cui è stato commissionato un mandato;
- `PARTITA_IVA_O_CF`: partita iva o codice fiscale del beneficiario;
- `IMPORTO`: importo pagato (o ricevuto) dal Comune;
- `ATTO_DI_IMPEGNO`: codice contenente il tipo, il numero e l'anno di stipula dell'atto d'impegno;
- `DIREZIONE_SERVIZIO_RESP_PROC`: Direzione di Servizio del Comune di Firenze che ha commissionato il lavoro;
- `DOCUMENTO`: il tipo di documento (fattura/nota di credito).

Lo scopo di questo lavoro è integrare tra loro i *dataset* relativi ai bilanci annuali in modo da poter compiere successivamente delle analisi che non siano ristrette ad un singolo anno, ma che possano confrontare anche i dati su più anni di bilancio. Permetteremo

¹<http://opendata.comune.fi.it/>

così anche un'analisi temporale dei flussi di pagamento e di credito del Comune di Firenze. Ad esempio cercheremo di integrare i dati in modo tale da permettere il recupero di informazioni come la lista delle aziende a cui sono stati commissionati lavori più lunghi o come variano gli investimenti nei vari settori del Comune di Firenze tra il 2012 ed il 2013.

Lo strumento utilizzato in questa prima fase è *Kettle*: *Kettle* è un tool di *Data Integration* che permette di specificare, a partire dalle sorgenti dati, una serie di operazioni che produrranno i dati aggregati (o derivati) in output. Sia gli input che gli output possono essere tabelle di Database vere e proprie, oppure, come nel nostro caso, un file *.csv*, che costituisce un insieme di dati tra loro anche ridondanti.

3.1 Trasformazione

Vedremo adesso nel dettaglio la trasformazione *Kettle* che è stata costruita ed utilizzata per questo progetto.

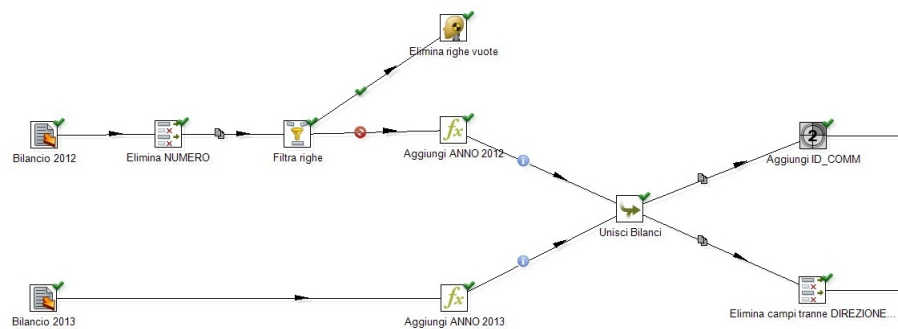


Figura 3.1: Trasformazione Kettle, prima parte.

In questa prima parte della trasformazione (Figura 3.1) si provvede innanzitutto a caricare le sorgenti dati dell'algoritmo collegate ai file dei vari *dataset* da integrare. Quindi si riportano i due *dataset* ad una struttura comune, eliminando dal bilancio 2012 il campo superfluo *NUMERO_MANDATO*, eliminando le eventuali righe vuote ed aggiungendo a ciascun bilancio il relativo anno di appartenenza (campo *ANNO_PAGAMENTO*). Infine i due flussi d'input vengono concatenati (funzione di *append*) l'uno di seguito all'altro. Si osservi che in questo caso, come per il resto della trasformazione, l'ordine di concatenazione dei vari flussi è indifferente, in quanto tra gli ultimi step della trasformazione verrà eseguito un ordinamento.

A partire dall'unico elenco di commissioni precedentemente generato si generano due flussi di esecuzione (Figura 3.2): nel primo si provvede ad aggiungere un identificativo per ogni commissione (ovvero per ogni riga presente), chiamato *ID_COMM*, e successivamente ad ordinare il tutto rispetto alla direzione di servizio (vincolo necessario per eseguire

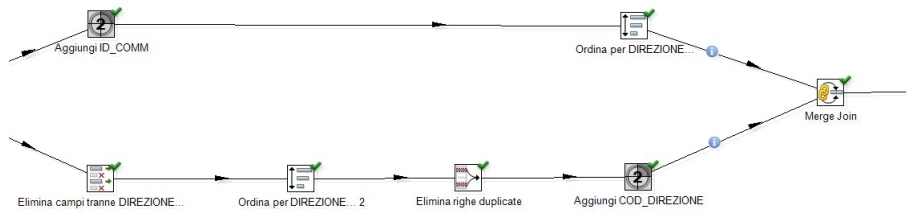


Figura 3.2: Trasformazione Kettle, seconda parte.

un merge-join); nel secondo flusso generato, invece, si considerano le sole direzioni di servizio, eliminando ogni altro campo, ed eliminando le voci duplicate (quindi mantenendo una riga per ogni direzione distinta), per poter poi aggiungere un identificativo (COD_DIREZIONE) ad ognuna di esse. Infine tramite l'utilizzo di un merge-join si provvede ad unire (e non concatenare, come nel caso precedente) i due flussi, così da avere al termine di questi passi per ogni commissione un proprio identificativo e oltre al campo con il nome della direzione di servizio sarà presente anche il codice corrispettivo.

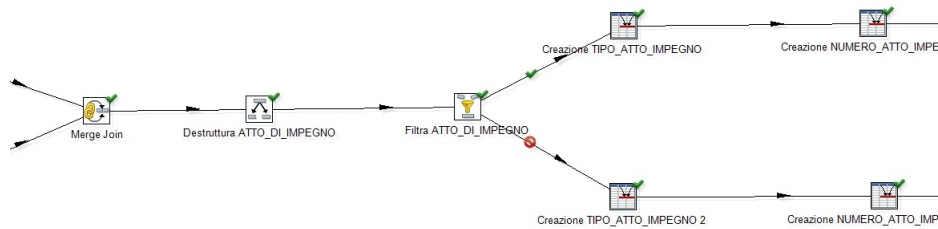


Figura 3.3: Trasformazione Kettle, terza parte.

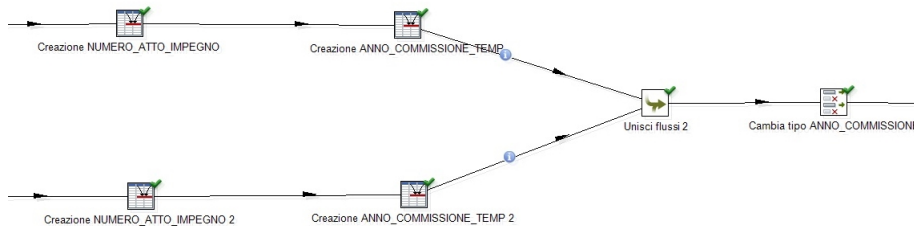


Figura 3.4: Trasformazione Kettle, quarta parte.

I passi presenti nelle Figure 3.3 e 3.4 hanno come obiettivo quello di destrutturare il campo ATTO_DI_IMPEGNO, nel quale è tenuto traccia dell'anno di commissione del lavoro, che risulterà utile in fase di analisi. Come si può intuire osservando le figure, l'atto d'impegno è destrutturato nei campi:

- **TIPO_ATTO_IMPEGNO**: suddiviso essenzialmente in *CC* (Consiglio Comunale) e *DD* (Decreto, o Determina, Dirigenziale), indica a che livello è stata deliberata la commissione;
- **NUMERO_ATTO_IMPEGNO**: identifica univocamente gli atti d'impegno (si osservi che più commissioni possono riferirsi allo stesso atto d'impegno);
- **ANNO_DI_COMMISSIONE_TEMP**: campo temporaneo in cui vengono registrate le ultime due cifre dell'anno di commissione.

Al termine di questi passi di destrutturazione vengono riuniti i flussi e viene associato un tipo intero al campo **ANNO_DI_COMMISSIONE_TEMP**.

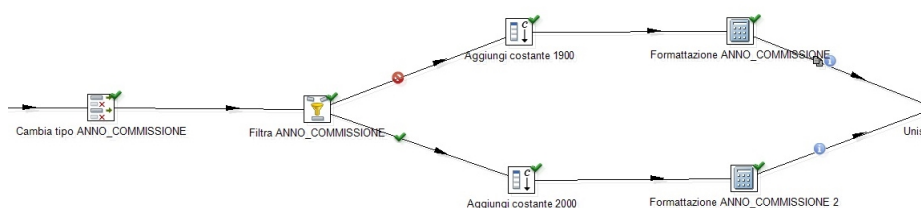


Figura 3.5: Trasformazione Kettle, quinta parte.

I passi rappresentati nella Figura 3.5 hanno invece come obiettivo quello di formattare nel modo corretto l'anno di commissione estratto dall'atto d'impegno: si provvede innanzitutto a discriminare, tramite un filtro, che l'anno sia relativo al 1900 o al 2000 (alcuni atti d'impegno risalgono a più di vent'anni fa!); quindi si generano due flussi di esecuzione, che provvedono ad aggiungere le costanti 1900 o 2000 al campo **ANNO_DI_COMMISSIONE_TEMP**, generando il nuovo campo **ANNO_DI_COMMISSIONE**; infine si provvede a riunire due flussi.

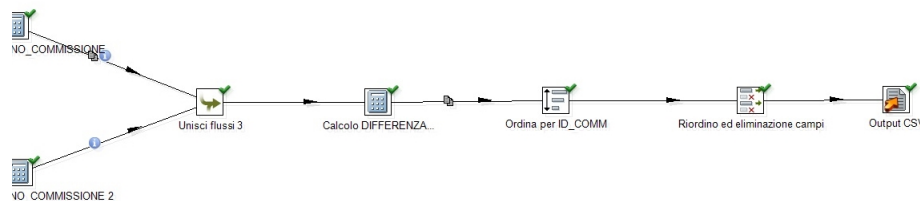


Figura 3.6: Trasformazione Kettle, sesta parte.

Gli ultimi passi della trasformazione (rappresentati in Figura 3.6) hanno come obiettivo quello di aggiungere un campo derivato, anch'esso necessario in fase d'analisi. Il campo in questione è **DIFFERENZA_PAGAMENTO_COMMISSIONE**, calcolato come differenza tra l'anno di pagamento (associato al bilancio stesso) e l'anno di commissione (precedentemente estratto dall'atto d'impegno).

Le ultime operazioni provvedono ad ordinare tutte le commissioni, ad eliminare i campi non necessari all'analisi e riordinare tra loro quelli utili.

Infine viene generato l'output della trasformazione, `Commissioni.csv`, che contiene tutte le commissioni registrate nei due anni e presenta in quest'ordine i seguenti campi:

- ID_COMM
- PARTITA_IVA_O_CF
- BENEFICIARIO
- COD_DIREZIONE
- DIREZIONE_SERVIZIO_RESP_PROC
- IMPORTO
- DOCUMENTO
- TIPO_ATTO_IMPEGNO
- NUMERO_ATTO_IMPEGNO
- ANNO_COMMISSIONE
- ANNO_PAGAMENTO
- DIFFERENZA_PAGAMENTO_COMMISSIONE

3.2 Job

La trasformazione presentata nella sezione precedente è quella che sta alla base del *job* che verrà poi eseguito anno per anno per integrare i nuovi *dataset* ai precedenti.

In generale, infatti, ogni volta che verrà rilasciato un nuovo *dataset* sarà necessario appendere al file `.csv`, contenente i dati aggregati dei bilanci precedenti, il nuovo bilancio opportunamente modificato, per avere una struttura coerente con le precedenti.

Il *job* di trasformazione è il seguente:

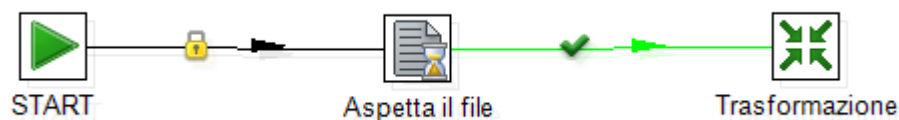
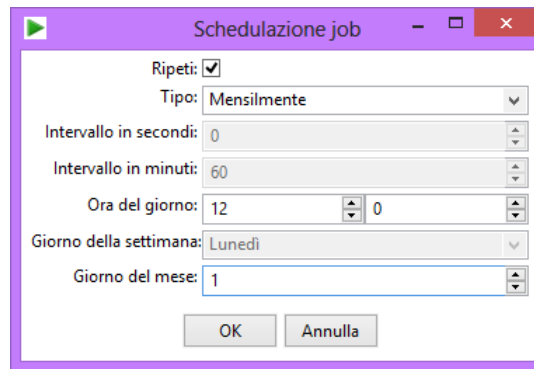


Figura 3.7: *Job* di trasformazione.



The image shows a Windows-style dialog box titled "Schedulazione job". It has a green play button icon in the top-left corner. The dialog contains several configuration options: a checked checkbox labeled "Ripeti:", a dropdown menu labeled "Tipo:" with "Mensilmente" selected, a numeric input field for "Intervallo in secondi:" set to 0, another numeric input field for "Intervallo in minuti:" set to 60, a time selection control for "Ora del giorno:" showing 12 and 0, a dropdown menu for "Giorno della settimana:" with "Lunedì" selected, and a numeric input field for "Giorno del mese:" set to 1. At the bottom, there are two buttons: "OK" and "Annulla".

Figura 3.8: Impostazioni di scheduling del job.

Dalla Figura 3.7 deduciamo subito la semplicità del *job*: esso è formato da uno step *start*, che definisce il punto di inizio e le impostazioni di scheduling, uno step di attesa, che blocca il flusso finché non è presente il *.csv* del nuovo bilancio da integrare, ed uno step *trasformazione*, che è esattamente la trasformazione *Kettle* precedentemente descritta in questo capitolo, opportunamente rivista per gestire un qualsiasi file di bilancio. Nella Figura 3.8 vediamo invece il dettaglio dello step *start*, dove vengono definite le impostazioni di scheduling, in modo da eseguire l'integrazione dei nuovi dati ad intervalli prefissati ed in modo del tutto automatico.

