GROUP C

Baldoni V., Boffa M., Fabiani F.,

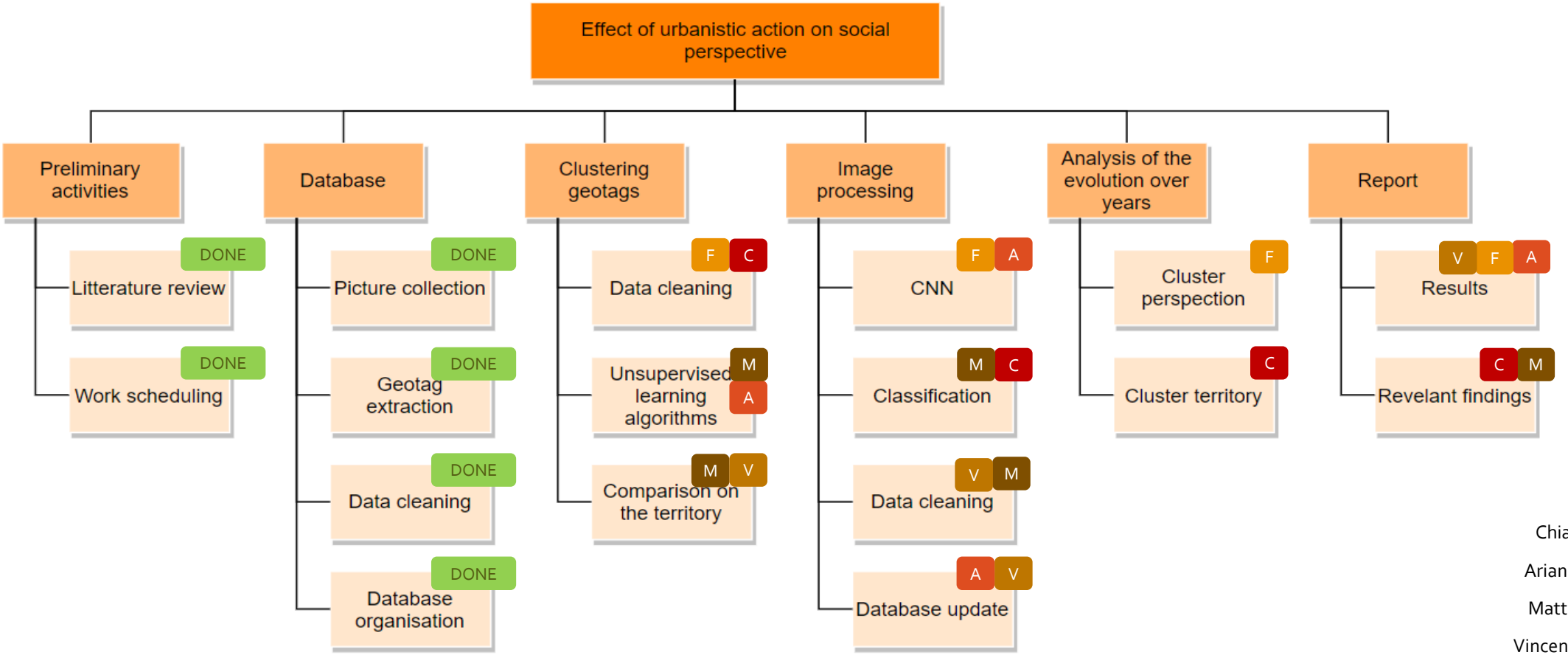
Lanza C., Ravera A.

Effect of urbanistic action on social perspective

Capture the user's perception of a city exploiting its digital footprints and study how the city zones evolved looking at the geo-tagged contents

Report and presentation for the administration to show the effectiveness of the adopted territorial policies and to suggest future improvements and investments

