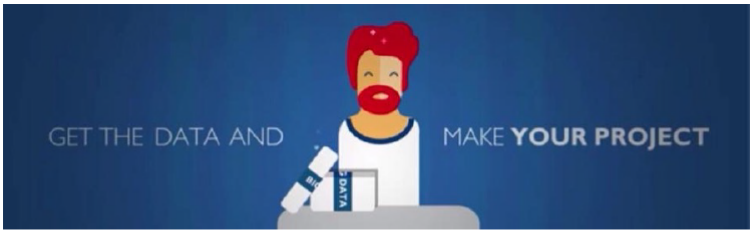
Relazione di approfondimento

Secondo Progetto nell'ambito del corso di Big Data
(Anno accademico 2014/2015)



BIG DATA

Viaggiare Serenamente

a cura di

Lorenzo Martucci, Claudia Raponi, Luca Tomaselli

ROMA TRE
UNIVERSITÀ DEGLI STUDI

Indice

Introduzione

1.1 Obiettivi

Il nostro secondo progetto, nell'ambito del corso di *Big Data*, è relativo allo sviluppo di uno strumento di valutazione dei viaggi effettuati nelle province del Lazio, della Campania e della Lombardia durante i mesi di febbraio, marzo, aprile e maggio 2015. I dati relativi ai viaggi sono stati incrociati con informazioni associate alle condizioni meteorologiche delle zone analizzate, per fornire una panoramica sulle scelte dei viaggiatori in base al clima del luogo di partenza e di destinazione dei loro spostamenti.

Il nostro lavoro è stato supportato dalla partecipazione al contest **Big Data Challenge 2015** promosso da *Telecom Italia*, grazie alla quale è stato possibile avere a disposizione una quantità molto vasta di dati da cui partire per lo sviluppo del progetto.

1.2 Attività svolte

La realizzazione dell'intero progetto ha previsto diverse fasi, che si sono evolute a partire dallo sviluppo dell'idea iniziale fino ad arrivare all'analisi dei risultati ottenuti.

Inizialmente ci siamo occupati di selezionare tra i numerosi dataset messi a disposizione dal contest, quelli che potessero essere di maggior interesse per l'elaborazione e questo ci ha portato alla scelta di utilizzare il dataset relativo ai viaggi effettuati dai veicoli monitorati da *Viasat*. Successivamente abbiamo cercato dati che fossero in grado di combinarsi con questi, in modo da arricchire l'elaborazione con ulteriori parametri e criteri di analisi. Per questo motivo abbiamo scelto di cercare dati inerenti alle condizioni meteorologiche, in modo da creare una corrispondenza tra le abitudini dei viaggiatori nell'intraprendere tratte più o meno brevi e il tempo atmosferico delle città di partenza e arrivo.

Una volta raccolti i dati di interesse ci siamo preoccupati di selezionare i campi più rilevanti per il lavoro di analisi che volevamo svolgere. Il processamento dei dati ha previsto quindi un'attività di filtraggio degli attributi più importanti e di esclusione di quelli meno interessanti, sia per quanto riguarda i dati dei viaggi che quelli del meteo. Successivamente le due tipologie di dati sono stati combinati, considerando come parametri comuni data e luogo analizzato. In questo modo siamo riusciti a definire un record completo di tutte le informazioni di interesse, che potesse essere elaborato per ricavare dati derivati.

La fase di elaborazione dei dati è stata supportata da *Apache Spark* che ci ha permesso di processare i Big Data fin qui collezionati, mappandoli e raccogliendoli secondo dei criteri di analisi da noi impostati e che verranno approfonditi nei capitoli successivi. Abbiamo scelto questo strumento sia per sperimentare una nuova tecnologia per la risoluzione di un problema di Big Data sia per testare la sua efficienza in termini di computazione rispetto ad altri sistemi. Infatti le stesse elaborazioni sono state effettuate anche sulla piattaforma *Hadoop*, che sfrutta il modello di programmazione *MapReduce*, al fine di confrontare le prestazioni dei due strumenti in base ai differenti approcci.

