



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΕΡΓΑΣΤΗΡΙΟ ΣΥΣΤΗΜΑΤΩΝ ΒΑΣΕΩΝ ΓΝΩΣΕΩΝ ΚΑΙ ΔΕΔΟΜΕΝΩΝ
Ακ. έτος 2023-2024, 6ο εξάμηνο, ΣΗΜΜΥ

**Βάσεις Δεδομένων
Εξαμηνιαία Εργασία**

Ονοματεπώνυμο : Αγγελική Σπανού - Καπαντώνη, Χαράλαμπος-Ιωάννης Σφυρής, Αγγελική Γεράρδου

ΑΜ: 03121090, 03121044, 03120110

Ημερομηνία: 19/05/2024

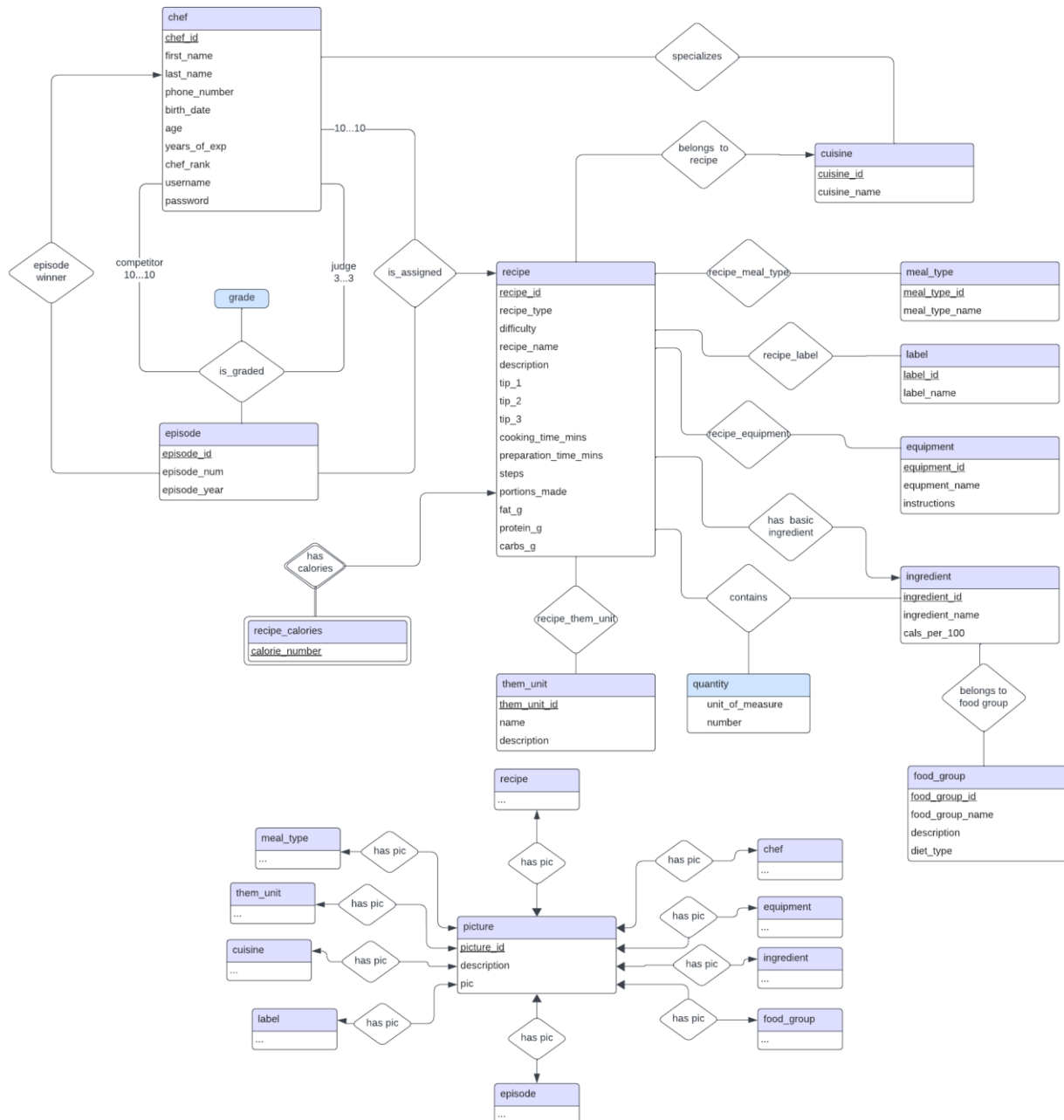
GitHub Repository: <https://github.com/harrissfiris/DB-Project-2024.git>