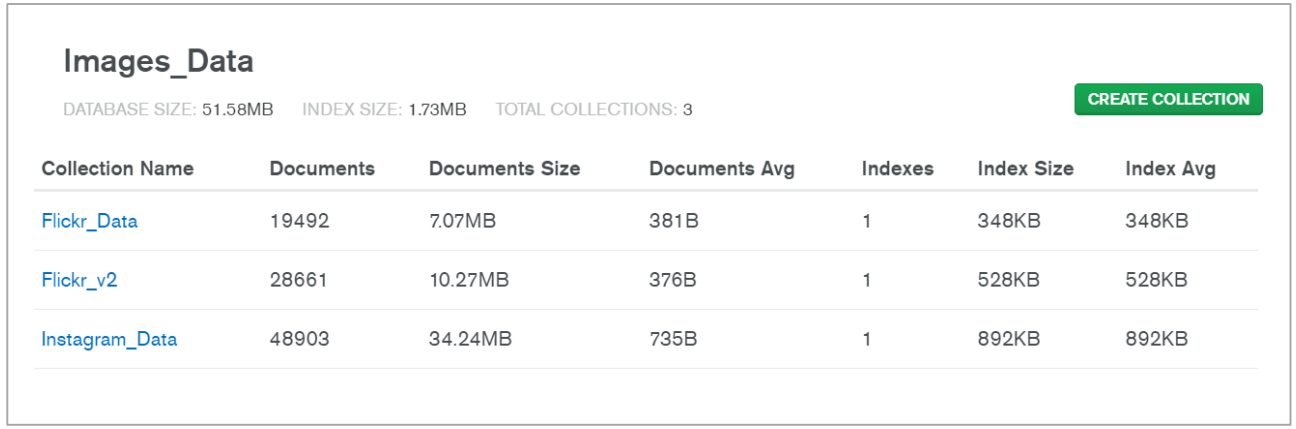
WORK PLAN


Effect of urbanistic action on social perspective
Preliminary activities
Litterature review
DONE
Work scheduling
DONE
Database
Picture collection
DONE
Geotag extraction
DONE
Data cleaning
DONE
Database organisation
DONE
Clustering geotags
Data cleaning
F C
Unsupervised learning algorithms
M A
Comparison on the territory
M V
Image processing
CNN
F A
Classification
M C
Data cleaning
V M
Database update
A V
Analysis of the evolution over years
Cluster perspection
F
Cluster territory
C
Report
Results
V F A
Revelant findings
C M
Chiara C
Arianna A
Matteo M
Vincenzo V
Federico F

System Requirement Review (SRR)

- **Data Collection**
  - **Estimated time** : 3 weeks
  - **Tools** : python, Flickr API, Instragram API, MongoDB, Dropbox
  - **Requirements** :
    - **DB organization**
    - Data downloading
    - Image downloading
    - Insert data into DB
    - Insert images into Cloud
    - Pulizia dati ridondanti, incompleti o errati

DB organization

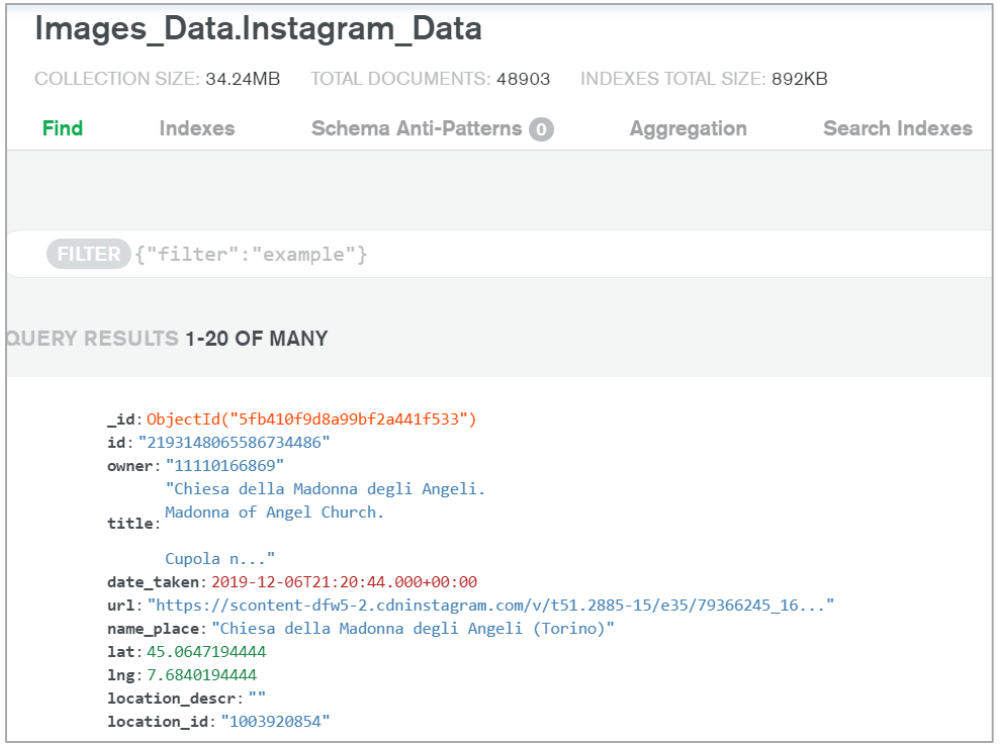


Images_Data

DATABASE SIZE: 51.58MB INDEX SIZE: 1.73MB TOTAL COLLECTIONS: 3

CREATE COLLECTION

Collection Name	Documents	Documents Size	Documents Avg	Indexes	Index Size	Index Avg
Flickr_Data	19492	7.07MB	381B	1	348KB	348KB
Flickr_v2	28661	10.27MB	376B	1	528KB	528KB
Instagram_Data	48903	34.24MB	735B	1	892KB	892KB



