## ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ ΤΜΗΜΑ ΜΗΧ. Η/Υ & ΠΛΗΡΟΦΟΡΙΚΗΣ

### ΠΡΟΧΩΡΗΜΕΝΑ ΘΕΜΑΤΑ ΤΕΧΝΟΛΟΓΙΑΣ ΚΑΙ ΕΦΑΡΜΟΓΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ

# ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ ΓΙΑ ΤΟ ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2021-2022

ΧΑΡΙΣΗΣ ΑΛΕΞΑΝΔΡΟΣ, 3361 ΔΗΜΗΤΡΗΣ ΒΑΜΠΙΡΗΣ 3186 ΚΩΝΣΤΑΝΤΙΝΟΣ ΖΙΟΥΔΑΣ 3225

ΤΕΛΙΚΗ ΑΝΑΦΟΡΑ

#### 1 ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Στην παρούσα άσκηση τα δεδομένα δίνονται με ένα συγκεκριμένο φορμάτ:

Country Name	Country Code	Indicator Name	Indicator Code	1960	1961	1962	1963	1964	1965	1966	1967
Argentina	ARG	Intentional hom	VC.IHR.PSRC.P5								
Argentina	ARG	Merchandise ex	TX.VAL.MRCH.RS	6.746363	8.794856	7.692308	7.241278	7.173774	10.18767	10.52797	7.211144
Argentina	ARG	Food exports (%	TX.VAL.FOOD.ZS.	.UN		71.51672	72.12415	77.22665	81.63538	77.91825	76.82996
Argentina	ARG	Merchandise im	TM.VAL.MRCH.R4	1.ZS				0.501561	2.095529	2.160768	0.556773
Argentina	ARG	Tariff rate, most	TM.TAX.TCOM.W	M.FN.ZS							
1 Argentina	ARG	Tariff rate, most	TM.TAX.MANF.W	/M.FN.ZS							
2 Argentina	ARG	International to	ST.INT.RCPT.XP.ZS								
Argentina	ARG	Completeness o	SP.REG.BRTH.FE.	ZS							
4 Argentina	ARG	Population ages	SP.POP.7579.FE.	0.999176	1.03297	1.071677	1.113946	1.156454	1.197503	1.220098	1.241792
5 Argentina	ARG	Population ages	SP.POP.5054.FE.	4.925295	4.97667	5.037093	5.097243	5.143884	5.170595	5.172239	5.157283
Argentina	ARG	Population ages	SP.POP.1564.MA	6560840	6653967	6753339	6857648	6964531	7072602	7171107	7271952
7 Argentina	ARG	Population ages	SP.POP.0004.FE.	10.88049	10.81417	10.72306	10.61618	10.51012	10.41663	10.3454	10.27441
Argentina	ARG	Contraceptive p	SP.DYN.CONM.Z	S							
9 Argentina	ARG	Children in emp	SL.WAG.0714.ZS								
0 Argentina	ARG	Unemployment	SL.UEM.BASC.MA	.ZS							
1 Argentina	ARG	Labor force with	SL.TLF.INTM.ZS								
2 Argentina	ARG	Labor force with	SL.TLF.ADVN.FE.Z	ZS							

Εικόνα 1: Μικρό απόσπασμα από τα στατιστικά της Αργεντίνας.

Παρατηρούμε πως χρειάζεται κανονικοποίηση των δεδομένων για μεγαλύτερη ευκολία στη χρήση και για την αποφυγή σφαλμάτων.

Σκοπός του τι θέλουμε να υλοποιήσουμε είναι η προβολή ή η σύγκριση, πολλών indicator ή χωρών μεταξύ τους, οπότε αποφασίσαμε να μετατρέψουμε και να αποθηκεύσουμε τα δεδομένα στο τραπέζι 'stats' στην παρακάτω μορφή:

Country_id	Year	Indicator1	Indicator2		IndicatorN
Country_id0	minYear	value	value		value
Country_id0	minYear+1	value	value		value
Country_id0	•••	value	value		value
Country_id0	maxYear	value	value		value
Country_id1	minYear	value	value	•••	value
Country_id1	minYear+1	value	value		value
Country_id1	•••	value	value		value
Country_id1	maxYear	value	value		value
	••••		•••	•••	•••
Country_idN	minYear	value	value		value
Country_idN	minYear+1	value	value		value
Country_idN		value	value		value
Country_idN	maxYear	value	value		value

Εικόνα 2: Stats table format

Κύριο κλειδί αποτελεί η χώρα και η χρονιά, συμβάλοντας στη κάθε χώρα να εμφανίζεται Ν φορές, όσα και τα χρόνια. Κάθε συνδυασμός χώρας και χρόνου συνοδεύεται από τις τιμές του κάθε indicator (index) για εκείνη τη συγκεκριμένη χρονιά και χώρα.

