

---

# ICTDays Summercamp

*Sito per l'ICT Days Summercamp*

**Team ICT Days Summercamp**

**Apr 03, 2022**

Copyright © 2022 by Team ICT Days Summercamp.

ICTDays Summercamp is available under the Creative Commons Attribution 4.0 International License, granting you the right to copy, redistribute, modify, and sell it, so long as you attribute the original to Team ICT Days Summercamp and identify any changes that you have made. Full terms of the license are available at:

<http://creativecommons.org/licenses/by/4.0/>

The complete book can be found online for free at:

<https://jupman.softpython.org/en/latest/>



## CONTENTS

	About . . . . .	1
	Vai alle Challenges 2019 . . . . .	1
	Vai alle Challenges 2018 . . . . .	2
<b>1</b>	<b>2018</b>	<b>3</b>
	Business Oriented Challenge . . . . .	3
	Mondiali Russia 2018 Challenge . . . . .	6
<b>2</b>	<b>2019</b>	<b>9</b>
	RiParco da Trento Challenge . . . . .	9
	Lavoro 4.0 Challenge . . . . .	13
	A Prova di Hacker Challenge . . . . .	16
	Real Time Transport Challenge . . . . .	19



## About

Questo sito raccoglie le sfide di data science presentate agli ICTDays Summercamp per ragazzi di Istituti superiori del Trentino in Alternanza Scuola-Lavoro presso l'Università degli Studi di Trento al Dipartimento di Ingegneria e Scienze dell'Informazione (DISI)<sup>1</sup>, in collaborazione con Hub Innovazione Trentino HIT<sup>2</sup> e la rete STAAR<sup>3</sup> nelle sedi del Clab Trento<sup>4</sup>.

## Vai alle Challenges 2019



UNIVERSITÀ DEGLI STUDI  
DI TRENTO  
Dipartimento di Ingegneria  
e Scienza dell'Informazione



**S.T.A.A.R.R.**



<sup>1</sup> <https://www.disi.unitn.it>

<sup>2</sup> <https://www.trentinoinnovation.eu>

<sup>3</sup> <http://www.staarr.it/>

<sup>4</sup> <https://international.unitn.it/blt/clab-trento>

## Vai alle Challenges 2018



## Business Oriented Challenge



**Sponsor:** SpazioDati<sup>5</sup>.

Gli enti pubblici come i comuni hanno costantemente bisogno che siano svolti diverse attività, come per esempio pavimentazione stradale, gestione impianti sportivi, manutenzione aree verdi, etc. A tal fine aprono **bandi di gara** per chiedere alle imprese di offrire questi servizi al prezzo più basso. In Italia ci sono circa 8000 comuni, e per un'impresa **monitorare tutti i bandi** può richiedere un notevole sforzo. SpazioDati offre un servizio di ricerca di imprese chiamato Atoka, che permette di trovare aziende per partita IVA, settore e contenuti dei loro siti web. Per esempio, permette di capire rapidamente i prodotti e servizi venduti dalle aziende di un certo territorio. Tra i vari filtri, è possibile ricercare aziende che siano risultate vincitrici di bandi pubblici. SpazioDati<sup>6</sup> ci chiede di **aggiungere filtri** riguardanti i bandi di gara, implementando funzionalità come **estrazione di parole chiave**, classificazione dei bandi, estrazione di date, importi e luoghi.

In Trentino sono disponibili su [dati.trentino](http://dati.trentino.it)<sup>7</sup> bandi di gara per vari comuni forniti da ComunWeb, negli esempi ci concentriamo su Trento. Per avere un'idea di cosa è disponibile, si può guardare l'interfaccia di ricerca [sul sito del Comune di Trento](http://www.comune.trento.it)<sup>8</sup>, che permette di cercare per

- titolo
- servizi o ufficio di competenza
- argomento
- tipologia (lavori pubblici, servizi, forniture)
- fase (aperto, in esame, aggiudicato)
- data pubblicazione

Per i fini di questa challenge, ci limiteremo ad usare il dataset dei bandi di gara da [dati.trentino.it](http://dati.trentino.it) con un dataset di aziende trentine ricavato dal database di Atoka di SpazioDati. Il vostro compito sarà **integrare i dataset** ed estenderli

<sup>5</sup> <http://spaziodati.eu>

<sup>6</sup> <https://spaziodati.eu>

<sup>7</sup> <http://dati.trentino.it/dataset/bandi-di-gara-del-comune-di-trento>

<sup>8</sup> <http://www.comune.trento.it/Amministrazione-Trasparente/Bandi-di-gara-e-contratti/Atti-delle-amministrazioni-aggiudicatrici-e-degli-enti-aggiudicatori-distintamente>  
Atti-relativi-alle-procedure-per-l'affidamento-di-appalti-pubblici-di-servizi-forniture-lavori-e-opere-di-concorsi-pubblici-di-progettazione-di-concorsi-di-idee-e-di-co  
-Compresi-quelli-tra-enti-nell-mabito-del-settore-pubblico-di-cui-all-art/Bandi-di-gara

ulteriormente con altre colonne come keywords e importo, implementando infine un **prototipo di motore di ricerca** che permetta di filtrare in base a tali colonne.

## .1 a. Analisi

Quante e quali aziende potrebbero essere interessate al nostro servizio ?

## .2 b. Ricerca base

Un primo approccio semplice potrebbe essere replicare la ricerca già presente sul sito del comune. Per iniziare, basterebbe creare delle funzioni python che permettono di specificare i vari argomenti di ricerca. Fatto ciò, se rimane tempo e se ne hanno le competenze, si potrebbe pensare di costruire una interfaccia HTML rudimentale.

## .3 c. Ricerca avanzata

Si potrebbe migliorare le funzionalità di ricerca permettendo di filtrare:

- keyword
- importo complessivo
- sottocategorie (pavimentazioni, reti idrauliche ...)
- area geografica, comune

Per realizzare quanto sopra, si potrebbe effettuare analisi semantica del testo usando il servizio Dandelion di SpazioDati con l'API Entity Extraction (vedi [esempio visuale](#)<sup>9</sup> e [documentazione API](#)<sup>10</sup>)

Esempi di ricerca:

- impresa edile vuole sapere quando esce un bando per lavori pubblici da almeno 2 milioni di euro in un certo territorio, filtrando se possibile tra lavori di pavimentazioni, reti idrauliche, impianto di illuminazione pubblica, etc. Esempio [Lavori Pubblici - Area ex Michelin](#)<sup>11</sup> CIG n. 70813914B7

Dalla descrizione è possibile estrarre diverse parole chiave che identificano i lavori effettivamente richiesti (pavimentazioni, reti idrauliche, impianto di illuminazione pubblica etc). [Esempio estrazione su Dandelion](#)<sup>12</sup> (leggermente editato per stare nei limiti del sito). L'estrazione dell'importo complessivo può essere invece fatta con l'uso di [regex](#)<sup>13</sup> Altre imprese potrebbero essere imprese di pulizie, servizi alla persona, etc...