Περιεχόμενα

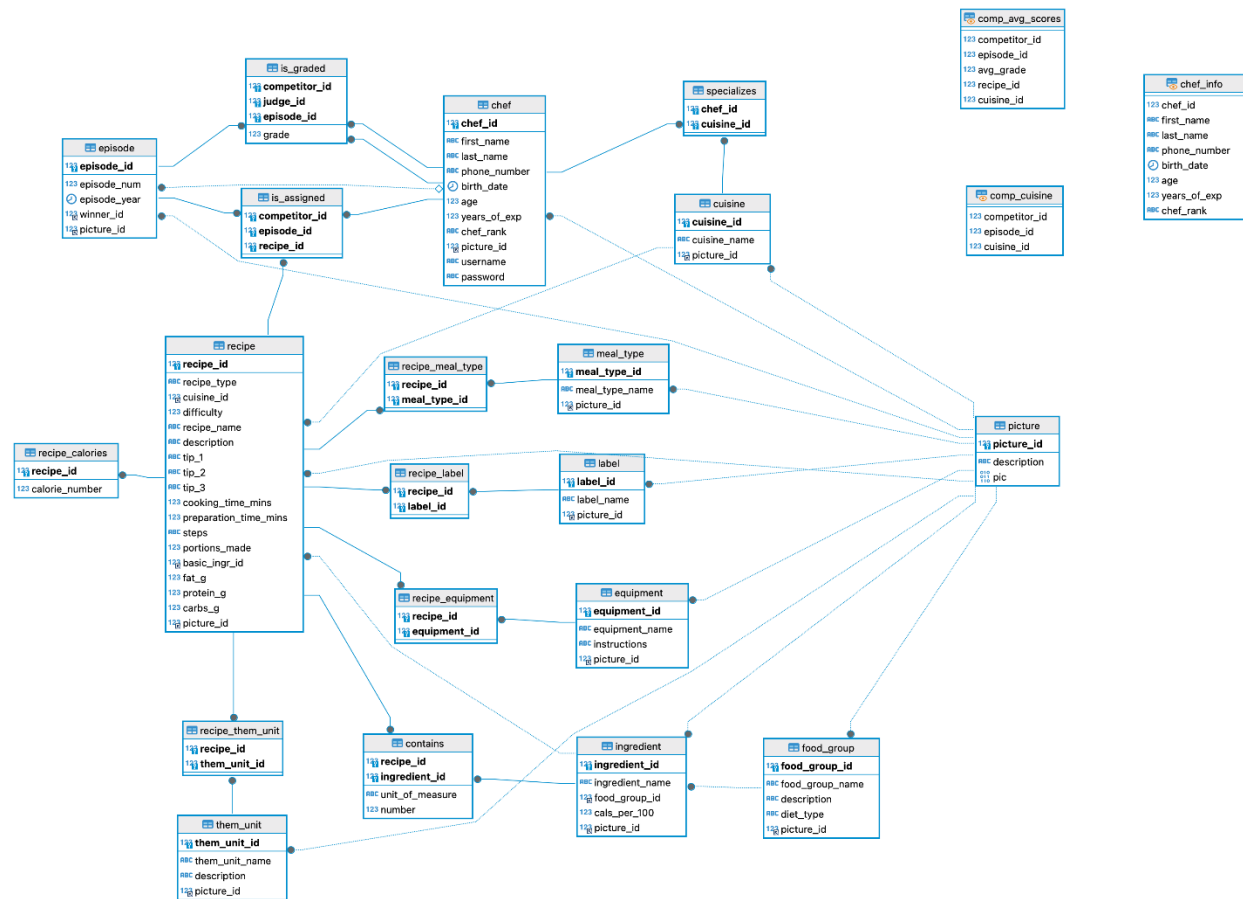
Περιεχόμενα	2
Βάση Δεδομένων	3
Διάγραμμα Οντοτήτων-Συσχετίσεων (Entity Relationship).....	3
Σχεσιακό Διάγραμμα (Relational Schema)	4
Σχολιασμός Διαγράμματος ER και Relational Schema.....	4
Ευρετήρια	6
DDL Script	9
DML Script.....	9
Οδηγίες Εγκατάστασης	10