Προσφέρουμε κατ' αυτόν τον τρόπο την δυνατότητα εύρεσης πολλαπλών (ή όλων) των στατιστικών που συσχετίζονται με μία χρονιά και μία χώρα, γεγονός που μας επιτρέπει πιο εύκολα να συγκρίνουμε πολλά indicator την ίδια στιγμή.

#### Ανάλυση του τραπεζιού:

Κάθε κελί αποτελείται από μοναδική τιμή με μοναδικούς συνδυασμούς κλειδιών, οπότε το τραπέζι είναι μορφής 1NF.

Έπειτα, επειδή και η χρονιά αλλά και η χώρα καθορίζουν την τιμή μίας ένδειξης, το τραπέζι βρίσκεται και σε μορφή 2NF.

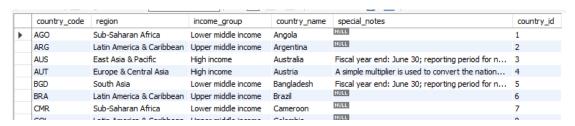
Τέλος, καμία στήλη από τα indicators δεν εξαρτάται από άλλη πέρα του primary key.

Άρα το τραπέζι μας stats είναι 3NF. Αυτό σημαίνει πως είναι απαλλαγμένο από ανωμαλίες εισαγωγής, ενημέρωσης και διαγραφής των δεδομένων της.

#### Βοηθητικά τραπέζια:

Για να είναι πιο κατηγοριοποιημένη η σχεδίαση αλλά και να έχουμε στη διάθεσή μας περισσότερες πληροφορίες, δημιουργήσαμε δύο βοηθητικά τραπέζια.

Πρώτο και πιο σχετικό τραπέζι είναι το 'countries', το οποίο περιέχει κάθε χώρα ως όνομα, κωδικό, country\_id και λοιπές πληροφορίες ή σχόλια.



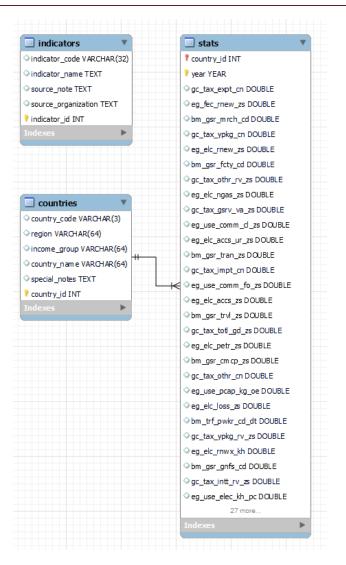
Το country\_id αποτελεί το primary key, και αυτό περνιέται και ως foreign key στο τραπέζι stats. Αποφασίσαμε να δηλώσουμε αυτό ως κλειδί διότι αποτελεί έναν απλό αριθμό και στο φόρτωμα όλου του τραπεζιού 'stats' χρειάζεται διασταύρωση του εκείνου του κλειδιού με του τραπεζιού 'countries', πράγμα που το καθιστά πιο γρήγορο άμα είναι αριθμητικό.

Επίσης είναι λιγότερο πιθανό να έχει τον ίδιο κωδικά με άλλη χώρα, όπως κάνει το country\_code που αποτελείται από 3 γράμματα και έχει περιορισμένους συνδυασμούς.

