
Πείραμα 2 και ημερολόγιο συναντήσεων μετά τις 28 Φεβρουαρίου 2023

Κώστας Παπαδόπουλος

Πειραιάς, 2030

Contents

1	Επανεκίνηση μετά τα Χριστούγεννα	4
2	Δοκιμές για διαφορετικά βάρη στις παραμέτρους	4
3	Για τη 31η Γενάρη	5
4	Για την 28η Φεβρουαρίου	10
4.1	Αλλαγή 1	10
4.2	Μεσαία χώρα ανά τα έτη	10
4.3	Όλες οι χώρες για όλα τα έτη	11
5	Για τη 14η Μαρτίου	14
5.1	Πίνακας r^2 με βέλτιστα επιλεγμένα βάρη	14
5.2	Οπτικοποίηση δεδομένων	16
5.2.1	Πληθυσμός	16
5.2.2	Πληθωρισμός	18
5.2.3	Total energy Supply	20
5.2.4	GDP per capita	20
5.2.5	Verified emissions	21
5.2.6	Agriculture	24
5.2.7	Industry	26
5.2.8	Manufacturing	26
6	Για την 21η Μαρτίου	28
6.1	Γιατί η Πολωνία και η Γαλλία αποτυγχάνουν να εξηγήσουν τις άλλες πριν το 2012;	28
6.2	Τα βέλτιστα βάρη για όλα τα σενάρια	29
7	Για την 28η Μαρτίου	30
7.1	Τι όντως έκανα	30
7.2	PPS (Purchasing Power Standards)	30
7.2.1	Χρήση PPS	30

7.2.2	Υπολογισμός	30
7.3	Power Intensity	31
7.4	Για την επόμενη φορά	31
7.4.1	Σε δεύτερο χρόνο	31
7.4.2	Άμεσα, για το τετρασέλιδο	32
8	Για τις 4 Απριλίου	32
8.1	Συσταδοποίηση (έπρεπε...)	32
8.1.1	Πλήθος συστάδων	32
8.1.2	Βέλτιστες συστάδες	34
8.1.3	Συστάδες, αν όλα τα δεδομένα ήταν κατά κεφαλήν και κανονικοποιημένα ύστερα	34
8.2	Ικανότητα εξήγησης υπολοίπων μέσα στην ίδια συστάδα	35
8.3	Υπάρχει κάποια εύκολη σχέση μεταξύ των δωρεάν αδειών και κάποιου feature;	36
8.4	Ύστερα από αλλαγή στον τρόπο της κανονικοποίησης	38
8.5	To do list	39
9	Για την 2η Μαΐου	43
10	Για τις 16 Μαΐου	44
10.1	To do	44
10.2	Δεν μοιάζει να μπορεί να προσομοιωθεί απλά το Energy Intensity	44
10.3	Διαφορετικά μέτρα απόδοσης / αποτελεσματικότητας	46
10.4	Για την επόμενη φορά	46
11	Για τις 23 Μαΐου	46
11.1	21-99 Cannot be used a proxy	46
11.2	Χρήση του GDPpps	48
11.3	Χωρίς Αυστρία, Λιθουανία και Φινλανδία	48
12	Για τη 13η Μαΐου	50
12.1	Το πρόβλημα με το πλασματικό energy Intensity	50
12.2	Clustering se sectors	51
12.3	Τι ειπώθηκε στην κλίση	52
13	Παύση Καλοκαιριού	53
14	Επανεκκίνηση Μετά την παύση του καλοκαιριού	53
14.1	Polution Permits: Efficiency by Design	53
14.2	Παρουσίαση για το paper	53
14.3	1 sector	53
15	Συνάντηση στο ΜΟΠ 10/10/23	53
15.1	Προκύπτοντα ερωτήματα	53
15.2	Προτεινόμενοι στόχοι	53
16	Συνάντηση στο ΜΟΠ 31/10/23	54
16.1	Σημαντικά Ειπωθέντα	54
16.2	Στόχοι	54

17 Συνάντηση μέσω Skype στις 5 Νοεμβρίου 2023	54
17.1 Δημιουργία Συνθετικών Δεδομένων	54
17.2 Ανάλυση Εταιρικών Συναρτήσεων	55
17.3 Στόχοι για την Επόμενη Συνάντηση	55
18 Συνάντηση στο ΜΟΠ στις 7 Νοεμβρίου 2023	55
18.1 Πριν την συνάντηση	55
18.1.1 Ορισμοί	55
18.1.2 Πιθανές συναρτήσεις παραγωγής	55
18.1.3 Abatement cost	56
18.2 Τι είπαμε στην συνάντηση - Ορισμός προβλήματος	56
18.2.1 Βήμα 1	57
18.2.2 Βήμα 2	57
18.2.3 Βήμα 3 - Βελτιώσεις	57
19 Συνάντηση 14 Νοεμβρίου 2023 στο ΜΟΠ	58
20 Συνάντηση 28 Νοεμβρίου 2023	58
21 Πηγές, από 23/5/23	58

1 Επανεκίνηση μετά τα Χριστούγεννα

Έγινε μία καλύτερη επιλογή των πηγών για τα δεδομένα, το οποίο οδήγησε, χωρίς να μπορώ να το εξηγήσω σε κάτι πολύ περίεργες συμπεριφορές. Τα καινούρια δεδομένα ήταν τα:

1. population ίδιο
2. GDP per capita ίδιο
3. Total energy supply ίδιο
4. inflation ίδιο
5. verified emissions ίδιο
6. manufacturing σε δισεκατομύρια δολλάρια, άλλαξε και πλέον δεν έχουμε δεδομένα μόνο για 2010 και 2020, αλλά έχουμε ξεχωριστά για κάθε χρονιά
7. industry σε δισεκατομύρια δολλάρια, ομοίως
8. Agriculture σε δισεκατομύρια δολλάρια, ομοίως

Ένα πιθανό πρόβλημα το οποίο ίσως προκύπτει είναι το ότι $(6 + 7 + 8) \approx 1 * 2$. Όμως ρεαλιστικά: $1 * 2 - (6 + 7 + 8) \approx \text{services}$. Οπότε ελπίζω πως αυτό δεν είναι σοβαρό πρόβλημα.

Το δεύτερο που παρατηρώ είναι πως για κάποιο λόγο μετά τις βελτιώσεις στον κώδικα, πλέον αν βάλω τα δεδομένα απλώς να λογαριθμιστούν, τα αποτελέσματα είναι λογικά. Αν βάλω τα αποτελέσματα να κανονικοποιηθούν, πάλι είναι λογικά, αλλά προσεγγίζουν λιγότερο την ευθεία γραμμή. Αν όμως τα βάλω και τα δύο, τότε τα δεδομένα δείχνουν να μην έχουν κάποια γραμμική συσχέτιση, το οποίο δεν το καταλαβαίνω.

Επίσης, λόγω έλλειψης δεδομένων για την παραγωγή της Βουλγαρίας, τα δεδομένα υπολογίστηκαν ως GDP - Agriculture - services - industry.

2 Δοκιμές για διαφορετικά βάρη στις παραμέτρους

Ξέρω πως δεν είπαμε να κάνω αυτό, αλλά όταν ξεκίνησα να το κάνω, δεν μπορούσα να σταματήσω τις δοκιμές. Αρχικά. Όλα όσα θα παρουσιαστούν παρακάτω είναι έχοντας ενεργοποιημένη μόνο την κανονικοποίηση και όχι την λογαρίθμηση, γιατί για αυτήν είχα απορία.

Για να κρίνω το πόσο σωστά είναι τα βάρη, διάλεξα να προσπαθώ να μεγιστοποιήσω το r squared της παλινδρόμησης. Εγιναν λοιπόν οι παρακάτω δοκιμές:

Population	GDP per Capita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	r ²
1	1	1	1	1	1	1	1	0.76
4	0	0	0	10	0	3	10	0.93
0	0	0	0	6.8	0	3.2	10	0.9496
100	0	0	50	626	0	400	1000	0.9483
0	0	0	60	570	0	360	981	0.953

Εδώ είναι πολύ εμφανής η τεράστια εξάρτηση από τα verified emissions, το οποίο είναι εντελώς αναμενόμενο. Στην επόμενη δοκιμή, τα verified emissions είχαν αυστηρά τιμή 0.

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	r^2
1	1	1	1	1	1	1	0	0.70
0	0	0	0	1000	50	500	0	0,913
0	0	0	500	10000	0	2700	0	0,92

Αντίστοιχα, εδώ φαίνεται μία πολύ μεγάλη εξάρτηση από την παράμετρο του ακαθάριστου προϊόντος η οποία αφορά στην βιομηχανία. Αν την αφαιρέσουμε και αυτήν από το παιχνίδι, προσπαθώντας να βρούμε ένα λίγο διαφορετικό mix βλέπουμε πως:

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	r^2
1	1	1	1	0	1	1	0	0,63
0.5	0.5	0.5	0.5	0	1	10	0	0.84
0	0	0	0	0	50	1000	0	0.86
0	0	0	0	0	4300	99500	0	0.8645

Η τελευταία δοκιμή για να δούμε αν τα δεδομένα έστω και λίγο βγάζουν κάποιο νόημα είναι αυτήν την οποία βγάζουμε και το total energy supply το οποίο κυριαρχεί ξανά:

Population	GDP perCapita	Inflation	Agriculture GDP	Industry GDP	Manufacturing GDP	Total Energy Supply	Verified emissions	r^2
1	1	1	1	0	1	0	0	0.549
200	0	0	0	0	100	0	0	0.767
6640	30	10	350	0	3720	0	0	0.768

3 Για τη 31η Γενάρη

Δεν ξέρω αν έχει νόημα αυτό το pdf να γίνει λίγο σαν ημερολόγιο, αλλά ελπίζω να μην είναι μεγάλο πρόβλημα. Την τελευταία φορά είπαμε:

1. Η σύγκριση να γίνει μόνο με μία χώρα.
2. Είπε ο κ. Φωτάκης δύο εικασίες. Πρώτη πως το allocation είναι το αποτέλεσμα ενός optimization αλγόριθμου. Δεύτερη, πως είναι ένα πολυκριτηριακό optimization. Αν αυτά ισχύουν θα ήταν έξυπνο / προφανές να βρούμε, συγκρίνοντάς το με άλλα αντίστοιχα, ποιο πρόβλημα προσπαθεί να λύσει. Στη συνέχεια θα είναι αιτιολογημένο και πιο λογικό να προσθέσουμε πάνω σε αυτό οποιοδήποτε επιπλέον constraint θέλουμε.
3. "Policy options to improve the effectiveness of the EU emissions trading system: A multi-criteria analysis, stefan clo 2013" - Είναι αρκετά γενικό αλλά ίσως είναι χρήσιμο για να δούμε την στοχοθεσία της ΕΕ.
4. Για να κρίνουμε μία παλινδρόμηση, καλό είναι να ελέγχονται 3 κριτήρια: r^2 , MSE, MAE
5. Είναι λογικό ο πληθωρισμός για αυτά τα δεδομένα να παίζει ελάχιστο ρόλο.

Λοιπόν, το πρώτο βήμα είναι να βάλουμε μία χώρα με την οποία θα συγκρίνουμε όλες τις υπόλοιπες. Επομένως, πλέον δεν κοιτάμε αν για να γίνει η σύγκριση με την μεσαία χώρα πρέπει πρώτα να οριστεί η μεσαία χώρα. Αυτή

προκύπτει ως εξής: Δίνουμε σε κάθε χώρα πόντους ίσους με το άθροισμα των θέσεών της ως προς κάθε feature. Στη συνέχεια διαλέγουμε αυτή με τη μεσαία βαθμολογία. Κάναμε όμως και δοκιμές για άλλες χώρες για σιγουριά. Μία πρώτη εντύπωση είναι πως όσο πιο ακραία μικρή ή ακραία μεγάλη είναι η οικονομία της χώρας, τόσο πιο γραμμική είναι η σχέση της απόστασης.

2017	r^2
Hungary	0.75
Germany	0.9147
Greece	0.6694
Malta	0.8331

Στη συνέχεια, αν συγκρίνουμε όλες τις χώρες με την Ουγγαρία, τότε προκύπτει αυτό το διάγραμμα. Συγκριμένα, βλέπουμε:

- Στον x άξονα: Την απόσταση της εκάστοτε χώρας με την Ουγγαρία ως προς το feature set που έχουμε επιλέξει. (GDP per capita, inflation, Population, manufacturing, industry, verified emissions, agriculture, total energy supply)
- Στον y άξονα: Την απόσταση της εκάστοτε χώρας με την Ουγγαρία ως προς το σύνολο των δωρεάν αδειών που έλαβαν μέσα στο 2017 εταιρείες με έδρα την Ουγγαρία.
- Τα δεδομένα κανονικοποιήθηκαν στο $[0,1]$ πριν τον υπολογισμό των αποστάσεων, αλλά δε λογαριθμίστηκαν.



Figure 1: Hungary 2017

Το οποίο μπορούμε να το αντιπαράθεσουμε με τις 3 άλλες χώρες που είδαμε και πριν.



(a) Malta 2017



(b) Greece 2017



(c) Germany 2017

Παράλληλα μπορούμε να δούμε μέσα στις χρονιές επαναλαμβάνοντας το ίδιο πείραμα για όλες τις χώρες, τι θα προκύψει.

year	slope	mean r_squared	min_r_squared	max_r_squared
2008	0.34	0.65	0.00	0.79
2009	0.32	0.60	0.02	0.78
2010	0.33	0.58	0.04	0.79
2011	0.33	0.62	0.04	0.78
2012	0.34	0.64	0.07	0.82
2013	0.37	0.72	0.51	0.92
2014	0.36	0.71	0.46	0.93
2015	0.37	0.75	0.48	0.94
2016	0.35	0.70	0.50	0.91
2017	0.36	0.73	0.47	0.91
2018	0.37	0.74	0.43	0.92

Συμπεράσματα:

- Η κλίση αυτής της ευθείας δείχνει να παραμένει πολύ σταθερή
- Μέχρι το 2012 (δηλαδή και το τέλος της φάσης II) βλέπουμε πως υπάρχουν χώρες για τις οποίες δεν ισχύει το συμπέρασμα το οποίο είχαμε βγάλει, περί δικαιοσύνης του συστήματος.
- Μετά τη φάση III ακόμα υπάρχουν χώρες για τις οποίες η μέθοδος αυτή δε δίνει σημαντικά καλές τιμές.

Ένα ερώτημα που νομίζω πως είναι λογικό να προκύψει ξαφνικά είναι το "ποιο κριτήριο κάνει μία χώρα να είναι κατάλληλη για αυτή τη δουλειά; Πότε μπορούμε να διαλέξουμε μία χώρα για να είναι χρήσιμες οι αποστάσεις της;". Οπότε τα βάζουμε όλα μαζί στο ίδιο γράφημα και έχουμε:

Figure 3: 2017 r^2 vs GDP

Δοκιμάζοντας και άλλα χαρακτηριστικά δεν κατέληξα σε κάτι πιο ενδιαφέρον από το παραπάνω διάγραμμα. Η Γερμανία δείχνει να είναι πολύ καλή για να κάνουμε αυτή τη δουλειά, μάλλον γιατί στον "χώρο των αποστάσεων"

είναι τόσο μακριά από όλους τους άλλους που οι υπόλοιποι καταλήγουν να είναι συγκριτικά κοντά. Αυτό το λέω σαν διαισθητική παρατήρηση, δεν ξέρω αν στηρίζεται μαθηματικά.

Το επόμενο βήμα είναι να δούμε το προηγούμενο με τα βέλτιστα βάρη.

4 Για την 28η Φεβρουαρίου

Την τελευταία φορά προτείναμε να κάνουμε/δούμε/διαβάσουμε, χωρίς κάποια σειρά:

1. Να διαβάσω τον αλγόριθμο του benchmark, πώς προκύπτει ακριβώς.
2. Να εξετάσουμε αν τα αποτελέσματα σχετικά με το "ποιες χώρες εξηγούν καλύτερα τα δεδομένα είναι consistent μέσα στον χρόνο.
3. Υπάρχει κάποιο clustering μεταξύ των χωρών που εξηγούν καλά ή κακά τις άλλες χώρες; Έχουν κάποια κοινά χαρακτηριστικά όλες αυτές;
4. Γιατί είναι η Ολλανδία τόσο διαφορετική;
5. Να επαναληφθούν οι δοκιμές της προηγούμενης φοράς, αλλά για τις φάσεις 1,2 (πιθανώς θα προκύψει πρόβλημα με τα δεδομένα για αυτές τις περιόδους, αλλά θα το δούμε)
6. Κάνε δοκιμές και με μία fictional χώρα η οποία να έχει ως τιμές της μόνο τα medians όλων των χωρών.
7. Να γίνει καλύτερα η σύγκριση μεταξύ κάθε δύο παλινδρομήσεων, αξιολογώντας και το r^2 , αλλά και άλλες παραμέτρους, όπως το p-value και MSE/MAE.
8. Τι συμβαίνει στα πρώτα χρόνια με την Πολωνία και την Γαλλία;
9. Να χρησιμοποιηθεί ggplot για τα διαγράμματα και να έχουν πάντα grid.
10. Για διάβασμα : A multi-criteria decision analysis model for carbon emission quota allocation in China's east coastal areas: Efficiency and Equity.

