Capstone Project 1
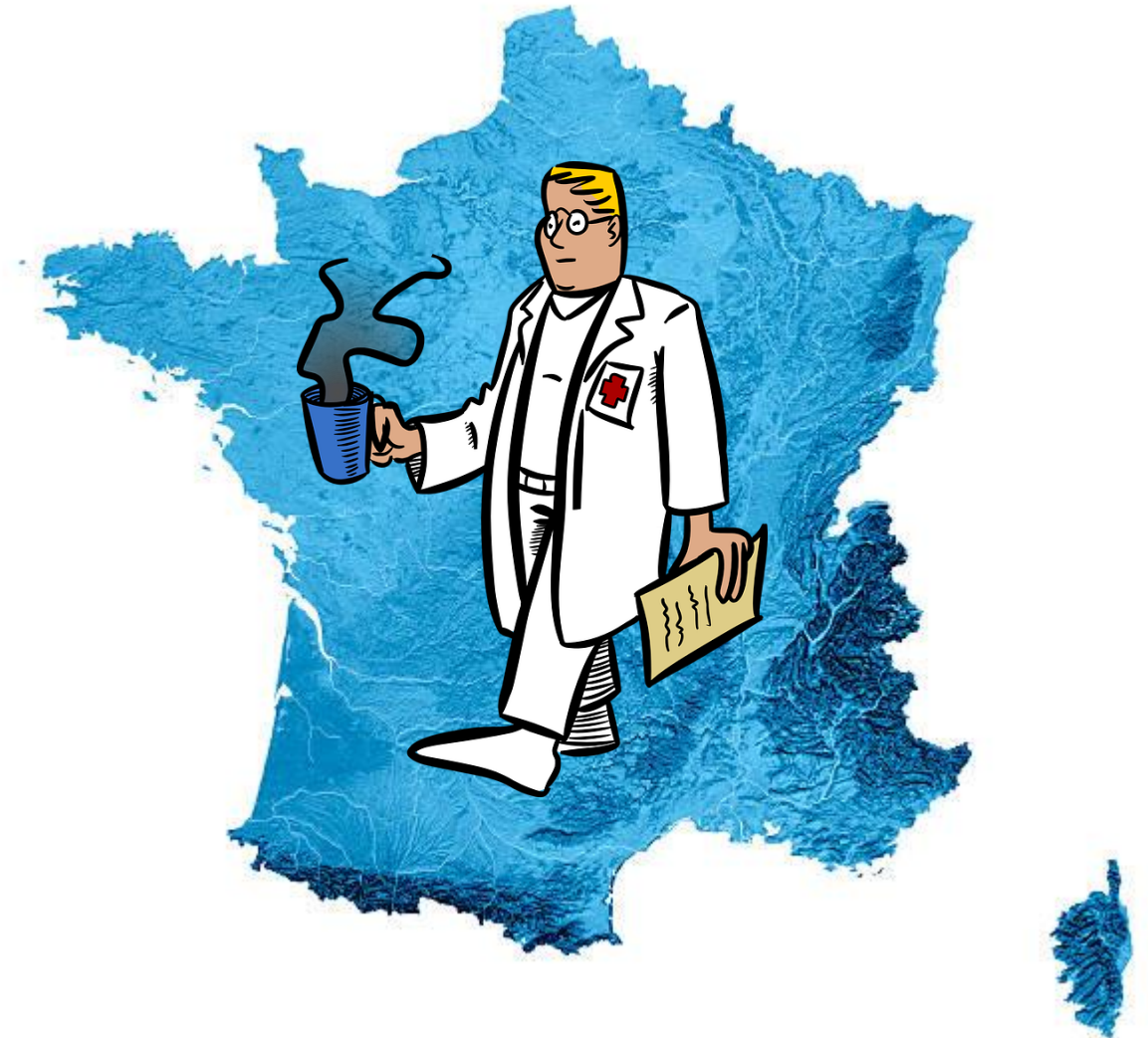
Distribution of doctors in France

Diane Deroualle – September 2019

➢ A problematic healthcare professional distribution in France

➢ **Purpose: build a model with machine learning to predict the distribution of healthcare professionals in France considering the location, the population of municipalities, the specialty of doctors….**

➢ Which areas lack healthcare professionals? Where future doctors should settle?