<sup>9</sup> [https://dandelion.eu/semantic-text/entity-extraction-demo/?text=%22Sottopasso+stradale+e+strada+di+collegamento+sull%27area+ex+Aziende+Agrarie%E2%80%9D.Le+opere+che+formano+oggetto+dell%27appalto+possono+riassumersi+in+via+puramente+indicativa+come+di+seguito%3A-+demolizioni%2C+rimozioni+e+scavi%3B-+formazioni+di+rilevati+e+massicciate%3B-+calcestruzzo+e+acciai+per+c.a.%3B-+micropali+e+fondazioni+speciali%3B-+opere+di+abbassamento+falda%3B-+infissione+e+traslazione+del+monolite+ferroviario%3B-+sistema+provvisorio+di+sostegno+dei+binari%3B-+pavimentazioni%3B-+reti+idrauliche%3B-+impianto+di+illuminazione+pubblica.Importo+complessivo+di+appalto%3A+euro+2.137.992%2C87+di+cui+euro+73.115%2C84+per+oneri+di+sicurezza+non+soggetti+a+ribasso.&lang=it&min\\_confidence=0&country=IT&exec=true#results](https://dandelion.eu/semantic-text/entity-extraction-demo/?text=%22Sottopasso+stradale+e+strada+di+collegamento+sull%27area+ex+Aziende+Agrarie%E2%80%9D.Le+opere+che+formano+oggetto+dell%27appalto+possono+riassumersi+in+via+puramente+indicativa+come+di+seguito%3A-+demolizioni%2C+rimozioni+e+scavi%3B-+formazioni+di+rilevati+e+massicciate%3B-+calcestruzzo+e+acciai+per+c.a.%3B-+micropali+e+fondazioni+speciali%3B-+opere+di+abbassamento+falda%3B-+infissione+e+traslazione+del+monolite+ferroviario%3B-+sistema+provvisorio+di+sostegno+dei+binari%3B-+pavimentazioni%3B-+reti+idrauliche%3B-+impianto+di+illuminazione+pubblica.Importo+complessivo+di+appalto%3A+euro+2.137.992%2C87+di+cui+euro+73.115%2C84+per+oneri+di+sicurezza+non+soggetti+a+ribasso.&lang=it&min_confidence=0&country=IT&exec=true#results)

<sup>10</sup> <https://dandelion.eu/docs/api/datatxt/nex/v1/>

<sup>11</sup> <http://www.comune.trento.it/Amministrazione-Trasparente/Bandi-di-gara-e-contratti/Atti-delle-amministrazioni-aggiudicatrici-e-degli-enti-aggiudicatori-distintamente-Atti-relativi-alle-procedure-per-l'affidamento-di-appalti-pubblici-di-servizi-forniture-lavori-e-opere-di-concorsi-pubblici-di-progettazione-di-concorsi-di-idee-e-di-co>

<sup>12</sup> [https://dandelion.eu/semantic-text/entity-extraction-demo/?text=%22Sottopasso+stradale+e+strada+di+collegamento+sull%27area+ex+Aziende+Agrarie%E2%80%9D.Le+opere+che+formano+oggetto+dell%27appalto+possono+riassumersi+in+via+puramente+indicativa+come+di+seguito%3A-+demolizioni%2C+rimozioni+e+scavi%3B-+formazioni+di+rilevati+e+massicciate%3B-+calcestruzzo+e+acciai+per+c.a.%3B-+micropali+e+fondazioni+speciali%3B-+opere+di+abbassamento+falda%3B-+infissione+e+traslazione+del+monolite+ferroviario%3B-+sistema+provvisorio+di+sostegno+dei+binari%3B-+pavimentazioni%3B-+reti+idrauliche%3B-+impianto+di+illuminazione+pubblica.Importo+complessivo+di+appalto%3A+euro+2.137.992%2C87+di+cui+euro+73.115%2C84+per+oneri+di+sicurezza+non+soggetti+a+ribasso.&lang=it&min\\_confidence=0&country=IT&exec=true#results](https://dandelion.eu/semantic-text/entity-extraction-demo/?text=%22Sottopasso+stradale+e+strada+di+collegamento+sull%27area+ex+Aziende+Agrarie%E2%80%9D.Le+opere+che+formano+oggetto+dell%27appalto+possono+riassumersi+in+via+puramente+indicativa+come+di+seguito%3A-+demolizioni%2C+rimozioni+e+scavi%3B-+formazioni+di+rilevati+e+massicciate%3B-+calcestruzzo+e+acciai+per+c.a.%3B-+micropali+e+fondazioni+speciali%3B-+opere+di+abbassamento+falda%3B-+infissione+e+traslazione+del+monolite+ferroviario%3B-+sistema+provvisorio+di+sostegno+dei+binari%3B-+pavimentazioni%3B-+reti+idrauliche%3B-+impianto+di+illuminazione+pubblica.Importo+complessivo+di+appalto%3A+euro+2.137.992%2C87+di+cui+euro+73.115%2C84+per+oneri+di+sicurezza+non+soggetti+a+ribasso.&lang=it&min_confidence=0&country=IT&exec=true#results)

<sup>13</sup> <http://it.softpython.org/search/regex-sol.html>



#### .4 d. Bandi simili

Dato un bando, un'impresa potrebbe voler cercare bandi simili, per esempio per capire quali sono stati i criteri di selezione e per individuare quanti e quali potenziali aziende concorrenti hanno partecipato al bando.

#### .5 Dati bandi di gara

Dataset su dati.trentino: <http://dati.trentino.it/dataset/bandi-di-gara-del-comune-di-trento>

Esempi API per comune di Trento (per documentazione parametri query vedere documentazione ComunWeb):

JSON [www.comune.trento.it/api/opendata/v2/content/search/classes+bando+offset+30](http://www.comune.trento.it/api/opendata/v2/content/search/classes+bando+offset+30)<sup>14</sup>

CSV: API: [www.comune.trento.it/exportas/custom/csv\\_search?classes=bando](http://www.comune.trento.it/exportas/custom/csv_search?classes=bando)<sup>15</sup>

#### .6 Dati Atoka

Questo dataset è fornito da SpazioDati<sup>16</sup> tramite il servizio Atoka<sup>17</sup>. Riportiamo qui un esempio dei dati (in verticale).

Per i dati completi chiedere a [david.leoni@unitn.it](mailto:david.leoni@unitn.it)

id	f4b138b59562	e3e9ded1bc72
vat_id	00437240229	01669210229
legal_name	LABORATORIO SOCIALE - SOCIETA' COOPERATIVA SOCIALE	COOPERATIVA SOCIALE ASSISTENZA
ateco_code	88	88.1
ateco_label	ASSISTENZA SOCIALE NON RESIDENZIALE	ASSISTENZA SOCIALE NON RESIDENZIALE PER ANZIANI E DISABILI
headquarters_municipality	Trento	Tione di Trento
headquarters_address	Via Giambattista Unterveger, 6, 38122, Trento (TN)	Via Damiano Chiesa, 2/A, 38079, Tione di Trento (TN)
headquarters_lat	46,09469	46,03307
headquarters_lon	11,10958	10,72513
number_employees	81	81
date_activity_start	30-03-1977	04-10-1999
website	<a href="http://www.laboratoriosociale.it">http://www.laboratoriosociale.it</a>	<a href="http://www.coopassistenza.org">http://www.coopassistenza.org</a>

<sup>14</sup> <http://www.comune.trento.it/api/opendata/v2/content/search/classes+bando+offset+30>

