

Analisi dei dati relativi alle entrate ed alle spese nelle regioni italiane





Rosario Laface Andrea Messina Andrea Pascarella Giuseppe Sofo

1. Introduzione

Questo eleborato si pone come obiettivo quello di estrarre conoscenza tramite l'analisi dei dati delle entrate e delle spese pubbliche relative alle regioni italiane. In particolare, i dati saranno analizzati attraverso l'utilizzo dei software Qlik Sense e Tableau, i quali permettono di effettuare analisi descrittiva e diagnostica. Come conseguenza, verranno realizzate delle dashboard, ovvero schermate contenenti grafici di ogni tipo (istogrammi, diagrammi a torta, etc.).

Il fine della attività descrittiva è quello di analizzare la realtà di riferimento, descrivendone eventi passati. Quello della attività diagnostica, invece, è effettuare una diagnosi della realtà precedentemente descritta, preoccupandosi di trovare la causa che ha portato al verificarsi di un particolare fenomeno.

2. Struttura del Dataset

La sorgente da cui sono stati estrapolati i dati è l'Unità Tecnica Centrale dei Conti Pubblici Territoriali (CPT) e Nuclei Regionali della Rete CPT.

In partenza, sono stati ricavati due differenti dataset, uno relativo alle entrate e l'altro alle spese. Essi riportano i flussi di spese/entrate erogati/percepiti dalla totalità degli enti pubblici di una determinata regione italiana, per uno specifico anno finanziario. In particolare, la copertura temporale dei dataset interessa gli anni che intercorrono dal 2000 al 2015.

Inoltre è stato ricavato un terzo dataset, che raccoglie la popolazione relativa ai Comuni italiani.

Universo di riferimento	Tipologia di Soggetti	Anno di Riferimento	Macro Area a 2	Regione	Categoria livello 5	E - Consolidato PA Euro/000
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 00Z - TRIBUTI PROPRI	28.229,50
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 01 - Imposte dirette	15.342,64
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 02 - Imposta sul reddito persone fisiche (IRPEF)	10.662,33
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 03 - Imposte indirette	12.836,67
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 04 - Imposta sul valore aggiunto (IVA)	7.599,81
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 05 - Altri tributi propri	50,19
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 06 - Redditi da capitale	1.089,00
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 07 - Contributi sociali	12.286,24
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 08 - Vendita di beni e servizi	255,05
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 08Z - TRASFERIMENTI IN CONTO CORRENTE	319,56
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 09 - Trasf. in conto corrente da Unione Europea e altre istituzioni estere	0,06
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 10 - Trasf. in conto corrente da famiglie e istitituzioni sociali	67,2
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 11 - Trasf. in conto corrente da imprese private	252,29
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 12 - Trasf. in conto corrente da imprese pubbliche nazionali	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 12Z - TRASF. IN CONTO CORRENTE DA ENTI PUBBLICI	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 13 - Trasf. in conto corrente da Stato	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 14 - Tributi devoluti da altre Amministrazioni pubbliche	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 15 - Trasf. in conto corrente da altri enti dell'amministrazione centrale	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 16 - Trasf. in conto corrente da Regioni e Provincie Autonome	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 17 - Trasf. in conto corrente da Province e Citta' metropolitane	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 18 - Trasf. in conto corrente da Comuni	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 19 - Trasf. in conto corrente da A.S.L., Aziende ospedaliere e IRCCS	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 20 - Trasf. in conto corrente da Consorzi e Forme associative	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 21 - Trasf. in conto corrente da Aziende, Istituzioni, Societa' e fondazioni part	
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	E - 22 - Trasf. in conto corrente da Comunita' Montane e altre Unioni di enti loca	

2.1. Struttura del primo Dataset

Figura 1- Dataset delle entrate

Nel dettaglio, il dataset delle entrate fornisce i seguenti attributi:

- **Universo di riferimento** tale campo contiene dati relativi alla Pubblica Amministrazione (PA);
- **Tipologie di soggetti** distingue le varie tipologie di Uffici Pubblici (Amministrazioni Centrali, Amministrazioni Locali e Amministrazioni Regionali);
- Anno di riferimento;

- Macro Area a 2 settore regionale coinvolto (Centro Nord e Mezzogiorno);
- Regione;
- **Categoria livello 5** tipologia di settore relativo alle entrate (Imposte dirette, Contributi sociali, IVA, etc.);
- E Consolidato PA Euro/000 valore espresso in migliaia di euro.

2.2. Struttura del secondo Dataset

Universo di riferimento	Tipologia di Soggetti	Anno di Riferimento	Macro Area a 2	Regione	Settore	Categoria livello 5	S - Consolidato PA Euro/000
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 06 - Spese di personale	218,32
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 08 - Oneri sociali	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 09 - Ritenute IRPEF	30,02
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 12 - Acquisto di Beni e Servizi	262,58
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 13 - Contratti di servizio pubblico con IPN e/o IPL	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 14Z - Trasferimenti in conto corrente	48,88
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 15 - Trasf. in conto corrente a famiglie e istituzion	48,77
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 16 - Trasf. in conto corrente a imprese private	0,12
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 17 - Trasf. in conto corrente a imprese pubbliche	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 19Z - Trasf. in conto corrente a enti pubblici	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 21 - Trasf. in conto corrente a Stato	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 22 - Trasf. in conto corrente a altri enti dell'am	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 23 - Trasf. in conto corrente a Regioni e Provin	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 24 - Trasf. in conto corrente a Province e Citta'	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 25 - Trasf. in conto corrente a Comuni	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 26 - Trasf. in conto corrente a A.S.L., Aziende os	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 27 - Trasf. in conto corrente a Consorzi e Form	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 28 - Trasf. in conto corrente a Aziende, Istituzio	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 29 - Trasf. in conto corrente a Comunita' Mont	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 30 - Trasf. in conto corrente a enti dipendenti	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 31 - Trasf. in conto corrente a altri enti dell'Am	0
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 33 - Interessi passivi	176,74
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 35 - Poste correttive e compensative delle entrate	1.837,45
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 37 - Somme di parte corrente non attribuibili	16,13
PA	Amministrazioni Centrali	2000	91 - Centro-Nord	01 - Piemonte	00001 - Amministrazione Generale	S - 40Z - TOTALE SPESE CORRENTI	2.560,09