Images_Data.Instagram_Data

COLLECTION SIZE: 34.24MB TOTAL DOCUMENTS: 48903 INDEXES TOTAL SIZE: 892KB

Find Indexes Schema Anti-Patterns 0 Aggregation Search Indexes

FILTER {"filter":"example"}

QUERY RESULTS 1-20 OF MANY

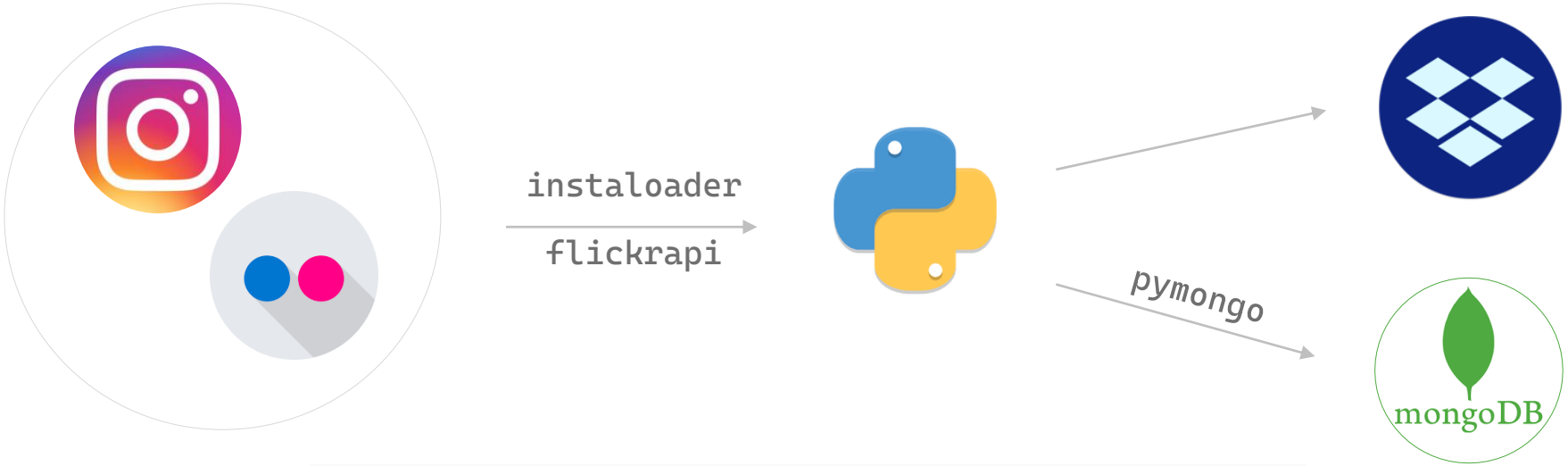

```
_id: ObjectId("5fb410f9d8a99bf2a441f533")
id: "2193148065586734486"
owner: "11110166869"
title: "Chiesa della Madonna degli Angeli.
Madonna of Angel Church.
Cupola n..."
date_taken: 2019-12-06T21:20:44.000+00:00
url: "https://scontent-dfw5-2.cdninstagram.com/v/t51.2885-15/e35/79366245_16..."
name_place: "Chiesa della Madonna degli Angeli (Torino)"
lat: 45.0647194444
lng: 7.6840194444
location_descr: ""
location_id: "1003920854"
```


- Shared cluster and online storage
- Unstructured data
- Large amount of data
- Developer-friendly maintenance

System Requirement Review (SRR)

- **Data Collection**
  - **Estimated time** : 3 weeks
  - **Tools** : python, Flickr API, Instragram API, MongoDB, Dropbox
  - **Requirements** :
    - DB organization
    - **Data downloading**
    - **Image downloading**
    - **Insert data into DB**
    - **Insert images into Cloud**
    - Pulizia dati ridondanti, incompleti o errati