<sup>15</sup> [http://www.comune.trento.it/exportas/custom/csv\\_search?classes=bando](http://www.comune.trento.it/exportas/custom/csv_search?classes=bando)

<sup>16</sup> <http://spaziodati.eu>

<sup>17</sup> <http://atoka.io>

[ ]:

## Mondiali Russia 2018 Challenge



**Sponsor:** U-Hopper<sup>18</sup>

Daniele, CEO di U-Hopper<sup>19</sup>, ha deciso di diventare ricco per poter andare a fare il baby-pensionato ad Antigua. Dato che però diventare ricco facendo l'imprenditore si è rivelata una strada lunga e tortuosa (più del previsto, almeno), ha avuto un'idea geniale (a suo parere, ndr). Con i soldi dell'azienda assumerà un data scientist, e lo metterà a lavorare in segreto su un progetto speciale. Da appassionato di **calcio**, Daniele conosce bene il mondo delle **scommesse online**. E da ex ricercatore ne sa abbastanza di big data e intelligenza artificiale per sapere che un algoritmo ben congegnato da un brillante data scientist può facilmente battere la logica (e gli algoritmi) delle varie piattaforme di scommesse online. E il caso di Soccermatics<sup>20</sup>, con un ritorno del 1.800% in circa un anno gli conferma che è sulla buona strada. L'occasione per fare il botto è dietro l'angolo, ed è rappresentata dai **mondiali di Russia 2018**. Con una stima di 5.45 miliardi di euro **investiti in scommesse**<sup>21</sup> durante i mondiali, è chiaro che una **intelligenza artificiale** ben allenata può fare ricco il suo creatore. E l'idea di Daniele è di sfruttare, oltre ai dati statistici consueti (ranking FIFA, storico delle partecipazioni ai Mondiali etc.) anche dati dalle news e dai social media per intercettare quei piccoli **segnali** che potrebbero **predire il risultato**. Con un budget iniziale di 1.000 euro, quanto riuscirà a realizzare Daniele durante i mondiali, usando il prodigioso algoritmo progettato dal suo data scientist?

### .1 a. Analisi

Se fossimo in grado di prevedere le partite con buona accuratezza e li pubblicassimo, potremmo pensare di attrarre una buona quantità di scommettitori sul nostro sito. Quanti potrebbero essere?

- descrivere / capire la strategia di investimento
- descrivere il modello (e differenza tra parametrico / non parametrico)

---

<sup>18</sup> <http://u-hopper.com/>

<sup>19</sup> <http://u-hopper.com/>

<sup>20</sup> <https://medium.com/@Soccermatics/if-you-had-followed-the-betting-advice-in-soccermatics-you-would-now-be-very-rich-1f643a4f5a23>

<sup>21</sup> <https://www.sportytrader.com/en/news/world-cup-infographic/>

## .2 b. Pulizia di dati

- pulizia dei dati (*data cleaning*) : convertire il risultato di ogni partita i.e. 5-3 in vinta (V), persa (P), pareggiata (X) in modo da semplificare il modello

## .3 c. Integrazione dati

Vi sono disponibili diverse statistiche sulle squadre. Per essere effettivamente riutilizzabili andrebbero unite, per esempio si potrebbe unire il ranking degli ultimi anni ai risultati delle squadre

## .4 d. Sviluppo del modello di predizione

Le attività precedenti sarebbero da supporto all'effettivo sviluppo di un modello che permetta di predire i risultati. In termini più tecnici, bisognerebbe effettuare il *training* del il modello e variare i pesi di ogni caratteristica (*feature*) in modo da ottimizzare i risultati.

## .5 e. Sviluppo chatbot

Se avanza tempo, si potrebbe provare a creare una specie di chatbot in Jupyter, in cui l'utente può domandare al bot l'esito delle partite future. (nota: non sarà necessario creare un servizio web).

## .6 f. Lavorare alla presentazione

Forse più che nelle altre challenge sarà necessario lavorare ad una presentazione convincente, che sia un po' più elaborata del "Col mio sistema diventi ricco sfondato". Dovrebbero esserci considerazioni sulla strategia di investimento. E' conservativa? Aggressiva? A quale pubblico si rivolge? Il pensionato? Il giovane con la propensione al rischio ?

## .7 Dati

Come dati useremo i FIFA Soccer Rankings dal 1993 to 2018

### FIFA Soccer Rankings

[fifa-ranking.csv.zip](#)

### Risultati match internazionali

Risultati match internazionali dal 1872 to 2018

Scarica il file: [results.csv.zip](#)

## FIFA World Cup 2018 data set

Scarica il file: [world-cup-2018.csv](#)

## Dati news

I dati news sono forniti da SpazioDati. Per il file completo chiedere a [david.leoni@unitn.it](mailto:david.leoni@unitn.it)

[ ]:

## RiParco da Trento Challenge



UNIVERSITÀ DEGLI STUDI  
DI TRENTO  
Dipartimento di Ingegneria  
e Scienza dell'Informazione



**Sponsor:** Dipartimento di Ingegneria e Scienze dell'Informazione (DISI)<sup>22</sup> e [dati.trentino.it](http://dati.trentino.it)<sup>23</sup>

I parchi sono un importante luogo di **aggregazione sociale**, e consentono a famiglie e giovani di riunirsi in spazi all'aperto. Le occasioni di ritrovo possono essere molteplici: a carattere **sportivo** come partite di calcio, ping pong, criquet, skateboard, **attività ludiche** nei parchi giochi per bambini, ritrovi informali di giocolieri di strada, suonatori di bonghi, o **eventi organizzati** come sagre rionali e concerti.

Non sempre tutte queste occasioni sono note, perché o informali o non sufficientemente pubblicizzate. A volte, gli incontri neppure nascono perché si crede non vi sia un sufficiente numero di persone interessato a svolgere una certa attività.

Si richiede pertanto di **creare un sito** che mostri una **mappa dei parchi** di Trento arricchita con le attività in essere. La visualizzazione sarà basata su [OpenStreetMap](https://www.openstreetmap.org/)<sup>24</sup>, la mappa del mondo realizzata da volontari. Qualora le attività siano informali, si dovrà provvedere a scattare **foto** e caricarle manualmente sulla mappa. Dato che nelle foto di attività pubbliche è sempre probabile la presenza di volti, per questioni di privacy si dovrà provvedere a sviluppare un **software per anonimizzare** le facce. Per attività pubbliche (es. sagre, concerti), si dovranno cercare opportuni dataset che le descrivano (da fonti come [dati.trentino.it](http://dati.trentino.it)<sup>25</sup>) e sviluppare programmi che **integrino i dati** nella mappa. Il sito dovrà inoltre prevedere un'**interfaccia grafica** per aggiungere eventi, caricare le foto ed organizzare nuovi incontri. A tal fine, potrebbe essere interessante integrare dati meteo da [dati.trentino.it](http://dati.trentino.it).

