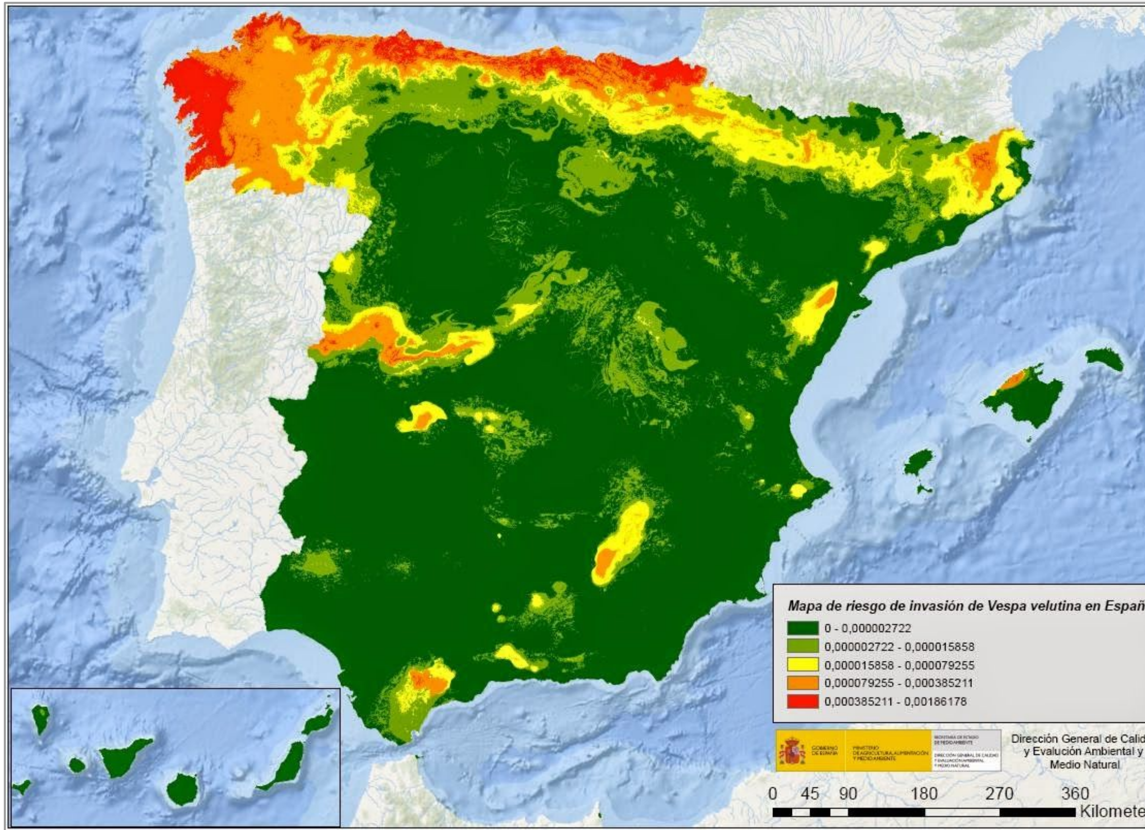
Tracking Vespa Velutina

Eco Analytics

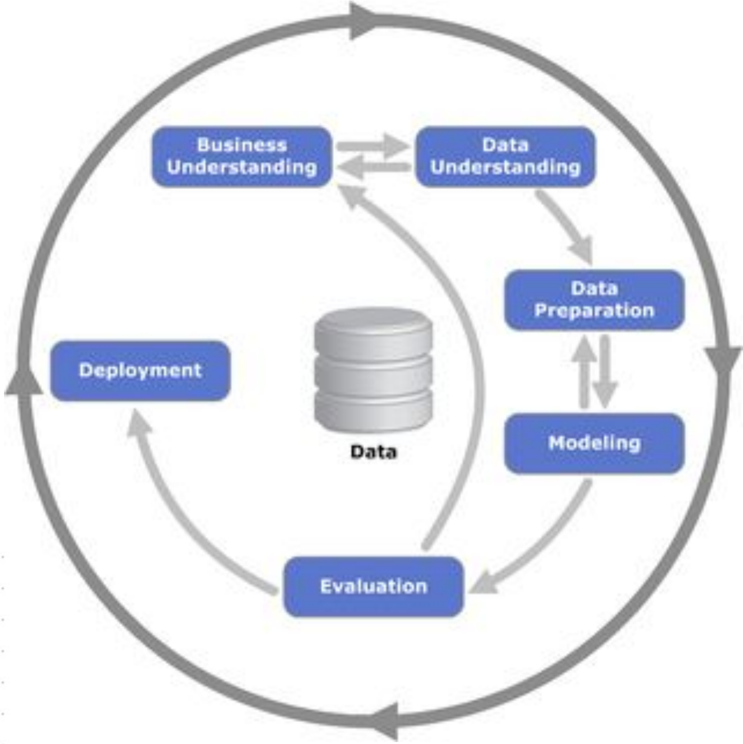
Problem Refresh

- *Vespa Velutina*, indigenous to SE Asia
- Area of concern: Basque, NW Spain
- **Invasive species**: Kills honey bees


Mapa de riesgo de invasión de Vespa velutina en Españ
🟩 0 - 0,000002722
🟩 0,000002722 - 0,000015858
🟨 0,000015858 - 0,000079255
🟧 0,000079255 - 0,000385211
🟥 0,000385211 - 0,00186178
GOBIERNO DE ESPAÑA
MINISTERIO DE AGRICULTURA, ALIMENTACIÓN Y MEDIO AMBIENTE
SECRETARÍA DE ESTADO DE MEDIO AMBIENTE
DIRECCIÓN GENERAL DE CALIDAD Y EVALUACIÓN AMBIENTAL Y MEDIO NATURAL
Dirección General de Calid
y Evaluación Ambiental y
Medio Natural
0
45
90
180
270
360
Kilomete

Proposal Outline

- Cross Industry Standard Process for Data Mining (CRISP-DM) Framework
- **Tracking** the Vespa Velutina:
- Predictive model trained to predict # of Asian Hornets in each municipality.



Eco Analytics

Table of Contents

01

Merging Data

02

Insights

03

Pre-processing

04

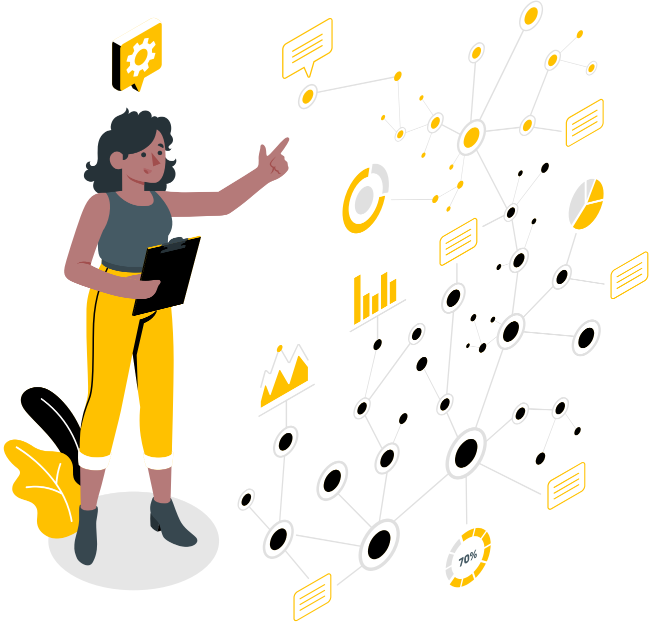
Model

05

Results

Merging Data

01

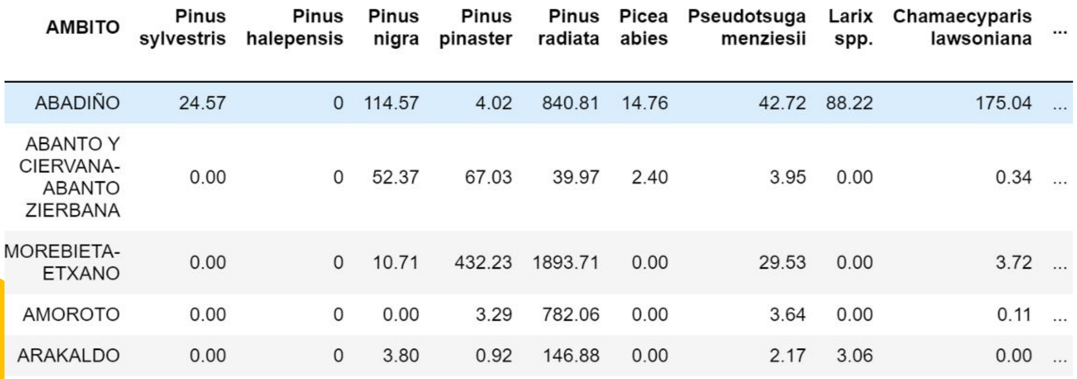


Forest Dataset

Variables

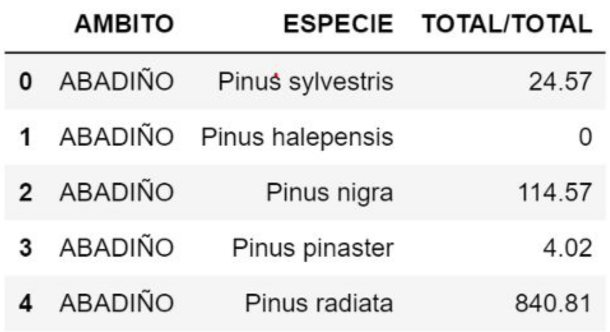
AMBITO = AREA
ESPECIE = TREE SPECIES
TOTAL = TREE COUNT

Modified



AMBITO	Pinus sylvestris	Pinus halepensis	Pinus nigra	Pinus pinaster	Pinus radiata	Picea abies	Pseudotsuga menziesii	Larix spp.	Chamaecyparis lawsoniana	...
ABADIÑO	24.57	0	114.57	4.02	840.81	14.76	42.72	88.22	175.04	...
ABANTO Y CIERVANA-ABANTO ZIERBANA	0.00	0	52.37	67.03	39.97	2.40	3.95	0.00	0.34	...
MOREBIETA-ETXANO	0.00	0	10.71	432.23	1893.71	0.00	29.53	0.00	3.72	...
AMOROTO	0.00	0	0.00	3.29	782.06	0.00	3.64	0.00	0.11	...
ARAKALDO	0.00	0	3.80	0.92	146.88	0.00	2.17	3.06	0.00	...