Στη συνέχεια θα δούμε το μεμονομένο τραπέζι 'indicators' το οποίο συγκρατεί πληροφορίες σχετικά με τα indicators όπως όνομα, κωδικό κλπ. Κι εκεί έχουμε βάλει ως primary key το indicator\_id για να αποφύγουμε τη χρήση μεγάλων αλφαρηθμιτικών.

	indicator_code	indicator_name	source_note	source_organization	indicator_id
•	GC.TAX.EXPT.CN	Taxes on exports (cu	Taxes on exports	International Monetary Fu	1
	EG.FEC.RNEW.ZS	Renewable energy co	Renewable energy	World Bank, Sustainable E	2
	BM.GSR.MRCH.CD	Goods imports (BoP,	Goods imports refe	International Monetary Fu	3
	GC.TAX.YPKG.CN	Taxes on income, pro	Taxes on income,	International Monetary Fu	4
	EG.ELC.RNEW.ZS	Renewable electricity	Renewable electric	IEA Statistics © OECD/IEA	5
	BM.GSR.FCTY.CD	Primary income paym	Primary income pa	International Monetary Fu	6
	GC.TAX.OTHR.RV.ZS	Other taxes (% of re	Other taxes includ	International Monetary Fu	7
	EG.ELC.NGAS.ZS	Electricity production	Sources of electrici	IEA Statistics © OECD/IEA	8
	GC.TAX.GSRV.VA.ZS	Taxes on goods and	Taxes on goods an	International Monetary Fu	9

#### 1.1 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΛΟΓΙΚΟ ΕΠΙΠΕΔΟ



Σχήμα 1.1 Σχεσιακό σχήμα της βάσης δεδομένων του συστήματος

#### **Database creation command:**

CREATE DATABASE `dbms` /\*!40100 DEFAULT CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci \*/ /\*!80016 DEFAULT ENCRYPTION='N' \*/;

#### Countries table creation command:

CREATE TABLE `countries` (
 `country\_code` varchar(3) DEFAULT NULL,
 `region` varchar(64) DEFAULT NULL,
 `income\_group` varchar(64) DEFAULT NULL,
 `country\_name` varchar(64) DEFAULT NULL,
 `special\_notes` text,
 `country\_id` int NOT NULL AUTO\_INCREMENT,
 PRIMARY KEY (`country\_id`)

) ENGINE=InnoDB AUTO\_INCREMENT=26 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

#### **Indicators table creation command:**

```
CREATE TABLE 'indicators' (
   'indicator_code' varchar(32) DEFAULT NULL,
   'indicator_name' text,
   'source_note' text,
   'source_organization' text,
   'indicator_id' int NOT NULL AUTO_INCREMENT,
   PRIMARY KEY ('indicator_id')
) ENGINE=InnoDB AUTO_INCREMENT=56 DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4_0900_ai_ci;
```

#### Stats table creation command:

```
CREATE TABLE `stats` (
`country_id` int NOT NULL,
'year' year NOT NULL,
 `gc_tax_expt_cn` double DEFAULT NULL,
 'eg_fec_rnew_zs' double DEFAULT NULL,
 `bm_gsr_mrch_cd` double DEFAULT NULL,
 `gc_tax_ypkg_cn` double DEFAULT NULL,
 `eg_elc_rnew_zs` double DEFAULT NULL,
 `bm_gsr_fcty_cd` double DEFAULT NULL,
 `gc_tax_othr_rv_zs` double DEFAULT NULL,
 `eg_elc_ngas_zs` double DEFAULT NULL,
 `gc_tax_gsrv_va_zs` double DEFAULT NULL,
 'eg_use_comm_cl_zs' double DEFAULT NULL,
 `eg_elc_accs_ur_zs` double DEFAULT NULL,
 `bm_gsr_tran_zs` double DEFAULT NULL,
 `gc_tax_impt_cn` double DEFAULT NULL,
 `eg_use_comm_fo_zs` double DEFAULT NULL,
 `eg_elc_accs_zs` double DEFAULT NULL,
 `bm_gsr_trvl_zs` double DEFAULT NULL,
 `gc_tax_totl_gd_zs` double DEFAULT NULL,
```