4.1 Αλλαγή 1

Το πρώτο βήμα που έγινε είναι το να ξαναγίνουν πολλά από τα πειράματα, μόνο που αυτή τη φορά, για να θεωρηθεί μία γραμμική παλινδρόμηση καλύτερη από μία άλλη, θα πρέπει να είναι αληθή τα παρακάτω κριτήρια:

- Έχει μεγαλύτερο r^2 .
- Έχει p-value μικρότερο του 0.05.
- Το Mean Square Error δεν είναι πολύ μεγαλύτερο από την προηγούμενη μέγιστη τιμή (1.5 φορά).

4.2 Μεσαία χώρα ανά τα έτη

Εδώ βλέπουμε την πιο μεσαία χώρα κάθε χρονιάς και το πόσο καλά αποδίδει στο να εξηγεί τις άλλες χώρες. Οι τιμές στο p-value βγήκαν όντως 0, η μεγαλύτερη είχε την τιμή 0.00002581...

Year	R ²	p-value	MSE	Country
2008	0.77	0.00	0.01	Denmark
2009	0.69	0.00	0.01	Denmark
2010	0.57	0.00	0.01	Greece
2011	0.65	0.00	0.02	Denmark
2012	0.75	0.00	0.01	Ireland
2013	0.71	0.00	0.01	Portugal
2014	0.58	0.00	0.02	Hungary
2015	0.77	0.00	0.01	Portugal
2016	0.70	0.00	0.01	Portugal
2017	0.75	0.00	0.01	Hungary
2018	0.74	0.00	0.01	Hungary

4.3 Όλες οι χώρες για όλα τα έτη

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.70	0.62	0.66	0.68	0.71	0.74	0.73	0.82	0.72	0.68	0.74	0.00	0.02
Belgium	0.62	0.62	0.62	0.64	0.65	0.67	0.67	0.65	0.57	0.60	0.59	0.00	0.01
Bulgaria	0.75	0.67	0.73	0.67	0.78	0.84	0.85	0.86	0.69	0.81	0.85	0.00	0.01
Cyprus	0.79	0.71	0.69	0.71	0.78	0.67	0.58	0.73	0.72	0.75	0.78	0.00	0.02
Denmark	0.77	0.69	0.61	0.65	0.74	0.71	0.72	0.73	0.72	0.68	0.69	0.00	0.02
Estonia	0.79	0.74	0.67	0.69	0.66	0.64	0.68	0.78	0.68	0.82	0.81	0.00	0.02
Finland	0.71	0.64	0.66	0.73	0.68	0.73	0.77	0.80	0.72	0.69	0.76	0.00	0.01
France	0.08	0.08	0.20	0.19	0.25	0.78	0.79	0.78	0.78	0.73	0.76	0.19	0.01
Germany	0.76	0.75	0.79	0.76	0.80	0.92	0.93	0.94	0.91	0.91	0.92	0.00	0.01
Greece	0.62	0.57	0.57	0.60	0.57	0.51	0.46	0.69	0.65	0.67	0.71	0.00	0.02
Hungary	0.74	0.52	0.58	0.69	0.69	0.67	0.58	0.78	0.73	0.75	0.74	0.00	0.02
Ireland	0.70	0.76	0.47	0.70	0.75	0.73	0.69	0.77	0.73	0.63	0.64	0.00	0.02
Italy	0.63	0.61	0.65	0.64	0.60	0.86	0.82	0.75	0.74	0.72	0.78	0.00	0.01
Latvia	0.76	0.70	0.62	0.66	0.68	0.77	0.74	0.75	0.78	0.84	0.81	0.00	0.02
Lithuania	0.77	0.78	0.63	0.70	0.74	0.79	0.76	0.74	0.74	0.79	0.81	0.00	0.02
Luxembourg	0.79	0.67	0.59	0.77	0.73	0.68	0.65	0.75	0.60	0.74	0.70	0.00	0.02
Malta	0.76	0.65	0.62	0.68	0.78	0.75	0.72	0.82	0.71	0.83	0.82	0.00	0.02
Netherlands	0.51	0.53	0.53	0.44	0.54	0.60	0.60	0.51	0.50	0.47	0.43	0.00	0.01
Poland	0.00	0.02	0.04	0.04	0.07	0.60	0.60	0.48	0.54	0.69	0.69	0.92	0.02
Portugal	0.71	0.66	0.66	0.60	0.59	0.71	0.73	0.77	0.70	0.75	0.77	0.00	0.02
Romania	0.53	0.46	0.42	0.58	0.45	0.60	0.75	0.80	0.64	0.66	0.67	0.00	0.01
Slovenia	0.78	0.59	0.59	0.67	0.70	0.76	0.73	0.77	0.78	0.78	0.81	0.00	0.02
Spain	0.63	0.61	0.63	0.57	0.60	0.84	0.74	0.76	0.72	0.77	0.81	0.00	0.01
Sweden	0.70	0.56	0.61	0.59	0.67	0.74	0.75	0.81	0.75	0.77	0.77	0.00	0.02
United Kingdom	0.75	0.71	0.77	0.78	0.82	0.63	0.69	0.66	0.66	0.72	0.71	0.00	0.01

Το οποίο γραφικά φαίνεται και ως εξής:

R² values for each country and year

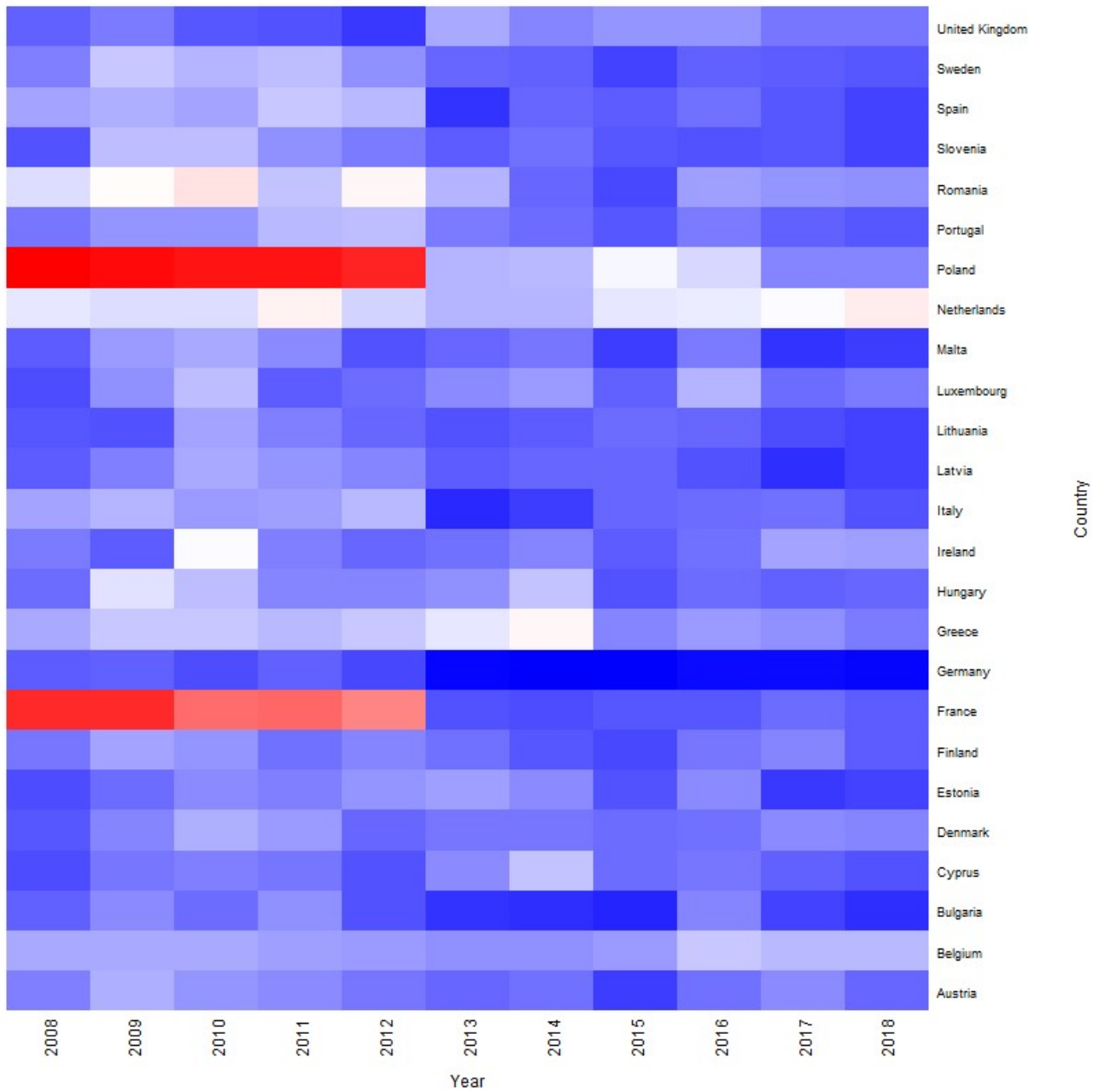


Figure 4: r^2

Το πιο λογικό λοιπόν βήμα είναι το να ελέγξουμε το γιατί αυτές οι δύο χώρες αλλάζουν τόσο πολύ το 2013. Οπότε ας δούμε τα δεδομένα που χρησιμοποιούμε για αυτές:

	2012	2013	2012	2013
GEO	Poland	Poland	France	France
Total_energy_supply	0.3083121	0.3038433	0.8215192	0.8080548
GDPpc	0.1163326	0.1141371	0.3630366	0.3550416
Population	0.4732704	0.4716958	0.8164021	0.8183792
Inflation	0.57990986	0.07563843	0.28530815	0.19301719
Verified_emissions	0.4212666	0.4212431	0.2421093	0.2422275
Agriculture	0.3402106	0.3644831	1.0000000	0.8928814
Industry	0.1558893	0.1482026	0.4982965	0.5053396
Manufacturing	0.1166064	0.1091561	0.3911272	0.3913494

Όπως είναι εμφανές, στα δεδομένα των χωρών δεν υπάρχει κάτι προφανές το οποίο να δικαιολογεί αυτήν την τόσο απότομη αλλαγή. Τα δεδομένα τα βλέπουμε κανονικοποιημένα.

Επίσης, για το ποιες χώρες τα πηγαίνουν καλύτερα, μπορούμε να αφτιάξουμε αυτήν την κατάταξη, η οποία είναι αρκετά αντιπροσωπευτική.

	Country	Average R^2
1	Germany	0.85
2	Bulgaria	0.77
3	Lithuania	0.75
4	Malta	0.74
5	Latvia	0.74
6	Estonia	0.72
7	Slovenia	0.72
8	Cyprus	0.72
9	United Kingdom	0.72
10	Finland	0.72
11	Italy	0.71
12	Austria	0.71
13	Sweden	0.70
14	Denmark	0.70
15	Spain	0.70
16	Luxembourg	0.70
17	Portugal	0.70
18	Ireland	0.69
19	Hungary	0.68
20	Belgium	0.63
21	Greece	0.60
22	Romania	0.60
23	Netherlands	0.52
24	France	0.49
25	Poland	0.34

5 Για τη 14η Μαρτίου

Είπαμε να διορθωθούν τα παρακάτω:

- Οι πίνακες με δεδομένα να έχουν 2 δεκαδικά ψηφία ή όσα είναι αναγκαία, όχι παραπάνω.
- Να προστίθενται γραμμές στις αναφορές ανά 5 στοιχεία, ώστε να είναι ευδιάκριτα τα νούμερα.
- Να φτιαχτεί ένα πινακάκι με descriptive statistics για τα δεδομένα τα οποία έχουμε, να περιλαμβάνει min, max, quantiles, median και boxplot.
- Να γίνει η παραπάνω παρουσίαση για τα αρχικά δεδομένα, αλλά και για τα κανονικοποιημένα δεδομένα.
- Πιθανώς να δούμε την διαφορά που θα είχε η χρήση κάποιου διαφορετικού τρόπου κανονικοποίησης. (Μέχρι τώρα η κανονικοποίηση γίνεται διαιρώντας με την μέγιστη τιμή. Ένας άλλος τρόπος θα ήταν ο $\frac{X - X_{min}}{X_{max} - X_{min}}$)

5.1 Πίνακας r^2 με βέλτιστα επιλεγμένα βάρη

Εδώ φαίνεται ο πίνακας της προηγούμενης εβδομάδας, με την διαφορά ότι κάθε χώρα μπορούσε να επιλέξει το "mix" δεδομένων που θα χρησιμοποιήσει ώστε να είναι βέλτιστο το r^2 . Εδώ βλέπουμε πως η Γαλλία έχει πολύ πιο μικρό πρόβλημα.

Ο αλγόριθμος ο οποίος χρησιμοποιήθηκε για να βγει αυτός ο πίνακας είναι πολύ κακός. Θεωρεί πως υπάρχει κάποια γραμμική συσχέτιση της ικανότητας μίας χώρας να περιγράψει τις άλλες και του βάρους κάθε διαφορετικού δεδομένου (για παράδειγμα, τα verified emissions έχουν μία γεννήτως αύξουσα σχέση με την ικανότητα της χώρας να περιγράψει τις άλλες.) Αυτό δεν ισχύει και ίσως είναι ενδιαφέρον το να δούμε πώς μοιάζει η συνάρτηση αυτή.

Πάντως ο αλγόριθμος αυτός κάνει τα παρακάτω:

1. Διαλέγει ένα από τα βάρη των διαφορετικών δεδομένων.
2. Δοκιμαστικά "ανεβάζει" και "κατεβάζει" το αντίστοιχο βάρος και βλέπει ποιο οδηγεί σε καλύτερη παλινδρόμηση.
3. Επαναλαμβάνει το προηγούμενο βήμα μέχρι να χτυπήσει κάποιο όριο ή να είναι χειρότερες και οι δύο επιλογές.
4. Προχωρά στο επόμενο βάρος το οποίο μπορεί να βελτιστοποιήσει.
5. Οι παράμετροι ήταν:
 - Ελάχιστο βάρος: 0
 - Μέγιστο βάρος: 1000
 - Βήμα: 10
 - Τιμή εκκίνησης: 50
 - Υπάρχουν 8 βάρη, τυχαία διαλέγει το βάρος το οποίο θα προσπαθήσει να βελτιστοποιήσει 100 φορές

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.99	0.98	0.98	0.98	0.99	0.98	0.98	0.98	0.98	0.97	0.97	0.00	0.00
Belgium	0.98	0.98	0.98	0.99	0.98	0.91	0.90	0.91	0.89	0.88	0.89	0.00	0.00
Bulgaria	0.98	0.98	0.98	0.98	0.99	0.97	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Cyprus	0.99	0.98	0.98	0.98	0.99	0.97	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Denmark	0.99	0.98	0.97	0.98	0.99	0.97	0.97	0.97	0.97	0.97	0.96	0.00	0.00
Estonia	0.99	0.98	0.98	0.98	0.98	0.96	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Finland	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.00	0.00
France	0.08	0.08	0.97	0.96	0.98	0.98	0.98	0.96	0.98	0.98	0.98	0.19	0.01
Germany	1.00	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.00	0.00
Greece	0.98	0.98	0.98	0.98	0.98	0.96	0.96	0.97	0.98	0.97	0.97	0.00	0.00
Hungary	0.99	0.97	0.97	0.98	0.98	0.97	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Ireland	0.99	0.98	0.97	0.99	0.99	0.97	0.97	0.97	0.97	0.96	0.96	0.00	0.00
Italy	0.96	0.93	0.97	0.98	0.99	0.96	0.96	0.97	0.97	0.97	0.96	0.00	0.00
Latvia	0.98	0.98	0.97	0.98	0.98	0.97	0.97	0.97	0.98	0.98	0.97	0.00	0.00
Lithuania	0.99	0.98	0.97	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.97	0.00	0.00
Luxembourg	0.98	0.97	0.97	0.98	0.98	0.96	0.96	0.97	0.97	0.97	0.97	0.00	0.00
Malta	0.99	0.98	0.97	0.98	0.99	0.97	0.97	0.97	0.97	0.98	0.97	0.00	0.00
Netherlands	0.97	0.97	0.97	0.97	0.98	0.81	0.81	0.80	0.79	0.73	0.69	0.00	0.00
Poland	0.00	0.02	0.04	0.04	0.07	0.94	0.94	0.94	0.95	0.95	0.94	0.92	0.02
Portugal	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.97	0.98	0.97	0.97	0.00	0.00
Romania	0.95	0.95	0.93	0.96	0.96	0.98	0.98	0.98	0.97	0.97	0.97	0.00	0.00
Slovenia	0.99	0.97	0.97	0.98	0.98	0.97	0.97	0.97	0.97	0.97	0.97	0.00	0.00
Spain	0.98	0.97	0.98	0.97	0.98	0.95	0.94	0.95	0.95	0.95	0.95	0.00	0.00
Sweden	0.99	0.98	0.97	0.98	0.99	0.97	0.98	0.98	0.97	0.97	0.97	0.00	0.00
United Kingdom	0.99	0.99	0.99	0.99	0.99	0.92	0.94	0.97	0.98	0.98	0.98	0.00	0.00

Σε αυτόν τον πίνακα τα μέσα βάρη ήταν:

- Population: 59.63
- GDP per capita: 19.85
- Inflation: 15.12
- Agriculture: 69.09
- Industry: 407.60
- Manufacturing: 12.69
- Total energy supply: 170.25
- Verified emissions: 849.81

5.2 Οπτικοποίηση δεδομένων

5.2.1 Πληθυσμός

- Source: <https://data.worldbank.org/indicator/SP.POP.TOTL>
- Year: 2011 - 2021
- Unit: Persons

Για τον πληθυσμό δεν έχω τίποτα εντυπωσιακό να πω. Ίσως είναι κάπως εύκολο να δούμε όλα τα δεδομένα ταυτόχρονα και να παρατηρήσουμε τι αλλαγή επιφέρει η κανονικοποίηση στα δεδομένα. Όπως φαίνεται στα figure 7 και 8. Τα σχήματα είναι ολόιδια. Η διαφορά είναι πως το ένα έχει διααιρεθεί με το 1000000 ενώ το άλλο έχει διααιρεθεί με το μέγιστο κάθε χρονιάς, οπότε είναι ελαφρώς διαφορετικές οι σχέσεις μεταξύ των αριθμών κάθε χώρας.

