



Πανεπιστήμιο Θεσσαλίας  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

## **Εξαγωγή Πληροφορίας από Δεδομένα Έργων Τέχνης με χρήση Μεθόδων Αναλυτικής Επεξεργασίας**

**Διπλωματική Εργασία**

**ΕΥΑΓΓΕΛΙΑΣ ΚΩΝΣΤΑΝΤΙΝΟΥ**

**Επιβλέπων**

Μιχαήλ Βασιλακόπουλος  
Αναπληρωτής Καθηγητής

Βόλος, Οκτώβριος 2019





Πανεπιστήμιο Θεσσαλίας

Πολυτεχνική Σχολή

Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

# Εξαγωγή Πληροφορίας από Δεδομένα Έργων Τέχνης με χρήση Μεθόδων Αναλυτικής Επεξεργασίας

## Διπλωματική Εργασία

**ΕΥΑΓΓΕΛΙΑΣ ΚΩΝΣΤΑΝΤΙΝΟΥ**

Επιτροπή επίβλεψης

Επιβλέπων

Μιχαήλ Βασιλακόπουλος  
Αναπληρωτής Καθηγητής

Συνεπιβλέπουσα

Ελένη Τουσίδου  
Μέλος Ε.ΔΙ.Π.

Συνεπιβλέπουσα

Τσαλαπάτα Χαρίκλεια  
Μέλος Ε.ΔΙ.Π.

Βόλος, Οκτώβριος 2019



Πανεπιστήμιο Θεσσαλίας  
Πολυτεχνική Σχολή  
Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών Υπολογιστών

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή / της φοιτήτριας που την εκπόνησε. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

Ο/Η συγγραφέας αυτής της εργασίας βεβαιώνει ότι κάθε βοήθεια την οποία είχε για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης βεβαιώνει ότι έχει αναφέρει τις όποιες πηγές από τις οποίες έκανε χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται επακριβώς, είτε παραφρασμένες.



University of Thessaly  
Faculty of Engineering  
Department of Electrical & Computer Engineering

# **Extracting Information from Artwork Data using OLAP**

## **Diploma Thesis**

**EVANGELIA KONSTANTINOY**

**Supervisor**

Michael Vassilakopoulos  
Associate professor

Volos, October 2019



# Περίληψη

Τα τελευταία χρόνια οι περισσότερες επιχειρήσεις έχουν δημιουργήσει αποθήκες δεδομένων, ώστε να αποθηκεύουν ιστορικά δεδομένα και συνοπτικές πληροφορίες που είναι χρήσιμες στη διαδικασία λήψης αποφάσεων. Ως εκ τούτου, έχει παρουσιαστεί η ανάγκη για εργαλεία ανάλυσης που διευκολύνουν την επεξεργασία τέτοιου είδους δεδομένων, αλλά και την εξαγωγή πληροφορίας από αυτά. Η παρούσα διπλωματική εργασία έχει ως αντικείμενο, την εξαγωγή πληροφορίας από δεδομένα έργων τέχνης του μουσείου MoMa (Museum of Modern Art), με τη μέθοδο της άμεσης αναλυτικής επεξεργασίας (OLAP). Σκοπός της εργασίας είναι, η πληροφορία που θα προκύψει να μπορεί να αξιοποιηθεί είτε από στελέχη/αναλυτές επιχειρήσεων για την υποστήριξη λήψης αποφάσεων, είτε από Διεθνείς οργανισμούς ή Υπουργεία Πολιτισμού για την δημιουργία πολιτισμικών στατιστικών.

Για την αναλυτική επεξεργασία των δεδομένων χρησιμοποιείται το OLAP framework της προγραμματιστικής γλώσσας Python, Cubes. Η εργασία περιέχει μία εκτεταμένη περιγραφή του παραπάνω εργαλείου καθώς και του τρόπου χρήσης του. Για την οπτικοποίηση των δεδομένων και την εξαγωγή πληροφορίας από αυτά, χρησιμοποιήθηκε το εργαλείο CubesViewer. Στο τέλος της εργασίας, παρουσιάζονται τα γραφήματα των αποτελεσμάτων της αναλυτικής επεξεργασίας των δεδομένων και αναλύονται τα συμπεράσματα που προκύπτουν από αυτά.

## Λέξεις Κλειδιά

Άμεση Αναλυτική Επεξεργασία, Αποθήκες Δεδομένων, Λήψη Αποφάσεων, Python, Cubes Framework, Οπτικοποίηση Δεδομένων, CubesViewer tool





# Abstract

In recent years, most businesses have created data warehouses to store historical data and summary information, useful in the decision-making process. Therefore, there is a need for analytical tools that facilitate the process and extraction of such data. The subject of this thesis is to extract information from Museum of Modern Art (MoMa) artwork data, using the method of Online Analytical Processing (OLAP). The purpose of this thesis is to provide information that can be used by business executives/analysts to support decision-making, or by international organizations or Ministries of Culture for the creation of cultural statistics.

The OLAP framework Cubes, of the Python programming language, is used for data processing. The thesis contains an extensive description of the above tool as well as how to use it. The CubesViewer tool was used to visualize the data and extract information from it. Finally, the graphs of the processed data are presented, along with the analyzed conclusions derived from them.

## Keywords

OLAP, Data Warehouses, Decision Making, Python, Cubes Framework, Data Visualization, CubesViewer tool



*Στους αγαπημένους μου φίλους Lucky, Leo και Keani.*



# Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τους καθηγητές κ.Βασιλακόπουλο Μιχαήλ, κ.Τουσίδου Ελένη και κ.Τσαλαπάτα Χαρίκλεια, για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσαν να την εκπονήσω. Ιδιαίτερα, θα ήθελα να ευχαριστήσω την κ.Τουσίδου για την πολύτιμη βοήθεια και την καθοδήγηση της, καθ' όλη τη διάρκεια εκπόνησης της εργασίας αυτής.

Ένα μεγάλο ευχαριστώ θα ήθελα να εκφράσω στην οικογένειά μου, στους γονείς μου, Σπύρο και Βιβή και στην αδερφή μου Νεφέλη, για την αγάπη και την συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια. Πάντα ήταν στο πλάι μου και φρόντισαν να θέσουν γερές βάσεις για την ανάπτυξή μου. Ακόμα θα ήθελα να ευχαριστήσω όλους τους φίλους μου για τις όμορφες στιγμές που περάσαμε. Τέλος, θα ήθελα να ευχαριστήσω τον Λευτέρη, για την αγάπη και την στήριξή του όλο αυτό το διάστημα.



# Πρόλογος

Η παρούσα διπλωματική εργασία, εκπονήθηκε ως το τελευταίο βήμα για την απόκτηση του διπλώματος του τμήματος Ηλεκτρολόγων Μηχανικών & Μηχανικών Ηλεκτρονικών Υπολογιστών του Πανεπιστημίου Θεσσαλίας. Η εργασία πραγματοποιήθηκε υπό την επίβλεψη των καθηγητών κ. Βασιλακόπουλου Μιχαήλ και κ. Τουσίδου Ελένης. Συνεπιβλέπουσα ήταν η καθηγήτρια κ. Τσαλαπάτα Χαρίκλεια.





# Περιεχόμενα

<b>Περίληψη</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Ευχαριστίες</b>	<b>vii</b>
<b>Πρόλογος</b>	<b>ix</b>
<b>Περιεχόμενα</b>	<b>xi</b>
<b>Κατάλογος σχημάτων</b>	<b>xiii</b>
<b>Κατάλογος πινάκων</b>	<b>xv</b>
<b>1 Εισαγωγή</b>	<b>1</b>
1.1 Αντικείμενο της διπλωματικής . . . . .	1
1.1.1 Συνεισφορά . . . . .	2
1.2 Οργάνωση του τόμου . . . . .	2
<b>2 Θεωρητικό Υπόβαθρο</b>	<b>3</b>
2.1 Εισαγωγή . . . . .	3
2.2 Ανάκτηση Πληροφορίας . . . . .	3
2.3 Μέθοδοι Αναλυτικής Επεξεργασίας για Υποστήριξη Λήψης Αποφάσεων . . . . .	4
2.4 Εξαγωγή Πληροφορίας και Cultural Analytics . . . . .	5
2.5 Αποθήκες Δεδομένων (Data Warehouses) . . . . .	6
2.6 Άμεση Αναλυτική Επεξεργασία - OLAP . . . . .	6
<b>3 Cubes: Python Framework, OLAP Server</b>	<b>13</b>
3.1 Εισαγωγή . . . . .	13
3.2 Λογικό Μοντέλο (Model) . . . . .	13
3.3 Περιηγητής (Browser) . . . . .	19
3.4 Εξυπηρετητής (Slicer Server) . . . . .	22
3.5 Backends . . . . .	23
3.6 Εγκατάσταση του Cubes . . . . .	26

<b>4</b>	<b>Αναλυτική Επεξεργασία Δεδομένων Έργων Τέχνης με τη χρήση του Cubes</b>	<b>27</b>
4.1	Εισαγωγή . . . . .	27
4.2	Σετ Δεδομένων Έργων Τέχνης . . . . .	27
4.3	Μετατροπή του CSV αρχείου σε SQLite Table . . . . .	29
4.4	Υλοποίηση Λογικού Μοντέλου . . . . .	30
4.5	Υλοποίηση του Εξυπηρετητή Slicer . . . . .	36
<b>5</b>	<b>Χρήση του CubesViewer</b>	<b>39</b>
5.1	Εισαγωγή . . . . .	39
5.2	Παρουσίαση του Cubes Viewer . . . . .	39
5.2.1	Εγκατάσταση και τρόπος λειτουργίας του CubesViewer . . . . .	40
5.3	Οπτικοποίηση του κύβου δεδομένων μέσω του εργαλείου CubesViewer . . . . .	43
<b>6</b>	<b>Εξαγωγή Πληροφορίας από την Ανάλυση Δεδομένων</b>	<b>49</b>
6.1	Γενικά Συμπεράσματα . . . . .	49
6.2	Εξαγωγή Πληροφορίας από τις Αναλυτικές Καθόδους . . . . .	50
6.2.1	Αναλυτική Κάθοδος κατά Κατηγορία . . . . .	50
6.2.2	Αναλυτική Κάθοδος κατά Τρίμηνα του Έτους . . . . .	51
6.2.3	Αναλυτική Κάθοδος κατά Εθνικότητα . . . . .	53
6.2.4	Αναλυτική Κάθοδος κατά Ηλικιακό γκρούπ . . . . .	57
6.2.5	Αναλυτική Κάθοδος κατά Φύλο . . . . .	59
<b>7</b>	<b>Επίλογος</b>	<b>63</b>
7.1	Σύνοψη και συμπεράσματα . . . . .	63
7.2	Μελλοντικές επεκτάσεις . . . . .	64
	<b>Βιβλιογραφία</b>	<b>65</b>
	<b>Συντομογραφίες</b>	<b>67</b>
	<b>Ορολογία - Γλωσσάρι</b>	<b>69</b>

# Κατάλογος σχημάτων

2.1	Λήψη αποφάσεων [2]	4
2.2	Cultural Analysis[1]	6
2.3	Κύβος OLAP [13].	7
2.4	Ιεραρχίες Διαστάσεων[9].	8
2.5	Εφαρμογή Αναλυτικής καθόδου και Συναθροιστικής ανόδου[14].	8
2.6	Κόψιμο σε φέτες (slice)[13].	9
2.7	Τεμαχισμός σε κύβους (dice).[13]	9
2.8	Λειτουργία περιστροφής (pivot).[13]	10
2.9	Σχήμα αστέρα και σχήμα χιονονιδάφας.	10
3.1	Παράδειγμα ενός data Cube (Πηγή: [12])	14
3.2	Μεταδεδομένα Λογικού Μοντέλου (Πηγή: [12])	15
3.3	Παράδειγμα ενός κελιού του κύβου. (Πηγή: [12])	19
3.4	Παραδείγματα τύπων τομών. (Πηγή: [11])	20
3.5	Παραδείγματα συναθροίσεων και αναλυτικών καθόδων. (Πηγή: [12])	21
3.6	Σχήματα αστέρα, χιονονιφάδας και μη κανονικοποιημένου πίνακα. (Πηγή: [11])	25
3.7	Αντιστοίχιση Λογικού με Φυσικό Μοντέλο. (Πηγή: [11])	26
4.1	Οι διαστάσεις του Λογικού Μοντέλου	31
5.1	Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (grouped).	43
5.2	Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stacked).	44
5.3	Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stream stacked).	44
5.4	Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stream expanded).	45
5.5	Απεικόνιση μόνο του συνολικού αριθμού έργων κατηγορίας “Prints & Illustrated Books”(areas stacked).	45
5.6	Απεικόνιση των facts data για γυναίκες καλλιτέχνες με γερμανική εθνικότητα.	46
5.7	Απεικόνιση γυναικών καλλιτεχνών με γερμανική εθνικότητα ανά ηλικιακό γκρουπ.	47
6.1	Εφαρμογής της συνάρτησης Aggregate σε όλο τον κύβο.	49
6.2	Ποσοστό έργων τέχνης ανά κατηγορία.	50
6.3	Ποσοστό καλλιτεχνών ανά κατηγορία.	51

6.4	Συνολικός αριθμός έργων ανά τρίμηνο. . . . .	52
6.5	Συνολικός αριθμός έργων ανά κατηγορία κατά τρίμηνο. . . . .	53
6.6	Ποσοστό έργων τέχνης ανά εθνικότητα. . . . .	54
6.7	Ποσοστό καλλιτεχνών ανά εθνικότητα. . . . .	54
6.8	Καλλιτέχνες κατά εθνοκότητα και ανά κατηγορία. . . . .	55
6.9	Αριθμός Αμερικανών καλλιτεχνών ανά κατηγορία. . . . .	56
6.10	Αριθμός Ιταλών καλλιτεχνών ανά κατηγορία. . . . .	56
6.11	Αριθμός έργων τέχνης ανά ηλικιακό γκρούπ. . . . .	57
6.12	Αριθμός καλλιτεχνών ανά ηλικιακό γκρούπ. . . . .	58
6.13	Αριθμός καλλιτεχνών ανά κατηγορία, κατά ηλικιακό γκρούπ. . . . .	59
6.14	Ποσοστό έργων τέχνης ανά φύλο. . . . .	59
6.15	Ποσοστό καλλιτεχνών ανά φύλο. . . . .	60
6.16	Ποσοστό καλλιτεχνών κατά φύλο και ανά κατηγορία. . . . .	61
6.17	Αριθμός Αμερικανών καλλιτεχνών κατά φύλο και ανά ηλικιακά groups. . . . .	61

# Κατάλογος πινάκων

4.1 Γνωρίσματα του σετ δεδομένων . . . . .	28
--	----



# Κεφάλαιο 1

## Εισαγωγή

Τα τελευταία χρόνια, οι τεχνολογικές εξελίξεις στη συλλογή και αποθήκευση δεδομένων δίνουν την δυνατότητα στους οργανισμούς να αποθηκεύουν τεράστιες ποσότητες δεδομένων. Με τη σειρά τους αυτές οι εξελίξεις, έχουν δημιουργήσει την ανάγκη για ανάλυση αυτού του μεγάλου όγκου δεδομένων καθώς και για την εξαγωγή πληροφορίας που μελλοντικά θα έχει αξία για την επιχείρηση. Οι οργανισμοί πλέον δίνουν όλο και μεγαλύτερη έμφαση, στην δημιουργία συγκεντρωτικών αποθηκών δεδομένων, που περιέχουν ιστορικές και συνοπτικές πληροφορίες, και κατ' επέκταση σε ισχυρά εργαλεία ανάλυσης. Η ανάλυση των δεδομένων, με τη βοήθεια των κατάλληλων εργαλείων για την επεξεργασία και οπτικοποίησή τους, μπορεί να οδηγήσει σε πολύτιμα συμπεράσματα και πληροφορίες, τα οποία με τη σειρά τους κρίνονται υψίστης σημασίας για τη διαδικασία λήψης αποφάσεων.

### 1.1 Αντικείμενο της διπλωματικής

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η άμεση αναλυτική επεξεργασία (OLAP) δεδομένων έργων τέχνης με σκοπό την εξαγωγή πληροφορίας, η οποία μπορεί να αξιοποιηθεί: α) από στελέχη της επιχείρησης για την λήψη αποφάσεων, β) από Διεθνείς Οργανισμούς ή Υπουργεία Πολιτισμού για την δημιουργία πολιτιστικών στατιστικών, γ) από απλούς λάτρεις της μοντέρνας τέχνης.

Προκειμένου να γίνει αυτό, θα χρησιμοποιηθεί το framework Cubes, το οποίο αποτελεί μέρος των εργαλείων της προγραμματιστικής γλώσσας Python για την επεξεργασία και ανάλυση δεδομένων. Το Cubes διαθέτει ενσωματωμένο έναν HTTP Server, ώστε να μπορεί να χρησιμοποιηθεί από εφαρμογές για τις βασικές λειτουργίες της αναλυτικής επεξεργασίας. Η εργασία παρέχει σε έναν αρχάριο την δυνατότητα να χρησιμοποιήσει το εργαλείο Cubes, εφόσον παρουσιάζονται αναλυτικά η δομή του και ο τρόπος λειτουργίας του. Το σετ δεδομένων έργων τέχνης του μουσείου MoMa (Museum of Modern Art, Νέα Υόρκη) που υπήρχε διαθέσιμο προς χρήση, επεξεργάστηκε κατάλληλα ώστε να συμπληρωθούν όλες οι ελλείψεις ή λανθασμένες τιμές. Ακόμα δημιουργήθηκαν κάποιες πρόσθετες στήλες με βάση τις πληροφορίες που μας ενδιέφερε να αναλύσουμε.

Η τελική βάση δεδομένων που προκύπτει από την επεξεργασία περιλαμβάνει πληροφορίες για τα έργα τέχνης του μουσείου αλλά και για τους καλλιτέχνες. Το σετ δεδομένων μοντελοποιείται ως

υπερκύβος, ώστε να είναι σε κατάλληλη μορφή για την χρήση του από το Cubes. Για την υλοποίηση των ερωτημάτων (queries) από τον χρήστη, αλλά και την οπτικοποίηση των αποτελεσμάτων σε γραφήματα, επιλέγεται το εργαλείο CubesViewer. Το συγκεκριμένο εργαλείο οπτικοποίησης, δίνει την δυνατότητα πολλών διαφορετικών τρόπων αναπαραστάσεως ενός γραφήματος. Η πληροφορία που εξάγεται από τα δεδομένα με την βοήθεια του CubesViewer αναλύεται διεξοδικά προς το τέλος της εργασίας.

### 1.1.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Παρουσιάζεται η σύνδεση της επιστήμης δεδομένων με τη διαδικασία λήψης αποφάσεων.
2. Επεξηγείται η χρησιμότητα της εξαγωγής πληροφορίας στον τομέα Cultural Analysis.
3. Αναλύονται βασικές θεωρητικές έννοιες της άμεσης αναλυτικής επεξεργασίας και των αποθηκών δεδομένων.
4. Περιγράφεται αναλυτικά το εργαλείο Cubes και ο τρόπος χρήσης του.
5. Χρησιμοποιήθηκε και επεξεργάστηκε πραγματικό, μεγάλου όγκου σετ δεδομένων από σημαντικό οργανισμό.
6. Παρουσιάζεται ολόκληρη η διαδικασία προετοιμασίας του σετ δεδομένων για την χρήση του από το εργαλείο Cubes.
7. Γίνεται παρουσίαση του εργαλείου CubesViewer, που χρησιμοποιείται για την οπτικοποίηση των αποτελεσμάτων.
8. Υλοποιείται η οπτικοποίηση των αποτελεσμάτων της αναλυτικής επεξεργασίας.
9. Παρουσιάζονται τα βασικά συμπεράσματα της ανάλυσης.

## 1.2 Οργάνωση του τόμου

Στο Κεφάλαιο 2 παρουσιάζονται αναλυτικά οι βασικές θεωρητικές έννοιες που θα χρησιμοποιηθούν στην διπλωματική αυτή εργασία. Η περιγραφή του εργαλείου Cubes, οι οδηγίες χρήσης και η εγκατάστασή του περιγράφονται εκτενώς στο Κεφάλαιο 3. Το Κεφάλαιο 4 περιέχει την επεξεργασία της βάσης δεδομένων και τη δημιουργία του λογικού μοντέλου, ώστε να χρησιμοποιηθεί από το Cubes για την άμεση αναλυτική επεξεργασία των δεδομένων. Το επόμενο Κεφάλαιο περιέχει την περιγραφή του εργαλείου CubesViewer που χρησιμοποιείται για την οπτικοποίηση των αποτελεσμάτων καθώς και μερικά παραδείγματα χρήσης του. Η εξαγωγή πληροφορίας από τα δεδομένα αλλά και η μελέτη των συμπερασμάτων περιλαμβάνονται στο Κεφάλαιο 6.