In questa sezione abbiamo descritto come produrre una sorgente dati singola, dettagliata, esauriente e di alta qualità che possa alimentare il DW a partire da dati operazionali reperiti all'interno di *dataset*.

Le operazioni svolte, o che è possibile svolgere, tramite *Kettle* sono estrazione, pulitura, trasformazione e caricamento. Nel nostro caso abbiamo semplicemente aggregato, estratto e trasformato alcune informazioni, come ad esempio la suddivisione del campo *ATTO_DI_IMPEGNO* o l'introduzione del nuovo campo *ANNO_PAGAMENTO*. Non è stato necessario infatti normalizzare, standardizzare o correggere in quanto i *dataset* messi a disposizione dal Comune di Firenze risultavano in quest'ottica già puliti e correttamente trasformati.

Osserviamo inoltre che l'output a seguito dell'utilizzo di questo strumento può essere visto come un *Refresh* per popolare il DW in un primo momento, oppure può essere interpretato come un *Update* se utilizzando il *job* presentato viene aggiunto alla sorgente dati finale solo il nuovo *dataset* appena rilasciato.

Capitolo 4

Costruzione e alimentazione del Data Warehouse

La tabella `Commissioni` serve per poter popolare il Data Warehouse (DW): per ogni riga di tale tabella si possono individuare le misure d'interesse, come ad esempio `IMPORTO` o `ID_COMM`, e le dimensioni di analisi, ad esempio `BENEFICIARIO` o `ANNO_PAGAMENTO`.

Pentaho mette a disposizione uno strumento per fare *On-Line Analytical Processing (OLAP)*, permettendo di creare un DW importando dati da database o da file `.csv`. Questo strumento è reso disponibile tramite un'interfaccia web, dalla quale è possibile:

- importare nuovi dati sorgenti;
- modificare i modelli dei dati importati;
- svolgere analisi;
- creare report.

È necessario fare una considerazione preliminare: l'elaborazione del file `Commissioni.csv`, realizzato nei passaggi precedenti, è computazionalmente onerosa. In uno scenario di lavoro reale il *dataset* dovrebbe risiedere in un database o in una struttura ad accesso rapido. In questo lavoro è stato scelto di usare un file `.csv` per praticità e semplicità di uso, in quanto l'enfasi del progetto è rivolta verso il DW e le tecniche di analisi, piuttosto che sul relazionale.

A partire dalla vista ottenuta considerando tutto il *dataset* `Commissioni` possiamo creare il nostro cubo con Pentaho semplicemente selezionando a quali dimensioni e misure siamo interessati; su queste sarà poi possibile utilizzare tecniche per ridurre la quantità di dati ed ottenere informazioni utili, ad esempio per *restrizione* o *aggregazione*. Con *restrizione* s'intende il ritagliare una porzione del cubo circoscrivendo il campo d'analisi

(*selezioni e/o proiezioni*), mentre l'*aggregazione* consiste nel raggruppare una o più misure rispetto ad una dimensione, ad un livello di dettaglio inferiore. Le principali *funzioni d'aggregazione* utilizzate sono le seguenti:

- **SUM**: la misura aggregata è calcolata come somma dei singoli valori. Questa funzione viene utilizzata sul campo **IMPORTO** per calcolare costi e spese aggregate secondo certi criteri.
- **COUNT**: viene contato il numero di righe che hanno un valore nel campo corrispondente (da non confondere con **COUNT DISTINCT** che conta soltanto il numero di valori distinti). Questa funzione è utilizzata con il campo **ID_COMM** per contare quante commissioni corrispondono ai criteri specificati nel filtro di query.
- **MAX**: la misura aggregata è il valore massimo tra quelli che corrispondono a certi criteri. Questa funzione viene utilizzata sul campo **DIFFERENZA_PAGAMENTO_COMMISSIONE**, in modo che venga restituita la massima differenza di anni presente per una certa Azienda.

Oltre alle funzioni sopra citate, Pentaho mette a disposizione anche **MIN**, che restituisce il minimo, e **AVERAGE**, che restituisce la media dei singoli valori.

La vista di analisi resa disponibile dalla tabella **Commissioni**, una volta selezionate le dimensioni e le misure d'interesse, è costituita da una rappresentazione navigabile del cubo *OLAP* in cui è possibile fare operazioni di *drill down* e *drill up* sulle dimensioni stesse. Quindi la costruzione del Data Warehouse, utilizzando Pentaho, è estremamente facile. Una volta individuato il fatto d'interesse occorre solo specificare mediante un'interfaccia web quali sono le dimensioni e le misure che lo identificano e lo caratterizzano.

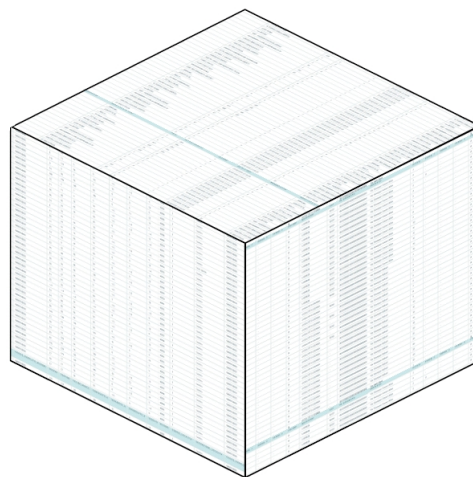


Figura 4.1: Schema *MOLAP*.

L'alimentazione del DW consiste invece nel mantenere aggiornato il Data Warehouse nel tempo (in genere ad intervalli stabiliti). Per alimentare il Data Warehouse occorre utilizzare il *job* descritto in Sezione 3.2. Quindi dopo un primo caricamento, verrà eseguito il *job* ad intervalli regolari (ad esempio ogni anno) con input il *dataset* del bilancio annuale, così da inserire nel Data Warehouse dati sempre aggiornati e permettere analisi più dettagliate ed aggiornate. L'unico requisito richiesto è che i *dataset* dei bilanci annuali dovranno essere già stati ripuliti e validati, così da richiedere a Pentaho di eseguire le trasformazioni necessarie per produrre una struttura compatibile con le precedenti.

Capitolo 5

Alcune analisi

In quest'ultimo capitolo ci occuperemo, finalmente, di analizzare i dati presenti nel DW costruito. Le analisi che proponiamo sono solo alcune delle possibili analisi effettuabili sui bilanci del Comune di Firenze, ma sono piuttosto significative, in quanto mostrano come, da due bilanci separati, si possano ottenere misure e considerazioni non immediate, esaltando la potenza degli strumenti di Data Warehouse.

Le analisi presenti in questo capitolo saranno sia *qualitative* che *quantitative*: ci potremo chiedere infatti se il Comune di Firenze ha investito o meno in un certo settore oppure quanti euro sono costati i servizi di una certa Azienda al Comune. Inoltre, alcune delle analisi qui presentate sono anche di carattere *comparativo*, ovvero mettono in luce certi aspetti che intercorrono tra i due bilanci (quindi di tipo *temporale*).

Tutte le analisi, oltre ad una breve descrizione ed alcune considerazioni, sono costituite da un grafico (a barre o a linee) e da una tabella, che riporta i valori numerici del grafico per una miglior lettura. In alcuni casi (solitamente nelle analisi che coinvolgono aziende) sarà presente anche una tabella disaggregata, che mostrerà più nel dettaglio alcuni campi coinvolti nell'analisi. Ad esempio si mostrerà spesso in queste tabelle disaggregate il nome (o i nomi) di un'azienda associato ad una certa partita iva, per una miglior comprensione dell'analisi.

5.1 Totale Costi

In questa prima sezione vedremo alcune analisi molto semplici ma significative, in quanto possono già darci un'idea delle potenzialità degli strumenti di DW, permettendoci di trarre anche le prime conclusioni.

Vedremo, in particolare, alcune analisi di tipo totale che coinvolgano i pagamenti effettuati.

5.1.1 Confronto: Totale Costo 2012/2013

La prima di queste analisi consiste in un confronto tra il costo totale (calcolato come pagamenti+crediti) dell'anno 2012 e del 2013.

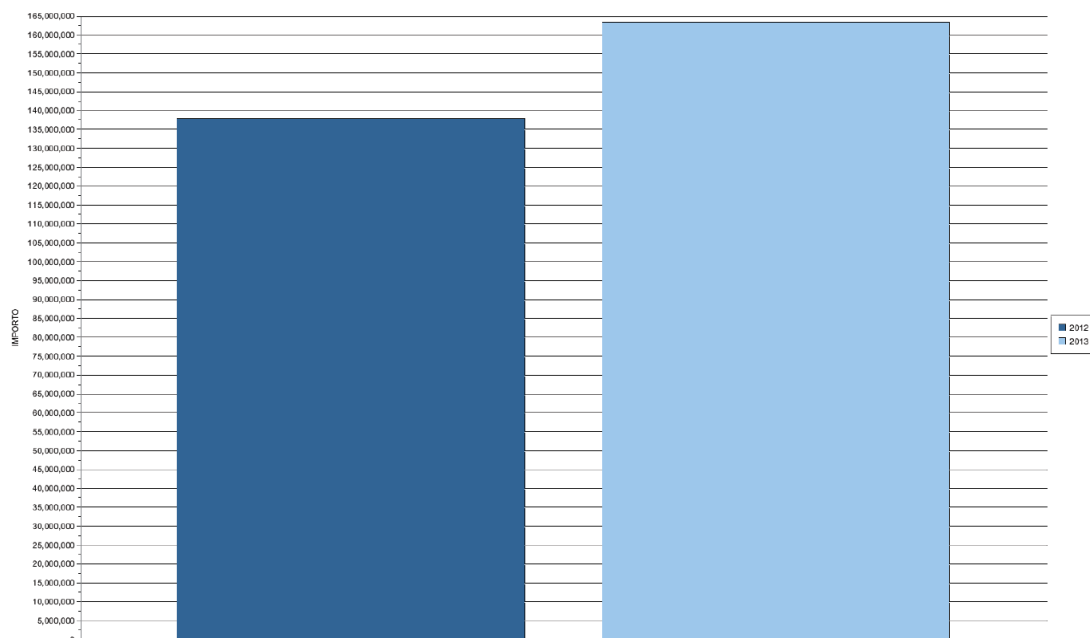


Figura 5.1: Totale Costo 2012/2013, grafico.

ANNO PAGAMENTO	
2012	2013
IMPORTO	IMPORTO
137947908	163481601

Figura 5.2: Totale Costo 2012/2013, tabella.

Nonostante il bilancio 2013 non sia definitivo, vediamo come quest'ultimo superi già il bilancio 2012 in termini di costo di circa 25 milioni di euro. Questa disparità può essere spiegata col fatto che il bilancio 2012 è, a sua volta, non completo, in quanto, come già accennato nel Capitolo 3, esso inizia a partire dal 26 giugno 2012.

5.1.2 Confronto: Totale Pagamenti/Crediti 2012/2013

In questa analisi partiamo dalla precedente ed entriamo nel dettaglio di Pagamenti e Crediti, disaggregando rispetto al campo DOCUMENTO, che indica se una commissione è una *fattura* o una *nota di credito*.

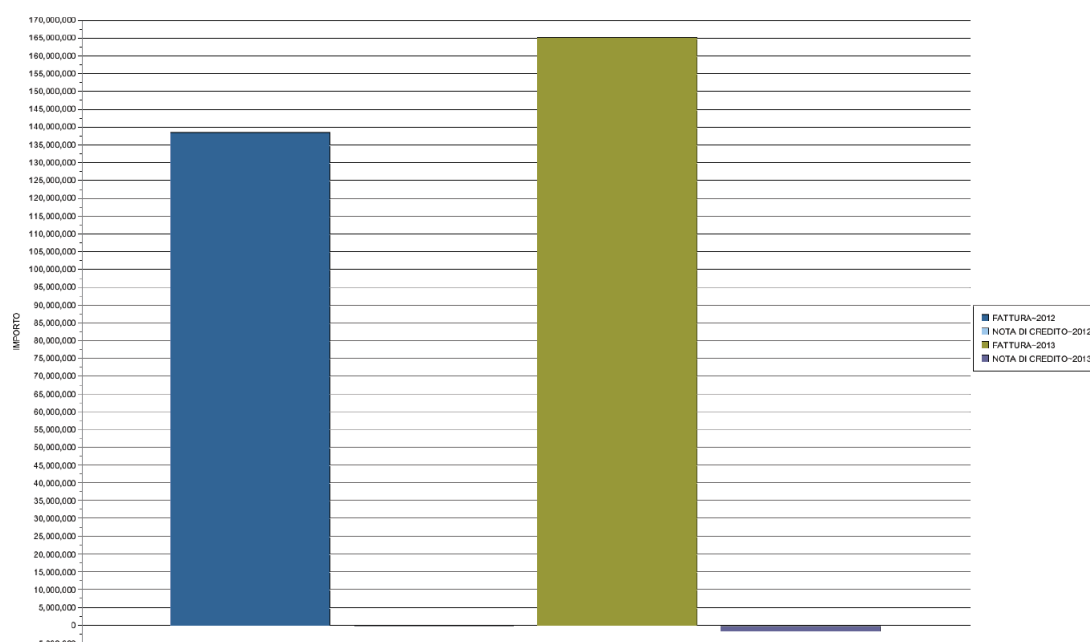


Figura 5.3: Totale Pagamenti/Crediti 2012/2013, grafico.

ANNO PAGAMENTO			
DOCUMENTO			
2012		2013	
FATTURA	NOTA DI CREDITO	FATTURA	NOTA DI CREDITO
IMPORTO	IMPORTO	IMPORTO	IMPORTO
138389342	-44143	165224948	-1743347

Figura 5.4: Totale Pagamenti/Crediti 2012/2013, tabella.

Come nel caso precedente, notiamo che il bilancio 2013 presenta valori più alti (in valore assoluto), sia per Pagamenti che per Crediti, probabilmente per la stessa considerazione fatta precedentemente.