Population	min	25-quantile	median	75-quantile	max	Std
Austria	8.32	8.38	8.48	8.69	8.84	0.19
Belgium	10.71	10.97	11.16	11.30	11.43	0.24
Bulgaria	7.03	7.15	7.27	7.37	7.49	0.15
Cyprus	1.08	1.12	1.14	1.17	1.19	0.03
Denmark	5.49	5.56	5.61	5.71	5.79	0.10
Estonia	1.31	1.32	1.32	1.33	1.34	0.01
Finland	5.31	5.38	5.44	5.49	5.52	0.07
France	64.37	65.19	66.00	66.64	67.10	0.93
Germany	80.27	80.81	81.78	82.23	82.91	0.91
Greece	10.73	10.80	10.97	11.09	11.12	0.15
Hungary	9.78	9.83	9.89	9.99	10.04	0.09
Ireland	4.49	4.57	4.62	4.73	4.87	0.12
Italy	58.83	59.33	60.23	60.58	60.79	0.73
Latvia	1.93	1.97	2.01	2.08	2.18	0.08
Lithuania	2.80	2.89	2.96	3.06	3.20	0.13
Luxembourg	0.49	0.51	0.54	0.58	0.61	0.04
Malta	0.41	0.42	0.43	0.45	0.48	0.03
Netherlands	16.45	16.65	16.80	16.99	17.23	0.25
Poland	37.97	37.98	38.04	38.06	38.15	0.06
Portugal	10.28	10.34	10.46	10.56	10.57	0.12
Romania	19.47	19.76	19.98	20.20	20.54	0.33
Slovenia	2.02	2.05	2.06	2.06	2.07	0.01
Spain	45.95	46.46	46.58	46.68	46.80	0.24
Sweden	9.22	9.41	9.60	9.86	10.18	0.31
United Kingdom	61.81	63.01	64.13	65.36	66.46	1.55

Table 1: Population in millions 2008-2018

Παράλληλα, μπορούμε να παρατηρήσουμε τα δεδομένα κάπως καλύτερα, όπου φαίνεται πως η κανονικοποίηση δεν επηρεάζει καθόλου την σχέση μεταξύ των τιμών:



(a)



(b)

5.2.2 Πληθωρισμός

- Source: <https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG>

- Year: 1960-2021

- Unit: %

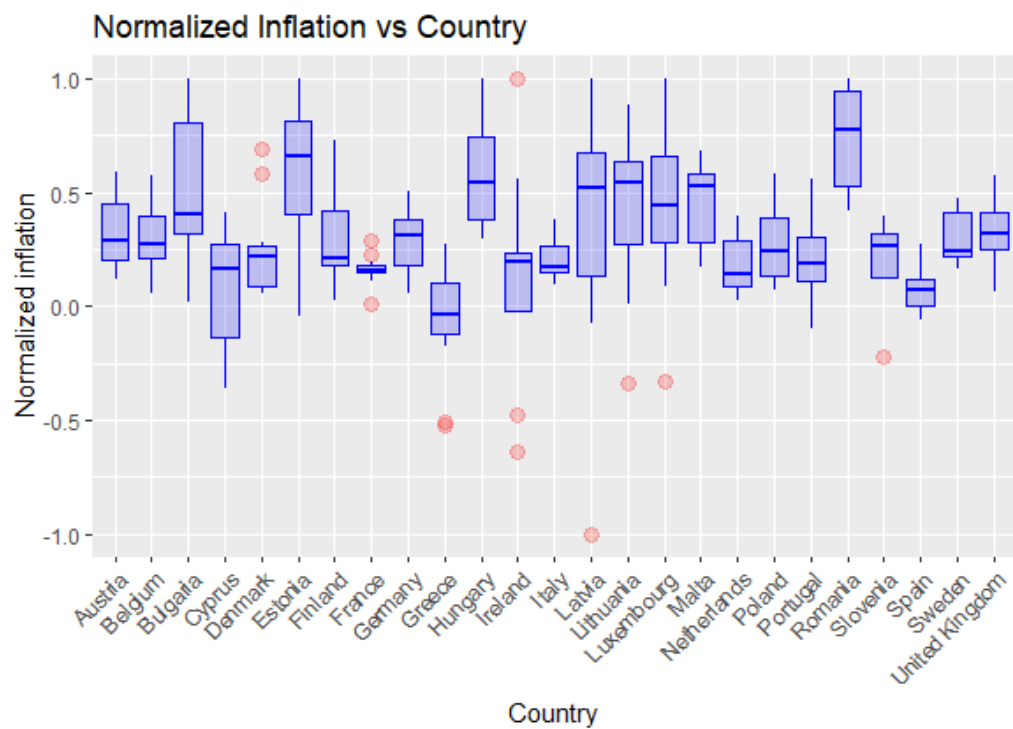
Inflation	min	25-quantile	median	75-quantile	max	Std
Austria	0.87	1.71	1.85	2.01	2.30	0.45
Belgium	0.53	1.30	1.81	1.90	1.96	0.47
Bulgaria	0.07	1.23	3.32	4.52	8.10	2.45
Cyprus	-1.33	-0.66	1.00	1.65	4.73	1.74
Denmark	0.25	0.58	0.89	1.78	4.13	1.27
Estonia	-0.39	2.00	3.59	4.05	6.80	2.02
Finland	0.09	1.22	1.77	2.59	3.04	1.02
France	0.07	0.55	0.95	1.10	2.37	0.58
Germany	0.65	1.20	1.50	1.87	1.97	0.46
Greece	-2.05	-0.44	-0.18	0.62	4.34	1.85
Hungary	1.32	2.66	2.89	4.11	4.85	1.15
Ireland	-4.62	-0.28	0.71	1.22	7.70	3.10
Italy	0.44	0.92	1.13	1.58	2.40	0.54
Latvia	-9.67	0.49	1.92	3.77	11.65	5.17
Lithuania	-3.30	1.06	2.53	3.88	9.71	3.30
Luxembourg	-1.11	1.94	2.28	3.36	6.61	1.99
Malta	1.12	2.06	2.22	2.75	4.22	0.82
Netherlands	0.19	0.35	0.94	1.36	2.44	0.79
Poland	0.30	0.75	1.65	2.82	3.89	1.34
Portugal	-0.39	0.67	1.51	1.78	2.25	0.90
Romania	1.80	3.33	3.77	4.38	16.02	3.88
Slovenia	-1.03	0.68	1.04	1.86	4.47	1.49
Spain	-0.22	0.06	0.32	0.90	2.25	0.76
Sweden	0.93	1.04	1.74	2.25	3.24	0.75
United Kingdom	0.51	1.60	1.82	2.03	3.23	0.65

Table 2: Πληθωρισμός μεταξύ 2008-2018

Εδώ βέβαια ίσως έχει σημασία να δούμε λίγο τι συμβαίνει στα δεδομένα με την κανονικοποίηση. Ούτε εδώ αλλάζουν σημαντικά οι σχέσεις μεταξύ τους.



(a)



(b)

5.2.3 Total energy Supply

- Source: https://ec.europa.eu/eurostat/databrowser/view/nrg_bal_s_1/default/table?lang=en
- Year: 2011 - 2020
- Unit: Thousand tonnes of oil equivalent

	min	25-quantile	median	75-quantile	max	Std
Austria	32011.29	33031.17	33177.75	33562.30	34166.28	660.54
Belgium	52238.38	53015.79	55039.87	55272.03	59313.06	2255.37
Bulgaria	16923.38	17726.43	18234.79	18722.35	19823.90	813.92
Cyprus	1955.96	2121.56	2262.44	2440.09	2616.93	221.87
Denmark	16374.16	16838.22	17383.65	18304.45	19558.31	1128.72
Estonia	4399.62	5378.58	5648.48	5846.32	5978.41	496.84
Finland	32022.99	33165.71	33582.86	34469.06	36251.74	1195.84
France	248383.41	250072.60	256292.91	259845.13	266394.54	6206.33
Germany	305036.83	310746.75	313107.91	318940.66	335474.27	9403.81
Greece	22748.62	23240.71	23407.97	27231.17	30404.91	2854.85
Hungary	23652.51	24816.51	25609.63	26395.50	26900.70	1096.99
Ireland	12776.63	13293.33	13625.33	14225.86	15022.14	727.58
Italy	146769.88	152859.01	156093.49	168758.58	181736.20	10969.89
Latvia	4259.46	4307.46	4407.65	4465.42	4640.00	129.86
Lithuania	6946.51	7067.95	7290.53	7647.47	9553.50	819.46
Luxembourg	3682.05	3785.50	3948.40	4127.44	4212.94	195.48
Malta	594.33	687.68	780.25	834.21	881.48	97.20
Netherlands	71379.09	73882.75	75619.51	77123.41	82743.78	3132.34
Poland	93773.11	96236.75	97971.41	101131.11	108970.23	4555.35
Portugal	21439.29	22060.46	22651.14	23433.93	24716.20	1092.37
Romania	31378.66	31668.07	33454.78	34845.41	39485.27	2432.37
Slovenia	6473.51	6722.90	6875.85	7127.16	7982.69	419.35
Spain	114522.76	119335.11	125486.61	126434.02	138166.04	6411.76
Sweden	44092.81	48432.69	49103.98	49874.22	50118.89	1857.82
United Kingdom	174024.39	177000.66	186386.17	191789.70	208268.88	11384.30

Table 3: Total energy supply μεταξύ 2008-2018

5.2.4 GDP per capita

- Source: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- Year: 1960 - 2021
- Unit: US\$

	min	25-quantile	median	75-quantile	max	Std
Austria	44.20	47.17	48.56	51.46	51.92	2.75
Belgium	41.01	44.19	44.76	47.48	48.30	2.43
Bulgaria	6.85	7.17	7.57	7.88	9.45	0.74
Cyprus	23.41	26.89	28.91	31.57	35.40	3.57
Denmark	53.25	57.83	58.51	61.67	64.32	3.39
Estonia	14.66	17.40	18.20	19.66	23.06	2.45
Finland	42.80	46.46	47.71	50.16	53.77	3.25
France	36.65	39.73	41.59	42.84	45.52	2.75
Germany	41.10	41.89	44.65	46.50	48.02	2.61
Greece	17.92	19.17	21.79	26.10	32.13	4.84
Hungary	12.72	13.09	13.72	14.46	16.43	1.21
Ireland	48.66	51.83	55.60	62.44	79.11	9.57
Italy	30.24	33.51	35.56	36.63	40.94	3.17
Latvia	11.42	13.56	14.33	15.72	17.87	1.87
Lithuania	11.82	14.32	14.94	16.14	19.19	2.11
Luxembourg	105.46	109.81	112.58	119.51	123.68	6.14
Malta	21.08	22.37	24.77	26.19	31.57	3.23
Netherlands	45.19	49.37	52.20	52.97	57.88	3.66
Poland	11.53	12.60	13.70	13.94	15.47	1.08
Portugal	19.25	21.03	22.10	23.18	24.95	1.68
Romania	8.21	8.76	9.55	10.24	12.40	1.23
Slovenia	20.89	23.07	23.53	24.96	27.60	1.92
Spain	25.74	28.25	29.50	31.11	35.51	2.74
Sweden	46.95	52.42	54.59	59.03	61.13	4.45
United Kingdom	38.95	41.18	42.69	44.56	47.79	2.92

Table 4: GDP per capita in thousands USD

5.2.5 Verified emissions

- Source: EU ETS Database
- Table: eutl.compliance
- Column: verified

Από όποια μονάδα και να ήταν, εδώ πάνε 1.000.000 φορές κάτω:

	min	25-quantile	median	75-quantile	max	Std
Austria	27.36	29.61	30.51	30.87	32.08	1.27
Belgium	45.03	45.17	45.99	46.29	55.46	3.16
Bulgaria	31.28	33.24	34.57	35.95	40.00	2.63
Cyprus	4.19	4.58	4.61	4.92	5.58	0.41
Denmark	15.50	17.05	19.46	23.71	26.55	4.05
Estonia	10.38	13.50	13.89	14.76	16.00	1.55
Finland	26.18	27.81	30.68	34.72	41.30	4.74
France	101.40	104.86	110.90	114.46	124.13	7.10
Germany	428.29	448.66	462.35	469.31	489.86	18.24
Greece	47.34	50.74	58.84	61.09	69.85	7.10
Hungary	20.08	21.23	22.40	22.61	27.24	1.91
Ireland	15.77	18.88	23.63	27.08	28.53	4.72
Italy	148.37	157.09	166.78	187.42	220.68	21.98
Latvia	2.43	2.59	2.74	2.96	3.24	0.26
Lithuania	5.61	5.92	6.23	6.65	7.56	0.59
Luxembourg	1.73	1.83	2.06	2.17	3.62	0.52
Malta	0.84	1.09	1.89	1.92	2.28	0.51
Netherlands	79.97	82.27	89.14	92.79	96.47	6.31
Poland	191.17	198.28	199.73	203.07	206.35	4.12
Portugal	24.17	25.64	26.99	28.75	31.42	2.23
Romania	40.53	42.21	43.07	48.72	63.82	6.83
Slovenia	6.18	6.60	7.45	8.03	8.86	0.91
Spain	121.48	128.26	132.69	140.52	163.46	11.31
Sweden	17.49	21.05	22.51	22.63	22.86	1.71
United Kingdom	141.76	172.90	220.88	236.57	265.06	42.22

Table 5: Verified emissions

Εδώ επειδή δεν φαίνεται τίποτα, θα βάλω log απλά για να μπορούμε να διακρίνουμε κάτι, χωρίς κάποια κανονικοποίηση.