Infine, i risultati prodotti sono stati analizzati per identificare, per ciascun caso osservato, conclusioni consistenti che fossero in grado di spiegare ed individuare le preferenze e le scelte dei viaggiatori.

Elaborazione

2.1 Big Data

I Big Data necessari per la realizzazione dell'intero progetto sono stati recuperati da due diverse fonti:

- *Big Data Challenge 2015*: contest curato da *Telecom Italia*, dal quale sono stati prelevati i dati inerenti ai viaggi effettuati e messi a disposizione da *Viasat*;
- *ilMeteo.it*: principale sito d'informazione sul meteo, dal quale sono stati prelevati i dati inerenti alle condizioni meteorologiche;

Per quanto riguarda le informazioni sui tragitti effettuati, nel dataset ogni record rappresenta un viaggio e si presenta nel formato:
{Kind of vehicle, Starting sample, Starting time, Starting street type,Starting CAP, Starting City, Starting Province, Ending sample, Ending time, Ending street type, Ending CAP, Ending City, Ending Province, Kind of Trip, Distance, Average Speed, Samples, Samples in movement} [1]
E' stato quindi possibile considerare questi attributi rispetto a tre categorie: informazioni sulla partenza, sull'arrivo e dettagli aggiuntivi.

Un esempio di questa tipologia di dati è:
{3,1,20150404165425,U, 00135, Roma, RM, 3, 20150404175715, U, 00043, Ciampino, RM, 3, 40900, 40,60,11}.

Per quanto riguarda invece i dati associati al meteo sono state effettuate delle richieste Http al *Meteo.it* in modo da poter estrarre le condizioni climatiche dei luoghi osservati. Sulla base del giorno e della città considerata è stato possibile arricchire il record precedente con ulteriori parametri indicativi del meteo del luogo di partenza e di destinazione del viaggio. L'aggiunta dell'attributo associato alla condizione meteorologica è stata effettuata attraverso una gestione particolare dei dati, in grado di ottimizzare i tempi di esecuzione.

2.1.1 Estrazione dei dati meteorologici

I dataset relativo ai viaggi contiene circa 18 milioni record, ciascuno dei quali rappresenta una tratta percorsa da un determinato viaggiatore. Effettuare

[1] Nella parte conclusiva della relazione è presente un'approfondimento relativo a questi dati

una richiesta Http per ogni record considerato, al fine di conoscere il meteo da associare al viaggio avrebbe comportato uno spreco di risorse in termini prestazionali. Per questo motivo, l'estrazione delle condizioni meteorologiche dei luoghi osservati è stata gestita cercando di minimizzare tali richieste, considerando anche il fatto che molte delle città presenti nel dataset ricorrevano con una certa continuità.

Ogni richiesta Http utilizza come chiavi di ricerca il luogo e la data, ottenendo in risposta il meteo associato ai parametri di input. Ogni risposta è salvata in modo da creare un'archivio di informazioni in cui sono gestite le diverse città rintracciate con le condizioni meteorologiche e i fenomeni associati a ciascuna data. Ad esempio un piccolo estratto del meteo associato alla città di *Roma* è il seguente:

0304, nubi sparse, pioggia
0302, nubi sparse, pioggia
0301, poco nuvoloso, nessuno
0308, poco nuvoloso, nessuno
0314, poco nuvoloso, nessuno
0322, coperto, pioggia
0307, poco nuvoloso, nessuno
0313, poco nuvoloso, nessuno

Organizzando i dati secondo questa gestione ogni volta che è necessario associare ad un viaggio il meteo della città di partenza o destinazione, è necessario effettuare la richiesta Http se e solo se l'informazione desiderata non è già presente nell'archivio. Così facendo si riducono i tempi di esecuzione per il recupero dei dati.

2.2 Processamento con Spark

I record ottenuti dalla combinazione dei viaggi, con l'aggiunta delle informazioni meteorologiche, sono stati successivamenti processati grazie all'utilizzo di *Apache Spark*, un framework open-source di cloud computing. Le elaborazioni effettuate con Spark hanno seguito tutte lo stesso modello di programmazione, che ha permesso di raggrupare i dati sulla base di criteri stabiliti. Tale modello prevede l'attività di misurazione e conteggio dei record sulla base di criteri definiti e caratterizzanti rispetto a vari casi di uso.