Figura 2 - Dataset delle Spese

Il dataset delle spese consta dei seguenti attributi:

- **Universo di riferimento** tale campo contiene dati relativi alla Pubblica Amministrazione (PA);
- *Tipologie di soggetti* distingue le varie tipologie di Uffici Pubblici (Amministrazioni Centrali, Amministrazioni Locali e Amministrazioni Regionali);
- Anno di riferimento;
- Macro Area a 2 settore regionale coinvolto (Centro Nord e Mezzogiorno);
- Regione;
- Settore settore di attività dell'intervento pubblico (Cultura, Ambiente, Giustizia, etc.);
- **Categoria livello 5** tipologia di settore relativo alle spese (Spese di personale, Oneri sociali, Ritenute IRPEF, etc.);
- S Consolidato PA Euro/000 valore espresso in migliaia di euro.

2.3. Struttura del terzo Dataset

	COMUNIVERSO: Tutti i Comuni d'Italia [elaborazione Ancitel 2016]								
Codice ISTAT	Regione	Тіро	Provincia/Città Metropolitana	Comune	Superficie(kmq)	Pop.residente(Istat 2016)	Densità(ab/kmq		
069001	Abruzzo	Provincia	Chieti	Altino	15,33	3062,00	199,74		
069002	Abruzzo	Provincia	Chieti	Archi	28,55	2203,00	77,18		
069003	Abruzzo	Provincia	Chieti	Ari	11,39	1170,00	102,70		
069004	Abruzzo	Provincia	Chieti	Arielli	11,72	1133,00	96,70		
069005	Abruzzo	Provincia	Chieti	Atessa	110,98	10639,00	95,87		
069006	Abruzzo	Provincia	Chieti	Bomba	17,26	816,00	47,27		
069007	Abruzzo	Provincia	Chieti	Borrello	14,51	354,00	24,39		
069008	Abruzzo	Provincia	Chieti	Bucchianico	38,08	5242,00	137,66		
069010	Abruzzo	Provincia	Chieti	Canosa Sannita	13,91	1376,00	98,90		
069011	Abruzzo	Provincia	Chieti	Carpineto Sinello	29,85	608,00	20,37		
069012	Abruzzo	Provincia	Chieti	Carunchio	32,56	600,00	18,42		
069013	Abruzzo	Provincia	Chieti	Casacanditella	12,54	1307,00	104,24		
069014	Abruzzo	Provincia	Chieti	Casalanguida	13,67	922,00	67,45		
069015	Abruzzo	Provincia	Chieti	Casalbordino	46,02	6166,00	134,00		
069016	Abruzzo	Provincia	Chieti	Casalincontrada	16,00	3074,00	192,11		
069017	Abruzzo	Provincia	Chieti	Casoli	67,04	5664,00	84,49		
069018	Abruzzo	Provincia	Chieti	Castel Frentano	21,89	4347,00	198,62		
069019	Abruzzo	Provincia	Chieti	Castelguidone	15,07	382,00	25,34		
069020	Abruzzo	Provincia	Chieti	Castiglione Messer Marino	47,99	1790,00	37,30		
069021	Abruzzo	Provincia	Chieti	Celenza sul Trigno	22,68	915,00	40,35		
069022	Abruzzo	Provincia	Chieti	Chieti	59,57	51815,00	869,85		
069023	Abruzzo	Provincia	Chieti	Civitaluparella	22,46	351,00	15,63		
069024	Abruzzo	Provincia	Chieti	Civitella Messer Raimondo	12,72	848,00	66,66		
069025	Abruzzo	Provincia	Chieti	Colledimacine	11,30	202,00	17,88		

Figura 3 - Dataset Comuni Italiani

Come da immagine, gli attributi di interesse in questo dataset sono *Regione* e *Pop.residente*.

Tramite questo dataset sarà possibile ottenere successivamente, nella fase di ETL, un ulteriore dataset contenente il totale della popolazione per ogni regione italiana.

Quest'ultimo dataset viene ricavato, perché nelle analisi è necessario ottenere il valore delle entrate e delle spese pro capite per ogni regione.

3. Fase di Extract, Trasform e Load (ETL)

Il primo passo della fase di ETL consiste nell'unire, in un'unica tabella, le tre distinte tabelle relative alle entrate, facenti riferimento a quinquenni differenti (2000-2004,2005-2009,2010-2015). La stessa operazione è svolta relativamente alle spese.

Una volta eseguita l'unione, si è proceduto con la rinomina dei nomi degli attributi, in particolare:

- Macro Area a 2 è rinominato in Macro-Area;
- Categoria livello 5 diventa Categoria economica;
- da Settore a Settore economico;
- E-Consolidato PA Euro/000 in Entrate;
- S-Consolidato PA Euro/000 in Spese.