(a)



(b)

5.2.6 Agriculture

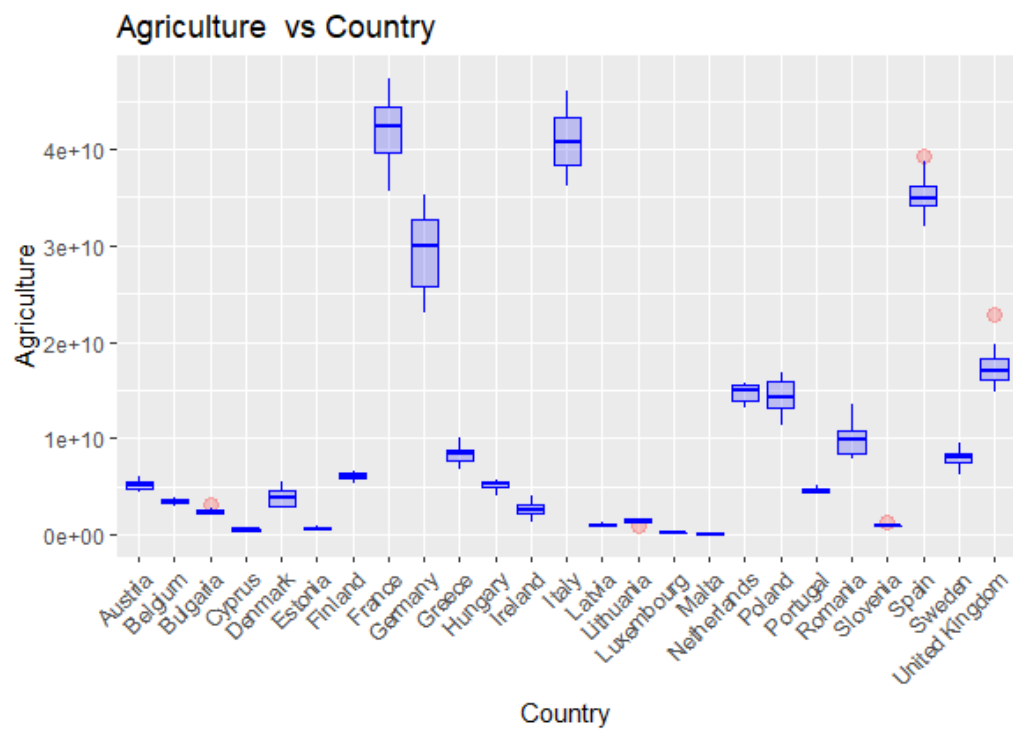
- Source: <http://wdi.worldbank.org/table/4.2>

- Year: 2020

- Unit: Agricultural in Billions USD.

	min	25-quantile	median	75-quantile	max	Std
Austria	4.32	4.79	5.14	5.46	6.05	0.55
Belgium	2.99	3.22	3.39	3.62	3.88	0.27
Bulgaria	2.04	2.20	2.39	2.56	3.20	0.33
Cyprus	0.37	0.44	0.49	0.52	0.59	0.07
Denmark	2.68	3.02	3.87	4.49	5.41	0.91
Estonia	0.49	0.63	0.65	0.80	0.85	0.13
Finland	5.29	5.82	6.03	6.41	6.59	0.43
France	35.54	39.60	42.43	44.42	47.28	3.91
Germany	22.99	25.78	29.91	32.78	35.20	4.54
Greece	6.79	7.71	8.33	8.85	9.99	0.90
Hungary	4.01	4.87	5.31	5.51	5.67	0.60
Ireland	1.32	2.28	2.61	3.06	3.96	0.71
Italy	36.20	38.32	40.70	43.21	45.94	3.17
Latvia	0.85	0.96	0.97	1.07	1.24	0.11
Lithuania	0.95	1.38	1.55	1.66	1.71	0.24
Luxembourg	0.13	0.14	0.15	0.17	0.20	0.02
Malta	0.09	0.10	0.11	0.11	0.13	0.01
Netherlands	13.20	13.89	15.07	15.49	15.71	0.92
Poland	11.25	13.13	14.22	15.96	16.76	1.91
Portugal	4.16	4.38	4.64	4.77	5.17	0.32
Romania	7.87	8.47	9.88	10.88	13.51	1.74
Slovenia	0.89	0.91	0.92	1.01	1.22	0.10
Spain	31.90	34.18	34.84	36.07	39.19	2.23
Sweden	6.27	7.53	8.11	8.47	9.59	0.87
United Kingdom	14.74	16.11	17.05	18.33	22.84	2.28

Table 6: Agriculture



(a)

5.2.7 Industry

- Source: <http://wdi.worldbank.org/table/4.2>
- Year: 2020
- Unit: Industry in Billions USD

	min	25-quantile	median	75-quantile	max	Std
Austria	96.16	102.24	106.03	111.06	116.48	6.50
Belgium	90.94	98.13	100.59	104.53	111.72	6.42
Bulgaria	12.04	13.16	13.62	14.01	14.87	0.87
Cyprus	2.03	2.45	2.95	3.63	4.98	0.90
Denmark	60.50	64.40	68.89	69.76	79.84	5.40
Estonia	4.61	5.63	5.97	6.42	7.36	0.81
Finland	54.66	61.38	63.49	65.70	84.47	7.78
France	431.14	459.74	479.72	506.08	551.30	36.97
Germany	843.80	934.68	1000.01	1014.25	1085.27	69.42
Greece	27.44	28.64	35.83	42.48	55.73	9.66
Hungary	32.23	32.95	33.99	36.13	40.79	2.98
Ireland	51.79	57.18	62.99	110.85	141.75	32.57
Italy	383.06	431.67	449.66	474.26	568.48	50.60
Latvia	4.90	5.37	5.81	6.01	7.81	0.78
Lithuania	9.38	11.12	12.22	13.05	13.97	1.52
Luxembourg	5.99	6.42	6.93	7.48	7.80	0.62
Malta	1.36	1.48	1.54	1.59	1.84	0.14
Netherlands	138.29	155.19	166.89	173.29	204.75	19.28
Poland	129.76	146.41	150.02	156.35	169.32	10.65
Portugal	38.81	41.82	43.54	47.46	53.70	4.42
Romania	54.81	61.02	63.42	67.76	78.81	7.80
Slovenia	12.08	12.77	13.77	13.95	16.62	1.31
Spain	240.11	268.98	278.66	328.16	429.02	57.54
Sweden	97.80	114.87	122.21	126.94	135.67	10.52
United Kingdom	448.57	473.78	494.10	522.75	583.44	40.55

Table 7: Industry

5.2.8 Manufacturing

- Source: <http://wdi.worldbank.org/table/4.2>
- Year: 2020
- Unit: Manufacturing in Billions USD.

Εδώ φαίνεται εμφανώς πως έχω κάνει ένα τεράστιο λάθος. Λείπουν οι τιμές για το Manufacturing στην Βουλγαρία και νόμιζα πως το είχα βάλει να παίρνει έναν μέσο όρο και πως αυτό δεν θα άλλαζε σημαντικά τα αποτελέσματα. Όμως είναι εμφανές πως η Βουλγαρία όχι μόνο δεν έπαιρνε τις τιμές που ήθελα, αλλά και να συνέβαινε αυτό είναι πολύ πιθανό και πάλι να μην ήταν αρκετά λογικό το αποτέλεσμα. Η Βουλγαρία μάλλον θα πρέπει να αφαιρεθεί από τις δοκιμές, μαζί με χώρες όπως η Σλοβακία.

	min	25-quantile	median	75-quantile	max	Std
Austria	63.75	66.61	70.28	72.47	76.57	4.24
Belgium	58.69	62.46	64.21	66.70	72.39	4.02
Bulgaria	-470.74	-453.31	-432.37	-420.07	-399.33	22.79
Cyprus	0.85	0.95	1.12	1.33	1.52	0.23
Denmark	35.21	37.47	40.47	41.64	46.70	3.45
Estonia	2.41	3.20	3.35	3.56	4.13	0.47
Finland	34.44	38.11	39.73	42.29	59.40	6.76
France	254.30	268.11	278.31	292.38	325.40	20.87
Germany	603.23	697.07	743.97	755.59	796.43	56.32
Greece	15.64	16.78	18.19	22.95	30.27	4.63
Hungary	22.45	24.56	25.39	27.39	29.71	2.23
Ireland	43.26	47.00	49.10	100.06	126.39	31.76
Italy	264.39	290.81	301.69	308.97	372.63	28.24
Latvia	2.57	2.85	3.20	3.32	3.63	0.32
Lithuania	5.65	7.20	8.00	8.15	8.91	0.96
Luxembourg	2.53	2.88	3.14	3.41	3.93	0.42
Malta	0.83	0.96	1.00	1.06	1.23	0.11
Netherlands	82.70	89.39	91.54	95.12	109.07	7.43
Poland	71.81	82.05	85.70	88.28	98.64	7.14
Portugal	24.20	25.62	27.24	27.71	31.35	2.12
Romania	35.95	37.70	39.72	44.53	49.91	4.70
Slovenia	8.43	8.69	9.28	9.90	11.00	0.87
Spain	135.09	148.07	155.06	166.17	207.17	19.84
Sweden	60.02	69.65	72.99	77.49	83.80	6.77
United Kingdom	218.33	242.14	251.42	268.16	286.56	20.19

Table 8: Manufacturing

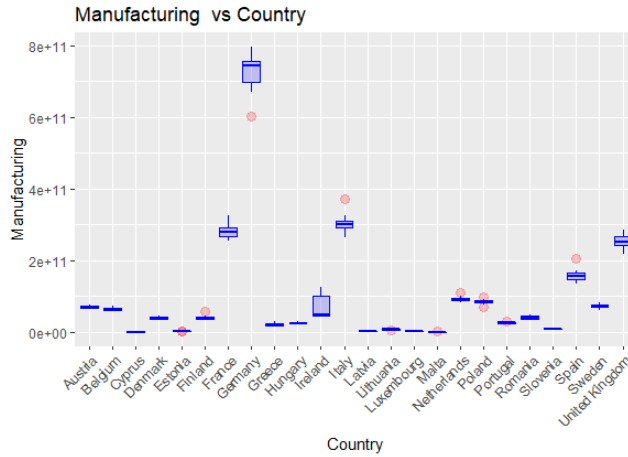


Figure 9

6 Για την 21η Μαρτίου

Είχαμε πει, να βάλουμε κάποια πρόταση για δουλειά, ώστε να αρχίσει να ετοιμάζεται το paper. Αυτά που έπρεπε να κάνω εγώ είναι:

- Να δούμε αν τα free ήταν το πρόβλημα για την πολωνία και την Γαλλία για το 2012 και το 2013.
- Να βρεθούν βάρη τα οποία να λειτουργούν για όλους καλά. Χωρίς να αλλάζουν κάθε φορά
- Διάβασμα για Nass Welfare
- Να αναφερθούν ελλείψεις στα data

6.1 Γιατί η Πολωνία και η Γαλλία αποτυγχάνουν να εξηγήσουν τις άλλες πριν το 2012;

Τελικά η απάντηση ήταν απίστευτα απλή. Κάτι πάει λάθος με τις δωρεάν άδειες, ή απλά έπαιρναν υπέρβολικά πολλές στην προηγούμενη φάση. όλες οι χώρες έχασαν πολλές δωρεάν άδειες από το 2012 στο 2013. Αλλά ας δούμε τι ποσοστό έχασαν και όλα θα βγάλουν νόημα. Spoiler, δεν βγάζει. Όχι μόνο δεν είναι παό τις πιο "ακραίες", αλλά μάλιστα είναι και σχετικά στην μέση.:

Country	2012 free in Millions	2013 free in Millions	Percentage drop in %
Malta	2.37	0.18	-92.52
Cyprus	6.70	1.15	-82.89
Estonia	14.30	3.11	-78.25
Greece	65.92	14.50	-78.01
Bulgaria	43.08	10.68	-75.22
United Kingdom	283.33	72.30	-74.48
Slovenia	8.31	2.32	-72.04
Poland	213.47	63.63	-70.20
Luxembourg	4.80	1.45	-69.79
Romania	75.61	24.06	-68.18
Portugal	35.04	12.37	-64.69
Ireland	28.76	10.54	-63.36
Germany	467.32	173.23	-62.93
Spain	163.59	67.11	-58.98
Italy	197.62	87.81	-55.56
Hungary	25.79	12.24	-52.54
Denmark	25.17	12.65	-49.72
Netherlands	99.37	50.24	-49.44
Latvia	5.30	2.77	-47.83
France	159.96	84.38	-47.25
Finland	40.28	23.05	-42.77
Belgium	61.61	37.56	-39.04
Austria	35.38	22.88	-35.33
Lithuania	8.43	6.55	-22.33
Sweden	25.72	30.21	17.47

Άρα καμία απάντηση ακόμα...

6.2 Τα βέλτιστα βάρη για όλα τα σενάρια

Το πρώτο βήμα είναι λογικά να δοκιμαστούν τα πρώτα τα βάρη που προέκυψαν ως μέσος όρος όλων των βέλτιστων βαρών. Σε δεύτερο βήμα, πιθανότατα έχει νόημα να δοκιμάσουμε να μεγιστοποιήσουμε το γινόμενο όλων.

- Population: 59.63
- GDP per capita: 19.85
- Inflation: 15.12
- Agriculture: 69.09
- Industry: 407.60
- Manufacturing: 12.69
- Total energy supply: 170.25
- Verified emissions: 849.81

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	max p-value	max MSE
Austria	0.94	0.93	0.94	0.94	0.95	0.97	0.97	0.97	0.96	0.95	0.95	0.00	0.00
Belgium	0.91	0.90	0.93	0.93	0.93	0.86	0.83	0.82	0.80	0.78	0.78	0.00	0.00
Bulgaria	0.92	0.91	0.93	0.92	0.94	0.96	0.96	0.97	0.95	0.96	0.96	0.00	0.00
Cyprus	0.95	0.94	0.94	0.95	0.96	0.95	0.94	0.96	0.96	0.96	0.96	0.00	0.00
Denmark	0.95	0.94	0.93	0.94	0.96	0.96	0.96	0.96	0.96	0.95	0.95	0.00	0.00
Estonia	0.94	0.94	0.94	0.94	0.95	0.94	0.95	0.96	0.95	0.96	0.96	0.00	0.00
Finland	0.94	0.93	0.94	0.94	0.95	0.96	0.97	0.96	0.97	0.96	0.95	0.00	0.00
France	0.58	0.53	0.73	0.77	0.83	0.90	0.86	0.85	0.85	0.90	0.89	0.00	0.01
Germany	0.92	0.92	0.93	0.94	0.95	0.97	0.96	0.95	0.95	0.94	0.93	0.00	0.00
Greece	0.87	0.88	0.89	0.89	0.91	0.94	0.94	0.96	0.97	0.96	0.96	0.00	0.00
Hungary	0.94	0.92	0.93	0.94	0.95	0.96	0.95	0.97	0.96	0.96	0.96	0.00	0.00
Ireland	0.95	0.94	0.93	0.95	0.95	0.96	0.96	0.95	0.95	0.94	0.94	0.00	0.00
Italy	0.83	0.80	0.87	0.91	0.93	0.88	0.90	0.94	0.94	0.94	0.93	0.00	0.00
Latvia	0.94	0.94	0.94	0.94	0.95	0.96	0.96	0.96	0.96	0.97	0.96	0.00	0.00
Lithuania	0.94	0.94	0.94	0.94	0.95	0.97	0.97	0.97	0.97	0.97	0.96	0.00	0.00
Luxembourg	0.94	0.93	0.93	0.94	0.95	0.94	0.94	0.95	0.94	0.95	0.95	0.00	0.00
Malta	0.94	0.93	0.94	0.94	0.96	0.96	0.96	0.96	0.96	0.97	0.96	0.00	0.00
Netherlands	0.88	0.87	0.88	0.86	0.88	0.76	0.74	0.71	0.72	0.67	0.62	0.00	0.01
Poland	0.19	0.21	0.28	0.35	0.45	0.81	0.86	0.89	0.93	0.93	0.92	0.04	0.01
Portugal	0.94	0.93	0.94	0.94	0.95	0.96	0.97	0.97	0.96	0.96	0.96	0.00	0.00
Romania	0.84	0.82	0.83	0.86	0.87	0.96	0.98	0.98	0.96	0.95	0.96	0.00	0.00
Slovenia	0.95	0.93	0.93	0.94	0.95	0.96	0.96	0.96	0.96	0.96	0.96	0.00	0.00
Spain	0.93	0.92	0.89	0.88	0.90	0.80	0.87	0.90	0.88	0.92	0.90	0.00	0.00
Sweden	0.94	0.93	0.93	0.93	0.94	0.94	0.94	0.94	0.94	0.93	0.93	0.00	0.00
United Kingdom	0.92	0.94	0.95	0.94	0.94	0.84	0.89	0.89	0.96	0.94	0.94	0.00	0.00

Εδώ πλέον γίνεται πολύ αμφισβητήσιμο το αν χρειάζεται να ψάξουμε για κάτι καλύτερο. Με εξαίρεση την Πολωνία και την Γαλλία, οι υπόλοιποι δείχνουν να τα πηγαίνουν πολύ καλά.

7 Για την 28η Μαρτίου

Είπαμε να γίνουν τα παρακάτω:

- Να δούμε αν η Πολωνία είναι τόσο ανώμαλη, εξαιτίας του κανόνα 10c.
- Να προστεθούν στα features ξεχωριστά η πράσινη ενέργεια και ξεχωριστά η υπόλοιπη ενέργεια.
- Να οριστεί και να προσομοιωθεί το παρακάτω πρόβλημα:

$$- \max \sum u_i(x_i) * GDP_i$$

$$- \sum x_i \leq \text{Cap.}$$

- Όπου x_i είναι το allocation στην χώρα i .

- Η u_i είναι ξεχωριστή για κάθε χώρα και δηλώνει την ικανοποίηση που λαμβάνει η αντίστοιχη χώρα με το αντίστοιχο allocation.

- Πιθανότατα η u_i θα μπορούσε να είναι το energy intensity.
- Μετά μπορούν να προστεθούν επιπλέον constraints
- Να προστεθεί το energy Intensity στα features της κάθε χώρας.

7.1 Τι όντως έχανα

- Δεν υλοποίησα κανένα κομμάτι του κώδικά.
- Το power intensity δεν είναι απλώς μία τιμή.