5.1.3 Confronto: Totale Fatture/Note di Credito 2012/2013

Per avere una conferma delle supposizioni fatte nelle ultime due analisi, andiamo ad osservare quanti mandati sono stati commissionati negli anni 2012/2013, mantenendo la distinzione tra fatture e note di credito.

Come ci aspettavamo, si nota che nel 2013 sono stati commissionati più lavori (sia a credito che non), probabilmente proprio a causa delle mancanze presenti nel bilancio 2012 prima menzionate.

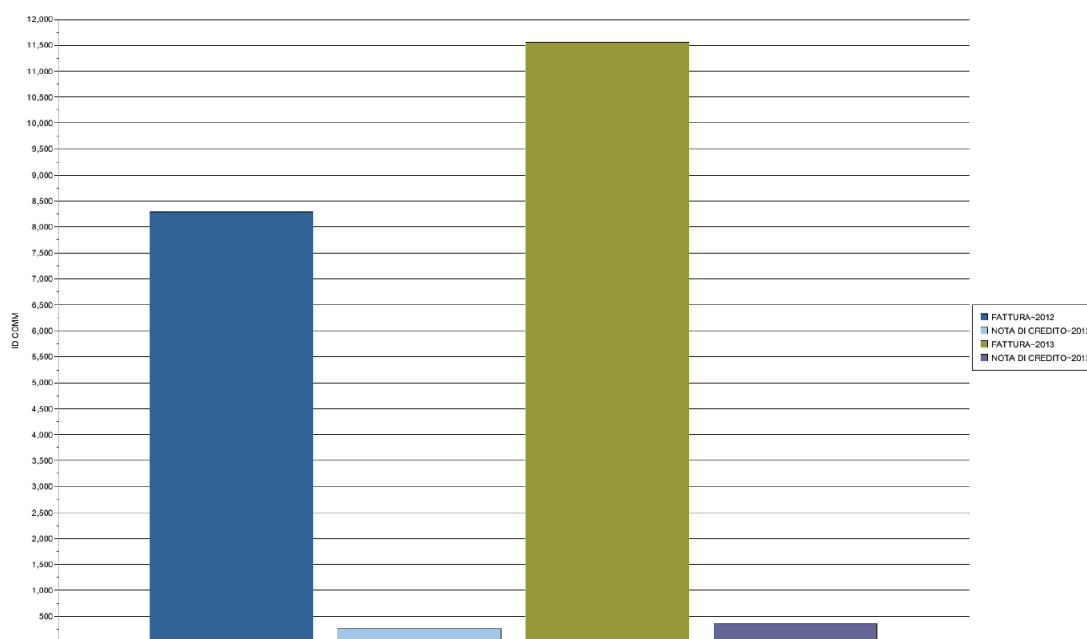


Figura 5.5: Totale Fatture/Note di Credito 2012/2013, grafico.

ANNO PAGAMENTO			
DOCUMENTO			
2012		2013	
FATTURA	NOTA DI CREDITO	FATTURA	NOTA DI CREDITO
ID COMM	ID COMM	ID COMM	ID COMM
8294	266	11568	367

Figura 5.6: Totale Fatture/Note di Credito 2012/2013, tabella.

5.2 TOP 10 Pagamenti per Aziende

Le analisi di questa sezione saranno orientate ad individuare le Aziende più pagate dal Comune di Firenze: vedremo innanzitutto la Top-10 delle Aziende più pagate nel 2012, quindi nel 2013 ed infine verranno confrontati i pagamenti tra 2012 e 2013 per le Aziende relative alla prima di queste analisi.

5.2.1 TOP 10 Pagamenti per Aziende 2012

Vediamo in questa analisi quali sono state le 10 Aziende più pagate dal Comune di Firenze durante l'anno 2012.

Si nota subito l'enorme disparità che intercorre tra la prima Azienda, ovvero la *Quadrifoglio s.p.a.*, e tutte le altre presenti nella Top-10.

La *Quadrifoglio s.p.a.* è un'Azienda di Firenze che si occupa di servizi ambientali nel

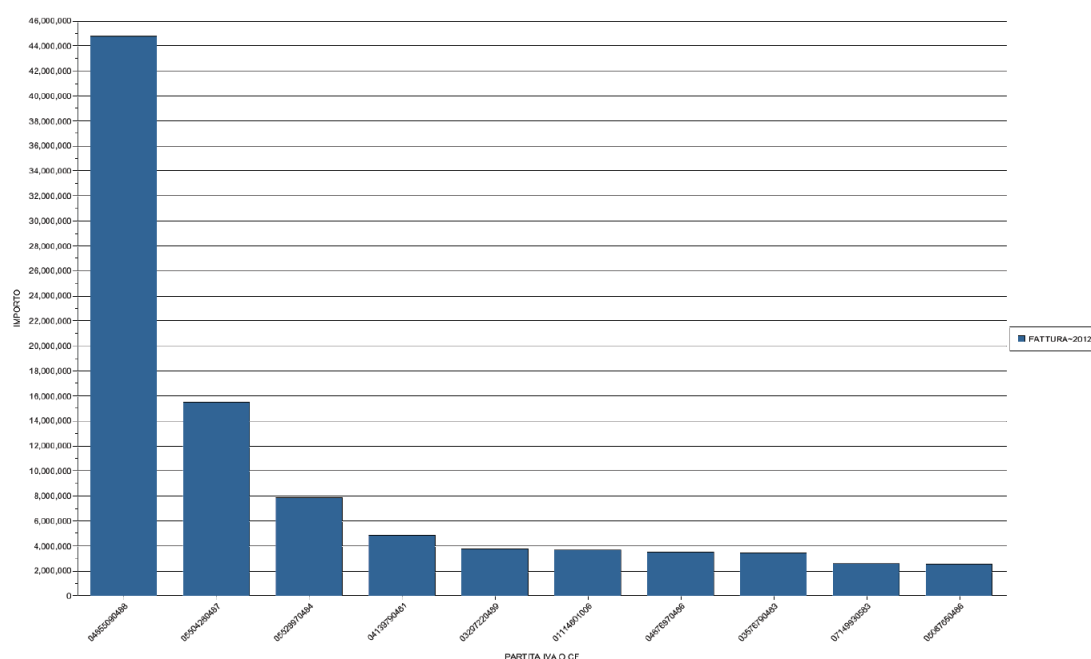


Figura 5.7: TOP 10 Pagamenti per Aziende 2012, grafico.

PARTITA IVA O CF	ANNO PAGAMENTO
	DOCUMENTO
	2012
	FATTURA
	IMPORTO
04855090488	44789799
05504280487	15515294
05529970484	7855174
04139790481	4880540
03297220489	3798317
01114601006	3677300
04876970486	3539871
03576790483	3443045
07149930583	2586918
05087650486	2538400

Figura 5.8: TOP 10 Pagamenti per Aziende 2012, tabella.

Comune di Firenze e nei comuni limitrofi, offrendo servizi come raccolta e smaltimento rifiuti, in primo luogo, ma anche come disinfestazione (zanzare, topi, ecc...) o centro di analisi e ricerca. È quindi un'Azienda molto importante, ormai onnipresente nell'ambiente fiorentino, che giustifica tutte le spese osservate.

Com'è facile aspettarsi, al secondo e terzo posto troviamo altre due Aziende onnipresenti nello scenario fiorentino: *ATAF&Linea S.C.R.L.* e *Tram di Firenze*, che si occupano, rispettivamente, del servizio di autobus e della tramvia fiorentina. Tra le altre Aziende, troviamo altri nomi ben conosciuti come *Consorzio Zenit*, che si occupa di interventi educativi, formativi e sanitari, e le *Poste Italiane*.

		ANNO PAGAMENTO
		DOCUMENTO
		2012
		FATTURA
PARTITA IVA O CF	BENEFICIARIO	IMPORTO
01114601006	POSTE ITALIANE S.P.A. - A.T.G.C.	3471666
	POSTE ITALIANEC. CONTR 30006059-048	170956
	POSTE ITALIANE S.P.A.	23329
	POSTE ITALIANE S.P.AD.C.G.C./S.C.A.	9680
	POSTE ITALIANE S.P.A.C. CONTR. 30006059-044	1669
03297220489	ASP FIRENZE MONTEDOMINEX S.AMBROGIO/FULIGNO/BIGALLO	3798317
03576790483	SILFI S.P.A.	3443045
04139790481	R.T.I.ZENIT - DI VITTORIO COOP.ZENIT-COOP. VITTORIO	3283471
	ATI CONSORZIO ZENIT E COOP. ELFO COOPERATIVA DI VITTORIO	511289
	CONSORZIO ZENIT SCRL RSA LORENZO	157389
	CONSORZIO ZENIT CENTRO G.PINOCCHIO	138634
	CONSORZIO ZENIT C.P.A. GIANBURRASCA	133479
	CONSORZIO ZENIT RSA CHIOCCIOLA	130224
	CONSORZIO ZENIT COM.EDUC.DEMIDOFF	124852
	CONSORZIO ZENIT R.S.A. S. GIUSEPPE	79849
	CONSORZIO ZENIT C.D. LA TENDA	72426
	CONSORZIO ZENIT COM.EDUC.S.FILIPPO	71887
	CONSORZIO ZENIT R.S.A. LA TENDA	61833
	CONSORZIO ZENIT R.S.A. DEMIDOFF	43396
	CONSORZIO ZENIT C.D. ALZHEIMER	37338
	CONSORZIO ZENIT C.D. DEMIDOFF	34473
04855090488	QUADRIFOGLIO SPA	44789799
04876970486	RTC CO&SO E CONSORZIO FABRICA S.C.	1368756
	CO & SO FIRENZE CONS.COOP.SOLIDARIEEX CO & SO CONS.COOP.SOLIDARIETA'	1060524
	A.T.I.CO&SO-COOPCULTURE- COPLAT-IFNET-COOP.ARCHEOLOGIA	528218
	A.T.I.CO&SO METROPOLI	422278
	CO & SO FIRENZE CONS. COOP. SOL.	160095
05087650486	SERVIZI ALLA STRADA S.P.A.	2538400
05504280487	ATAF&LINEA S.C.R.L.	15515294
05529970484	TRAM DI FIRENZE	7855174
07149930583	RTI COFELY-ASTRIM-OPEROSA-COLSER IPRAMS	2586918

Figura 5.9: TOP 10 Pagamenti per Aziende 2012, tabella disaggregata.

5.2.2 TOP 10 Pagamenti per Aziende 2013

Ripetiamo qui l'analisi precedente sull'anno 2013 anziché 2012.

Al primo e secondo posto troviamo ancora la *Quadrifoglio s.p.a.* e *ATAF&Linea S.C.R.L.*, esattamente come nel 2012. Al terzo posto troviamo invece *Gala Energia Pulita*, che si occupa di fornitura energetica. Tra le altre, ritroviamo il *Consortio Zenit* ed altre Aziende.

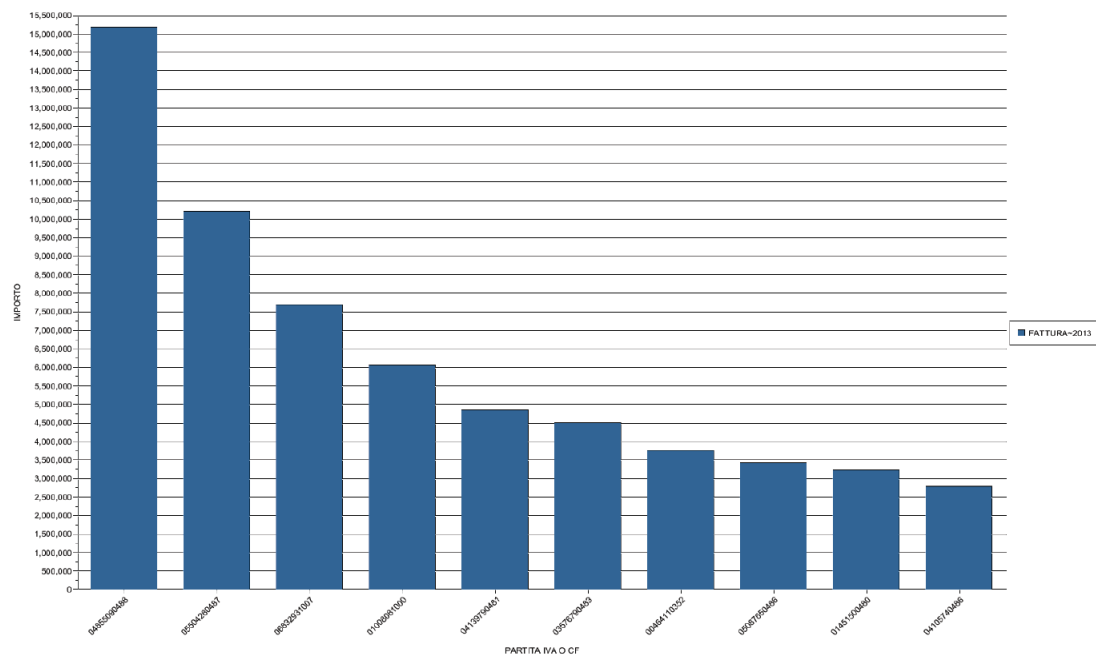


Figura 5.10: TOP 10 Pagamenti per Aziende 2013, grafico.

PARTITA IVA O CF	ANNO PAGAMENTO
	DOCUMENTO
	2013
	FATTURA
	IMPORTO
04855090488	15179955
05504280487	10221227
06832931007	7706345
01008081000	6067025
04139790481	4856351
03576790483	4520071
00464110352	3772330
05087650486	3441868
01451500480	3240812
04105740486	2789819

Figura 5.11: TOP 10 Pagamenti per Aziende 2013, tabella.