Βάση Δεδομένων

Διάγραμμα Οντοτήτων-Συσχετίσεων (Entity Relationship)



Σχεσιακό Διάγραμμα (Relational Schema)



Σχολιασμός Διαγράμματος ER και Relational Schema

Η Βάση Δεδομένων που καλούμαστε να υλοποιήσουμε αφορά δημοφιλή διαγωνισμό μαγειρικής. Πιο συγκεκριμένα, ζητείται να σχεδιάσουμε και να υλοποιήσουμε ένα σύστημα αποθήκευσης και διαχείρισης των πληροφοριών που απαιτούνται για τη λειτουργία του διαγωνισμού αναφορικά με τις συνταγές, τα υλικά, τον εξοπλισμό, τους συμμετέχοντες και τα επεισόδια.

Στη βάση περιέχονται **5 κύριες οντότητες**:

1. Συνταγές (recipes):

Κεντρική οντότητα που περιλαμβάνει πληροφορίες όπως το όνομα, το id, μια σύντομη περιγραφή, το επίπεδο δυσκολίας (από 1 έως 5), την εθνική κουζίνα στην οποία ανήκει κ.ά.. Κάθε συνταγή αναφέρεται σε ειδικές κατηγορίες γευμάτων και μπορεί να έχει επιπλέον ετικέτες για πιο εξειδικευμένη κατηγοριοποίησή της.

2. Συστατικά (ingredients):

Κάθε συνταγή έχει συγκεκριμένα υλικά, τα οποία κατηγοριοποιούνται σύμφωνα με τις Ομάδες Τροφίμων.

3. Εξοπλισμός (equipment):

Οι συνταγές χρειάζονται συγκεκριμένο εξοπλισμό για να υλοποιηθούν με τις απαραίτητες οδηγίες χρήσης.

4. Μάγειρες (chefs):

Κάθε μάγειρας μπορεί να εκτελέσει συνταγές που ανήκουν στις Εθνικές Κουζίνες που έχει εξειδίκευση. Στο διαγωνισμό κάποιος μάγειρας μπορεί να αναλάβει τον ρόλο του κριτή και του διαγωνιζόμενου αλλά όχι και τους δύο στο ίδιο επεισόδιο.

5. Επεισόδια (episodes):

Κάθε σεζόν του διαγωνισμού αποτελείται από 10 επεισόδια. Σε κάθε επεισόδιο διαγωνίζονται 10 μάγειρες και υπάρχουν 3 κριτές. Από τον μέσο όρο των βαθμολογιών των κριτών προκύπτει ο νικητής του κάθε επεισοδίου.

και άλλες οντότητες που σχετίζονται με αυτές:

- **Εθνικές κουζίνες (cuisine)**
- **Μορφές γεύματος (meal_type)**
- **Ετικέτες (labels)**
- **Θεματικές ενότητες (them_unit)**
- **Ομάδες Τροφίμων (food_group)**
- **Φωτογραφίες (picture)**
- **Θερμίδες Συνταγών (recipes_calories)** που έχει προστεθεί σαν *weak entity set* στον πίνακα *recipe*.

Κάθε οντότητα έχει **Primary Keys** για να εξασφαλίζεται η μοναδικότητα κάθε εγγραφής, καθώς και **Foreign Keys** για να συνδέεται με μια άλλη οντότητα στη βάση.

Οι οντότητες στη βάση συνδέονται μεταξύ τους με τα παρακάτω junction tables:

- **Contains:** Many to Many σχέση μεταξύ recipe και ingredient. Μια συνταγή μπορεί να περιλαμβάνει πολλά συστατικά και ένα συστατικό μπορεί να χρησιμοποιηθεί σε πολλές συνταγές. Αυτή η σχέση διαχειρίζεται τις ποσότητες

των συστατικών που απαιτούνται για κάθε συνταγή, γι' αυτό και προσθέτουμε επιπλέον attributes στο junction table "contains".

