

Un'ondata di colore sui "capelli d'argento"

Autori

Linda Pimpinelli, Ignazia Vitali - Giallo Dat@
Guido Cuzzocrea, Salvina Piccarreta - NUNATAC

Abstract

Chi stava andando "verso la vetta" dove è arrivato? I "single in carriera" hanno deciso di mettere la testa a posto? Infine, i "top & vip" sono ancora al loro posto?. Giallo Dat@ ha realizzato un progetto per rinfrescare la colorazione delle 320.000 microzone censuarie italiane, a quasi dieci anni dall'ultimo censimento ISTAT e quasi cinque dall'uscita sul mercato della Segmentazione Geodemografica Cluster SEAT, basata su tali dati.

Le fonti di dati a disposizione per il raggiungimento dell'ambizioso obiettivo sono eterogenee come livello di dettaglio, coprono diversi ambiti informativi, sono misurate su campioni estremamente numerosi, ma non sovrapposti.

Il paper presenterà i risultati ottenuti, descrivendo la metodologia che ha permesso di utilizzare al meglio le diverse fonti e di costruire un processo di classificazione al contempo robusto e ricco di contenuti informativi: gli utenti SAS potranno seguire un percorso realizzativo che sfrutta al massimo le potenzialità di integrazione e la completezza delle tecniche di analisi disponibili nel sistema. Per gli utenti marketing non mancheranno certo le sorprese, un'ondata di colore sui "capelli d'argento".

1. Premessa

A partire dal 1999 Giallo Dat@ ha avviato un progetto di raccolta di informazioni denominato **Ricerca Europea** sugli Stili di Consumo, che ha portato alla rilevazione di circa **400.000 questionari**. Le informazioni raccolte sono di tipo *anagrafico* (età, professione, titolo di studio...), *comportamentale* (sport, letture, modalità di pagamento ...), *di consumo* (prodotti acquistati, beni posseduti, marche preferite...). Il canale di distribuzione dei questionari utilizzato è stato, prevalentemente, l'inserimento nelle Pagine Gialle. Come incentivo alla compilazione sono stati proposti sorteggi di premi vari.

Oltre alla Ricerca Europea, Giallo Dat@ ha a disposizione ulteriori fonti informative aggiornate, classificate rispettivamente in fonti interne e fonti esterne. Nello specifico,

Fonti interne:

- Archivio delle utenze telefoniche private SEAT, anno 2000;
- Classificazione socio-demografica CLUSTER91 delle microzone;
- Archivio degli operatori economici SEAT, anno 2000;

Fonti esterne:

- Banca dati dei possessori di auto e relative caratteristiche (fonte Targa Services).

- Indagine telefonica ad-hoc allo scopo di validare la rappresentatività a livello totale Italia della Ricerca Europea.

A distanza di nove anni dalla data di rilevazione dei dati censuari e a meno di cinque anni dall'uscita sul mercato della Segmentazione Geodemografica CLUSTER91, basata su tali dati, è nata l'esigenza di aggiornare tale classificazione sfruttando il patrimonio di informazioni raccolte. Il progetto si pone come obiettivo quello di produrre una nuova segmentazione a livello di microzona basata sia su variabili socio-demo, garantendo così continuità rispetto alla segmentazione '91, sia sugli stili di consumo rilevati dalla Ricerca Europea. In questo modo è possibile colorare ciascuna microzona in uno dei 19 Segmenti ottenuti identificandone lo stile di consumo e gli aspetti demografici delle famiglie che in essa risiedono. Il nuovo prodotto diventa un utile strumento sia per le attività di Geomarketing che per quelle di Target Analysis quali, ad esempio:

- qualificare porzioni di territorio
- localizzare clienti potenziali
- costruire aree di potenzialità commerciale
- ottimizzare la composizione dell'offerta di prodotti nei punti vendita
- segmentare archivi di prospect quali il file telefonico
- segmentare ed arricchire base dati clienti

2. Presentazione di Nunatac e Giallo Dat@

2.1 Giallo Dat@ è la nuova società di Seat Pagine Gialle Spa, in cui sono confluite le attività di business information e direct marketing della divisione Seat Direct. Con la sua rete commerciale, presente capillarmente sul territorio nazionale, è in grado di fornire ai propri clienti servizi completi e innovativi di marketing diretto e di business information ad aziende di qualsiasi dimensione e settore di attività dal Largo Consumo al Finanziario, dalla Pubblica Amministrazione al No Profit.

