



PROGETTO TRENITALIA

Anastasia Marzi 847848
Alessandro Riboni 847160
Federico Signoretta 847343

INDICE

01

INTRODUZIONE

02

DATA MANAGEMENT

03

DATA VISUALIZATION

04

CONCLUSIONI



INTRODUZIONE

Considerata l'enorme quantità di persone che tutti i giorni utilizzano i mezzi pubblici per i propri spostamenti, i ritardi dei suddetti possono costituire un grosso problema.

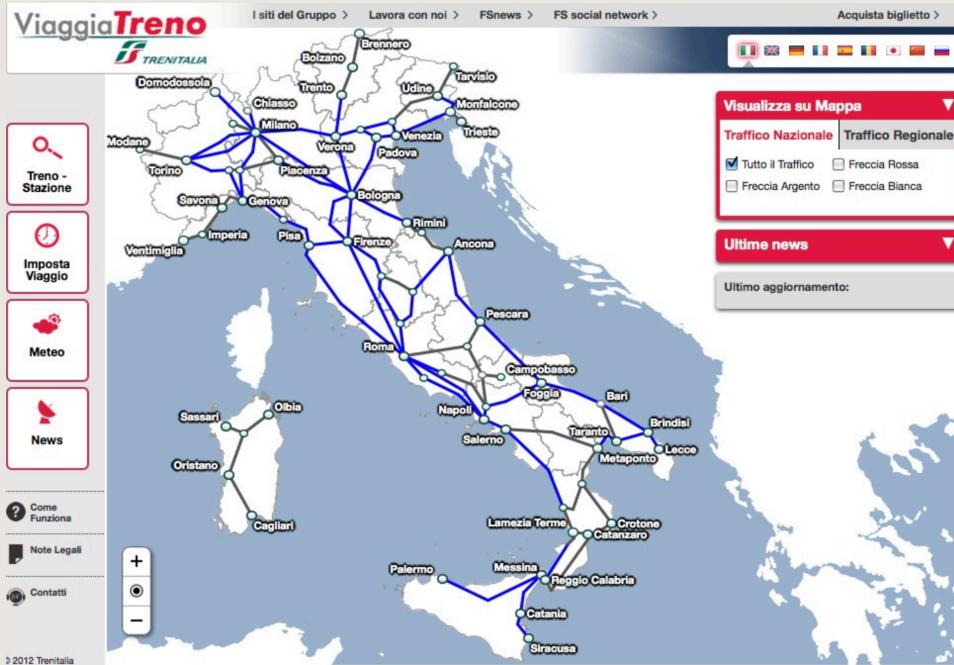
Nel nostro progetto, abbiamo deciso di focalizzare la nostra attenzione sui ritardi compiuti dai treni sulla rete ferroviaria italiana.



