

Population Evolution

Scop si Setup

Ne vom pune urmatoarea problema. Putem prezice evolutia populatiei unei asezari umane, pe 10 ani? Raspunsul este putem aproxima suficient.

Trebuie mai intai sa generam un set de date, sa alegem o formula. Vom genera 1000 de cazuri, si vom folosi 700 pt train si 300 pt test. Vom folosi regresia pntru a calcula populatia peste 10 ani a unui asezari umane.

Vom salva in X_train, X_test, Y_train s Y_test, toate fiind .csv, tabele de valori si numere.

Caracteristici

Deci avem urmatoarele date:

base_population	int64
city_type	object
birth_rate	float64
death_rate	float64
migration_rate	float64
health_index	float64
gdp_per_capita	float64
education_index	float64
resulting_population	float64

Populatia va fi un numar intreg, evident, apoi tipul de oras va apartine uneia din urmatoarele: rural, industrial, modern. Apoi avem doar variabile de tip float, rata de natalitate, mortalitate, bilantul de migratie. Acestea impreuna sunt bilantul total al populatiei. Apoi avem niste factori, considerati de mine personal importanti in calculul evolutiei unei societati, educatia, PIB per capita si sanatatea.

Formula aleasa

Vom calcula un `growth_rate`, pe care il vom inmulti la final ca sa obtinem noua populatie.

Acesta se calculeaza in felul urmatoare: Notam cu `X` `growth_rate`

$$X = N - M + BM$$

Unde `N` este `birth_rate`, `M` este `death_rate`, `BM` este `migration rate`. (nascuti vii, morti, bilant migratoriu).

$$X = X / 1000$$

Pt fiecare tip de asezare avem : Notam numarul random interval cu `nr`

$$X = X * nr$$

Unde intervalul este 1.1 1.2 pt rural, 1.0 1.1 pt industrial 1.5 2.0 pt modern

Apoi

$$X = X * 0.001 * \text{health_index}$$

$$X = X * 0.002 * \text{education_index}$$

$$X = X * 0.001 * (\text{gdp_per_capita} / 1000)$$

Apoi trebuie sa luam in calcul un factor random deci vom avea

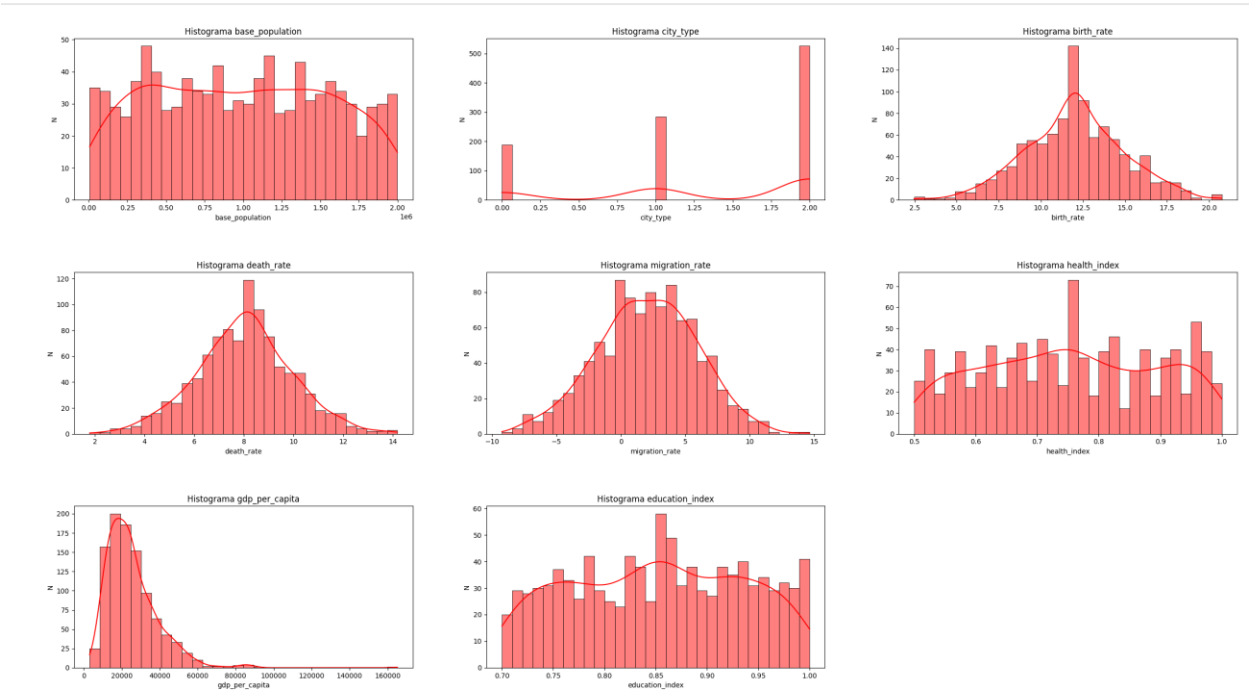
$$\text{Populatia_rezultanta} = \text{Populatia_start} * (1 + X + nr(0, 0.1)).$$