PARTITA IVA O CF	BENEFICIARIO	ANNO PAGAMENTO
		DOCUMENTO
		2013
		FATTURA
		IMPORTO
00464110352	CIR FOOD ARL	3772330
01008081000	RETE FERROVIARIA ITALIANA S.P.A.DIR COMP. INFRASTRUTTURA FIRENZE	6067025
01451500480	ATAF SPA	3240812
03576790483	SILFI S.P.A.	4520071
04105740486	NIVI CREDIT SRL	2789819
04139790481	R.T.I.ZENIT - DI VITTORIOCOOP.ZENIT- COOP. VITTORIO	3199365
	ATI CONSORZIO ZENIT E COOP. ELFOCOOPERATIVA DI VITTORIO	302129
	CONSORZIO ZENIT RSA CHIOCCIOLA	270722
	CONSORZIO ZENIT C.D. LA TENDA	172635
	CONSORZIO ZENIT SCRL RSA LORENZO	161501
	CONSORZIO ZENIT R.S.A. S. GIUSEPPE	131176
	CONSORZIO ZENIT CENTRO G.PINOCCHIO	103899
	CONSORZIO ZENIT C.P.A. GIANBURRASCA	89866
	CONSORZIO ZENIT COM.EDUC.DEMIDOFF	81494
	CONSORZIO ZENIT C.D. ALZHEIMER	78994
	CONSORZIO ZENIT R.S.A. DEMIDOFF	75257
	CONSORZIO ZENIT C.D. DEMIDOFF	71576
	CONSORZIO ZENIT R.S.A. LA TENDA	63181
	CONSORZIO ZENIT COM.EDUC.S.FILIPPO	54556
04855090488	QUADRIFOGLIO SPA	15179955
05087650486	SERVIZI ALLA STRADA S.P.A.	3441868
05504280487	ATAF&LINEA S.C.R.L.	10221227
06832931007	GALA ENERGIA PULITA	7706345

Figura 5.12: TOP 10 Pagamenti per Aziende 2013, tabella disaggregata.

5.2.3 Confronto: TOP 10 Pagamenti per Aziende 2012/2013

Vediamo adesso come le Aziende presenti nella Top-10 di Pagamenti del 2012 vengono pagate tra il 2012 ed il 2013.

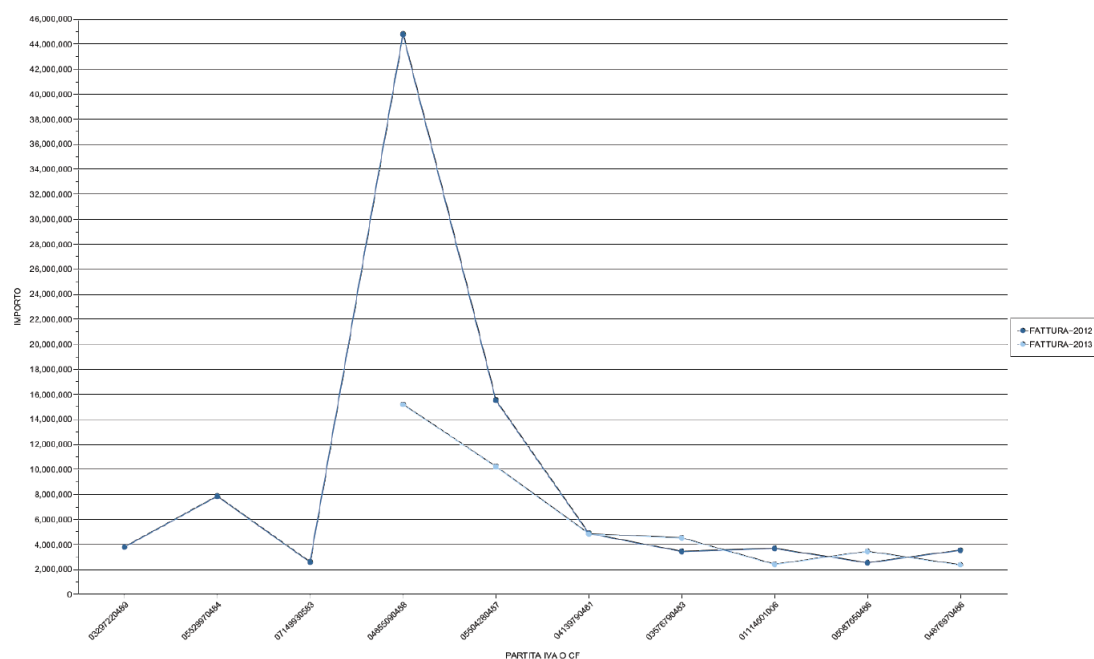


Figura 5.13: TOP 10 Pagamenti per Aziende 2012/2013, grafico.

PARTITA IVA O CF	ANNO PAGAMENTO	
	DOCUMENTO	
	2012	2013
	FATTURA	FATTURA
	IMPORTO	IMPORTO
04855090488	44789799	15179955
05504280487	15515294	10221227
04139790481	4880540	4856351
03576790483	3443045	4520071
05529970484	7855174	-
01114601006	3677300	2405362
05087650486	2538400	3441868
04876970486	3539871	2362093
03297220489	3798317	-
07149930583	2586918	-

Figura 5.14: TOP 10 Pagamenti per Aziende 2012/2013, tabella.

Si notano innanzitutto tre aziende (*Tram di Firenze*, *RTI Cofely-Astrim-Operosa-Colser Iprams* e *ASP Firenze Montedomini*) presenti nella Top-10 2012 e del tutto assenti nel bilancio 2013, indice, probabilmente, di Aziende a cui erano stati assegnati lavori una

PARTITA IVA O CF	BENEFICIARIO	ANNO PAGAMENTO	
		DOCUMENTO	
		2012	2013
		FATTURA	FATTURA
		IMPORTO	IMPORTO
01114601006	POSTE ITALIANE S.P.A. - A.T.G.C.	3471666	2254663
	POSTE ITALIANEC. CONTR 30006059-048	170956	149021
	POSTE ITALIANE S.P.A.	23329	-
	POSTE ITALIANE S.P.A.D. C.G.C./S.C.A.	9680	-
	POSTE ITALIANE S.P.A.C. CONTR. 30006059-044	1669	1678
03297220489	ASP FIRENZE MONTEDOMINIEX S.AMBROGIO/FULIGNO/BIGALLO	3796317	-
03576790483	SILFI S.P.A.	3443045	4520071
04139790481	R.T.I.ZENIT - DI VITTORIO COOP.ZENIT-COOP. VITTORIO	3283471	3199365
	ATI CONSORZIO ZENIT E COOP. ELFOCOOPERATIVA DI VITTORIO	511269	302129
	CONSORZIO ZENIT RSA CHIOCCIOLA	130224	270722
	CONSORZIO ZENIT SCCRL RSA LORENZO	157389	161501
	CONSORZIO ZENIT C.D. LA TENDA	72426	172635
	CONSORZIO ZENIT CENTRO G.PINOCCHIO	138634	103899
	CONSORZIO ZENIT C.P.A. GIANBURRASCA	133479	89866
	CONSORZIO ZENIT R.S.A. S. GIUSEPPE	79849	131176
	CONSORZIO ZENT COM.EDUC.DEMIDOFF	124852	81494
	CONSORZIO ZENIT COM.EDUC.S.FILIPPO	71887	54556
	CONSORZIO ZENIT R.S.A. LA TENDA	61833	63181
	CONSORZIO ZENIT R.S.A. DEMIDOFF	43396	75257
	CONSORZIO ZENIT C.D. ALZHEIMER	37338	78994
	CONSORZIO ZENIT C.D. DEMIDOFF	34473	71576
04855090488	QUADRIFOGLIO SPA	44789799	15179955
04876970486	CO & SO FIRENZE CONS.COOP.SOLIDARIEEX CO & SO CONS.COOP.SOLIDARIETA'	1060524	1279179
	RTC CO&SO E CONSORZIO FABRICA S.C.	1368756	-
	A.T.I.CO&SO-COOPCULTURE- COPLAT-IFNET-COOP.ARCHEOLOGIA	528218	357297
	A.T.I.CO&SO METROPOLI	422278	440984
	CO & SO FIRENZE CONS. COOP. SOL.	160095	284633
05087650486	SERVIZI ALLA STRADA S.P.A.	2538400	3441868
05504280487	ATAF&LINEA S.C.R.L.	15515294	10221227
05529970484	TRAM DI FIRENZE	7855174	-
07149930583	RTI COFELY-ASTRIM-OPEROSA-COLSER IPRAMS	2586918	-

Figura 5.15: TOP 10 Pagamenti per Aziende 2012/2013, tabella disaggregata.

tantum. Si nota inoltre la disparità dei pagamenti verso *Quadrifoglio s.p.a.* che, nonostante sia comunque la più pagata delle Aziende, ha registrato comunque un forte crollo rispetto al 2012. Per quanto riguarda le altre Aziende, si registra invece un andamento piuttosto regolare nel biennio 2012/2013.

5.3 TOP 10 Pagamenti per Direzioni

Dal momento che le Direzioni di Servizio sono delle classificazioni per area di competenza del Comune di Firenze, possiamo vedere i pagamenti approvati da una certa Direzione come gli investimenti del Comune nelle rispettive aree.

Eseguiamo in questa sezione le stesse analisi effettuate in Sezione 5.2, con l'unica differenza che i pagamenti saranno aggregati per Direzione di Servizio e non più per Azien-

da. Potremo analizzare in questo modo in quali aree ha scelto di investire il Comune di Firenze e come queste scelte si siano evolute tra il 2012 ed il 2013.

5.3.1 TOP 10 Pagamenti per Direzioni 2012

La prima analisi da effettuare è relativa alle 10 Direzioni di Servizio che, nel 2012, hanno pagato di più rispetto alle altre.

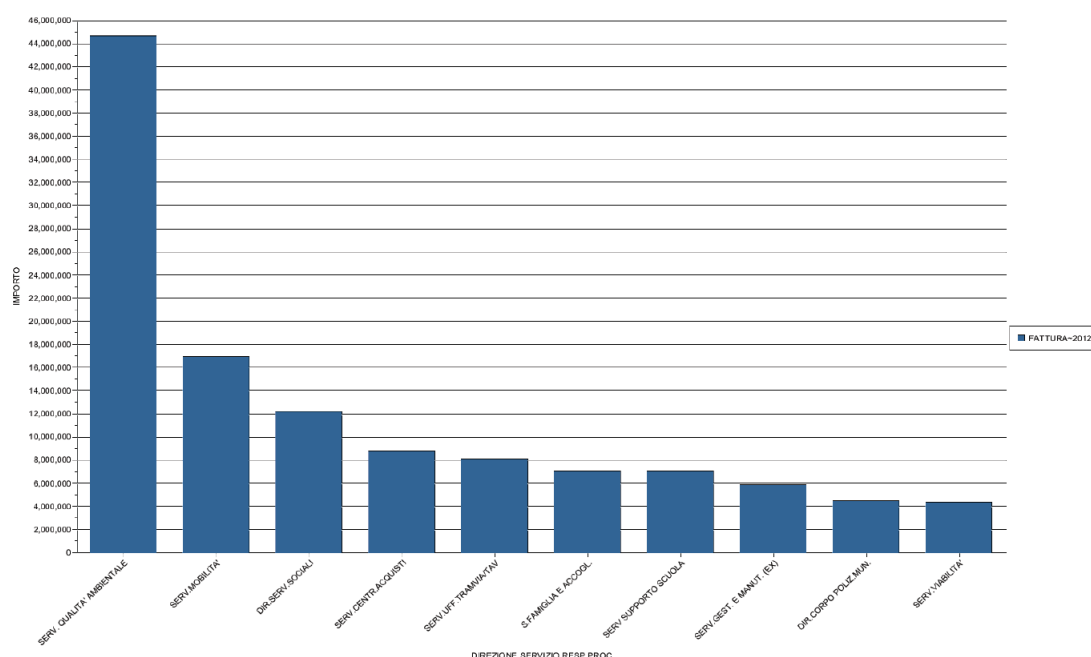


Figura 5.16: TOP 10 Pagamenti per Direzioni 2012, grafico.

DIREZIONE SERVIZIO RESP PROC	ANNO PAGAMENTO
	DOCUMENTO
	2012
	FATTURA
	IMPORITO
SERV. QUALITA' AMBIENTALE	44703190
SERV. MOBILITA'	16995241
DIR. SERV. SOCIALI	12201501
SERV. CENTR. ACQUISTI	8815575
SERV. UFF. TRAMVIA/TAV	8125691
S. FAMIGLIA E ACCOGL.	7074200
SERV. SUPPORTO SCUOLA	7058086
SERV. GEST. E MANUT. (EX)	5908707
DIR. CORPO POLIZ. MUN.	4483238
SERV. VIABILITA'	4352445

Figura 5.17: TOP 10 Pagamenti per Direzioni 2012, tabella.

Come era intuibile, le Direzioni ad aver speso di più nel 2012 sono *Servizi Qualità Ambientale* e *Servizi Mobilità* (ricordando dalle analisi in Sezione 5.2 che le aziende

più pagate nel 2012 erano la *Quadrifoglio s.p.a.*, *Ataf* e *Tram di Firenze*). Tra le altre Direzioni presenti nella Top-10 troviamo altri settori molto importanti, come *Servizi Supporto Scuola*, *Direzione Corpo Polizia Municipale* e *Servizi Viabilità*.

5.3.2 TOP 10 Pagamenti per Direzioni 2013

Ripetiamo adesso le analisi per l'anno 2013.