- **Recipe_equipment:** Many to many σχέση που συνδέει κάθε συνταγή με τον απαραίτητο εξοπλισμό.
- **Recipe_label:** Many to Many σχέση μεταξύ συνταγής και ετικετών.
- **Is_assigned:** Many to Many σχέση μεταξύ episode και chef. Ένα επεισόδιο μπορεί να έχει πολλούς συμμετέχοντες μάγειρες και ένας μάγειρας μπορεί να συμμετέχει σε πολλά επεισόδια.
- **Is_graded:** Many to Many σχέση μεταξύ Chef και episode. Στο junction table έχει προστεθεί το attribute "grade" που αφορά τον βαθμό που έχει λάβει κάθε διαγωνιζόμενος από κάποιον κριτή.
- **Recipe_meal_type:** Many to Many σχέση μεταξύ recipe και meal type
- **Specializes:** Many to Many σχέση μεταξύ chef και cuisine, καθώς κάθε μάγειρας μπορεί να έχει εξειδίκευση σε παραπάνω από μία εθνικές κουζίνες.
- **Recipe_them_units:** Many to One σχέση μεταξύ recipe και them_units , καθώς κάθε συνταγή αντιστοιχεί σε μία θεματική ενότητα και κάθε θεματική ενότητα αντιστοιχεί σε πολλές συνταγές.

Οι παραπάνω σχέσεις υλοποιούνται μέσω των junction tables που περιλαμβάνουν τα αντίστοιχα ID των οντοτήτων που αφορούν.

Ευρετήρια

Τα ευρετήρια αυξάνουν την απόδοση της βάσης δεδομένων. Τα ερωτήματα της εργασίας απαιτούν αναζήτηση πολλών στηλών πινάκων της βάσης η οποία θα μπορούσε για κάθε ερώτημα ξεχωριστά να βελτιστοποιηθεί μέσω ενός index. Εμείς επιλέξαμε να μειώσουμε την χρήση τους καθώς σε πολλές περιπτώσεις η ενημέρωση των στοιχείων του ευρετηρίου μπορεί να αποδειχθεί χρονοβόρα. Για αυτό προσπαθήσαμε να προσθέσουμε ευρετήρια μόνο για δεδομένα πινάκων τα οποία εμφανίζονται συχνά σε ερωτήματα της βάσης , πέρα από τα ήδη υπάρχοντα ευρετήρια για τα primary keys.

Αρχικά έχουν δημιουργηθεί ευρετήρια (indexes) τύπου B+ Tree για όλα τα primary keys και για όλα τα attributes με UNIQUE constraint. Επιπρόσθετα από αυτά, δημιουργήσαμε τα παρακάτω:

- **Competitor** και **Recipe** από το table **is_assigned**
Για τα ερωτήματα 3.1, 3.3.
- **Competitor** από το table **is_assigned**

Για τα ερωτήματα που μας ενδιαφέρει μόνο ποιοι chef έχουν διαγωνιστεί (3.7 , 3.12 ,3.13) .

- **Episode** από το table **is_assigned** για το ερώτημα 3.8 στο εναλλακτικό query plan χρησιμοποιήθηκε ως forced index και για το ερώτημα 3.9 .
- **Recipe** και **Episode** από το table **is_assigned**
Για ερωτήματα στα οποία μας ενδιαφέρουν οι συνταγές που έχουν μαγειρευτεί στον διαγωνισμό αλλά όχι οι chef που τις μαγείρεψαν (3.6, 3.8, 3.10, 3.14, 3.15).
- **Chef_id , First_name** και **Last_name** από το table **chef**
Για ερωτήματα στα οποία ζητείται να εμφανιστούν τα στοιχεία των chef που ικανοποιούν μια συγκεκριμένη συνθήκη (3.1, 3.2, 3.3, 3.4, 3.5, 3.7, 3.11).
- **Episode_id** και **Year** από το table **episode**
Για ερωτήματα στα οποία αναζητούμε μόνο το έτος του κάθε επεισοδίου (3.5, 3.9, 3.10, 3.12, 3.13)
- **Ingredient_id** και **Cals_per_100** από το table **ingredient**
Για τον υπολογισμό των θερμίδων της κάθε συνταγής.
- **Recipe_id** και **Cuisine_id** από το table **recipe**
Για ερωτήματα στα οποία αναζητούμε τις κουζίνες που μαγειρεύτηκαν στον διαγωνισμό χωρίς να θέλουμε πληροφορίες για τις συνταγές (3.1, 3.2, 3.10).
- **carbs_g** από το table **recipe** για το ερώτημα 3.9
- **difficulty** από το table **recipe** για το ερώτημα 3.12
- **chef_rank** από το table **chef** για το ερώτημα 3.13