Relativamente a quanto detto nel paragrafo precedente, il terzo dataset è stato utile per ottenere un ulteriore dataset:

Comuniverso.Regione	Totale pop per regione
13 - Abruzzo	1.326.513.00
17 - Basilicata	573.694.00
18 - Calabria	1.970.521.00
15 - Campania	5.850.850.00
08 - Emilia Romagna	4.448.146.00
06 - Friuli Venezia Giulia	1.221.218.00
12 - Lazio	5.888.472.00
07 - Liguria	1.571.053.00
03 - Lombardia	10.008.349.00
11 - Marche	1.543.752.00
14 - Molise	312.027.00
01 - Piemonte	4.404.246.00
16 - Puglia	4.077.166.00
20 - Sardegna	1.658.138.00
19 - Sicilia	5.074.261.00
09 - Toscana	3.744.398.00
21 - Provincia Autonoma di Trento	538.223.00
10 - Umbria	891.181.00
02 - Valle d'Aosta	127.329.00
05 - Veneto	4.915.123.00
22 - Provincia Autonoma di Bolzano	521.825.00

Figura 4 - Dataset Popolazione

esso riporta una tabella composta dalla regione con la corrispondente popolazione. Va sottolineato come, nel dataset in questione, è riportata una distinzione politica, e non geografica, delle varie regioni: infatti compaiono Provincia Autonoma di Trento e Provincia Autonoma di Bolzano.

In seguito, sono stati risolti alcuni problemi di sinonimie tra attributi dei diversi dataset. Infine, i dataset Entrate, Spese e Popolazione sono stati combinati tramite operazione di Join, al fine di ottenere un unico, grande database contenente tutte le informazioni necessarie alle analisi.

4. Dimensioni e Misure

Descritti i dataset, all'interno degli stessi si è fatta distinzione tra dimensioni e misure. In particolare, i dati sono stati replicati al fine di costruire le varie gerarchie necessarie ai fini delle analisi. Le dimensioni individuate sia per le entrate sia per le spese sono:

- Anno;
- Biennio;
- Quadriennio;
- Regione;
- Macro-Area;
- Tipologia di soggetti;
- Universo di riferimento;
- Categoria economica;
- Settore.

Anno, Biennio e Quadriennio sono legate da una relazione di gerarchia, così come Regione e Macroarea, e Tipologia di soggetti e Universo di riferimento.

Allo stesso modo, sono state individuate le seguenti misure:

- Entrate;
- Spese;
- Entrate pro capite;
- Spese pro capite;
- Differenza Entrate/Spese pro capite;
- Spese Regione Calabria;
- Spese per Ente.

La scelta del modello logico a supporto del modello multidimensionale è ricaduta sullo Schema a Stella per due differenti ragioni: lo Schema a Stella si presta al caso di analisi batch-oriented (e non real time), come quello in questione, e, in più, Qlik Sense è un software di Big Data BI. Di seguito, si riportano gli Schemi a Stella individuati.



Figura 6 - Schema a stella Entrate

Figura 5 - Schema a Stella Spese

Questi schemi sono stati implementati tramite Qlik Sense e Tableau e si è proceduto con lo svolgimento delle analisi.

5. Analisi

5.1. I Analisi – Entrate e Spese Pro capite

In questa analisi, sono state evidenziate le entrate e le spese relative al singolo cittadino di ogni regione in un arco di tempo selezionabile a partire dalla opzione Quadriennio all'opzione Anno. Le dimensioni coinvolte sono sia quelle temporali (dunque Anno, Biennio e Quadriennio) sia quelle spaziali (Regione e Macro area). Di seguito si riportano le due analisi:



Figura 7 - Entrate pro capite su Qlik Sense



Figura 8- Entrate pro capite su Tableau

In *Figura 7 e 8* è mostrato un istogramma in cui vengono riportati i valori delle entrate relativi ad ogni regione elencati separatamente per ogni quadriennio. In particolare, è facile notare che i quadrienni di maggiore picco sono relativi alla Valle d'Aosta (2004-2007) e alla Provincia Autonoma di Bolzano (2008-2011).

Tramite operazione di drill-down è possibile indagare su quali siano i bienni che contribuiscono maggiormente a tali massimi: per la prima il biennio 2006-2007, per la seconda 2010-2011.

Il minimo livello di Entrate si è registrato per la regione Calabria nel biennio 2012-2013.



Figura 9 - Spese pro capite su Qlik Sense



Figura 10 - Spese pro capite su Tableau

In *Figura 9* e *10* è mostrata un'analisi del tutto analoga alla precedente ma relativa al denaro speso per ogni singolo cittadino nello stesso periodo temporale. Si evince che la regione che ha investito di più è la Valle d'Aosta nel quadriennio 2008-2011. Di contro, quelle più "restie" all'investimento sono la Campania, nel biennio 2000-2001, e la Puglia nello stesso arco di tempo.

5.2. II Analisi – Differenza tra Entrate e Spese pro capite

Sulla base dei risultati delle due precedenti analisi, si è pensato di sfruttarli al fine di condurre un'ulteriore analisi incentrata sulla differenza tra entrate e spese pro capite. In altre parole si vuole calcolare l'importo che ogni singola regione investe su ogni cittadino. Sono emerse situazioni di bilancio positivo (in verde) e negativo (in rosso) per le diverse regioni italiane.