Original



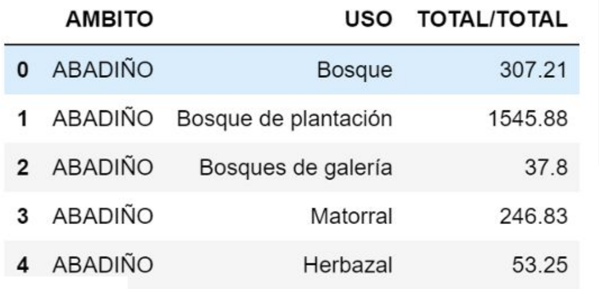
	AMBITO	ESPECIE	TOTAL/TOTAL
0	ABADIÑO	Pinus sylvestris	24.57
1	ABADIÑO	Pinus halepensis	0
2	ABADIÑO	Pinus nigra	114.57
3	ABADIÑO	Pinus pinaster	4.02
4	ABADIÑO	Pinus radiata	840.81

Terrain Dataset

Variables

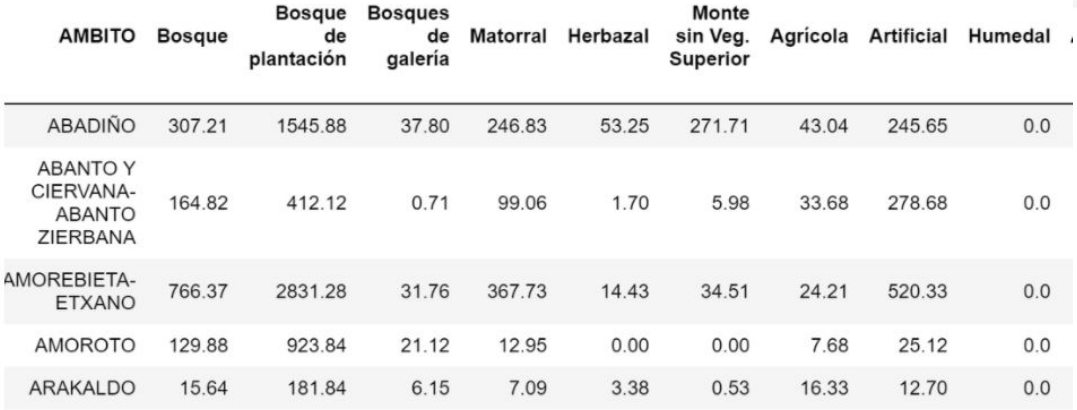
AMBITO = AREA
USO = USES INSIDE THE AREA
TOTAL = COUNT OF EACH USE

Original



	AMBITO	USO	TOTAL/TOTAL
0	ABADIÑO	Bosque	307.21
1	ABADIÑO	Bosque de plantación	1545.88
2	ABADIÑO	Bosques de galería	37.8
3	ABADIÑO	Matorral	246.83
4	ABADIÑO	Herbazal	53.25

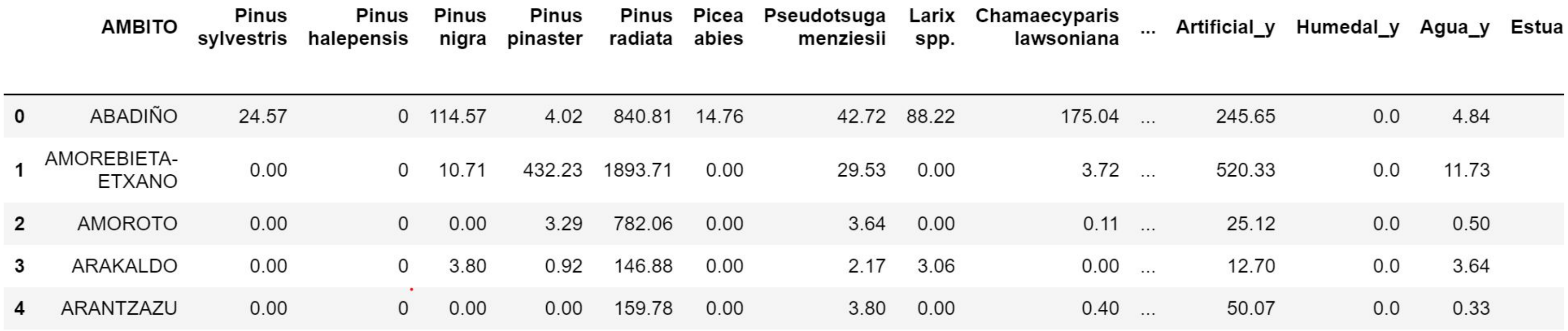
Modified



AMBITO	Bosque	Bosque de plantación	Bosques de galería	Matorral	Herbazal	Monte sin Veg. Superior	Agrícola	Artificial	Humedal
ABADIÑO	307.21	1545.88	37.80	246.83	53.25	271.71	43.04	245.65	0.0
ABANTO Y CIERVANA-ABANTO ZIERBANA	164.82	412.12	0.71	99.06	1.70	5.98	33.68	278.68	0.0
AMOREBIETA-ETXANO	766.37	2831.28	31.76	367.73	14.43	34.51	24.21	520.33	0.0
AMOROTO	129.88	923.84	21.12	12.95	0.00	0.00	7.68	25.12	0.0
ARAKALDO	15.64	181.84	6.15	7.09	3.38	0.53	16.33	12.70	0.0

Forest & Terrain

109 rows, 97 columns



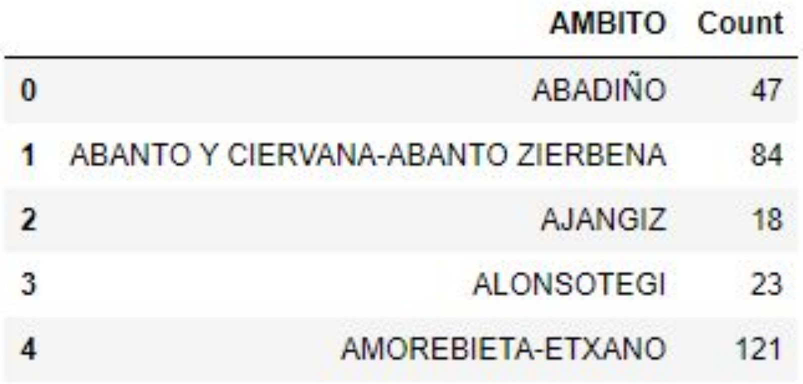
	AMBITO	Pinus sylvestris	Pinus halepensis	Pinus nigra	Pinus pinaster	Pinus radiata	Picea abies	Pseudotsuga menziesii	Larix spp.	Chamaecyparis lawsoniana	...	Artificial_y	Humedal_y	Agua_y	Estua
0	ABADIÑO	24.57	0	114.57	4.02	840.81	14.76	42.72	88.22	175.04	...	245.65	0.0	4.84	
1	AMOREBIETA-ETXANO	0.00	0	10.71	432.23	1893.71	0.00	29.53	0.00	3.72	...	520.33	0.0	11.73	
2	AMOROTO	0.00	0	0.00	3.29	782.06	0.00	3.64	0.00	0.11	...	25.12	0.0	0.50	
3	ARAKALDO	0.00	0	3.80	0.92	146.88	0.00	2.17	3.06	0.00	...	12.70	0.0	3.64	
4	ARANTZAZU	0.00	0	0.00	0.00	159.78	0.00	3.80	0.00	0.40	...	50.07	0.0	0.33	

Merged on **municipality**

Nest Dataset

Variables

AMBITO = AREA
COUNT = # OF VESPA VELUTINA NESTS

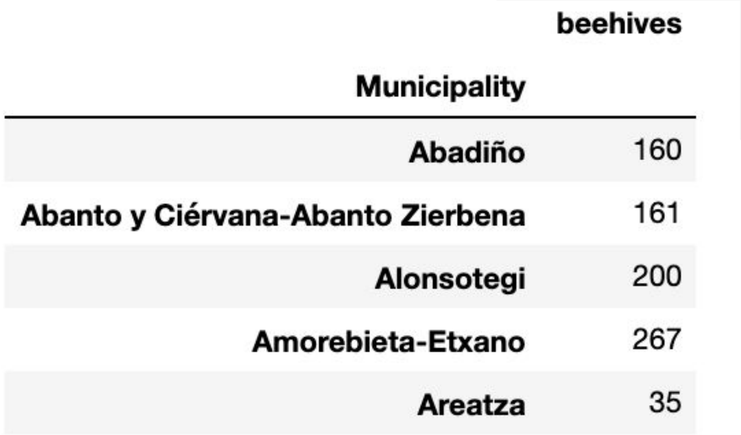


	AMBITO	Count
0	ABADIÑO	47
1	ABANTO Y CIERVANA-ABANTO ZIERBENA	84
2	AJANGIZ	18
3	ALONSOTEGI	23
4	AMOREBIETA-ETXANO	121

- Filtered by Vespa Velutina
- Creation of Count variable
- Left join onto master table on municipality

Beekeeping Dataset

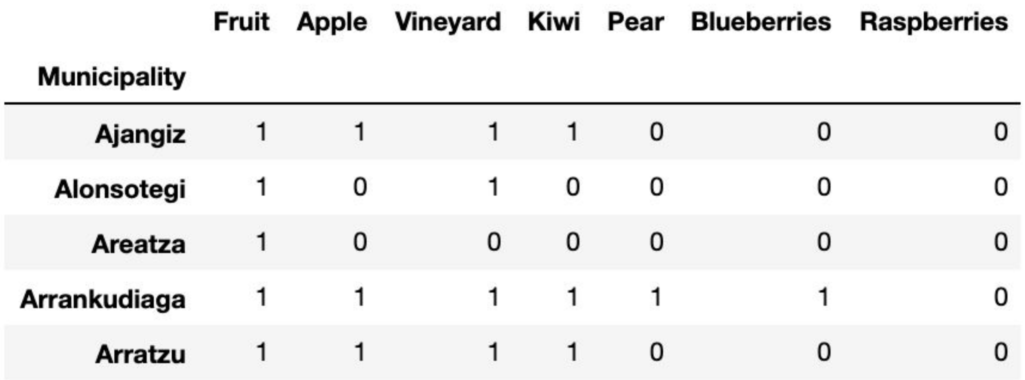
- Grouped number of beehives by municipality (removed other variables in dataset)
- Left join to master table
- Filled in missing values



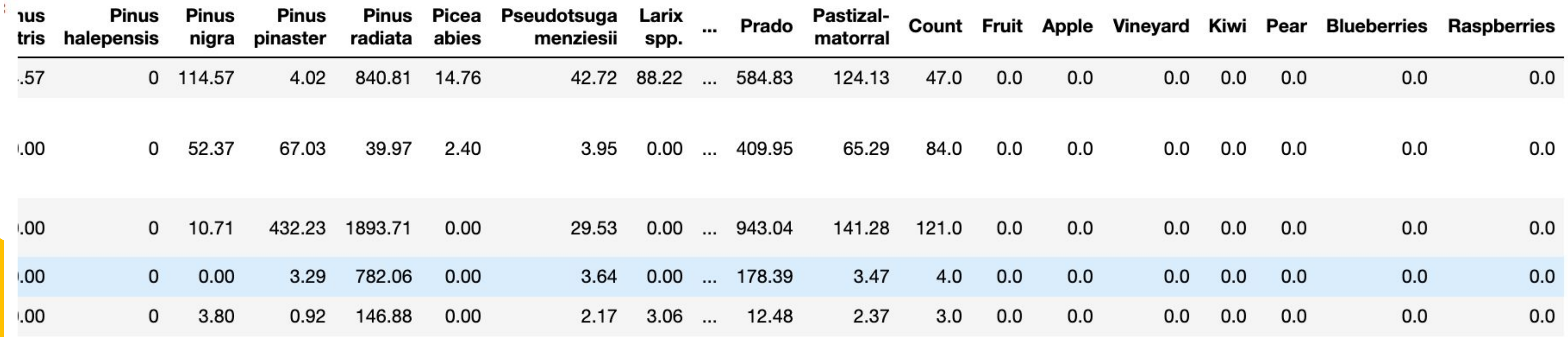
	beehives
Municipality	
Abadiño	160
Abanto y Ciérvana-Abanto Zierbena	161
Alonsotegi	200
Amorebieta-Etxano	267
Areatza	35

Fruit Trees Dataset

- Duplicate value cleaning
- Structured dataset by grouping it by municipalities



Municipality	Fruit	Apple	Vineyard	Kiwi	Pear	Blueberries	Raspberries
Ajangiz	1	1	1	1	0	0	0
Alonsotegi	1	0	1	0	0	0	0
Areatza	1	0	0	0	0	0	0
Arrankudiaga	1	1	1	1	1	1	0
Arratzu	1	1	1	1	0	0	0



ius tris	Pinus halepensis	Pinus nigra	Pinus pinaster	Pinus radiata	Picea abies	Pseudotsuga menziesii	Larix spp.	...	Prado	Pastizal-matorral	Count	Fruit	Apple	Vineyard	Kiwi	Pear	Blueberries	Raspberries
.57	0	114.57	4.02	840.81	14.76	42.72	88.22	...	584.83	124.13	47.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.00	0	52.37	67.03	39.97	2.40	3.95	0.00	...	409.95	65.29	84.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.00	0	10.71	432.23	1893.71	0.00	29.53	0.00	...	943.04	141.28	121.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.00	0	0.00	3.29	782.06	0.00	3.64	0.00	...	178.39	3.47	4.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
.00	0	3.80	0.92	146.88	0.00	2.17	3.06	...	12.48	2.37	3.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