Επιπλέον πρέπει να σημειωθεί ότι για τα Queries 3.6 και 3.8 δημιουργήθηκαν εναλλακτικά Query Plans με Forced indexes. Αυτό γίνεται με σκοπό να βελτιστοποιηθεί η εκτέλεση του query επιλέγοντας αναγκαστικά ποια indexes θα χρησιμοποιηθούν. Σε πολλές περιπτώσεις παρατηρείται βελτίωση ενώ σε άλλες όχι καθώς τα ίδια indexes θα χρησιμοποιούνταν εξ αρχής από το βασικό query plan.

Για να δούμε αν στην περίπτωση των queries 3.6 και 3.8 έχουμε βελτίωση παρατηρούμε τις αναζητήσεις που έγιναν με και χωρίς forced indexes.

3.6 χωρίς:

123 id	ABC select_type	ABC table	ABC partitions	ABC type	ABC possible_keys
1	SIMPLE	rl1	[NULL]	index	PRIMARY,recipe_label_label_FK
1	SIMPLE	rl2	[NULL]	ref	PRIMARY,recipe_label_label_FK
1	SIMPLE	ia	[NULL]	ref	is_assigned_recipe_FK,is_assigned_recipe_id

ABC key	ABC key_len	ABC ref	123 rows	123 filtered	ABC Extra
recipe_label_label_FK	4	[NULL]	117	100	Using index; Using temporary; Using filesort
PRIMARY	4	masterchef.rl1.recipe_id	1	33.33	Using where; Using index
is_assigned_recipe_FK	4	masterchef.rl1.recipe_id	8	100	Using index

3.6 με forced indexes:

123 id	ABC select_type	ABC table	ABC partitions	ABC type	ABC possible_keys
1	SIMPLE	rl1	[NULL]	index	PRIMARY
1	SIMPLE	rl2	[NULL]	ref	PRIMARY
1	SIMPLE	ia	[NULL]	ref	is_assigned_recipe_FK

ABC key	ABC key_len	ABC ref	123 rows	123 filtered	ABC Extra
PRIMARY	8	[NULL]	117	100	Using index; Using temporary; Using filesort
PRIMARY	4	masterchef.rl1.recipe_id	1	33.33	Using where; Using index
is_assigned_recipe_FK	4	masterchef.rl1.recipe_id	8	100	Using index

3.8 χωρίς:

123 id	ABC select_type	ABC table	ABC partitions	ABC type	ABC possible_keys	ABC key
1	SIMPLE	e	[NULL]	index	PRIMARY,episode_unique	PRIMARY
1	SIMPLE	ia	[NULL]	ref	is_assigned_episode_FK,i	is_assigned_episode_FK
1	SIMPLE	re	[NULL]	ref	PRIMARY	PRIMARY

ABC key_len	ABC ref	123 rows	123 filtered	ABC Extra
4	[NULL]	57	100	Using temporary; Using filesort
4	masterchef.e.episode_id	9	100	Using index
4	masterchef.ia.recipe_id	2	100	Using index

3.8 με forced indexes:

123 id	ABC select_type	ABC table	ABC partitions	ABC type	ABC possible_keys	ABC key
1	SIMPLE	e	[NULL]	index	PRIMARY	PRIMARY
1	SIMPLE	ia	[NULL]	ref	is_assigned_episode_FK	is_assigned_episode_FK
1	SIMPLE	re	[NULL]	ref	PRIMARY	PRIMARY

ABC key_len	ABC ref	123 rows	123 filtered	ABC Extra
4	[NULL]	57	100	Using temporary; Using filesort
4	masterchef.e.episode_id	9	100	Using index
4	masterchef.ia.recipe_id	2	100	Using index

Καταλήγουμε στο συμπέρασμα ότι τα forced indexes δεν βελτιστοποίησαν περεταίρω τον κώδικα καθώς οι αναζητήσεις με και χωρίς αυτά είναι ίσες και για τα δύο queries. Επομένως και με το βασικό query plan χρησιμοποιούνταν τα κατάλληλα indexes.