Spark permette di gestire i dati attraverso degli oggetti **RDD** (*resilient distributed dataset*) ovvero come insiemi di elementi partizionati tra i nodi del cluster che sono in grado di operare in parallelo. Nel nostro caso, è stato creato un *JavaRDD<String>* che fosse in grado di memorizzare record caratterizzati dalle informazioni più rilevanti in termini di viaggi e meteo[2]. Questo è stato possibile grazie all'utilizzo della *FlatMapFunction*, funzione predefinita da Spark, adatta per l'organizzazione di dati strutturati.

[2]Negli esempi di codice che seguiranno tale elemeno è memorizzato nella variabile *record*

La gestione e l'aggregazione di queste informazioni è stata sottoposta a due attività: una di Mapping e una di Reduce.
Nella fase di *Mapping*, iterando sui record, è stato possibile definire delle coppie chiave-valore per associare ad ogni record il valore 1, attività necessaria per la successiva fase di raggruppamento dei dati. In questo caso, le strutture dati utilizzate sono state delle *Tuple2* predefenite da Spark ed utilizzate come mostrato di seguito:

```
JavaPairRDD<String, Integer> pairs =
     record.mapToPair(new PairFunction<String, String, Integer>() {

     public Tuple2<String, Integer> call(String s) {
         return new Tuple2<String, Integer>(s, 1);
     }
 });
```

Nella fase di *Reduce* invece l'attività principale è stata quella di aggregare i dati sulla base della chiavi comuni in modo da fornire una stima di quanti viaggi fossero caratterizzati dagli stessi parametri descrittivi. Questo invece è stato implementato dalle seguenti istruzioni:

```
JavaPairRDD<String, Integer> counts =
     pairs.reduceByKey(new Function2<Integer, Integer, Integer>() {

     public Integer call(Integer a, Integer b) {
         return a + b;
     }
 });
```

Le analisi che seguiranno sono state realizzate sulla base di questo modello implementativo e si differenziano le une dalle altre per la scelta dei parametri chiave considerati per l'aggregazione.

2.3 Analisi effettuate

Sono state implementate sei tipi di analisi che hanno permesso di considerare tre principali casi d'uso mentre gli ulteriori tre casi sono stati ottenuti apportando modifiche aggiuntive rispetto a quelli principali.
In particolare abbiamo condotto analisi relativamente a:

- **Caso A**: luoghi raggiunti in base alle condizioni meteorologiche della città di destinazione, analizzati rispetto ad una data specifica;
- **Caso B**: luoghi raggiunti in base alle condizioni meteorologiche della città di destinazione, analizzati rispetto a diverse fasce orarie considerate;
- **Caso C**: luoghi raggiunti in base alle condizioni meteorologiche della città di destinazione, analizzati rispetto a giorni festivi/feriali;
- **Caso D**: luoghi raggiunti in base alle condizioni meteorologiche delle città di partenza e di destinazione, analizzati rispetto ad una data specifica;

- **Caso E**: luoghi raggiunti in base alle condizioni meteorologiche delle città di partenza e di destinazione, analizzati rispetto a diverse fasce orarie considerate;
- **Caso F**: luoghi raggiunti in base alle condizioni meteorologiche delle città di partenza e di destinazione, analizzati rispetto a giorni festivi/feriali;

2.3.1 Caso A: Data specifica

Nel primo caso sono stati raccolti tutti quei viaggi effettuati in diverse località tenendo conto di una data specifica. A questi viaggi è stato associato il meteo della sola città di destinazione in modo da valutare quali potessero essere i luoghi raggiunti in presenza di determinate condizioni climatiche proprie della meta dello spostamento.
Un'estratto dell'output relativo a questo caso d'uso è il seguente:

(2015-04-15 Ceriano Laghetto sereno,38)
(2015-03-02 Pompei coperto,32)
(2015-03-30 Bonate Sopra poco nuvoloso,48)
(2015-03-06 Colturano sereno,4)
(2015-04-17 Boffalora Sopra Ticino coperto,42)