## Κεφάλαιο 2

# Θεωρητικό Υπόβαθρο

### 2.1 Εισαγωγή

Αρχικά, στο κεφάλαιο αυτό γίνεται κατανοητή η σύνδεση της επιστήμης δεδομένων με την υποστήριξη λήψης αποφάσεων και μετέπειτα περιγράφεται η χρησιμότητα της εξαγωγής πληροφορίας, από την συγκεκριμένη εργασία, στην ανάλυση πολιτιστικών στατιστικών (culture analysis). Ακόμη επεξηγούνται αναλυτικά θεωρητικές έννοιες που χρειάζονται να κατανοηθούν από τον αναγνώστη, ώστε να προχωρήσει στη συνέχεια και στο τεχνικό μέρος της εργασίας.

### 2.2 Ανάκτηση Πληροφορίας

Η ανάκτηση πληροφορίας (Information Retrieval) είναι η επιστημονική περιοχή που μελετά τα προβλήματα που σχετίζονται με την αναπαράσταση, την οργάνωση και την επεξεργασία στοιχείων πληροφορίας, με στόχο την αποτελεσματική αναζήτηση πληροφοριών ή κάθε τύπου δεδομένων [3]. Η ανάκτηση πληροφοριών στηρίζεται στη θεωρία των βάσεων δεδομένων, σε κατάλληλα υπολογιστικά συστήματα και σε μαθηματικές μεθόδους, ενώ εφαρμόζεται στην επιστήμη πληροφορίας και στον Παγκόσμιο Ιστό (μηχανές αναζήτησης).

Σύμφωνα με τον όγκο και τη φύση των δεδομένων που χρησιμοποιούνται στη συγκεκριμένη εργασία επιλέχθηκαν ως εργαλεία ανάλυσης και επεξεργασίας, τα συστήματα OLAP (On Line Analytical Processing). Όπως αναφέρεται και στο τεχνικό έγγραφο [6], τα παραπάνω συστήματα προσφέρουν μια δυναμική και πολυδιάστατη ανάλυση δεδομένων. Η πληροφορία που μπορεί να ανακτηθεί μέσω της πολυδιάστατης ανάλυσης που προσφέρουν τα εργαλεία OLAP καθίσταται ιδιαίτερα χρήσιμη στον τομέα λήψης αποφάσεων ενός οργανισμού. Περισσότερα για τα συστήματα αναλυτικής επεξεργασίας (OLAP) αλλά και για την υποστήριξη λήψης αποφάσεων ακολουθούν στην συνέχεια του παρόντος κεφαλαίου.

## 2.3 Μέθοδοι Αναλυτικής Επεξεργασίας για Υποστήριξη Λήψης Αποφάσεων

Οι σύγχρονοι οργανισμοί κατακλύζονται από ένα τεράστιο όγκο δεδομένων, τα οποία προέρχονται από εσωτερικές και εξωτερικές πηγές. Τα δεδομένα αυτά είναι πολύ σημαντικά για την εξαγωγή πληροφορίας και την λήψη αποφάσεων, στο πλαίσιο της κάθε επιχείρησης. Γι αυτό το λόγο οι σύγχρονοι οργανισμοί του 21ου αιώνα, έχουν ενσωματώσει τεχνολογίες επεξεργασίας και ανάλυσης των ιστορικών τους δεδομένων. Τα ιστορικά αυτά δεδομένα αναλύονται και επεργάζονται με κατάλληλο τρόπο, ώστε να δημιουργούν χρήσιμες τάσεις και συνόψεις δεδομένων για την υποστήριξη αποφάσεων (decision support) σε όλα τα επίπεδα.

Πολλές επιχειρήσεις πλέον δημιουργούν αποθήκες δεδομένων (data warehouses) που περιέχουν ιστορικά δεδομένα και συνοπτικές πληροφορίες. Αυτή η οργανωμένη αποθήκευση των δεδομένων οδήγησε με τη σειρά της στην αύξηση της χρήσης εργαλείων ανάλυσης [9]. Σε συνδυασμό με τις αποθήκες δεδομένων τα εργαλεία αναλυτικής επεξεργασίας (OLAP), είναι σχεδιασμένα για την υποστήριξη λήψης αποφάσεων. Τα συστήματα OLAP, μετατρέπουν τα ακατέργαστα δεδομένα σε αξιοποιήσιμη πληροφορία στρατηγικής φύσεως και ιδιαίτερα χρήσιμη για τα επιχειρηματικά στελέχη. Αυτό επιτυγχάνεται λόγω της δυνατότητας τους να επεξεργάζονται δεδομένα μεγάλου όγκου, με γρήγορη ταχύτητα και από πολλές διαστάσεις.



Σχήμα 2.1: Λήψη αποφάσεων [2]

Στην παρούσα εργασία χρησιμοποιήθηκαν εργαλεία OLAP για την επεξεργασία και την ανάλυση των δεδομένων. Στην συγγενική εργασία των Ραβασόπουλου, Παπαϊωάννου και Βουτσινά [6] χρησιμοποιείται η αναλυτική επεξεργασία (OLAP) για την υποστήριξη αποφάσεων των υπευθύνων περιβαλλοντικής εκπαίδευσης των Διευθύνσεων Εκπαίδευσης. Παρουσιάζεται η λειτουργία των συστημάτων OLAP και στη συνέχεια, με την βοήθεια της οπτικοποίησης των δεδομένων, αναλύονται κάποιες περιπτώσεις για την διεξαγωγή συμπερασμάτων. Τα αποτελέσματα αυτής της πολυδιάστατης ανάλυσης μπορούν να φανούν χρήσιμα στους υπευθύνους των Διευθύνσεων Εκπαίδευσης, ώστε να πάρουν αποφάσεις που αφορούν στην υλοποίηση προγραμμάτων για την ανά-

πτυξη της περιβαλλοντικής εκπαίδευσης στις σχολικές μονάδες.

Το σετ δεδομένων που χρησιμοποιήθηκε στη συγκεκριμένη εργασία, είναι ενός από τα μεγαλύτερα μουσεία μοντέρνας τέχνης του κόσμου, το MoMa (Νέα Υόρκη). Τα μουσεία είναι οργανισμοί, και σαν οργανισμοί χρειάζονται πληροφορίες που θα υποστηρίξουν την λήψη αποφάσεων. Όπως κάθε επιχείρηση με οικονομικές ανάγκες, τα μουσεία συλλέγουν και αναλύουν τα δεδομένα τους με σκοπό τη βελτίωση της εμπειρίας των χρηστών και την αύξηση των επισκεπτών και των εσόδων. Ακόμη χρησιμοποιούν αναλυτικά στοιχεία για να κατανοήσουν πώς να εξυπηρετήσουν καλύτερα τον κύκλο των επισκεπτών τους.

Η πληροφορία που θα εξαχθεί από το συγκεκριμένο σετ δεδομένων, μπορεί να φανεί ιδιαίτερα χρήσιμη στους υπευθύνους για την λήψη αποφάσεων. Σίγουρα φαίνεται ο τρόπος με τον οποίο κάθε αναλυτής θα επιλέξει να επεξεργαστεί τα δεδομένα που τον ενδιαφέρουν κάθε φορά, ώστε να εξάγει την πληροφορία που του είναι χρήσιμη.

Εκτός των άλλων, οι πληροφορίες που προκύπτουν από τα δεδομένα ενός μουσείου δεν είναι χρήσιμες μόνο στον τομέα λήψης αποφάσεων. Μπορούν να φανούν εξίσου χρήσιμες και σε στατιστικά χωρών σχετικά με τον πολιτισμό και την τέχνη.

## 2.4 Εξαγωγή Πληροφορίας και Cultural Analytics

Από μια διαφορετική σκοπιά, η πληροφορία που εξάγεται από την ανάλυση δεδομένων των έργων και των καλλιτεχνών ενός μουσείου, είναι ιδιαίτερα χρήσιμη στην δημιουργία πολιτισμικών στατιστικών.

Η πολιτισμική ανάλυση (cultural analysis) βασίζεται στη χρήση ερευνητικών μεθόδων για τις τέχνες, τις ανθρωπιστικές και κοινωνικές επιστήμες (όπως εθνογραφία, ανθρωπολογία κ.α) και τη συλλογή δεδομένων σχετικά με πολιτιστικά φαινόμενα. Η ερμηνεία των φαινομένων αυτών και η ανάλυση των δεδομένων, οδηγεί στην απόκτηση νέων γνώσεων και στην κατανόηση των πολιτιστικών διαδικασιών.

Η Ευρωπαϊκή Ένωση κυκλοφορεί ειδικά rocketbooks με την συλλογή και την ανάλυση πολιτισμικών στατιστικών, με σκοπό την ενίσχυση των στατιστικών γνώσεων σχετικά με τον πολιτισμό και τη σύγκριση πολιτιστικών δεδομένων σε όλη την Ευρώπη. Όπως αναφέρεται στο rocketbook [5], τα πολιτισμικά στατιστικά καλύπτουν πολλές πτυχές της οικονομικής και κοινωνικής ζωής. Γι' αυτόν ακριβώς τον λόγο, μπορούν να χρησιμεύσουν στην υποστήριξη του σημαντικά αυξανόμενου ενδιαφέροντος των πολιτικών ιθυνόντων για τον πολιτισμό και τον ρόλο του στην κοινωνία, στην οικονομία, καθώς και στη συνοχή μιας χώρας ή μιας ένωσης κρατών.

Επομένως, αφού τα μουσεία αποτελούν την καρδιά της πολιτιστικής συντήρησης, οι πληροφορίες που μπορούν να ανακτηθούν από τα δεδομένα των έργων τους είναι ιδιαίτερα χρήσιμες στην δημιουργία πολιτισμικών στατιστικών από Υπουργεία Πολιτισμού ή Διεθνείς οργανισμούς (π.χ. UNESCO).



Σχήμα 2.2: Cultural Analysis[1]

## 2.5 Αποθήκες Δεδομένων (Data Warehouses)

Η αποθήκη δεδομένων (Data Warehouse- DW) αποτελεί μία βάση δεδομένων με σκοπό την στήριξη αποφάσεων και διατηρείται ξεχωριστά από τα λειτουργικά δεδομένα ενός οργανισμού. Τα δεδομένα της περιέχουν συναθροιστικές πληροφορίες που είναι χρήσιμες για ανάλυση και εξαγωγή συμπερασμάτων, καθώς και για λήψη αποφάσεων. Σύμφωνα με τον ορισμό του Inmon [15], μια αποθήκη δεδομένων είναι μια θεματικά προσανατολισμένη (object-oriented), ολοκληρωμένη (integrated), με χρονική διαφοροποίηση (time-variant), μη ευμετάβλητη (non volatile) συλλογή δεδομένων, η οποία χρησιμοποιείται για την υποστήριξη των διαδικασιών λήψης αποφάσεων. Ουσιαστικά οι αποθήκες δεδομένων αποτελούνται από ιστορικές πληροφορίες που αναφέρονται σε βάθος χρόνου, αποτελώντας στιγμιότυπα του οργανισμού, με σκοπό την άμεση απάντηση ερωτημάτων και την ανάλυση.

Όπως προαναφέρθηκε, οι περισσότεροι οργανισμοί πλέον έχουν ανάγκη από τα παραπάνω πληροφοριακά συστήματα ώστε να συγκεντρώνουν, να αποθηκεύουν και να αναλύουν τα απαραίτητα δεδομένα για τη λήψη αποφάσεων. Οι αποθήκες δεδομένων σχετίζονται άμεσα με τα συστήματα Αναλυτικής Επεξεργασίας Δεδομένων (On Line Analytical Processing (OLAP)), τα οποία παρέχουν ευέλικτη πρόσβαση και ανάλυση μεγάλου όγκου δεδομένων.[16]

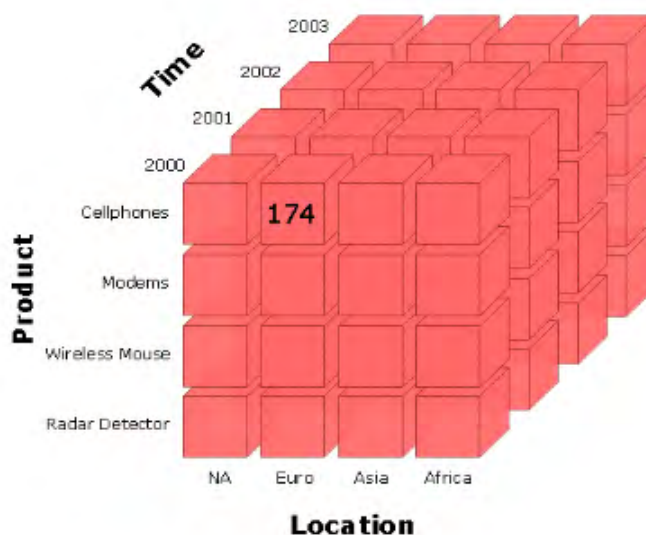
## 2.6 Άμεση Αναλυτική Επεξεργασία - OLAP

Η άμεση αναλυτική επεξεργασία είναι η τεχνολογία πίσω από πολλές εφαρμογές Επιχειρηματικής Ευφυΐας (Business Intelligence (BI)). Ο όρος OLAP χρησιμοποιείται για να περιγράψει την ανάλυση πολύπλοκων δεδομένων από αποθήκες δεδομένων. Τα συστήματα OLAP χρησιμοποιούνται από αναλυτές και υψηλόβαθμα στελέχη επιχειρήσεων, για την διεξαγωγή αναλύσεων και μετέπειτα, την υποστήριξη λήψης αποφάσεων [4]. Εξασφαλίζουν την εύκολη και γρήγορη

πρόσβαση σε μεγάλους όγκους δεδομένων και επιτρέπουν την πολυδιάστατη επεξεργασία τους.

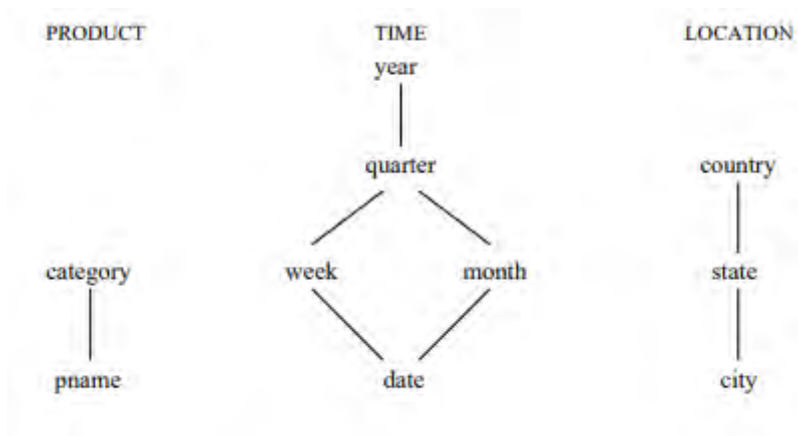
Τα εργαλεία OLAP δίνουν την δυνατότητα στον χρήστη να υποβάλλει σύνθετα ερωτήματα που απαιτούν συναθροίσεις, επικεντρώνοντας σε ζητήματα που τον ενδιαφέρουν και αυξομειώνοντας τον βαθμό γενίκευσης. Για την επίτευξη των παραπάνω, στο επίκεντρο κάθε συστήματος OLAP βρίσκεται η έννοια του κύβου ή υπερκύβου. Ο κύβος OLAP ή αλλιώς OLAP Cube, είναι μία πολυδιάστατη όψη αθροιστικών δεδομένων, που συνήθως βρίσκονται αποθηκευμένα σε μια αποθήκη δεδομένων, και αντιστοιχεί στα δεδομένα που θέλει ο χρήστης να αναλύσει. Κάθε κύβος συγκροτείται από μία συλλογή αριθμητικών μετρήσεων, των μέτρων (measures), και κάθε μέτρο εξαρτάται από ένα σύνολο διαστάσεων (dimensions) [6]. Ουσιαστικά μία διάσταση μοντελοποιεί τους διαφορετικούς εκείνους τρόπους, με τους οποίους τα δεδομένα μπορούν να συναθροιστούν σε σχέση με μια συγκεκριμένη παράμετρο του περιεχομένου τους (πχ.ημερομηνία ή τοποθεσία).

Αν υποθέσουμε ότι χρησιμοποιούμε ένα παράδειγμα βασισμένο σε δεδομένα πωλήσεων, το measure στο παράδειγμα αυτό είναι το sales (πωλήσεις). Οι διαστάσεις όπως φαίνεται και στο Σχήμα 2.3 είναι οι Product (προϊόν), Time (χρόνος) και Location (τοποθεσία).



Σχήμα 2.3: Κύβος OLAP [13].

Κάθε διάσταση μπορεί να εμπεριέχει μία ιεραρχία. Έτσι η διάσταση “Location” ενός κύβου μπορεί να εμπεριέχει την ιεραρχία “country”, “state”, “city”, όπως φαίνεται στο Σχήμα 2.4. Ο προσδιορισμός των διαστάσεων γίνεται με βάση το επίκεντρο της ανάλυσης. Για παράδειγμα, αν στο επίκεντρο ανάλυσης βρίσκεται μία πώληση, μπορούν να επιλεγθούν ως διαστάσεις η τοποθεσία και η ημερομηνία της. Με αυτόν τον τρόπο κάθε συνδυασμός των τιμών των διαστάσεων προσδιορίζει ένα κελί του κύβου. Τα περιεχόμενα κάθε κελιού αντικατοπτρίζουν την τιμή του χαρακτηριστικού που επιλέγουμε προς ανάλυση.[8]

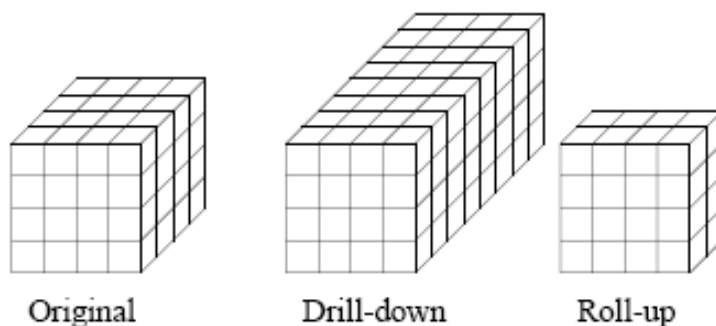


Σχήμα 2.4: Ιεραρχίες Διαστάσεων[9].

Κύριο χαρακτηριστικό της πολυδιάστατης ανάλυσης των δεδομένων είναι η συνάθροιση αυτών (aggregation) με πολλούς διαφορετικούς τρόπους. Στο παράδειγμα με τις πωλήσεις, θα μπορούσε να ζητηθεί ως συνάθροιση το συνολικό εισόδημα των πωλήσεων για ένα συγκεκριμένο προϊόν και για ένα συγκεκριμένο έτος[8].

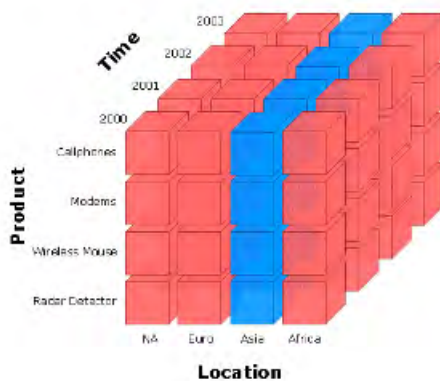
Οι βασικές πράξεις των συστημάτων OLAP είναι οι εξής:

- **Συναθροιστική Ανόδος (Roll-up):** Παράγει έναν κύβο δεδομένων με μειωμένες διαστάσεις ή με διαστάσεις που έχει επιλεγθεί το ανώτερο επίπεδο ιεραρχίας τους (Συσσώρευση).
- **Αναλυτική Κάθοδος (Drill-down):** Παράγει έναν κύβο δεδομένων με αυξημένες διαστάσεις ή με διαστάσεις που έχει επιλεγθεί το κατώτερο επίπεδο ιεραρχίας τους (Εμβάθυνση).