- `eg\_elc\_petr\_zs` double DEFAULT NULL,
- `bm\_gsr\_cmcp\_zs` double DEFAULT NULL,
- `gc\_tax\_othr\_cn` double DEFAULT NULL,
- 'eg\_use\_pcap\_kg\_oe' double DEFAULT NULL,
- 'eg\_elc\_loss\_zs' double DEFAULT NULL,
- `bm\_trf\_pwkr\_cd\_dt` double DEFAULT NULL,
- `gc\_tax\_ypkg\_rv\_zs` double DEFAULT NULL,
- 'eg\_elc\_rnwx\_kh' double DEFAULT NULL,
- `bm\_gsr\_gnfs\_cd` double DEFAULT NULL,
- `gc\_tax\_intt\_rv\_zs` double DEFAULT NULL,
- `eg\_use\_elec\_kh\_pc` double DEFAULT NULL,
- `eg\_elc\_hyro\_zs` double DEFAULT NULL,
- `bm\_trf\_prvt\_cd` double DEFAULT NULL,
- `gc\_tax\_intt\_cn` double DEFAULT NULL,
- `eg\_use\_crnw\_zs` double DEFAULT NULL,
- 'eg\_elc\_fosl\_zs' double DEFAULT NULL,
- `bm\_klt\_dinv\_wd\_gd\_zs` double DEFAULT NULL,
- `gc\_tax\_gsrv\_rv\_zs` double DEFAULT NULL,
- 'eg\_imp\_cons\_zs' double DEFAULT NULL,
- 'eg\_elc\_accs\_ru\_zs' double DEFAULT NULL,
- `bm\_gsr\_totl\_cd` double DEFAULT NULL,
- `gc\_tax\_expt\_zs` double DEFAULT NULL,
- `eg\_gdp\_puse\_ko\_pp` double DEFAULT NULL,
- `eg\_cft\_accs\_zs` double DEFAULT NULL,
- `bm\_gsr\_nfsv\_cd` double DEFAULT NULL,
- `gc\_tax\_gsrv\_cn` double DEFAULT NULL,
- 'eg\_gdp\_puse\_ko\_pp\_kd' double DEFAULT NULL,
- `eg\_egy\_prim\_pp\_kd` double DEFAULT NULL,
- `bm\_gsr\_royl\_cd` double DEFAULT NULL,
- `gc\_tax\_totl\_cn` double DEFAULT NULL,
- 'eg\_elc\_nucl\_zs' double DEFAULT NULL,
- `gc\_tax\_ypkg\_zs` double DEFAULT NULL,
- 'eg\_elc\_rnwx\_zs' double DEFAULT NULL,

```
`bm_gsr_insf_zs` double DEFAULT NULL,
```

PRIMARY KEY ('country\_id', 'year'),

CONSTRAINT `stats\_ibfk\_1` FOREIGN KEY ('country\_id') REFERENCES `countries` ('country\_id')

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

<sup>`</sup>gc\_tax\_impt\_zs` double DEFAULT NULL,

<sup>`</sup>eg\_use\_comm\_gd\_pp\_kd` double DEFAULT NULL,

<sup>`</sup>eg\_elc\_coal\_zs` double DEFAULT NULL,

<sup>`</sup>bm\_klt\_dinv\_cd\_wd` double DEFAULT NULL,

#### 1.2 ΣΧΕΣΙΑΚΟ ΣΧΗΜΑ ΣΕ ΦΥΣΙΚΟ ΕΠΙΠΕΔΟ

#### 1.2.1 ΡΥΘΜΙΣΗ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ DBMS

storage engine, memory allocation (of various kinds), ...

Η InnoDB είναι μια μηχανή αποθήκευσης που εξισορροπεί την υψηλή αξιοπιστία και την υψηλή απόδοση. Συγκεκριμένα έχει μεγαλύτερη ταχύτητα γραφής, γεγονός που συμφέρει δεόντος στην περίπτωσή μας που φορτώνουμε πολλά δεδομένα από εξωτερικά αρχεία.

Γαι την επίτρεψη φόρτωσης από τοπικά αρχεία υπήρχαν διάφορες αλλαγές που έπρεπε να γίνουν:

- 1. Εκτέλεση της mysql εντολής: SET GLOBAL local\_infile = 'ON'; (Αυτό γίνεται μέσω του κώδικα).
- 2. Τερματισμός της mysql μέσω του terminal με την εντολή: net stop MySQL80
- 3. Άνοιγμα της επιλογής εμφάνισης hidden files,
- 4. Άνοιγμα του αρχείου" C:\ProgramData\MySQL\MySQL Server 8.0\my.ini" και
- 5. Αλλαγή της μεταβλητής secure-file-priv σε "" (0-length string).
- 6. Έναρξη της mysql μέσω του terminal με την εντολή: net start MySQL80