2.3.2 Caso B: Fasce orarie

Nel secondo caso sono stati aggregati i viaggi rispetto a diverse fasce orarie considerate, associando a questi la particolare condizione meteo relativa al luogo di arrivo. Tale analisi ha richiesto un processamento del parametro *Ending time*, dal quale è stato necessario separare i valori associati alla data rispetto a quelli relativi all'orario vero e proprio. In particolare sono state considerate quattro diverse fasce orarie:

Fascia oraria	Intervallo temporale
(0) Notte	00:00 - 06:00
(1) Mattina	06:00 - 12:00
(2) Pomeriggio	12:00 - 18:00
(3) Sera	18:00 - 24:00

In questo modo è stato possibile valutare le città più raggiunte in precisi momenti della giornata e quali spostamenti, all'interno di queste categorie temporali, fossero più frequenti associati a particolari condizioni meteorologiche della città di destinazione.
Un'estratto dell'output relativo a questo caso d'uso è il seguente:

(Cisterna di Latina poco nuvoloso Fascia Oraria: 0,806)
(Magnago poco nuvoloso Fascia Oraria: 0,47)
(Cerro al Lambro sereno Fascia Oraria: 3,55)
(Monterotondo nubi sparse Fascia Oraria: 1,924)
(Casaluce poco nuvoloso Fascia Oraria: 0,95)

2.3.3 Caso C: Giorni festivi/feriali

Nel terzo caso sono stati valutati i viaggi rispetto ai giorni festivi e feriali in cui tali spostamenti sono stati monitorati. Associando a tali tratte anche il meteo del luogo di destinazione, abbiamo considerato quali fossero le mete maggiormente raggiunte in presenza di particolati condizioni climatiche durante i giorni di vacanza rispetto a quelli lavorativi. Sono stati considerati festivi tutte le domeniche e tutti i giorni di festa presenti nel quattro mesi monitorati (febbraio, marzo, aprile, maggio). Sono quindi stati considerati festivi anche i giorni:

- 6 aprile: *Lunedì dell'Angelo*
- 25 aprile: *Festa della Liberazione*
- 1 maggio: *Festa dei Lavoratori*

Tale analisi ha permesso quindi di valutare le abitudini dei viaggiatori durante giorni di festività, riuscendo in questo modo ad identificare anche le mete più raggiunte proprio in concomitanza con questi giorni di assenza dal lavoro.
Un'estratto dell'output relativo a questo caso d'uso è il seguente:

(Marano di Napoli nubi sparse Festivo,20)
(Morlupo nubi sparse Festivo,387)
(Saronno nubi sparse Feriale,2179)
(Seregno nubi sparse Feriale,2628)
(San Giuliano Milanese nebbia Festivo,12)

2.3.4 Casi D, E, F: luogo e meteo di partenza

Gli ulteriori tre casi d'uso analizzati sono stati elaborati sulla base di quelli precedentemente definiti. Come ulteriori parametri valutativi, per ciascuna delle tre situazioni, sono stati aggiunti anche il luogo e il meteo della città di partenza. Questa scelta è stata fatta per permettere di derivare informazioni sulle abitudini dei viaggiatori in modo dipendente dalle condizioni meteo della zona in cui si trovano.
Un'estratto degli output relativi a questi casi d'uso è il seguente:

Caso D: Data specifica
(2015-03-29 Aversa sereno 2015-03-29 Sant'Arpino sereno,15)
(2015-04-24 Frattamaggiore nubi sparse 2015-04-24 Scafati poco nuvoloso,1)
(2015-03-04 Inzago 2015-03-04 Pozzo d'Adda ,1)
(2015-03-06 San Giuliano Milanese sereno 2015-03-06 Liscate ,1)
(2015-03-31 Castel Gandolfo poco nuvoloso 2015-03-31 Marino poco nuvoloso,2)

Caso E: Fasce orarie
(Roma sereno Fascia Oraria: 2 Fiumicino sereno Fascia Oraria: 3,9)
(Nocera Inferiore nubi sparse Fascia Oraria: 1 Nola coperto Fascia Oraria: 1,2)
(Rocca di Papa Fascia Oraria: 0 Gallicano nel Lazio nubi sparse Fascia Oraria: 0,1)
(Bergamo coperto Fascia Oraria: 1 Vaprio d'Adda Fascia Oraria: 1,1)
(Carate Brianza nubi sparse Fascia Oraria: 1 Milano nubi sparse Fascia Oraria: 1,41)