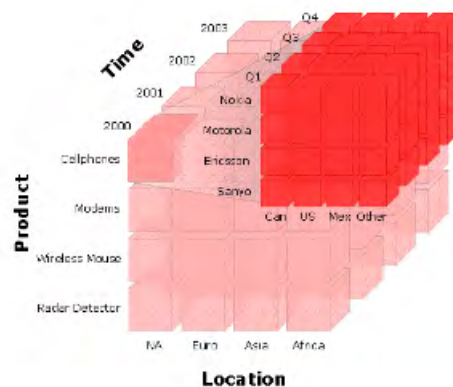
Σχήμα 2.5: Εφαρμογή Αναλυτικής καθόδου και Συναθροιστικής ανόδου[14].

- **Κόψιμο σε φέτες (Slicing):** Παράγει έναν κύβο δεδομένων όπου έχει επιλεγθεί μόνο μία διάσταση. Επιτρέπει την εξαγωγή πληροφορίας για μια συγκεκριμένη τιμή μιας διάστασης.



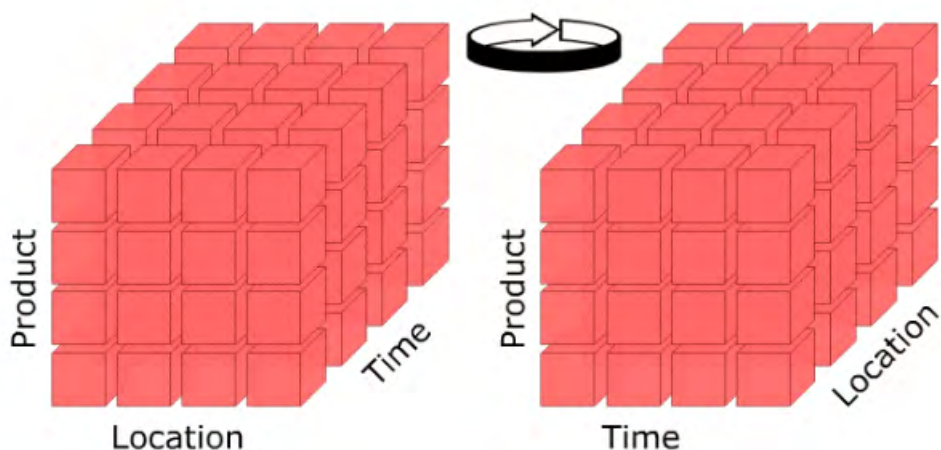
Σχήμα 2.6: Κόψιμο σε φέτες (slice)[13].

- **Τεμαχισμός σε κύβους (Dicing):** Παράγει έναν κύβο δεδομένων όπου έχει επιλεγθεί μία ή και περισσότερες διαστάσεις. Υποσύνολο από όλες τις διαστάσεις για συγκεκριμένες τιμές των διαστάσεων (zoom).



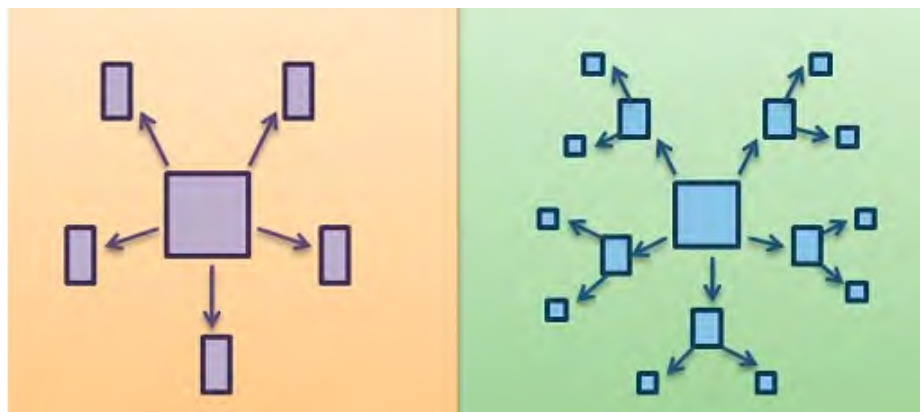
Σχήμα 2.7: Τεμαχισμός σε κύβους (dice).[13]

- **Περιστροφή (Pivoting):** Παράγει έναν κύβο δεδομένων, του οποίου η διάταξη των διαστάσεων έχει αλλαχθεί.



Σχήμα 2.8: Λειτουργία περιστροφής (pivot).[13]

Τα συστήματα OLAP που αποθηκεύουν όλη την πληροφορία με τη μορφή σχέσεων ονομάζονται σχεσιακά συστήματα OLAP (relational OLAP systems - ROLAP). Τα πιο βασικά σχήματα βάσεων δεδομένων που έχουν σχεδιαστεί για αναλυτική επεξεργασία είναι αυτά του αστερά και της χιονονιφάδας. Στο σχήμα αστερά (star schema), ο πίνακας γεγονότων, η πινακοειδής δηλαδή αναπαράσταση των δεδομένων, βρίσκεται στο κέντρο. Οι πίνακες διαστάσεων (περιέχουν τις διαστάσεις) είναι τοποθετημένοι γύρω του. Το σχήμα χιονονιφάδας διαφέρει ουσιαστικά απ' το σχήμα αστερά στο γεγονός ότι οι πίνακες διαστάσεων διασπώνται σε περισσότερους από έναν πίνακες.[9]



Σχήμα 2.9: Σχήμα αστερά και σχήμα χιονονιφάδας.

Τέλος, αξίζει να αναφερθεί πως υπάρχουν διαθέσιμα στην αγορά αρκετά εργαλεία, τα οποία παρέχουν τις λειτουργίες OLAP. Κάποια από τα πιο δημοφιλή είναι τα IBM Cognos, Micro Strategy, Palo OLAP Server, Apache Kylin, OBBIE κ.α. Για την συγκεκριμένη εργασία χρησιμοποιήθηκε το framework **Cubes**, το οποίο είναι μέρος του Data Brewery, μιας σειράς από εργαλεία της γλώσσας Python, με σκοπό την επεξεργασία και την ανάλυση δεδομένων. Επιλέχθηκε το συγκεκρι-



κριμένο framework καθώς είναι ένα ελεύθερο και ελαφρύ εργαλείο σε σχέση με τα πιο δημοφιλή, βαριά επαγγελματικής χρήσης, εργαλεία που προαναφέρθηκαν. Να προστεθεί ότι, ο όγκος των δεδομένων που χρησιμοποιείται στο πειραματικό στάδιο αυτής της εργασίας δεν είναι τόσο μεγάλος όσο αυτών που διαχειρίζεται μία επιχείρηση. Στο επόμενο κεφάλαιο περιγράφεται διεξοδικά το εργαλείο Cubes και αναλύεται ο τρόπος χρήσης του.



## Κεφάλαιο 3

# Cubes: Python Framework, OLAP Server

### 3.1 Εισαγωγή

Για την υλοποίηση της διπλωματικής αυτής εργασίας χρησιμοποιούμε το Cubes, ένα ελαφρύ, ανοιχτού κώδικα framework της Python και OLAP HTTP Server. Το εργαλείο αυτό ανήκει στο Data Brewery, ένα σύνολο από frameworks και εργαλεία της προγραμματιστικής γλώσσας Python τα οποία χρησιμοποιούνται για επεξεργασία και ανάλυση δεδομένων. Πρόκειται ουσιαστικά για ένα εύχρηστο εργαλείο για Data Warehousing το οποίο μπορεί να χρησιμοποιηθεί από εφαρμογές αναφοράς δεδομένων και modules για την μοντελοποίηση και εξέταση δεδομένων σε πολλαπλά επίπεδα. Το Cubes είναι σχεδιασμένο για να παρέχει σε έναν αναλυτή ή σε οποιονδήποτε χρήστη κάποιας εφαρμογής έναν γρήγορο και κατανοητό τρόπο αναφοράς πολυδιάστατων δεδομένων, βασίζόμενο στην έννοια των data Cubes. [12]

Η δομή του Cubes αποτελείται από τα εξής τέσσερα μέρη:

1. Λογικό μοντέλο (model)
2. Περιηγητή (browser)
3. Σύστημα υποστήριξης (backend)
4. Εξυπηρετητή ( Slicer Server)

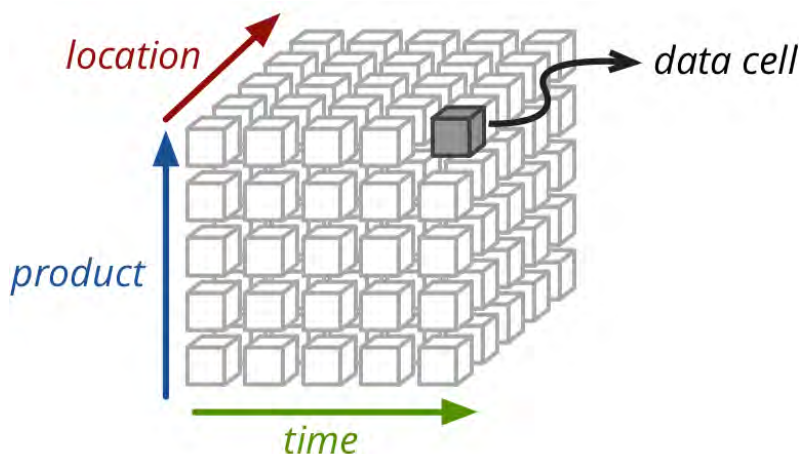
Στη συνέχεια θα αναλύσουμε διεξοδικά τα τέσσερα μέρη που προαναφέρθηκαν ώστε να γίνει κατανοητή η δομή καθώς και η διαδικασία με την οποία λειτουργεί το Cubes.

### 3.2 Λογικό Μοντέλο (Model)

Το Λογικό μοντέλο περιγράφει τα δεδομένα από την οπτική του αναλυτή ή της επιχείρησης και δεν εξαρτάται από την φυσική τους υλοποίηση. Λόγω της ανεξαρτησίας αυτής γίνεται ευκο-

λότερη η έμφαση στα δεδομένα καθαυτά, παρά στους τρόπους με τους οποίους τα δεδομένα θα μετατραπούν σε κατανοητή μορφή. Ουσιαστικά στο μοντέλο καθορίζουμε τον τρόπο με τον οποίο μας ενδιαφέρει να εξετάσουμε και να αναλύσουμε τα δεδομένα "κρύβοντας" τη φυσική δομή της βάσης δεδομένων και τον τρόπο με τον οποίο η εφαρμογή τα χρησιμοποιεί. [12]

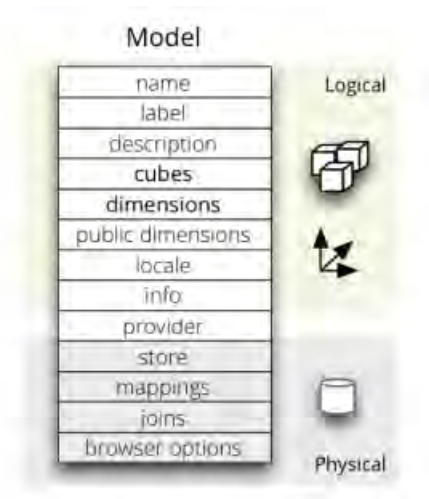
Για να προχωρήσουμε στη διαδικασία δημιουργίας του μοντέλου θα πρέπει να αναφέρουμε πρώτα τις βασικές έννοιες ενός OLAP Cube. Τα δεδομένα μοντελοποιούνται σαν κύβος με πολλαπλές διαστάσεις, όπως βλέπουμε στο Σχήμα 3.1.



Σχήμα 3.1: Παράδειγμα ενός data Cube (Πηγή: [12])

- Fact: Η πιο μικρή λεπτομερής μονάδα του κύβου ονομάζεται fact και μπορεί να είναι π.χ. για ένα συμβόλαιο ή ένα έργο. Αποτελεί το αντικείμενο που θέλουμε να μελετήσουμε.
- Cube: Το σύνολο των δεδομένων που προσδιορίζουν μία συγκεκριμένη ομάδα από facts.
- Dimension: Οι διαστάσεις παρέχουν το εννοιολογικό πλαίσιο στα facts και χρησιμοποιούνται στο φιλτράρισμα των ερωτήσεων. Μία διάσταση μπορεί να έχει περισσότερα από ένα επίπεδα ιεραρχίας (hierarchies).
- Measure: Ένα χαρακτηριστικό το οποίο μετράμε για ένα fact, όπως μπορεί να είναι για παράδειγμα το ποσό μίας αγοράς.

Το λογικό μοντέλο περιγράφεται από ένα λεξικό μεταδεδομένων όπως μπορούμε να δούμε και στο Σχήμα 3.2. Τα πρώτα οχτώ χαρακτηριστικά αντιστοιχούν στο λογικό μέρος του μοντέλου ενώ τα τέσσερα τελευταία στο φυσικό μέρος. Ακόμη υπάρχουν και άλλα χαρακτηριστικά τα οποία σχετίζονται με τις υπόλοιπες πληροφορίες της περιγραφής του μοντέλου.



Σχήμα 3.2: Μεταδεδομένα Λογικού Μοντέλου (Πηγή: [12])

Λογικό μέρος περιγραφής μοντέλου:

- name: όνομα μοντέλου
- label: όνομα που εμφανίζεται στον χρήστη
- description: περιγραφή μοντέλου
- cubes: λίστα μεταδεδομένων του κύβου
- dimensions: λίστα μεταδεδομένων των διαστάσεων του κύβου
- locale: έχει χρήση σε localizable μοντέλα

Φυσικό μέρος περιγραφής μοντελού:

- store: όνομα του datastore όπου είναι αποθηκευμένοι οι κύβοι του μοντέλου
- mappings: backend- λεξικό αντιστοιχίας λογικών με φυσικά δεδομένα· κληρονομείται απ' όλους τους κύβους του μοντέλου
- joins: backend- καταγράφει όλα τα joins· κληρονομείται απ' όλους τους κύβους του μοντέλου
- browser options: οι επιλογές του περιηγητή

Θα χρησιμοποιήσουμε ως παράδειγμα δεδομένα από πωλήσεις, ώστε στη συνέχεια να επεξηγήσουμε ευκολότερα κάποιες ακόμα έννοιες. Για την περιγραφή του λογικού μοντέλου θα πρέπει να δημιουργηθεί ένα αρχείο JSON, όπως αυτό που φαίνεται στο παρακάτω παράδειγμα.

```
{
    "name": "sales",
    "label": "Sales",
    "cubes": [...],
    "dimensions": [...],
    "mappings": [...]
}
```

Πιο συγκεκριμένα στο παρακάτω κομμάτι κώδικα γίνεται ο ορισμός των διαστάσεων και των αντίστοιχων ιεραρχιών του παραδείγματος των πωλήσεων που χρησιμοποιούμε. Υπάρχουν τρεις διαστάσεις: η “product”, η “price” και η “date”, η οποία περιλαμβάνει τρία επίπεδα (“year”, “month”, “day”). Η “default” ιεραρχία για την διάσταση “date” περιλαμβάνει τα επίπεδα με σειρά “year”, “month”. Ενώ στην ιεραρχία “ymd” τα επίπεδα παρουσιάζονται με σειρά “year”, “month”, “day”.

```
"hierarchies": [
    {
        "name": "default",
        "levels": ["year", "month"]
    },
    {
        "name": "ymd",
        "levels": ["year", "month", "day"]
    }
]
```

```
"dimensions": [  
  {  
    "name" : "product",  
    "label": "Product Type"  
  },  
  {  
    "name": "price",  
    "label": "Price"  
  },  
  {  
    "name": "date",  
    "levels": [  
      {  
        "name": "year",  
        "label": "Year"  
      },  
      {  
        "name": "month",  
        "label": "Month"  
      },  
      {  
        "name": "day",  
        "label": "Day"  
      }  
    ]  
  }  
]
```

Τα μεταδεδομένα που χρησιμοποιούνται για την περιγραφή του κύβου αναλύονται στον παρακάτω πίνακα.

name	Όνομα κύβου
label	Όνομα που εμφανίζεται στον χρήστη
description	Περιγραφή κύβου
Info	Ειδικές πληροφορίες
dimensions	Διαστάσεις κύβου
measures	Λίστα με τα μέτρα του κύβου
aggregates	Λίστα των συναθροιστικών μέτρων
details	Λίστα λεπτομερειών, χαρακτηριστικά που δε σχετίζονται με τις συναθροίσεις
joins	Καταγραφή συνδέσεων
mappings	Αντιστοίχιση με φυσικά δεδομένα
browser_options	Επιλογές περιηγητή
store	Όνομα χώρου αποθήκευσης κύβου

Για το παράδειγμα με τις πωλήσεις που χρησιμοποιείται ορίζεται ένας κύβος με όνομα “sales” και τρεις διαστάσεις (“product”, “price”, “date”), όπως αναλύσαμε και παραπάνω. Υπάρχει ένα measure, το “price”, το οποίο αντιστοιχεί στην τιμή κάθε πώλησης. Επίσης έχει οριστεί μία συναθροιση (aggregate) η οποία ονομάζεται “total price” και χρησιμοποιεί την συνάρτηση (function) “sum” πάνω στο measure “price” ώστε να αθροίσει τις τιμές όλων των πωλήσεων του κύβου. Τέλος χρησιμοποιούνται mappings ώστε να υπάρχει απόλυτη αντιστοιχία των ονομάτων των στηλών που υπάρχουν στη βάση δεδομένων, υποθέτοντας ότι είναι “year”, “month” και “day”.

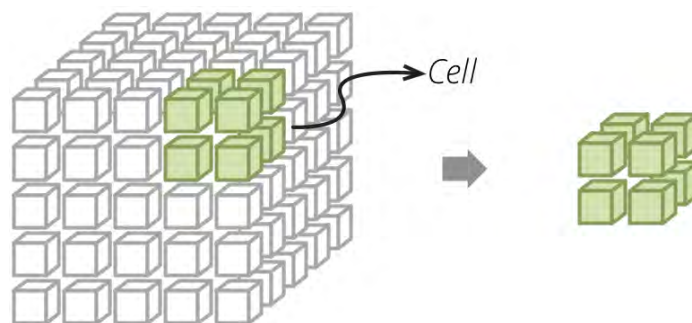


```
"cubes": [  
  {  
    "name": "sales",  
    "dimensions": ["product", "price", "date"],  
    "measures": ["price"],  
    "aggregates": [  
      {  
        "name": "total_price",  
        "function": "sum",  
        "measure": "price"  
      }  
    ],  
    "mappings": {  
      "date.year": "year",  
      "date.month": "month",  
      "date.day": "day"  
    }  
  }  
]
```

### 3.3 Περιηγητής (Browser)

Ο περιηγητής (Aggregation Browser) είναι αυτός, που εφαρμόζοντας το μοντέλο πάνω στα δεδομένα υπολογίζει τις συναθροίσεις (Aggregations) μέσα στον κύβο, ενώ επίσης δίνει την δυνατότητα, να «περιηγηθούμε» μέσα στα δεδομένα και να εξάγουμε τις πληροφορίες που χρειαζόμαστε.

Η βασική έννοια του περιηγητή είναι το κελί (Cell). Το κελί αποτελεί ένα κομμάτι του κύβου το οποίο καθορίζει το μέρος εκείνο του κύβου που ενδιαφέρει τον χρήστη. [12]



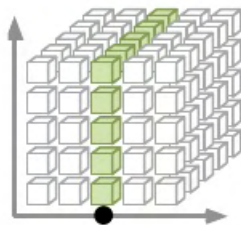
Σχήμα 3.3: Παράδειγμα ενός κελιού του κύβου. (Πηγή: [12])

Το εκάστοτε κελί που μας ενδιαφέρει καθορίζεται από την εφαρμογή των κατάλληλων τομών (Cuts) πάνω στον αρχικό κύβο. Αυτή η διαδικασία ονομάζεται Slicing and Dicing. Υπάρχουν τρεις διαφορετικοί τύποι τομών τους οποίους υποστηρίζει το Cubes:

1. Point Cut: Καθορίζει ένα συγκεκριμένο σημείο σε κάποια διάσταση.
2. Range Cut: Καθορίζει ένα εύρος σημείων σε κάποια διατεταγμένη διάσταση.
3. Set Cut: Καθορίζει ένα σύνολο συγκεκριμένων σημείων.