Data & Image insertion and downloading



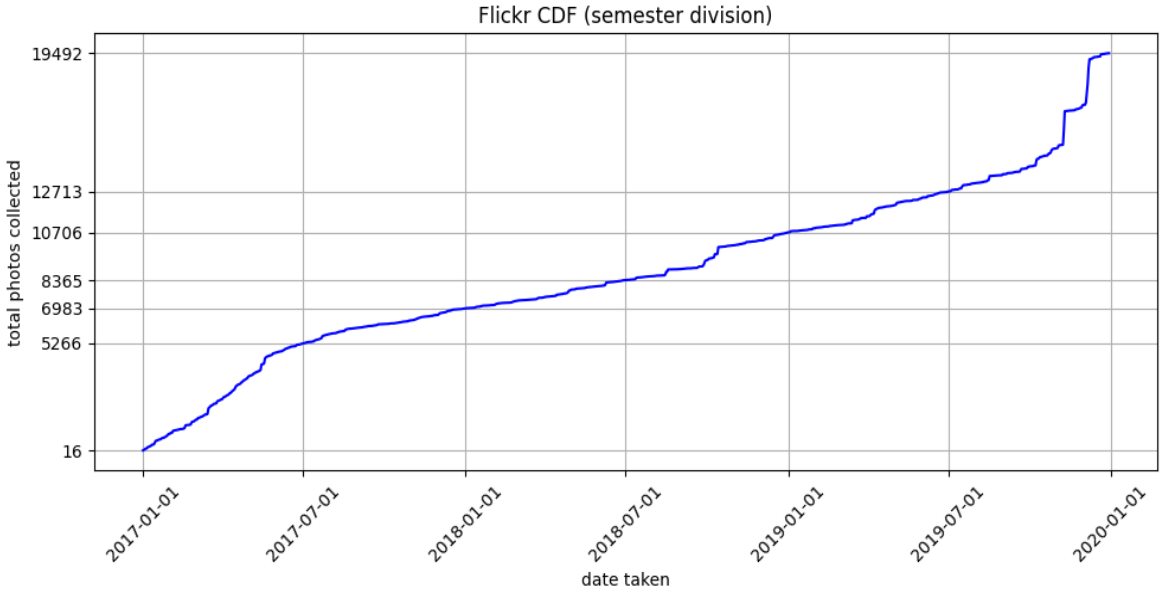

```python
#GET ALL THE PICTURES UPLOADED FROM A LOCATION
pictures_uploaded=pymongo_database.find({"location_id":location_id})

if picture not in pictures_uploaded:
    """
    DOWNLOAD IMAGE
        &
    COLLECT METADATA HERE
    """
    pymongo_database.insert_one(METADATA,ID_PICTURE)
    dropbox_folder.files_upload(STREAMED_PICTURE, DROPBOX_PATH, ID_PICTURE)
```

System Requirement Review (SRR)

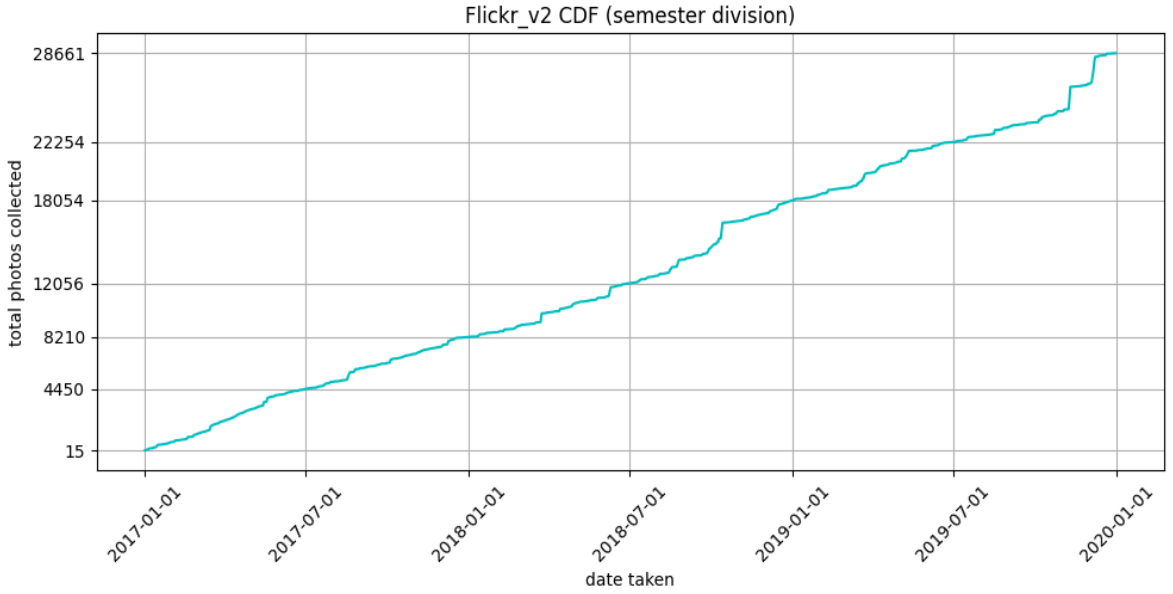
- **Data Collection**
  - **Estimated time** : 3 weeks
  - **Tools** : python, Flickr API, Instragram API, MongoDB, Dropbox
  - **Requirements** :
    - DB organization
    - Data downloading
    - Image downloading
    - Insert data into DB
    - Insert images into Cloud
    - **Data cleaning**