Datasets

❑ **First Dataset: Healthcare professional by zipcodes in France**
- Sex of doctors
- Specialisation of doctors

➔ **https://public.opendatasoft.com**

❑ **Second Dataset: Population by zipcodes in France**
- Population
- Surface Area
- Latitudes
- Longitudes
- Regions

➔ **https://data.opendatasoft.com**

❑ **Two extra datasets:**
- Poverty rate by region in France
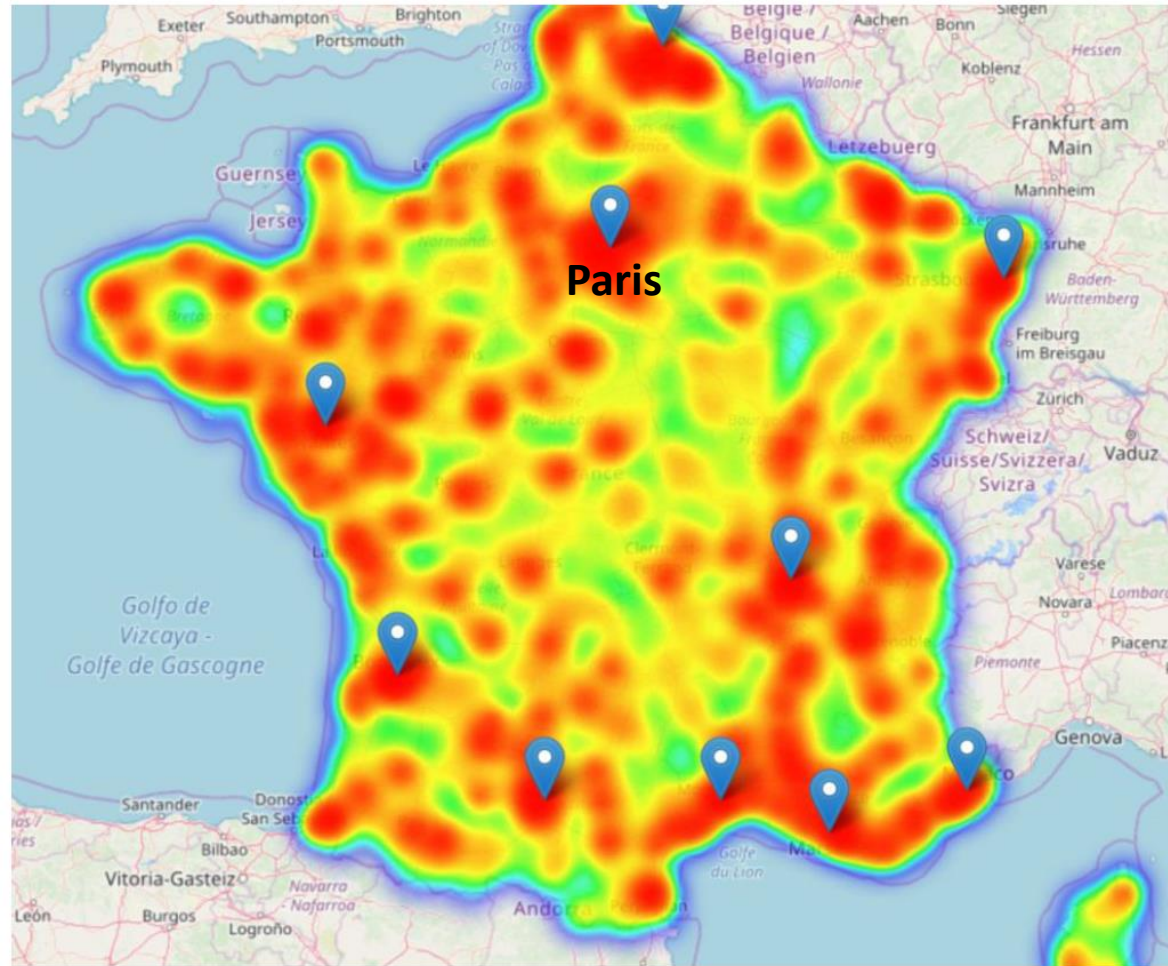- Unemployment rate by region in France