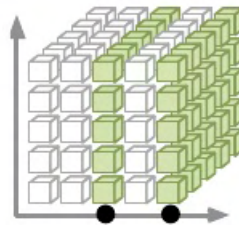


## cut types



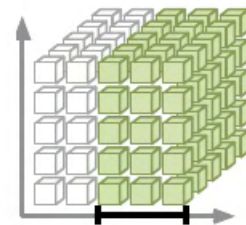
point

```
[2010]
```



set

```
[[2010, 10],  
 [2010, 12]]
```

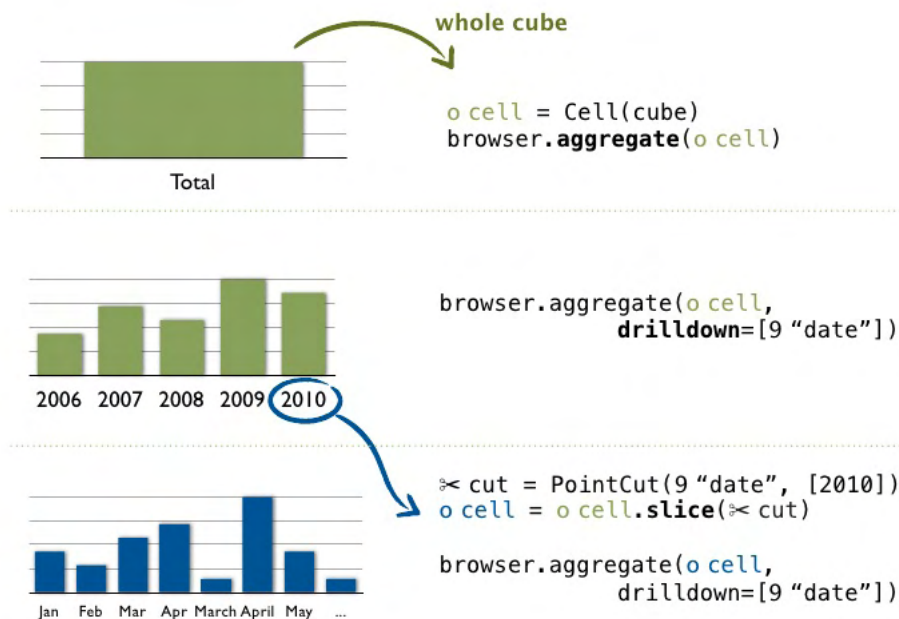


range

```
from=[2010, 10]  
to=[2010, 12]
```

Σχήμα 3.4: Παραδείγματα τύπων τομών. (Πηγή: [11])

Μια ακόμη πολύ σημαντική δυνατότητα που προσφέρει ο browser είναι η αναλυτική κάθοδος (drilldown) κατά την οποία ο αναλυτής μπορεί να λάβει περισσότερες λεπτομέρειες αφού το αποτέλεσμα ομαδοποιείται ως προς κάποια διάσταση. Λόγου χάριν, στο παράδειγμα των πωλήσεων μπορεί να εμφανίσει τα αποτελέσματα ανα είδος προϊόντος ή ανα ημερομηνία. Εφόσον η κάθοδος γίνει στην διάσταση “date” η οποία έχει πολλαπλά επίπεδα, θα εφαρμοστεί για το αμέσως επόμενο επίπεδο ιεραρχίας από αυτό στο οποίο έχει γίνει ίσως, μια βαθύτερη τομή.



Σχήμα 3.5: Παραδείγματα συναθροίσεων και αναλυτικών καθόδων. (Πηγή: [12])

Βασικό κομμάτι της υλοποίησης του browser αποτελεί ο ορισμός του χώρου εργασίας (Workspace), ο οποίος ουσιαστικά περιγράφει ποιο λογικό μοντέλο χρησιμοποιείται και επάνω σε ποια δεδομένα. Στο παρακάτω κομμάτι κώδικα σε Python φαίνεται η υλοποίηση του Workspace. Υποθέτουμε ότι χρησιμοποιείται SQLite.

```
# Create Workspace

ws= Workspace()

ws.register_default_store("sql", url="sqlite:///data.sqlite")

ws.import_model("sales.json")
```

### 3.4 Εξυπηρετητής (Slicer Server)

Το Cubes διαθέτει ενσωματωμένο έναν απλό HTTP OLAP Server, τον Slicer Server, ο οποίος καλύπτει το μεγαλύτερο μέρος των λειτουργιών που διαθέτει ο περιηγητής συναθροίσεων (Slicing and Dicing, Aggregations, Drilldowns, κ.α.). Κατ'αυτον τον τρόπο προσφέρεται η δυνατότητα χρήσης του Cubes χωρίς να απαιτείται άμεσα γνώση της Python από τον χρήστη/εφαρμογή, κάτι το οποίο καθιστά το framework αυτό ιδιαίτερα αυτόνομο και εύχρηστο.[12]

Για την χρήση του Cubes αρκεί ο χρήστης να υποβάλλει ένα απλό HTTP Request και θα λάβει την απάντηση σε μορφότυπο JSON. Η μορφοποιημένη αυτή απάντηση δίνει την δυνατότητα αξιοποίησης της από διάφορες εφαρμογές. Μερικά από τα πιο βασικά HTTP Requests είναι τα εξής[12]:

- GET /info: Επιστρέφει πληροφορίες για τον server και για τα δεδομένα του server
- GET /cubes: Επιστρέφει τους κύβους του μοντέλου
- GET /cube/<name>/model: Επιστρέφει πληροφορίες του συγκεκριμένου κύβου με όνομα "name"
- GET /cube/<name>/members/<dim>: Επιστρέφει τα μέλη μιας συγκεκριμένης διάστασης
- GET /cube/<name>/facts: Επιστρέφει τα facts του συγκεκριμένου κύβου
- GET /cube/<cube>/aggregate: Επιστρέφει τα αποτελέσματα των συναθροίσεων

Για να οριστεί το κατάλληλο κομμάτι του κύβου πάνω στο οποίο θέλουμε να πραγματοποιήσουμε τις τομές και τις καθόδους θα πρέπει να προσθέσουμε στις παραπάνω αιτήσεις τις σωστές παραμέτρους. Αυτό γίνεται βάζοντας ένα αγγλικό ερωτηματικό(?) ακολουθούμενο από τις ζητούμενες παραμέτρους. Στη συνέχεια φαίνονται κάποια παραδείγματα εκτέλεσης τομών και αναλυτικών καθόδων:

- Πωλήσεις του προϊόντος jeans  
GET /cube/sales/aggregate?cut=product:jeans
- Πωλήσεις του προϊόντος jeans και coats (Set Cut)  
GET /cube/sales/aggregate?cut=product:jeans;coats
- Πωλήσεις του προϊόντος jeans με τιμή 50 ευρώ  
GET /cube/sales/aggregate?cut=product:jeans|price:50
- Πωλήσεις απο το 2015 έως το 2018 (Range Cut)  
GET /cube/sales/aggregate?cut=date:2015-2018
- Πωλήσεις ανά προϊόν  
GET /cube/sales/aggregate?drilldown=product

- Πωλήσεις ανά μήνα  
GET /cube/sales/aggregate?drilldown=date:month
- Πωλήσεις ανά προϊόν με τιμή 50 ευρώ  
GET /cube/sales/aggregate?drilldown=product&cut=price:50

Για την κατάλληλη αρχικοποίηση του Slicer Server θα πρέπει να δημιουργηθεί ένα αρχείο .ini το οποίο θα περιέχει πληροφορίες που θα χρησιμοποιεί ο εξυπηρετητής, όπως το αρχείο στο οποίο έχει οριστεί το μοντέλο καθώς και το μέρος που βρίσκονται τα δεδομένα. Υποθέτουμε ξανά ότι τα δεδομένα αποθηκεύτηκαν στο SQLite:

```
[store]

type: sql

url: sqlite:///data.sqlite

[models]
model: sales.json
```

Για να «τρέξει» ο Εξυπηρετητής χρειάζεται να δοθεί η εξής εντολή:

→ slicer serve slicer.ini

## 3.5 Backends

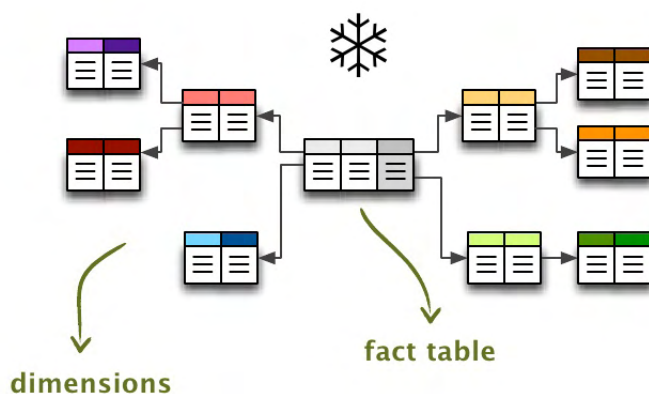
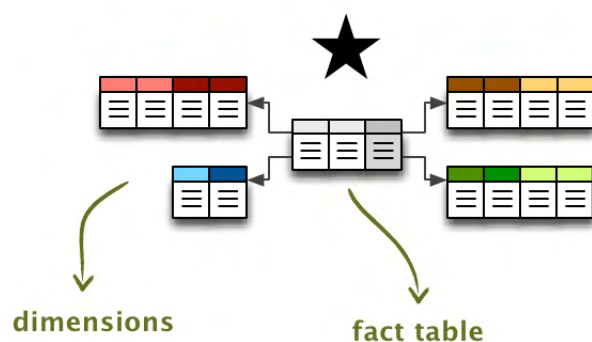
Το Backends, ή αλλιώς εσωτερικό μέρος, είναι αυτό που διαχειρίζεται τα «φυσικά» δεδομένα αποτελώντας το χαμηλότερο επίπεδο του framework. Το Cubes υποστηρίζει πληθώρα από database backends και λόγω αυτού πολλούς διαφορετικούς τρόπους υπολογισμού του κύβου αλλά και περιήγησης των συναθροιστικών δεδομένων [12]. Περιλαμβάνει τα εξής backends:

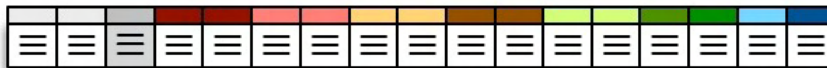
1. SQL Backend
2. MongoDB Backend
3. Mixpanel Backend
4. Slicer Server
5. Google Analytics Backend

Ιδιαίτερη έμφαση θα δοθεί στο SQL backend μιας και χρησιμοποιήθηκε για την υλοποίηση αυτής της εργασίας.

Το SQL Backend χρησιμοποιεί το SQLAlchemy το οποίο υποστηρίζει πληθώρα σχεσιακών βάσεων δεδομένων όπως οι MySQL, Oracle, PostgreSQL, SQLite κ.α. Το SQLAlchemy είναι ένα Python SQL Toolkit και Object Relational Mapper το οποίο είναι σχεδιασμένο για αποτελεσματική και υψηλής απόδοσης πρόσβαση σε βάσεις δεδομένων. Το συγκεκριμένο Backend μπορεί να υποστηρίξει απλές συναθροιστικές συναρτήσεις όπως είναι οι count, sum, min, max, avg, variance κ.α. [12].

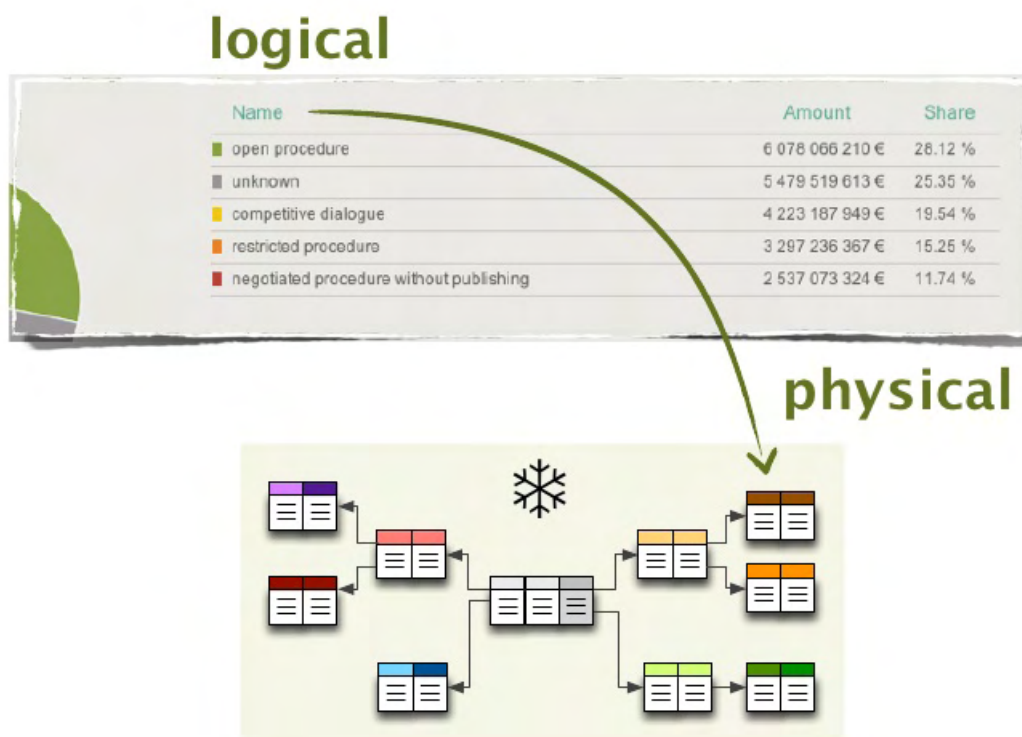
Το σχήμα των βάσεων δεδομένων που μπορεί να υποστηρίξει το SQL Backend είναι είτε αστέρα είτε χιονονιφάδας, καθώς και οποιονδήποτε μη κανονικοποιημένο πίνακα (denormalized table). Εάν τα δεδομένα δεν είναι σε καμία από τις παραπάνω μορφές θα πρέπει να εφαρμόσουμε πάνω τους denormalization ώστε να χρησιμοποιηθούν από το Cubes. Στο Σχήμα 3.7 φαίνονται τα τρία επιτρεπόμενα σχήματα βάσεων δεδομένων.





Σχήμα 3.6: Σχήματα αστέρα, χιονονιφάδας και μη κανονικοποιημένου πίνακα. (Πηγή: [11])

Σημαντικό κρίνεται για την λειτουργία του framework να υπάρχει αντιστοιχία των χαρακτηριστικών του λογικού μοντέλου με το φυσικό, δηλαδή της βάσης δεδομένων. Οι διαστάσεις και τα μέτρα που έχουν οριστεί στο μοντέλο θα πρέπει να αντιστοιχίζονται με την κατάλληλη στήλη του πίνακα της βάσης[12]. Για να επιτευχθεί αυτή η απόλυτη αντιστοιχία δεδομένων θα πρέπει η βάση δεδομένων να έχει τη μορφή αστέρα και τα ονόματα των στηλών και των πινάκων να ταιριάζουν με εκείνα του λογικού μοντέλου. Στις περιπτώσεις που δεν έχουμε τις παραπάνω προδιαγραφές κρίνεται αναγκία η δήλωση των αντιστοιχίσεων (mappings) καθώς και των κατάλληλων συνδέσεων (joins).



Σχήμα 3.7: Αντιστοίχιση Λογικού με Φυσικό Μοντέλο. (Πηγή: [11])

### 3.6 Εγκατάσταση του Cubes

Για να εγκατασταθεί το Cubes θα πρέπει προηγουμένως να έχουμε εγκατεστημένο ήδη τον pip installer. Εάν στο σύστημα μας έχουμε ήδη εγκατεστημένη κάποια έκδοση Python 2  $\geq 2.7.9$  ή Python 3  $\geq 3.7$  τότε ο pip είναι ήδη εγκατεστημένος. Έτσι για το Cubes δίνουμε στην γραμμή εντολών την εξής εντολή:

```
-> pip install cubes[all]
```

Για τον κώδικα του Cubes ακολουθούμε τον παρακάτω σύνδεσμο:

<https://github.com/DataBrewery/cubes>



## Κεφάλαιο 4

# Αναλυτική Επεξεργασία Δεδομένων Έργων Τέχνης με τη χρήση του Cubes

### 4.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα παρουσιαστεί ο τρόπος με τον οποίο επεξεργαζόμαστε το σετ από δεδομένα έργων τέχνης, ώστε να χρησιμοποιηθεί από το εργαλείο Cubes με σκοπό την ανάκτηση χρήσιμης πληροφορίας. Αρχικά παρουσιάζονται το σετ δεδομένων και η δημιουργία του λογικού μοντέλου που θα χρησιμοποιήσει το Cubes Framework, ενώ στη συνέχεια κάποια βασικά παραδείγματα λειτουργίας του.

### 4.2 Σετ Δεδομένων Έργων Τέχνης

Για την συγκεκριμένη εργασία έχει χρησιμοποιηθεί το dataset του Μουσείου Μοντέρνας Τέχνης (**MoMa**), το οποίο βρίσκεται στη Νέα Υόρκη. Το dataset υπάρχει διαθέσιμο προς χρήση στην ιστοσελίδα Kaggle και περιέχει πληροφορίες για τα έργα τέχνης και τους δημιουργούς τους.

Στην παρούσα εργασία έχουμε επιλέξει να δουλέψουμε με δεδομένα που υπάρχουν για το έτος 2016 τα οποία αποτελούνται από 1873 εγγραφές για τα έργα τέχνης και 495 εγγραφές για τους καλλιτέχνες. Για να μπορέσουμε να χειριστούμε σωστά το συγκεκριμένο σετ δεδομένων χρειάστηκε να προβούμε σε καθαρισμό δεδομένων (data cleaning) αλλά και σε δημιουργία όψεων (views) από τα δεδομένα, ώστε να μπορέσουμε να αποσπάσουμε την απαραίτητη πληροφορία.

Το τελικό dataset που προέκυψε μετά την κατάλληλη επεξεργασία των δεδομένων φαίνεται στον Πίνακα 4.1. Αποτελείται από έναν ενιαίο πίνακα με 16 γνωρίσματα (στήλες) ο οποίος παρέχει πληροφορίες για τα έργα τέχνης που απέκτησε το μουσείο την χρονιά 2016, καθώς και πληροφορίες για τους καλλιτέχνες που τα δημιούργησαν.

Γνώρισμα	Περιγραφή
artwork id	id έργου τέχνης
title	τίτλος έργου τέχνης
date	ημερομηνία δημιουργίας έργου τέχνης
medium	μέσο στο οποίο είναι υλοποιημένο το έργο
dimensions	διαστάσεις έργου τέχνης
acquisition date	ημερομηνία απόκτησης έργου από το μουσείο
quarters	κατηγοριοποίηση ημερομηνίας απόκτησης του έργου σε τρίμηνα του έτους 2016
credit	ο τρόπος που απέκτησε το μουσείο το έργο
department	τμήμα στο οποίο ανήκει το έργο
classification	κατάταξη του έργου σε κατηγορία
artist id	id του καλλιτέχνη του έργου
name	όνομα καλλιτέχνη
nationality	εθνικότητα καλλιτέχνη
gender	φύλο καλλιτέχνη
age	ηλικία καλλιτέχνη
age groups	ηλικιακή ομάδα καλλιτέχνη

Πίνακας 4.1: Γνώρισμα του σετ δεδομένων

Κάποιες στήλες του παραπάνω πίνακα δημιουργήθηκαν με τον ορισμό των κατάλληλων views (όψεων) ώστε να μας βοηθήσουν στην εξαγωγή συγκεκριμένων συμπερασμάτων. Ένα παράδειγμα είναι το γνώρισμα “quarters” το οποίο μπορεί να πάρει τέσσερις τιμές (quarter1, quarter2, quarter3, quarter4) όσα δηλαδή και τα τρίμηνα που μπορεί να χωριστεί ένα έτος. Επιλέξαμε να κατατάξουμε την ημερομηνία απόκτησης του έργου από το μουσείο (“acquisition date”) με κριτήριο σε ποιο τρίμηνο του έτους 2016 αποκτήθηκε, καθώς αποφασίσαμε ότι δεν μας ενδιαφέρει η ακριβής ημερομηνία απόκτησης αλλά μια γενικότερη κατάταξη. Ακόμη ένα τέτοιο παράδειγμα είναι το γνώρισμα “age groups”, το οποίο αποτελεί μια ομαδοποίηση της ηλικίας (“age”) του καλλιτέχνη, εφόσον η ακριβής ηλικία δεν περιείχε σημαντική πληροφορία όσο μια ηλικιακή ομαδοποίηση.