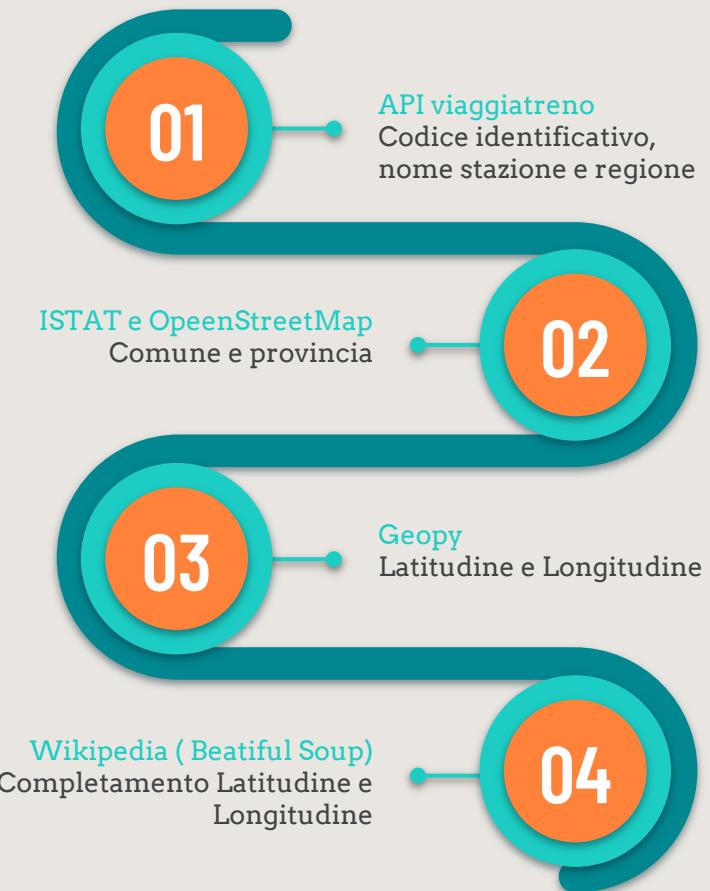
DATA MANAGEMENT



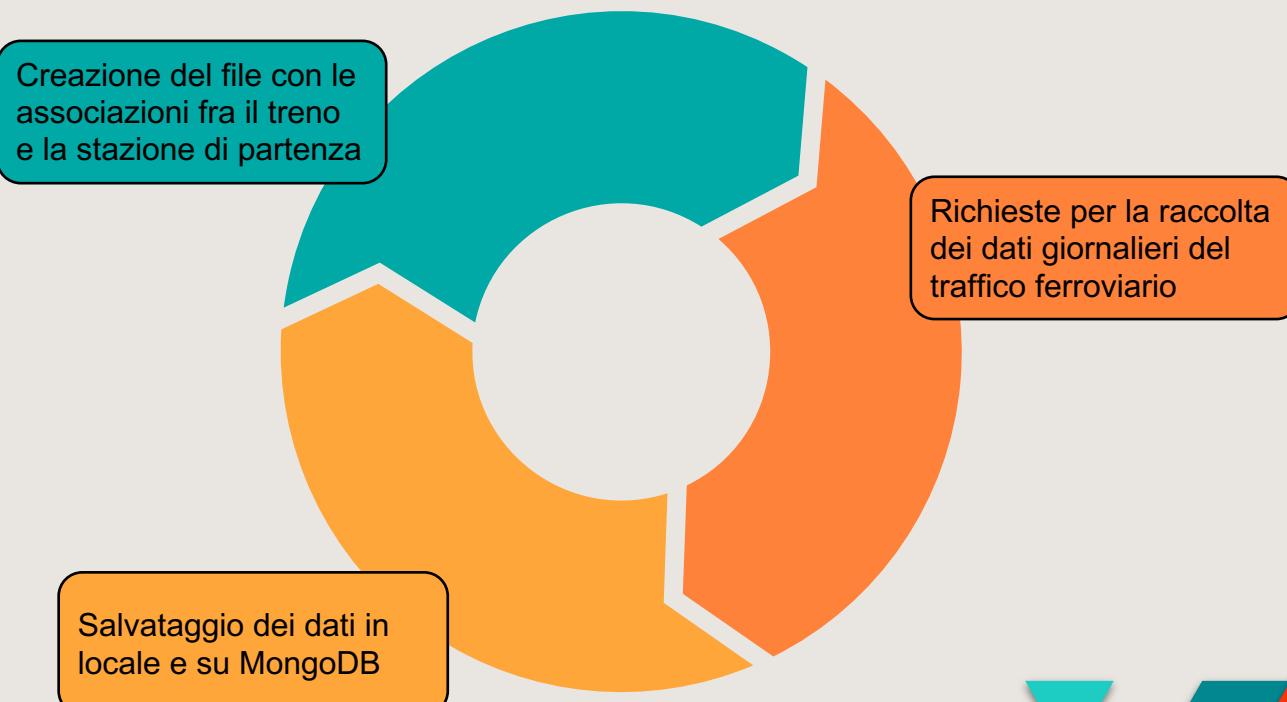
VIAGGIATRENO - TRENITALIA



STAZIONI.CSV



RACCOLTA DATI GIORNALIERA

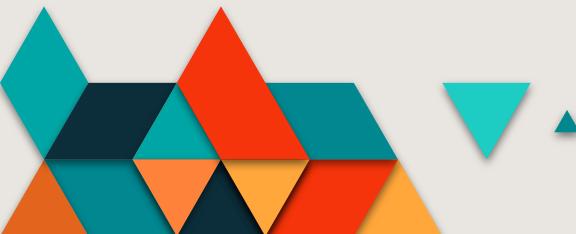


Creazione del file con le associazioni fra il treno e la stazione di partenza

Richieste per la raccolta dei dati giornalieri del traffico ferroviario

Salvataggio dei dati in locale e su MongoDB

DATA MANIPULATION



MONGODB - Regolare

```
_id: ObjectId("5caccbcc6422510291b3283b")
codice_treno: "REG1077"
categoria: "REG"
ritardo_finale: 0
data: 2019-01-01T00:00:00.000+00:00
origine: "UDINE"
destinazione: "VENEZIA SANTA LUCIA"
tratta: "UDINE-VENEZIA SANTA LUCIA_84"
> fermate: Array
  numero_fermate: 22
  durata: 135
  giorno_settimana: 3
  tipo_giorno: "feriale"
  stato: "regolare"
    > fermate: Array
      > 0: Object
        id_stazione: "S03026"
        orario: 2019-01-01T18:31:00.000+00:00
        ritardo: 5
        tipo_fermata: "P"
        nome_stazione: "UDINE"
        provincia: "UDINE"
        regione: "FRIULI VENEZIA GIULIA"
        lat: "46.055791"
        lon: "13.242003"
      > 1: Object
        id_stazione: "S02832"
        orario: 2019-01-01T18:39:00.000+00:00
        ritardo: 6
        tipo_fermata: "P"
        nome_stazione: "BASILIANO"
        provincia: "UDINE"
        regione: "FRIULI VENEZIA GIULIA"
        lat: "46.012737"
        lon: "13.109233"
      > 21: Object
        id_stazione: "S02593"
        orario: 2019-01-01T20:46:00.000+00:00
        ritardo: 0
        tipo_fermata: "A"
        nome_stazione: "VENEZIA SANTA LUCIA"