7.2 PPS (Purchasing Power Standards)

Το PPS είναι ένα εικονικό νόμισμα. Το οποίο προσπαθεί να πετύχει να είναι αποπληθωρισμένο, ενώ ταυτόχρονα προσπαθεί να αφαιρέσει διαφορές λόγω άλλης αγοραστικής δύναμης.

7.2.1 Χρήση PPS

Το PPS χρησιμοποιείται προκειμένου να κάνουμε συγκρίσεις μεταξύ διαφορετικών χωρών. Το βρίσκουμε συνήθως στα dataset της eurostat ως MPPS (Million Purchasing Power Standarts).

7.2.2 Υπολογισμός

Ένα PPS ισοδυναμεί με ένα καλάθι στην εκάστοτε χώρα. Το καλάθι αυτό έχει ένα μείγμα από διάφορα προϊόντα και υπηρεσίες. Το κόστος αυτών συμπεριλαμβάνεται σε έναν σταθμισμένο μέσο, εκ του οποίου προκύπτει μία τιμή, η οποία είναι το κόστος του καλαθιού. Στην συνέχεια, μπορούμε να χρησιμοποιήσουμε αυτήν την τιμή για να κανονικοποιήσουμε διαφορετικές "αξίες" μεταξύ διαφορετικών χωρών ώστε να μπορούμε να κάνουμε συγκρίσεις με βάση την αγοραστική δύναμη και όχι απλά ευρώ.

7.3 Power Intensity

Energy intensity can be considered as an approximation of the energy efficiency of a country's economy, and shows the amount of energy needed to produce a unit of GDP. There are various reasons for observed improvements in energy intensity: a general shift from industry towards a service-based economy in Europe, a shift within industry to less energy-intensive activities and production methods, the closure of inefficient units, and more energy-efficient appliances.

To Power Intensity υπολογίζεται ως

$$Power_Intensity = \frac{Units_of_Energy}{Units_of_GDP}$$

Αυτό λοιπόν μπορούμε να το υπολογίσουμε με δύο διαφορετικούς τρόπους.

- Chain Linked Volumes. Χρησιμοποιείται για να κάνουμε συγκρίσεις για μία χώρα σε βάθος χρόνου. Εδώ είναι σημαντικό το να σημειωθεί πως είναι πιθανό να μην αρκεί μία τιμή για να κάνουμε την δουλειά μας. Οι τιμές προκύπτουν σε επίπεδο έτους. Όμως οι οικονομικές αλλαγές στις χώρες (όπως αλλαγές σε φορολογίες) δημιουργούν μεγάλες ανωμαλίες στα δεδομένα, επομένως είναι καλύτερο να χρησιμοποιούνται περισσότερα από 4 data points.
- Η άλλη επιλογή είναι το να χρησιμοποιήσουμε τα δεδομένα σε PPS, αυτή η επιλογή μας επιτρέπει να βγάλουμε συμπεράσματα μεταξύ χωρών.

7.4 Για την επόμενη φορά

7.4.1 Σε δεύτερο χρόνο

- Ορίζουμε ένα πρόβλημα βελτιστοποίησης. Χρησιμοποιούμε το power intensity ως το εργαλείο το οποίο θα μας εξηγήσει ποια χώρα είναι σε θέση να διαχειριστεί με πιο αποτελεσματικό τρόπο τους ρίπους. Μας ενδιαφέρει το power intensity όπως αυτό βρίσκεται στην eurostat στο NRG_IND_EI, από όπου θα αναζητήσουμε την έτοιμη τιμή PPS.

Στην συνέχεια θα υλοποιηθούν τα παρακάτω:

- Vanilla version. Θα είναι λογικά ίδιο με αυτό που είχαμε κάνει στο παρελθόν, μόνο που εκεί αντί για το power intensity είχαμε τον λόγο $\frac{GDP}{verified}$.
 - Διάφορα constraints και συνδυασμοί αυτών.
 - Να πλησιάσουμε να βρούμε συνδυασμό από περιορισμούς οι οποίοι να προσεγγίζει τις τιμές που προκύπτουν από το EU ETS.
- Διάβασμα το "Cap and Trade and Emission Clustering : A spatial-temporal analysis of the EU ETScheme".

Ορίζουμε ένα πρόβλημα βελτιστοποίησης. Χρησιμοποιούμε το power intensity ως το εργαλείο το οποίο θα μας εξηγήσει ποια χώρα είναι σε θέση να διαχειριστεί με πιο αποτελεσματικό τρόπο τους ρίπους. Μας ενδιαφέρει το power intensity όπως αυτό βρίσκεται στην eurostat στο NRG_IND_EI, από όπου θα αναζητήσουμε την έτοιμη τιμή PPS.

Στην συνέχεια θα υλοποιηθούν τα παρακάτω:

- Vanilla version. Θα είναι λογικά ίδιο με αυτό που είχαμε κάνει στο παρελθόν, μόνο που εκεί αντί για το power intensity είχαμε τον λόγο $\frac{GDP}{verified}$.
- Διάφορα constraints και συνδυασμοί αυτών.
- Να πλησιάσουμε να βρούμε συνδυασμό από περιορισμούς οι οποίοι να προσεγγίζει τις τιμές που προκύπτουν από το EU ETS.

7.4.2 Άμεσα, για το τετρασέλιδο

Για κάθε χώρα έχουμε έναν συνδυασμό από features. Με βάση αυτά τα δεδομένα μπορούμε να τις ομαδοποιήσουμε σε 3 ή 4 κατηγορίες οι οποίες θα έχουν κοινά χαρακτηριστικά. Έστερα, να απαντηθούν τα παρακάτω ερωτήματα:

- Να προστεθούν και άλλα features για τις χώρες αυτές, όπως είναι το power intensity, το ποσοστό της πράσινης ενέργειας που παράγουν.
- Οι χώρες οι οποίες είμαι μέσα στο ίδιο cluster μπορούν να εξηγήσουν καλά τις υπόλοιπες του cluster τους;
- Μήπως χώρες διαφορετικών χωρών μπορούν να εξηγήσουν καλύτερη η μία την άλλη; (Υπόθεση Κώστα, μπορεί να ελεγχθεί τρέχοντας το ίδιο πρόβλημα, αλλά αφαιρώντας όλες τις άλλες χώρες του ίδιου cluster)
- Αν πάρω την χώρα που δρα ως κέντρο σε κάθε cluster τι προκύπτει;
- Μήπως μέσα στο ίδιο cluster παρατηρείται κάποια απλή σχέση των free με κάποιον από τα features ή με κάποιο συνδυασμό αυτών;

8 Για τις 4 Απριλίου

Καλό μήνα!

8.1 Συσταδοποίηση (έπρεπε...)

8.1.1 Πλήθος συστάδων

Στην R υπάρχει ένα πολύ όμορφο εργαλείο το οποίο προτείνει τον αριθμό των βέλτιστων clusters με 30 διαφορετικούς δείκτες (elbow, silhouette κλπ). Με βάση τα δεδομένα μας, νομίζω πως οι 3 συστάδες είναι η βέλτιστη λύση.

* Among all indices:

- * 6 proposed 2 as the best number of clusters
- * 9 proposed 3 as the best number of clusters
- * 1 proposed 4 as the best number of clusters
- * 1 proposed 5 as the best number of clusters
- * 5 proposed 9 as the best number of clusters
- * 2 proposed 10 as the best number of clusters

***** Conclusion *****

* According to the majority rule, the best number of clusters is 3 Συγκεκριμένα, για όλες τις διαφορετικές μεθόδους προκύπτουν τα παρακάτω:

Clusters	KL	CH	Hartigan	CCC	Scott	Marriot	TrCovW	TraceW	Friedman	Rubin	Cindex	DB	Silhouette
2	4.16	30.55	10.06	5.49	140.15	0.00	0.91	5.56	221.35	4.54	0.35	0.76	0.56
3	6.73	25.81	2.74	5.51	191.04	0.00	0.30	3.87	252.61	6.53	0.27	0.90	0.41
4	0.13	19.34	13.76	4.39	219.59	0.00	0.25	3.44	276.56	7.34	0.26	1.27	0.31
5	4.75	26.15	3.77	6.65	303.07	0.00	0.12	2.08	590.66	12.16	0.33	0.83	0.35
6	3.09	24.33	1.65	6.22	370.12	0.00	0.09	1.75	1,054.45	14.45	0.30	0.70	0.40
7	0.16	21.14	7.77	5.22	383.36	0.00	0.08	1.61	1,044.11	15.70	0.30	0.79	0.34
8	1.56	25.54	6.22	6.54	459.86	0.00	0.03	1.12	1,440.67	22.48	0.24	0.82	0.32
9	3.17	29.46	2.31	7.46	587.08	0.00	0.02	0.82	3,137.27	30.71	0.20	0.70	0.42
10	5.57	28.34	0.68	6.99	613.29	0.00	0.02	0.72	3,743.78	35.14	0.19	0.71	0.38

Clusters	Duda	Pseudot2	Beale	Ratkowsky	Ball	Ptbiserial	Frey	McClain	Dunn	Hubert	SDindex	Dindex	SDbw
2	0.56	14.33	3.99	0.47	2.78	0.76	2.96	0.20	0.40	0.16	8.57	0.42	0.64
3	0.74	4.29	1.74	0.46	1.29	0.59	6.15	0.66	0.14	0.16	8.25	0.33	0.45
4	0.63	2.99	2.37	0.42	0.86	0.43	-0.07	1.53	0.14	0.16	10.58	0.31	0.54
5	2.03	-3.05	-2.30	0.41	0.42	0.46	0.36	1.46	0.21	0.18	8.82	0.25	0.21
6	1.03	-0.17	-0.13	0.38	0.29	0.46	14.61	1.58	0.21	0.19	9.19	0.22	0.15
7	0.35	14.73	7.30	0.36	0.23	0.41	0.10	2.04	0.14	0.19	9.54	0.21	0.15
8	3.69	-3.64	-2.57	0.34	0.14	0.41	0.35	2.04	0.21	0.20	9.06	0.18	0.13
9	5.57	-3.28	-0.00	0.32	0.09	0.39	1.45	2.27	0.32	0.21	11.70	0.15	0.08
10	5.93	-1.66	-2.93	0.31	0.07	0.36	-1.40	2.77	0.25	0.21	11.91	0.14	0.07

8.1.2 Βέλτιστες συστάδες



Χώρα	partition
France	1
Germany	1
Italy	1
Spain	1
United Kingdom	1
Bulgaria	2
Estonia	2
Hungary	2
Latvia	2
Lithuania	2
Romania	2
Austria	3
Belgium	3
Cyprus	3
Denmark	3
Finland	3
Greece	3
Ireland	3
Luxembourg	3
Malta	3
Netherlands	3
Poland	3
Portugal	3
Slovenia	3
Sweden	3
EUROZONE	

Figure 10: Συστάδες

Είναι λοιπόν σαν να έχουμε 3 βασικές κατηγορίες και σε αυτό ίσως φτάνει τα δεδομένα τα οποία έχω χρησιμοποιήσει. Μοιάζει σαν να είναι:

- Εκβιομηχανισμένες χώρες
- Μέρος του ανατολικού μπλοκ
- Υπόλοιποι

8.1.3 Συστάδες, αν όλα τα δεδομένα ήταν κατά κεφαλήν και κανονικοποιημένα ύστερα

Αυτό δεν είχαμε πει να γίνει, όμως ένοιωσα την ανάγκη να το δοκιμάσω.

Εδώ λοιπόν φαίνεται πως κάτι πήγε πολύ λάθος. Δεν ξέρω τι κοινό έχει η πρώτη συστάδα. Μπορώ να δεχτώ πως η Βουλγαρία είναι μία κατηγορία από μόνης της, αλλά τα μπλε δεν ξέρω πώς σχετίζονται. Το περίεργο είναι πως δεν είναι ο πληθυσμός η ειδοποιός διαφορά, καθώς το Λουξεμβούργο, η Μάλτα και η Βουλγαρία είναι σε 3 διαφορετικές ομάδες. Σημείωση, αυτή τη φορά 15 από τα 23 κριτήρια πρότειναν τις 3 συστάδες.



Figure 11: Συστάδες αλλά όλα τα δεδομένα είναι κατά κεφαλήν.

8.2 Ικανότητα εξήγησης υπολοίπων μέσα στην ίδια συστάδα

Θα διατηρήσουμε τις συστάδες έτσι όπως διαμορφώθηκαν από την πρώτη διαδικασία:

Εκβιομηχανισμένες	Μισό ανατολικό μπλοκ	Λοιποί
Γαλλία	Βουλγαρία	Αυστρία
Γερμανία	Εσθονία	Βέλγιο
Ιταλία	Ουγγαρία	Κύπρος
Ισπανία	Λετονία	Δανία
Ηνωμένο Βασίλειο	Λιθουανία	Φινλανδία
	Ρουμανία	Ελλάδα
		Ιρλανδία
		Λουξεμβούργο
		Μάλτα
		Ολλανδία
		Πολωνία
		Πορτογαλία
		Σλοβενία
		Σουιδία

Εδώ λοιπόν αφήνουμε τον αλγόριθμο μόνο του να ψάξει για τη μεσαία χώρα και στην συνέχεια να προσπαθήσει μέσω αυτής να εξηγήσει τις υπόλοιπες της ίδιας συστάδας. Έχουμε λοιπόν:

1. Εκβιομηχανισμένες, διάγραμμα 14

- Μεσαία χώρα: Ιταλία
- Ικανότητα εξήγησης των άλλων r^2 : 0.99966 Ουάου, μπορεί να περάσει ευθεία από 3 σημεία όταν τα δύο είναι δίπλα... Απίστευτο

2. Ανατολικό Μπλοκ, διάγραμμα 12

- Μεσαία χώρα: Λιθουανία
- Ικανότητα εξήγησης των άλλων r^2 : 0.0117745

3. Λοιποί, διάγραμμα 13

- Μεσαία χώρα: Δανία
- Ικανότητα εξήγησης των άλλων r^2 : 0.5614167

8.3 Υπάρχει κάποια εύκολη σχέση μεταξύ των δωρεάν αδειών και κάποιου feature;

Εδώ ένας εύκολος τρόπος να προβλέψουμε την σχέση αυτή είναι να αναζητήσουμε τα καταλληλότερα βάρη ώστε να γίνεται βέλτιστη η γραμμική παλινδρόμηση. Οπότε τρέχοντάς το προκύπτει:

Τιμή r^2	Εκβιομηχανισμένες	Ανατολικό Μπλοκ	Λοιποί
	0.9999999	0.976670	0.9463548
Βάρη			
Population	9	0	0
GDPpc	50	1	1
Inflation	50	0	0
Agriculture	59	100	5
Industry	50	100	0
Manufacturing	50	0	0
Total Energy Supply	50	0	100
Verified Emissions	50	1	1
Πιθανή αναλογία free	Οτιδήποτε	Industry + Agriculture	Total energy supply