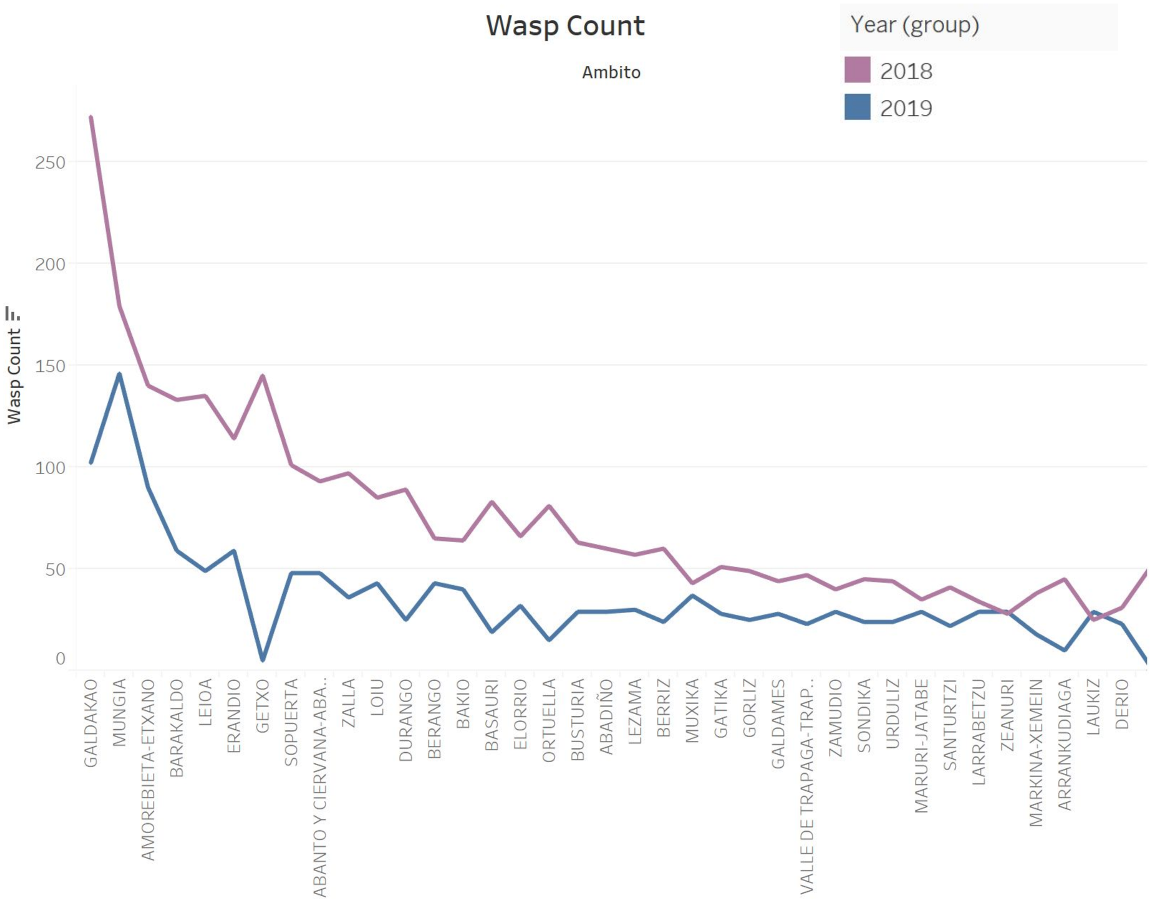
Weather Dataset

- **Precipitation**, **Temperature**, **Wind** data
- 18 tables across 2 years
- Reduced to annual values
- Merged on nearest Weather Station
- Two records per Municipality: 2018, 2019

Insights

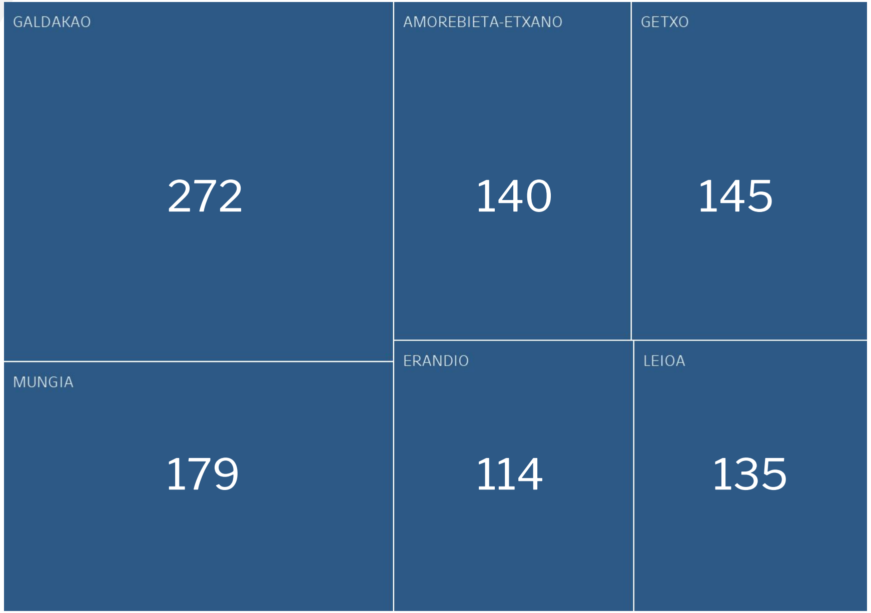
02

Wasp Nests



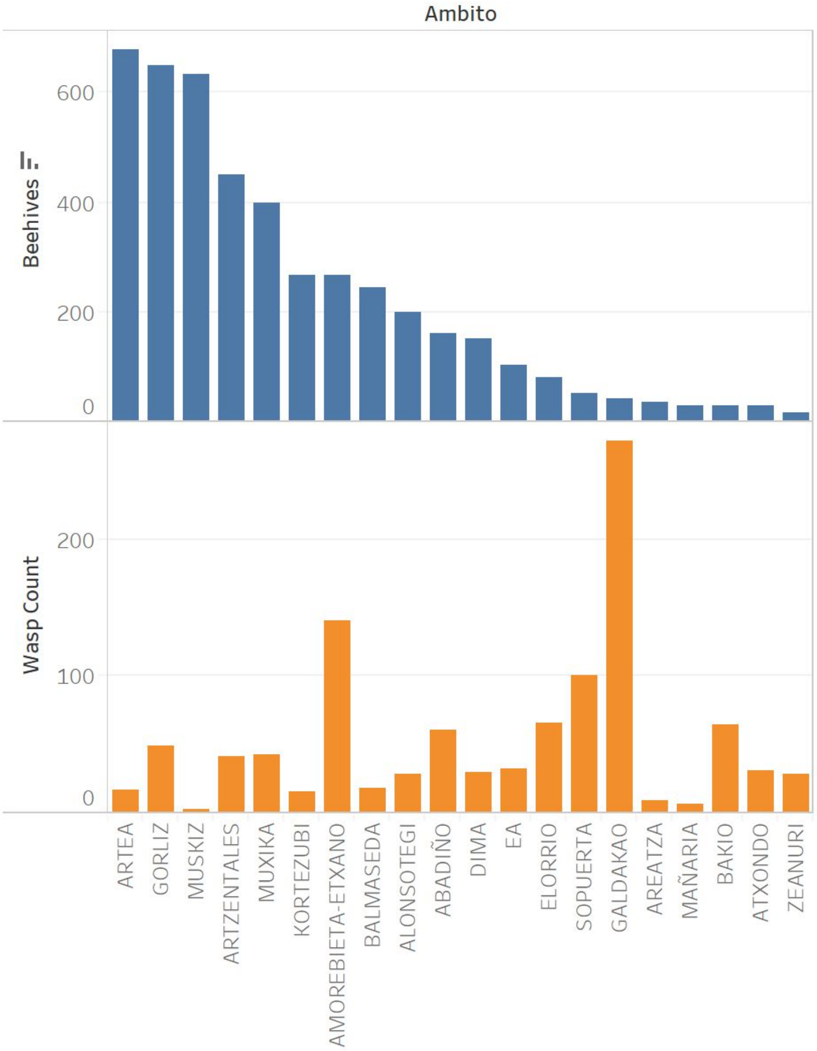
Wasp Count
Ambito
Year (group)
2018
2019
Wasp Count
250
200
150
100
50
0
GALDAKAO
MUNGIA
AMOREBIETA-ETXANO
BARAKALDO
LEIOA
ERANDIO
GETXO
SOPUERTA
ABANTO Y CIERVANA-ABA..
ZALLA
LOIU
DURANGO
BERANGO
BAKIO
BASAURI
ELORRIO
ORTUELLA
BUSTURIA
ABADIÑO
LEZAMA
BERRIZ
MUXIKA
GATIKA
GORLIZ
GALDAMES
VALLE DE TRAPAGA-TRAP..
ZAMUDIO
SONDIKA
URDULIZ
MARURI-JATABE
SANTURTZI
LARRABETZU
ZEANURI
MARKINA-XEMEIN
ARRANKUDIAGA
LAUKIZ
DERIO

TOP LOCATIONS



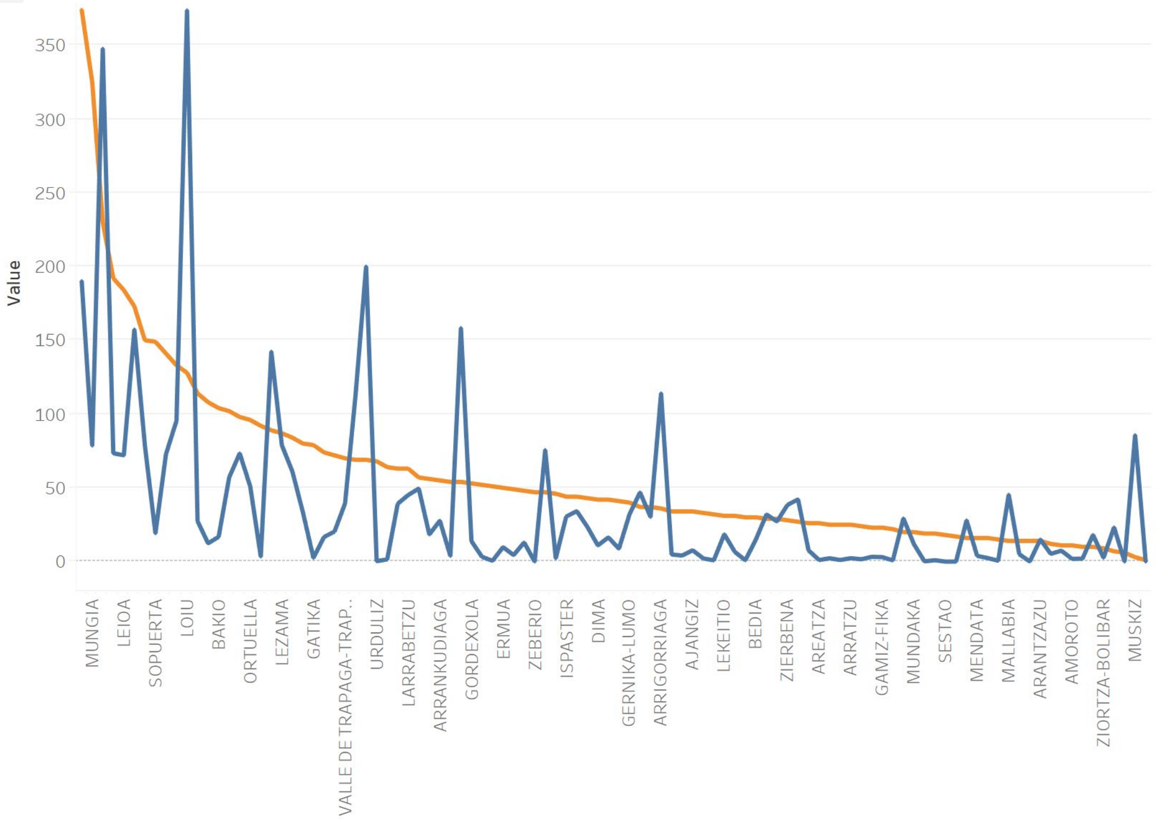
1. **GALDAKAO**
2. **MUNGIA**
3. **AMOREBIETA-ETXANO**