Relation between Population & Doctors Number
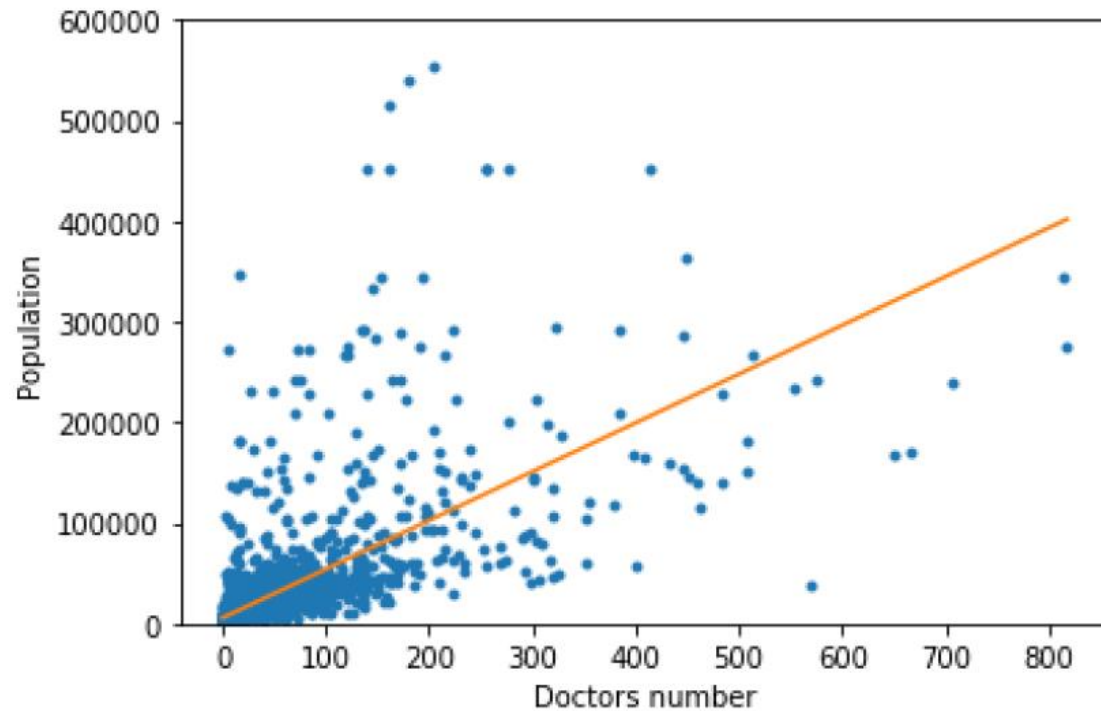
➔ Positive correlation between Population