Statistici de descriere

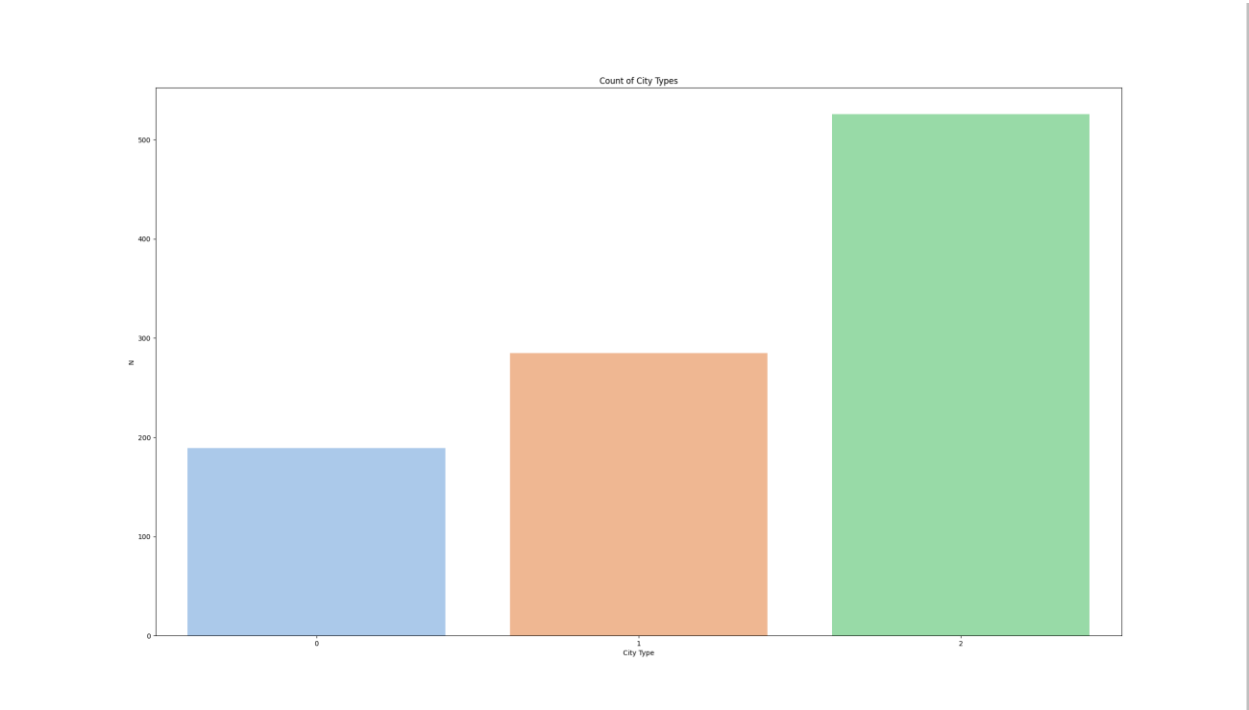
	base_population	city_type	birth_rate	death_rate	migration_rate	health_index	gdp_per_capita	education_index	resulting_population
count	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00	1000.00
mean	980257.15	1.34	12.03	8.02	2.02	0.75	25378.74	0.85	1019458.35
std	566389.08	0.78	2.92	1.92	3.86	0.14	13884.94	0.08	598151.71
min	5940.00	0.00	2.47	1.79	-9.24	0.50	3105.03	0.70	6600.00
25%	473298.00	1.00	10.14	6.82	-0.45	0.64	15762.87	0.78	495650.00
50%	968763.25	2.00	12.03	8.02	2.02	0.75	22580.78	0.85	1009041.00
75%	1460555.00	2.00	13.88	9.18	4.70	0.87	31711.92	0.92	1524597.25
max	1998902.00	2.00	20.76	14.18	14.65	1.00	165151.69	1.00	2425756.00