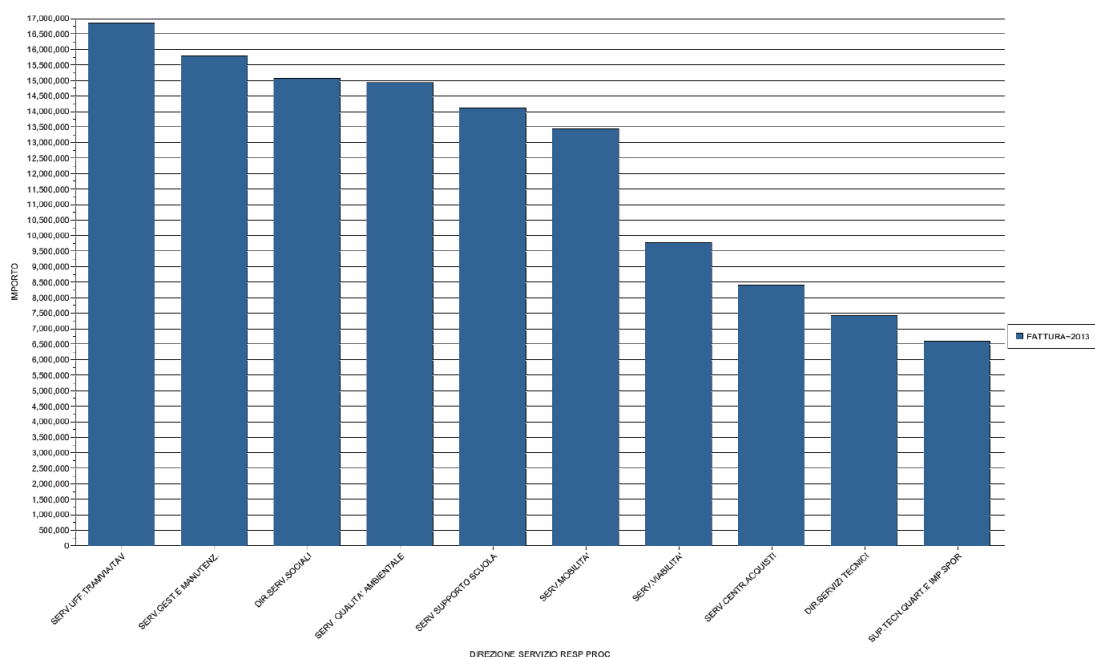


Figura 5.18: TOP 10 Pagamenti per Direzioni 2013, grafico.

DIREZIONE SERVIZIO RESP PROC	ANNO PAGAMENTO
	DOCUMENTO
	2013
	FATTURA
	IMPORTO
SERV.UFF. TRAMVIA/TAV	16852593
SERV.GEST.E MANUTENZ.	15788546
DIR.SERV.SOCIALI	15063219
SERV. QUALITA' AMBIENTALE	14940874
SERV.SUPPORTO SCUOLA	14125175
SERV.MOBILITA'	13452782
SERV.VIABILITA'	9765161
SERV.CENTR.ACQUISTI	8408478
DIR.SERVIZI TECNICI	7440052
SUP.TECN.QUART.E IMP.SPOR	6596964

Figura 5.19: TOP 10 Pagamenti per Direzioni 2013, tabella.

Vediamo in questo grafico (Figura 5.18 che gli importi pagati dalle prima 10 Direzioni nel 2013 sono piuttosto regolari, senza sbalzi o irregolarità troppo evidenti. Rispetto al 2012, gli investimenti sono stati indirizzati più o meno nelle stesse direzioni.

5.3.3 Confronto: TOP 10 Pagamenti per Direzioni 2012/2013

Confrontiamo adesso le Direzioni con più investimenti nel 2012 con gli investimenti nel 2013 (relativamente alle stesse Direzioni di Servizio).

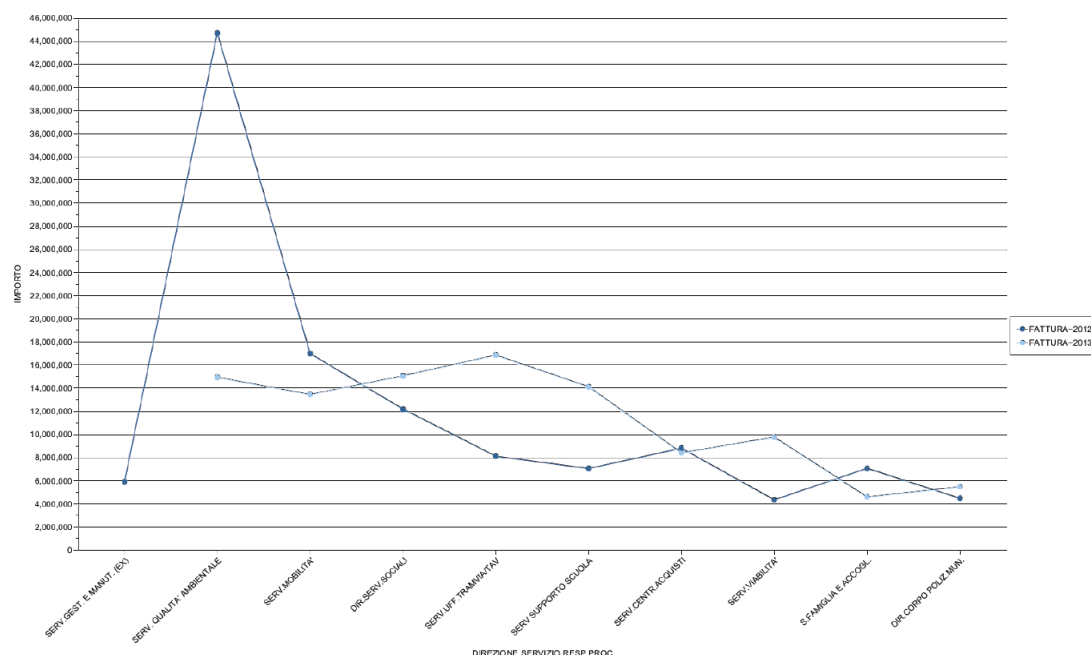


Figura 5.20: TOP 10 Pagamenti per Direzioni 2012/2013, grafico.

DIREZIONE SERVIZIO RESP PROC	ANNO PAGAMENTO	
	DOCUMENTO	
	2012	2013
	FATTURA	FATTURA
	IMPORTO	IMPORTO
SERV. QUALITA' AMBIENTALE	44703190	14940874
SERV.MOBILITA'	16995241	13452782
DIR.SERV.SOCIALI	12201501	15063219
SERV.UFF.TRAMVIA/TAV	8125691	16852593
SERV.SUPPORTO SCUOLA	7058086	14125175
SERV.CENTR.ACQUISTI	8815575	8408478
SERV.VIABILITA'	4352445	9765161
S.FAMIGLIA E ACCOGL.	7074200	4595523
DIR.CORPO POLIZ.MUN.	4483238	5474256
SERV.GEST. E MANUT. (EX)	5908707	-

Figura 5.21: TOP 10 Pagamenti per Direzioni 2012/2013, tabella.

A parte *Servizi Gestione e Manutenzione*, che non è presente nel bilancio 2013, notiamo nel grafico in Figura 5.20 un netto distacco nei finanziamenti 2012 e 2013. Il cambiamento più eclatante risulta essere quello relativo ai *Servizi Qualità Ambientale*, in accordo con quanto osservato in Sezione 5.2, che registra un calo di quasi 30 milioni di euro negli investimenti, pari a circa 66% degli investimenti registrati nel 2012.

5.4 BOTTOM 10 Pagamenti per Direzioni

Avendo osservato le Direzioni con più investimenti, analizziamo in questa sezione le aree con meno investimenti.

5.4.1 BOTTOM 10 Pagamenti per Direzioni 2012

Questa analisi rappresenta la Bottom-10 degli investimenti nel 2012, ovvero le 10 Direzioni di Servizio del Comune di Firenze che hanno speso meno nell'anno 2012.

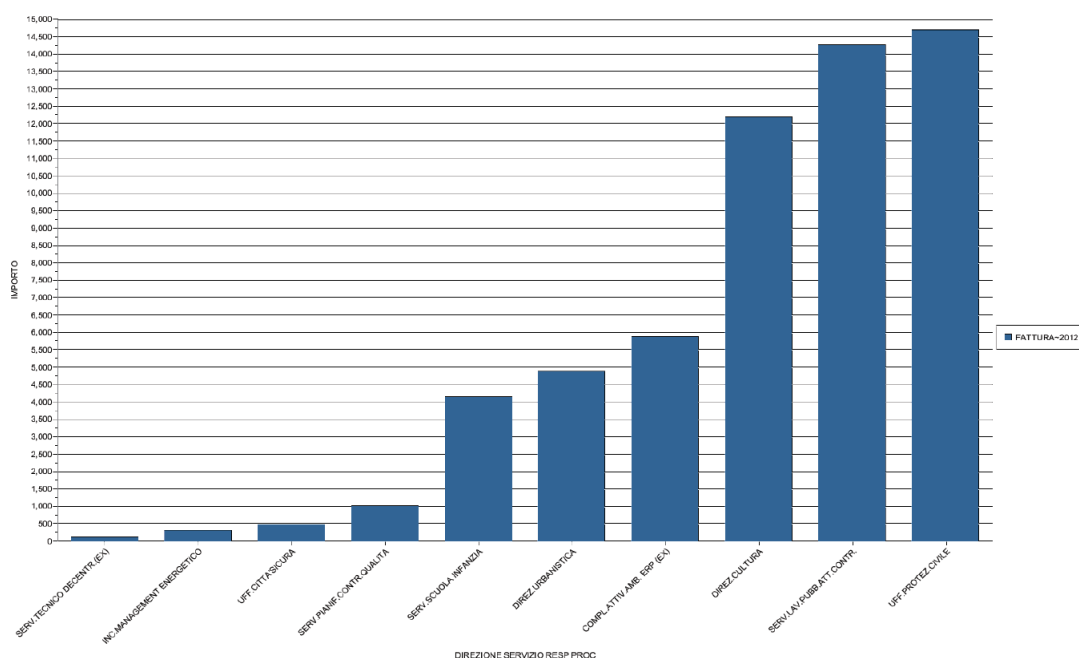


Figura 5.22: BOTTOM 10 Pagamenti per Direzioni 2012, grafico.

Osservando il grafico di quest'analisi, si può vedere come alcuni importanti settori, come *Protezione Civile*, *Cultura* o *Scuola dell'Infanzia*, siano stati alquanto trascurate per quanto riguarda gli investimenti promossi dal Comune di Firenze. Anche se si tratta, senza dubbio, di aree molto importanti, quella con più investimenti tra le 10 in questione registra poco più di 14mila euro di investimenti, mentre la prima registra soltanto 130 euro.

	ANNO PAGAMENTO
	DOCUMENTO
	2012
	FATTURA
DIREZIONE SERVIZIO RESP PROC	IMPORTO
SERV.TECNICO DECENTR.(EX)	130
INC.MANAGEMENT ENERGETICO	314
UFF.CITTA'SICURA	484
SERV.PIANIF.CONTR.QUALITA	1012
SERV.SCUOLA INFANZIA	4156
DIREZ.URBANISTICA	4892
COMPL.ATTIV.AMB. ERP (EX)	5876
DIREZ.CULTURA	12200
SERV.LAV.PUBB.ATT.CONTR.	14274
UFF.PROTEZ.CIVILE	14703

Figura 5.23: BOTTOM 10 Pagamenti per Direzioni 2012, tabella.

5.4.2 BOTTOM 10 Pagamenti per Direzioni 2013

Analizziamo adesso gli investimenti più bassi nel 2013.

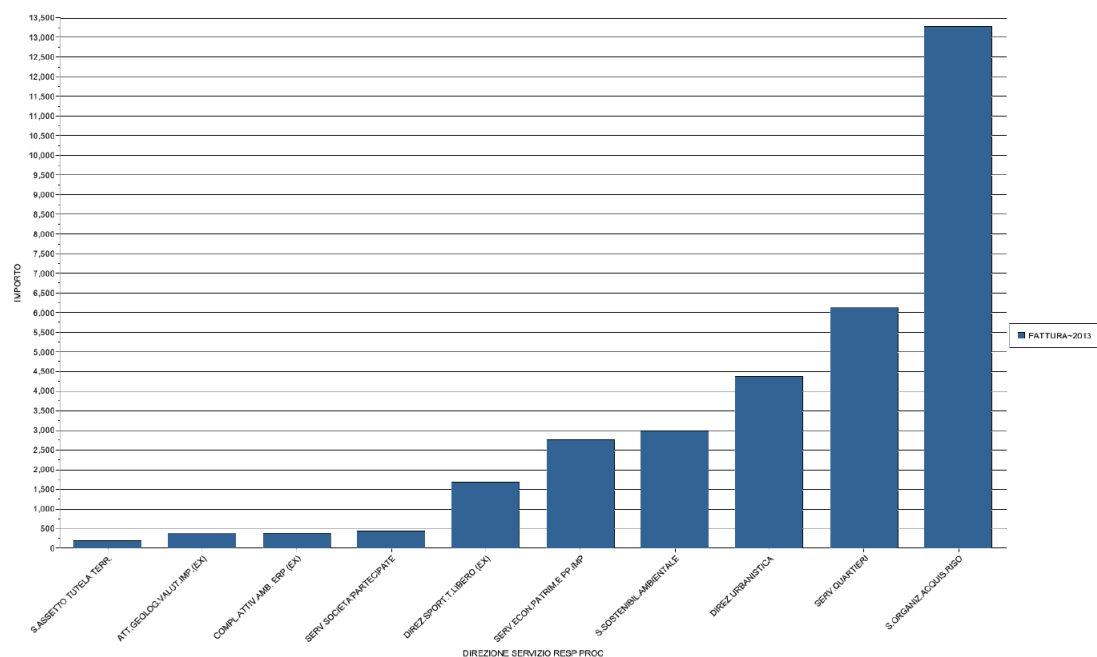


Figura 5.24: BOTTOM 10 Pagamenti per Direzioni 2013, grafico.

Fortunatamente, dal grafico e dalla tabella risulta che durante il 2013 gli investimenti più bassi del Comune di Firenze sono relativi a settori ed aree minori, che giustificano i bassi costi riscontrati.

	ANNO PAGAMENTO
	DOCUMENTO
	2013
	FATTURA
DIREZIONE SERVIZIO RESP PROC	IMPORTO
S.ASSETTO TUTELA TERR.	201
ATT.GEOLOG.VALUT.IMP.(EX)	366
COMPL.ATTIV.AMB. ERP (EX)	385
SERV.SOCIETA'PARTECIPATE	452
DIREZ.SPORT.T.LIBERO (EX)	1684
SERV.ECON.PATRIM.E PP.IMP	2780
S.SOSTENIBIL.AMBIENTALE	3000
DIREZ.URBANISTICA	4383
SERV.QUARTIERI	6117
S.ORGANIZ.ACQUIS.RISO	13293

Figura 5.25: BOTTOM 10 Pagamenti per Direzioni 2013, tabella.