Data consistency

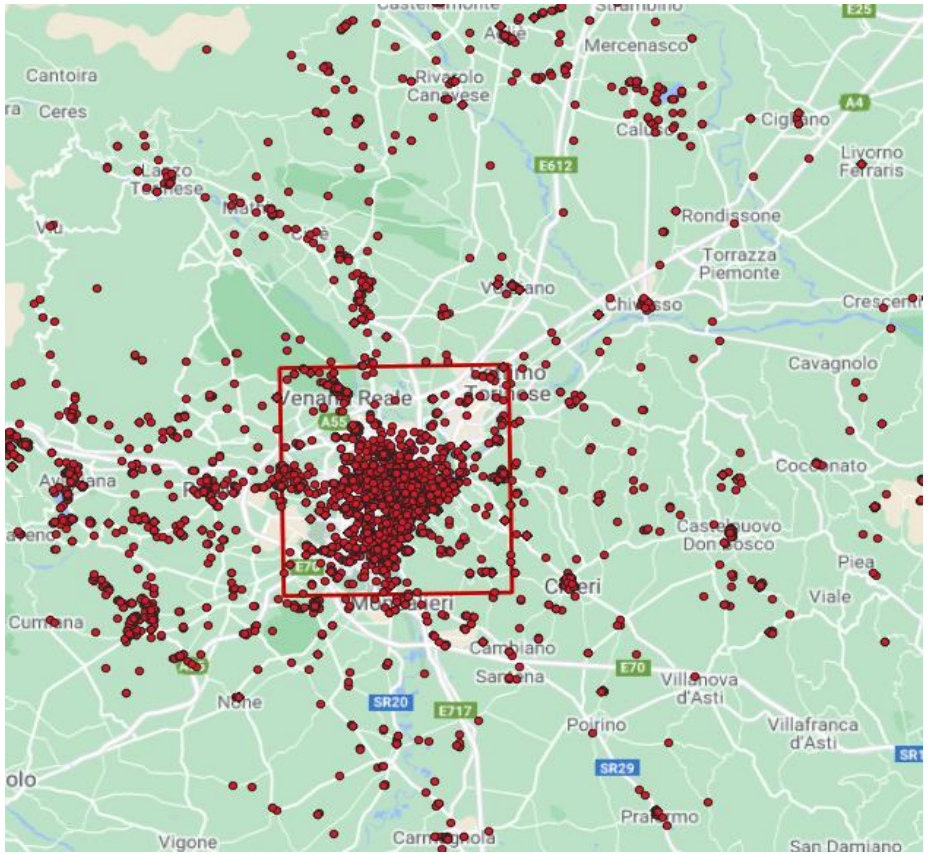


Before

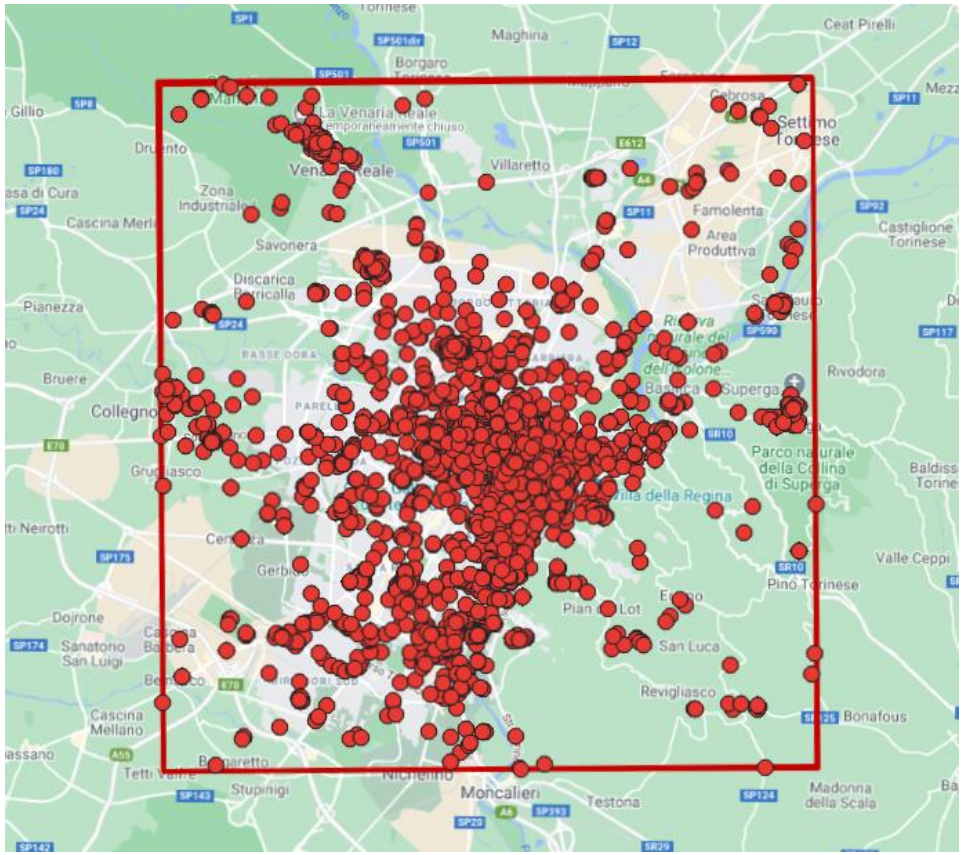
After