Distributia Variabilelor

Histograma

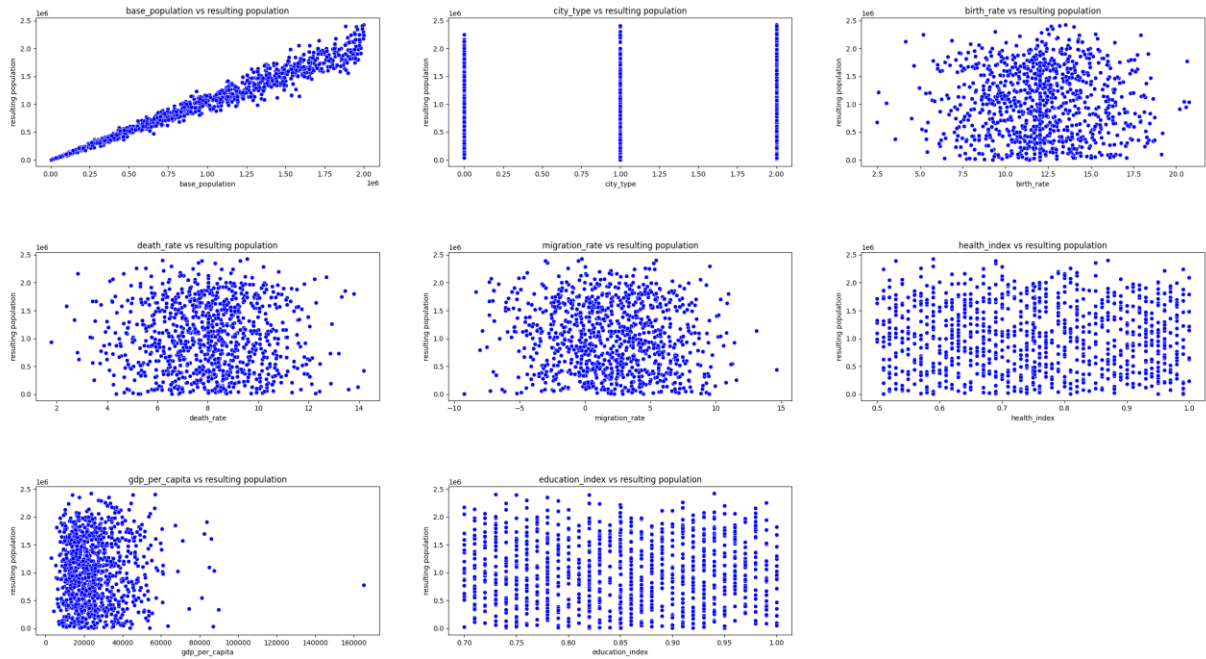


Club pt Orase

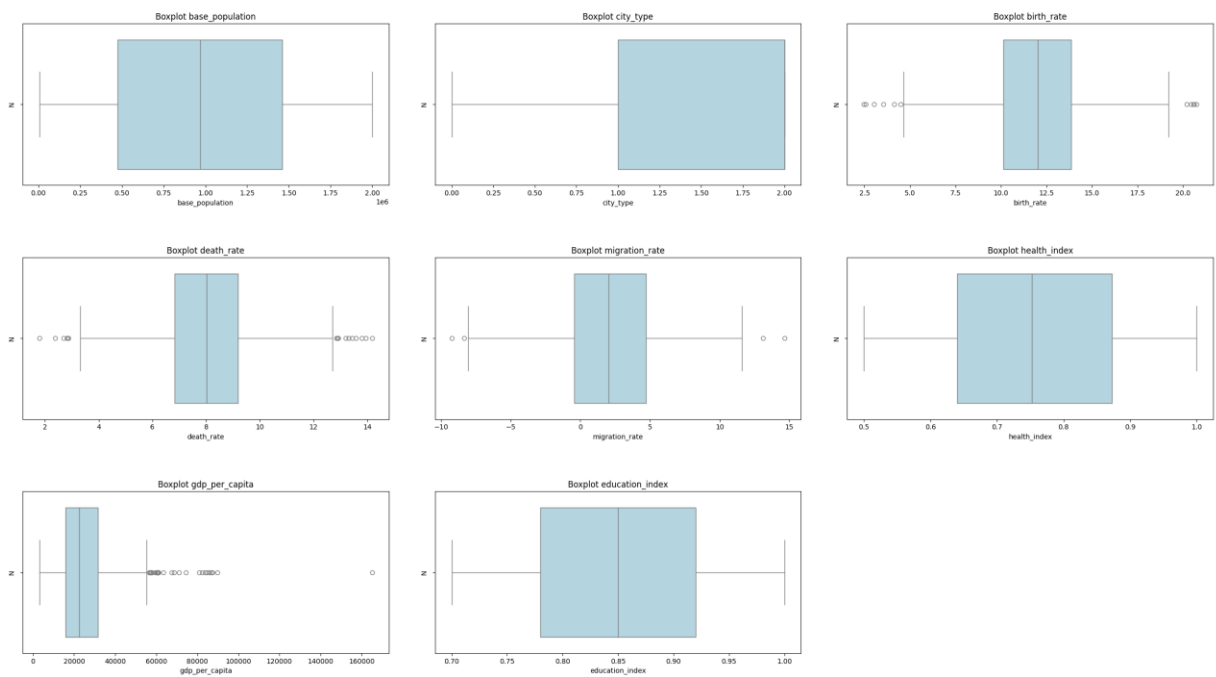


Outlieri

Scatter

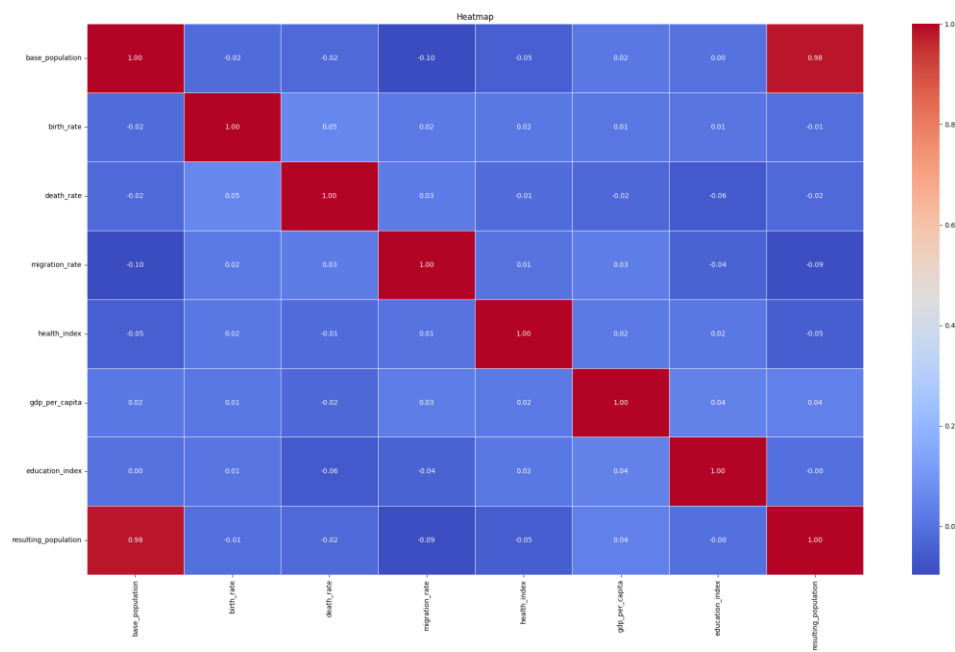


Box