5.4.3 Confronto: BOTTOM 10 Pagamenti per Direzioni 2012/2013

Confrontiamo adesso le Direzioni che hanno pagato di meno nel 2012, mostrando la loro evoluzione tra il 2012 ed il 2013.

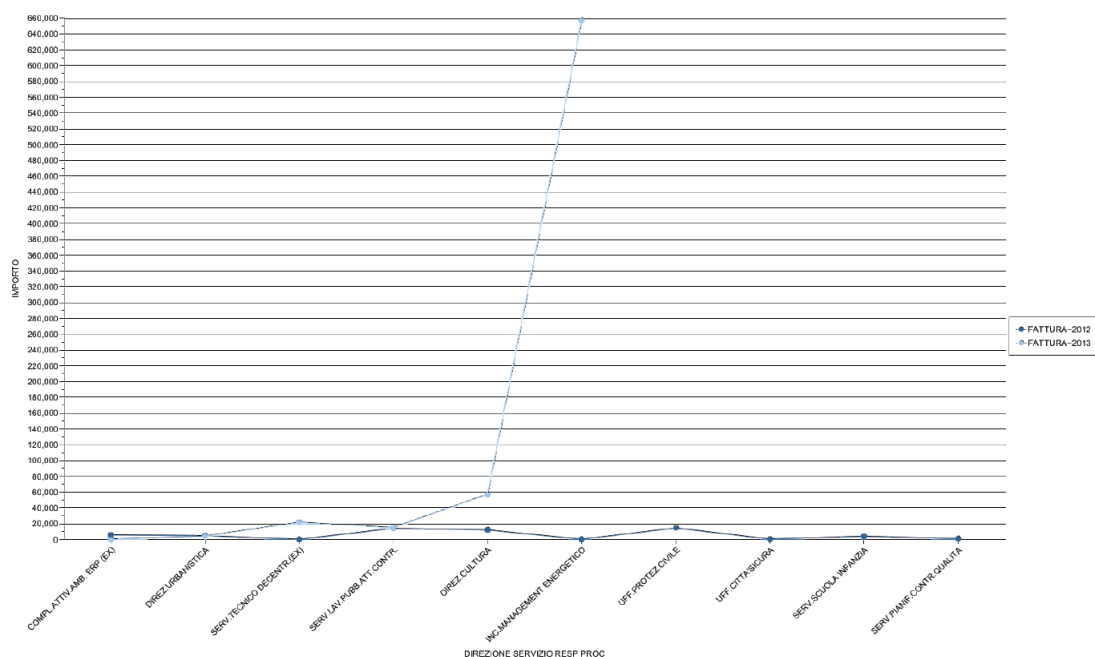


Figura 5.26: BOTTOM 10 Pagamenti per Direzioni 2012/2013, grafico.

La cosa che risalta subito all'occhio è l'enorme distacco presente negli investimenti nel settore *Management Energetico*, che registra un sostanziale aumento di finanziamenti nel 2013 rispetto al 2012. Inoltre si nota anche un aumento degli investimenti di circa il 369% per quanto riguarda il settore della *Cultura*.

	ANNO PAGAMENTO	
	DOCUMENTO	
	2012	2013
	FATTURA	FATTURA
DIREZIONE SERVIZIO RESP PROC	IMPORTO	IMPORTO
UFF.CITTA'SICURA	484	-
SERV.PIANIF.CONTR.QUALITA	1012	-
SERV.SCUOLA INFANZIA	4156	-
COMPL.ATTIV.AMB. ERP (EX)	5876	385
DIREZ.URBANISTICA	4892	4383
UFF.PROTEZ.CIVILE	14703	-
SERV.TECNICO DECENTR.(EX)	130	21921
SERV.LAV.PUBB.ATT.CONTR.	14274	15122
DIREZ.CULTURA	12200	57330
INC.MANAGEMENT ENERGETICO	314	657119

Figura 5.27: BOTTOM 10 Pagamenti per Direzioni 2012/2013, tabella.

Per quanto riguarda le altre Direzioni, o esse sono assenti nel bilancio 2013, o gli investimenti risultano essere più o meno gli stessi del 2012.

5.5 TOP 10 Accoppiate Direzione-Azienda

Proponiamo in questa sezione un'analisi piuttosto interessante: vogliamo ottenere la lista delle 10 accoppiate Azienda-Direzione di Servizio più frequenti nel biennio 2012/2013.

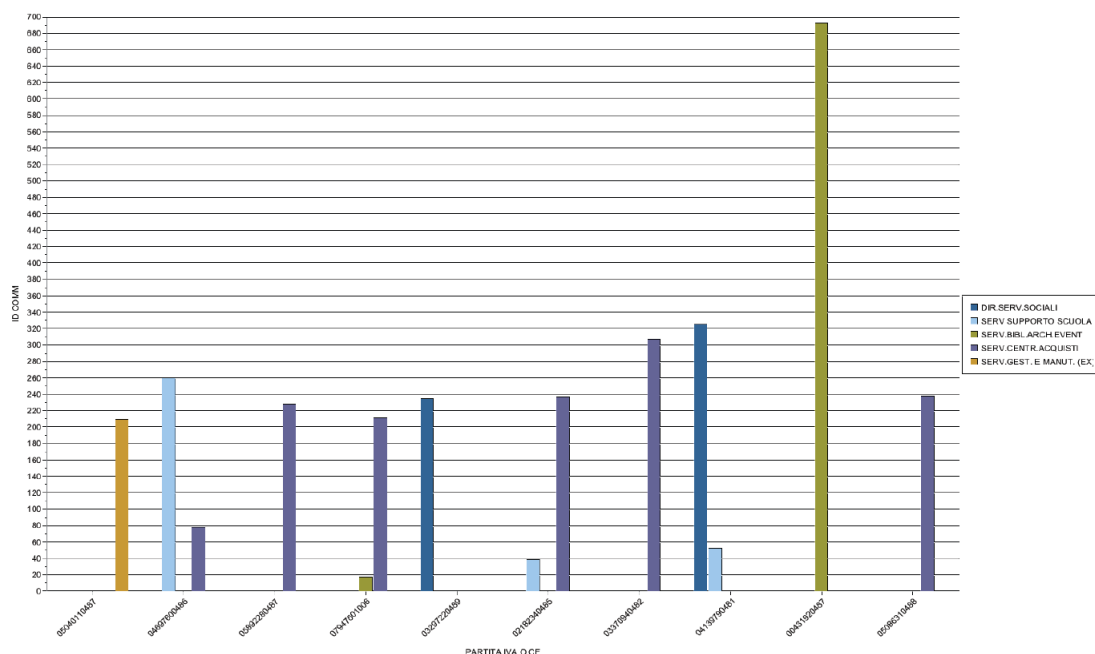


Figura 5.28: TOP 10 Accoppiate Direzione-Azienda, grafico.

DIREZIONE SERVIZIO RESP PROC					
	DIR.SERV.SOCIALI	SERV.SUPPORTO SCUOLA	SERV.BIBL.ARCH.EVENT	SERV.CENTR.ACQUISTI	SERV.GEST. E MANUT. (EX)
PARTITA IVA O CF	ID COMM	ID COMM	ID COMM	ID COMM	ID COMM
00431920487	-	-	693	-	-
04139790481	325	53	-	-	-
03370940482	1	-	1	307	-
04697600486	-	260	-	78	-
05086310488	-	-	-	238	-
02182340485	-	39	-	237	-
03297220489	235	-	-	-	-
05892280487	-	-	-	228	-
07947601006	-	-	17	211	-
05040110487	-	-	-	-	210

Figura 5.29: TOP 10 Accoppiate Direzione-Azienda, tabella.

Al primo posto, con ben 693 commissioni, troviamo la Direzione *Servizi Biblioteche e Archivistica* e l'Azienda *Licosa*. *Licosa* è infatti un'Azienda fiorentina che si occupa di abbonamenti a riviste, fornitura e ricerca di libri e pubblicazioni.

PARTITA IVA O CF	BENEFICIARIO	DIREZIONE SERVIZIO RESP PROC				
		DIR.SERV.SOCIALI	SERV.SUPPORTO SCUOLA	SERV.BIBLARCH.EVENT	SERV.CENTR.ACQUISTI	SERV.GEST. E MANUT. (EX)
		ID COMM	ID COMM	ID COMM	ID COMM	ID COMM
00431920487	LICOSA	-	-	693	-	-
02182340485	FARMACIE FIORENTINE AFAM SPA	-	39	-	237	-
03297220489	ASP FIRENZE MONTEDOMINIEX S.AMBROGIO/FULIGNO/BIGALLO	235	-	-	-	-
03370940482	NUOVADATA S.R.L.	1	-	1	307	-
04139790481	ATI CONSORZIO ZENIT E COOP. ELFOCOOPERATIVA DI VITTORIO	61	-	-	-	-
	CONSORZIO ZENIT C.D. ALZHEIMER	18	-	-	-	-
	CONSORZIO ZENIT C.D. DEMIDOFF	31	-	-	-	-
	CONSORZIO ZENIT C.D. LA TENDA	8	-	-	-	-
	CONSORZIO ZENIT CENTRO G.PINOCCHIO	3	-	-	-	-
	CONSORZIO ZENIT COM.EDUC.S.FILIPPO	1	-	-	-	-
	CONSORZIO ZENIT R.S.A. DEMIDOFF	18	-	-	-	-
	CONSORZIO ZENIT R.S.A. LA TENDA	8	-	-	-	-
	CONSORZIO ZENIT R.S.A. S. GIUSEPPE	12	-	-	-	-
	CONSORZIO ZENIT RSA CHIOCCIOLA	32	-	-	-	-
	CONSORZIO ZENT COM.EDUC.DEMIDOFF	2	-	-	-	-
	CONZORZIO ZENIT SCCRL RSA LORENZO	47	-	-	-	-
	R.T.I.ZENIT - DI VITTORIOCOOP.ZENIT-COOP. VITTORIO	84	53	-	-	-
04697600486	CECIONI CATERING S.R.L.	-	260	-	78	-
05040110487	PUBLIACQUA SPA	-	-	-	-	210
05086310488	SETTI MARCO S.R.L.	-	-	-	238	-
05892280487	S.D.M.	-	-	-	228	-
07947601006	TEAM SERVICE SOC.CONSORTILE A R.L.	-	-	17	211	-

Figura 5.30: TOP 10 Accoppiate Direzione-Azienda, tabella disaggregata.

Al secondo posto troviamo invece la coppia *Direzione Servizi Sociali* con il *Consorzio Zenit* (che avevamo già citato in Sezione 5.2 tra le Aziende più pagate nel 2012/2013).

5.6 TOP 10 Anni di Pagamento per Aziende

Adesso ci proponiamo invece di individuare a quali Aziende, tra quelle presenti nei bilanci 2012/2013, sono stati assegnati i lavori più lunghi (e presumibilmente più impegnativi).

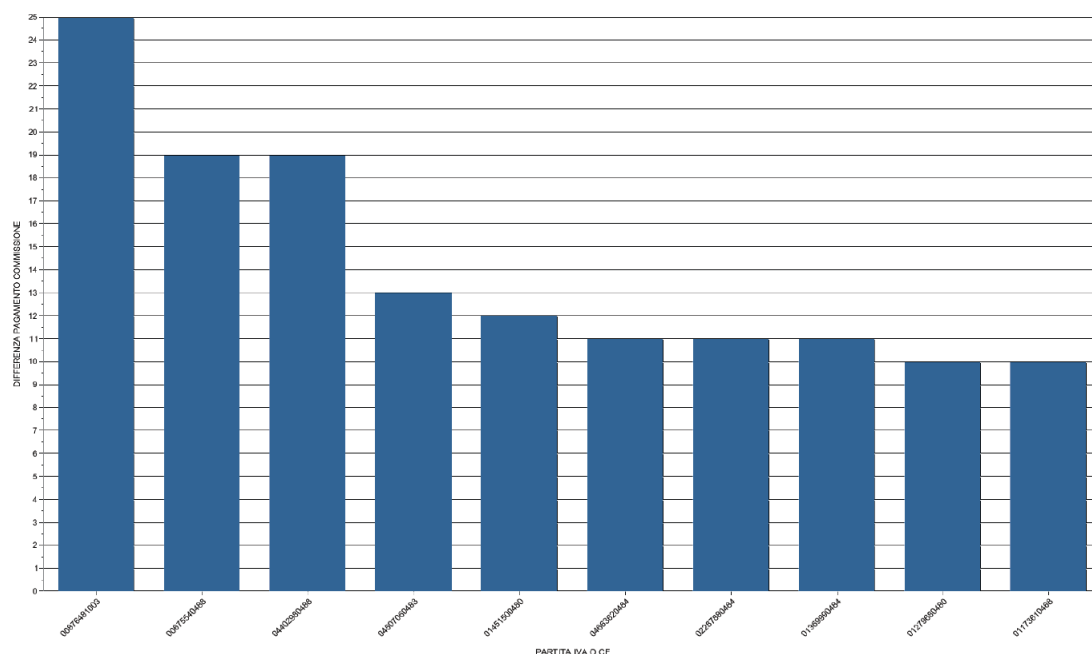


Figura 5.31: TOP 10 Anni di Pagamento per Aziende, grafico.

PARTITA IVA O CF	DIFFERENZA PAGAMENTO COMMISSIONE
00876481003	25
00675540488	19
04402980488	19
04507060483	13
01451500480	12
01369990484	11
02287880484	11
04663620484	11
01173610468	10
01279680480	10

Figura 5.32: TOP 10 Anni di Pagamento per Aziende, tabella.