In particolare i principali servizi offerti sono:

- Commercializzazione di banche dati e informazioni business-to-business e business to consumer
- Gestione delle banche dati dei propri clienti e realizzazione di campagne di direct marketing
- Consulenza e supporto decisionale per l'elaborazione di strategie di marketing, attraverso strumenti applicativi di target analysis, geomarketing, analisi statistico-territoriali e cartografiche.

2.2 Nunatac

NUNATAC è una società Quality Partner a livello europeo di SAS Institute, costituita da un gruppo di consulenti con competenze di:

- metodi statistici applicati (datamining),
- sviluppi informatici per i sistemi di supporto alle decisioni,
- target marketing.

La caratteristica distintiva di NUNATAC è la combinazione di competenze specifiche e di capacità realizzativa in ambito Database Marketing: avere responsabilità progettuale significa, nel nostro approccio, proporre soluzioni attraverso la consulenza direzionale, realizzare sistemi e formare i nostri clienti sui metodi utilizzati. Durante gli ultimi 7 anni, Nunatac ha sviluppato soluzioni verticali di Database Marketing nei seguenti settori: Banche, Assicurazioni, Vendita per Corrispondenza e Industria Automobilistica.

3. Il progetto di riclassificazione delle microzone

3.1 Considerazioni preliminari

I dati utilizzati per la costruzione dei cluster comportamentali sono stati attinti da fonti GIALLO DAT@ e fonti esterne, con il proposito di sfruttare al massimo il patrimonio informativo sugli Stili di Consumo rilevato attraverso la Ricerca Europea.

Considerazioni effettuate e validate nella fase di impostazione del progetto hanno portato alle seguenti conclusioni:

⇒ effettuare la Segmentazione a livello territoriale di microzona.

⇒ effettuare un'approfondita analisi di copertura campionaria, per poi individuare un percorso di analisi finalizzato alla classificazione delle microzone, nel rispetto di criteri di robustezza e affidabilità statistica.

⇒ per considerazioni inerenti il time-to-market, la disponibilità di dati nell'immediato e la presenza di sub-mercati eterogenei nel mercato Italia, suddividere il progetto in due successive fasi: la prima fase riguarda la classificazione delle microzone appartenenti all'insieme delle 15 città capoluogo, ritenute le piazze più importanti. La seconda fase riguarda la classificazione delle microzone del resto d'Italia.

⇒ va detto che le 15 città, i restanti capoluoghi di provincia e il resto dell'Italia rappresentano una stratificazione dell'universo, rispetto alle caratteristiche socio-demografiche e abitative descritte attraverso la segmentazione CLUSTER91. Il trattamento separato dei sub-universi è perciò un fatto auspicabile dal punto di vista metodologico, oltre che commerciale.

⇒ la problematica di copertura campionaria a livello di microzone, unitamente alla trattazione separata di alcuni sub-universi, richiede l'individuazione di un percorso di classificazione stabile e ripetibile: non si tratta semplicemente di clusterizzare dinamicamente un insieme di unità statistiche, bensì di individuare una regola generale di classificazione, applicabile ai diversi strati dell'universo, mantenendo un denominatore comune nell'interpretazione dei segmenti.

3.2 Completezza informativa e robustezza dei risultati dell'analisi

A partire dalle considerazioni precedenti, possono essere tratte alcune conclusioni sulle scelte metodologiche più adatte per ottenere una soluzione soddisfacente alla problematica di classificazione del progetto di riclassificazione delle microzone.