        provincia: "VENEZIA"
        regione: "VENEZO"
        lat: "45.441569"
        lon: "12.320882"
```

Treno regionale Udine – Venezia S. Lucia

Campi principali:

- codice_treno
- data
- tratta
- fermate [array]
- ritardo_finale



MONGODB – Parzialmente soppresso

Treno regionale Bologna C.LE- Milano C.LE

Modifiche della struttura:

- destinazione_teorica
- origine_teorica
- fermate [array]

```
_id: ObjectId("5caccbcc6422510291b338f1")
codice_treno: "REG2274"
categoria: "REG"
ritardo_finale: 65
data: 2019-01-01T00:00:00.000+00:00
stato: "parzialmente soppresso"
origine_teorica: "BOLOGNA C.LE"
origine: "BOLOGNA C.LE"
destinazione_teorica: "MILANO CENTRALE"
destinazione: "PIACENZA"
tratta: "BOLOGNA C.LE-MILANO CENTRALE_258"
fermate: Array
  numero_fermate: 12
  durata: 118
  giorno_settimana: 3
  tipo_giorno: "feriale"

  ▼ 0: Object
    id_stazione: "S05043"
    orario: 2019-01-01T05:50:00.000+00:00
    ritardo: 2
    tipo_fermata: "P"
    nome_stazione: "BOLOGNA C.LE"
    provincia: "BOLOGNA"
    regione: "EMILIA ROMAGNA"
    lat: "44.50626"
    lon: "11.342267"

  ▼ 1: Object
    id_stazione: "S05386"
    orario: 2019-01-01T06:02:00.000+00:00
    ritardo: 2
    tipo_fermata: "F"
    nome_stazione: "ANZOLA"
    provincia: "BOLOGNA"
    regione: "EMILIA ROMAGNA"
    lat: "44.546367"
    lon: "11.194608"

  ▼ 11: Object
    id_stazione: "S05000"
    orario: 2019-01-01T07:48:00.000+00:00
    ritardo: 65
    tipo_fermata: "A"
    nome_stazione: "PIACENZA"
    provincia: "PIACENZA"
    regione: "EMILIA ROMAGNA"
    lat: "45.052238"
    lon: "9.70633"
```