Είναι εμφανές ότι η διαμόρφωση της βάσης δεδομένων έχει γίνει με κριτήριο την εστίαση σε συγκεκριμένα συμπεράσματα που επιθυμούμε να παράγουμε. Σίγουρα ο κάθε αναλυτής μπορεί να επεξεργαστεί το σετ δεδομένων ανάλογα με την πληροφορία που επιθυμεί να ανακτήσει, αφού όμως πρώτα εστιάσει μόνο στα δεδομένα που επιθυμεί κατά την δημιουργία του λογικού μοντέλου.

### 4.3 Μετατροπή του CSV αρχείου σε SQLite Table

Το σετ δεδομένων που προαναφέρθηκε βρίσκεται σε μορφή csv, μια μορφή την οποία δεν μπορεί να διαχειριστεί το Cubes. Γι αυτόν τον λόγο το αρχείο csv θα μετατραπεί σε SQLite table με τον κατάλληλο κώδικα σε Python. Το σύστημα διαχείρισης σχεσιακών βάσεων δεδομένων είναι επιλογή του χρήστη. Συγκεκριμένα σε αυτή την εργασία έχει επιλεγεί το SQLite. Στον παρακάτω αρχείο `preparation.py` φαίνεται αυτή η μετατροπή.

```
from sqlalchemy import create_engine
from cubes.tutorial.sql import create_table_from_csv

engine = create_engine('sqlite:///data.sqlite')

create_table_from_csv(engine,
    "v_artworks.csv",
    table_name="artworks",
    fields=[
        ("artwork_id", "integer"),
        ("title", "string"),
        ("date", "string"),
        ("medium", "string"),
        ("dimensions", "string"),
        ("acquisition_date", "date"),
        ("quarters", "string"),
        ("credit", "string"),
        ("department", "string"),
        ("classification", "string"),
        ("artist_id", "integer"),
        ("name", "string"),
        ("nationality", "string"),
        ("gender", "string"),
        ("age", "string"),
        ("age_groups", "integer")],
    create_id= False
)
```

Για την μετατροπή του csv αρχείου σε SQLite table χρησιμοποιείται η συνάρτηση `create_table_from_csv`, η οποία δημιουργεί έναν πίνακα με τις επιθυμητές στήλες του αρχείου csv χρησιμοποιώντας εδώ το SQLite. Η παραπάνω συνάρτηση έχει τις εξής πέντε παραμέτρους:

1. Η database engine που θα χρησιμοποιηθεί (π.χ. SQLite).
2. Το αρχείο csv στο οποίο βρίσκονται τα δεδομένα.
3. Το όνομα του πίνακα που θα δημιουργηθεί.
4. Οι στήλες του πίνακα που θα δημιουργηθούν. Πρώτα δηλώνεται το όνομα της στήλης και στην συνέχεια ο τύπος δεδομένων (integer, string, κ.λπ.).
5. Η παράμετρος “create\_id”. Αν αρχικοποιηθεί σε “True” δημιουργεί μία επιπλέον στήλη “id” ώστε να χρησιμοποιηθεί σαν πρωτεύον κλειδί του πίνακα.

Κατ’ αυτόν τον τρόπο έχουμε δημιουργήσει ένα αρχείο sqlite μέσα στο οποίο περιέχονται τα δεδομένα με τη μορφή ενός μη κανονικοποιημένου πίνακα με όνομα “artworks”.

#### 4.4 Υλοποίηση Λογικού Μοντέλου

Με βάση τα δεδομένα που προαναφέρθηκαν έχει υλοποιηθεί το λογικό μοντέλο το οποίο θα χρησιμοποιήσει το Cubes για την αναλυτική επεξεργασία, από τη “σκοπιά” που ο αναλυτής επιλέγει να επεξεργαστεί τις πληροφορίες. Το λογικό μοντέλο διαμορφώνεται αναλόγως, σύμφωνα με τα συμπεράσματα στα οποία ο χρήστης επιθυμεί να καταλήξει.

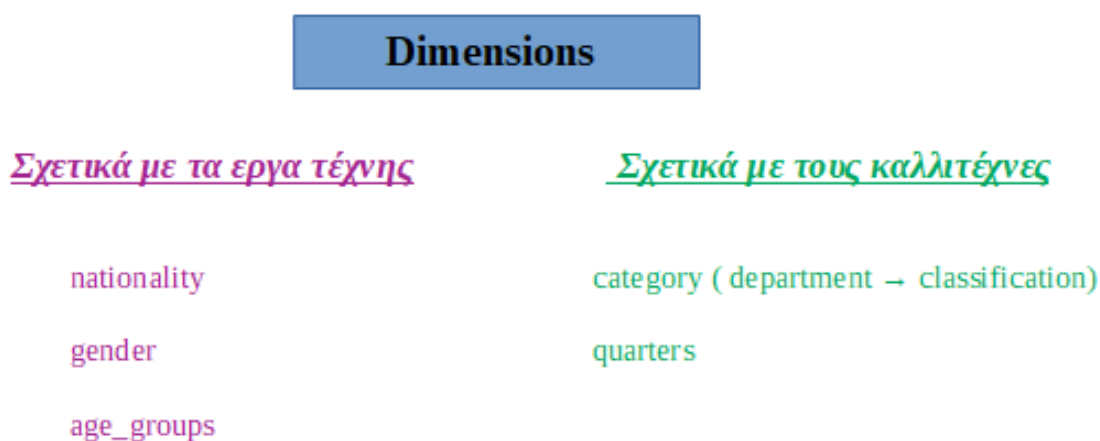
Το συγκεκριμένο μοντέλο αποτελείται από έναν (υπερ-)κύβο με πέντε διαστάσεις, με τις δύο εκ των οποίων να παρέχουν πληροφορίες σχετικά με τα έργα τέχνης (artworks) και τις υπόλοιπες τρεις για τους καλλιτέχνες (artists). Ακόμη περιέχει δύο μέτρα (measures), ένα που αφορά τα έργα τέχνης και ένα τους καλλιτέχνες. Στο τμήμα κώδικα που ακολουθεί φαίνονται τα μεταδεδομένα του μοντέλου (name, label, cubes, dimensions) καθώς και ο ορισμός του σε αρχείο JSON.

```
{  
  
    "name": "moma_artworks_model",  
  
    "label": "Moma Artworks 2016",  
  
    "dimensions": [...],  
  
    "cubes": [...]  
}
```

Για τον κύβο έχουν οριστεί οι εξής πέντε διαστάσεις:

1. “nationality”: Η εθνικότητα του καλλιτέχνη.
2. “gender”: Το φύλο του καλλιτέχνη.
3. “age\_groups”: Η ηλικιακή ομάδα στην οποία ανήκει ο καλλιτέχνης.

4. “category”: Η κατηγορία στην οποία ανήκει ένα έργο τέχνης. Αυτή η διάσταση είναι διεπίπεδη, πρώτα υπάρχει το επίπεδο “department”, το τμήμα δηλαδή στο οποίο κατατάσσεται το έργο (film, architecture, paint, κ.α.) και δεύτερο το επίπεδο “classification”, μια πιο ειδική κατηγορία κατάταξης του έργου.
5. “quarters”: Η κατηγοριοποίηση ημερομηνίας απόκτησης του έργου από το μουσείο σε τρίμηνα του έτους 2016 (quarter\_1, quarter\_2, quarter\_3, quarter\_4).



Σχήμα 4.1: Οι διαστάσεις του Λογικού Μοντέλου

Οπώς είναι προφανές από την παραπάνω περιγραφή και από το Σχήμα 4.1, οι διαστάσεις “nationality”, “gender”, “age\_groups” αφορούν τους καλλιτέχνες των έργων ενώ οι “category” και “quarters” τα ίδια τα έργα τέχνης. Σχετικά με την διάσταση “category” η οποία αποτελείται από δύο επίπεδα (“department”, “classification”) έχουν οριστεί δύο διαφορετικές ιεραρχίες των επιπέδων της. Η πρώτη είναι η “default” η οποία θα περιέχει τα δύο επίπεδα με την σειρά την οποία έχουν οριστεί, ενώ η δεύτερη είναι η “only\_classification” η οποία παραλείπει το “department” και περιέχει μόνο το επίπεδο “classification” με την πιο ειδική κατάταξη του έργου στο αντίστοιχο τμήμα έργων τέχνης που ανήκει. Παρακάτω φαίνονται τα μεταδεδομένα που έχουν οριστεί για τις διαστάσεις του κύβου.

```
"dimensions":[  
  {  
    "name": "nationality",  
    "label": "nationality",  
    "attributes": ["nationality"]  
  },  
  {  
    "name": "gender",  
    "label": "gender",  
    "attributes": ["gender"]  
  },  
  {  
    "name": "quarters",  
    "label": "quarters",  
    "attributes": ["quarters"]  
  },  
  {  
    "name": "age_groups",  
    "label": "age_groups",  
    "attributes": ["age_groups"]  
  },  
]
```

```
{
  "name": "category",
  "levels": [
    {
      "name": "department",
      "label": "department",
      "attributes": ["department"]
    },
    {
      "name": "classification",
      "label": "classification",
      "attributes": ["classification"]
    }
  ],
  "hierarchies": [
    {
      "name": "default",
      "levels": ["department", "classification"]
    },
    {
      "name": "only_classification",
      "levels": ["classification"]
    }
  ]
}]
```

Ο κύβος του μοντέλου ονομάζεται “artworks” και όπως προαναφέρθηκε έχει πέντε διαστάσεις και δύο μέτρα. Για τον κύβο επιπλέον έχουν οριστεί στο αρχείο JSON τα μεταδεδομένα aggregates, mappings και details, τμήμα του κώδικα των οποίων θα ακολουθήσει στη συνέχεια. Παρακάτω παρουσιάζονται τα μεταδεδομένα που έχουν οριστεί για τον κύβο.

```

"cubes": [
{
"name": "artworks",
"label": "Momarworks",
"dimensions": ["age_groups", "nationality", "gender", "category", "quarters"],
"measures": [.....],
"aggregates": [.....],
"details": [.....],
"mappings": {.....}
}]

```

Πιο αναλυτικά, ο κύβος περιέχει τα μέτρα (measures) “artwork\_id” και “artist\_id” τα οποία ουσιαστικά χρησιμεύουν στην συνολική καταμέτρηση των έργων τέχνης αλλά και των δημιουργών τους αντιστοίχα. Σχετικά με τις συναθροίσεις (aggregates) του μοντέλου θα πρέπει να έχει οριστεί το μέτρο αλλά και η συνάρτηση (function) για τον υπολογισμό τους. Το συγκεκριμένο κομμάτι κώδικα που ακολουθεί αντιστοιχεί στον ορισμό των measures και aggregates του κύβου. Όπως μπορούμε να δούμε στις συναθροίσεις που ορίζονται χρησιμοποιείται η συνάρτηση “count\_distinct” (μετράει μόνο μοναδικές τιμές) με την οποία υπολογίζεται ο συνολικός αριθμός των έργων και των καλλιτεχνών αντίστοιχα.

```

"measures": ["artwork_id", "artist_id"],

"aggregates": [
{
"name": "total number of artworks",
"function": "count_distinct",
"measure": "artwork_id"
},

{
"name": "total number of artists",
"function": "count_distinct",
"measure": "artist_id"
}]

```



Παρακάτω θα δούμε ότι έχουμε ορίσει ως “details” κάποιες στήλες της βάσης δεδομένων, όπως “title”, “dimensions” και “credit”, τις οποίες δεν επιθυμούσαμε να χρησιμοποιήσουμε ως παραμέτρους της πληροφορίας προς ανάκτηση. Παρ’ όλα αυτά έχουν δηλωθεί ως λεπτομέρειες και εμφανίζονται από τη συνάρτηση “facts”, η οποία επιστρέφει πληροφορίες για τα στοιχεία που βρίσκονται στο κελί του κύβου που ορίζουμε κατά την OLAP.

```
"details": [  
  {  
    "name": "title",  
    "label": "artwork title"  
  },  
  {  
    "name": "dimensions",  
    "label": "dimensions"  
  },  
  {  
    "name": "credit",  
    "label": "credit"  
  }  
]
```

Τέλος, σημαντικό κομμάτι της δημιουργίας του μοντέλου αποτελούν οι αντιστοιχίσεις (mappings), δηλαδή ουσιαστικά η σύνδεση μεταξύ του λογικού μοντέλου και των κατάλληλων στηλών των πινάκων της βάσης δεδομένων. Επειδή η βάση δεδομένων βρίσκεται σε έναν μοναδικό αποκανονικοποιημένο πίνακα, και προφανώς οι διαστάσεις υπάρχουν ως στήλες στον πίνακα αυτό, θα πρέπει να οριστούν οι αντιστοιχίσεις για τις διαστάσεις με πολλαπλά επίπεδα. Αυτό συμβαίνει διότι το backend περιμένει έναν fact table “artworks”, ο οποίος περιέχει όλες τις διαστάσεις σε διαφορετικούς πίνακες. Κάτι τέτοιο δεν συμβαίνει στο συγκεκριμένο μοντέλο που μελετάμε, άρα θα πρέπει να οριστούν ρητά οι αντιστοιχίσεις για την διάσταση “category” η οποία αποτελείται, όπως είδαμε νωρίτερα, από τα επίπεδα “department” και “classification”. Ακολουθεί η δήλωση των mappings του μοντέλου.

```
"mappings": {  
  "category.department": "department",  
  "category.classification": "classification"  
}
```

## 4.5 Υλοποίηση του Εξυπηρετητή Slicer

Έχοντας ολοκληρώσει την διαδικασία υλοποίησης του λογικού μοντέλου για να την άμεση αναλυτική επεξεργασία των δεδομένων, απαραίτητο κομμάτι είναι η αρχικοποίηση του ενσωματωμένου HTTP Server του Cubes, του Slicer Server. Αν και στη συγκεκριμένη εργασία γίνεται χρήση του Cubes από το Cubes Viewer για την επεξεργασία και οπτικοποίηση των δεδομένων από τον χρήστη, ωστόσο είναι αρκετά χρήσιμο να αναλυθεί η λειτουργία του Slicer Server, αφού προσφέρει στον χρήστη άμεση χρήση του εργαλείου χωρίς την απαραίτητη υλοποίηση σε Python.

Ξεκινώντας θα πρέπει να δημιουργηθεί το αντίστοιχο αρχείο `.ini`, όπως αναφέρθηκε στο προηγούμενο κεφάλαιο, για τον ορισμό των πληροφοριών σχετικά με τον εξυπηρετητή. Όπως μπορούμε να δούμε και παρακάτω στο αντίστοιχο αρχείο, έχει δηλωθεί το μοντέλο στο αρχείο `.json` καθώς και το μέρος στο οποίο βρίσκονται τα δεδομένα. Στη συγκεκριμένη περίπτωση πρόκειται για το αρχείο `.sqlite` που δημιουργήσαμε στην υποενότητα 4.3.

```
[store]

type: sql

url: sqlite:///data.sqlite

[models]

model: moma_model.json
```

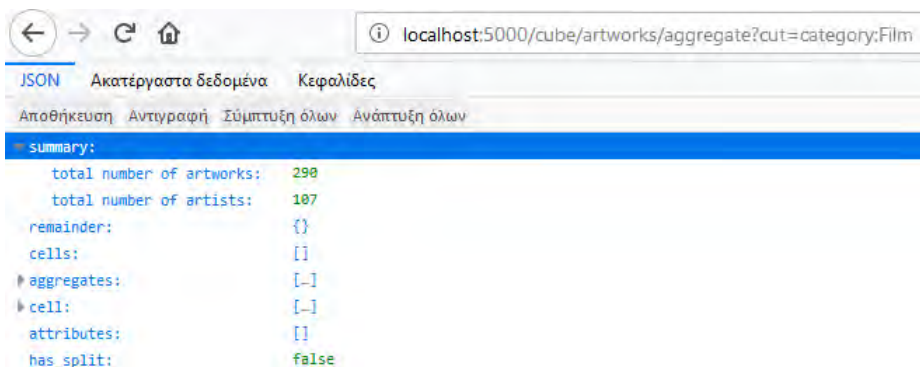
Για την εκτέλεση στον Server αρκεί να εισέλθουμε στον φάκελο που βρίσκονται όλα τα σχετικά αρχεία που έχουμε δημιουργήσει μέχρι στιγμής ( `.sqlite`, `.json`, `.ini` ) και να “τρέξουμε” την παρακάτω εντολή στην γραμμή εντολών.

→ **slicer serve slicer.ini**

Ο χρήστης πλέον έχει την δυνατότητα να μεταβεί στην ιστοσελίδα **`http://localhost:5000/cube/artworks`** στην οποία τρέχει ο Server και να δοκιμάσει να υποβάλλει ορισμένα HTTP Requests, σαν αυτά που μελετήθηκαν στο προηγούμενο κεφάλαιο. Στην συνέχεια παρατίθενται κάποια βασικά παραδείγματα queries(ερωτημάτων) και αντίστοιχα responses (αποκρίσεις) του εξυπηρετητή:

- **aggregate?cut=category:Film**

Με αυτό το query επιθυμούμε να λάβουμε ως αποτέλεσμα όλες τις συναθροίσεις που έχουν οριστεί επιλέγοντας ως κελί το υποσύνολο των δεδομένων που αφορά έργα τέχνης που ανήκουν στην κατηγορία “Film”. Παρακάτω βρίσκεται το αποτέλεσμα του ερωτήματος σε μορφή JSON, το οποίο περιέχει τις συναθροίσεις για τα δεδομένα που πληρούν τα συγκεκριμένα κριτήρια.



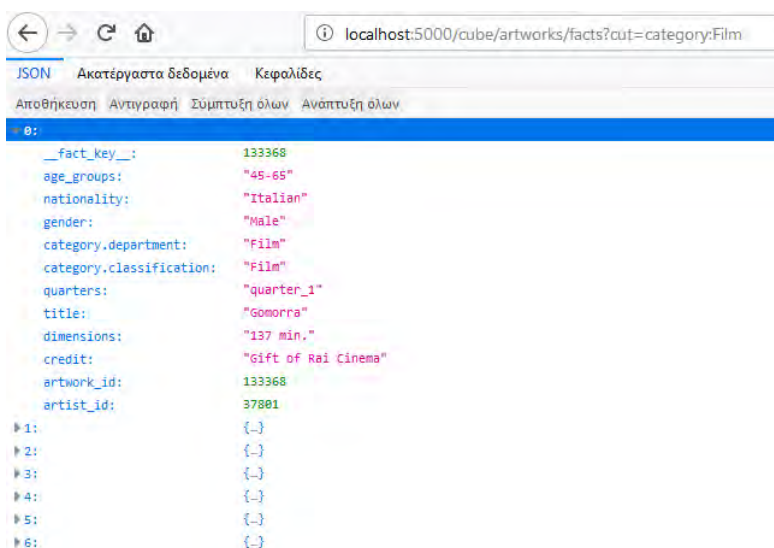
```

{
  "summary": {
    "total number of artworks": 290,
    "total number of artists": 107,
    "remainder": {},
    "cells": [],
    "aggregates": [],
    "cell": [],
    "attributes": [],
    "has_split": false
  }
}

```

- **facts?cut=category:Film**

Με αυτό το query επιθυμούμε να λάβουμε ως αποτέλεσμα τα στοιχεία (έργα τέχνης) που ανήκουν στο κελί του προηγούμενου ερωτήματος. Τα έργα τέχνης που πληρούν το κριτήριο είναι 290, αλλά στο συγκεκριμένο απόσπασμα του αποτελέσματος φαίνονται αναπτυγμένα μόνο τα δεδομένα του πρώτου έργου τέχνης, ενώ όλα τα υπόλοιπα παρουσιάζονται συμπυκνμένα.



```

{
  "0": {
    "__fact_key__": 133368,
    "age_groups": "45-65",
    "nationality": "Italian",
    "gender": "Male",
    "category.department": "Film",
    "category.classification": "Film",
    "quarters": "quarter_1",
    "title": "Gomorra",
    "dimensions": "137 min.",
    "credit": "Gift of Rai Cinema",
    "artwork_id": 133368,
    "artist_id": 37801
  },
  "1": {},
  "2": {},
  "3": {},
  "4": {},
  "5": {},
  "6": {}
}

```

- **aggregate?cut=category:Film&cut=gender:Female**

Σε αυτό το query εκτός από την τομή στην διάσταση “category” (δηλαδή στο πρώτο επίπεδο της, “department”) πραγματοποιούμε ταυτόχρονα και μία τομή στη διάσταση “gender” με συγκεκριμένη τιμή “Female”.

```

{
  "summary": {
    "total number of artworks": 32,
    "total number of artists": 21,
    "remainder": {},
    "cells": []
  },
  "aggregates": {
    "0": "total number of artworks",
    "1": "total number of artists",
    "cell": [ "- ]",
    "attributes": [],
    "has_split": false
  }
}

```

- **aggregate?cut=category:Film&drilldown=gender**

Σε αυτό το query εκτός από την τομή στην διάσταση “category”, που υπάρχει και σε όλα τα προηγούμενα παραδείγματα, πραγματοποιούμε παράλληλα αναλυτική κάθοδο (drilldown) στην διάσταση “gender”. Στο αποτέλεσμα που παρουσιάζεται φαίνονται οι συναθροίσεις που έχουν οριστεί από το κελί που αντιστοιχεί στα έργα τέχνης που ανήκουν στην κατηγορία “Film” ομαδοποιημένα κατά το φύλο των καλλιτεχνών (“gender”).

```

{
  "summary": {
    "remainder": {}
  },
  "cells": {
    "0": {
      "gender": "Female",
      "total number of artworks": 32,
      "total number of artists": 21
    },
    "1": {
      "gender": "Male",
      "total number of artworks": 216,
      "total number of artists": 80
    },
    "2": {
      "gender": "unknown",
      "total number of artworks": 47,
      "total number of artists": 6
    }
  },
  "total_cell_count": 3,
  "aggregates": [ "- ]",
  "cell": [ "- ]",
  "levels": { "- }",
  "attributes": [ "- ]",
  "has_split": false
}

```

## Κεφάλαιο 5

# Χρήση του CubesViewer

### 5.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα περιγράψουμε τη χρήση του εργαλείου Cubes Viewer το οποίο χρησιμοποιείται για οπτικοποίηση δεδομένων με την μέθοδο της άμεσης αναλυτικής επεξεργασίας. Θα αναπτύξουμε τις βασικές δυνατότητες του εργαλείου αυτού καθώς και την εφαρμογή του σε συνδυασμό με το μοντέλο δεδομένων που δημιουργήθηκε στο προηγούμενο κεφάλαιο με σκοπό την ανάκτηση πληροφορίας.