---

<sup>22</sup> <https://www.disi.unitn.it>

<sup>23</sup> <http://dati.trentino.it>

<sup>24</sup> <https://www.openstreetmap.org/#map=14/46.0681/11.1181>

<sup>25</sup> <http://dati.trentino.it>

## .1 a. Acquisire foto

**ATTENZIONE: LEGGETE QUANTO SEGUE IN OGNI SUA PARTE, COMPORTAMENTI INADEGUATI E MANCANZE NON SARANNO TOLLERATI**

Nel **vostro tempo libero**, o in pausa pranzo, vi si chiede di visitare parchi di Trento fotografando le attività presenti, come giochi per bambini, giochi cricket, ping-pong, skateboard, giocoleria, musicali, etc. Nel caso nella scena vi siano persone, per non destare preoccupazioni si raccomanda di scattare le foto da lontano.

Le foto devono essere scattate da voi, **non** scaricate foto a caso da Internet.

Ai soli fini di testare il software di anonimizzazione che produrrete, vi si richiede di fare foto dei **vostri** (e **solo** i vostri) volti mentre simulate lo svolgimento di qualche attività (es. giocare a ping pong). Queste foto (e **solo** queste) dovranno essere sufficientemente da vicino affinché i volti siano riconoscibili.

Durante gli scatti, assicuratevi che il GPS sia attivo e che le coordinate geografiche di dove siete siano memorizzate nei file delle foto.

## .2 b. Estrazione dati foto

Se avete scattato le foto come richiesto, all'interno dei file dovrebbero essere presenti le coordinate geografiche. Come primo passo, produce una tabella con i seguenti dati, ed eventualmente altri se li ritenete rilevanti

Tra i dati potenzialmente utili, considerate :

- percorso file foto (es: `pics/parco-s-chiara/DSC10132323.jpg`)
- latitudine, longitudine
- nome file `DSC10132323.jpg`
- nome foto “Prato”
- descrizione “Prato usato per cricket e calcio”

La lista **non** è esaustiva e **non** è necessario mettere tutte le colonne indicate, stabilite voi cosa è meglio per il vostro progetto.

Concretamente, dovrete estrarre metadati GPS dalle foto, che sono codificati nel cosiddetto formato EXIF. Come spunto per farlo con Python, potete guardare [questo tutorial](#)<sup>26</sup> (in inglese). Per avere le coordinate in formato usabile, dovrete convertirle in formato decimale (guardare paragrafo *Decimal form*)

Il tutorial usa la libreria [pillow](#)<sup>27</sup>, che potete installare così:

**Anaconda:**

Apri Anaconda Prompt (per istruzioni su come trovarlo o se non hai idea di cosa sia, prima di proseguire [leggi sezione interprete Python nell'introduzione](#)<sup>28</sup>) ed esegui:

```
conda install -c anaconda pillow
```

**Linux/Mac:**

- installa ipywidgets (`--user` installa nella propria home):

```
python3 -m pip install --user pillow
```

---

<sup>26</sup> <https://www.sylvaindurand.org/gps-data-from-photos-with-python/>

<sup>27</sup> <https://python-pillow.org>

<sup>28</sup> <https://it.softpython.org/intro/intro-sol.html#L'interprete-Python>

### .3 c. Anonimizzare foto

Aggiungere una funzionalità di anonimizzazione volti, implementandolo seguendo il tutorial [Computer vision su SoftPython](#)<sup>29</sup> (che ha sua volta con la [libreria opencv](#)<sup>30</sup> se non riuscite a fare tutto in automatico, eventualmente potete dividere in due step: nel primo fornite agli utenti per segnare manualmente i poligoni sulle facce, e nel secondo si svolge l'anonimizzazione in base al poligono fornito).

### .4 d. Realizzare sito

Realizzare un prototipo di sito con una mappa, implementando nell'ordine, le seguenti funzionalità.

- **NOTA 1** ricordatevi che questo è un prototipo, dovete comunicare un'idea, non fare un sito professionale
- **NOTA 2:** NON serve fare anche app per smartphone, realizzarle bene è difficile e ancora più difficile è mostrarle durante una demo al pubblico (non vorrete mica girare con lo smartphone tra i presenti ...)

Per farlo, seguire il tutorial [Interfacce grafiche](#)<sup>31</sup> e guardare in particolare il [sottocapitolo sulle mappe](#)<sup>32</sup>

#### d.1 mostrare e caricare foto

i punti di interesse (POI) con foto, e consenta di caricarne altri. Al caricamento di una foto, questa dovrà essere automaticamente anonimizzata

#### d.2 mostrare eventi pubblici

Mostrare sulla mappa attività pubbliche (es. sagre, concerti). A tal fine, si dovranno cercare opportuni dataset che le descrivano (da fonti come [dati.trentino.it](#)<sup>33</sup>) e sviluppare un programma che integri i dati nella mappa.

#### d.3 interfaccia per organizzare eventi

Realizzare un'interfaccia per organizzare eventi informali. Ogni evento avrà un luogo con le coordinate all'interno parco, la tipologia (es. partita cricket), un numero minimo di partecipanti. Chi vuole può iscriversi all'evento. Per esempio, raggiunto un certo numero di partecipanti l'evento viene mostrato con un colore diverso sulla mappa

Ricordatevi che questo è un prototipo, quindi fate una cosa semplice senza login e notifiche.

## .5 Dataset

### Foto

Le foto **le farete voi** - NON scaricate foto da internet.

<sup>29</sup> <https://it.softpython.org/computer-vision/computer-vision-sol.html>

<sup>30</sup> <https://opencv.org/>

<sup>31</sup> <https://it.softpython.org/gui/gui-sol.ipynb>

<sup>32</sup> <https://it.softpython.org/gui/gui-sol.html#Mappe>

<sup>33</sup> <http://dati.trentino.it>

### EXTRA: Concorso fotografico Wiki Loves Monuments

Già che ci siete e solo se volete / vi interessa (non è parte della valutazione), dal 1° al 30 settembre potete partecipare al concorso fotografico [Wiki Loves Monuments](#)<sup>34</sup>, è semplice:

Guardate la [lista dei monumenti autorizzati per il concorso](#)<sup>35</sup> (in particolare, [vedere lista per Trento](#)<sup>36</sup>) e trovate il soggetto che più vi piace (nel nostro caso della challenge ICTDays, sceglietene uno in un parco !!!). Dal 1° al 30 settembre caricate una o più foto dei beni culturali fotografati su Wikimedia Commons, l'archivio multimediale di Wikimedia: potete andare a caccia di nuove immagini nei 30 giorni del concorso ma anche utilizzare scatti che hai già realizzato in passato!

**ATTENZIONE:** Se decidete di aderire, **NON** mandate foto con facce, nemmeno le vostre!

**ATTENZIONE:** Wikimedia Commons accetta **solamente** foto pubblicate con licenza libera CC BY-SA ([Creative Commons attribuzione condividi allo stesso modo](#)<sup>37</sup>). Per i dettagli leggete il [regolamento](#)<sup>38</sup>

### OpenStreetMap

[OpenStreetMap](#)<sup>39</sup> è la mappa libera del mondo realizzata da volontari.

Licenza: [ODBL](#)<sup>40</sup>

### [dati.trentino.it](#)

Catalogo dati trentino: [dati.trentino.it](#)<sup>41</sup>

Cercate dataset che riportino informazioni su avvenimenti nei parchi.

Licenza: in genere i dataset sono Creative Commons [CC-0](#)<sup>42</sup> o [CC-BY](#)<sup>43</sup>, comunque controllate.