Figure 12

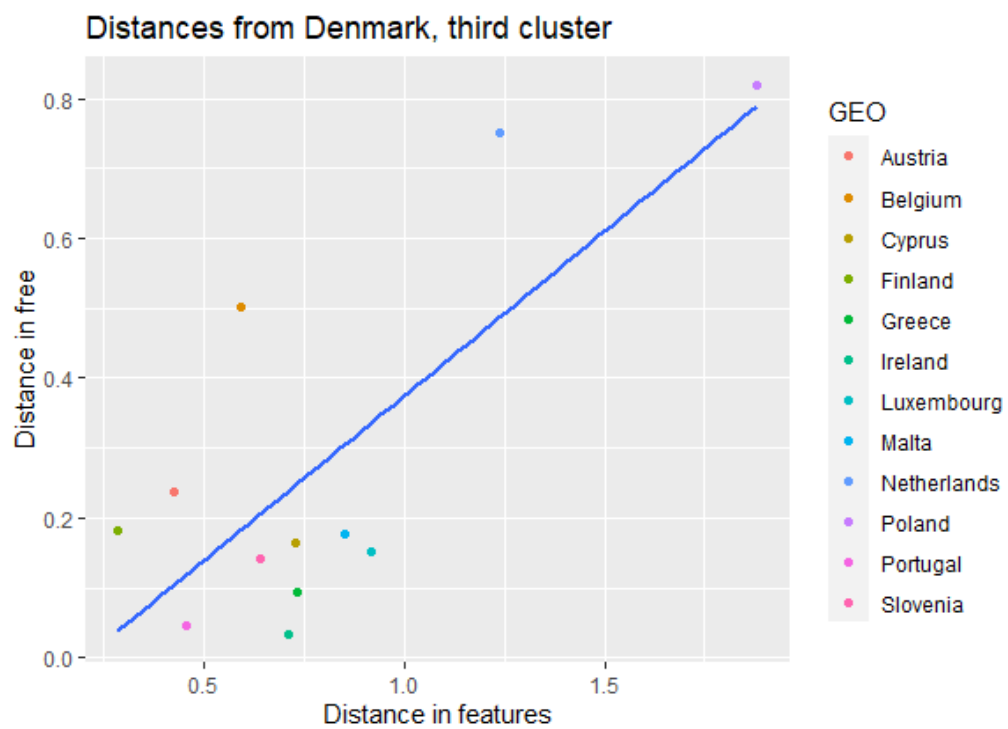


Figure 13

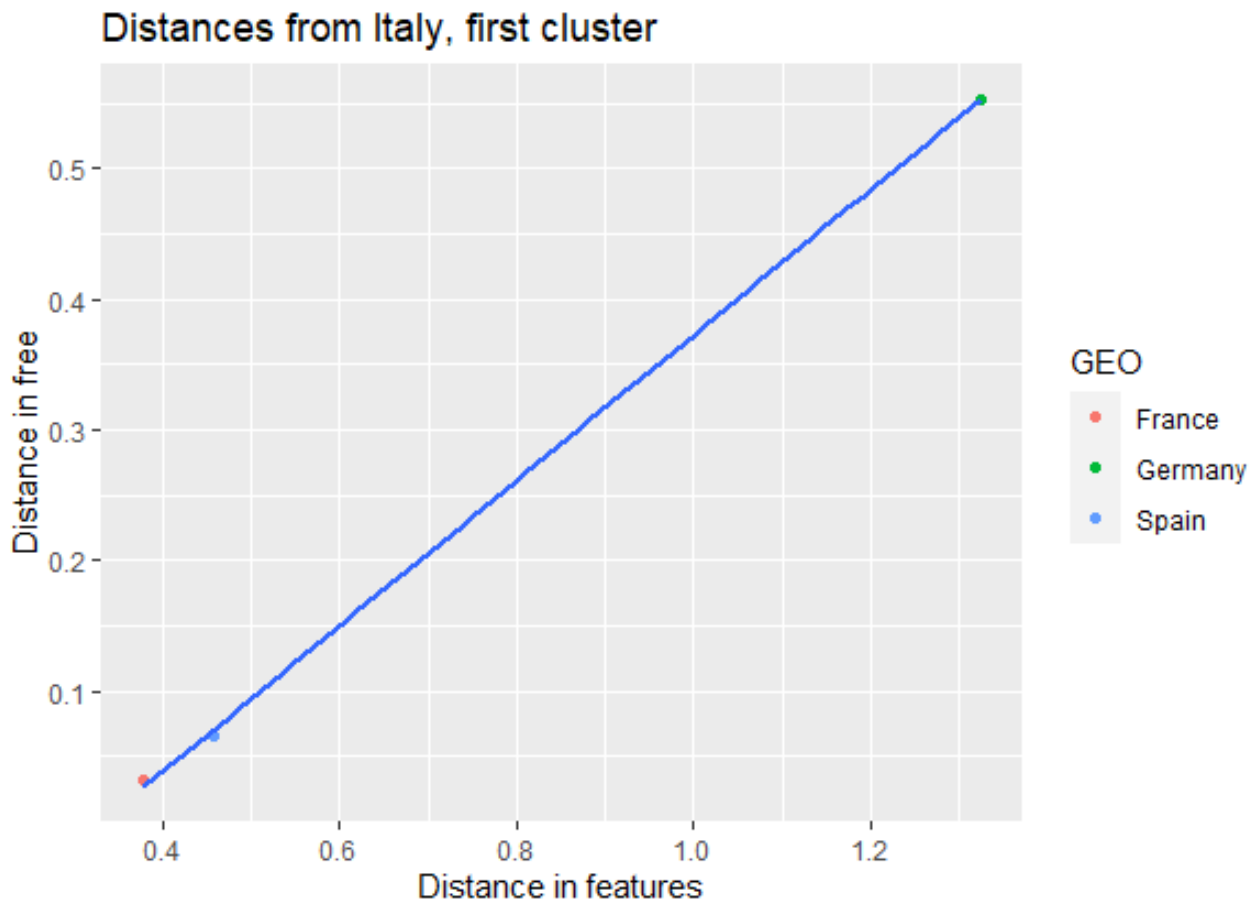


Figure 14

8.4 Ύστερα από αλλαγή στον τρόπο της κανονικοποίησης

Μέχρι τώρα η κανονικοποίηση γίνεται χρησιμοποιώντας την μέγιστη τιμή σε κάθε feature. Τα επόμενα όμως θα γίνουν διαιρώντας τα πάντα με την τιμή που εμφανίζεται στην Γερμανία. Οπότε προκύπτουν τα παρακάτω:

Σε περίπτωση που κάποιος ήθελε να δει τα νέα κανονικοποιημένα δεδομένα με βάση την Γερμανία, έχουμε: Όπως είναι εμφανές, η μόνη χώρα που άλλαξε κατά την αλλαγή αυτή είναι η Πολωνία. Αντίστοιχα λοιπόν προκύπτει:

GEO	Tot.en.sup	GDPpc	Pop.	Infl.	Verified.em.	Agr.	Industry	Manu.	partition
France	0.80	0.87	0.81	0.35	0.25	1.27	0.44	0.35	1
Germany	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1
Italy	0.50	0.73	0.73	0.48	0.35	1.23	0.41	0.39	1
Poland	0.33	0.31	0.46	1.24	0.45	0.50	0.15	0.12	1
Spain	0.40	0.63	0.56	0.86	0.32	1.16	0.26	0.20	1
United Kingdom	0.56	0.92	0.80	1.21	0.33	0.51	0.47	0.32	1
Bulgaria	0.06	0.19	0.09	3.20	0.08	0.08	0.01	-0.57	2
Estonia	0.02	0.46	0.02	2.39	0.03	0.02	0.01	0.00	2
Hungary	0.08	0.33	0.12	2.68	0.05	0.17	0.04	0.04	2
Latvia	0.01	0.35	0.02	1.96	0.01	0.03	0.01	0.00	2
Lithuania	0.02	0.38	0.03	2.82	0.01	0.05	0.01	0.01	2
Romania	0.11	0.24	0.24	3.11	0.09	0.30	0.06	0.06	2
Austria	0.11	1.06	0.11	0.67	0.07	0.16	0.10	0.09	3
Belgium	0.18	0.99	0.14	1.22	0.10	0.11	0.10	0.08	3
Cyprus	0.01	0.60	0.01	0.69	0.01	0.01	0.00	0.00	3
Denmark	0.05	1.29	0.07	0.79	0.04	0.14	0.07	0.06	3
Finland	0.11	1.04	0.07	0.54	0.06	0.19	0.06	0.05	3
Greece	0.07	0.42	0.13	0.19	0.11	0.25	0.03	0.02	3
Ireland	0.04	1.56	0.06	0.62	0.06	0.13	0.12	0.14	3
Luxembourg	0.01	2.47	0.01	1.43	0.00	0.00	0.01	0.00	3
Malta	0.00	0.65	0.01	1.40	0.00	0.00	0.00	0.00	3
Netherlands	0.24	1.09	0.21	0.84	0.21	0.49	0.15	0.12	3
Portugal	0.07	0.48	0.12	1.01	0.07	0.15	0.04	0.04	3
Slovenia	0.02	0.53	0.02	0.97	0.01	0.03	0.01	0.01	3
Sweden	0.16	1.20	0.12	1.42	0.05	0.25	0.12	0.09	3

Τιμή	Εκβιομηχανισμένες	Ανατολικό Μπλοκ	Λοιποί
r^2	0.9388624	0.987699	0.891066
Βάρη			
Population	0	0	100
GDPpc	14	1	1
Inflation	50	50	50
Agriculture	0	79	1
Industry	1	1	0
Manufacturing	0	0	0
Total Energy Supply	100	0	100
Verified Emissions	100	1	1
Πιθανή αναλογία free	Tot energy + Verified	Agriculture	Pop + Tot energy

Εδώ είναι πολύ σημαντικό να γίνει αλλαγή ώστε ο αλγόριθμος να βάζει στο 0 κάτι που δεν αλλάζει σημαντικά το αποτέλεσμα.

8.5 To do list

- Αλλαγή αλγορίθμου υπολογισμού βαρών για να ελέγξουμε το 50 που εμφανίζεται στα βέλτιστα βάρη.

- Εκτέλεση αλλαγή ώστε να βλέπουμε απευθείας τα free ως συνάρτηση των features. Όχι αυτό με τις αποστάσεις.
- Βρες το σωστό τρόπο να γίνει κανονικοποίηση στα δεδομένα.
- Δοκιμαστικά να γίνει το ίδιο πράγμα, αλλά κρατώντας τυχαία χώρες από διαφορετικά clusters.

	GEO	modifier	sol
1	Austria	35.61	0.00
2	Belgium	29.62	0.00
3	Bulgaria	10.06	0.00
4	Cyprus	22.23	0.00
5	Denmark	47.83	0.00
6	Estonia	7.08	0.00
7	Finland	23.43	0.00
8	France	62.70	0.00
9	Germany	22.94	0.00
10	Greece	14.22	0.00
11	Hungary	29.68	0.00
12	Ireland	30.83	0.00
13	Italy	37.10	0.00
14	Latvia	53.59	0.00
15	Lithuania	35.18	0.00
16	Luxembourg	91.99	1.00
17	Malta	48.35	0.00
18	Netherlands	23.44	0.00
19	Poland	12.90	0.00
20	Portugal	25.24	0.00
21	Romania	29.71	0.00
22	Slovenia	26.63	0.00
23	Spain	30.28	0.00
24	Sweden	54.00	0.00
25	United Kingdom	50.14	0.00

Table 11: GDP per capita PPS

	GEO	modifier	last_year	actual	solution
1	Austria	35.6065	0.0273	0.0266	0.0300
2	Belgium	29.6232	0.0453	0.0437	0.0407
3	Bulgaria	10.0627	0.0179	0.0160	0.0161
4	Cyprus	22.2329	0.0032	0.0029	0.0028
5	Denmark	47.8350	0.0116	0.0107	0.0127
6	Estonia	7.0784	0.0066	0.0044	0.0059
7	Finland	23.4298	0.0247	0.0233	0.0223
8	France	62.7004	0.0971	0.0927	0.1068
9	Germany	22.9363	0.2071	0.1996	0.1864
10	Greece	14.2248	0.0198	0.0194	0.0178
11	Hungary	29.6837	0.0144	0.0141	0.0157
12	Ireland	30.8299	0.0139	0.0137	0.0153
13	Italy	37.0977	0.0954	0.0935	0.1049
14	Latvia	53.5870	0.0026	0.0024	0.0029
15	Lithuania	35.1782	0.0077	0.0073	0.0085
16	Luxembourg	91.9900	0.0018	0.0017	0.0020
17	Malta	48.3540	0.0002	0.0002	0.0003
18	Netherlands	23.4438	0.0610	0.0601	0.0549
19	Poland	12.9007	0.0964	0.0892	0.0867
20	Portugal	25.2398	0.0153	0.0150	0.0137
21	Romania	29.7109	0.0367	0.0285	0.0404
22	Slovenia	26.6280	0.0025	0.0024	0.0022
23	Spain	30.2807	0.0829	0.0810	0.0912
24	Sweden	53.9986	0.0329	0.0311	0.0362
25	United Kingdom	50.1382	0.0760	0.0733	0.0836

Table 12: GDP per capita PPS

Country	efficiency	last_year	low_free	up_free	pop	min	max	forecasted	change
Luxembourg	107.5132	0.0017	0.0014	0.0021	0.0012	0.0002	0.0036	0.0014	-20 %
France	65.1380	0.0984	0.0788	0.1181	0.1361	0.0272	0.4084	0.1181	20 %
Latvia	63.4842	0.0024	0.0019	0.0029	0.0040	0.0008	0.0119	0.0019	-20 %
Sweden	62.0909	0.0322	0.0258	0.0386	0.0205	0.0041	0.0614	0.0258	-20 %
United Kingdom	52.8965	0.0735	0.0588	0.0882	0.1344	0.0269	0.4031	0.0882	20 %
Ireland	52.0695	0.0071	0.0057	0.0086	0.0098	0.0020	0.0293	0.0057	-20 %
Denmark	49.7555	0.0115	0.0092	0.0138	0.0117	0.0023	0.0352	0.0092	-20 %
Italy	38.1931	0.0959	0.0767	0.1150	0.1231	0.0246	0.3694	0.1150	20 %
Austria	36.5967	0.0277	0.0222	0.0333	0.0179	0.0036	0.0537	0.0222	-20 %
Lithuania	35.5613	0.0080	0.0064	0.0097	0.0058	0.0012	0.0173	0.0064	-20 %
Hungary	32.7181	0.0140	0.0112	0.0168	0.0199	0.0040	0.0597	0.0112	-20 %
Spain	31.7875	0.0816	0.0653	0.0979	0.0948	0.0190	0.2843	0.0746	-8.549 %
Belgium	30.6643	0.0464	0.0371	0.0557	0.0231	0.0046	0.0694	0.0371	-20 %
Romania	30.3832	0.0379	0.0303	0.0454	0.0398	0.0080	0.1195	0.0303	-20 %
Slovenia	27.0485	0.0025	0.0020	0.0030	0.0042	0.0008	0.0126	0.0020	-20 %
Portugal	26.3706	0.0152	0.0122	0.0183	0.0210	0.0042	0.0629	0.0122	-20 %
Finland	24.4140	0.0251	0.0201	0.0301	0.0112	0.0022	0.0336	0.0201	-20 %
Netherlands	24.1686	0.0618	0.0494	0.0741	0.0348	0.0070	0.1045	0.0494	-20 %
Germany	23.4216	0.2088	0.1670	0.2506	0.1681	0.0336	0.5044	0.2506	20 %
Cyprus	22.7209	0.0033	0.0026	0.0039	0.0024	0.0005	0.0072	0.0026	-20 %
Greece	14.5357	0.0198	0.0158	0.0237	0.0219	0.0044	0.0656	0.0158	-20 %
Poland	12.9609	0.0998	0.0799	0.1198	0.0772	0.0154	0.2317	0.0799	-20 %
Bulgaria	10.1351	0.0184	0.0147	0.0220	0.0144	0.0029	0.0432	0.0147	-20 %
Estonia	7.0940	0.0068	0.0055	0.0082	0.0027	0.0005	0.0080	0.0055	-20 %

Table 13: GDP per capita PPS

R ²	First	Second	Third
Verified_emissions	0.822	0.825	0.786
Total_energy_supply	0.433	0.689	0.971
GDPpc	0.076	0.477	0.000
Population	0.281	0.794	0.801
Inflation	0.003	0.065	0.001
Agriculture	0.088	0.561	0.004
Industry	0.364	0.572	0.019
Manufacturing	0.601	0.417	0.004
Energy_Intensity	0.067	0.042	0.041
actual_agri	0.000	0.663	0.692
actual_ind	0.539	0.630	0.683
actual_manu	0.650	0.594	0.491
tot_and_EI	0.427	0.570	0.914

9 Για την 2η Μαΐου

Έγιναν τα παρακάτω:

1. Γραφθηκε πιο αναλυτικά το: LP εδώ
 2. Συγκέντρωση των στόχων από εδώ και πέρα
- Επόμενοι αναπάντητα ερωτήματα, στόχοι και πλάνα:
1. Χρήση Nash Equilibrium για το Allocation, το οποίο όμως θα πρέπει κάπως να προσαρμοστεί στο να παίρνει υπόψην του το μέγεθος της χώρας.
 2. Τι συμβαίνει στην Γαλλία και στην Πολωνία, κυρίως στα δωρεάν τους. Μήπως αυτά τρώνε μεγάλη αλλαγή; Παίζει ρόλο η πυρηνική ενέργεια;
 3. Διάβασμα: Κινεζικο σύστημα, paper. A multi-criteria decision analysis model for carbon emission quota allocation in China's east coastal areas: Efficiency and equity
 4. Να σουλουπωθεί ο κώδικας. Είναι αίσχος.
 5. Να χωριστεί το total energy supply σε πράσινο και σε ριπογόνο energy supply.
 6. Να δούμε αν κάποιος συνδυασμός με constraints οδηγεί σε κάποιο allocation κοντινό στο πραγματικό. Προφανώς στις πρώτες φάσεις το απλό grandfathering μπορεί να προσομοιωθεί με στενά a1 και a2, οπότε δεν μας νοιάζει αυτό.
 7. Υπάρχουν κάποιες μορφής αντιπρόσωποι μέσα στα cluster?
 8. Όλα τα παραπάνω μπορούν να γίνουν αναφορικά με το δίκτυο;
 9. Έχει νόημα κάποιο Ιεραρχικό clustering σε αυτό;
 10. Διάβασμα το "Cap-and-trade and emissions clustering: A spatial-temporal analysis of the European Union Emissions Trading Scheme
 11. Διάβασμα για ERGM, Exponential family random graph models
 12. Διάβασμα για διαφορετικούς τρόπους κανονικοποίησης.
 13. Τι θα συμβεί εάν επαναλάβουμε τα πειράματα αυτά με δεδομένα μόνο population. total energy supply και Verified emissions με βάρη 1,1,0.5.
 14. Να γίνει PCA (Principle Component Analysis) στα δεδομένα.
 15. Προσθήκη νέων δεδομένων στην βάση.
 16. Ποια είναι τα ειδοποιη χααρακτηριστικά των clusters?
 17. Τι συμβαίνει με τους άλλους stakeholders? Το περιβάλλον, ο μέσος πολίτης κλπ είναι δίκαιο το σύστημα σε αυτούς;
 18. Να κατανοήσουμε τι συμβαίνει με τα διαφορετικά clusters και τις διαφορές που βλέπουμε στην απόδοση του regression.
 19. Γιατί κάποιες χώρες παίρνουν παραπάνω ή παρακάτω από την καμπύλη
 20. Να χρησιμοποιήσουμε το μαθηματικό μοντέλο που βάζει στο παιχνίδι και την οντότητα του κράτους για να εξισορροπήσουμε τη διανομή και σε επίπεδο κρατών (το benchmarking το επιτυγχάνει αυτό μόνο μεταξύ εταιρειών στο ίδιο sector).