and Doctors Number. There is a higher Number of Doctors for the more populated territories.
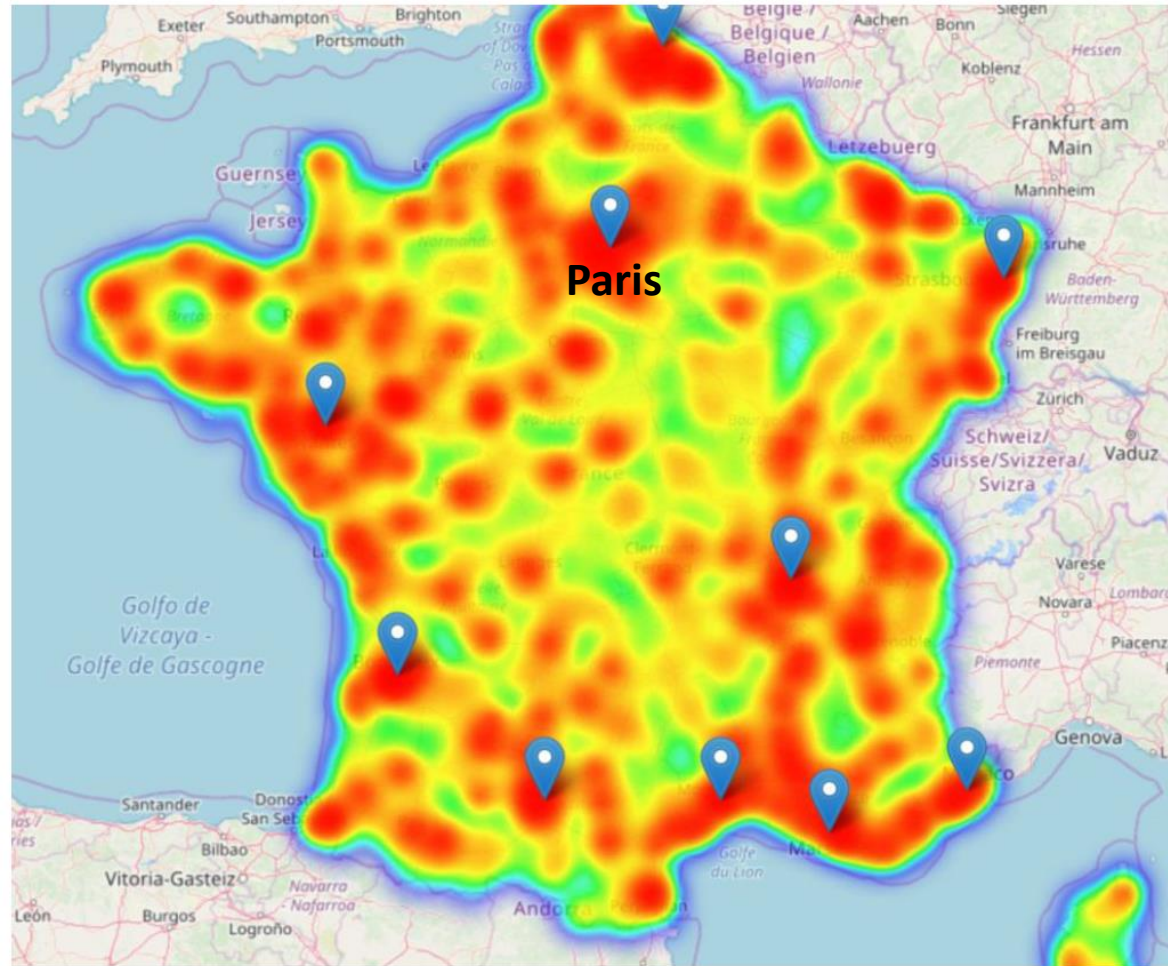