Al primo posto, con un lavoro durato ben 25 anni, troviamo l'*ANSA (Agenzia Nazionale Stampa Associata)*, mentre al secondo e terzo posto, a pari merito con 19 anni, troviamo l'Azienda *S.I.L.V.E. s.p.a.*, che si occupa di impianti di illuminazione elettrica, ed l'Azienda *Zucchetti Franco*, i cui servizi ci sono ignoti. Le altre commissioni presenti in questa Top-10 vano invece dai 13 ai 10 anni di durata.

PARTITA IVA O CF	BENEFICIARIO	DIFFERENZA PAGAMENTO COMMISSIONE
00675540488	S.I.L.V.E. S.P.A.	19
00876481003	ANSAAGENZIA NAZIONALE STAMPA ASSOCIATA	25
01173610468	BICICCHI FELICE S.R.L.BICICCHI FELICE S.R.L.	10
01279680480	UNIV. DI FIRENZE DIP. BIOLOGIADIPARTIMENTO BIOLOGIA "LEO PARDI"	10
	DIP.TO SCIENZA POLITICA-SOCIOL.POL.UNIVERSITA' STUDI DI FIRENZE	6
	UNIVERSITA DEGLI STUDI DI FIRENCECENTRO SERVIZI INFORMATICI E TELEM.	6
	UNIVERSITA STUDI	1
	UNIVERSITA' DEGLI STUDI DI FIRENZE	
	UNIVERSITA' DEGLI STUDI DI FIRENZEMUSEO DI STORIA NATURALE	
01369990484	ALEXANIAN ARTHUR	11
01451500480	ATAF SPA	12
02287880484	GEO ECO PROGETTI	11
04402980488	ZUCCHETTI FRANCO	19
04507060483	DE RENZIS GIANCARLO	13
04663620484	SERTEC SAS DI C. MESSINA E L. PAOLINI & C.	11

Figura 5.33: TOP 10 Anni di Pagamento per Aziende, tabella disaggregata.

5.7 Analisi per tipo di Atto di Impegno

Concludiamo questa serie di analisi con alcune osservazioni riguardanti il tipo di Atto di Impegno indicato per ogni commissione, ovvero sul campo **TIPO_ATTO_IMPEGNO** del nostro file **.csv**.

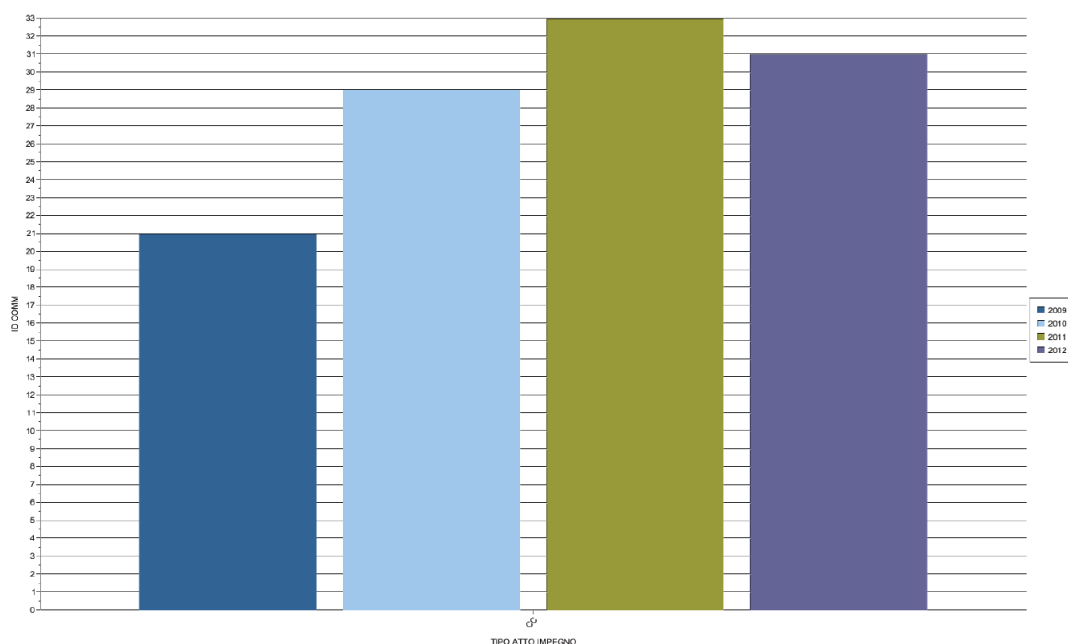
Il tipo di Atto di Impegno è una sigla, di due o tre lettere, che indica con quale priorità è stato commissionato un certo mandato, ovvero l'importanza attribuita ad ogni commissione. A parte sigle poco utilizzate o deprecate, le due più importanti in questo senso sono *DD* e *CC*, che stanno, rispettivamente, per *Decreto* (o *Determina*) *Dirigenziale* e per *Consiglio Comunale*.

Un *Decreto Dirigenziale* indicherà quindi tutte quelle decisioni prese da un dirigente del settore (solitamente il dirigente della Direzione di Servizio che commissiona il lavoro), che non necessitano consiglio o approvazione da parte di altri settori; al contrario, un lavoro commissionato dal *Consiglio Comunale* sarà sicuramente un lavoro molto più importante o impegnativo, in quanto richiede che tutto il Consiglio Comunale sia d'accordo sui dettagli e le modalità di quel particolare mandato.

5.7.1 Consigli Comunali negli anni

Una prima, semplice analisi in questo senso può essere la determinazione di quanti mandati sono stati commissionati dal Consiglio Comunale negli anni, per capire in quale anno sono state prese più "decisioni importanti".

Effettuiamo quindi un'analisi su due dimensioni: il tipo di Atto d'Impegno, che sarà fissato su *CC*, e l'Anno di Commissione, che spazierà sull'intero dominio.

Figura 5.34: *Consigli Comunali* negli anni, grafico.

TIPO ATTO IMPEGNO	ANNO COMMISSIONE			
	2009	2010	2011	2012
	ID COMM	ID COMM	ID COMM	ID COMM
CC	21	29	33	31

Figura 5.35: *Consigli Comunali* negli anni, tabella.

Si osserva innanzitutto che prima del 2009 non è stato fatto (o più ragionevolmente, non è stato registrato) alcun *Consiglio Comunale*, come anche nel 2013. Quindi per quanto riguarda gli anni che vanno dal 2009 al 2012, si vede che la differenza tra i vari anni è minima, raggiungendo un picco nel 2011, con 33 *Consigli Comunali*.

5.7.2 Confronto: CC-DD nel 2009-2013

Volendo entrare adesso più nel dettaglio rispetto all'analisi precedente, possiamo includere anche il numero, per ogni anno, di lavori approvati per *Decreto Dirigenziale*, mettendo in relazione quanti lavori sono stati commissionati per *Consiglio Comunale* e quanti per *Decreto Dirigenziale*.

Come ci si poteva aspettare, i *Decreti Dirigenziali*, associati a lavori più semplici o meno importanti, sono molti di più rispetto ai *Consigli Comunali*, come possiamo vedere sia dai grafici che dalla tabella. Inoltre, se osserviamo i grafici a torta in Figura 5.37, si nota anche che in percentuale i *Consigli Comunali* sono andati a diminuire negli

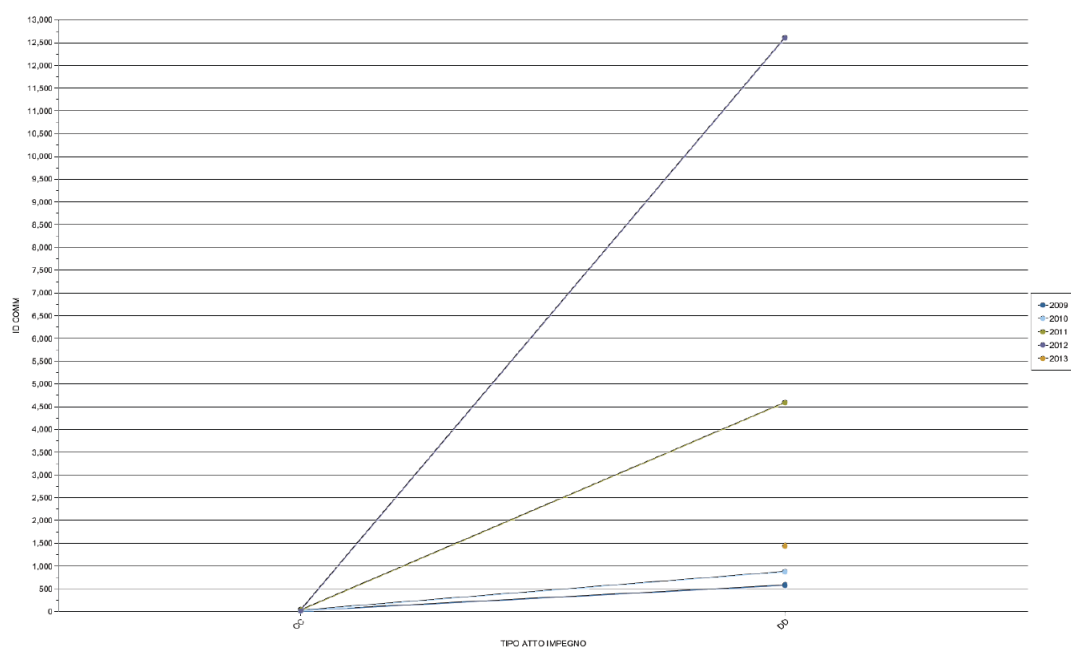


Figura 5.36: *Consigli Comunali* e *Decreti Dirigenziali* negli anni, grafico a linee.

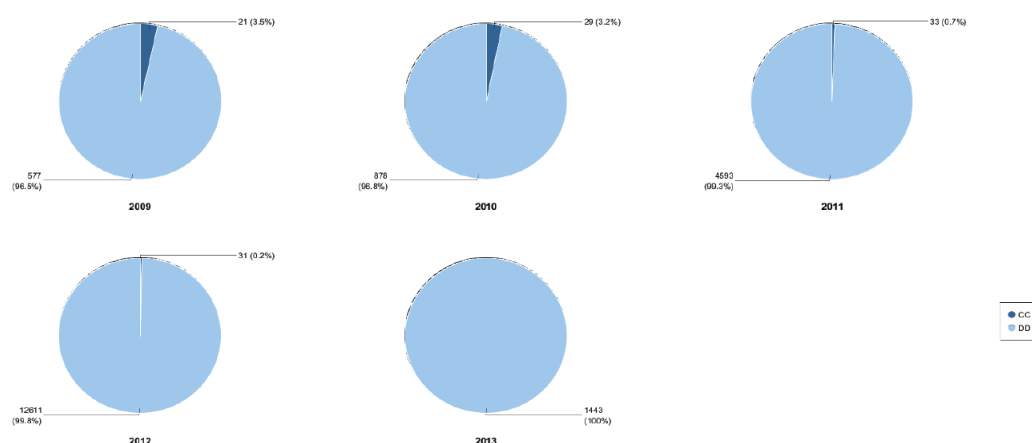


Figura 5.37: *Consigli Comunali* e *Decreti Dirigenziali* negli anni, grafici a torta.

anni rispetto ai *Decreti Dirigenziali*. Escludendo motivi tecnici del settore (sui quali non possiamo fare speculazioni di alcun tipo), una possibile spiegazione potrebbe essere che, essendo i *Consigli Comunali* molto più importanti di un qualsiasi *Decreto Dirigenziale*, questi sono stati tracciati e registrati in modo più accurato e preciso rispetto ai *Decreti Dirigenziali*, che molto probabilmente venivano persi tra le scartoffie e la burocrazia del Comune, almeno fino al Decreto Sviluppo 2012.

TIPO ATTO IMPEGNO	ANNO COMMISSIONE				
	2009	2010	2011	2012	2013
	ID COMM	ID COMM	ID COMM	ID COMM	ID COMM
CC	21	29	33	31	-
DD	577	878	4593	12611	1443

Figura 5.38: *Consigli Comunali e Decreti Dirigenziali* negli anni, tabella.

Conclusioni

Chiudiamo questo elaborato con alcune conclusioni generali sul progetto, le modalità e gli strumenti utilizzati, facendo anche alcune considerazioni personali. Esporremo brevemente anche un altro progetto universitario nel quale abbiamo utilizzato e sfruttato le conoscenze apprese durante il corso di Data Warehousing.

Il progetto che abbiamo seguito per la compilazione di questo elaborato è stato molto interessante, sia per quello che ci ha insegnato (come teoria, tecniche e strumenti) che per le possibilità che ci ha aperto nell’ambito dell’analisi *OLAP* e del Data Warehousing in generale. La cosa interessante è che in questo progetto abbiamo potuto studiare ed applicare le tecniche di DW non ad un problema “giocattolo” (come spesso accade in ambito scolastico/universitario) ma ad un problema reale e non banale, come l’analisi del bilancio del Comune di Firenze. Ovviamente non ci siamo dilungati troppo sulle conclusioni delle singole analisi, in quanto non fa parte delle nostre competenze, ma abbiamo invece mostrato come sia semplice, tramite l’utilizzo di queste tecniche e strumenti, mettere a disposizione anche ad un esperto del settore tutti i dati e le analisi ricavate.

Oltre a queste motivazioni, questo progetto ci ha anche “insegnato” ad utilizzare il sistema Pentaho che, come indicato più avanti, abbiamo già saputo sfruttare in altri ambiti.

Alcune considerazioni

Alcune considerazioni vanno fatte riguardo i *dataset utilizzati* ed il sito degli OpenData del Comune di Firenze¹: senza ombra di dubbio il Comune di Firenze ha fatto grandi passi, negli ultimi anni, in ambiente informatico e la scelta degli OpenData è senz’altro la direzione giusta da seguire. Tuttavia il Comune è ancora lontano dalla “perfezione”: la maggior parte dei *dataset* pubblicati sul sito sono parziali, incompleti, inconsistenti e molto probabilmente anche pieni di errori. Inoltre molti *dataset* presenti hanno un livello aggregazione talmente elevato da impedire qualsiasi analisi su di essi che non sia priva di significato.