### 5.2 Παρουσίαση του Cubes Viewer

Το Cubes Viewer αποτελεί ένα web-based εργαλείο/εφαρμογή για ανάλυση και οπτικοποίηση OLAP δεδομένων που εξυπηρετείται από το Cubes OLAP Framework [10].



Μπορεί να χρησιμοποιηθεί για εξερεύνηση δεδομένων, δημιουργία αναφορών απο αναλυτές, σχεδιασμό γραφημάτων και ενσωμάτωση, αλλά και ως απλή εφαρμογή ανάλυσης για κάποια εταιρεία [7].

Στη συγκεκριμένη εργασία χρησιμοποιείται για την οπτικοποίηση των δεδομένων και την εξαγωγή συμπερασμάτων που θα οδηγήσουν με την σειρά τους στην επιθυμητή ανάκτηση πληροφορίας από την βάση δεδομένων που διαθέτουμε.

Το Cubes Viewer διαθέτει αρκετές δυνατότητες και μπορεί είτε να ενσωματωθεί σε μια υπάρχουσα εφαρμογή για αναλυτική επεξεργασία δεδομένων, είτε να σταθεί μόνο του σαν εφαρμογή τέτοιου σκοπού. Μερικές από τις δυνατότητες του είναι οι εξής [7]:

- OLAP Browser: Εξερεύνηση των δεδομένων κατα μήκος διαστάσεων. Δυνατότητα φιλτράρισματος του αποτελέσματος και αναλυτικής καθόδου.
- Charts: Είναι διαθέσιμο ένα ευρύ φάσμα τρόπων οπτικοποίησης δεδομένων. Υπάρχει εύκολη δυνατότητα αλλαγής αναπαράστασης.
- Exports: Δυνατότητα εξαγωγής γραφισμάτων αλλά και δεδομένων.
- Undo, Redo, Clone: Δυνατότητα επανάληψης, ακύρωσης και κλωνοποίησης. Εύκολη υλοποίηση παραλλαγών σε μία όψη(view).
- Embed Views: Δυνατότητα ενσωμάτωσης των όψεων που θα προκύψουν από την επεξεργασία σε άλλες ιστοσελίδες.
- CubesViewer Server: Προαιρετικό multi-user (πολλαπλών χρηστών) backend για αποθήκευση και διαμοιρασμό όψεων.

### 5.2.1 Εγκατάσταση και τρόπος λειτουργίας του CubesViewer

Αρχικά, το CubesViewer αποτελεί μία ανοιχτού κώδικα (open source) εφαρμογή και ο κώδικας της οποίας βρίσκεται στον παρακάτω σύνδεσμο:

→ <https://github.com/jjmontesl/cubesviewer>

Το πακέτο που θα χρησιμοποιήσουμε στη συγκεκριμένη εργασία είναι το CubesViewer (τελευταία έκδοση 2.0.2), client-side, το οποίο μπορεί κάποιος να κατεβάσει είτε από τον παραπάνω σύνδεσμο, είτε από την ιστοσελίδα [www.cubesviewer.com](http://www.cubesviewer.com).

Για να συνδεθεί το Cubes Framework με το CubesViewer, ώστε να “σερβίρουμε” το μοντέλο δεδομένων που έχουμε δημιουργήσει στο Κεφάλαιο 4, χρειάζονται κάποια επιπρόσθετα βήματα τα οποία παρουσιάζονται αναλυτικά παρακάτω.

Εφόσον έχουμε “κατεβάσει” το πακέτο του CubesViewer πρέπει να προσθέσουμε στο αρχείο .ini της αρχικοποίησης του Slicer Server κάποιες ακόμη παραμέτρους σχετικά με τις πληροφορίες για τον server. Ακολουθεί το αρχείο .ini όπως έχει προκύψει τελικώς.

```
[server]

host: localhost
port: 5000
reload: yes
log_level: info
allow_cors_origin: *

[store]

type: sql
url: sqlite:///data.sqlite

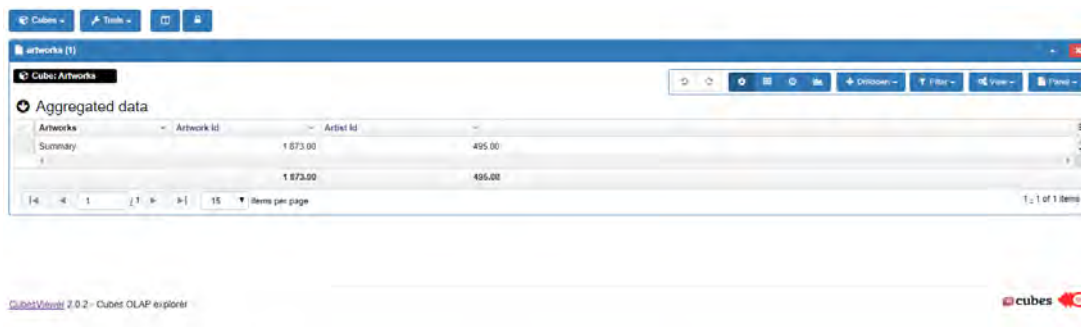
[models]

model: onetable.json
```

Όπως φαίνεται και στο αρχείο, για να είναι δυνατό οι χρήστες του CubesViewer να συνδεθούν στον Slicer Server του Cubes, έχει προστεθεί κάτω από το τμήμα του server το setting (ρύθμιση) “allow\_cors\_origin: \*”. Η συγκεκριμένη ρύθμιση επιτρέπει στον Server του Cubes τον μηχανισμό cross origin resource sharing (CORS) ώστε να είναι εφικτό να χρησιμοποιούνται κάποια resources (πόροι) του από άλλες ιστοσελίδες με διαφορετικό domain, στη συγκεκριμένη περίπτωση από το CubesViewer.

Στη συνέχεια “τρέχουμε” τον Cubes OLAP server με την διαδικασία που παρουσιάστηκε στην Υποενότητα 4.5, και αφού σιγουρευτούμε ότι “τρέχει”, χρησιμοποιούμε τον εξυπηρετητή για να φορτώσουμε το αρχείο studio.html, το οποίο βρίσκεται στο πακέτο του CubesViewer, μέσα στον φάκελο με όνομα “html”. Όταν ζητηθεί, επιλέγουμε την προκαθορισμένη τοποθεσία που “τρέχει” ο Cubes Server, δηλαδή **http://localhost:5000**. Κατ’ αυτόν τον τρόπο χρησιμοποιούμε ένα αντίγραφο του CubesViewer για να έχουμε πρόσβαση στον τοπικό Cubes Server.

Αφού έχει προηγηθεί όλη η παραπάνω διαδικασία εμφανίζεται το περιβάλλον του CubesViewer στη διεύθυνση **http://localhost:5000** και πατώντας κλικ πάνω στο “Cubes” φορτώνουμε τον κύβο δεδομένων με όνομα “artworks” που έχουμε δημιουργήσει. Ακολουθεί ένα στιγμιότυπο του προαναφερθέντος περιβάλλοντος στο οποίο φαίνεται ο κύβος δεδομένων όπως εμφανίζεται από το CubesViewer.



Σε αυτό το σημείο θα περιγράψουμε τις βασικές επιλογές που δίνονται από το CubesViewer για την εξερεύνηση και οπτικοποίηση των δεδομένων.

Ξεκινώντας από τα αριστερά, υπάρχει το κουμπί “Cubes”, που αναφέρθηκε παραπάνω, μέσα απ’ το οποίο επιλέγεται ο εκάστοτε κύβος δεδομένων που θέλουμε να διερευνήσουμε. Στη συνέχεια υπάρχει το κουμπί “Tools”, το οποίο παρέχει διάφορες δυνατότητες, όπως πληροφορίες για τον server, διαμόρφωση αποτελέσματος οπτικοποίησης κ.α.

Στην δεξιά πλευρά βρίσκονται όλα τα κουμπιά που σχετίζονται με την αναλυτική επεξεργασία του κύβου δεδομένων. Πιο συγκεκριμένα, ακολουθεί μία λίστα με τις δυνατότητες του εργαλείου που αφορούν την ανάλυση αλλά και την οπτικοποίηση των δεδομένων.

- Drilldown : αναλυτική κάθοδος ως προς τις διαστάσεις του μοντέλου
- Filter : Dimension filter (τομές στις διαστάσεις), Date filter
- View : Export table , Export facts, Chart Types  
Εξαγωγή του πίνακα ή των facts που θα προκύψουν απ’ την επεξεργασία ως αρχείο .csv  
Επιλογή τύπου γραφήματος. Υπάρχουν οι επιλογές pie (πίτα), bars vertical (κάθετες μπάρες), bars horizontal(οριζόντιες μπάρες), lines (γραμμές) κ.α.
- Facts : Εμφανίζει τα δεδομένα facts και τα στοιχεία τους
- Series : Εμφανίζει τα αποτελέσματα της επεξεργασίας με τη μορφή series table
- Charts : Εμφανίζει τα αποτελέσματα της επεξεργασίας με γραφήματα. Υπάρχει η επιλογή το γράφημα να είναι ομαδοποιημένο(grouped), σωρού (stacked), stream κ.α.
- Panel : Δίνει την δυνατότητα αλλαγής ονόματος του πάνελ (rename), κλωνοποίησης των δεδομένων κ.α.
- Undo, Redo : Ακύρωση ενέργειας, Επανάληψη ενέργειας

Αυτές είναι μόνο μερικές από τις δυνατότητες που προσφέρει το CubesViewer στον χρήστη. Για την συγκεκριμένη εργασία θα χρησιμοποιηθούν σε γενικά πλαίσια οι περισσότερες από τις παραπάνω.



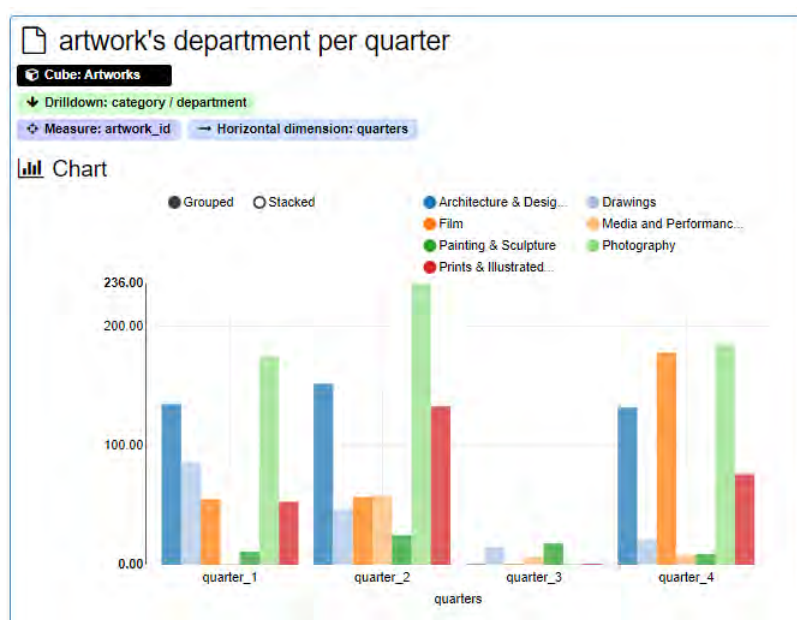
### 5.3 Οπτικοποίηση του κύβου δεδομένων μέσω του εργαλείου CubesViewer

Η οπτική αναπαράσταση των δεδομένων διευκολύνει τον άμεσο εντοπισμό σχέσεων και μοτίβων που υπάρχουν μέσα στα δεδομένα κάτι το οποίο καθιστά ευκολότερη και γρηγορότερη την ερμηνεία των δεδομένων και κατ' επέκταση την ανάκτηση πληροφορίας από αυτά. Προηγουμένως αναλύθηκε η σύνδεση του Cubes Server (OLAP Server) με την οπτική διεπαφή CubesViewer. Σε αυτή την Υποενότητα θα παρουσιαστούν μερικά βασικά παραδείγματα οπτικοποίησης των δεδομένων του κύβου “artworks”, που έχουμε δημιουργήσει για να επιτύχουμε την αναλυτική επεξεργασία των δεδομένων.

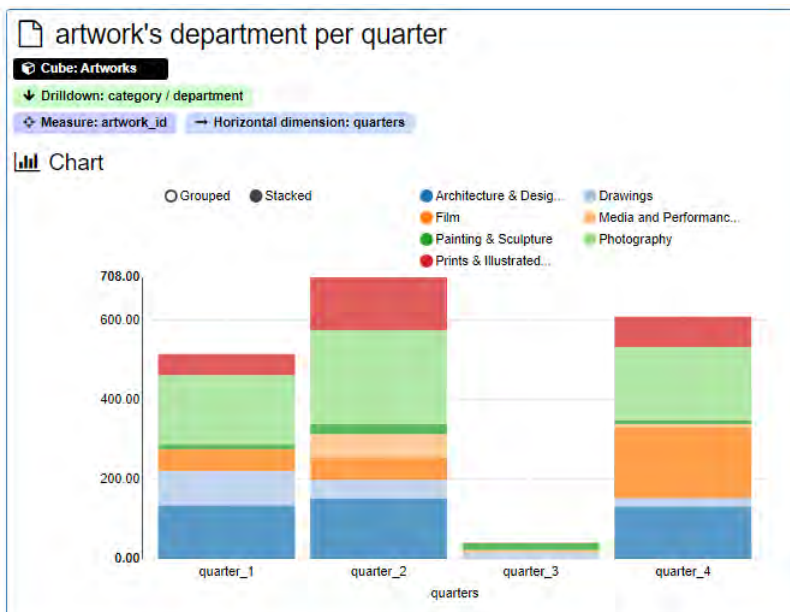
#### *Παράδειγμα πρώτο:*

Στις παρακάτω εικόνες 5.1 και 5.2 φαίνονται τα διαγράμματα διαφορετικών τύπων (bars vertical grouped και bars vertical stacked) που απεικονίζουν τον συνολικό αριθμό έργων τέχνης κάθε τμήματος του μουσείου ανά τέταρτα του ενός έτους. Αυτό το αποτέλεσμα επιτυγχάνεται ορίζοντας τα εξής:

- αναλυτική κάθοδο στη διάσταση “category” και στο πρώτο επίπεδό της “department”
- οριζόντια διάσταση η “quarters”
- μέτρο το “artworks\_id”

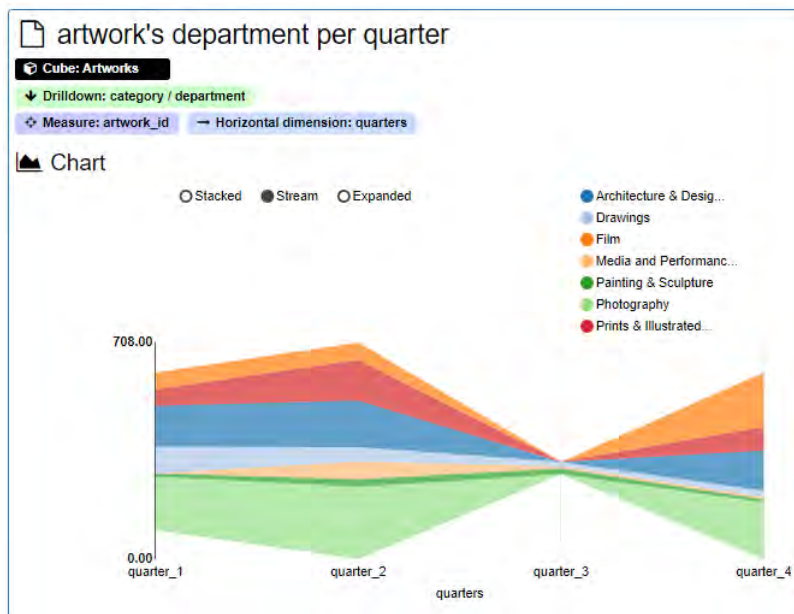


Σχήμα 5.1: Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (grouped).

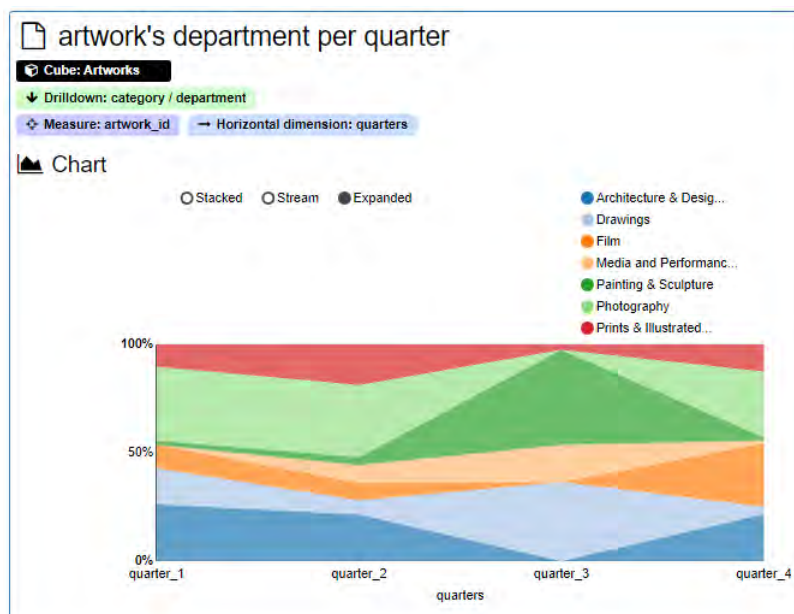


Σχήμα 5.2: Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stacked).

Επιπροσθέτως για το ίδιο παράδειγμα ακολουθούν δύο διαφορετικοί τύποι διαγραμμάτων στα Σχήματα 5.3 και 5.4 (stream stacked και stream expanded), ώστε να αναδειχθούν οι πολλαπλές δυνατότητες οπτικοποίησης δεδομένων που προσφέρει το CubesViewer στον χρήστη.

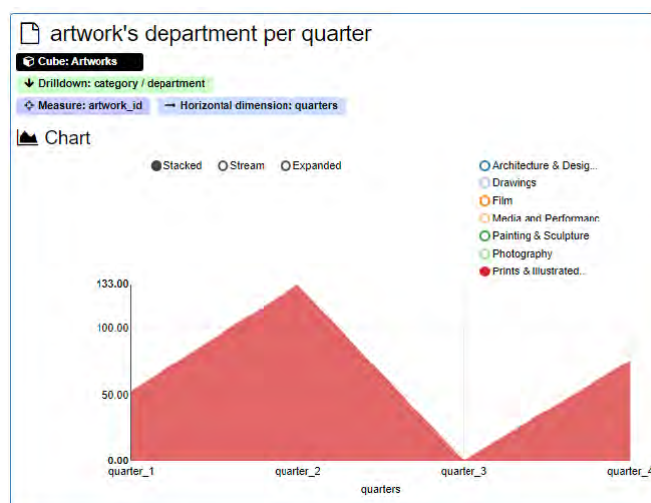


Σχήμα 5.3: Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stream stacked).