Density of the population of the France
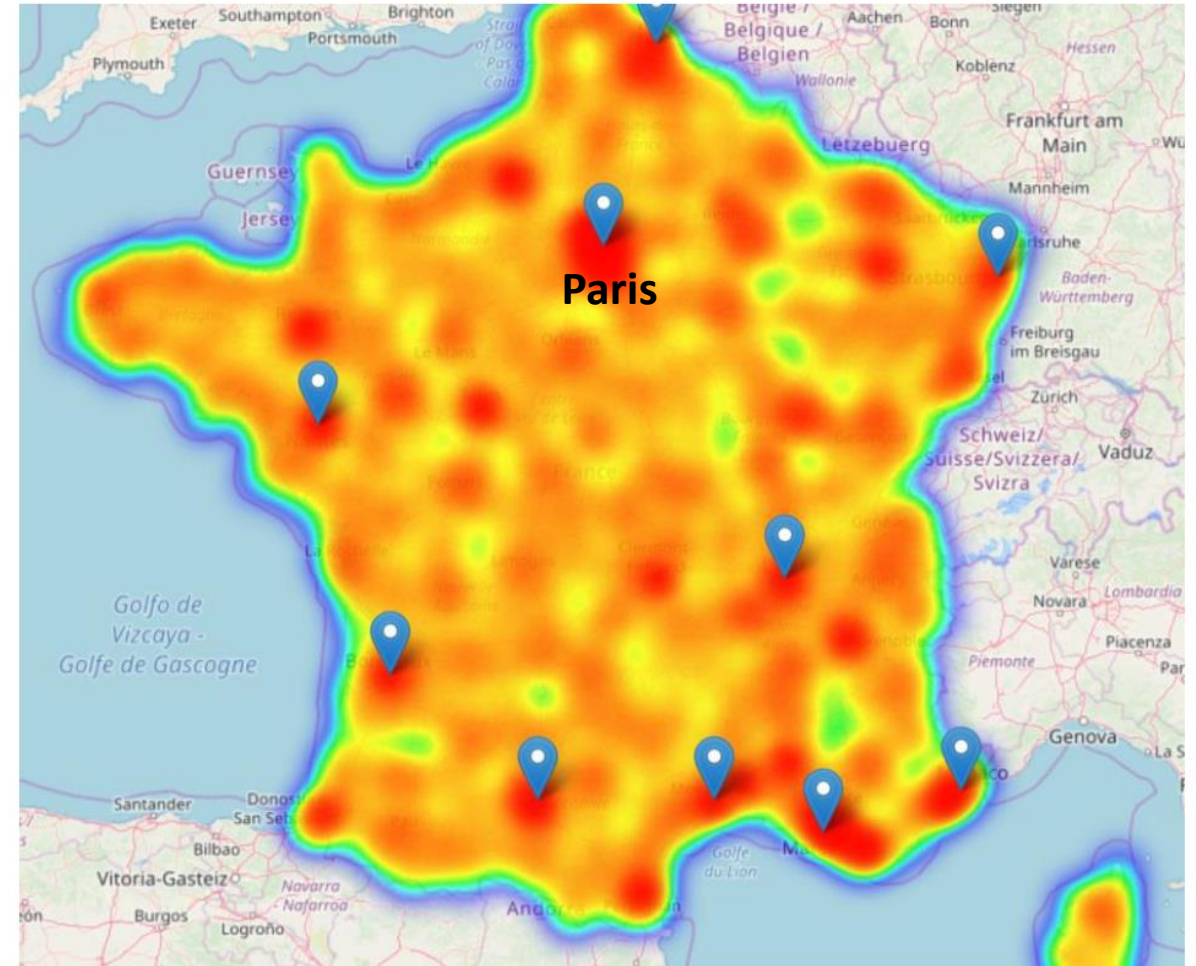
Heatmap Population vs. Doctors number