Caso F: Giorni festivi/feriali
(Succivo sereno FerialePartenza Caivano sereno FerialeArrivo,8)
(Melegnano poco nuvoloso FestivoPartenza Melegnano poco nuvoloso FestivoArrivo,67)
(Colturano poco nuvoloso FerialePartenza Rozzano poco nuvoloso FerialeArrivo,2)
(Dairago coperto FerialePartenza Parabiago nubi sparse FerialeArrivo,2)
(Triuggio nebbia FerialePartenza Albiate nebbia FerialeArrivo,14)

Osservazioni

3.1 Tempi di esecuzione

Le stime dei tempi di esecuzione del processamento dei dati sono state condotte secondo due diversi punti di vista. Inizialmente infatti sono state valutate le prestazioni computazionali rispetto alla sola elaborazione con Spark. In una fase successiva invece, le operazioni implementate in Spark sono state riadatte anche in *Hadoop* in modo da poter confrotare il lavoro in locale dei due diversi sistemi di elaborazione.

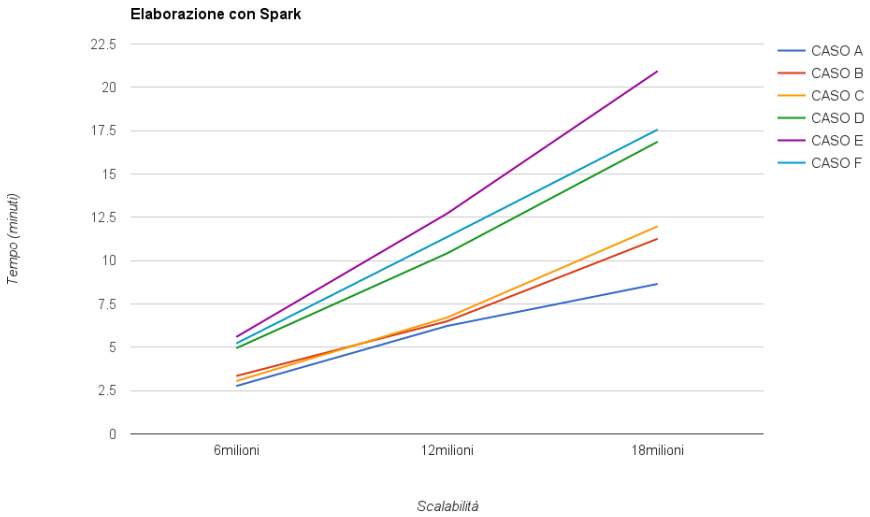
3.1.1 Spark

Le valutazioni rispetto all'elaborazione con Spark sono state effettuate considerando diversi input crescenti. Questo ha permesso di fornire una visione d'insieme sulla scalabilità del sistema sulla base di dati in ingresso di varie dimensioni. Il grado di scalabilità considerata, rispetto ai complessivi 18 milioni di dati in input forniti dal contest *Big Data Challenge 2015*, è il seguente:

- 6 milioni di dati;
- 12 milioni di dati;
- 18 milioni di dati;

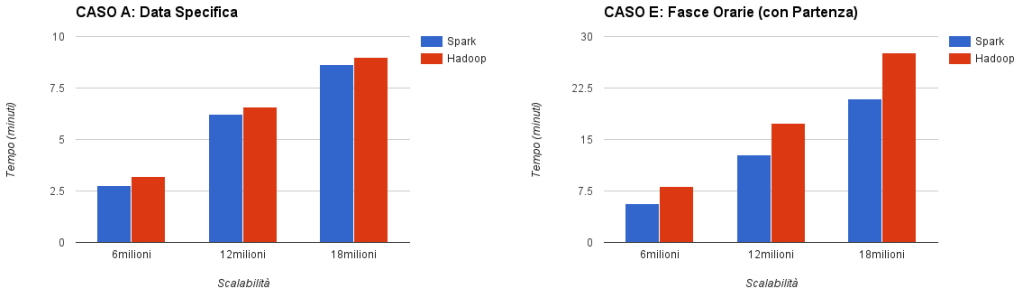
Questi tre livelli di input sono stati testati per ciascuno dei sei casi illustrati in precedenza. In questo modo sono state effettuate nel complesso 18 valutazioni.