Τέλος, θέτουμε το μέγεθος του buffer της mysql στο ιδανικότερο για την εφαρμογή, εκτελώντας την mysql εντολή:

- SET GLOBAL innodb\_buffer\_chuck\_size = 1.048.576;
- SET GLOBAL innodb\_buffer\_pool\_size = 1.048.576;

#### 1.2.2 ΡΥΘΜΙΣΗ ΤΟΥ ΦΥΣΙΚΟΥ ΣΧΗΜΑΤΟΣ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

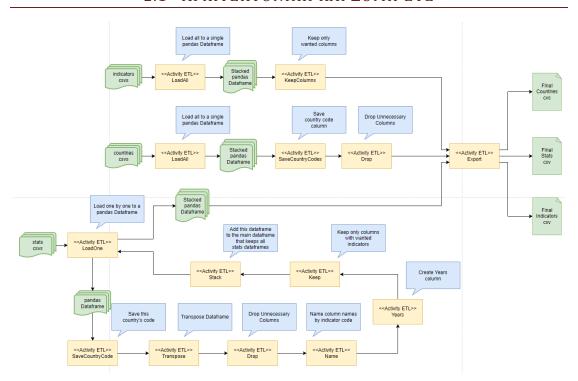
Όπως υπώθηκε προηγουμένως έχουμε δημιουργήσει τη βάση με τον σκοπό της στο μυαλό μας, αντιπροσωπεύοντάς την καλύτερα στις ανάγκες/απαιτήσεις της εφαρμογής.

#### 1.2.3 ΡΥΘΜΙΣΗ ΑΣΦΑΛΕΙΑΣ

Έχει δημιουργηθεί ένας backup φάκελος με τα τραπέζια του dmbs ως .csv αρχεία, μέσω του export wizard του mysql workspace.

#### 2 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΛΟΓΙΣΜΙΚΟΥ

#### 2.1 APXITEKTONIKH KAI ΔOMH ETL



Εικόνα 3: Αρχιτεκτονική φόρτωσης και μετατροπής δεδομένων

Κάθε .zip αρχείο που κατεβάζαμε περιείχε 3 csv αρχεία, το καθένα με διαφορετικά δεδομένα. Τα χωρίσαμε σε 3 φακέλους βάση κατηρογίας.



Έτσι κάθε χώρα έχει από ένα αρχείο σε κάθε αντιπροσωπευτικό φάκελο.

Εικόνα 4: /dbms/data/original/

Βάση κάθε φακέλου παράγουμε ένα αντίστοιχο τελικό αρχείο, περιέχοντας όλες τις ζητούμενες πληροφορίες από τα αρχεία εκείνου του φακέλου.

Η διαδικασία μετατροπής περιγράφεται παραπάνω.

#### Χειροκίνητες αλλαγές στο σύστημα:

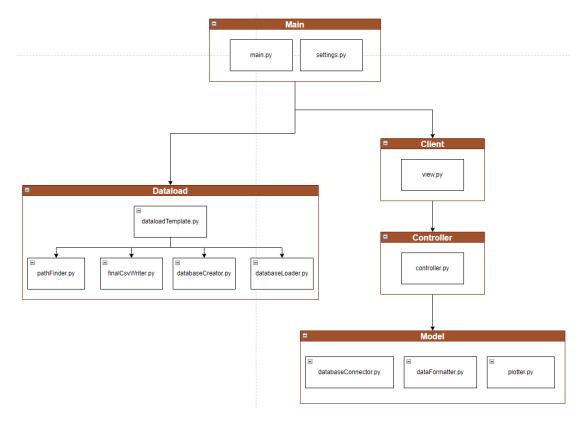
Για τη μετατροπή των αρχείων σε csv τελικής μορφής δεν απαιτούνται αλλαγές στον κώδικα.

Ωστόσο για τη σύνδεση στην βάση χρειάζεται αλλαγή των παραμέτρω σύνδεσης, δεδομένου ότι ο χρήστης χρησιμοποιεί διαφορετικό user & password.