Beehives



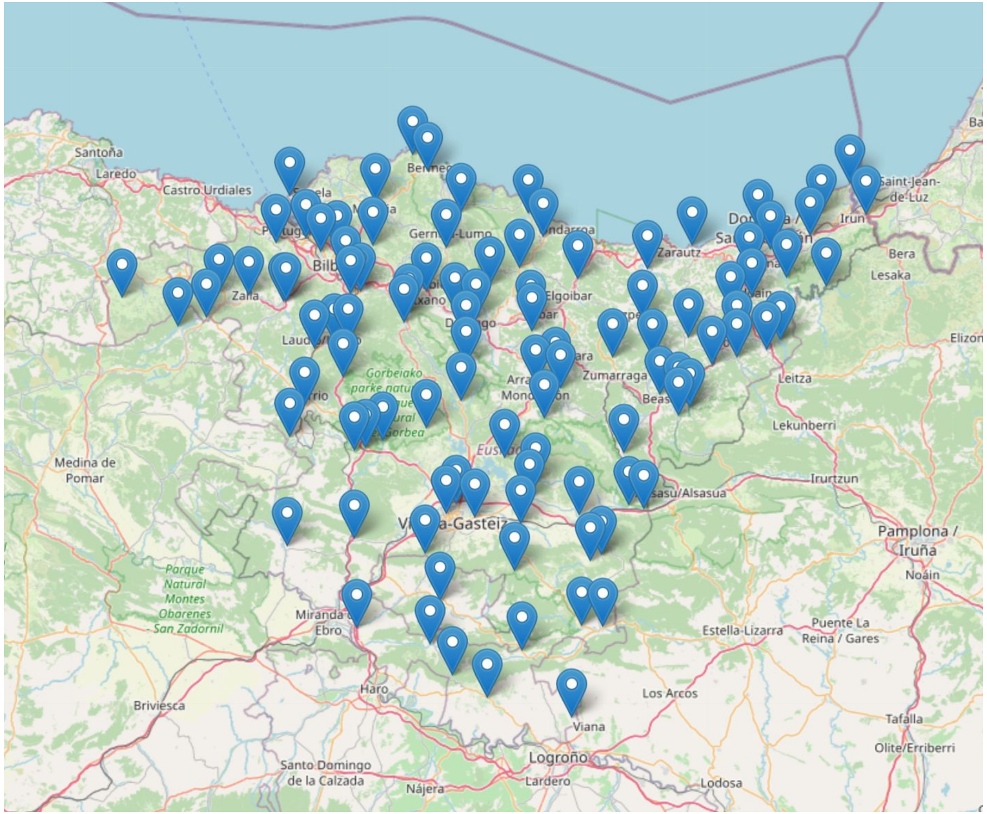
- 80%(88) municipality don’t have beehives

Terrain

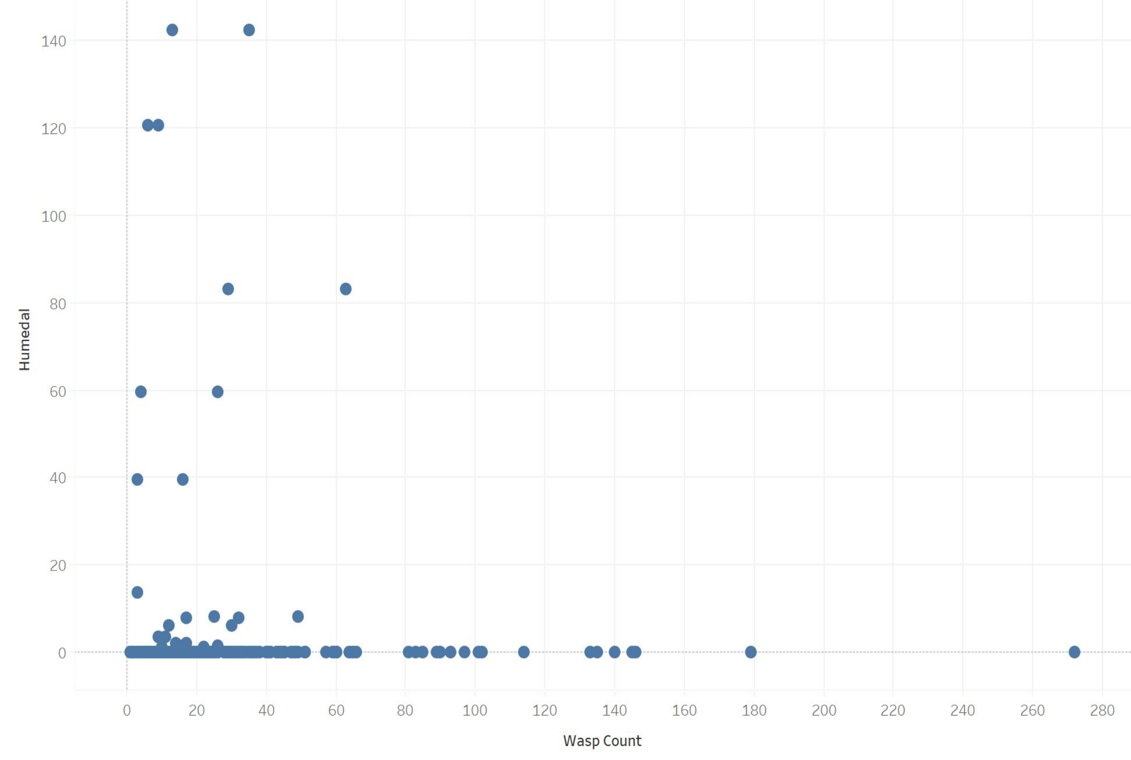


- Negative relationship: Freeways - Asian Hornet nests

WEATHER STATIONS

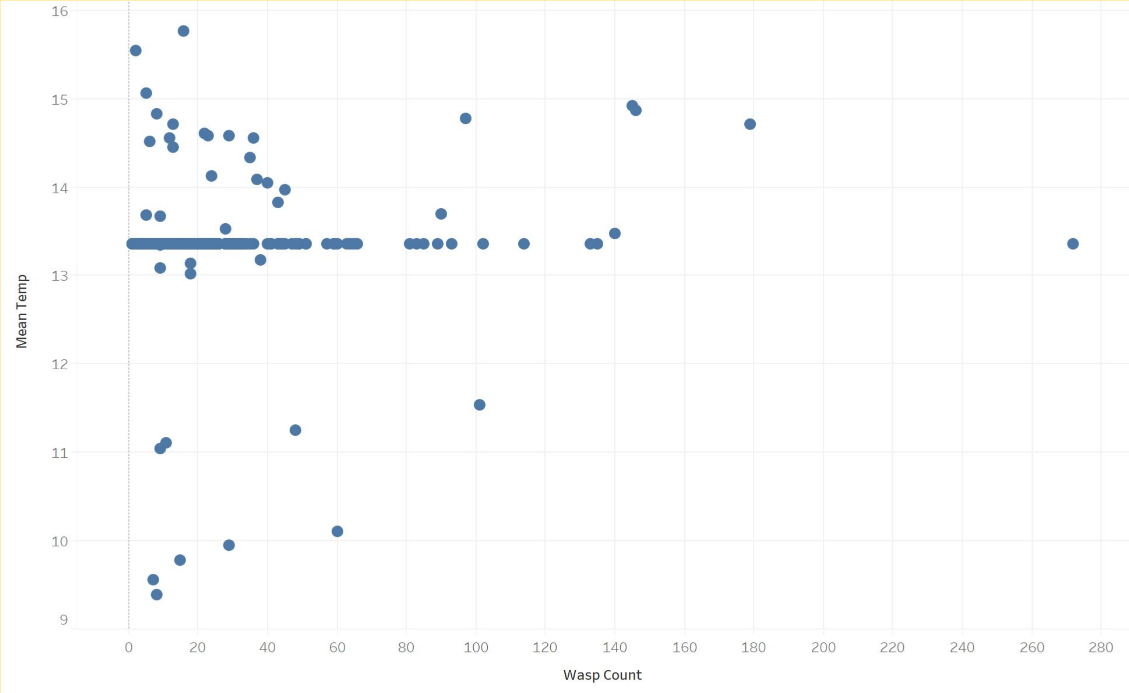


Humidity



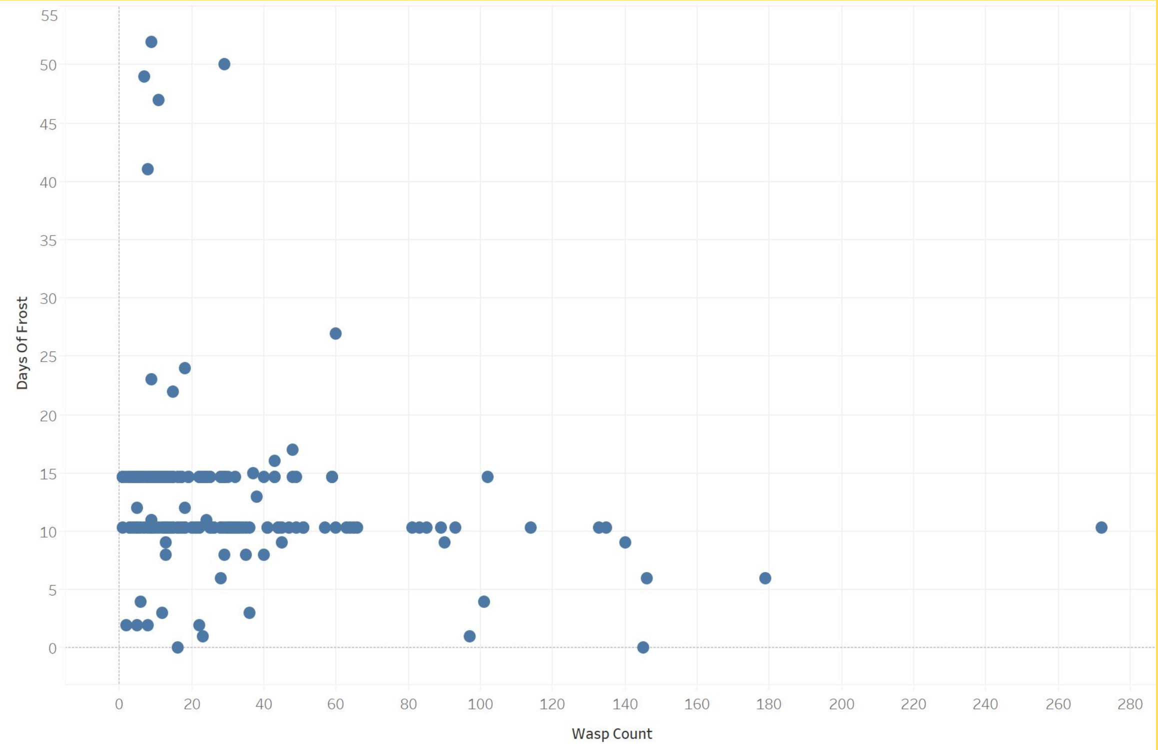
- Average humidity = 0.0
- It appears wasps do not like humid areas

Mean Temperature



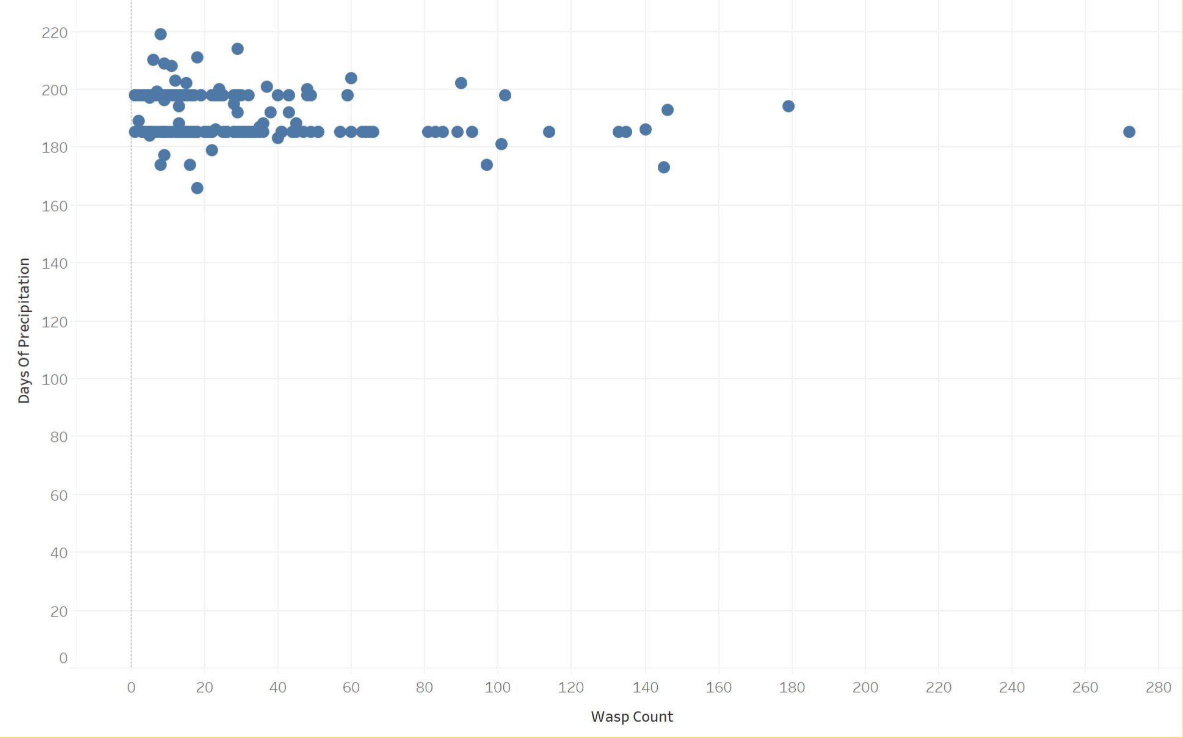
- Most wasps are located in places with an average temperature of 13 to 16 degrees