Per capire come leggere i vari formati che potete trovare, potete trarre spunto dal tutorial [Formati dati](#)<sup>44</sup>

---

<sup>34</sup> <https://wikilovesmonuments.wikimedia.it/concorso-fotografico-wlm/>

<sup>35</sup> [https://it.wikipedia.org/wiki/Progetto:Wiki\\_Loves\\_Monuments\\_2019/Monumenti](https://it.wikipedia.org/wiki/Progetto:Wiki_Loves_Monuments_2019/Monumenti)

<sup>36</sup> [https://it.wikipedia.org/wiki/Progetto:Wiki\\_Loves\\_Monuments\\_2019/Monumenti/Trentino-Alto\\_Adige/Provincia\\_autonoma\\_di\\_Trento/](https://it.wikipedia.org/wiki/Progetto:Wiki_Loves_Monuments_2019/Monumenti/Trentino-Alto_Adige/Provincia_autonoma_di_Trento/N-Z#Trento)

N-Z#Trento

<sup>37</sup> <https://creativecommons.org/licenses/by-sa/4.0/deed.it>

<sup>38</sup> <https://wikilovesmonuments.wikimedia.it/wp-content/uploads/2019/05/Regolamento-WLM-2019.pdf>

<sup>39</sup> <https://www.openstreetmap.org/#map=16/46.0662/11.1266>

<sup>40</sup> <https://www.openstreetmap.org/copyright>

<sup>41</sup> <https://dati.trentino.it>

<sup>42</sup> <https://creativecommons.org/publicdomain/zero/1.0/deed.it>

<sup>43</sup> <https://creativecommons.org/licenses/by/4.0/deed.it>

<sup>44</sup> <https://it.softpython.org/#formats>



## Lavoro 4.0 Challenge



**Sponsor:** Agenzia del Lavoro - Provincia Autonoma di Trento<sup>45</sup> ed Engineering<sup>46</sup>

L'offerta di servizi per l'incontro domanda/offerta di lavoro da parte delle Pubbliche Amministrazioni italiane è in costante evoluzione per venire incontro alle sempre mutevoli esigenze del mercato del lavoro nonché alle novità introdotte sul fronte tecnologico. Da anni la Provincia Autonoma di Trento, tramite l'Agenza per il Lavoro provinciale, ha creato, in collaborazione con il proprio partner IT Engineering, *Trentino Lavoro*<sup>47</sup>. Si tratta di un portale di servizi online per il lavoro che raccoglie al proprio interno quanto a disposizione nel territorio della Provincia Autonoma di Trento per fruire dei servizi afferenti l'area "Lavoro". Al proprio interno, infatti, tutti i soggetti coinvolti a vario titolo nella filiera del mercato del lavoro (cittadini, aziende, operatori pubblici e privati) operano in autonomia senza necessità di presentarsi presso gli sportelli dei Centri per l'Impiego.

In particolare, uno dei principali servizi offerti riguarda la gestione della **domanda e offerta di lavoro**: le aziende, autonomamente o per tramite degli operatori dei Centri per l'Impiego, caricano su *Trentino Lavoro* delle opportunità di lavoro (vacancy), alle quali i cittadini, interessati a ricercare lavoro possono candidarsi utilizzando i loro curriculum vitae (CV) e le loro lettere di presentazione creati e salvati all'interno del Portale stesso. Terminato il periodo di apertura delle candidature, l'azienda procederà quindi a valutare le candidature ricevute. Parimenti, le aziende possono effettuare delle ricerche nella banca dati dei CV anche in assenza di specifiche offerte di lavoro.

Punto focale del servizio offerto è, quindi, la ricerca che il Portale consente di effettuare. Ciascuna offerta di lavoro e ciascun CV, **strutturati secondo modelli predefiniti**, contengono sia campi tabellari sia campi testuali. Questi ultimi, tuttavia, difficilmente possono essere usati in modo esaustivo in considerazione dell'uso di parole dal significato simile ma non identiche tra loro (per es. cercando "cuoco" non troverebbe offerte di lavoro per "cuoco capo partita"). Oggi quindi il portale si limita ad utilizzare i campi tabellari contenuti nei CV e nelle vacancy, riducendo di molto le possibilità di incrocio fra la domanda e l'offerta che oggi sempre di più tendono ad esprimere i loro bisogni in un **linguaggio naturale**. Per ciascuna categoria di utenti (**Cittadini, Aziende, CPI/Agenzia**) la sfida è quella di ridisegnare i canali digitali offerti su *Trentino Lavoro* dall'Agenzia, andando poi ad innestarvi nuove modalità di servizio, **potenziate tramite l'uso dell'Intelligenza Artificiale**.

Da qui, un primo passo è costituito dalla modifica delle modalità con cui il Portale effettua la ricerca, basandosi quindi non sulle sole parole contenute in offerte di lavoro e CV ma sul **significato semantico** che esse trasmettono, ampliando così il numero e la potenziale rilevanza dei risultati ottenuti.

### Challenge

La semplice ricerca testuale per parole coincidenti sulla vacancy rispetto a quanto presente nel o nei CV quasi mai risulta essere efficace ed esaustivo nella ricerca di tutti i risultati potenzialmente desiderati e/o desiderabili. Infatti, tipicamente, il modo di procedere di un essere umano è quello di effettuare una ricerca per concetti e analogie in cui le possibilità di matching, tenendo conto delle distanze semantiche tra i concetti trovati tra CV e offerta di lavoro (es. un cv in cui ci sia la

<sup>45</sup> <https://www.agenzialavoro.tn.it/>

<sup>46</sup> <https://www.eng.it>

<sup>47</sup> <https://www.sil.provincia.tn.it/welcomepage/>

qualifica di “assistente alla cucina” potrà essere applicabile, con una opportuna valutazione di distanza, per una vacancy in cui si richieda un “cuoco”), diventano molto più ampie.

Ci viene quindi chiesto di sviluppare un prototipo che arricchisca i testi con il loro *significato semantico* al fine di ampliare le possibilità di associazione ed i risultati delle ricerche, o anche renderle più precise con filtri aggiuntivi. Tra le possibili attività da svolgere, identificazione delle sezioni del cv e delle vacancy relative ad esperienze lavorative e/o aspirazioni e/o formazione specifica, **estrazione di parole chiave** e loro estensione semantica (concetti limitrofi e/o sinonimi semantici), estrazione di impieghi passati dai CV, etc (vedere [teoria similarità del testo](#)<sup>48</sup>)

Concretamente, si dovranno **integrare i dataset relativi** a CV e Vacancy ed estenderli con ulteriori “colonne” in cui verranno riportate, ad es. le parole oggetto di estensione semantica, in modo da poter successivamente coinvolgere anche questi “nuovi” termini nella ricerca, permettere di fornire delle analisi di distanza e di effettuare filtri più raffinati.

### .1 a. Analisi

I dataset sono complessi, in particolare quello dei CV (nel paragrafo finale sui [Dataset](#) sono descritti meglio). Dovrete quindi esaminare attentamente i campi utili da estrarre, in base alle esigenze che credete sia più importante soddisfare nei tempi limitati a disposizione. In ogni caso, ricordatevi che dovete sviluppare un prototipo che comunichi un'idea, non un prodotto completo.