10 Για τις 16 Μαΐου

10.1 To do

Τα to do είναι:

- Δημιουργία ενός proxy για το energy Intensity το οποίο θα μπορούμε μετά να το εξειδικεύσουμε ανά sector, ούτως ώστε να μπορούμε να φτιάξουμε το LP.
- Reagration για 1 sector σε όλες τις χώρες. Τι βγάζει; Γιατί οι ίδιοι κλάδοι σε άλλες χώρες θα έχουν διαφορετικές τιμές;
- Regration για κάθε συνδυασμό χώρας και sector. Εδώ τι συμπεράσματα υπάρχουν;
- Όλα τα παλιά :P

10.2 Δεν μοιάζει να μπορεί να προσομοιωθεί απλά το Energy Intensity

Εδώ έχει υπολογισθεί ως:

$$\frac{\frac{VerifiedEmissions}{100 - PercentageGreen}}{GDP}$$

Όμως η τιμή της EU δείχνει να είναι κάτι αρκετά διαφορετικό.



Figure 15: Calculated Emission Intensity - Energy Intensity

Residuals:

Min	1Q	Median	3Q	Max
-3.173e-06	-8.898e-07	-2.022e-07	5.596e-07	4.043e-06

Coefficients:

Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.065e-06	1.258e-06	-0.847 0.4061
dat_Energy_Intensity	2.374e-08	9.301e-09	2.553 0.0181 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.713e-06 on 22 degrees of freedom

Multiple R-squared: 0.2285, Adjusted R-squared: 0.1934

F-statistic: 6.516 on 1 and 22 DF, p-value: 0.01815

Το βασικό πρόβλημα είναι πως δεν κατάφερα να το κάνω να μοιάζει περισσότερο σε καμία εκδοχή. Ακόμα και αν βάλω στην μία πλευρά το energy supply



Figure 16: Calculated Energy Intensity - Energy Intensity

Residuals:

Min	1Q	Median	3Q	Max
-1.315e-07	-3.408e-08	-1.516e-08	3.144e-08	1.561e-07

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.028e-08	4.200e-08	0.483	0.6339
dat\$Energy_Intensity	7.652e-10	3.106e-10	2.463	0.0221 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.722e-08 on 22 degrees of freedom

Multiple R-squared: 0.2162, Adjusted R-squared: 0.1806

F-statistic: 6.068 on 1 and 22 DF, p-value: 0.02206

10.3 Διαφορετικά μέτρα απόδοσης / αποτελεσματικότητας

- **Energy Efficiency** Αυτό μετρείται σε Mtoe και ειλικρινά δεν μπορώ να καταλάβω από την περιγραφή του πώς διαφέρει από κάποια μορφή του total energy consumed. Έχει να κάνει με τους στόχους της EE https://ec.europa.eu/eurostat/cache/metadata/en/nrg_ind_eff_esmsip2.htm

- **Energy Productivity**

$$\frac{EconomicOutput}{RawEnergy}$$

https://ec.europa.eu/eurostat/cache/metadata/en/nrg_ind_ep_esmsip2.htm

- **Energy Intensity** Τα γνωστά

10.4 Για την επόμενη φορά

- Να γίνει ο ίδιος δείκτης Energy Intensity, όμως αυτή τη φορά να μην έχει μέσα την ηλεκτροπαραγωγή, καθώς αυτή διπλομετράται. Άρα προκύπτουν δύο πειράματα:
 1. Όλα τα sectors εκτός από την αεροπορία και την ηλεκτροπαραγωγή συμμετέχουν.
 2. Μόνο η ηλεκτροπαραγωγή συμμετέχει.
- Clustering σε βάθος χρόνου ώστε να δούμε αν η αλλαγές στον χρόνο επηρεάζουν το υπόλοιπο.
- Clustering ανά sector
- Αφαίρεσε την Μάλτα, λολ
- Να ξεκινήσει το LP ακόμα και με σκατά δεδομένα για να δούμε τι βγάζει!

11 Για τις 23 Μαΐου

11.1 21-99 Cannot be used a proxy

Αρχικά κατεβάζοντας από το <https://www.eea.europa.eu/data-and-maps/data/european-union-emissions-trading> Country:Allcountries τους παραδοθέντες ρίπους, παρατηρούμε πως έχουν αυτήν την κατηγοριοποίηση:

- [1] "20 Combustion of fuels"
- [2] "42 Production of bulk chemicals"
- [3] "20-99 All stationary installations"
- [4] "43 Production of hydrogen and synthesis gas"
- [5] "25 Production or processing of ferrous metals"
- [6] "28 Production or processing of non-ferrous metals"
- [7] "21-99 All industrial installations (excl. combustion)"
- [8] "23 Metal ore roasting or sintering"
- [9] "36 Production of paper or cardboard"
- [10] "33 Manufacture of mineral wool"
- [11] "10 Aviation"
- [12] "26 Production of primary aluminium"
- [13] "27 Production of secondary aluminium"
- [14] "29 Production of cement clinker"
- [15] "41 Production of ammonia"
- [16] "21 Refining of mineral oil"

- [17] "34 Production or processing of gypsum or plasterboard"
- [18] "35 Production of pulp"
- [19] "38 Production of nitric acid"
- [20] "32 Manufacture of ceramics"
- [21] "24 Production of pig iron or steel"
- [22] "30 Production of lime, or calcination of dolomite/magnesite"
- [23] "31 Manufacture of glass"
- [24] "45 Capture of greenhouse gases under Directive 2009/31/EC"
- [25] "99 Other activity opted-in under Art. 24"
- [26] "22 Production of coke"
- [27] "44 Production of soda ash and sodium bicarbonate"
- [28] "37 Production of carbon black"
- [29] "39 Production of adipic acid"
- [30] "40 Production of glyoxal and glyoxylic acid"

όμως κανένα από αυτά δεν αναφέρεται στην ηλεκτροπαραγωγή συγκεκριμένα. Το πιο κοντινό ίσως είναι τα combustion fuels. Όμως κάνοντας δοκιμές για το συγκεκριμένο, καταλήξουμε πως αφαιρώντας αυτό δεν μπορούμε να καταλήξουμε σε κάποιο χρήσιμο proxy:



Συγκεκριμένα διαβάζουμε [1]:

"Most installations covered by the EU ETS will be found in the NACE categories C (Mining and quarrying), D (Manufacturing) and E (Electricity, gas and water supply). However, the activity \combustion of fuels" can occur in all types of NACE categories, not only industrial ones. Examples of such non-industrial installations are combustion units in greenhouses, hospitals, universities and office buildings, booster stations in natural gas transport networks etc"

Επομένως, δεν είναι καθόλου περίεργο. Αυτό το οποίο όντως χρειάζεται να μπει σε αυτήν την διαδικασία είναι το Nace code της εταιρείας. Συγκεκριμένα έχουμε [2]:

NACE codes are the standard European nomenclature of productive economic activities. They break down the universe of economic activities in such a way that a NACE code can be associated with a statistical unit carrying out the activity it designates. There is an economic activity when resources { such as capital goods { are combined to produce specific goods or services. All activity is characterized by the input of resources, a production process and an output of products (goods or services).

11.2 Χρήση του GDPpps

Εδώ βλέπουμε την πρώτη σημαντική βελτίωση στο αποτέλεσμα:



συγκεκριμένα, η βελτίωση είναι της τάξης του:

Residual standard error: 3.18e-07 on 21 degrees of freedom

Multiple R-squared: 0.3127, Adjusted R-squared: 0.2799

F-statistic: 9.553 on 1 and 21 DF, p-value: 0.005539

ΣΕ

Residual standard error: 2.269e-07 on 21 degrees of freedom

Multiple R-squared: 0.4208, Adjusted R-squared: 0.3932

F-statistic: 15.26 on 1 and 21 DF, p-value: 0.0008128

Οι πιο έντονοι outliers στο παραπάνω είναι οι 3 πάνω κουκίδες που ανήκουν στην Αυστρία, Λιθουανία και την Φινλανδία από αριστερά προς τα δεξιά.

11.3 Χωρίς Αυστρία, Λιθουανία και Φινλανδία

Εδώ τα πράγματα βελτιώνονται, αλλά δεν είναι τέλεια:



Residual standard error: 1.315e-07 on 18 degrees of freedom
Multiple R-squared: 0.6578, Adjusted R-squared: 0.6388
F-statistic: 34.6 on 1 and 18 DF, p-value: 1.435e-05

12 Για τη 13η Μαΐου

12.1 Το πρόβλημα με το πλασματικό energy Intensity

Υπήρξε ένα σημαντικό κενό εδώ λόγω διαφόρων προσωπικών υποχρεώσεων.

Αρχικά. Σχετικά με την αποδόμηση των sector των ρύπων: Έγινε αυτή η αντιστοίχιση https://docs.google.com/spreadsheets/d/1mwmZSr8PBIFgsxemVE_FFEdaZS0d0eX3/edit?usp=sharing&ouid=105915891653467891041&rtprof=true&sd=true Το οποίο δείχνει με αρκετή σαφήνεια πως οι δύο τρόποι κατηγοριοποίησης των επιχειρήσεων είναι πολύ διαφορετικές.

Οι πρώτες δύο στήλες αναφέρονται στα δεδομένα έτσι όπως εμφανίζονται στην βάση δεδομένων του EU ETS, ενώ οι στήλες 3,4 δείχνουν πώς αυτές εμφανίζονται στα δεδομένα της Eurostat. Προφανώς όμως κάπου υπάρχουν αυτά τα δεδομένα, ίσως κάπου στο EU ETS, αλλά δεν ξέρω πού. Σε κάθε περίπτωση, η αντιστοίχιση θα μπορούσε να γίνει όπως φαίνεται στο παραπάνω excel. Να δύο παραδείγματα:

Ένα παράδειγμα στο οποίο στεκόμαστε τυχεροί:

Eu ETS Code	Meaning	Nace code	Meaning
2	Mineral oil refineries	C19.20	Manufacture of refined petroleum products
21	Refining of mineral oil	c19.10	Manufacture of Coke and Refined Petroleum Products
22	Production of coke	c19.10	Manufacture of Coke and Refined Petroleum Products
3	Coke ovens	c19.10	Manufacture of Coke and Refined Petroleum Products
		C19	Manufacture of coke and refined petroleum products

Εδώ είμαστε τυχεροί γιατί μπορούμε να δουλέψουμε για το sector c19, ενώνοντας τους κωδικούς του EU ETS 2, 21, 22, 3. Αυτό θα προσφέρει λιγότερη ακρίβεια, όμως τουλάχιστον δεν θα είναι εκ προοιμίου λάθος.

Προβληματική περίπτωση:

Eu ETS Code	Meaning	Nace code	Meaning
		C23	Manufacture of other non-metallic mineral products
29	Production of cement clinker	C23.5.1	Manufacture of cement
30	Production of lime, or calcination of dolomite/mag...	C23.5.2	Manufacture of lime and plaste
31	Manufacture of glass	C23.1	Manufacture of glass and glass products
32	Manufacture of ceramics	C23.4	Manufacture of other porcelain and ceramic products
33	Manufacture of mineral wool	C23.9.9	Manufacture of other non-metallic mineral products n.e.c.
34	Production or processing of gypsum or plasterboard	C23.5.2	Manufacture of lime and plaster
		C23.2	Manufacture of refractory products
		C23.3	Manufacture of clay building materials
		C23.6	Manufacture of articles of concrete, cement and plaster
		C23.7	Cutting, shaping and finishing of stone
		C23.9.1	Production of abrasive products

Εδώ έχουμε ένα σενάριο στο οποίο όχι μόνο δεν ταιριάζουν όλα απόλυτα, αλλά ακόμα και να την αριστερή στήλη

σε ένα για να την ταυτίσουμε με την δεξιά, θα υπάρχουν κατηγορίες τις οποίες αφήσαμε απ' έξω.

12.2 Clustering se sectors

Query για να πάρουμε τις Surrendered άδειες ανά sector για το 2016.

```
SELECT SUM(trnew.NbOfUnits), concat(mat.kwdikos , ' ' , mat.onoma) as ActivityType
FROM transactions_new AS trnew
JOIN eutl_accountholders AS ah ON trnew.TransferringAccountHolder = ah.holderName
JOIN eutl_accounts AS a ON ah.rawCode = a.rawCode
JOIN eutl_installations_orair AS io ON io.account = a.rawCode
JOIN mainactivitytype AS mat ON io.mainActivity = mat.kwdikos
WHERE trnew.TransactionType LIKE '%%-2' AND YEAR(trnew.TransactionDate) = 2016
GROUP BY mat.onoma
ORDER BY mat.kwdikos
```

Το οποίο επιστρέφει:

Surrendered	ActivityType
178093310	1 Combustion installations with a rated thermal in...
287282477	10 Aircraft operator activities
4028607	2 Mineral oil refineries
697984775	20 Combustion of fuels
6795214	21 Refining of mineral oil
187557	22 Production of coke
179735	23 Metal ore roasting or sintering
98482418	24 Production of pig iron or steel
21458739	25 Production or processing of ferrous metals
806566	26 Production of primary aluminium
2287866	27 Production of secondary aluminium
16525496	28 Production or processing of non-ferrous metals
9283720	29 Production of cement clinker
29011461	30 Production of lime, or calcination of dolomite/...
28483791	31 Manufacture of glass
113925371	32 Manufacture of ceramics
855132	33 Manufacture of mineral wool
2350152	34 Production or processing of gypsum or plasterbo...
34482572	35 Production of pulp
120320357	36 Production of paper or cardboard
3530168	37 Production of carbon black
10773361	38 Production of nitric acid
3665	39 Production of adipic acid
205592	40 Production of glyoxal and glyoxylic acid
518707	41 Production of ammonia
51838556	42 Production of bulk chemicals
8355618	43 Production of hydrogen and synthesis gas
4930431	44 Production of soda ash and sodium bicarbonate
10617	46 Transport of greenhouse gases under Directive 2...
1853546	5 Installations for the production o...ion) includ...
12267703	6 Installations for the production o...rotary kiln...
1787688	7 Installations for the manufacture of glass inclu...
18284476	8 Installations for the manufacture ...ks, tiles, ...
21850482	9 Industrial plants for the producti...ous materia...
42185334	99 Other activity opted-in pursuant to Article 24 ...

12.3 Τι ειπώθηκε στην κλίση

Στην κλίση τέθηκαν οι παρακάτω στόχοι:

- Να γίνει το LP χρησιμοποιώντας τους 5 sectors έτσι όπως έγιναν λόγω των περιορισμών των δεδομένων. Άρα όπως είναι στο excel τα διαφορετικά χρώματα: https://docs.google.com/spreadsheets/d/1mwmZSr8PBifgsxemVE_FFEdaZS0d0eX3/edit?usp=sharing&ouid=105915891653467891041&rtpof=true&sd=true
- Να γίνει clustering στους διαφορετικούς sectors. Χρησιμοποίησε Queries σαν αυτό που ανέφερες λίγο παραπάνω για να αποκτήσεις δεδομένα για τους διαφορετικούς κλάδους της οικονομίας. Παραδείγματα: Verified, Free, no of companies, avg transactions per company, μέγεθος εταιρειών

- Να προσπαθήσεις να δεις γιατί κάποιες χώρες δεν αντιστοιχίζονται τέλεια στα διαγράμματα τα οποία υποβάλαμε στο paper. Μερικές πιθανές συσχετίσεις: ποσοστό service στο ακαθάριστο προϊόν, πυκνότητα σε εταιρείες με πολύ μεγάλο ή πολύ μικρό P(carbon leakage).
- Να ξαναγίνει η μελέτη που γίνεται στο paper μόνο για έναν από τους sectors της οικονομίας και να δούμε αν αυτό μιδενίζει τις διαφορές μεταξύ των χωρών.
- Διάβασμα το paper: Pollution Permits: Efficiency by Design https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4453956

13 Παύση Καλοκαιριού

Ρεαλιστικά δεν έγιναν όσα πράγματα θα έπρεπε από μέρους μου. Προέκυψαν πολλές δουλειές άσχετες με την σχολή πάνω στην εξεταστική, με αποτέλεσμα αυτή να πάει πολύ άσχημα. Τα βασικά αίτια ήταν το ότι αναλαμβάνω πράγματα που δεν πρέπει και πάντα έχω μπροστά μου ένα βουνά που η αναβλητικότητα μου δεν με αφήνει να νικήσω. Θα το αλλάξω.