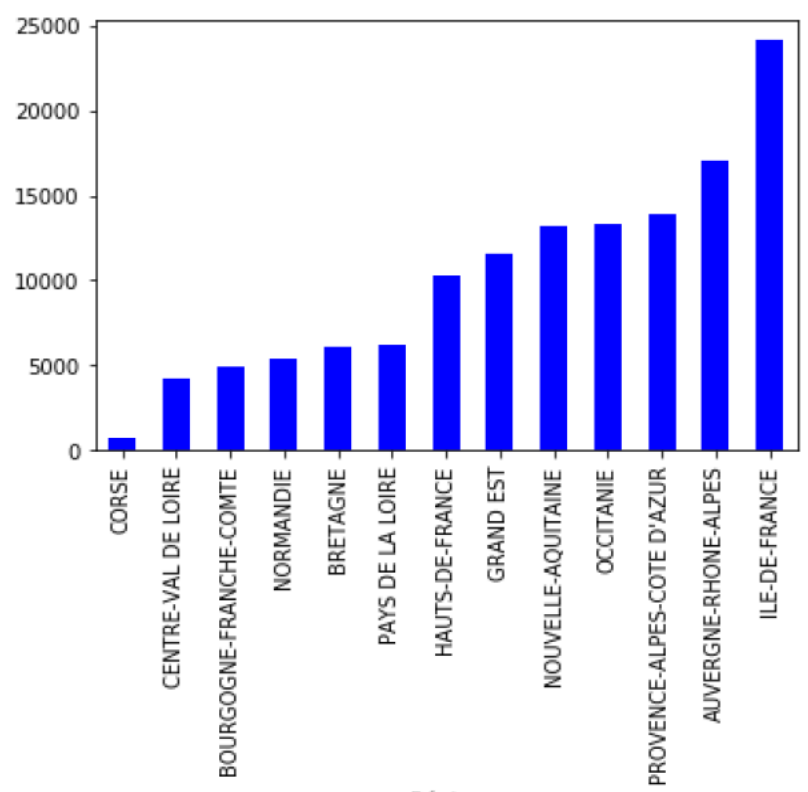**Density of the population of the France**
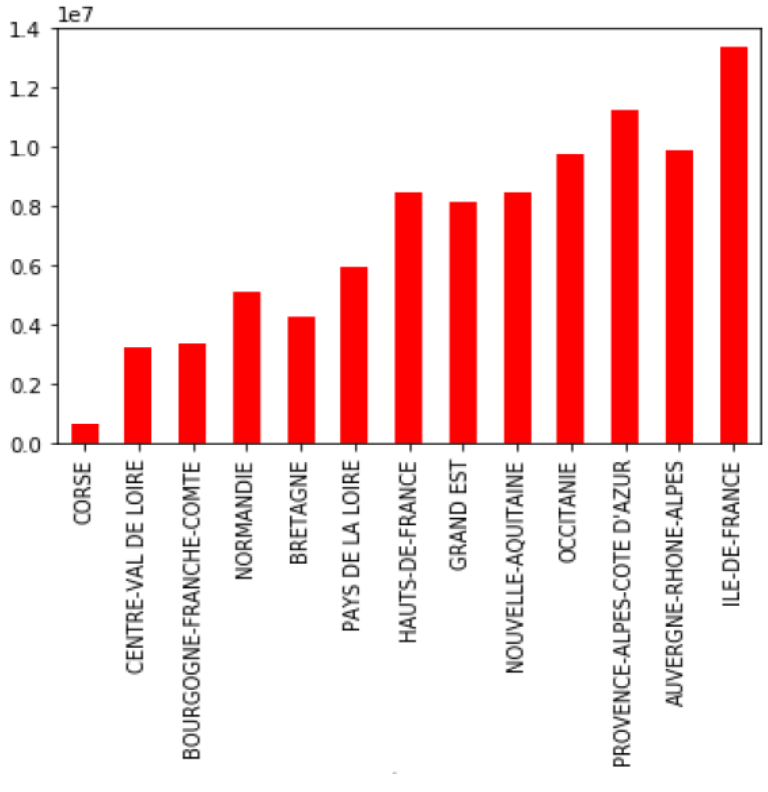