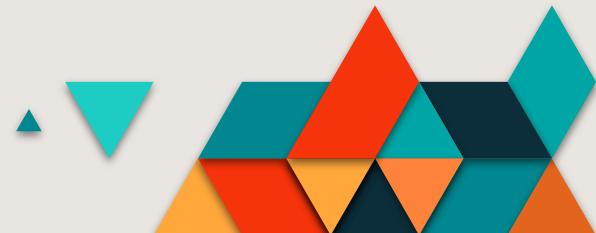
MONGODB - Cancellato

```
_id: ObjectId("5caccc8a6422510291b33e1e")
codice_treno: "REG20513"
categoria: "REG"
tratta: "CARPI-MODENA_82"
origine: "CARPI"
destinazione: "MODENA"
data: 2019-01-02T00:00:00.000+00:00
stato: "cancellato"
giorno_settimana: 4
```

Treno regionale Carpi - Modena

Modifiche della struttura:

- durata
- numero_fermate
- fermate [array]
- ritardo_finale



ESEMPIO DI QUERY

```
pipeline = [
    {"$group": {"_id": {"categoria": "$categoria", "stato": "$stato"},
                "totale": {"$sum": 1}}},
    {"$sort": SON([{"_id.categoria": -1}])}
]
lista = (list(db.trenitalia.aggregate(pipeline)))
```

	categoria	stato	totale		categoria	stato	totale
0	REG	regolare	211759	9	ES*	cancellato	4
1	REG	cancellato	2167	10	ES*	parzialmente soppresso	22
2	REG	parzialmente soppresso	1807	11	ES*	regolare	8557
3	MET	parzialmente soppresso	143	12	EN	cancellato	14
4	MET	cancellato	135	13	EN	regolare	185
5	MET	regolare	6344	14	EN	parzialmente soppresso	11
6	IC	cancellato	13	15	EC	cancellato	4
7	IC	regolare	3464	16	EC	regolare	1451
8	IC	parzialmente soppresso	30	17	EC	parzialmente soppresso	5



DATA VISUALIZATION



STRUMENTI UTILIZZATI



JavaScript Charts Maps
amCharts



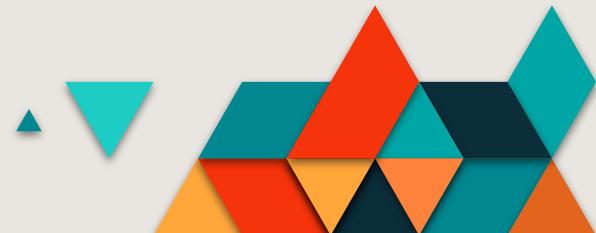
Python
Seaborn & Matplotlib



Tableau Desktop



R
ggplot2



LE INFOGRAFICHE

Infografica 1

Traffico ferroviario del mese
di Gennaio 2019



02

Infografica 3

Ritardi medi per regione e
per giorno



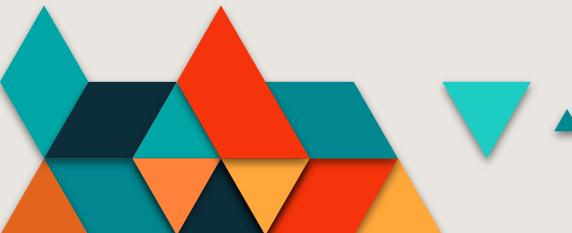
04

Infografica 2

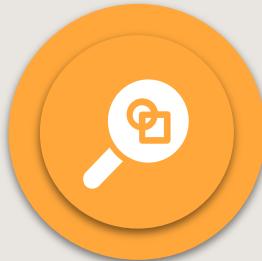
Ritardi mediani per giorno e per
provincia