Al fine di eseguire tale analisi, si è fatto utilizzo di una formula che permetta di aggiungere una nuova misura relativa proprio al parametro in oggetto, la differenza tra entrate e spese:

((Sum(Entrate)-Sum(Spese))/2)/[Totale pop per regione])*1000

L'operatore Sum viene utilizzato per sommare i valori di entrate/spese delle varie tuple aggregate per regione. La somma delle spese viene divisa per 2 in quanto all'interno della tabella è presente una tupla che riporta il totale delle spese di ciascuna regione.

Il valore finale dato dal rapporto con il numero di abitanti per regione viene moltiplicato per 1000 in quanto all'interno della tabella i valori (di entrate e spese) sono divisi per 1000 al fine di semplificarne la visualizzazione.

In seguito si riporta il grafico relativo a tale analisi (figura 11):



Figura 11 - Differenza Pro Capite Entrata-Spesa su Qlik Sense

Dal grafico, in cui viene considerato un arco temporale di 15 anni, è evidente come le regioni Molise, Sardegna, Sicilia, Basilicata e Calabria, abbiano un bilancio negativo, dunque hanno speso più di quanto hanno percepito. In *figura 12* è mostrato il particolare della regione Calabria.



Figura 12- Differenza Entrate-Spese Regione Calabria su Qlik Sense

5.3. III Analisi – Spese per Settore e Regione

La terza analisi si incentra sullo studio delle spese di ogni settore dell'amministrazione pubblica nelle aree geografiche, Centro Nord e Mezzogiorno. Il primo grafico include la dimensione "Settore" e, per ognuno di essi, mostra il valore complessivo delle spese per ognuna delle due Macro aree, come mostrato in *figura 13* e *14*.



Figura 13- Spese per Settore e Macro area su Qlik Sense



Figura 14 - Spese per Settore e Macro area su Tableau

La *figura 15* mostra, inoltre, un grafico a torta con le stesse informazioni, relativamente a tutto il territorio italiano.



Figura 15 - Diagramma a Torta su Spese massime per Settore

Tramite operazione di drill-down sul grafico in *figura 13*, è possibile scendere nel dettaglio di ogni Macro area e mostrare, dunque, per ogni particolare settore, le spese relative ad ogni singola regione, come mostrato in *figura 16*.



Figura 16 - Spese per ogni regione relative al settore dell'Amministrazione Generale

Dall'analisi si evince come la regione che investe maggiormente è la Lombardia, nel settore "Previdenza e Integrazioni Salariali". Al contrario, i settori con investimenti più bassi sono "Pesca Marittima e Agricoltura", "Turismo", "Interventi Igenico Sanitari", seguiti da "Commercio", "Opere Pubbliche" e "Lavoro".

5.4. VI Analisi – Analisi Tipologia di Soggetti

La quarta analisi è analoga alla precedente, ma è riferita ad un livello gerarchico superiore. Essa mostra le entrate e le spese dei tre settori dell'amministrazione pubblica (Amministrazioni Centrali, Locali e Regionali) distribuite tra le due Macro aree.

Come si evince dalla *figura 17*, in tutta Italia le maggiori spese sono dovute, naturalmente, all'Amministrazione Centrale in quanto è composta da un maggior numero di enti. L'informazione più rilevante che emerge da tale analisi è che il Mezzogiorno spende in misura molto minore rispetto al Centro-Nord e, inoltre, l'intera spesa del Mezzogiorno, non ricopre il solo investimento del Centro-Nord in merito all'Amministrazione Centrale.



Figura 17 - Analisi tipologia di Soggetti su Qlik Sense



Figura 18 Analisi Tipologia di Soggetti su Tableau

5.5. V Analisi – Calabria

La quinta e ultima analisi si focalizza sulla ricerca delle spese effettuate dalla Regione Calabria nei vari settori della pubblica amministrazione.

Il primo grafico (*Figura 19 e 20*) è un istogramma sul quale vengono riportate le spese di ogni settore per la Regione Calabria, sequenzialmente nei quattro quadrienni di riferimento (2000-2003, 2004-2007, 2008-2011, 2012-2015).



Figura 19 - Spese Regione Calabria su Qlik Sense



Figura 20 - Spese Regione Calabria su Tableau

Il grafico in *figura 21*, invece, permette di capire, in maniera immediata, in quale quadriennio ciascun settore dell'amministrazione pubblica si è visto il maggior picco di spese.



Figura 21 - Spese Regione Calabria su Qlik Sense

6. Analisi Predittiva

Tramite Tableau è possibile realizzare una forma di analisi predittiva sui dati in possesso.

In particolare si è cercato di prevedere l'andamento delle entrate pro capite per ciascuna regione, partendo dai dati storici utilizzati nelle precedenti analisi e relativi al periodo di tempo tra il 2000 e il 2015.

A tale scopo è stato necessario effettuare preventivamente delle operazioni di ETL sui dati, in quanto Tableau non riconosceva le date come tali ma come numeri interi. Dopo questa operazione è stato possibile rappresentare il possibile andamento delle entrate pro capite per regione, divise tra regioni del Centro-Nord e del Mezzogiorno, mediante il comando *Forecast*. I risultati di tale predizione sono mostrati in *figura 22* in funzione dell'anno, mostrando una previsione fino all'anno 2018.



Figura 22 - Analisi predittiva su entrate procapite realizzata tramite tableau

1. MongoDB

MongoDB è un DBMS NoSQL documentale utile per la memorizzazione e l'analisi di dati semistrutturati. Esso si allontana dalla struttura tradizionale basata su tabelle dei database relazionali in favore di documenti stile XML e JSON. MongoDB supporta ricerche e aggiornamenti sia sui documenti che sui singoli campi degli stessi e, inoltre, permette di includere funzioni definite dall'utente in JavaScript.