Il patrimonio informativo sugli stili di vita e i comportamenti di consumo degli italiani, rilevato attraverso la Ricerca Europea, costituisce la principale componente innovativa nei contenuti della segmentazione: mentre la segmentazione CLUSTER91 era basata unicamente sulle caratteristiche socio-

demografiche e abitative delle famiglie, questa volta disponiamo di un vasto insieme di indicatori riguardanti i seguenti ambiti:

- le abitudini culturali,
- l'utilizzo di prodotti e servizi finanziari,
- il livello di informatizzazione,
- la propensione alle donazioni ad associazioni non a fini di lucro,
- i canali distributivi preferiti per l'acquisto di generi alimentari e non,
- i principali hobbies del tempo libero e i modi di trascorrere le vacanze.

La prima scelta metodologica importante è la decisione di utilizzare i consumi e gli stili di vita come variabili di analisi, mentre le caratteristiche socio-demografiche come variabili supplementari, necessarie per meglio descrivere i profili comportamentali e renderli individuabili.

Nell'universo delle 15 città, la porzione di microzone aventi un livello di copertura campionaria considerato sufficiente, in termini di affidabilità statistica, è intorno al 20%. Al fine di irrobustire l'analisi, meglio sfruttare la quasi totalità dei questionari rilevati attraverso la Ricerca Europea e, infine, determinare una regola di segmentazione comportamentale estendibile all'intero universo Italia, si è deciso di operare come segue:

- è stato definito un concetto di Macrozone, strumentale all'analisi, aggregando le microzone aventi il medesimo CAP e il medesimo Segmento CLUSTER91. Le Macrozone così ottenute risultano omogenee sia rispetto a criteri di vicinanza geografica, sia rispetto alla somiglianza socio-demografica delle famiglie residenti.
- Sull'universo delle Macrozone così determinate, è possibile individuare le dimensioni rilevanti per caratterizzare i comportamenti di consumo e gli stili di vita. Una volta determinate tali dimensioni, i *factor score coefficients* ad esse associati possono essere applicati alle specifiche coordinate delle variabili originarie delle microzone: si tratta, infatti, di presenze percentuali di singole caratteristiche (e.g.: % lettori de Il Sole 24 Ore), normalizzate rispetto al numero di rispondenti nell'unità geografica considerata.
- Allo stesso modo ci si comporta per la regola di cluster, costruita a livello di Macrozone e applicata alle microzone, in base alle specifiche coordinate.
- La logica è che si stima una regola sull'universo più robusto delle Macrozone e la

si applica alle microzone, rispettando la tipicità di ciascuna di esse.

Con il vincolo di continuare a considerare il livello di copertura campionaria considerato minimale per l'affidabilità statistica dei risultati, la classificazione delle microzone è avvenuta come segue:

- in base alla regola generale di classificazione, stimata a livello di Macrozone, sono state classificate, considerando le proprie coordinate comportamentali, le microzone con un livello di copertura sufficiente (circa il 20% del totale microzone);
- sono state classificate, ereditando il cluster della Macrozone cui appartengono, le microzone con copertura campionaria insufficiente, facenti parte di Macrozone coperte (un ulteriore 60% del totale microzone).

3.3 Utilizzo delle banche dati esogene per la classificazione delle microzone non coperte.

Per quanto riguarda le unità territoriali aventi un livello di copertura campionaria considerato insufficiente, in termini di affidabilità statistica, sia a livello di microzone, sia a livello di macrozone, si è dovuto ricorrere a un metodo diverso per la clusterizzazione. A questo scopo si è deciso di costruire un modello discriminante basato su

- Banca dati dei possessori auto
- Archivio degli operatori economici SEAT
- Analisi socio-demografica effettuata nel '91,

Per maggiore robustezza dell'analisi, si è deciso di costruire il modello dal 20% di microzone sufficientemente coperte.

Attraverso il modello discriminante basato sulle fonti esterne è stato possibile classificare la quasi totalità delle microzone, che non era stato possibile colorare con i soli dati della Ricerca Europea.

3.4 La clusterizzazione a due stadi

Il patrimonio informativo connesso ai consumi e gli stili di vita, rilevato attraverso la Ricerca Europea è stato studiato utilizzando le tecniche di analisi fattoriale, al fine di individuare le dimensioni rilevanti dei fenomeni oggetto di studio.