Ciò che emerge da quest'analisi è che i primi tre casi, avendo meno dati da analizzare (solo luogo e meteo di destinazione), presentano tempi di esecuzione molto bassi che impiegano non più di 3 minuti per elaborare 6 milioni di dati. Inoltre il loro andamento crescente permette di eseguire elaborazioni su 18 milioni di dati con tempi che vanno dagli 8 ai 18 minuti.
Analizzando invece i casi più complessi (casi D, E ed F) i tempi di processamento risultano maggiori rispetto ai precedenti, mantenendo però sempre una crescita lineare che nel caso peggiore supera di poco i 20 minuti di esecuzione.
Le prestazioni sono riassunte nel grafico che segue:



3.1.2 Spark vs Hadoop

Le stesse operazioni effettuate con Spark sono state implementate anche con Hadoop per porre un confronto in termini prestazionali tra le due diverse esecuzioni in locale. In particolare sono stati approfonditi due situazioni: caso d'uso A e il caso d'uso E.



Analizzando il caso d'uso A (*Data Specifica*), che risulta meno oneroso dal punto di vista computazionale, ciò che emerge dal confronto è che le prestazioni dei due sistemi non mostrano rilevanti differenze. Osservando invece il caso d'uso E (*Fasce Orarie*), in cui vengono considerati anche i parametri di partenza, è possibile notare diversi rendimenti. Con dati più complessi quindi Spark offre dei risultati migliori rispetto ad Hadoop.

3.2 Risultati

In questa sezione conclusiva verrano illustrati i risultati derivati dall'osservazione dei dati processati. Infatti, per ognuno dei casi affrontati, è stato possibile dedurre informazioni in merito alle abitudini dei viaggiatori focalizzando l'attenzione sui diversi criteri imposti.

Data Specifica
I dati analizzati rispetto all'osservazione dei viaggi, in merito ad una data specifica (caso A), hanno evidenziato che i principali spostamenti monitorati sono stati effettuati verso le città più importanti per ciascuna delle regioni considerate. Filtrando infatti i risultati rispetto alle occorrenze maggiori di 10000, è emerso che le sole tre città di destinazione presenti sono proprio Roma, Napoli e Milano. I viaggi intrapresi verso questi tre luoghi sono stati effettuati in modo prevalente in presenza di meteo poco nuvoloso ma in generale la loro raggiungibilità è abbastanza indipendente dal tipo di condizione meteorologica. Inoltre è stata rilevata una particolare intensità di questi spostamenti nel mese di aprile. Risultati analoghi sono state dedotti dall'osservazione dei dati che tengono conto anche dei luoghi e del meteo di partenza (caso D). Infatti la maggior parte dei flussi monitorati risultano indipendenti dal tipo di meteo, presentano come luogo di partenza un capoluogo di regione e sono stati osservati nel mese di aprile.

Fasce orarie
I dati analizzati rispetto all'osservazione dei viaggi, in merito alle fasce orarie di percorrenza (caso B), hanno fatto emergere che il maggior numero dei viaggi sono stati effettuati nelle ore di piena attività di una giornata. Infatti le fasce 1 e 2 ovvero quelle relaive alla mattina e al pomeriggio sono quelle più rilevate. Gli spostamenti effettuati in questi momenti sono caratterizzati principalmente da tempo poco nuvoloso e sono diretti non solo verso città urbane, le quali sono raggiunte in qualsiasi orario, ma anche verso zone meno popolate come ad esempio Caserta, Nettuno e Tivoli. I viaggi monitorati con minore frequenza sono invece associati alle percorrenze verso piccoli comuni, in modo indipendente dal meteo ed effettuate principalmente nella fascia serale o notturna. Tenendo presente i luoghi di partenza dei viaggi (caso E), si riscontra una perfetta simmetria rispetto ai risultati precedenti, che evidenzia come le principali tratte vengano percorse nelle fasce giornaliere a partire da città non solo principali, in presenza di qualsiasi condizione metereologica.