Infografica 4

Analisi dei ritardi del traffico
dell'Alta Velocità Italiana



QUESTIONARI E TEST



Valutazione
Euristica



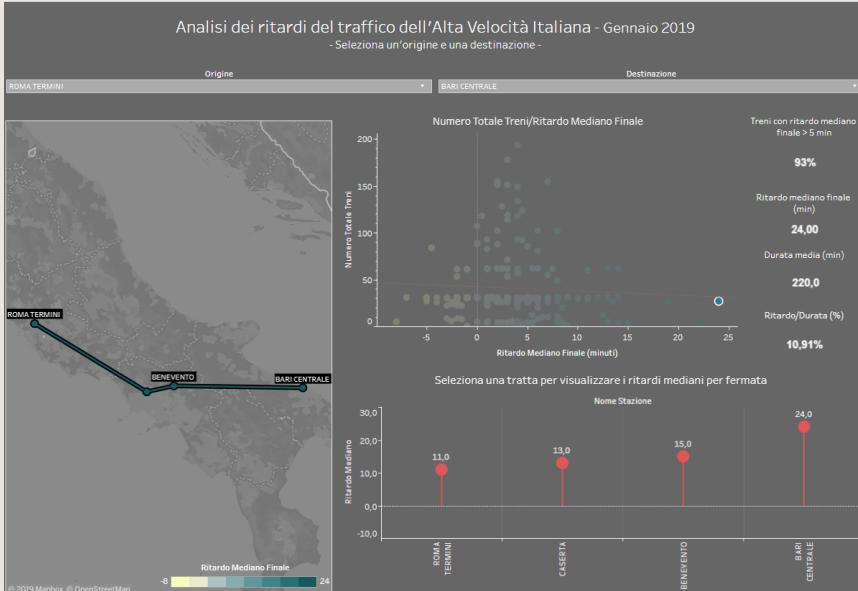
Questionario
Psicométrico



Test dell'Utente



UN ESEMPIO: INFOGRAFICA 4



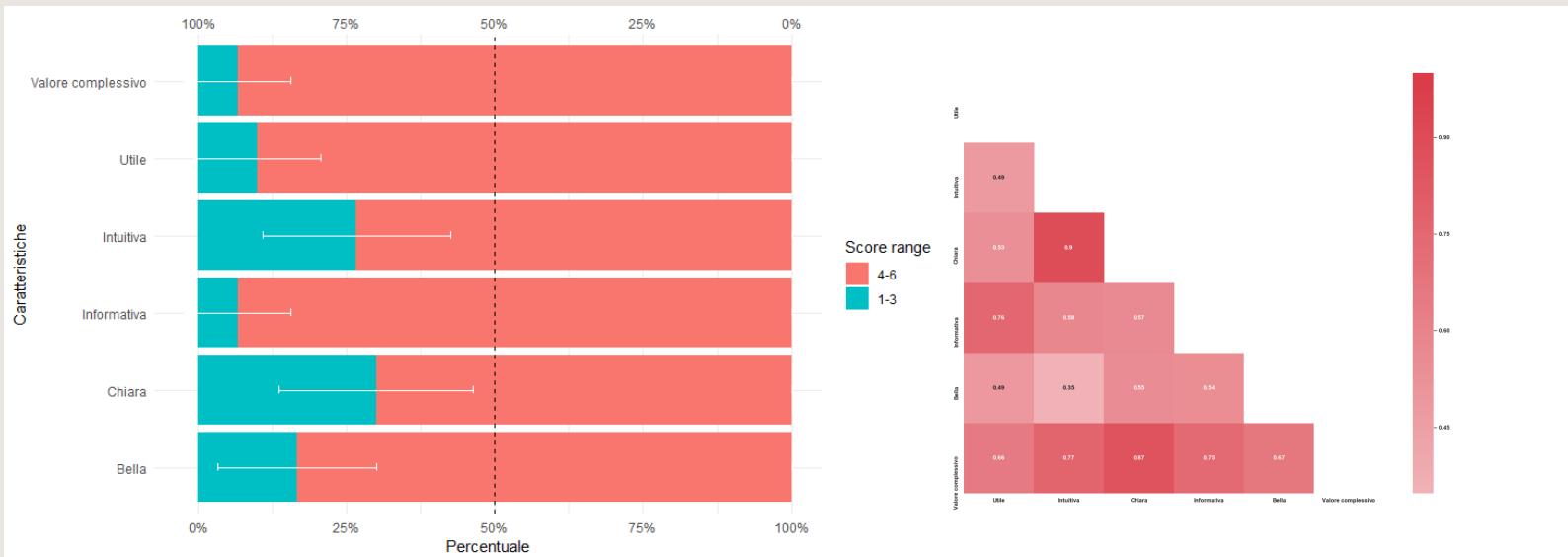
Valutazione euristica

Problematiche riscontrate:

- L'utente spesso tende ad utilizzare direttamente la cartina per selezionare le tratte, invece di utilizzare gli appositi spazi;
- Il bubble chart non viene quasi mai utilizzato perché non se ne capisce l'utilità;
- La legenda del bubble chart non viene quasi mai notata;
- Non è chiaro che la tratta è più facilmente selezionabile sul bubble chart, anziché sulla cartina.

UN ESEMPIO: INFOGRAFICA 4

Questionario Psicometrico



UN ESEMPIO: INFOGRAFICA 4

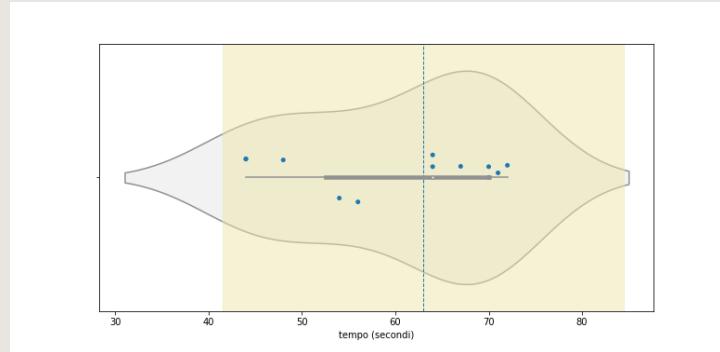
Test dell'Utente

Task:

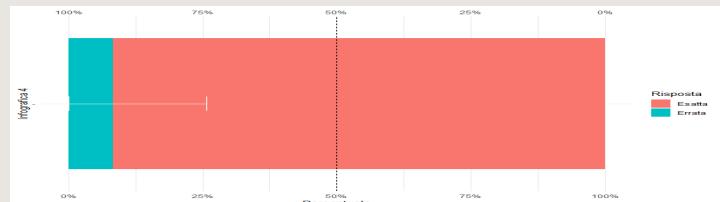
Considera il tragitto che ha come origine «Milano Centrale» e come destinazione «Napoli Centrale». Dopo aver selezionato il percorso che compie tale tragitto col ritardo mediano maggiore, qual è la stazione, all'interno di tale percorso, in cui si verifica il ritardo mediano maggiore?



Tempo di completamento:



Error rate:





CONCLUSIONI

- Ritardi mediani minori delle aspettative (Sole 24 Ore)
 - Scalare il lavoro compiuto su un arco temporale maggiore
 - Migliorare il servizio (individuazione punti strategici)
- 

**GRAZIE PER
L'ATTENZIONE**



“Gennaio_2019”: 20190101 - 20190131
“Febbraio_2019”: 20190201 - 20190128
“Marzo_2019”: 20190301 - 20190331

{ X : 20190103 } { X : 20190126 } { X : 20190212 } { X : 20190228 } { X : 20190303 }



Shard 1

Shard 2

Shard 3

Zone: [“Gennaio_2019”]

Zone: [“Febbraio_2019”]

Zone: [“Marzo_2019”]