Σχήμα 5.4: Συνολικός αριθμός έργων τέχνης κάθε τμήματος ανα τέταρτα (stream expanded).

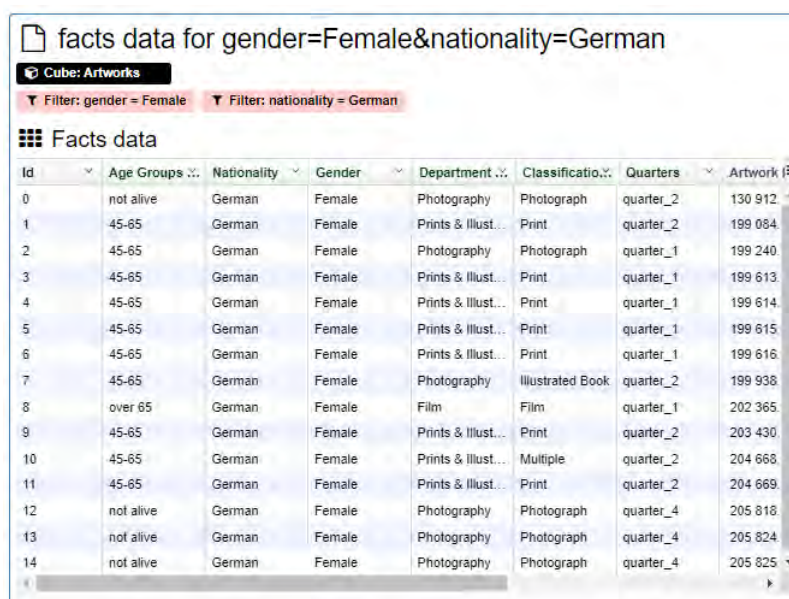
Αξίζει σε αυτό το σημείο να σημειωθεί ότι ο χρήστης έχει την δυνατότητα να “βλέπει” την ακριβή τιμή του μέτρου (measure), που έχει ορίσει να παρουσιάζεται σε κάθε διάγραμμα, σύροντας απλά το ποντίκι στο σημείο του γραφήματος που τον ενδιαφέρει. Ακόμη, κάνοντας κλικ πάνω σε μία συγκεκριμένη τιμή μιας διάστασης που απεικονίζεται, μπορεί να την απομονώσει και να πάρει το αντίστοιχο διάγραμμα που αντιστοιχεί σε αυτή τη συγκεκριμένη τιμή, αποκρύπτοντας τις υπόλοιπες. Ακολουθεί σχετικό παράδειγμα διαγράμματος (areas stacked) το οποίο περιέχει στην διάσταση “department” μόνο τα “Prints & Illustrated Books”.



Σχήμα 5.5: Απεικόνιση μόνο του συνολικού αριθμού έργων κατηγορίας “Prints & Illustrated Books”(areas stacked) .

**Παράδειγμα δεύτερο:**

Στο συγκεκριμένο παράδειγμα επιθυμούμε να λάβουμε ως αποτέλεσμα τα στοιχεία (facts data) που ανήκουν στο κελί που έχει οριστεί ως υποσύνολο δεδομένων που αφορά γυναίκες καλλιτέχνες ΚΑΙ γερμανικής εθνικότητας. Έχουν πραγματοποιηθεί δύο τομές, μία στη διάσταση “gender” και άλλη μία στη διάσταση “nationality”. Στο Σχήμα 5.6 απεικονίζεται το αποτέλεσμα (δεν είναι πλήρως ανεπτυγμένο για λόγους οικονομίας χώρου).

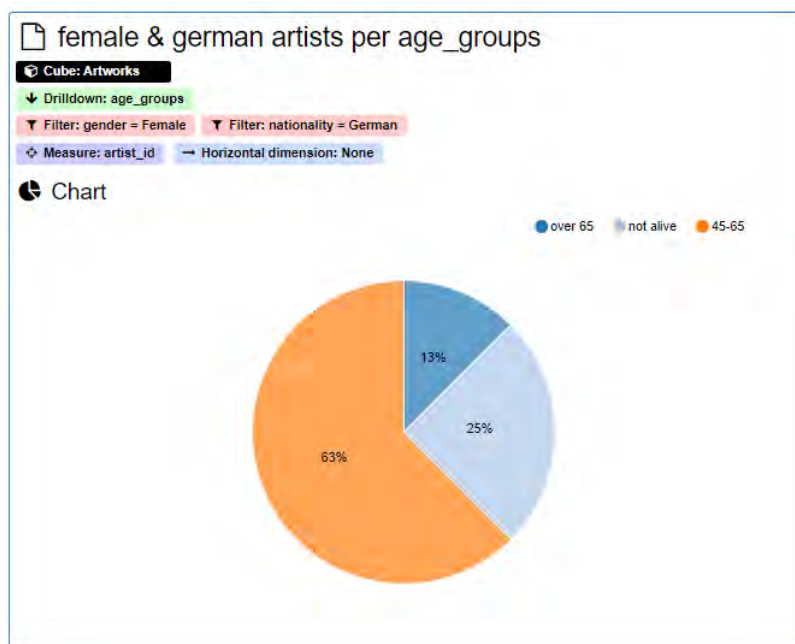


Id	Age Groups	Nationality	Gender	Department	Classification	Quarters	Artwork
0	not alive	German	Female	Photography	Photograph	quarter_2	130 912
1	45-65	German	Female	Prints & Illustration	Print	quarter_2	199 084
2	45-65	German	Female	Photography	Photograph	quarter_1	199 240
3	45-65	German	Female	Prints & Illustration	Print	quarter_1	199 613
4	45-65	German	Female	Prints & Illustration	Print	quarter_1	199 614
5	45-65	German	Female	Prints & Illustration	Print	quarter_1	199 615
6	45-65	German	Female	Prints & Illustration	Print	quarter_1	199 616
7	45-65	German	Female	Photography	Illustrated Book	quarter_2	199 938
8	over 65	German	Female	Film	Film	quarter_1	202 365
9	45-65	German	Female	Prints & Illustration	Print	quarter_2	203 430
10	45-65	German	Female	Prints & Illustration	Multiple	quarter_2	204 668
11	45-65	German	Female	Prints & Illustration	Print	quarter_2	204 669
12	not alive	German	Female	Photography	Photograph	quarter_4	205 818
13	not alive	German	Female	Photography	Photograph	quarter_4	205 824
14	not alive	German	Female	Photography	Photograph	quarter_4	205 825

Σχήμα 5.6: Απεικόνιση των facts data για γυναίκες καλλιτέχνες με γερμανική εθνικότητα.

**Παράδειγμα τρίτο:**

Επεκτείνοντας το προηγούμενο παράδειγμα, εκτός από τις δύο τομές που ορίσαμε πριν, προσθέτουμε μία αναλυτική κάθοδο στη διάσταση “age\_groups”. Άρα στο Σχήμα 5.7 που ακολουθεί απεικονίζονται οι γυναίκες καλλιτέχνες, γερμανικής καταγωγής ανά ηλικιακό γκρουπ (“age\_groups”) σε τύπο διαγράμματος “pie”(πίτα).



Σχήμα 5.7: Απεικόνιση γυναικών καλλιτεχνών με γερμανική εθνικότητα ανά ηλικιακό γκρουπ.

Κλείνοντας το παρόν Κεφάλαιο έχουν παρουσιαστεί οι τρόποι και οι δυνατότητες αναλυτικής επεξεργασίας και οπτικοποίησης των δεδομένων που χρησιμοποιούνται σε αυτή την εργασία. Παράλληλα έχει περιγραφεί διεξοδικά ο τρόπος ανάλυσης και επεξεργασίας δεδομένων, ώστε να δίνεται η δυνατότητα σε κάθε χρήστη των εργαλείων που χρησιμοποιούνται (Cubes Framework, CubesViewer), να επεξεργαστούν τα δεδομένα τους με τη μέθοδο OLAP και να εξάγουν την επιθυμητή πληροφορία από αυτά με την βοήθεια της οπτικοποίησης τους.




## Κεφάλαιο 6

# Εξαγωγή Πληροφορίας από την Ανάλυση Δεδομένων

Στο κεφάλαιο αυτό θα παρουσιαστεί η εξαγωγή πληροφορίας από τα δεδομένα έργων τέχνης που προήλθε απ'την αναλυτική επεξεργασία των δεδομένων αυτών και την οπτικοποίηση τους, όπως αναλύθηκε στο προηγούμενο κεφάλαιο. Αρχικά θα παρουσιαστούν γενικά συμπεράσματα που αφορούν στοιχεία όλου του κύβου και στη συνέχεια θα πραγματοποιηθεί εφαρμογή αναλυτικών καθόδων σε όλες τις διαστάσεις που έχουν οριστεί για τον κύβο.

### 6.1 Γενικά Συμπεράσματα

Ξεκινώντας, εάν απλά καλέσουμε τη συνάρτηση Aggregate για ολόκληρο τον κύβο δεδομένων, λαμβάνουμε ως αποτέλεσμα τον συνολικό αριθμό των έργων τέχνης που εκτέθηκαν στο Μομα κατά το έτος 2016, καθώς και τον συνολικό αριθμό καλλιτεχνών που σχετίζονται με τη δημιουργία των έργων αυτών. Στον πίνακα που ακολουθεί μπορούμε να διακρίνουμε ότι συνολικά στο μουσείο εκτέθηκαν 1873 έργα τέχνης, ενώ ο αριθμός των καλλιτεχνών στους οποίους ανήκουν τα έργα είναι συνολικά 495. Από τους παραπάνω αριθμούς είναι εμφανές ότι πολλά έργα προέρχονται από τους ίδιους καλλιτέχνες ή από σύμπραξη αυτών.



The screenshot shows a web application interface for 'aggregate'. It features a 'Cube: Artworks' dropdown menu and a section for 'Aggregated data'. Below this, there is a table with columns for 'Artworks', 'Artwork Id', and 'Artist Id'. The table contains a single row labeled 'Summary' with values 1 873.00 and 495.00. At the bottom, there are navigation controls including a page number '1', a total of '15' items per page, and a status '1 of 1 items'.

Artworks	Artwork Id	Artist Id
Summary	1 873.00	495.00

Σχήμα 6.1: Εφαρμογής της συνάρτησης Aggregate σε όλο τον κύβο.

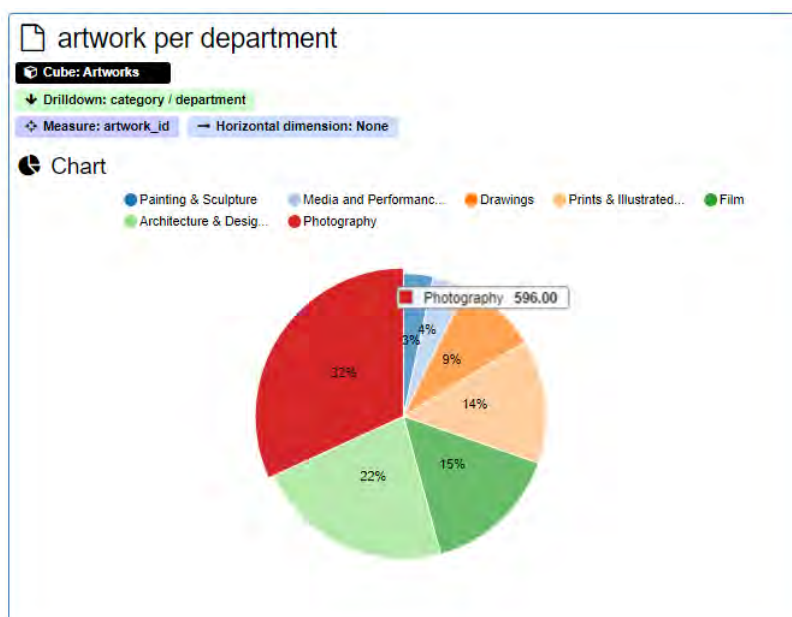
## 6.2 Εξαγωγή Πληροφορίας από τις Αναλυτικές Καθόδους

Εδώ περιγράφουμε τα συμπεράσματα που προκύπτουν από την εφαρμογή των αναλυτικών καθόδων (drilldown) σε κάθε μια από τις διαστάσεις του κύβου. Θα ακολουθήσουν αναλυτικά τα γραφήματα για κάθε μία απ' τις περιπτώσεις, καθώς και ο αντίστοιχος σχολιασμός της πληροφορίας που απορρέει από αυτά.

### 6.2.1 Αναλυτική Κάθοδος κατά Κατηγορία

Πραγματοποιώντας αναλυτική κάθοδο στη διάσταση “category”, και συγκεκριμένα στο πρώτο (default) επίπεδο της “department”, δημιουργούνται τα γραφήματα στα Σχήματα 6.2 και 6.3. Από το Σχήμα 6.2 είναι φανερό ότι τα περισσότερα αποκτήματα του μουσείου ανήκουν στην κατηγορία “Photography” με ποσοστό 32%. Οι αμέσως επόμενες κατηγορίες με μεγάλο ποσοστό έργων είναι οι “Architecture & Design” και “Film” με 22% και 15% αντίστοιχα. Ενδιαφέρον παρουσιάζει το γεγονός ότι η κατηγορία “Painting & Sculpture” έχει πολύ μικρό ποσοστό, μόλις 3%.

Από τα παραπάνω ποσοστά μπορούμε εύκολα να διακρίνουμε ότι ένα μουσείο μοντέρνας τέχνης δίνει βαρύτητα στην απόκτηση και έκθεση έργων τέχνης από πιο σύγχρονες εκφάνσεις της τέχνης όπως για π.χ. φωτογραφία, αρχιτεκτονική, design, έναντι πιο παλεών και κλασικών μορφών όπως η ζωγραφική και η γλυπτική. Ακόμη, έχοντας ως παράδειγμα ένα από τα μεγαλύτερα μουσεία μοντέρνας τέχνης στον κόσμο, μπορούμε να κατανοήσουμε καλύτερα τι είναι “Μοντέρνα Τέχνη”, αφού μπορούμε να δούμε ξεκάθαρα από την παραπάνω ανάλυση ποιες μορφές τέχνης αποκαλούνται “μοντέρνες” (είναι αυτές με τα μεγαλύτερα ποσοστά) και ποιες αποτελούν πιο κλασικές μορφές.

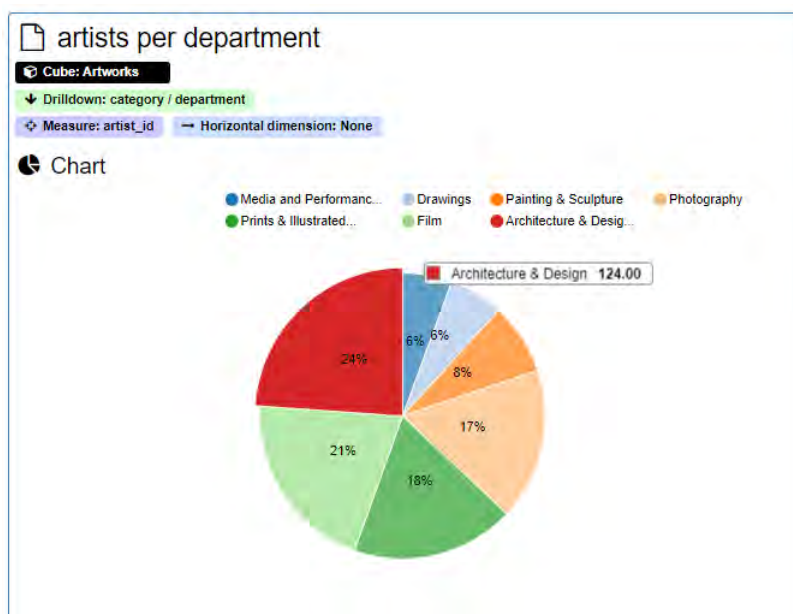


Σχήμα 6.2: Ποσοστό έργων τέχνης ανά κατηγορία.



Στο Σχήμα 6.3 φαίνεται η εφαρμογή της αναλυτικής καθόδου στη διάσταση “department”, αυτή τη φορά όμως με μέτρο τον συνολικό αριθμό των καλλιτεχνών ανά κατηγορία έργων τέχνης. Όπως μπορούμε να δούμε, το μεγαλύτερο ποσοστό καλλιτεχνών προέρχεται από το χώρο της αρχιτεκτονικής και του design με ποσοστό 24%, ενώ το αμέσως επόμενο ποσοστό είναι των καλλιτεχνών από τον χώρο των ταινιών (films). Από τα παραπάνω ποσοστά καταλαβαίνουμε ότι τα έργα τέχνης που ανήκουν στο τμήμα της φωτογραφίας, τα οποία είναι και τα περισσότερα, έχουν προέλθει σε ένα μεγάλο ποσοστό από τον ίδιο καλλιτέχνη ή σύμπραξη καλλιτεχνών, ή ακόμα και από έκθεση κάποιου καλλιτέχνη.

Σίγουρα και από τα δύο Σχήματα προκύπτει ότι οι τομείς της Αρχιτεκτονικής & Design αλλά και των Films αποτελούν κυρίαρχο τμήμα ενός μουσείου μοντέρνας τέχνης. Επιπροσθέτως φαίνεται πως για τους δημιουργούς έργων των παραπάνω τομέων υπάρχει ένα ευρύ πεδίο εργασίας σε τέτοιου είδους οργανισμούς.

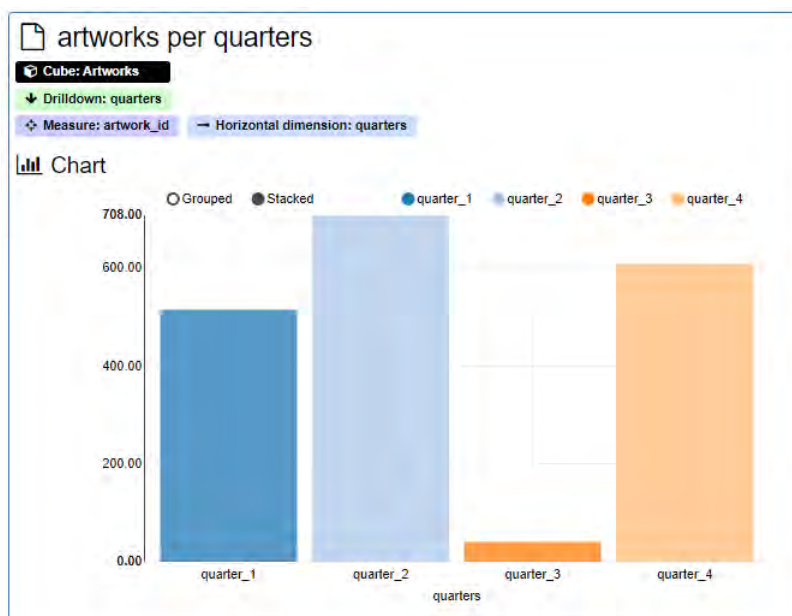


Σχήμα 6.3: Ποσοστό καλλιτεχνών ανά κατηγορία.

### 6.2.2 Αναλυτική Κάθοδος κατά Τρίμηνα του Έτους

Η πληροφορία που ανακτάται από την εφαρμογή της αναλυτικής καθόδου στη διάσταση “quarters”, μπορεί να φανεί ιδιαίτερα χρήσιμη για τη λήψη αποφάσεων και τη διαδικασία διοικήσεως ενός οργανισμού ή μιας επιχείρησης. Ουσιαστικά παρουσιάζει την κινητικότητα του μουσείου σε σχέση με κάποια ορισμένα χρονικά διαστήματα, και αν συνδιαστεί με οικονομικά στοιχεία του μουσείου, οδηγεί σε συμπεράσματα καθοριστικής σημασίας για την οικονομική διαχείριση της επιχείρησης.

Το συγκεκριμένο σετ δεδομένων που χρησιμοποιήθηκε για την εργασία αυτή δεν περιείχε οικονομικά στοιχεία του μουσείου ώστε να συνδυαστούν π.χ. με τα έσοδα και έξοδα που έχει το μουσείο ανά συγκεκριμένα διαστήματα του χρόνου. Εν τούτοις στο Σχήμα 6.2 φαίνεται αναλυτικά ο αριθμός των εκθεμάτων που απέκτησε το μουσείο ανά τρίμηνο του έτους.