Per fare esperimenti, si raccomanda di estrarre pochi cv e annunci (aprire tutti i cv in LibreOffice potrebbe pure risultare in una navigazione troppo lenta per essere agevole) e metterli in file separati.

Per una discussione generale, vedere [analisi dati](#)<sup>49</sup>

### .2 b. Ricerca base

Un primo approccio potrebbe essere la ricerca per termini chiave e quindi, implementare un processo di information retrieval con delle semplici funzioni python che permettono di specificare i vari argomenti di ricerca. Una possibile evoluzione potrebbe essere quella di realizzare una semplice interfaccia di utilizzo del motore di ricerca utilizzando, ad es., HTML.

Vedere [esempi](#)<sup>50</sup>

### .3 c. Ricerca avanzata

Si potrebbero migliorare le funzionalità di ricerca sfruttando i campi oggetto di arricchimento semantico, permettendo di filtrare:

- estrazione di CV e vacancy tra loro correlabili per concetti simili
- definizione di un ranking di matching sulla base di distanze tra i concetti
- area geografica, comune

Esempi di ricerca:

1. Cittadino vuole sapere le posizioni aperte per un certo tipo di lavoro
2. Azienda vuole sapere i candidati migliori per una posizione di cuoco

Per realizzare quanto sopra, si potrebbe effettuare analisi semantica del testo usando il servizio Dandelion di SpazioDati con l'API Entity Extraction, documentazione [API](#)<sup>51</sup> e [libreria Python](#)<sup>52</sup>) e l'uso di risorse come Wikipedia o dizionari

<sup>48</sup> <https://it.softpython.org/information-retrieval/information-retrieval-sol.html#Similarit%C3%A0-semantica>

<sup>49</sup> <https://it.softpython.org/jm-templates/project-NAME-SURNAME-ID/project.html#Analisi-dati>

<sup>50</sup> <https://it.softpython.org/information-retrieval/information-retrieval-sol.html#Costruiamo-il-nostro-motore-di-ricerca>

<sup>51</sup> <https://dandelion.eu/docs/api/datatxt/nex/v1/>

<sup>52</sup> <http://python-dandelion-eu.readthedocs.io/en/latest/datatxt.html>

per la gestione degli aspetti semantici (sinonimi, concetti limitrofi ed eventuali distanze). Per sapere come fare, puoi consultare gli [esempi sul sito SoftPython](#)<sup>53</sup>. Esempi di arricchimento con Dandelion:

Nel dataset delle vacancies, nel campo `Titolo` `annuncio` è presente il testo “OPERAIO PER CARPENTERIA METALLICA/SALDATORE” [Vedi testo arricchito](#)<sup>54</sup>

Nel dataset dei cv, nel campo `Desiderate: descrizione professione` è presente il testo “CAPOSQUADRA CARPENTIERE” [Vedi testo arricchito](#)<sup>55</sup>

Come si può notare, in uno è scritto “CARPENTERIA” mentre nell’altro è scritto “CARPENTIERE”. Nonostante le differenza categoria/mestiere, Dandelion ad entrambi i testi assocerà il singolo concetto ‘Carpentiere’ esplicitandolo con un link a Wikipedia: <http://it.wikipedia.org/wiki/Carpentiere>

## e. Interfaccia utente

Realizzare un prototipo di interfaccia grafica (anche rudimentale) che consenta di scegliere quale profilo essere (cittadino / impresa) ed effettuare ricerche immettendo filtri e testo libero

## .4 Dataset

Sono forniti due file, `dataset_CV.csv` e `dataset_vacancy.csv`

Per i dati completi chiedere a [david.leoni@unitn.it](mailto:david.leoni@unitn.it)

I campi hanno intestazioni parlanti, per cui non li descriveremo tutti nel dettaglio. In genere:

- in giallo vi sono i campi strutturati, provenienti da form di input in cui gli operatori, i cittadini o le aziende possono scegliere i valori in modo vincolato
- in verde vi sono le colonne che vengono compilate con testo libero
- Le colonne su cui si può soffermare l’attenzione e su cui effettuare l’analisi del testo sono evidenziate in verde con carattere rosso:
  - colonna `AI` per `dataset_CV.csv`
  - colonna `B` per `dataset_Vacancy.csv`

### `dataset_vacancy.csv`

Righe: 401

Dimensione: 110.5 KB

Campi: per le vacancy ci possono essere più ID uguali: rappresentano la stessa posizione espressa da una azienda per cui possono cambiare dei dettagli (es. cuoco per pranzo, cuoco per cena, ... oppure conoscenza arabo o conoscenza inglese)

<sup>53</sup> <https://it.softpython.org/information-retrieval/information-retrieval-sol.html#Prendiamo-le-distanze>

<sup>54</sup> [https://dandelion.eu/semantic-text/entity-extraction-demo/?text=OPERAIO+PER+CARPENTERIA+METALLICA%2FSALDATORE&lang=it&min\\_confidence=0&exec=true#1500677](https://dandelion.eu/semantic-text/entity-extraction-demo/?text=OPERAIO+PER+CARPENTERIA+METALLICA%2FSALDATORE&lang=it&min_confidence=0&exec=true#1500677)

<sup>55</sup> [https://dandelion.eu/semantic-text/entity-extraction-demo/?text=CAPOSQUADRA+CARPENTIERE&lang=it&min\\_confidence=0&exec=true#results](https://dandelion.eu/semantic-text/entity-extraction-demo/?text=CAPOSQUADRA+CARPENTIERE&lang=it&min_confidence=0&exec=true#results)

## dataset\_CV.csv

Righe: 102011

Dimensione: 151.5 MB

CV rappresentati: 133

- Un cv può presentare più righe: a ID (colonna A) uguali corrispondono più dettagli della stessa persona ovvero sono parte dello stesso cv
- È stata fatta una cernita dei soli CV che contenessero almeno una riga con campo a testo libero valorizzato (AI).

## A Prova di Hacker Challenge



**Sponsor:** SpazioDati<sup>56</sup>.

Nel mondo del **business**, quando si vuole fare affari con qualcuno, è importante capire se l'interlocutore con cui si sta trattando sia **affidabile** o meno. Per esempio, se la potenziale **azienda** partner dichiara di poterci fornire prodotti per centinaia di milioni di euro ma risulta avere pochissimi dipendenti e scarso fatturato, potrebbe venirci qualche dubbio sull'effettiva capacità di realizzare quanto promesso.

Per permettere di farsi rapidamente un'idea riguardo un'azienda, **SpazioDati** ha realizzato [Atoka.io](https://atoka.io/)<sup>57</sup>, un motore di ricerca di aziende italiane, che mostra in quale settore di mercato opera un'impresa, il fatturato, il numero di dipendenti, la presenza sul web, e altri fattori. SpazioDati raccoglie queste informazioni sfruttando numerose fonti dati, fra cui dati ufficiali dalla Camera di Commercio, OpenData, dati estratti dalle news e dai siti web.

Tra i fattori analizzati vi sono i **siti aziendali**, sempre più importanti nell'era di Internet. Purtroppo, spesso sono mantenuti senza la dovuta attenzione e quindi vulnerabili agli **attacchi di hacker** malintenzionati, che possono portare a sostituzione di pagine del sito con pubblicità indesiderata, rivelazione dei dati privati dei clienti, fuoriuscita di segreti industriali, e in generale gravi danni d'immagine all'azienda.