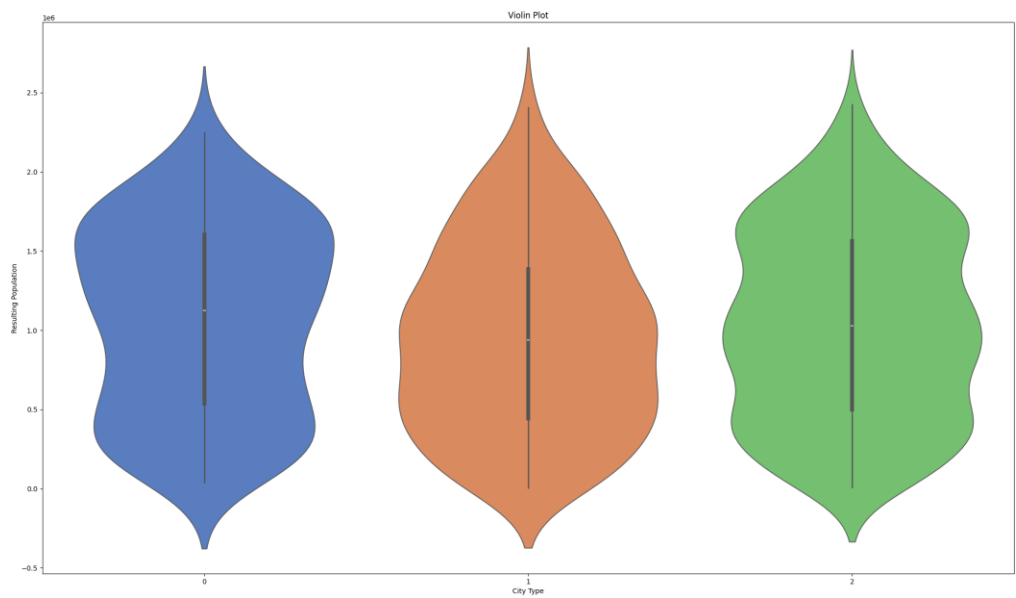


Corelatii

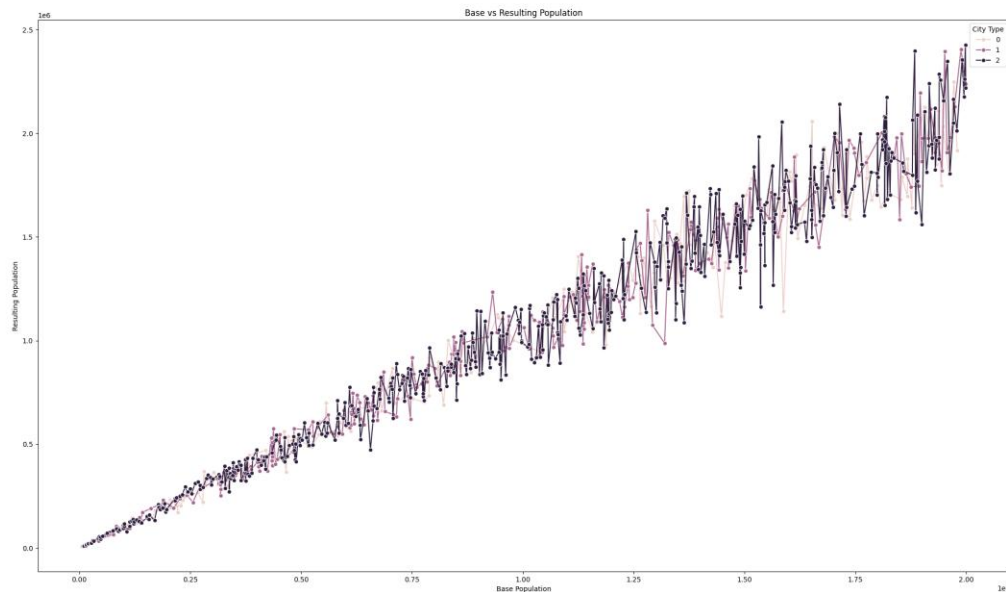
Heatmap



Violin



Lineplot pentru populatie la inceput vs dupa 10 ani, cu legenda tipului de asezare



Antrenare si regresia liniara

Folosim MAE pt regresie (Mean Absolute Error), vom obtine eroare de $73 * 10^4$, care pentru selectia de populatie (5000, 2000000), este una foarte buna (populatia este generata ca un numar intre 5000 si 200000).