Days of Frost



- Most wasps are located in places with the least days of frost

Days of Precipitation



- It seems wasps tend to live in places with the most days of precipitation

Pre-processing

03

Missing Values & Sparse Variables

Missing Values

Variables > 60% missing values were dropped

Variables <60% missing values were imputed with mean values

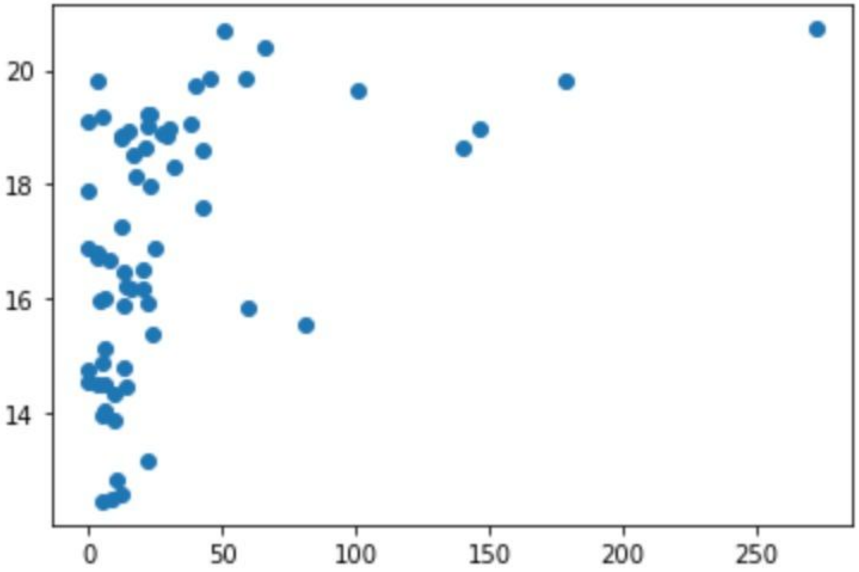
Variables dropped: **Populus nigra, Pinus halepensis, Raspberries**

Sparse Variables

Variables > 85% of data = 0 were dropped

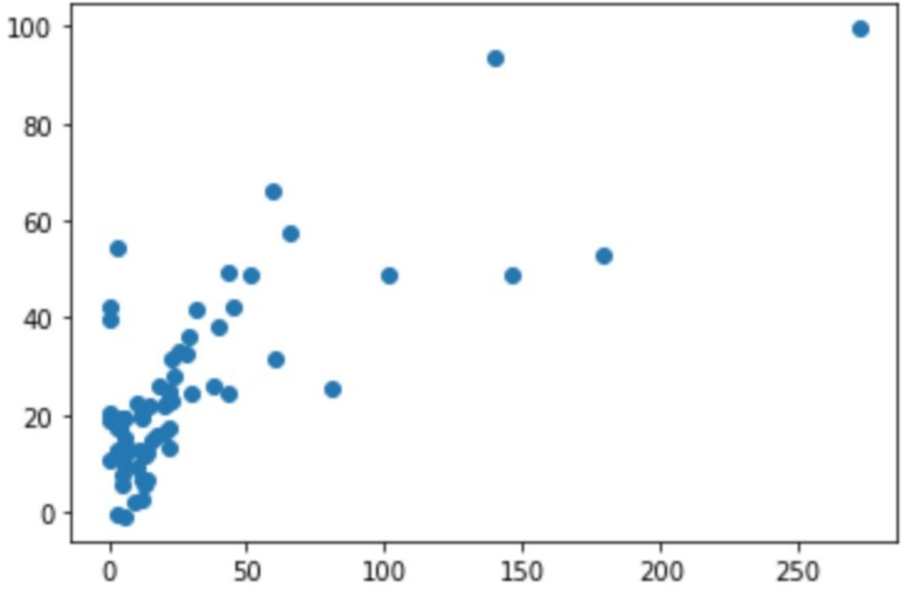
Variables dropped: **Quercus petraea, Bosque mixto de cantil, Humedal, Pear, Blueberries**

Support Vector Machine



- Min max scaler
- 20-fold Cross Validation
- MSE: 2175.5711

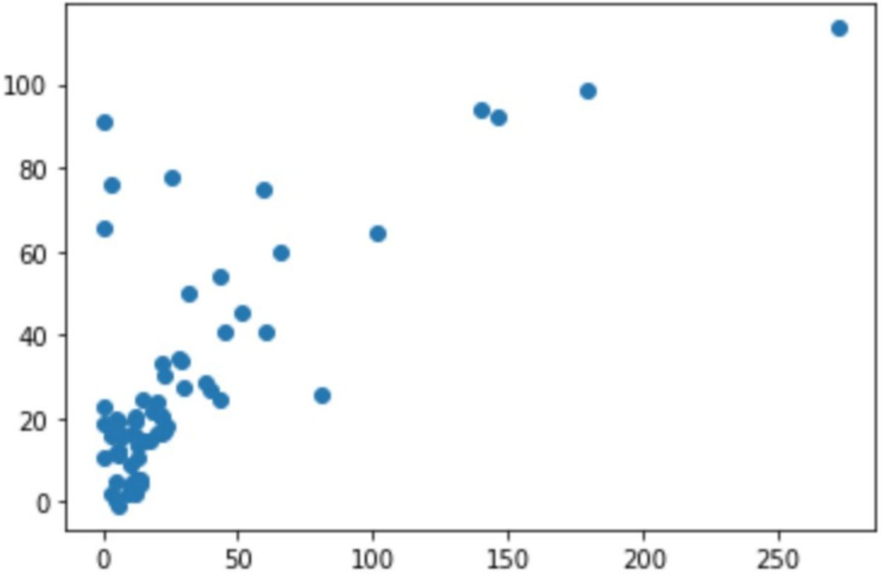
Support Vector Machine



kernel='rbf', C=100, gamma=0.1, epsilon=.1

MSE: 1115.167260334689

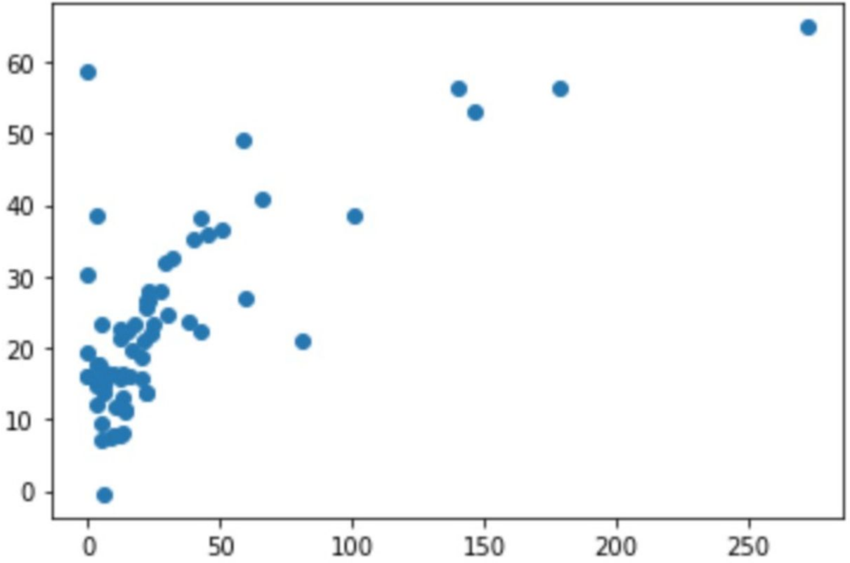
Support Vector Machine



kernel='linear', C=100, gamma='auto'