Nelle analisi condotte si è fatto utilizzo di MongoChef che offre un'interfaccia visuale sostituendosi al classico terminale a linea di comando. Tramite quest'ultimo è stato possibile effettuare delle comode operazioni di caricamento dei dati al fine di effettuare le analisi in oggetto.

Data la fonte dalla quale provengono i dati in nostro possesso, questi ultimi si presentano secondo il modello relazionale (formato CSV). La piattaforma MongoChef include dei comodi comandi per l'importazione di file strutturati e la successiva conversione in collezioni semistrutturate (vedi *figura 1*).

	- 0	×
Choose Import Options		
Import Target: dbSIA.Entrate @ ProgettoSIA localhost:27017		
Import Format:	Import Mode:	
◯ JSON (mongo shell / mongoexport)	Overwrite documents with same_id	
CSV	○ Always insert with new _id	
O Another Collection	○ Insert with new _id if _id exists	
	○ Skip documents with existing _id	
	○ Abort if _id already exists	
	< Back Next > Finish Cance	el

Figura 1 - Schermata di importazione dati esterni su MongoChef

I documenti importati prendono il nome di *Collezioni* e sono stati convertiti in formato JSON. Sono dunque pronti per essere elaborati attraverso i vari operatori di MongoDB.

2. STRUTTURA DEI DATI

I dati provengono dalla stessa sorgente dalla quale sono stati ricavati i dati delle analisi effettuate con Qlik Sense e Tableau. Si evita, perciò, di descriverne nuovamente la struttura in questo paragrafo.

In MongoDB le collezioni di documenti importati sono state organizzate in un unico database chiamato dbSIA (*Figura 2*). Ognuna di esse proviene da un singolo database relazionale e viene

✓ ₽rogettoSIA localhost:27017
🗸 📄 dbSIA
> [Entrate
> 🔟 EntrateRegione
> 順 Popolazione
> 🔟 Spese
> 🔟 UsciteRegione
> 🔟 join
> 🔟 temp
> 🔟 temp1
> 📄 local

Figura .	2 -	Collezioni	utilizzate
----------	-----	------------	------------

memorizzato in formato JSON. MongoDB formatta automaticamente i dati fornendo un identificativo univoco che identifica il singolo documento all'interno della collezione. Ogni tupla del documento relazionale sorgente viene convertita in un documento ed ogni suo attributo diventa un valore del documento. La *Figura 3* mostra un esempio della struttura dei vari documenti su MongoDB.



Figura 3 - Struttura dei documenti su MongoDB

Nei prossimi paragrafi verranno descritte le analisi effettuate su tali documenti e le relative query con le quali sono state implementate.

3. Analisi

3.1. I Analisi: Spese per Regione e Settore

La prima analisi va a ricercare la somma delle spese effettuate su ogni settore, separatamente per ogni regione. A tale scopo è necessario utilizzare l'operazione di aggregazione.

Tale operatore raggruppa valori provenienti da diversi documenti, al fine di effettuare una serie di possibili operazioni sui valori aggregati.

MongoDB prevede due paradigmi di aggregazione: la *Aggregation Pipeline* e la soluzione *MapReduce*. Nella prima analisi è stata utilizzata esclusivamente la prima.

Di seguito si riporta il codice relativo alla query in oggetto:

```
db.Spese.aggregate(
  [
   {$group : {_id : {Regione : '$Regione', Settore : '$Settore'},
        Spese :{$sum :'$Spese'}
        }
   }
}
```

Nella prima riga è presente l'operatore *aggregate*: esso introduce la pipeline, ossia una catena di operatori che è possibile applicare sui dati aggregati. In questo caso tale operatore va ad indicare che i vari documenti, relativi alla collezione *Spese*, vengono raggruppati (analogamente a ciò che avviene alle tuple di una tabella grazie all'operatore *group by* di SQL). In seguito, gli operatori che vengono utilizzati sono *group* e *sum*.

Il primo consente di assegnare un identificativo *_id* all'insieme dei valori secondo i quali è stato realizzato il raggruppamento (in questo caso Regione e Settore), dunque, di ottenere un documento identificato univocamente da ogni possibile coppia di tali attributi.

Il secondo operatore *sum* va a sommare l'attributo indicato (ossia Spese), separatamente per ogni singolo documento così realizzato.

Il risultato di tale query è mostrato in figura 4.

```
1 {
         "_id" : {
 2
               "Regione" : "08 - Emilia Romagna",
"Settore" : "00022 - Agricoltura"
 з
 4
 5
         },
"Spese" : 11618.459999999975
 6
 7 }
 8 {
         "_id" : {
 9
              " · ι
"Regione" : "11 - Marche",
"Settore" : "00020 - Viabilita'"
10
11
12
         },
"Spese" : 14043.95
13
14 }
15 {
16
         "_id" : {
              "Regione" : "16 - Puglia",
"Settore" : "00028 - Altre opere pubbliche"
17
18
19
20
21 }
           Spese": 1167.84999999999995
22 {
         "_id" : {
23
              יין
"Regione" : "01 - Piemonte",
"Settore" : "00012 - Servizio Idrico Integrato"
24
25
26
         },
"Spese" : 7497.429999999994
27
28 L
```

Figura 4 - Risultato prima analisi in formato JSON

MongoDB consente di visualizzare i risultati scegliendo tra diversi formati: JSON, albero e tabella. In questo modo è possibile avere una visione esaudiente dei risultati.