Sono state effettuate analisi fattoriali per singolo ambito informativo e per unione di

alcuni ambiti (vedi paragrafo 3.2). I fattori così ottenuti rappresentano l'input per la successiva analisi cluster. In particolare, la fattoriale sull'unione di alcuni ambiti è stata utilizzata: per ridurre il numero di variabili in input all'algoritmo di clusterizzazione, per assegnare un peso relativo maggiore agli ambiti Cultura, Tecnologia e Servizi Finanziari, altrimenti soffocati da un numero eccessivo di fattori "sovrastrutturali". I fattori più analitici, costruiti sui singoli ambiti, vengono comunque utilizzati per tracciare il profilo a posteriori dei segmenti ottenuti.

Sulla base di 27 fattori di sintesi dei consumi e degli stili di vita, si è pervenuti ad una regola di classificazione in 15 cluster. Tali cluster risultano avere una connotazione descrittiva socio-demografica molto ben definita, ad eccezione dei 4 cluster più numerosi: all'interno di ciascuno di questi, è stata operata un'ulteriore spaccatura, in base alle caratteristiche di numero componenti la famiglia e presenza delle diverse fasce di età all'interno del nucleo. Il numero totale di cluster cui si è pervenuti è 19.

4. Conclusioni e risultati ottenuti.

Per descrivere il profilo dei 19 cluster sono state impiegate sia le variabili utilizzate nell'analisi sia le variabili derivanti dalle altre banche dati. La disponibilità di un così vasto patrimonio informativo ha permesso di tracciare un profilo "a tutto tondo", comportamentale e socio-demografico, delle micro unità territoriali che caratterizzano il mercato italiano.

Di grande interesse è la possibilità di misurare i fenomeni migratori dai segmenti CLUSTER91 a quelli aggiornati: le nostre città, così come le province, sono cambiate insieme ai loro abitanti e alle loro caratteristiche.

Tra i risultati di sintesi, vale la pena di analizzare la distribuzione percentuale dei Segmenti 2000, ordinati per status economico, nelle 15 città capoluogo principali:

| Segmento | %Micro zone | %Famiglie | %Utenze Tel. |
|----------|-------------|-----------|--------------|
| RC_01 | 7,28% | 5,65% | 6,21% |
| RC_02 | 4,51% | 3,96% | 4,14% |
| RC_03 | 6,81% | 8,56% | 8,67% |
| RC_04 | 4,06% | 4,50% | 4,78% |
| RC_05 | 4,83% | 5,41% | 5,83% |
| RC_06 | 6,32% | 7,60% | 8,04% |
| RC_07 | 3,01% | 2,50% | 2,67% |
| RC_08 | 4,07% | 3,95% | 4,10% |
| RC_09 | 2,26% | 1,66% | 1,76% |
| RC_10 | 5,10% | 6,61% | 6,81% |
| RC_11 | 3,20% | 3,69% | 3,82% |
| RC_12 | 6,38% | 8,77% | 9,15% |
| RC_13 | 4,30% | 3,65% | 3,73% |
| RC_14 | 2,41% | 2,16% | 2,04% |
| RC_15 | 7,50% | 8,45% | 8,38% |
| RC_16 | 6,15% | 6,81% | 6,23% |
| RC_17 | 4,34% | 5,40% | 5,30% |
| RC_18 | 3,27% | 3,75% | 3,71% |
| RC_19 | 6,68% | 4,91% | 3,84% |
| NC | 7,53% | 2,01% | 0,78% |

In Figura è riportata un grafico, dove sono rappresentate le posizioni relative dei 19 cluster, rispetto a 4 variabili, costruite dalla combinazione lineare e dalla standardizzazione delle variabili usate. Più precisamente, sull'asse orizzontale è rappresentata una variabile legata alla ricchezza del cluster, su quello verticale, una variabile legata alla tendenza del cluster ad aderire ad iniziative di tipo culturale, mentre l'ampiezza del simbolo rappresenta la numerosità delle famiglie e il colore del simbolo diventa più chiaro, quando crescono le conoscenze tecnologiche manifestate attraverso le risposte al questionario.