Αρκεί ο χρήστης να πάει στον φάκελο '../DBMS/src/' στο αρχείο settings.py και να βάλει στα user, password (και άμα θέλει database) τα δικά του δεδομένα.

## 2.2 ΔΙΑΓΡΑΜΜΑΤΑ ΠΑΚΕΤΩΝ / ΥΠΟΣΥΣΤΗΜΑΤΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ

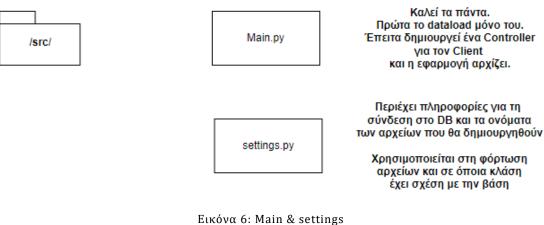
Το διάγραμμα για τα υποσυστήματα / πακέτα του λογισμικού που κατασκευάσατε ως κεντρική εφαρμογή επερώτησης. Ο στόχος είναι να φανεί η high-level αρχιτεκτονική του συστήματος, χωρίς λεπτομέρειες των επί μέρους κλάσεων. Κάποιος πολύ σύντομος σχολιασμός επίσης.

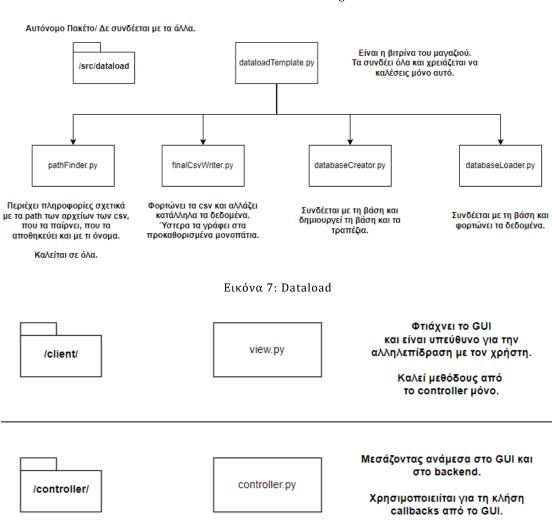


Εικόνα 5: High-level αρχιτεκτονική του συστήματος / packages.

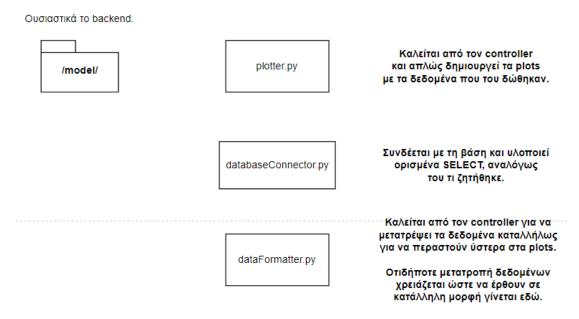
#### ΔΙΑΓΡΑΜΜΑ(ΤΑ) ΚΛΑΣΕΩΝ ΚΕΝΤΡΙΚΗΣ ΕΦΑΡΜΟΓΗΣ 2.3

Έχει χρησιμοποιηθεί αρχιτεκτονική Model-Controller-View για διαχωρισμό υπευθυνοτήτων και πιο εύκολη ανάπτυξη του συστήματος.