3.2. Il Analisi: Entrate e Spese Pro Capite

La seconda analisi va ad analizzare le entrate e le spese pro capite relative ad ogni regione italiana nell'intero arco di tempo considerato (2000-2015).

Di seguito, viene mostrato il codice relativo alla query che analizza le entrate:

```
db.Entrate.aggregate([
{$group : { id :{Regione :"$Regione"}, TotEntrate :{$sum:"$Entrate"}}},
{$lookup :{
    from : "Popolazione",
    localField: " id.Regione",
    foreignField: "ComuniversoRegione",
          as: "join1"
        } 

 { $unwind: "$join1" },
 {$project :{ Regione : "$Regione", TotEntrate : "$TotEntrate",
"join1.TotPopRegione" : 1}},
  {$out:"temp1"}
]);
db.temp1.find().forEach(function(t) {
     printjson("Regione= " +t. id.Regione+ "Totale Entrate Procapite= "
     +(t.TotEntrate/t.join1.TotPopRegione)*500);
 })
```

Anche in questo caso si utilizza il paradigma dell'*Aggregation Pipeline*. Trattandosi di una query che coinvolge più collezioni (Entrate e Popolazione) è necessario implementare un meccanismo di "join" tra documenti, non previsto normalmente in MongoDB in quanto database documentale.

Per tale procedura si fa utilizzo dell'operatore *lookup* che effettua una operazione di left outer join tra valori di due collezioni. Il parametro *from* specifica la collezione con cui effetture la join, *localField* indica il valore comune nella prima collezione (_id.Regione), infine, *foreignField* indica il corrispondente nella seconda collezione.

Il risultato della join viene salvato in una nuova collezione, indicata dal parametro *as* e chiamata join1.

Per ogni documento in ingresso della prima collezione, la funzione *lookup* aggiunge un nuovo campo che consiste in un array i cui elementi sono i campi del secondo documento, appartenente alla seconda collezione, specificata con il campo *from*.

Data la struttura, così ottenuta, risulta impossibile effettuare operazioni di lettura su tali risultati. E' dunque necessario scomporre gli elementi dell'array, ciò viene realizzato tramite l'operatore *unwind*.

Quest'ultimo scompone un array, contenuto in un documento, e restituisce diversi documenti, ciascuno contenenti tutti i campi non appartenenti all'array e un elemento dell'array stesso.

I comandi *lookup* e *unwind* consentono di realizzare una left outer join come se ciascun documento fosse del tutto simile ad una tabella del modello relazionale.

Nel caso in esame la prima tabella è rappresentata dai documenti della collezione Entrate, la seconda da quelli di Popolazione. *Lookup* permette di ottenere un array contenente gli attributi della seconda tabella, legati alla prima tramite l'attributo Regione.

Unwind scompone tale array inserendone gli elementi come attributi della prima tabella.

L'output così ottenuto viene indicizzato dall'operatore *group*, secondo gli attributi Regione e TotaleEntrate.

I documenti, a questo punto, contengono tutte le informazioni necessarie per poter eseguire le operazioni di raggruppamento desiderate secondo Regione, ottenendo così una nuova collezione contenente il campo Regione come attributo principale, alla quale poi vengono associati gli attributi relativi al valore della popolazione ed alle entrate complessive di ogni regione.

Per snellire il documento ed eliminare gli attributi non necessari viene utilizzato il comando *project*. Quest'ultimo, equivalente alla SELECT di SQL, seleziona solo determinati attributi, ossia: Regione, TotEntrate e TotPopRegione.

I risultati vengono esportati in una nuova collezione (denominata temp1) tramite l'operatore out.

In seguito vengono prelevati e visualizzati tramite una funzione JavaScript, invocata tramite la funzione .*find()* su ogni documento di temp1.

Le figure 5 e 6 mostrano due diversi tipi di rappresentazione dei risultati dell'analisi.



Figura 5 - Risultati seconda analisi in formato JSON

Aggregate Text 🗉	
1 "Regione= 07 - Liguria Totale Entrate Procapite= 452.68118580340774"	~
2 "Regione= 05 - Veneto Totale Entrate Procapite= 429.2573481884378"	
3 "Regione= 01 - Piemonte Totale Entrate Procapite= 452.6796981821637"	
4 "Regione= 09 - Toscana Totale Entrate Procapite= 427.56994582306595"	
5 "Regione= 02 - Valle d'Aosta Totale Entrate Procapite= 573.300151575838"	
6 "Regione= 03 - Lombardia Totale Entrate Procapite= 530.6526595945038"	
7 "Regione= 08 - Emilia Romagna Totale Entrate Procapite= 479.87515472738505"	
8 "Regione= 10 - Umbria Totale Entrate Procapite= 375.65655012842484"	
9 "Regione= 06 - Friuli Venezia Giulia Totale Entrate Procapite= 478.6415406585886"	
10 "Regione= 17 - Basilicata Totale Entrate Procapite= 285.52670064529116"	
11 "Regione= 11 - Marche Totale Entrate Procapite= 387.04423378884735"	
12 Regione= 15 - Campania Totale Entrate Procapite= 268.5108625242486"	
13 "Regione= 12 - Lazio Totale Entrate Procapite= 486.0305101221504"	
14 "Regione= 21 - Provincia Autonoma di Trento Totale Entrate Procapite= 461.747101108648"	
15 "Regione= 19 - Sicilia Totale Entrate Procapite= 262.6386620633031"	
16 "Regione= 22 - Provincia Autonoma di Bolzano Totale Entrate Procapite= 495.3708906242517"	·
17 "Regione= 13 - Abruzzo Totale Entrate Procapite= 364.48236089657604"	
18 "Regione= 14 - Molise Totale Entrate Procapite= 312.3695385335245"	
19 "Regione= 16 - Puglia Totale Entrate Procapite= 265.30711160644415"	
20 "Regione= 18 - Calabria Totale Entrate Procapite= 254.26656/06525853"	
21 "Regione= 20 - Sardegna lotale Entrate Procapite= 309.9556671398886"	
22	\vee
<	>