MSE: 1010.9830722878683

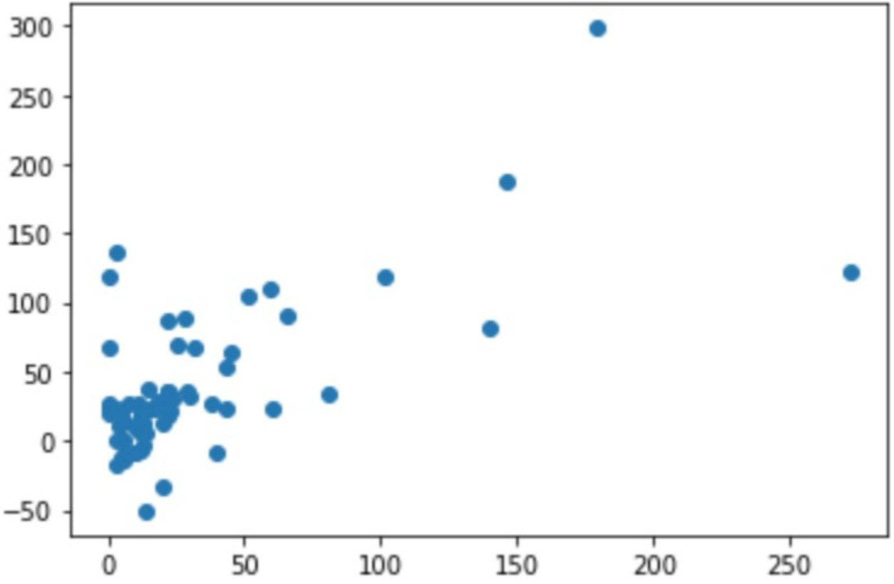
Support Vector Machine



kernel='poly', C=100, gamma='auto', degree=3, epsilon=.1, coef0=1

MSE: 1402.12189650883

Linear Models

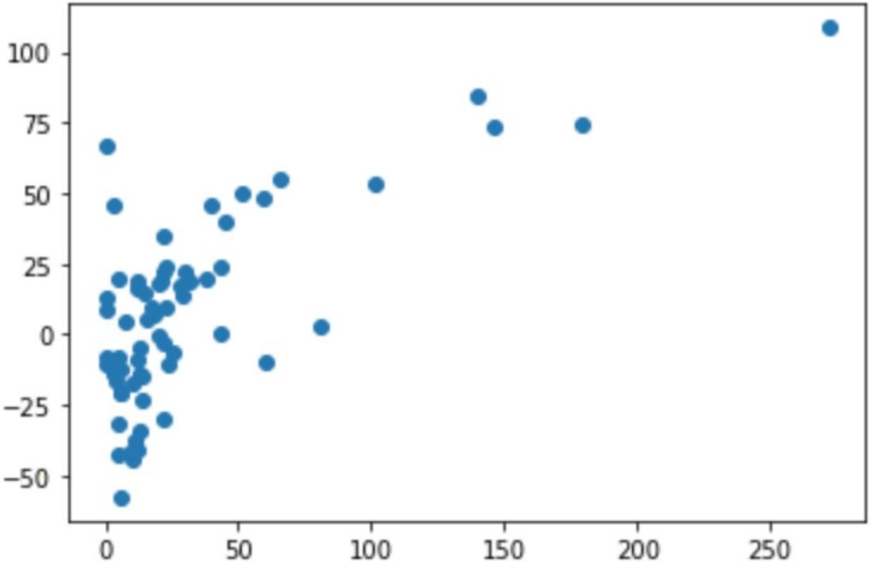


Standard scaler

20 fold cross validation

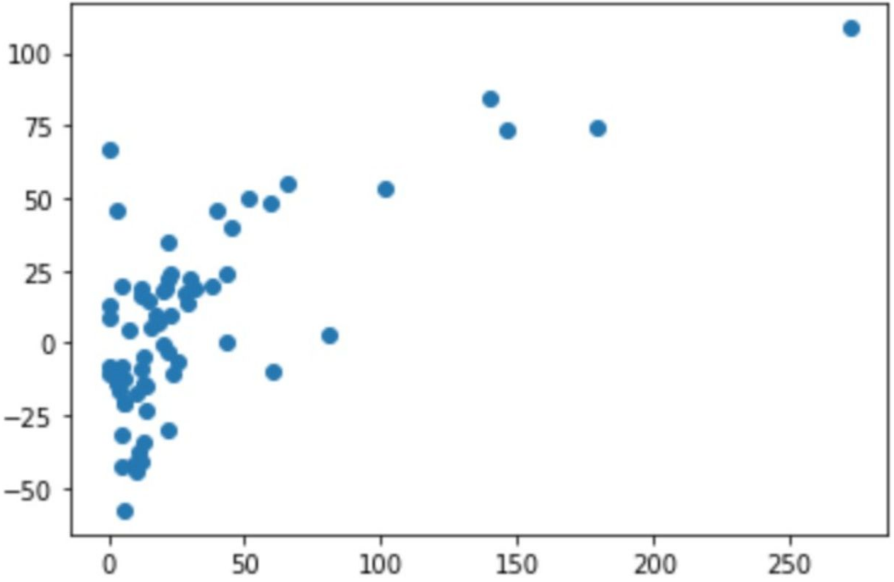
MSE: 1779.647293677089

Ridge Regression



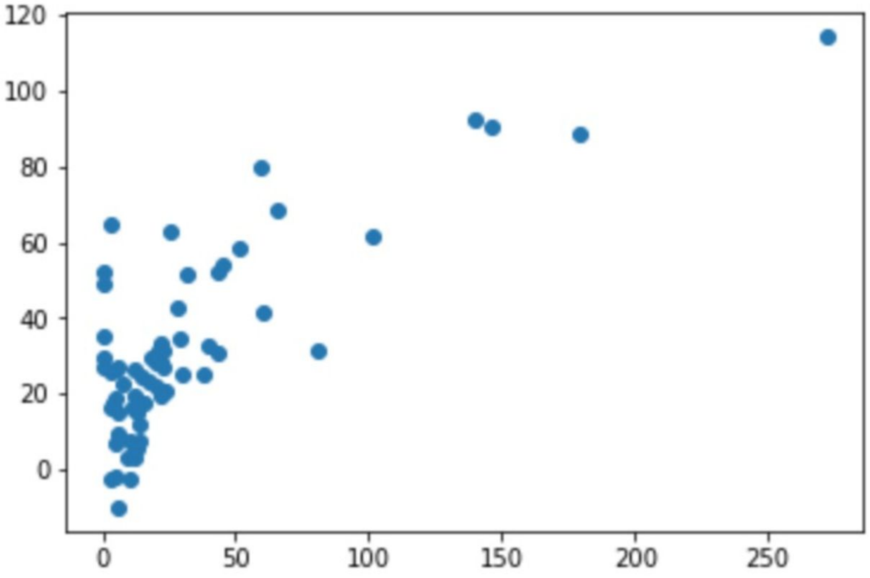
MSE: 1576.2136762604107

Lasso Regression



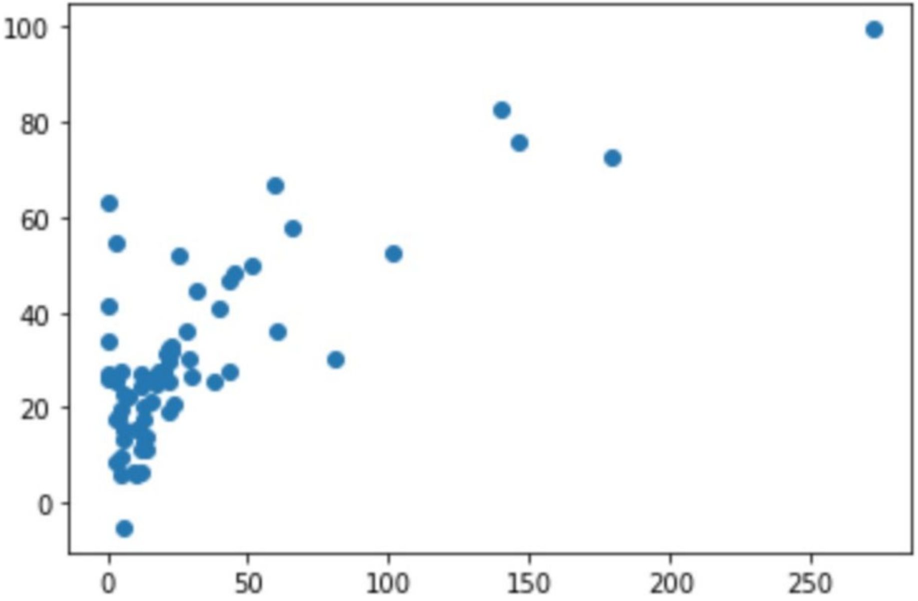
MSE: 1576.2136762604107

Bayesian Ridge



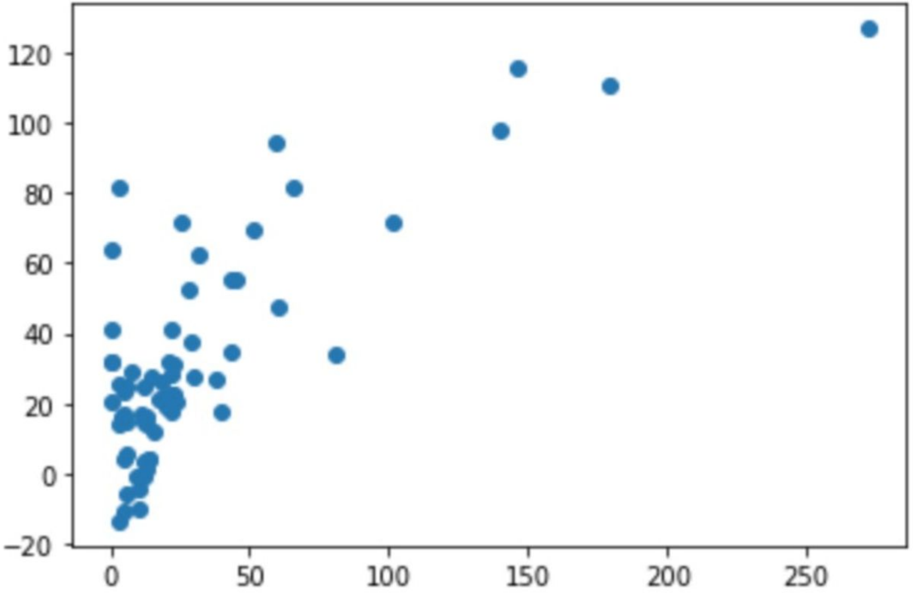
MSE: 934.0755392516393

ElasticNet



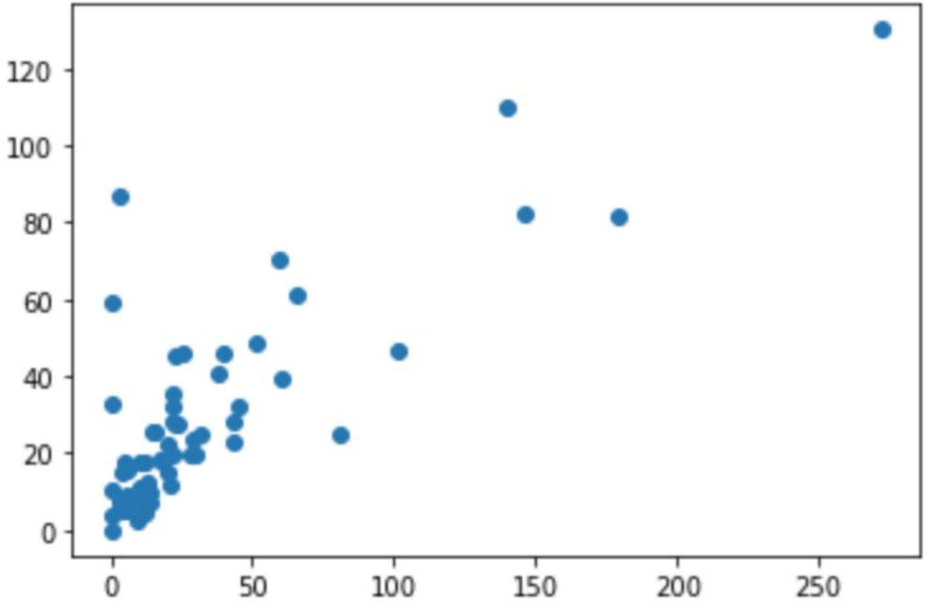
MSE: 1079.2563121328774

SGD Regressor



Loss = 'squared_loss'
MSE: 876.6228785209817

XGB Regressor

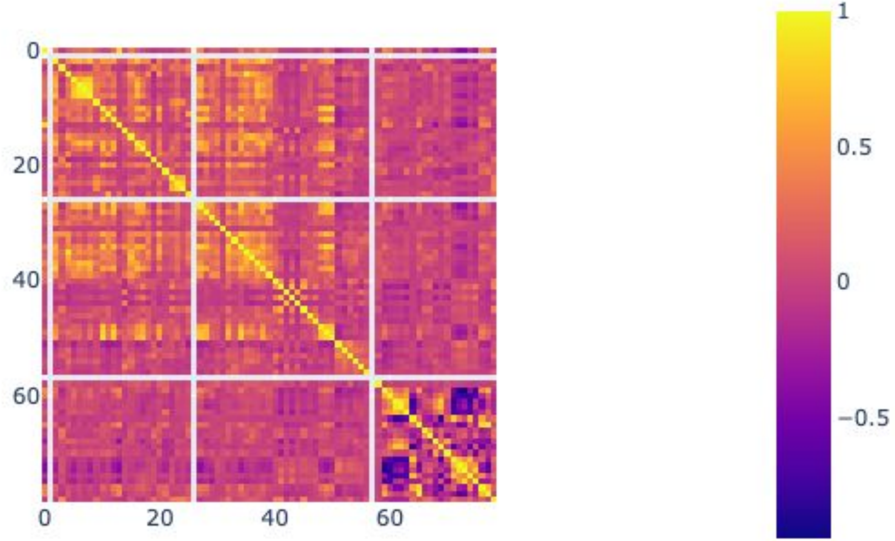


Standard scaler

MSE: 855.9152608451054

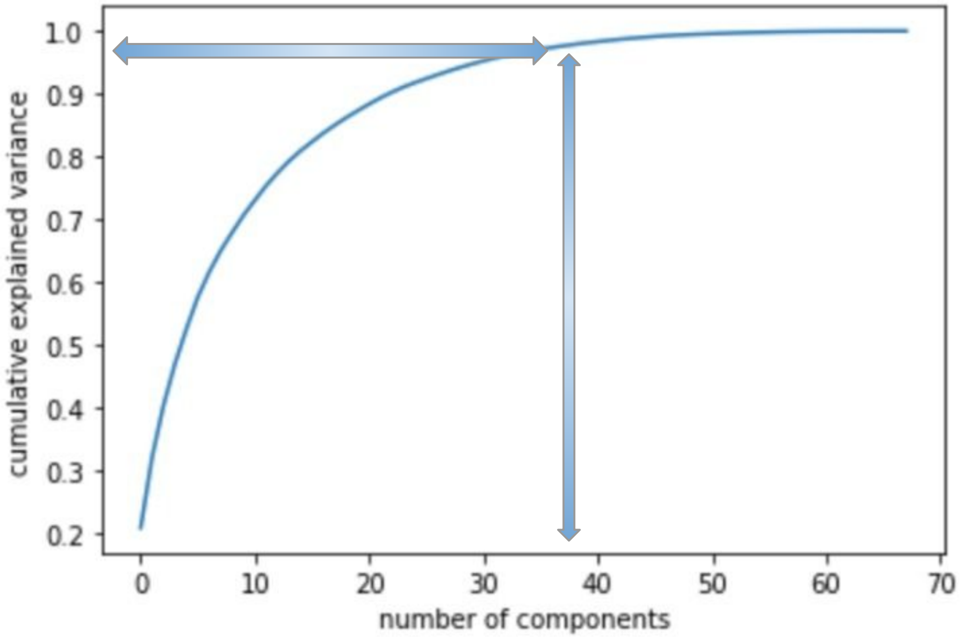
Relationship Between Variables

- 69 predictor variables
- Multicollinearity present:
- **Weather** features
- **Forest** features
- Principal Component Analysis (PCA) could aid this