14 Επανεκκίνηση Μετά την παύση του καλοκαιριού

14.1 Polution Permits: Efficiency by Design

Οι γνώσεις μου είναι υπερβολικά λίγες για να καταλάβω το περιεχόμενο αυτού του paper.

14.2 Παρουσίαση για το paper

14.3 1 sector

Ακριβώς ένα sector ίσως να γίνεται μόνο όσων αφορά στα δεδομένα της βάσης του ETS, οπότε δεν θα χρησιμοποιήσουμε άλλα

15 Συνάντηση στο ΜΟΠ 10/10/23

Στη συνάντηση που πραγματοποιήθηκε στις 10 Οκτωβρίου 2023 στο ΜΟΠ, συζητήθηκαν σημαντικά ζητήματα σχετικά με την ερευνητική προσέγγιση και τους πιθανούς στόχους της διπλωματικής.

15.1 Προκύπτοντα ερωτήματα

Κατά τη διάρκεια της συζήτησης προέκυψαν τα εξής ερωτήματα:

1. Είναι δύσκολο να εφαρμόσουμε τεχνικές clustering ανά εταιρεία λόγω της έλλειψης δεδομένων. Υπάρχει όμως η πιθανότητα για μεγάλες εταιρείες να έχουν δημοσιευμένα δεδομένα που μπορεί να χρησιμοποιηθούν;

15.2 Προτεινόμενοι στόχοι

Στη συνέχεια της συνάντησης, προτάθηκαν κάποιοι στόχοι για την ερευνητική διαδικασία:

1. Διαχωρισμός των προσεγγίσεων: Αρχικά, να καθοριστεί αν μπορούμε να παράγουμε απλώς συμπεράσματα ή αν μπορούμε να προτείνουμε συγκεκριμένες τεχνικές. Είναι σαφές ότι ο στόχος πρέπει να είναι η προσφορά συγκεκριμένων τεχνικών.

2. Επανάληψη των βημάτων του paper: Ακολουθώντας τις ίδιες διαδικασίες όπως στο paper, να εξετάσουμε αν η τεχνική Linear Programming (LP) μπορεί να λύσει τα διάφορα προβλήματα, ειδικά όταν ομαδοποιούμε τις εταιρείες.
3. Σύγκριση τεχνικών: Να δούμε αν η τεχνική "Allocating Emission Permits" παράγει αποτελέσματα που είναι κοντά στα αποτελέσματα του LP.
4. Ορισμός concept για τη διπλωματική: Είναι ζωτικής σημασίας να καθορίσουμε ένα σαφές concept για την διπλωματική, το οποίο θα κατευθύνει όλες τις ερευνητικές προσπάθειες.

16 Συνάντηση στο ΜΟΠ 31/10/23

Στη συνάντηση που πραγματοποιήθηκε στις 31 Οκτωβρίου 2023 στο ΜΟΠ, συζητήθηκαν οι εξής σημαντικές πτυχές:

16.1 Σημαντικά Ειπωθέντα

- Η Τζέλα ανέβασε στο drive ένα άρθρο που περιέχει τη διάσταση του χρόνου, με κλειδί λέξεις όπως uncertainty και αναφορά στην παράγραφο 3.
- Το abatement μπορεί να αναλυθεί σε δύο κατευθύνσεις: ρίποι και προϊόν.

16.2 Στόχοι

1. Δοκίμασε το τελευταίο paper για να δεις αν λειτουργεί εντός του LP και αν περιλαμβάνει κάποια μορφή fairness όπως έχει οριστεί.
2. Προσπάθησε να κατανοήσεις πλήρως το μοντέλο του paper και συγκρίνε το με το LP.
3. Ολοκλήρωσε τον καθαρισμό του κώδικα.
4. Δημιουργήστε ένα συνθετικό dataset. Θα χρειαστεί κανόνες για την επόμενη συνάντηση για να κάνεις το dataset να λειτουργεί, όπως κριτήρια συνοχής για τα συνθετικά sectors, GDP κ.λπ. Επίσης, η ισορροπία Cournot θα βοηθήσει στον καθορισμό της τιμής του προϊόντος ανά sector.

17 Συνάντηση μέσω Skype στις 5 Νοεμβρίου 2023

Η συζήτηση πραγματοποιήθηκε μέσω Skype στις 5 Νοεμβρίου 2023 και επικεντρώθηκε στη δημιουργία συνθετικών δεδομένων για τη διπλωματική.

17.1 Δημιουργία Συνθετικών Δεδομένων

Στόχος είναι η παραγωγή συνθετικών δεδομένων για τρεις χώρες και τρεις τομείς σε κάθε χώρα, με την εκπροσώπηση τεσσάρων εταιρειών ανά τομέα. Τα δεδομένα θα καλύπτουν τις παρακάτω πτυχές:

1. Energy Intensity
2. Ακαθάριστο Εγχώριο Προϊόν ανά τομέα
3. Εκπομπές ρύπων ανά τομέα

Επιπρόσθετα, η κλίμακα κάθε εταιρείας θα πρέπει να σχετίζεται με το μέγεθος της αντίστοιχης χώρας.

17.2 Ανάλυση Εταιρικών Συναρτήσεων

Η συζήτηση αφορούσε επίσης τις εταιρικές συναρτήσεις και τους κανόνες που τις καθορίζουν. Σημαντικά σημεία περιλαμβάνουν:

1. Η ανάγκη μελέτης του τρίτου κεφαλαίου του ΑΟΘ. Από εμένα, γιατί είμαι άσχετος.
2. Η συνάρτηση "abatement" θεωρείται ως φθίνουσα, με μηδενική τιμή στις εκπομπές Business as Usual και άπειρη τιμή στις μηδενικές εκπομπές.
3. Το κόστος παραγωγής αποτελείται από το άθροισμα σταθερών και μεταβλητών κοστών, με χαρακτηριστική φθίνουσα απόδοση.
4. Το τελικό κόστος ανά προϊόν αποτελεί αυξανόμενη κυρτή συνάρτηση.

17.3 Στόχοι για την Επόμενη Συνάντηση

Ως επόμενο βήμα, προτείνεται η εκπόνηση πέντε λογικών συναρτήσεων που θα εφαρμόζονται στα προαναφερθέντα δεδομένα.

18 Συνάντηση στο ΜΟΠ στις 7 Νοεμβρίου 2023

18.1 Πριν την συνάντηση

18.1.1 Ορισμοί

- Η παραγωγή q_i θα είναι η ελεύθερη μεταβλητή στο πρόγραμμά μας.
- Στο paper ορίζει το abatement ως non-decreasing και convex, σε αντίθεση με τον ορισμό του Σωτήρη, ο οποίος ορίζει πάλι κυρτή συνάρτηση αλλά φθίνουσα. Η διαφορά είναι πως η μία παίρνει για είσοδο την μείωση των ρίπων, ενώ η δεύτερη παίρνει για είσοδο τους τελικούς ρύπους.
- Η συνάρτηση απόδοσης των δωρεάν αδειών είναι κοίλη.
- x_i^* ορίζονται οι ρύποι ύστερα από τον συνυπολογισμό των δωρεάν αδειών για την εταιρεία i . q_i^* ομοίως είναι η παραγωγή μετά από τον υπολογισμό. Ορίζεται επίσης η διαφορά $q_i^* - x_i^*$ ως το μέγεθος του abatement.

18.1.2 Πιθανές συναρτήσεις παραγωγής

Η συνάρτηση για το κόστος παραγωγής μπορεί να έχει τις παρακάτω μορφές:

1. Γραμμική : $FC + Bx$, όπου A τα σταθερά κόστη και B το οριακό κόστος ανά μονάδα.
2. Κυρτή αύξουσα : $FC + Ae^{\lambda x}$
3. Αλλαγή κυρτή αύξουσα: $FC + Ax^\lambda$
4. Συνυπολογίζοντας τις οικονομίες κλίμακας (chat gpt): $TC(n) = FX + \frac{VC}{1+an} \times n$.
5. Πάλι από το GPT: $TC(n) = FC + nVC + SC(n)$, όπου το SC είναι το Scaling cost.

18.1.3 Abatement cost

Το abatement cost μπορεί να αναχθεί από τον ορισμό του Σωτήρη στον ορισμό που καταλαβαίνει το κεφάλι μου απλά με την αλλαγή $Sot(Actual, BAU) = C(BAU - Actual)$. Αν θέλουμε να παίρνει για είσοδο ποσοστό τότε η αντιστοιχία είναι:

$$Sot(Actual, BAU) = C\left(\frac{BAU - Actual}{BAU} \times 100\%\right)$$

Πιθανές συναρτήσεις για abatement είναι:

1. $g(x) = Ae^{\lambda x}$
2. $g(x) = ax^2 + bx + c$
3. GPT: $c(e^{bx} - 1), ax^3 \dots$
4. GPT: $\ln(x)$ το οποίο δεν βγάζει κανένα νόημα.

Συναρτήσεις που ταιριάζουν με τον ορισμό του Σωτήρη είναι:

1. $\frac{1}{x} - \frac{1}{BAU}$
2. $\frac{1}{x^2}$

18.2 Τι είπαμε στην συνάντηση - Ορισμός προβλήματος

Αρχικά βρήκα τον κύβο μου! (τον 11 M pro, ο i3 ακόμα αγνοείται) (ναι, δεν είναι τόσο σημαντικό αυτό, αλλά και είναι λίγο)

- N εταιρείες.
- C χώρες.
- S τομείς της αγοράς.
- τ κόστος της άδειας.
- $E_s : q_i \rightarrow x_i$ ρυθμός ρύπων για τον τομέα s
- $q_{i,y}$ παραγωγή της εταιρείας με id = i και έτος = y.
- q_i παραγωγή της εταιρείας με id = i, για ανάλυση μίας χρονιάς.
- $q_i^* = q_{i,y+1}$
- $\chi_{i,y}$ ρύποι της εταιρείας με id = i και έτος = y.
- χ_i ρύποι της εταιρείας με id = i, για ανάλυση μίας χρονιάς.
- $\chi_i^* = \chi_{i,y+1}$
- $\Phi_i(y)$ δωρεάν άδειες για το έτος y στην εταιρεία i.
- $C(i)$ = επιστρέφει την χώρα της εταιρείας i.
- $S(i)$ = επιστρέφει τον τομέα της εταιρείας i.
- $K_{c,s,y} = \sum x_{i,y}, \forall x_{i,y} | y = y \wedge c = C(i) \wedge S(i) = s$

- $Q_{c,s,y} = \sum q_{i,y}, \forall x_{i,y} | y = y \wedge c = C_i \wedge S(i) = s$
- $p_s(y)$ η αντίστροφη συνάρτηση ζήτησης για το προϊόν του τομέα s για το έτος y . Κούλη και φθίνουσα. $p_s : Q \rightarrow \mathbb{R}$, όπου \mathbb{R} εδώ είναι η τιμή πώλησης.
- $f_i(y)$ η συνάρτηση abatement, κυρτή, αύξουσα με $f(0) = 0 \wedge f'(0) = 0 \wedge f_i(x) \forall x \geq 0$
- $pr_i(q_{i,y})$ κόστος παραγωγής για την εταιρεία. Όπως είπαμε είναι ξεχωριστή συνάρτηση για την κάθε εταιρεία, κυρτή και αύξουσα. Σε αντίθεση με το paper [3] όπου το οριακό κόστος είναι σταθερό.

Κέρδος επιχείρησης:

$$p_s(\sum q_j^* \forall j | S(j) = S(i)) - \tau(x_i^* - \Phi_i) - pr_i(q_i^*) - f_i(E_s(q_i) - x_i^*)$$

18.2.1 Βήμα 1

Στη συνέχεια θέσαμε με πολλή σαφήνεια τους στόχους. Επομένως προκύπτει ο παρακάτω ορισμός προβλήματος:

- Θα φτιάξουμε 1 εταιρεία για κάθε j από τους 3 τομείς για κάθε μία χώρα i από τις 3 στα συνθετικά δεδομένα.
- Καλό είναι στα δεδομένα αυτά να έχουμε έναν μπουσουλα στην αρχή, οπότε οι 3 χώρες θα μοιάζουν με την Ιταλία, τη Γερμανία και την Ολλανδία.
- Αντίστοιχα, οι τομείς θα μοιάζουν με τους τομείς: τσιμέντα, ηλεκτροπαραγωγή, παραγωγή χαρτιού.
- Για κάθε τομέα θα χρησιμοποιήσουμε μία συνάρτηση ρυθμού ρύπων, άρα 3 συναρτήσεις.
- Θα χρειαστούμε επίσης 1 συνάρτηση ζήτησης για κάθε τομέα, άρα 3 συναρτήσεις φθίνουσες.
- Οι συναρτήσεις abatement θα είναι ξεχωριστές για όλους, άρα 9.
- Ομοίως, και οι 9 συναρτήσεις production.
- Η τιμή της άδειας θα θεωρείται σταθερή.
- Το κόστος παραγωγής θα είναι γραμμική αύξουσα.
- Το abatement θα ορίζεται ως μπλα μπλα.

18.2.2 Βήμα 2

Αφού οριστούν αυτά θα πρέπει να βρεθεί ο Cournot ισορροπία. Στο σύστημα.

Μετά τα δεδομένα αυτά θα feedαριστούν στο LP, τον Linear Mechanism του άρθρου [3] και ιδανικά και στο benchmarking.

18.2.3 Βήμα 3 - Βελτιώσεις

Χωρίς κάποια συγκεκριμένη σειρά πρέπει να γίνουν τα παρακάτω:

- Η τιμή της άδειας να ορίζεται από το Cap and Trade σύστημα.
- Η συνάρτηση κόστους παραγωγής να γίνει κυρτή. Πιθανώς πολυώνυμο τρίτου βαθμού καθώς το marginal cost του παναγιώτη ήταν δευτέρου.
- Η συνάρτηση του abatement να γίνει απειρίζουσα κυρτή.
- Να μπουν παραπάνω εταιρείες ανά sector.
- Παραπάνω χώρες.
- Παρακάτω τομείς.

19 Συνάντηση 14 Νοεμβρίου 2023 στο ΜΟΠ

1. Είπαμε πως στην αρχή το κόστος παραγωγής θα είναι 0. $h=0$, άρα δεν χρειάζεται να γίνει κάτι διαφορετικό από αυτό που κάνουν στο paper [3].
2. Βασάνισες πολύ την Τζέλα, χρωστάς καφέ / ποτό.
3. Πρέπει να διαβάσω το [3] για να καταλάβω καλύτερα το μηχανισμό του regulator.

Μερικές απορίες που έχω και οι οποίες δεν λύνονται νομίζω κάπως:

1. Στην ερώτηση είχα σημειώσει ένα συγκεκριμένο παράδειγμα, αλλά αυτό το ερώτημα θα μπορούσε να γραφτεί και ως: Πώς διαχειριζόμαστε τις διαφορές στην πληροφορία μεταξύ των εταιρειών;
2. Πώς μπορούμε να βάλουμε στο σύστημα επιπλέον στοιχεία ρεαλισμού, όπως την ανελαστικότητα της παραγωγής για κάποια προϊόντα ή το ότι αν μία επιχείρηση σταματήσει φέτος την παραγωγή της, ίσως να μην είναι εφικτό να την επαναφέρει την επόμενη χρονιά, οπότε ίσως να πρέπει να το συνυπολογίσει αυτό.

20 Συνάντηση 28 Νοεμβρίου 2023

Ακύρωσα την προηγούμενη Τρίτη. Διάβασα για το cournot competition. Αλλά και εκείνη την φορά πήγα από ντροπή. Δεν είχα κάτι καινούριο να πω ή δεν είχα κάνει κάτι παραπάνω. Μετά ακύρωσα και την επόμενη εβδομάδα, επειδή ήταν το Πρώτο Εθνικό πρωτάθλημα ταχυκυβισμού.

21 Συνάντηση στις 12/12/23

22 Πηγές, από 23/5/23

- [1] Guidance on Interpretation of Annex I of the EU ETS Directive (excl. aviation activities), https://www.mase.gov.it/sites/default/files/archivio/allegati/emission_trading/guidance_on_interpretation_annex_I_final.pdf
- [2] NACE Codes: What Are They and Why Do They Always Matter? https://connects.world/nace-codes/#What_are_NACE_codes_used_for
- [3] Allocating Emission Permits Efficiently via Uniform Linear Mechanisms, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4536951