Figura 6 - Risultato delle analisi in formato testuale

La stessa query è stata utilizzata per individuare le spese totali di ciascuna regione. Il codice è il seguente:

```
db.Spese.aggregate([
  {$group :{ id : {Regione :"$Regione"}, TotSpese :{$sum :"$Spese"}}},
   {$lookup:
       {
         from: "Popolazione",
         localField: " id.Regione",
         foreignField: "ComuniversoRegione",
          as: "join"
        } } ,
 { $unwind: "$join" },
 {$project :{ Regione : "$Regione", TotSpese : "$TotSpese",
"join.TotPopRegione" : 1}} ,
 {$out:"temp"}
]);
db.temp.find().forEach(function(t) {
printjson("Regione= " +t._id.Regione+ "Totale Spese Procapite= "
+(t.TotSpese/t.join.TotPopRegione)*500 );
 })
```

3.3. III Analisi: Differenza Pro Capite

Questa analisi va ad analizzare la differenza spese ed entrate pro capite, per ogni Regione italiana nell'intero periodo di tempo compreso tra il 2000 ed il 2015.

Trattandosi anch'essa di un'analisi che coinvolge più collezioni, anche in questo terzo caso è necessario effettuare un'operazione di join tra tabelle.

Tuttavia, si è volutamente preferito adottare un nuovo approccio, offerto dai possibili comandi di MongoDB, il quale consente di implementare il paradigma di *MapReduce*.

Esso consente di elaborare grosse quantità di dati, prelevandole da diverse fonti o dividendo grossi dataset e, successivamente, di riunire risultati, aggregandoli.

Un esempio di implementazione della versione MongoDB di *MapReduce* è mostrato in *figura 7*:



Figura 7- Esempio implementazione MapReduce di MongoDB

Inizialmente, MongoDB applica la fase Map ad ogni collezione fornita in input. Le varie funzioni di Map vengono eseguite separatamente (potrebbero essere, teoricamente, eseguite in modo parallelo in un eventuale ambiente distribuito) e possono ricevere in input collezioni che rispettano query prestabilite.

Le funzioni Map restituiscono delle coppie chiave-valore, da tali chiavi possono dipendere valori multipli.

MongoDB applica dunque la fase di Reduce che raccoglie e aggrega i dati di output delle funzioni Map, eventualmente eseguendo su di essi una logica precedentemente definita, ottenendo come risultato una nuova collezione. Tutte le funzioni Map in MongoDB sono scritte in JavaScript e vengono eseguite solo all'interno del processo Mongod.

Il codice relativo a l'implementazione di tale paradigma è il seguente:

```
db.Spese.aggregate(
     [
        {$group : {_id : {Regione:"$Regione"}, Spese:{$sum:"$Spese"} }},
        {$out: "UsciteRegione"}
     ]
  )
  db.Entrate.aggregate(
     [
        {$group:{ id : {Regione:"$Regione"},Entrate:{$sum:"$Entrate"} }},
        {$out: "EntrateRegione"}
     1
  )
  funzioneMap1 = function() {
    emit(this. id.Regione, {Entrate: this.Entrate, Uscite:0, Popolazione: 0});
  }
  funzioneMap2 = function() {
    emit(this. id.Regione, {Entrate: 0, Uscite: this.Spese, Popolazione: 0});
  }
  funzioneMap3 = function() {
      var Regione = this.ComuniversoRegione;
     emit(Regione , {Entrate: 0, Uscite: 0, Popolazione:
  this.TotPopRegione});
  }
  r = function(key, values) {
      var result = {Entrate: 0, Uscite: 0, Popolazione: 0};
      values.forEach(function(value) {
          if (value.Entrate !== 0) {
              result.Entrate = value.Entrate;
          }
        if (value.Uscite !== 0) {
            result.Uscite = value.Uscite;
        }
        if (value.Popolazione !== 0) {
            result.Popolazione = value.Popolazione;
        }
    });
    return result;
res = db.EntrateRegione.mapReduce(funzioneMap1, r, {out: {reduce: 'join'}})
```

}

```
res = db.UsciteRegione.mapReduce(funzioneMap2, r, {out: {reduce: 'join'}})
res = db.Popolazione.mapReduce(funzioneMap3, r, {out: {reduce: 'join'}})
db.join.find().forEach(function(t) {
    printjson("Regione= " +t._id+ "Differenza Procapite= "
    +((t.value.Entrate-t.value.Uscite)/t.value.Popolazione )*500);
})
```

Nella nostra analisi i dataset utilizzati sono: Spese ed Entrate. Su questi è stata effettuata preventivamente un'operazione di *group* per poter evidenziare la somma delle entrate e delle spese relative ad ogni Regione.

Le righe di codice seguenti mostrano l'implementazione delle tre funzioni Map (funzioneMap1, funzioneMap2, funzioneMap3) e della funzione reduce (chiamata r) in JavaScript.

Le prime prendono in input i risultati degli operatori *group* e restituiscono delle coppie chiave-valore dove, la chiave è la Regione, ed il valore può essere rappresentato dalle Entrate, dalle Spese e dal valore della Popolazione.