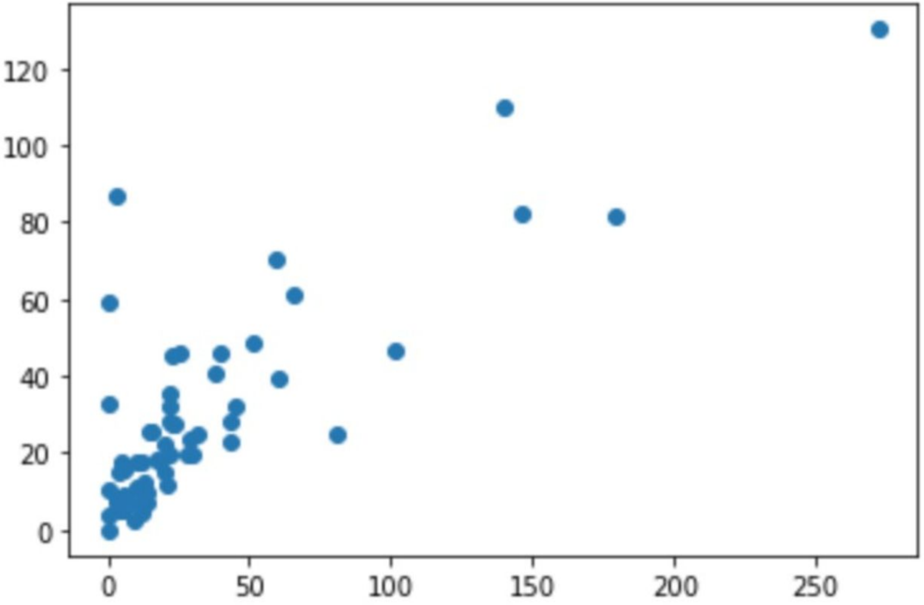


Principal Component Analysis (PCA)

- 38 Components → explains over 95% of variance in the dataset
- Attempt to model using these components later



XGB Regressor

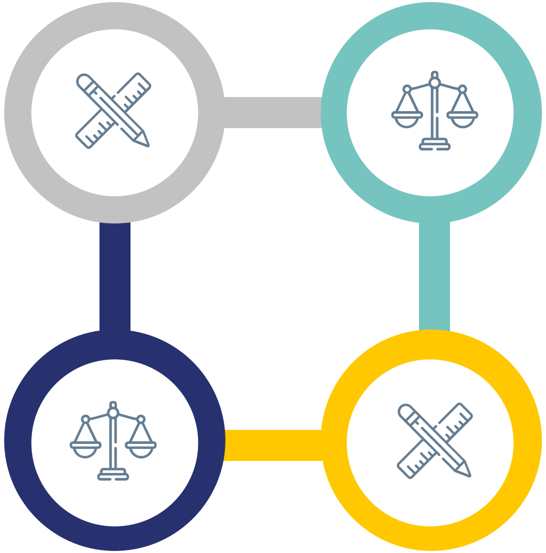


Standard scaler

PCA (38 components)

MSE: 518.740522668572971

Scaling the Data



Min Max Scaler

Standard Scaler

MaxAbs Scaler

Robust Scaler

Splitting the data

70-30 Split

Initial Approach

20-Fold Cross Validation

Eases training

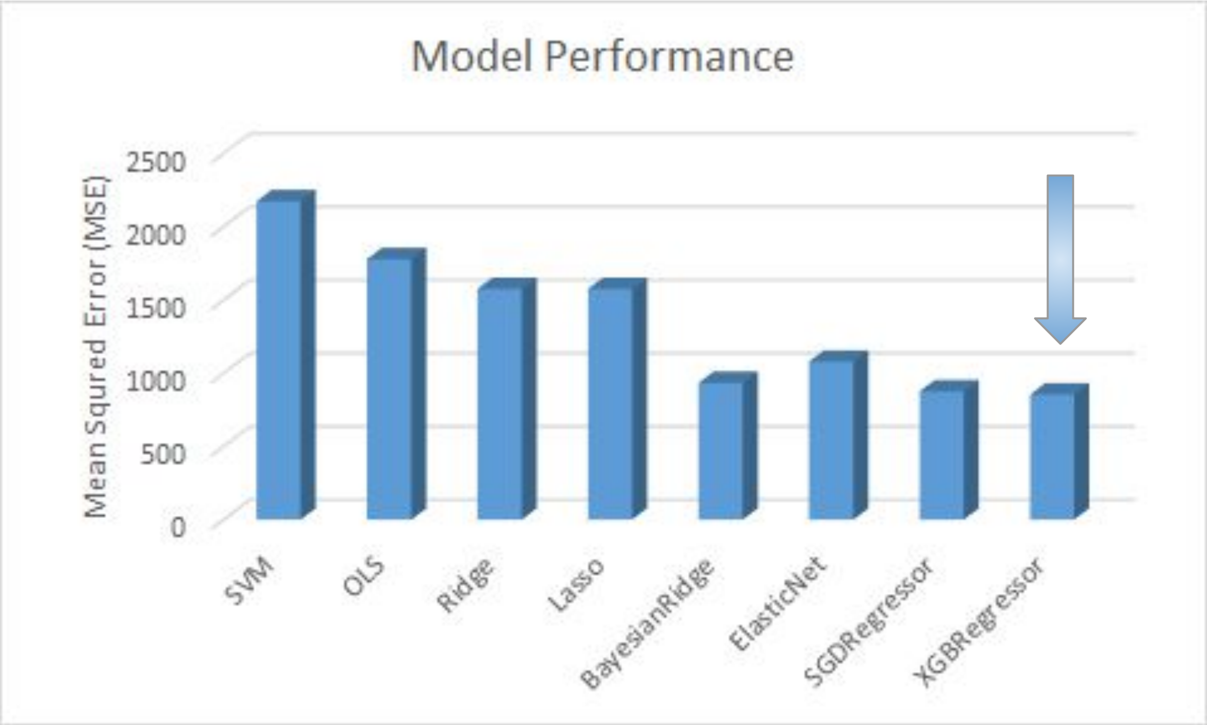
Split by year

Training set: 2017, 2018
Test set: 2019

Tuning the Model

04

Preliminary Results



Tuning Hyper-Parameters of XGBoost

Via Grid Search

Max Depth

Learning rate

Type of Booster

Colsample by Tree

Min Child Weight

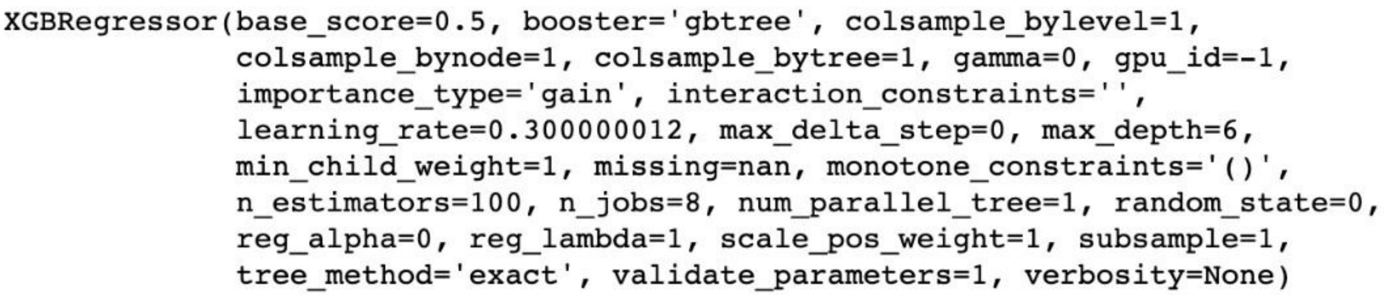
Gamma

TUNED RESULTS

XGBoost Regression
MSE = 384.16

Final Model Details

Standard Scaler, 38 PCA,
hyper-parameter tuning



```
XGBRegressor(base_score=0.5, booster='gbtree', colsample_bylevel=1,
             colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
             importance_type='gain', interaction_constraints='',
             learning_rate=0.300000012, max_delta_step=0, max_depth=6,
             min_child_weight=1, missing=nan, monotone_constraints='()',
             n_estimators=100, n_jobs=8, num_parallel_tree=1, random_state=0,
             reg_alpha=0, reg_lambda=1, scale_pos_weight=1, subsample=1,
             tree_method='exact', validate_parameters=1, verbosity=None)
```

Results

05

#6
MSE = 645.8112

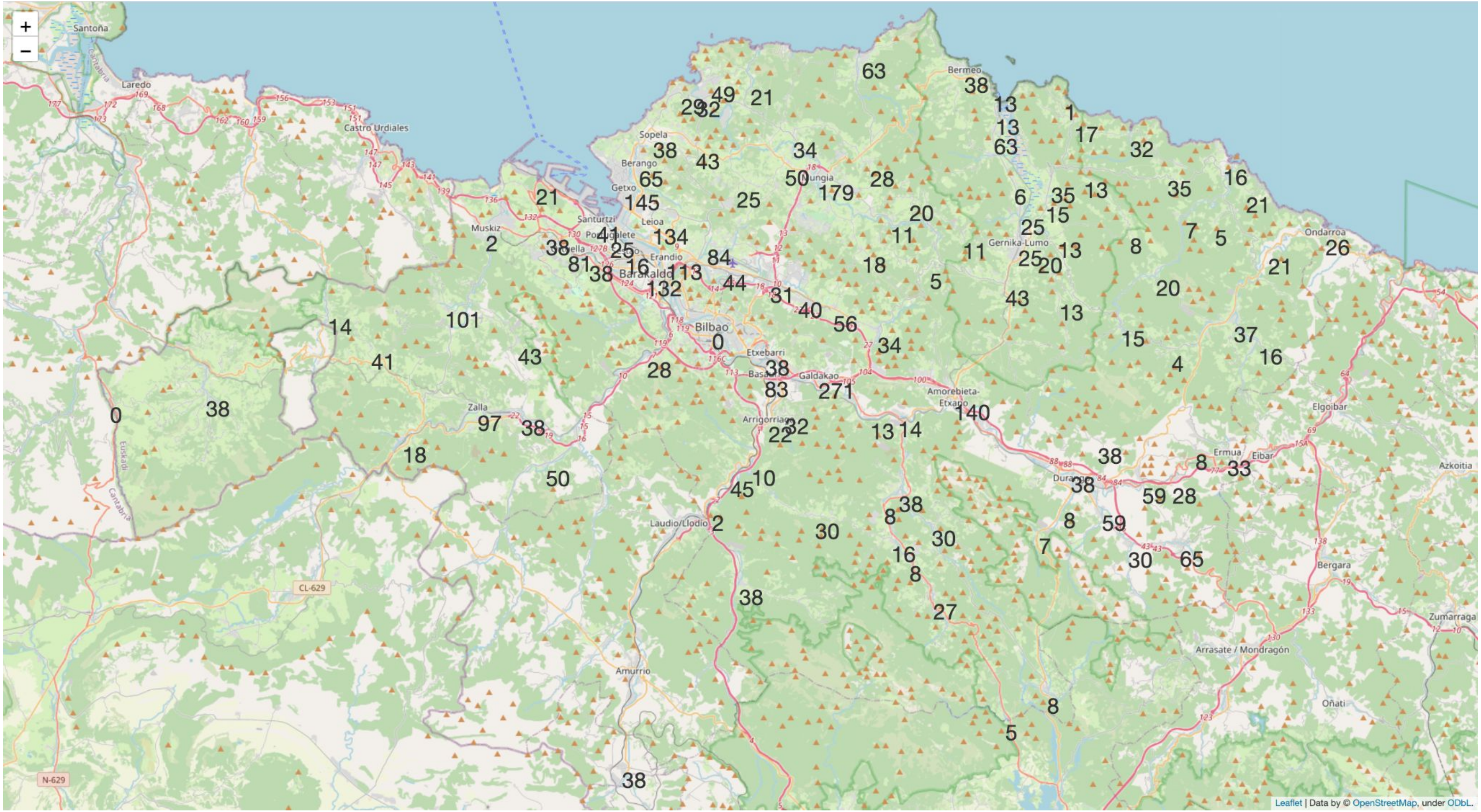
Visualization

Geopy

Geopy converts municipality name into latitude and longitude

Folium

Folium plots latitude and longitude on the map


+
−
Santoña
Laredo
Castro Urdiales
Muskiz
Santurtzi
Portugalete
Sopela
Berango
Getxo
Leioa
Erandio
Barakaldo
Bilbao
Etxebarri
Basauri
Galdakao
Arrigorriaga
Mungia
Bermeo
Gernika-Lumo
Amorebieta-Etxano
Durango
Ermua
Eibar
Elgoibar
Ondarroa
Azkoitia
Bergara
Zumarraga
Arrasate / Mondragón
Oñati
Zalla
Laudio/Llodio
Amurrio
CL-629
N-629
Leaflet | Data by © OpenStreetMap, under ODbL.