¹<http://opendata.comune.fi.it/>

Per quanto riguarda il nostro progetto non è stato facile individuare all'interno del sito questi due *dataset* che facessero al caso nostro, in quanto tutti gli altri impedivano, in un modo o nell'altro, delle analisi ragionevoli.

Inoltre potremmo aggiungere che, nonostante i due *dataset* dei bilanci 2012/2013 siano piuttosto ricchi e ben strutturati, si sarebbero potute aggiungere tante informazioni aggiuntive, come ad esempio se una o più commissioni fanno parte di un particolare progetto del Comune oppure il tipo di ogni commissione (ad esempio lavori di ampliamento, di ristrutturazione, di installazione di nuovi servizi, ecc. . .), dando la possibilità di effettuare molte più analisi su di essi.

Siamo quindi rimasti contenti dal progetto OpenData del Comune di Firenze, con la speranza che il lavoro in questo senso sia solo agli inizi e che negli anni avvenire si registri un incremento nell'affluenza di dati e magari un maggior controllo e lavoro di ripulitura su quest'ultimi.

Altri progetti

In quest'ultima sezione, riportiamo alcuni cenni ad un altro nostro utilizzo di Pentaho, nell'ambito di un altro progetto universitario. Il progetto in questione, relativo al corso di *AQS (Analisi Quantitativa dei Sistemi, Corso di Laurea Magistrale in Informatica)* aveva come obiettivo la caratterizzazione del comportamento temporale del tool *Snort*, un sistema di *Network Intrusion Detection (NIDS)*.

La prima parte del progetto si compone essenzialmente di un'analisi sperimentale, la quale produce una grossa mole di dati che, nella seconda parte del progetto, sarà necessario interpretare e analizzare per rispondere a quanto chiesto.

Per integrare e manipolare i dati raccolti abbiamo scelto di utilizzare Pentaho al fine di generare successivamente un sistema *OLAP* su cui effettuare tutte le analisi necessarie.

Senza entrare nel merito di quelli che sono i contenuti ed i significati dei file *.csv* generati in fase sperimentale, riportiamo qui la trasformazione *Kettle* eseguita su di essi. Utilizzando il tool *Kettle* è possibile infatti integrare e manipolare i file *.csv* generati in fase sperimentale, così da renderli in una forma più consona alla memorizzazione all'interno di un database: sarà possibile, ad esempio, modificare alcuni campi o aggiungerne di nuovi come campi calcolati.

Vediamo qui di seguito la trasformazione (per comodità divisa in quattro immagini) che permetterà di integrare tra loro i dati relativi ai vari esperimenti eseguiti:

Il file generato da *Kettle* (file di log *logCompleto.csv*), ancora in formato *.csv*, consiste nell'elenco di tutti i pacchetti loggati a cui erano associate tutte le informazioni registrate durante gli esperimenti.

A partire dal *.csv* generato da *Kettle*, utilizzando il tool *Mondrian* è stato possibile definire uno *star-schema*, con la struttura indicata in Figura C.5.

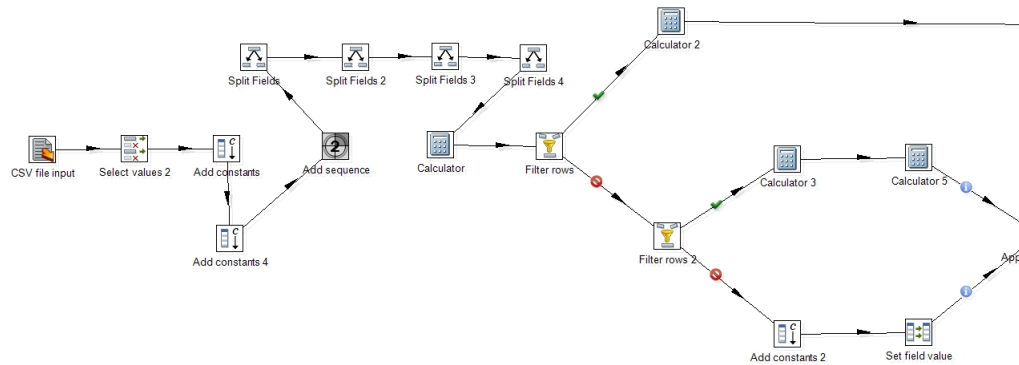


Figura C.1: AQS: trasformazione *Kettle*, prima parte.

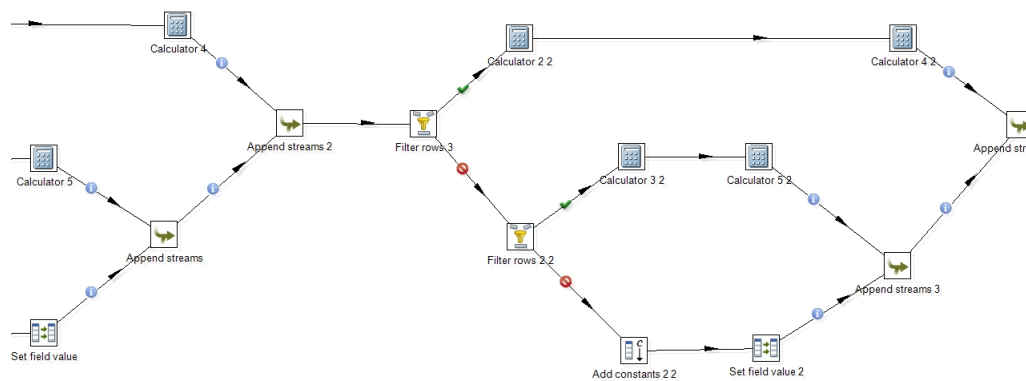


Figura C.2: AQS: trasformazione *Kettle*, seconda parte.

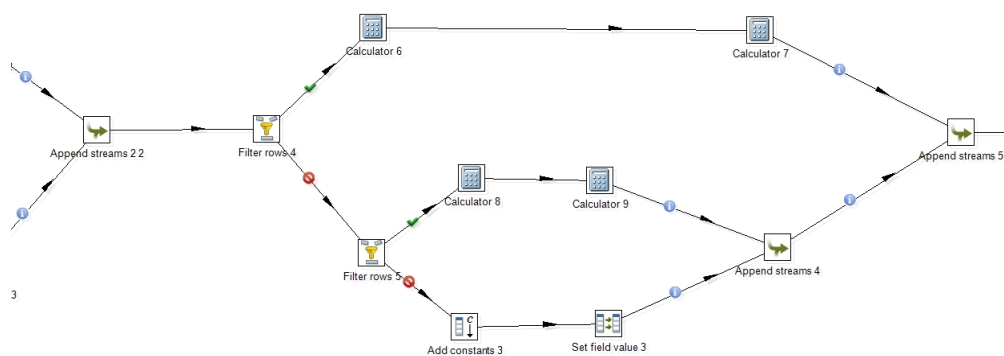


Figura C.3: AQS: trasformazione *Kettle*, terza parte.

Questa struttura ci ha permesso di modellare il fatto *Logged Packet*. L'analisi, infatti, è volta ad analizzare come Snort reagisce e varia le proprie prestazioni al variare sia di

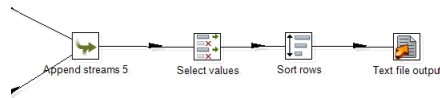


Figura C.4: AQS: trasformazione *Kettle*, quarta parte.

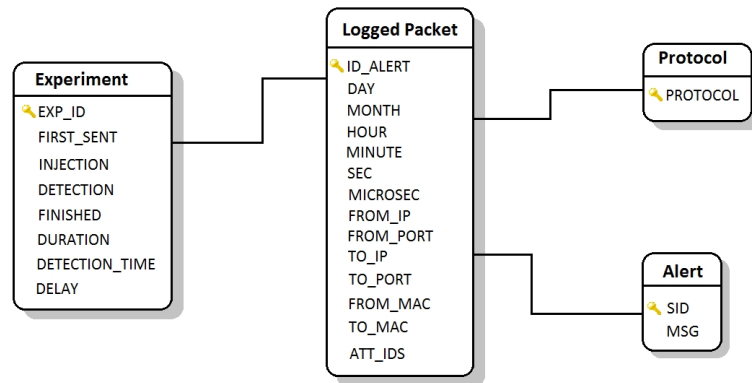


Figura C.5: *Schema a stella* costruito sul file di log finale.

tool di *fault injection* che, pur rimanendo all'interno dello stesso tool, delle opzioni attive. A partire da questo schema a stella, eseguendo sostanzialmente gli stessi procedimenti descritti nelle sezioni precedenti abbiamo eseguito tutte le analisi del caso, ottenendo grafici e tutte le considerazioni necessarie per una completa caratterizzazione del comportamento di Snort. Alcuni esempi a riguardo sono riportati nelle Figure C.6 e C.7.

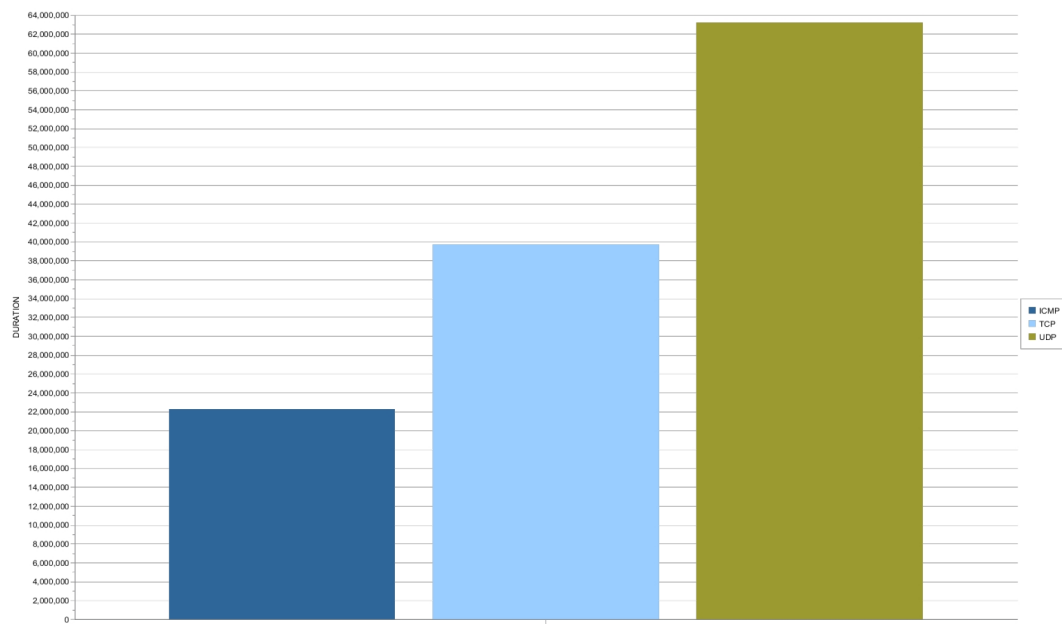
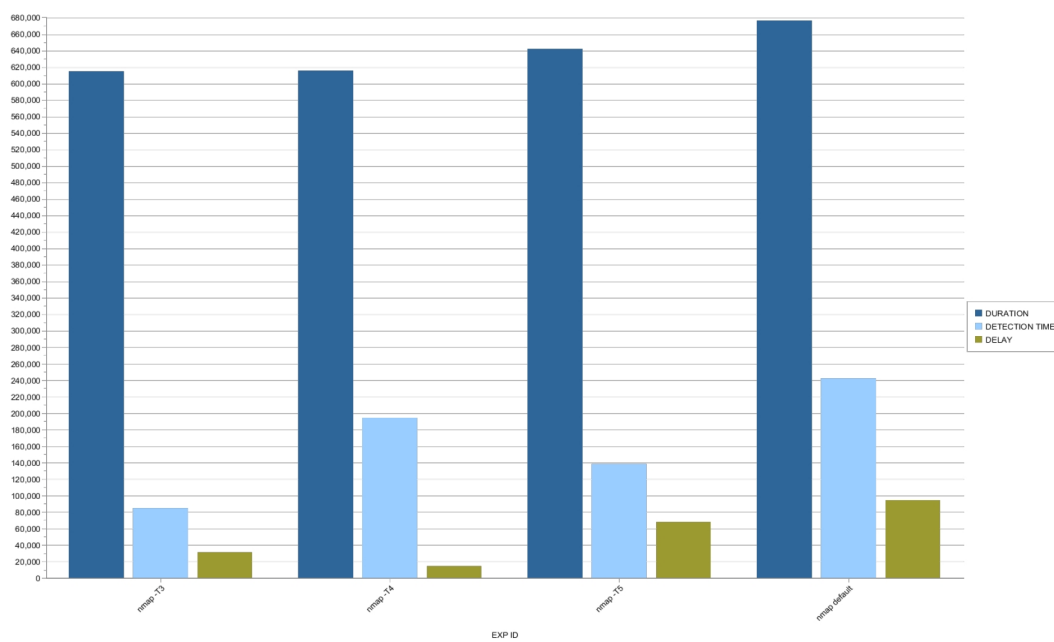


Figura C.6: Durata media degli attacchi per protocollo.

Figura C.7: Tempistiche dell'attacco *nmap* con opzione *-T*.

Bibliografia

- [1] Alessandro Gori. Data Warehousing (slides), 2012-2013. URL <http://e-l.unifi.it/course/view.php?id=2329>.
- [2] Stefano Rizzi and Matteo Golfarelli. *Data Warehouse. Teoria e pratica della progettazione*. McGraw-Hill Companies, 2006. ISBN 8838662916. URL <http://www.amazon.com/Warehouse-Teoria-pratica-della-progettazione/dp/8838662916%3FSubscriptionId%3D0JYN1NVW651KCA56C102%26tag%3Dtechkie-20%26linkCode%3Dxm2%26camp%3D2025%26creative%3D165953%26creativeASIN%3D8838662916>.
- [3] Wikipedia. Flat file database, Agosto 2013. URL http://en.wikipedia.org/wiki/Flat_file_database.