Tali coppie vengono fornite, infine, alla funzione Reduce, la quale effettua le operazioni richieste e restituisce i valori della differenza pro capite tra entrate e spese per ogni regione. I risultati sono mostrati nella *figura 8*.

Aggregate	Aggregat	te Text Te	ext Text	Text	Document	Document	Document	Text ≈		
1 "Region	e= 01 -	Piemonte	Di	fferei	nza Procap	ite= 103.8	2131356877	7342"		~
2 "Region	e= 02 -	Valle d'A	osta	Di	fferenza P	rocapite=	60.4182472	217840	3"	
3 "Region	e= 03 -	Lombardia	D	iffere	enza Proca	pite= 182.	9434824864	17648"		
4 "Region	e= 05 -	Veneto	Diff	erenza	a Procapit	e= 147.244	2815774927	79"		
5 "Region	e= 06 -	Friuli Ve	nezia Gi	ulia	Diff	erenza Pro	capite= 44	1.4952	82578534194"	
6 "Region	e= 07 -	Liguria	Dif	feren	za Procapi	te= 59.775	5900023680	955"		
7 "Region	e= 08 -	Emilia Ro	magna	D	ifferenza	Procapite=	153.28991	L67428	429"	
8 "Region	e= 09 -	Toscana	Dif	feren	za Procapi	te= 99.514	2570314383	39"		
9 "Region	e= 10 -	Umbria	Diff	erenza	a Procapit	e= 39.8624	353526295	5"		
10 "Region	e= 11 -	Marche	Diff	erenza	a Procapit	e= 82.0075	763464553			
11 "Region	e= 12 -	Lazio	Diffe	renza	Procapite	= 76.69127	661642086			
12 "Region	e= 13 -	Abruzzo	Dif	feren	za Procapi	te= 55.137	3073614736	595"		
13 "Region	e= 14 -	Molise	Diff	erenza	a Procapit	e= -5.3182	0964211207	75"		
14 "Region	e= 15 -	Campania	Di	ffere	nza Procap	ite= 15.16	64611124883	L835"		
15 "Region	e= 16 -	Puglia	Diff	erenza	a Procapit	e= 7.01561	8201462549)"		
16 "Region	e= 17 -	Basilicat	a	Diffe	renza Proc	apite= -12	.06060199	339956	4"	
17 "Region	e= 18 -	Calabria	Di	fferei	nza_Procap	ite= -36.8	34642284959	965"		
18 "Region	e= 19 -	Sicilia	Dit	feren:	za Procapi	te= -10.15	9324678012	2968"		
19 "Region	e= 20 -	Sardegna	D1	ttere	nza Procap	1te= -9.22	28242763876	343"		
20 "Region	e= 21 -	Provincia	Autonom	a di	Irento	Differe	enza Procap	oite=	58.040/56340/65235"	
21 "Region	e= 22 -	Provincia	Autonom	a di l	Bolzano	Ditter	enza Proca	apite=	103./8544531212091	L
22										
										\sim
<										>

Figura 8 - Risultati della funzione Reduce

3.4. IV Analisi: Calabria

Il quarto e ultimo caso in esame riguarda l'analisi delle spese della regione Calabria. Il codice utilizzato è il seguente:

Tramite ciò si vuole rappresentare il valore massimo delle spese selezionate per categoria economica all'interno della regione.

E' stato deciso di considerare il valore massimo in quanto i documenti sorgente contengono una lista delle spese effettuate in un settore, insieme al totale relativo ad esse.

Il codice utilizzato è molto simile a quello presente nella prima analisi. In aggiunta ad esso è stato utilizzato l'operatore *match* (anch'esso facente parte del paradigma di *aggregation pipeline*) che filtra la collezione in input restituendo solo i documenti indicati dalla condizione specificata (la regione Calabria).

Il valore relativo ad ogni Regione e Categoria Economica, ottenuto tramite l'operatore *max*, è mostrato in *figura 9*:

```
Aggregate 3
₩ 🛯 🕨 50

    Documents 1 to 50

                                                                                                                  JSON View 🖂
                                                                                                                               龘
 1{
                                                                                                                                ~
       "_id" : {
 2
            "Settore" : "00001 - Amministrazione Generale",
"Categoria_Economica" : "S - 09 - Ritenute IRPEF"
 3
 4
 5
        "MaxSpese" : 43.56
 6
 7}
 8 {
 9
       "_id" : {
            "Settore" : "00019 - Altri trasporti",
10
            "Categoria_Economica" : "S - 56 -
11
                                                          Trasf. in conto capitale a Regioni e Provincie Autonome"
12
       },
"MaxSpese" : 0.0
13
14 }
15 {
16
       " id" : {
            "Settore" : "00016 - Altri interventi igenico sanitari",
17
            "Categoria_Economica" : "S - 37 - Somme di parte corrente non attribuibili"
18
19
        "MaxSpese" : 0.47
20
21 }
22 {
       "_id" : {
23
            "Settore" : "00022 - Agricoltura",
24
            "Categoria_Economica" : "S - 57 -
25
                                                           Trasf. in conto capitale a Province e Citta' metropolitane"
26
27
       },
"MaxSpese" : 0.0
28 }
29 {
       "_id" : {
30
            "Settore" : "00001 - Amministrazione Generale",
31
            "Categoria_Economica" : "S - 06 - Spese di personale"
32
 33
       3.
```

Figura 9 - Risultati quarta analisi.