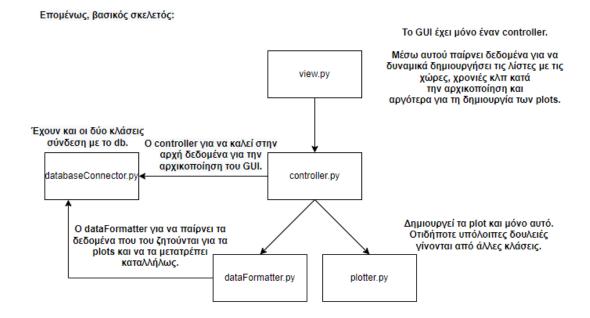




Εικόνα 8: Client & Controller



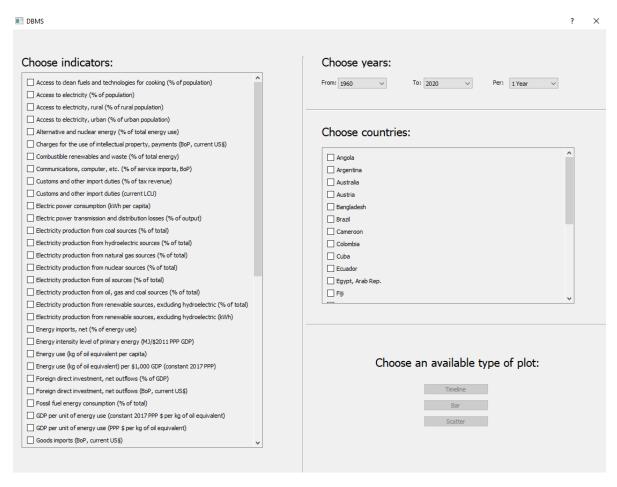
Εικόνα 9: Plotter, databaseConnector & dataFormatter



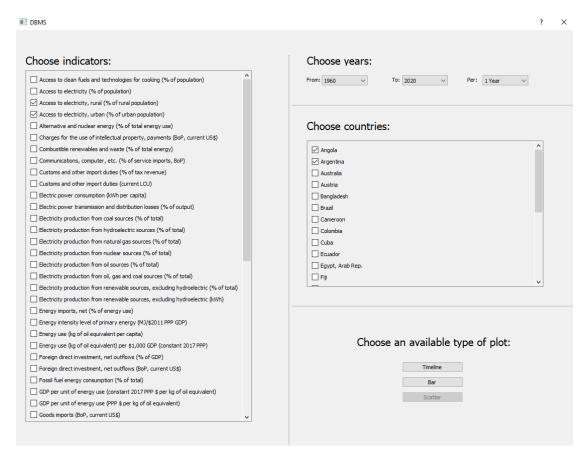
Εικόνα 10: Model package interactions

#### ΥΠΟΔΕΙΓΜΑΤΑ ΕΡΩΤΗΣΕΩΝ ΚΑΙ ΑΠΑΝΤΗΣΕΩΝ

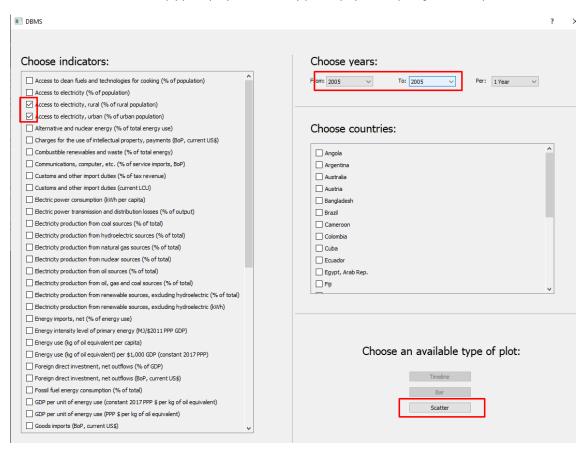
- 1) Κατεύθυνση στο directory .../DMBS/src/ και άνοιγμα της εφαρμογής με την εκκίνηση του main.py.
- 2) Ο χρήστης επιλέγει τις χώρες, χρονιές και στατιστικές της επιλογής του και του ανοίγουν οι αντίστοιχες επιλογές για δημιουργία plot.
- 3) Δυνατές επιλογές:
  - a. Timeline Plots: k-στατιστικές και k'-χώρες
  - b. Bar Plots: k-στατιστικές και k'-χώρες
  - Scatter Plots: 2 στατιστικές και
    - Ίδια χρονιά, καμία χώρα. (δείχνει για όλες μαζί εκείνη τη χρονιά)
    - ii. Διαφορετικές χρονιές, μία χώρα



Εικόνα 11: Αρχική κατάστασης της εφαρμογής



Εικόνα 12: Επιλογή μεταβλητών και ενεργοποίηση ανάλογων plot επιλογών.



Εικόνα 13: Επιλογή 2 χωρών και μίας χρονιάς δίχως χώρα για την ενεργοποίηση scatter

#### 4 ΛΟΙΠΑ ΣΧΟΛΙΑ

Τα Tests τρέχουν με το τρέξιμο του tests.py, το οποίο βρίσκεται στον ίδιο φάκελο με την main.