42

75th percentile

30

50th percentile

15

25th percentile

272

max

0

min

38

mode

MUNICIPIO Most Wasp Nest

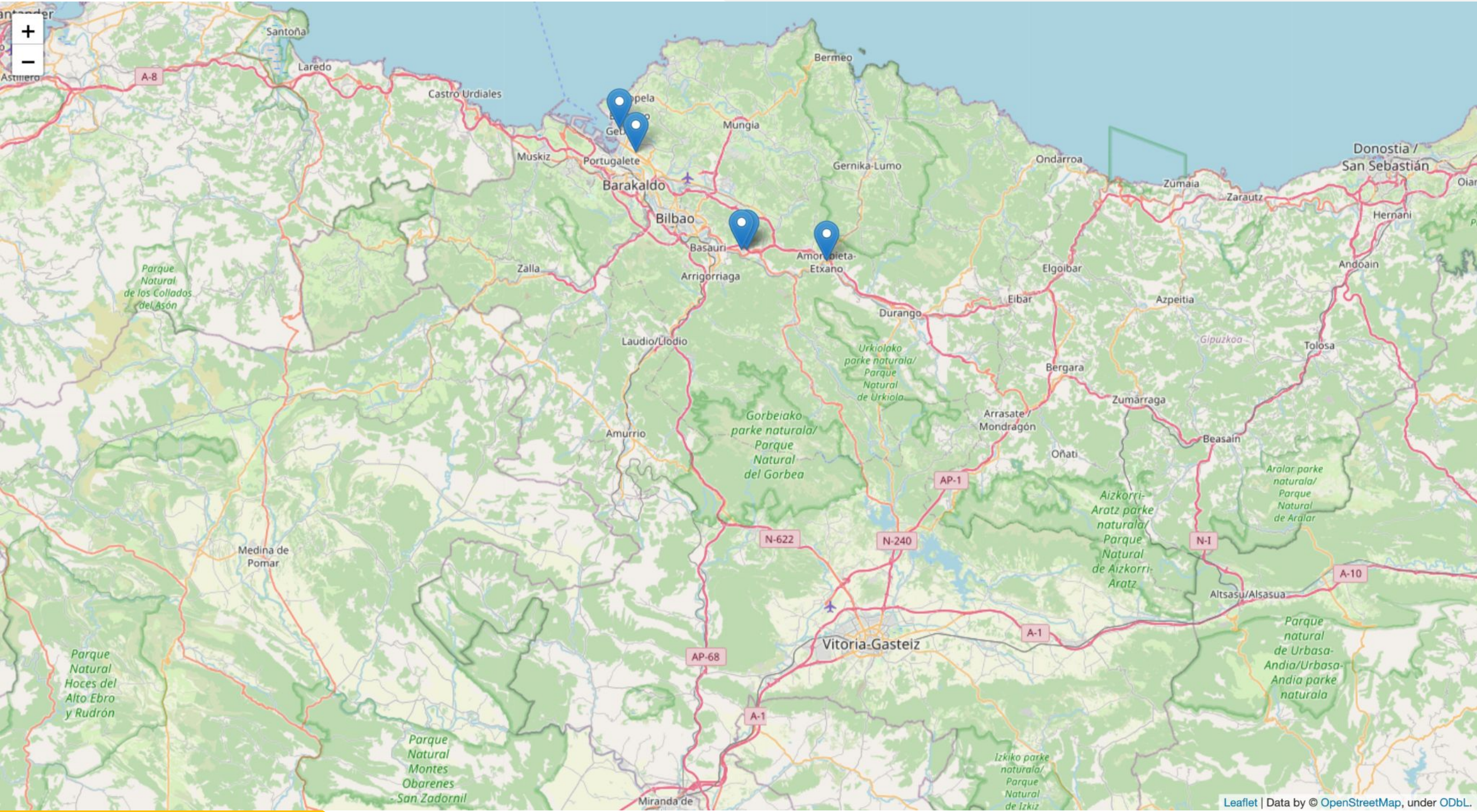
Galdakao: 272

Mungia: 179

Getxo: 145

Amorebieta-Etxano: 140

Leioa: 134



+
−
Santoña
Laredo
Astillero
A-8
Castro Urdiales
Bermeo
Mungia
Gernika-Lumo
Ondarroa
Donostia / San Sebastián
Zumaia
Zarautz
Hernani
Muskiz
Portugalete
Barakaldo
Bilbao
Basauri
Arrigorriaga
Amorebieta-Etxano
Zalla
Elgoibar
Eibar
Azpeitia
Andoain
Durango
Laudio/Llodio
Urkiolako parke naturala/ Parque Natural de Urkiola
Gipuzkoa
Tolosa
Bergara
Zumarraga
Arrasate / Mondragón
Oñati
Beasain
Gorbeiako parke naturala/ Parque Natural del Gorbea
Amurrio
Parque Natural de los Collados del Asón
AP-1
Aralar parke naturala/ Parque Natural de Aralar
Aizkorri-Aratz parke naturala/ Parque Natural de Aizkorri-Aratz
N-622
N-240
N-I
A-10
Medina de Pomar
Altsasu/Alsasua
Vitoria-Gasteiz
A-1
AP-68
Parque natural de Urbasa-Andia/Urbasa-Andia parke naturala
Parque Natural Hoces del Alto Ebro y Rudrón
Parque Natural Montes Obarenes - San Zadornil
Izkiko parke naturala/ Parque Natural de Izki
Miranda de

Leaflet | Data by © OpenStreetMap, under ODbL.

Regions With Least Nests

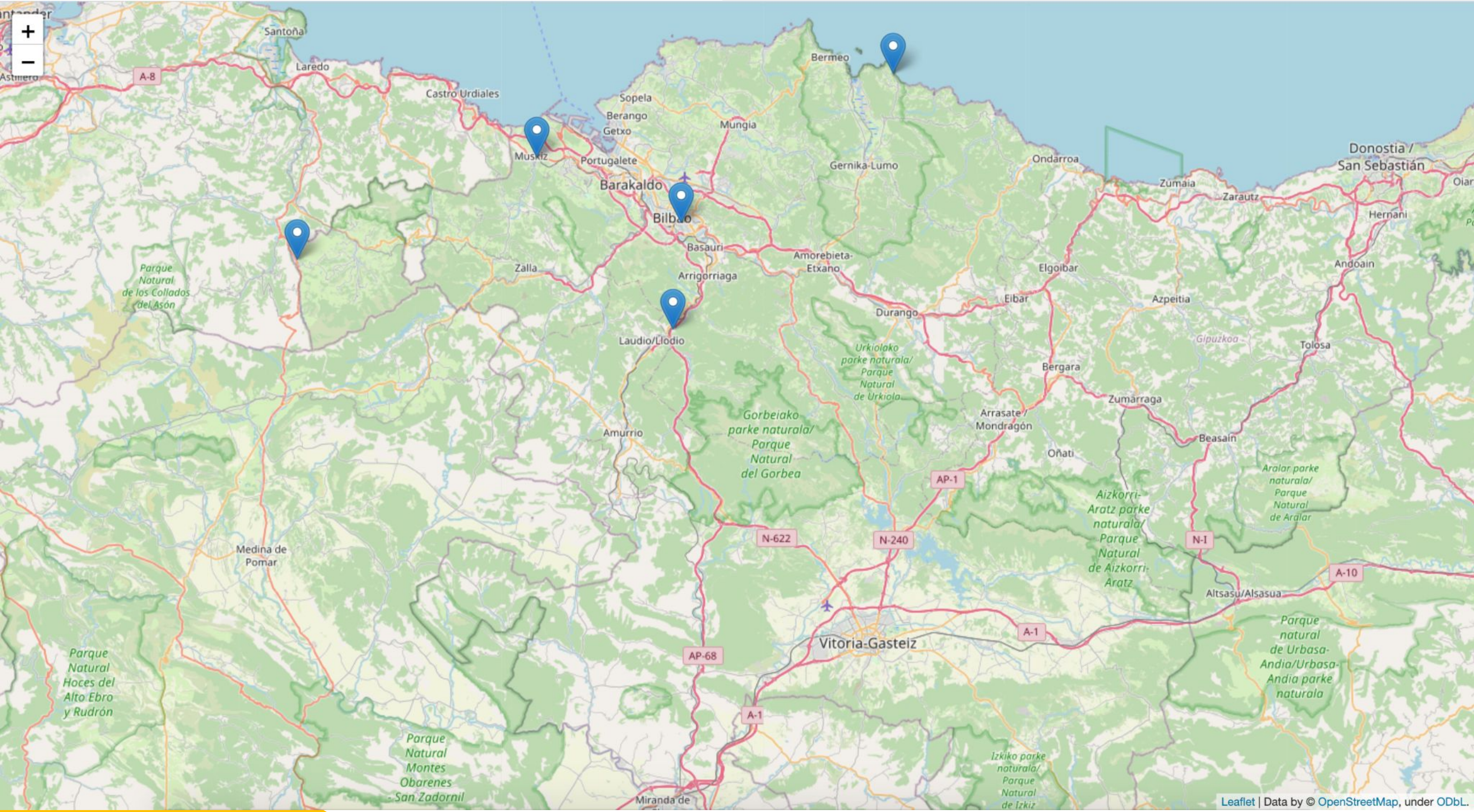
Bilbao: 0

Lanestosa: 0

Elantxobe: 1

Arakaldo: 3

Muskiz: 3


+
−
Santander
Santoña
Laredo
Astillero
A-8
Castro Urdiales
Sopela
Berango
Getxo
Mungia
Bermeo
Gernika-Lumo
Muskiz
Portugalete
Barakaldo
Bilbao
Basauri
Arrigorriaga
Amorebieta-Etxano
Zalla
Laudio/Llodio
Parque Natural de los Collados del Asón
Ondarroa
Zumaia
Zarautz
Donostia / San Sebastián
Oiartzun
Hernani
Andoain
Elgoibar
Eibar
Azpeitia
Durango
Gipuzkoa
Tolosa
Urkiolako parke naturala/ Parque Natural de Urkiola
Bergara
Zumarraga
Arrasate / Mondragón
Oñati
Beasain
Gorbeiako parke naturala/ Parque Natural del Gorbea
Amurrio
AP-1
Aizkorri-Aratz parke naturala/ Parque Natural de Aizkorri-Aratz
Aralar parke naturala/ Parque Natural de Aralar
N-622
N-240
N-I
A-10
Altsasu/Alsasua
Medina de Pomar
Vitoria-Gasteiz
A-1
AP-68
Parque natural de Urbasa-Andia/Urbasa-Andia parke naturala
Parque Natural Hoces del Alto Ebro y Rudrón
Parque Natural Montes Obarenes San Zadornil
Izkiko parke naturala/ Parque Natural de Izki
Miranda de
Leaflet | Data by © OpenStreetMap, under ODbL.

Thanks

Eco Analytics

Do you have
any questions?

kjuhl.ieu2018@student.ie.edu
rhageali.ieu2018@student.ie.edu
qxiang.ieu2018@student.ie.edu
vzaldivar.ieu2018@student.ie.edu