DDL Script

Το DDL Script είναι γραμμένο σε SQL και η εκτέλεση του κατασκευάζει τα tables, τα indexes, τα views, τα triggers και τα stored procedures της βάσης δεδομένων του διαγωνισμού με τα constraints και τα foreign/primary keys που έχουμε ορίσει. Ο κώδικας είναι μεγάλος για αυτό επισυνάπτουμε τον σύνδεσμο για το GitHub repository στο οποίο βρίσκεται.

https://github.com/harrissfiris/DB-Project-2024/blob/a9b0f8d8952fd1de4b5c5f765124312f9d14f239/manual%20download/masterchef_DDL.sql

DML Script

Το DML Script κάνει populate την βάση μας με τα δεδομένα που επιλέξαμε, τα οποία βρίσκονται και σε csv files στο GitHub. Ο κώδικας είναι μεγάλος για αυτό επισυνάπτουμε τον σύνδεσμο για το GitHub repository στο οποίο βρίσκεται.

https://github.com/harrissfiris/DB-Project-2024/blob/a9b0f8d8952fd1de4b5c5f765124312f9d14f239/manual%20download/masterchef_DML.sql

Οδηγίες Εγκατάστασης

Αρχικά πρέπει να εγκαταστήσετε το DBMS MySQL (<https://www.mysql.com/>, ανάλογα με το λογισμικό που χρησιμοποιείτε). Προαιρετικά μπορείτε να κατεβάσετε από το ίδιο μέρος και το Graphical Client MySQLWorkbench. Θα σας ζητηθεί να δώσετε password για τον root user. Στη συνέχεια, ως root user, συνδεθείτε στην mysql με την εντολή `mysql -u root -p -h localhost`. Τώρα που συνδεθήκατε στην mysql αρκεί να τρέξετε τα δύο SQL αρχεία `masterchef_DDL.sql`, `masterchef_DML.sql` για να δημιουργηθεί η βάση δεδομένων masterchef, να φτιαχτεί η δομή της και να γεμίσει με δεδομένα. Έπειτα τρέξτε το `create_users.sql` που δημιουργεί τους χρήστες της βάσης.

Επίσης για τη βάση μας κάνουμε χρήση 3 αρχείων python, το `episode_generator` το οποίο δημιουργεί τα επεισόδια (το έχουμε χρησιμοποιήσει και έχει ήδη 60 επεισόδια η βάση), το `DB_PIC_INSERT` για την εισαγωγή φωτογραφιών στον πίνακα `picture` της βάσης και το `show_picture` για την εμφάνιση φωτογραφιών από τον πίνακα `picture` της βάσης. Για να εκτελέσετε αυτά τα αρχεία χρειάζεται να εγκαταστήσετε την python3 και σε ένα virtual environment να εγκαταστήσετε την βιβλιοθήκη `mysql.connector`. Πιο αναλυτικά:

- Στο terminal δημιουργήστε ένα virtual environment και ενεργοποιήστε το:

```
python3 -m venv myenv
```

```
source myenv/bin/activate
```

- Κάνετε εγκατάσταση τη βιβλιοθήκη `connector`

```
pip install mysql-connector-python
```

- Και επιβεβαιώστε την εγκατάσταση

```
python -c "import mysql.connector; print('mysql-connector-python installed successfully')"
```

- Όταν τελειώστε απενεργοποιήστε το virtual environment:

```
deactivate
```

Εδώ να σημειώσουμε ότι εμείς χρησιμοποιήσαμε VS Code και χρειάστηκε επίσης να προσθέσουμε manually το virtual environment κάνοντας `Ctrl + Shift + P`, `Preferences: Open`

Workspace Settings, και στο αρχείο `settings.json` προσθέσαμε το `{ "python.pythonPath": ".../bin/python" }`

Και στη συνέχεια, *Ctrl + Shift + P, Python: Select Interpreter*, όπου επιλέξαμε αυτόν που χρησιμοποιεί το `myenv`.

Επίσης για να τρέξετε τα προγράμματα που εισάγουν ή εμφανίζουν φωτογραφίες από τη βάση χρειάζεται να εγκαταστήσετε τη βιβλιοθήκη *PIL*.