Giorni festivi/feriali
I dati analizzati rispetto all'osservazione dei viaggi, in merito alla presenza di giorni feriali o festivi (caso C), hanno sottolineato come le occorrenze maggiori si rilevano in presenza di spostamenti effettuati durante i giorni lavorativi. Anche in questo caso i viaggi sono stati effettuati prevalentemente in presenza di meteo poco nuvoloso verso le città principali. Focalizzando l'attenzione sull'analisi degli spostamenti durante i giorni festivi è emerso che, sebbene le città più importanti risultino le più frequentate con ogni tipo di condizione meteorologica, in presenza di tempo sereno tra le città più raggiunte sono presenti quelle turistiche di Ercolano e Angri. Valutazioni simili possono essere effettuate considerando anche luogo e meteo della città di partenza (caso F).

Approfondimenti

Nella seguente sezione è illustrata nel dettaglio la descrizione dei dati sui viaggi, presenti nel dataset messo a disposizione dal *Big Data Challenge 2015* (contest curato da *Telecom Italia*).[1]

- **Kind of vehicle**:
  - *3*: light vehicle;
  - *4*: heavy vehicle;
- **Starting sample**: the kind of the last sample belonging to the trip (see the description tab for the classification of the samples):
  - *1*: ignition sample;
  - *2*: intermediate sample;
  - *3*: switch off sample;
- **Starting time**: timestamp of generation of the first sample belonging to the trip. The timestamp has the following format: yyyymmddHHMMss;
- **Starting street type**: type of the street where the last sample belonging to the trip have been generated:
  - *U*: urban way;
  - *E*: express way;
  - *A*: highway;
- **Starting CAP**: the Postal Code (CAP) where the first sample of the trip has been generated;
- **Starting City**: the city where the first sample of the trip has been generated (this information can be inferred from Starting CAP);
- **Starting Province**: short description of the province (e.g. BA for Bari) where the first sample of the trip has been generated (this information can be inferred from Starting CAP and Starting City);
- **Ending sample**: the kind of the last sample belonging to the trip (see the description tab for the classification of the samples):

[1] Per ulteriori riferimenti consultare: http://www.telecomitalia.com/tit/it/bigdatachallenge.html

  - *1*: ignition sample;
  - *2*: intermediate sample;
  - *3*: switch off sample;

- **Ending time**: timestamp of generation of the last sample belonging to the trip. The timestamp has the following format: yyyymmddHHMMss;
- **Ending street type**: type of the street where the last sample belonging to the trip have been generated:
  - *U*: urban way;
  - *E*: express way;
  - *A*: highway;
- **Ending CAP**: the Postal Code (CAP) where the last sample of the trip has been generated;
- **Ending City**: the city where the last sample of the trip has been generated (this information can be inferred from Ending CAP);
- **Ending Province**: short description of the province (e.g. BA for Bari) where the last sample of the trip has been generated (this information can be inferred from Ending CAP and Ending City);
- **Kind of Trip**: tells if the trip is a complete trip or an incomplete trip and which kind of incomplete trip it is. See the description tab for the definition of complete trip and incomplete trip.
  - *1*: incomplete trip starting and ending with the same ignition sample;
  - *2*: incomplete trip starting with an ignition sample and ending with an intermediate sample;
  - *3*: complete trip;
  - *4*: incomplete trip starting and ending with intermediate samples;
  - *6*: incomplete trip starting with an intermediate sample and ending with a switch off sample;
  - *9*: incomplete trip starting and ending with the sample switch off sample;
- **Distance**: total distance travelled in the trip in meters (distance travelled as reported in the last sample - distance travelled as reported in the first sample);
- **Average Speed**: average speed in kilometers per hour - km/h (only considering the sample with speed greater than zero);
- **Samples**: number of samples produced during the trip;
- **Samples in movement**: number of samples produced during the trip with speed equal to zero;