Per migliorare la determinazione dell'affidabilità di un impresa, SpazioDati vuole quindi assegnare un **punteggio** ai siti aziendali sviluppando un indice (score) che misuri quanto sono vulnerabili. Dovrà anche essere realizzata una visualizzazione semplice da comprendere anche a **personale non-tecnico**, come dirigenti e figure nel ramo commerciale.

Per raccogliere dati, SpazioDati ha eseguito [Wappalyzer](https://github.com/AliasIO/Wappalyzer)<sup>58</sup> su milioni di siti web, estraendo le tecnologie utilizzate e le loro versioni. Vi chiediamo di incrociare i dati estratti dai siti web con le vulnerabilità espone in cataloghi specializzati (ad esempio da [itsecdb oval](https://www.itsecdb.com/oval/)<sup>59</sup>), e sviluppare uno score riassuntivo basato sulle versioni dei software (es: il software è vulnerabile o poco aggiornato) e sul contenuto testuale del sito.

---

<sup>56</sup> <http://spaziodati.eu>

<sup>57</sup> <https://atoka.io/>

<sup>58</sup> <https://github.com/AliasIO/Wappalyzer>

<sup>59</sup> <https://www.itsecdb.com/oval/>

## .1 a. Analisi

- Analizzare il mercato potenziale e la distribuzione delle tecnologie impiegate (non guardare solo il dataset fornito ! Cercate anche statistiche altrove ). Su quali tecnologie conviene concentrarsi? Quanto spesso vengono usate?
- Trovare dataset di vulnerabilità adatti allo scopo, ponendo particolare attenzione alla licenza dei dati (sono usabili commercialmente? Bisogna pagarli?). Sono aggiornati frequentemente? Ci sono versioni a pagamento più ricche di informazioni / aggiornate ?
- studio delle possibilità business (a chi si può vendere il servizio che andrete a creare? Come si può presentare in modo efficace?)

## .2 b. Integrazione

Incrociare i dati web di spaziati con il / i dataset di vulnerabilità trovati, producendo un'unica tabella csv (o json).

## .3 c. Sviluppo di un indice

Determinare un indice (*score*) che tenga in considerazione le vulnerabilità riscontrate assegnando un peso a ciascuna. **IMPORTANTE:** nel report tecnico deve essere chiaramente riportata e motivata la formula usata per determinare l'indice: presentare un software anche apparentemente funzionante non è sufficiente. Implementare il calcolo dell'indice e aggiungere una o più colonne per lo score (ed eventualmente sue componenti) ai dati integrati

## .4 d. Visualizzazione

Dato un sito, visualizzare il singolo score e presentare un riassunto

Quale può essere la visualizzazione più efficace che trasmette la maggior quantità di informazione senza confondere con tecnicismi da hacker? Tenete presente che i potenziali utilizzatori di tali grafici saranno personale non tecnico, come dirigenti e figure nel ramo commerciale.

## .5 Dataset

### dataset oval

Dataset vulnerabilità, scaricabile da [itsecdb.com/oval](https://www.itsecdb.com/oval/)<sup>60</sup>

Non ne forniamo dettagli, dovrete capire voi cosa (e in che formato) prendere la sito, capirne la licenza d'uso, se li possiamo usare, se è il caso di valutare alternative, etc.

---

<sup>60</sup> <https://www.itsecdb.com/oval/>

## rilevamento.jsonl

File fornito da SpazioDati contenente le tecnologie rilevate per ogni sito web.

Scarica file

Formato: JSONL , cioè una sequenza di JSON. Ogni riga è un json con un sito per riga. Per sapere come leggerli in Python, guardare guida [Formati dati](https://it.softpython.org/formats/formats3-json-sol.html)<sup>61</sup>

Righe: 3000

Dimensione: 44 Mb

Campi:

- domain
- text
- technologies: dizionario che contiene, per ogni categoria possibile, la lista di technologies estratte.

Esempio:

```
{
  "domain": "danzainfascia.it",
  "text": "Primary AGENDA",
  "technologies": {
    "font-scripts": [
      {
        "name": "Font Awesome",
        "categories": [
          "font-scripts"
        ],
        "confidence": 100
      }
    ],
    "web-servers": [
      {
        "name": "Apache",
        "categories": [
          "web-servers"
        ],
        "confidence": 100
      }
    ],
    "javascript-frameworks": [
      {
        "name": "jQuery",
        "categories": [
          "javascript-frameworks"
        ],
        "confidence": 100
      }
    ],
    "programming-languages": [
      {
        "name": "PHP",
        "version": "5.6.40",
        "categories": [
```

(continues on next page)

---

<sup>61</sup> <https://it.softpython.org/formats/formats3-json-sol.html>

(continued from previous page)

```

        "programming-languages"
      ],
      "confidence": 100
    }
  ],
  "blogs": [
    {
      "name": "WordPress",
      "version": "5.2.2",
      "categories": [
        "cms",
        "blogs"
      ],
      "confidence": 100
    }
  ],
  "cms": [
    {
      "name": "WordPress",
      "version": "5.2.2",
      "categories": [
        "cms",
        "blogs"
      ],
      "confidence": 100
    }
  ]
}

```

[ ]:

## Real Time Transport Challenge



**Sponsor:** U-Hopper<sup>62</sup> / Thinkin<sup>63</sup>

Nell'ambito dei **trasporti pubblici**, per comprendere appieno le dinamiche di afflusso dei passeggeri sugli **autobus** e quindi ottimizzare il numero di corse e i tragitti, è importante sapere il **numero di passeggeri** per linea durante la giornata. Una società di trasporti pubblici di una città che U-Hopper non è autorizzata a rivelare (a scanso di equivoci, la città non è comunque Trento), è interessata pertanto a verificare se sia possibile stimare il numero di passeggeri a bordo

<sup>62</sup> <https://u-hopper.com/>

<sup>63</sup> <https://thinkin.io>

di un autobus sulla base del numero di dispositivi (**smartphone**, essenzialmente) rilevati a bordo dell'autobus stesso tramite opportuni **sensori**. Una procedura di rilevazione di questo tipo presenta alcuni vantaggi rispetto a metodologie più tradizionali: per esempio, rispetto alla rilevazione manuale, non richiede l'impiego a tempo pieno di un addetto su ogni autobus che conteggi il numero di passeggeri a bordo nel corso tragitto; rispetto invece a una misurazione effettuata tramite controllo dei dati delle obliterate, una rilevazione tramite sensori consente di stabilire non solo quando un passeggero salga sull'autobus ma anche quando ne discenda.

Al riguardo, la società si è rivolta a U-Hopper per la programmazione e installazione dei sensori e, una volta acquisiti i dati, per la messa a punto di un procedimento (algoritmo) in grado di analizzare le misurazioni per effettuare una stima del numero di passeggeri. Per calibrare il sistema, la società di trasporti pubblici ha messo a disposizione di U-Hopper le rilevazioni del numero di passeggeri compiute manualmente da alcuni addetti a bordo degli autobus - questi ultimi ovviamente muniti anche di sensore - di varie linee urbane durante alcune giornate.