Density of the number of doctors in the France

→ There is a lot of doctors in more populated cities.

Population and Doctors number by Regions

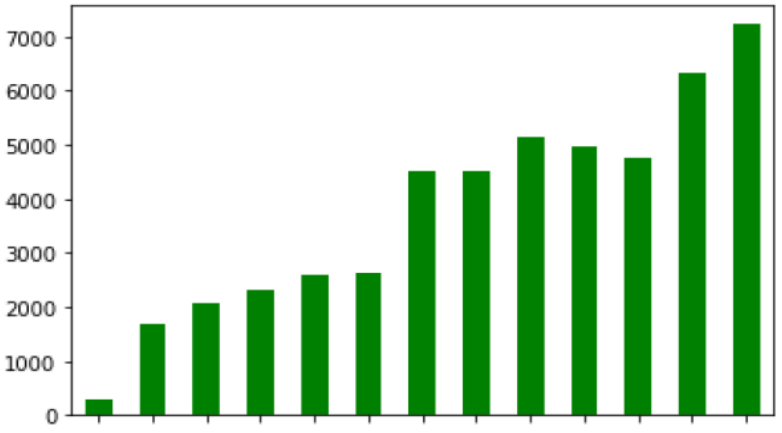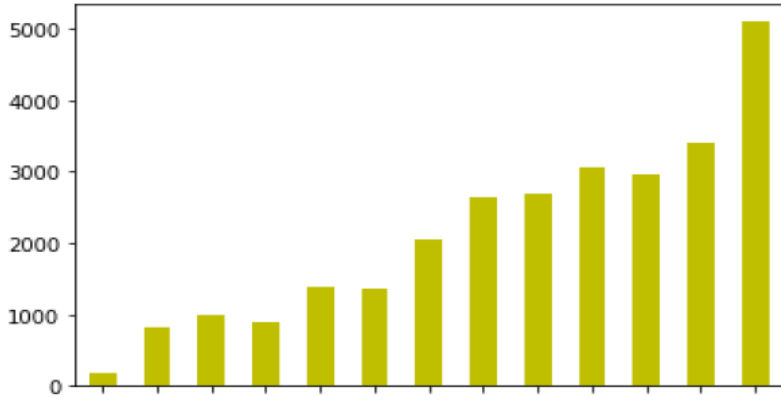➔ Similar distributions across French regions.

1st model: Predict the number of doctors per zip code

Linear Regression Model

$R^2 = 0.551$

Most significant features (p<0.001):

Features	+ / -	Coefficients
Surface Area	-	0.0004
Population	+	0.0007
General practitioner rate	-	0.99
Dental surgeon rate	-	0.96

Random Forest Regressor Model

R^2 training data = 0.93
R^2 test data = 0.78

Importance of features:

	importance
Population_Zipcode	0.633379
General_practitioner_rate	0.195719
SurfaceArea_Zipcode	0.047523
Longitude_from_Paris	0.041959
Dental_surgeon_rate	0.031124
Women_doctor_rate	0.018173
poverty_rate	0.014113
Latitude_from_Paris	0.010303
unemployment_rate	0.005956
North_South	0.001751

2nd model: Predict the population by General Practitioner by zip code

Linear Regression Model

$R^2 = 0.055$

Most significant features (p<0.001):

Features	+ / -	Coefficients
Surface Area	+	2.2
Latitude from Paris	+	5984.1
Male doctor rate	+	303.7

Random Forest Regressor Model

R^2 training data = 0.13
R^2 test data = 0.18

Importance of features:

	importance
SurfaceArea_Zipcode	0.407728
Latitude_from_Paris	0.240098
Women_doctor_rate	0.237319
Longitude_from_Paris	0.055437
poverty_rate	0.049127
unemployment_rate	0.008744
North_South	0.001545

1. **Model to predict the number of doctors by zip code**

➔ More doctors in the most populated cities

Is there the same ratio of general practitioner per habitant in both densely populated and less populated cities?

1. **Model to predict the number of doctors by zip code**

➔ More doctors in the most populated cities

Is there the same ratio of general practitioner per habitant in both densely populated and less populated cities?

2. **Model to predict the population by general practitioner by zip code.**

➔ Futur general practitioners should settle in priority in cities located in the south and east of Paris as indicated by the surface area, latitude and longitude from Paris features.

➔ We need more features that have an impact on our variable to predict.