Data cleaning

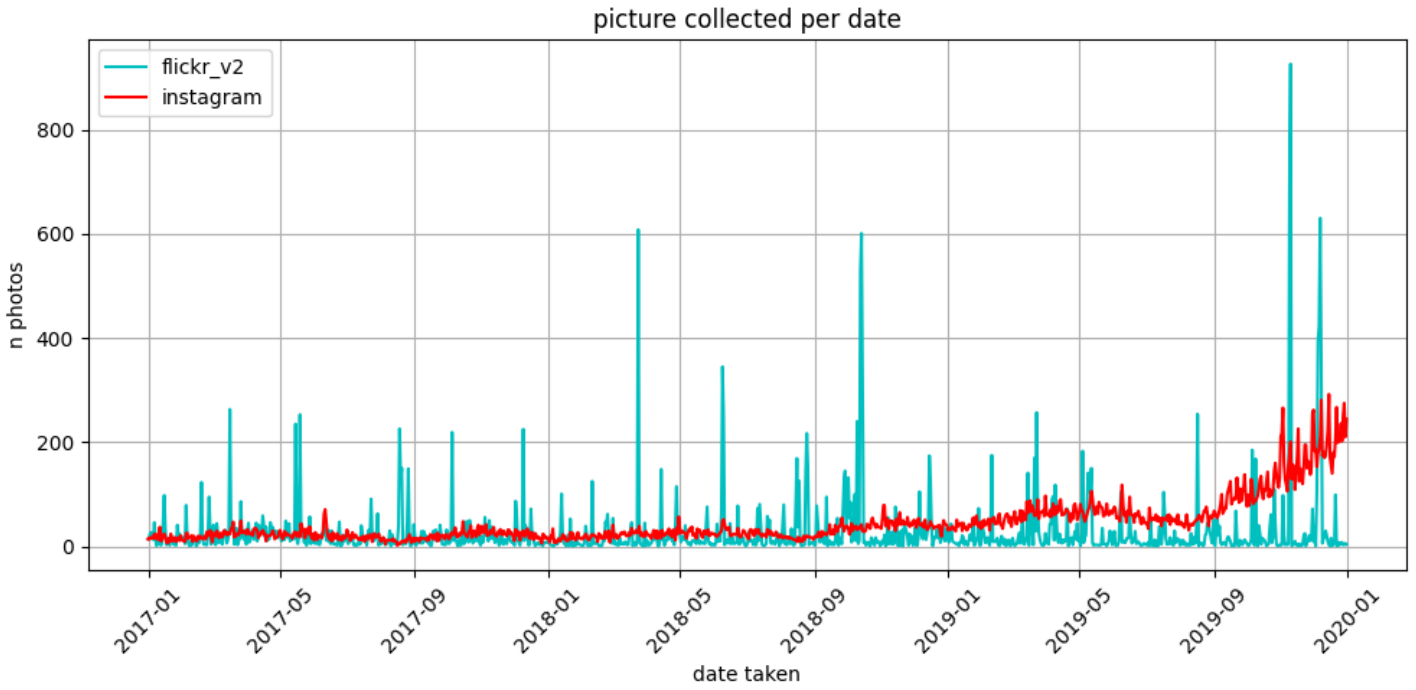


Before



After

Final data distributions



Number of images collected:

- Flickr : 28.661
- Instagram : 48.903

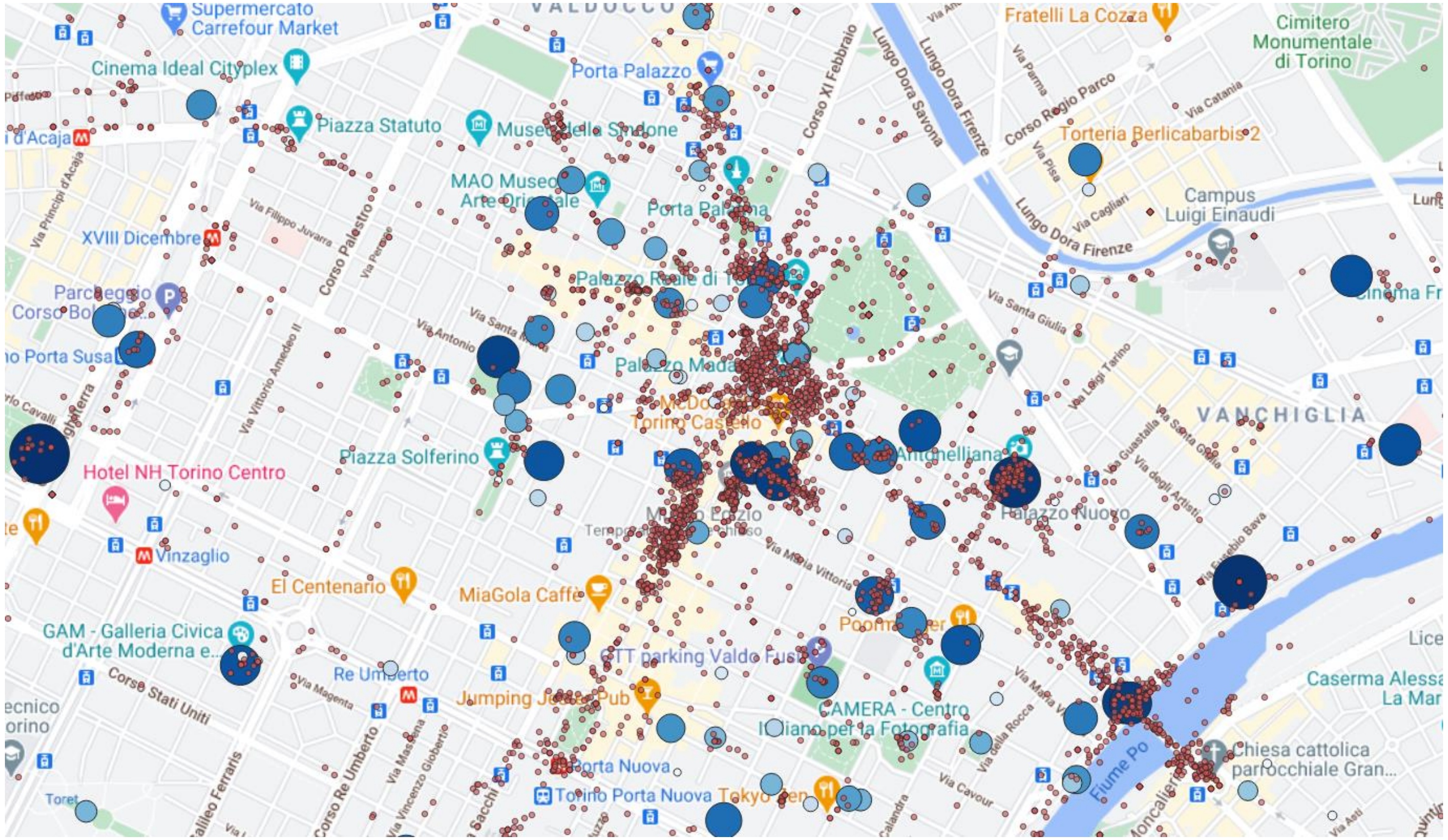
Period:

from Jan.2017 to Dec. 2019

System Requirement Review (SRR)

- **Clustering**
  - **Estimated time** : 7 weeks
  - **Tools** : Python, QGIS, Kmeans algorithm
  - **Requirements** :
    - **Study of different methods** (Kmeans, DBscan, Voronoi regions, etc.)
    - **Definition of the best method to be adopted** (Kmeans: faster, only one parameter to set, physically meaningful)
    - **Normalization of the datasets for the clustering phase**
    - Fine parameters' tuning (minimization of the inertia)
    - Data extraction and processing with the chosen method
    - Study and conclusive analysis of the results

Normalization of the datasets



LEGEND

Flickr :

Instagram :

- 1
- 203
- 506
- 1024

System Requirement Review (SRR)

- **Image processing**
  - **Estimated time** : 8 weeks
  - **Tools** : Python, Transfer learning, TensorFlow
  - **Requirements** :
    - Study of possible deep learning pre-trained models
    - Definition of the best model to be adopted
    - Complete image processing of the images into DB
    - Cleaning of non-useful images
    - Study and analysis of results

Deep Learning alternatives

Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Articles :

- Understanding tourists' urban images with geotagged photos using convolutional neural networks
- Rethinking the Inception Architecture for Computer Vision

Data & classes cleaning

Location:

Corso Inghilterra

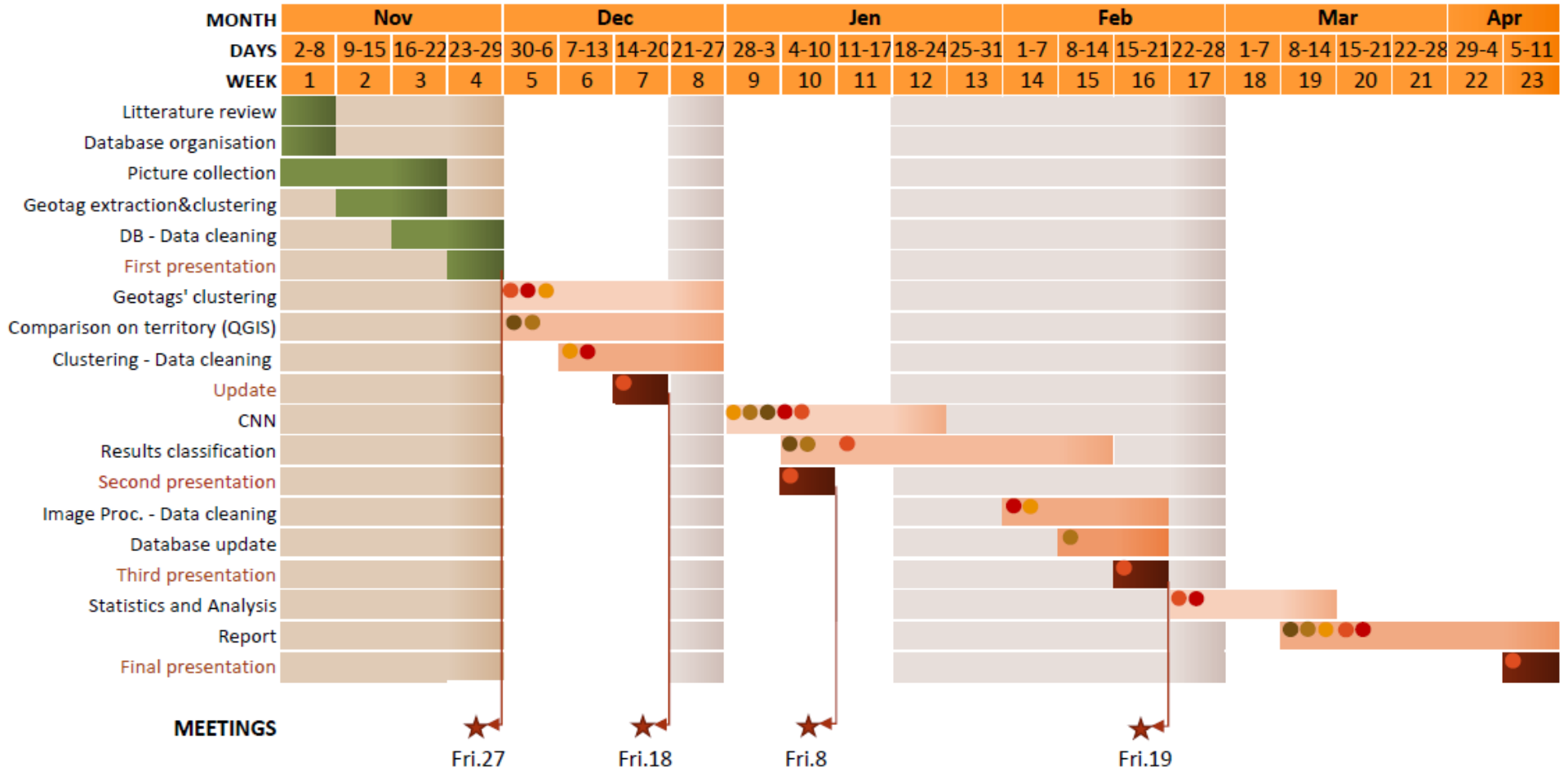
Title:

"Monday: going to work like…. yippee!!! #mondaymood #mondaymonster #t…"

1000 synsets for Task 2 (same as in ILSVRC2012)

kit fox, Vulpes macrotis
English setter
Australian terrier
grey whale, gray whale, devilfish, Eschrichtius gibbosus, Eschrichtius robustus
lesser panda, red panda, panda, bear cat, cat bear, Ailurus fulgens
Egyptian cat
ibex, Capra ibex
Persian cat
cougar, puma, catamount, mountain lion, painter, panther, Felis concolor
gazelle
porcupine, hedgehog
sea lion
badger
Great Dane
Scottish deerhound, deerhound
killer whale, killer, orca, grampus, sea wolf, Orcinus orca
mink
African elephant, Loxodonta africana

GANTT


MONTH
Nov
Dec
Jen
Feb
Mar
Apr
DAYS
2-8
9-15
16-22
23-29
30-6
7-13
14-20
21-27
28-3
4-10
11-17
18-24
25-31
1-7
8-14
15-21
22-28
1-7
8-14
15-21
22-28
29-4
5-11
WEEK
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
Litterature review
Database organisation
Picture collection
Geotag extraction&clustering
DB - Data cleaning
First presentation
Geotags' clustering
Comparison on territory (QGIS)
Clustering - Data cleaning
Update
CNN
Results classification
Second presentation
Image Proc. - Data cleaning
Database update
Third presentation
Statistics and Analysis
Report
Final presentation
MEETINGS
Fri.27
Fri.18
Fri.8
Fri.19
Chiara
Arianna
Matteo
Vincenzo
Federico

THANK YOU

for the attention