U-Hopper ha già provveduto a realizzare quanto richiesto dal cliente, ma in futuro potrebbe risultare utile produrre **visualizzazioni grafiche** per mettere in evidenza regolarità nell'affluenza dei passeggeri (*pattern*) tramite uno strumento che risulti **facilmente comprensibile** per qualunque interlocutore. Si richiede quindi di visualizzare il numero di device rilevati mediamente giorno per giorno sulle diverse linee nelle varie fasce orarie, comparandolo eventualmente con le rilevazioni manuali. Per aumentare la fruibilità, si potrebbe creare un'apposita **interfaccia utente** stile chatbot che consenta di scegliere per es. quale fascia oraria evidenziare.

### .1 a. Analisi

Esplorare i dati per individuare a quali linee urbane si riferiscano le misurazioni, in quali giorni siano state effettuate e in quali fasce orarie (ovviamente gli autobus non circolano per tutte le 24 ore) - in questa fase occorrerà procedere a una conversione del timestamp in modo che giorno e ora siano leggibili.

### .2 b. Preparazione

Importanti operazioni di pulizia sono state già svolte da U-Hopper, come la rimozione preliminare dai dati dei segnali spuri, cioè dei segnali provenienti da device non appartenenti a passeggeri dell'autobus (per esempio, il device dell'autista, quello del guidatore e/o passeggero dell'auto che transita a fianco dell'autobus, etc.).

Quello che dovrete fare voi sarà segmentare dei dati in modo che siano raggruppati in base alla linea dell'autobus e alla fascia oraria. Dato che i sensori effettuano rilevazioni a intervalli di qualche secondo, è naturale che il device di ogni passeggero compaia ripetutamente fintanto che il passeggero sia a bordo dell'autobus e di questo aspetto occorrerà tener opportunamente conto.

### .3 c. Visualizzazione

#### c.1 Utilizzo linee

Per ogni linea di autobus si potrebbero produrre una o più visualizzazioni grafiche (tabella, istogramma, diagramma a torta, etc.) relative alla frequentazione dell'autobus nelle varie fasce orarie delle diverse giornate (si potrebbe considerare un intervallo temporale di un'ora, per esempio), mettendo in evidenza quanto sia stato utilizzato l'autobus in ogni fascia rispetto alla media giornaliera della linea. Una visualizzazione di questo tipo sarebbe per esempio utile alla società di trasporti per verificare se sia il caso di intensificare - o rarefare - il numero di corse rispetto alla prassi attuale, tenuto anche conto delle variazioni di affluenza nei giorni festivi rispetto ai feriali.



## c.2 Differenze dati device / registrazioni manuali

Siccome non tutti i passeggeri dispongono necessariamente di un device, i numeri nelle visualizzazioni relative alle rilevazioni dei sensori non coincideranno con gli analoghi numeri che si potrebbero ottenere a partire dalle misurazioni manuali. Cionondimeno, un confronto tra i picchi rilevati in un caso e nell'altro potrebbe essere utile a segnalare eventuali anomalie/malfunzionamenti dei sensori (peraltro nuovi) o blackout, aiutando così U-Hopper o la società dei trasporti stessa a procedere con le necessarie verifiche e, nel caso, a intervenire opportunamente.

Si potrebbero quindi produrre visualizzazioni analoghe a quelle cui si fa riferimento nella sezione d.1, impiegando però i dati ottenuti tramite le misurazioni effettuate manualmente dagli addetti della società di trasporti, così da procedere poi al confronto qui sopra descritto.

## d. Interfaccia interattiva

Presentare il tutto in un chatbot o sito, includendo elementi interattivi: per esempio, un utente potrebbe inserire la linea urbana, il giorno della settimana e la fascia oraria e vedersi restituito il corrispondente numero di device rilevati, nuovamente evidenziato in maniera diversa a seconda di quanto elevato o basso sia il numero di device rapportato alla media giornaliera.

## e. Inferenza

Se vi avanza tempo, come extra guardate cosa si può inferire solo analizzando la distribuzione dei picchi nell'arco di una singola giornata. Potreste avanzare ipotesi sulla tipologia di area urbana servita dalla varie linee di autobus: per esempio, picchi al mattino presto e in tarda mattinata potrebbero far ipotizzare che la linea urbana in esame serva dei poli scolastici, picchi al mattino presto e nel tardo pomeriggio porterebbero invece a credere che la linea urbana serva aree di uffici e/o industrie. Provate quindi a realizzare una tabella che metta in corrispondenza le fermate con il tipo di luogo.

## .4 Dataset

I dataset sono forniti da U-Hopper<sup>64</sup>. Per i dati completi chiedere a [david.leoni@unitn.it](mailto:david.leoni@unitn.it)

Vi verranno forniti due file, `Rilevazioni_sensori.csv` e `Rilevazioni_manuali.csv`

Le rilevazioni - sia tramite sensori che tramite addetti della società - sono state effettuate su tre linee urbane di un Comune (che deve rimanere anonimo) nel corso di tre distinte giornate durante una settimana del mese di novembre dell'anno scorso. Più precisamente, il file con le rilevazioni dei sensori e quello con le misurazioni manuali contengono circa 110000 e 5300 righe rispettivamente, per un totale di circa 2.5 Mb e 210 Kb di memoria. Le rilevazioni coprono le fasce orarie dalle 5-6 del mattino (circa) fino alle 20-21 (circa), eccezion fatta per una linea di bus che, in due delle tre giornate in questione, ha dati solo tra le 5 e le 10 per un giorno e tra le 6 e le 16 per l'altro.

### Rilevazioni\_sensori.csv

Righe: 110000

Dimensione: 2.5 Mb

Colonne:

- `Line`: linea di autobus della rete urbana
- `Timestamp`: timestamp dell'istante in cui la rilevazione è stata effettuata
- `ID Device`: id del device rilevato (opportunamente anonimizzato)

<sup>64</sup> <https://www.u-hopper.com>

Esempio:

Line	Timestamp	ID_Device
11	1542859801.14977	53c5a73e-b1a3-5618-81f2-ec0d8a24cb10
11	1542859805.09879	53c5a73e-b1a3-5618-81f2-ec0d8a24cb10
11	1542860008.60541	0d36bc0a-c1c9-535a-92fe-187dcb670dc4
6	1542861382.94801	538072a4-722b-5e5a-afdd-46bcd9626245
6	1542861384.6065	5656b67a-67d3-5b31-8e53-637b1ef501e7
6	1542861384.89655	7493648b-4eef-5029-bfe3-f363aae59c2d

### **Rilevazioni\_manuali.csv**

Righe: 5300

Dimensione: 210 Kb

Colonne:

- linea dell'autobus
- timestamp
- Passengers: numero di passeggeri presenti sull'autobus una volta terminate le operazioni di salita e discesa a una fermata (è qui utile specificare che gli addetti hanno proceduto alla rilevazione alla ripartenza dell'autobus da ogni fermata)
- Gone\_up: numero di passeggeri saliti alla fermata
- Gone\_down: numero di passeggeri discesi

Esempio:

Line	Timestamp	Passengers	Gone_up	Went_down
6	1542870352	17	2	1
6	1542870960	12	0	2
6	1542871204	11	3	0
9	1542873885	5	0	0
9	1542873976	0	0	5
9	1542874068	18	18	0