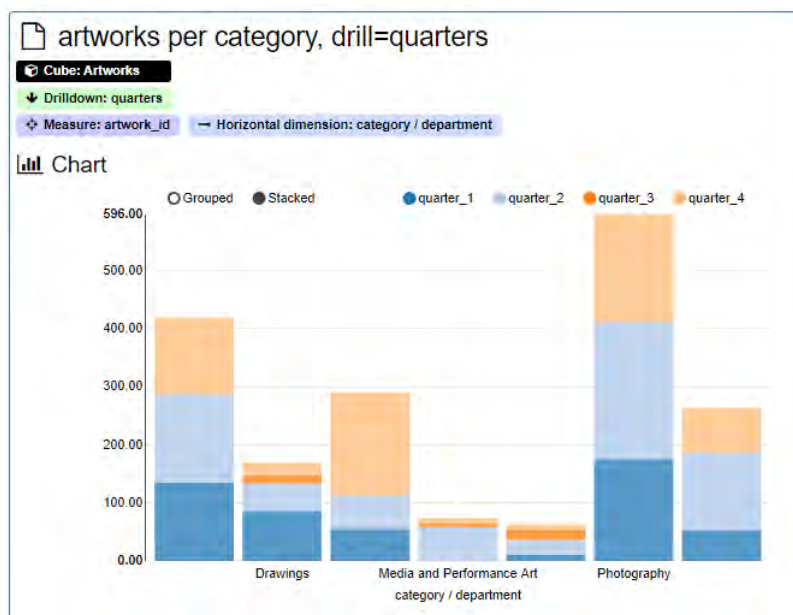


Σχήμα 6.4: Συνολικός αριθμός έργων ανά τρίμηνο.

Παρατηρείται έντονα ότι τα περισσότερα έργα τέχνης αποκτήθηκαν απ' το μουσείο κατά το δεύτερο τρίμηνο του έτους, δηλαδή τους μήνες από τον Απρίλιο έως και τον Ιούνιο, ενώ ιδιαίτερα μικρός είναι ο αριθμός των έργων το τρίτο τρίμηνο (Ιούλιος - Σεπτέμβριος) με μόλις 41 έργα. Τα υπόλοιπα δύο τρίμηνα έχουν επίσης μεγάλο αριθμό αποκτηθέντων έργων.

Μία πιο σύνθετη πληροφορία μπορεί να προκύψει από την εφαρμογή αναλυτικής καθόδου στη διάσταση "quarters", όπως στο Σχήμα 6.2, παράλληλα με την επιλογή της διάστασης "department" ως οριζόντιας διάστασης στο διάγραμμα. Το αποτέλεσμα που προκύπτει φαίνεται στο Σχήμα 6.5 που ακολουθεί και απεικονίζει τον αριθμό έργων καθε κατηγορίας που αποκτήθηκαν από το μουσείο ανα τρίμηνο.

Κατά αυτό τον τρόπο μπορούμε εξάγουμε πληροφορία σχετικά με το πόσα έργα της κάθε κατηγορίας αποκτήθηκαν ανά τρίμηνο. Ουσιαστικά το μουσείο και οι υπεύθυνοι διοικήσεως του έχουν την δυνατότητα να ελέγχουν από ποιον τομέα προέρχονται και πόσα ακριβώς είναι τα έργα τέχνης που αποκτούν ανά συγκεκριμένα χρονικά διαστήματα. Μία τέτοια πληροφορία μπορεί να φανεί χρήσιμη στη λήψη αποφάσεων.



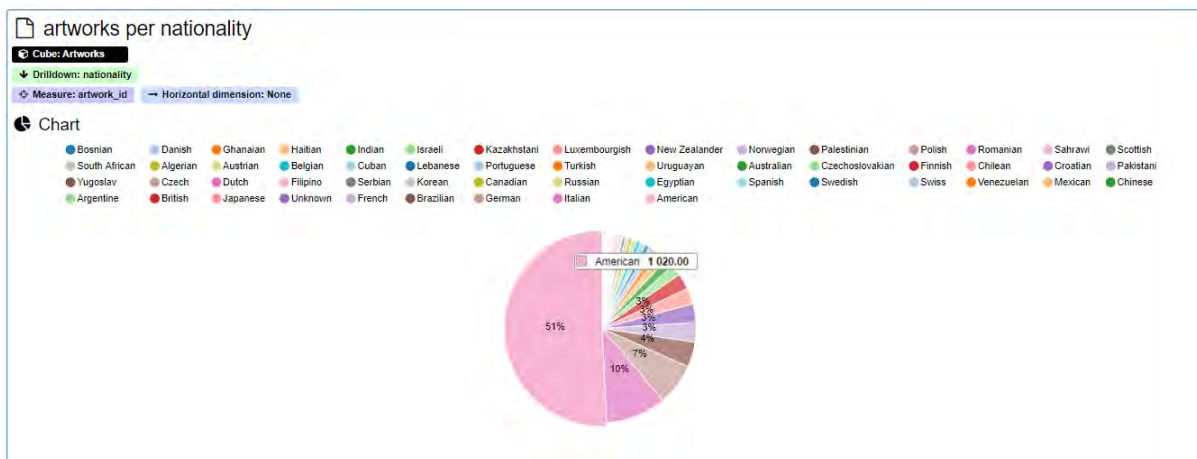
Σχήμα 6.5: Συνολικός αριθμός έργων ανά κατηγορία κατά τρίμηνο.

Συγκεκριμένα από το παραπάνω διάγραμμα, μιλώντας μόνο για τα τρία τρίμηνα που αποκτήθηκαν έργα (μη υπολογίζοντας το τρίτο τρίμηνο που ο αριθμός των έργων είναι σχεδόν αμελητέος σε σύγκριση με τα υπόλοιπα), ο αριθμός των έργων που αποκτήθηκαν από τους τομείς “Photography” και “Architecture & Design” ανά τρίμηνο είναι ισομοιρασμένος. Για τα έργα της κατηγορίας “Film” παρατηρούμε πως στο τέταρτο τρίμηνο αποκτήθηκαν περίπου τα τριπλάσια απ’ ότι στα υπόλοιπα δύο τρίμηνα.

### 6.2.3 Αναλυτική Κάθοδος κατά Εθνικότητα

Εφαρμόζοντας αναλυτική κάθοδο στην διάσταση “nationality” με μέτρο το συνολικό αριθμό έργων τέχνης μπορούμε να ανακτήσουμε πληροφορία που σχετίζεται με τον αριθμό έργων κατά εθνικότητα καλλιτέχνη. Από το Σχήμα 6.6 φαίνεται καθαρά ότι πάνω από τα μισά έργα (ποσοστό 51%) που αποκτήθηκαν απ’ το μουσείο προέρχονται από καλλιτέχνες Αμερικανικής Εθνικότητας. Με μεγάλη διαφορά ακολουθούν ακολουθούν καλλιτέχνες ιταλικής εθνικότητας, σε ποσοστό μόλις 11%, καθώς και γερμανικής, σε ποσοστό 7%. Αξίζει να σημειωθούν τα πολύ μικρά ποσοστά των έργων που προέρχονται απο τις χώρες Δανία, Πολωνία, Σκωτία, Ρουμανία, Ισραήλ, Ινδία, κ.α., με μόλις ένα έργο ανά χώρα.

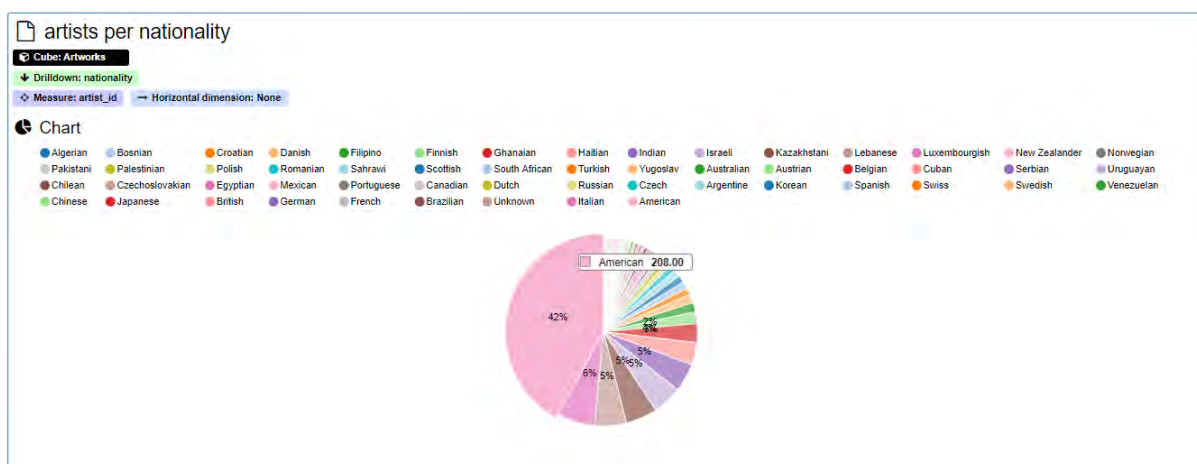
Είναι εμφανές το γεγονός ότι το μουσείο στηρίζεται ιδιαίτερα σε έργα που προέρχονται από Αμερικανούς καλλιτέχνες, κάτι το οποίο δικαιολογείται αφού η έδρα του είναι στη Νέα Υόρκη.



Σχήμα 6.6: Ποσοστό έργων τέχνης ανά εθνικότητα.

Στη συνέχεια θα μελετηθεί η εφαρμογή της αναλυτικής καθόδου στην διάσταση “nationality” με μέτρο τον αριθμό καλλιτεχνών. Στο Σχήμα 6.7 φαίνεται ότι το μεγαλύτερο ποσοστό του συνολικού αριθμού καλλιτεχνών, 42%, είναι Αμερικανικής εθνικότητας. Ακολουθούν με αισθητή διαφορά καλλιτέχνες από την Ιταλία (6%) αλλά και από τις Γερμανία, Γαλλία, Βραζιλία (5%) .

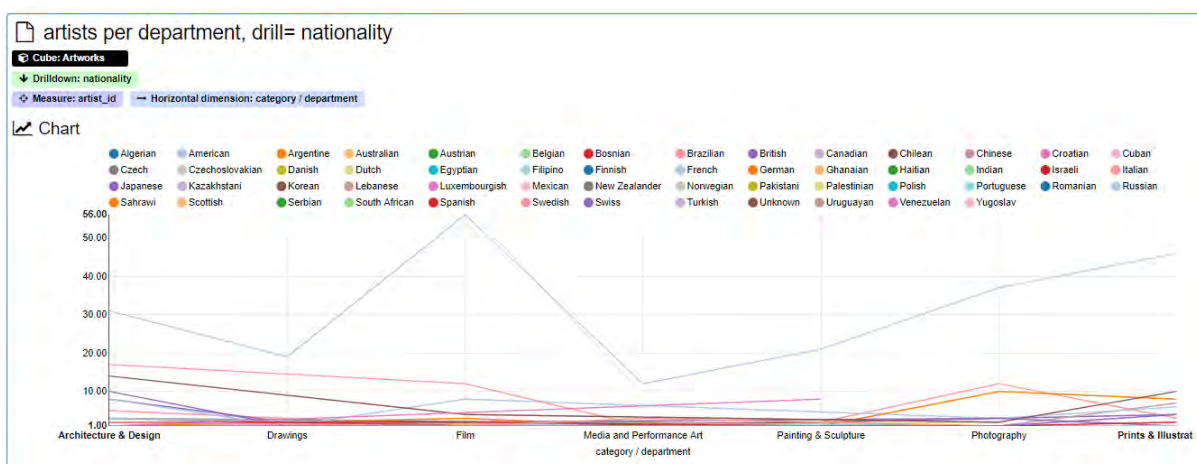
Αξίζει να αναφερθεί ότι ο αριθμός των έργων τέχνης που προέρχονται από Αμερικανούς είναι 1020 έργα συνολικά και αντιστοιχεί σε 208 καλλιτέχνες. Ο μικρότερος αριθμός καλλιτεχνών σε σχέση τα έργα δικαιολογείται αφού πολλά από τα παραπάνω έργα προέρχονται απο εκθέσεις. Σε αυτό το σημείο συμπεραίνουμε πως το μουσείο προτιμά εκθέσεις κυρίως Αμερικανών καλλιτεχνών. Σε πολύ μικρότερη κλίμακα παρατηρείται το ίδιο φαινόμενο και για καλλιτέχνες από την Ιταλία. Για όλες τις υπόλοιπες χώρες που εμφανίζονται, φαίνεται πως το ποσοστό των έργων είναι αρκετά χαμηλό.



Σχήμα 6.7: Ποσοστό καλλιτεχνών ανά εθνικότητα.

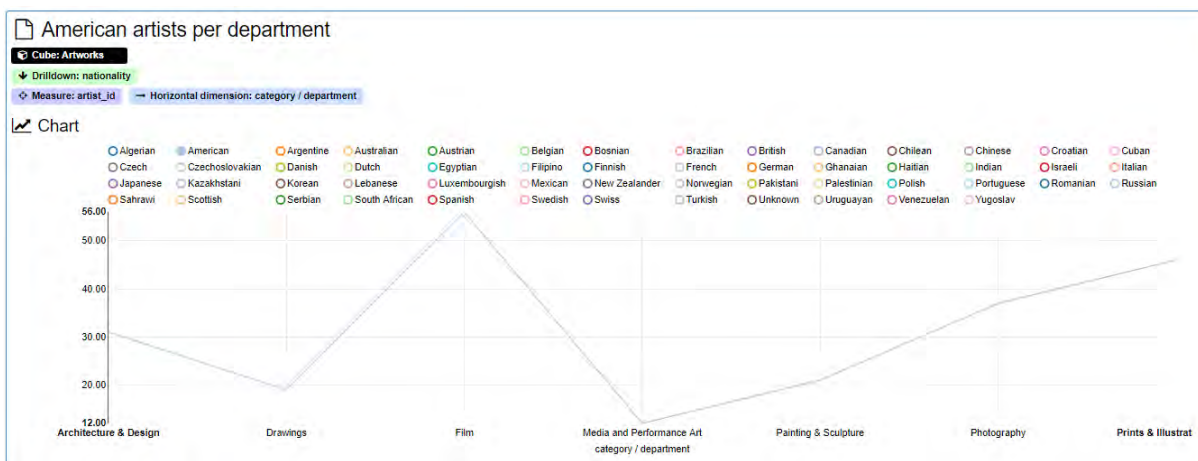
Ιδιαίτερο ενδιαφέρον παρουσιάζει το προηγούμενο παράδειγμα προσθέτοντας ως οριζόντια διάσταση την “department”, ώστε να μελετηθεί πόσοι καλλιτέχνες ανά εθνικότητα ανήκουν σε κάθε κατηγορία έργων τέχνης. Κατ’ αυτό τον τρόπο θα μπορούσαμε να συνδέσουμε την εθνικότητα των καλλιτεχνών με τις διαφορές μορφές τέχνης και να βγάλουμε συμπεράσματα που μπορούν να φανούν χρήσιμα σε πολιτιστικό επίπεδο, όπως για παράδειγμα σε ένα Υπουργείο Πολιτισμού.

Πιο συγκεκριμένα, από τα Σχήματα 6.8 και 6.9 είναι εμφανές ότι ο μεγαλύτερος αριθμός Αμερικανών καλλιτεχνών συγκεντρώνεται στον τομέα των έργων ενώ σε μεγάλα ποσοστά και στους τομείς “Photography” και “Prints & Illustrated Books”.

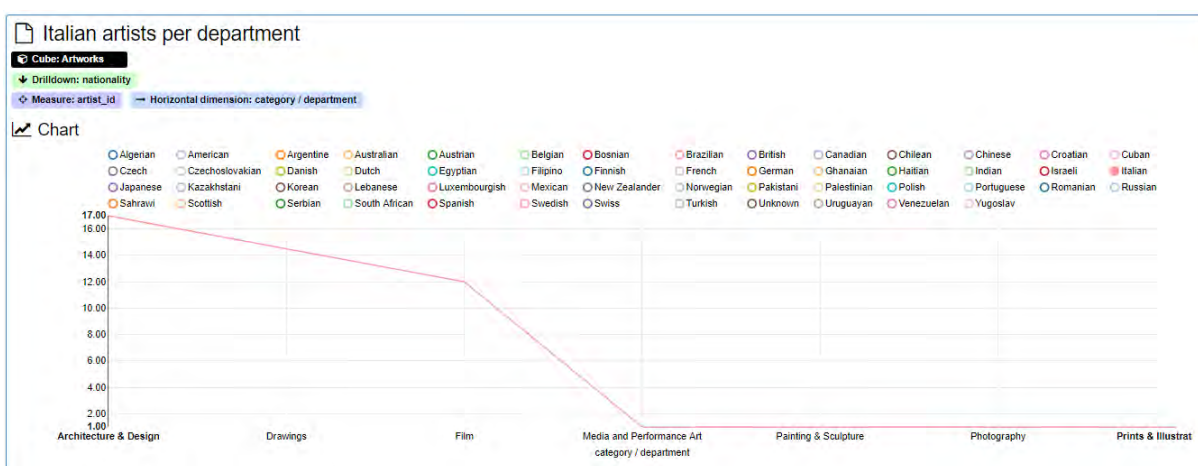


Σχήμα 6.8: Καλλιτέχνες κατά εθνικότητα και ανά κατηγορία.

Απομονώνοντας και τους Ιταλούς καλλιτέχνες (οι οποίοι κατέχουν το δεύτερο μεγαλύτερο ποσοστό σε αριθμό καλλιτεχνών), όπως φαίνεται και στο Σχήμα 6.10, παρατηρείται ότι οι περισσότεροι Ιταλοί καλλιτέχνες προέρχονται από τον χώρο “Architecture & Design και μετέπειτα από τους χώρους των “Drawings” και “Film”. Δεν παρατηρείται σημαντικός αριθμός στις υπόλοιπες κατηγορίες.



Σχήμα 6.9: Αριθμός Αμερικανών καλλιτεχνών ανά κατηγορία.



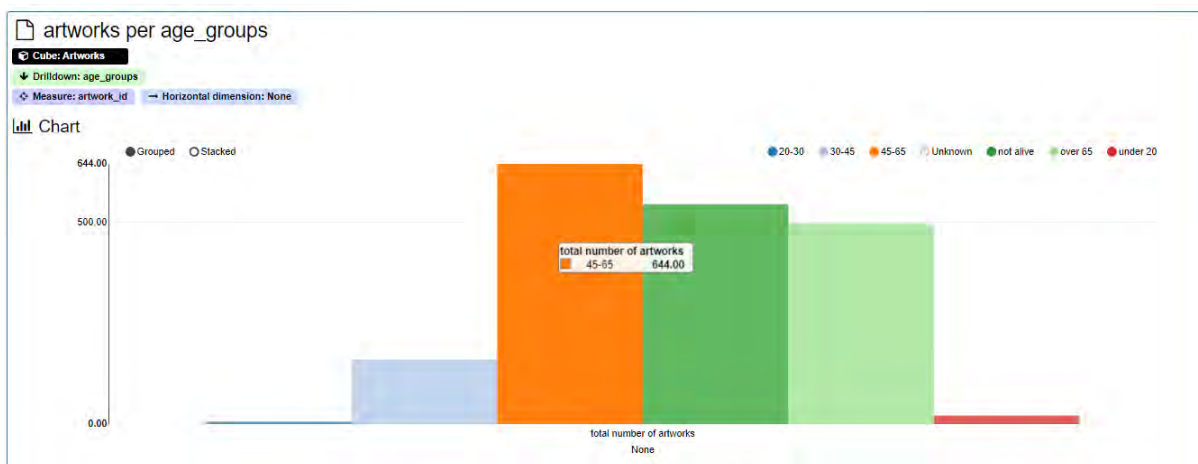
Σχήμα 6.10: Αριθμός Ιταλών καλλιτεχνών ανά κατηγορία.

Τέλος, αν κοιτάξουμε το Σχήμα 6.8 από μια γενικότερη σκοπιά, μπορούμε να σημειώσουμε τα εξής βασικά συμπεράσματα:

- Στην κατηγορία “Architecture & Design” συναντώνται, εκτός από Αμερικανούς καλλιτέχνες, αρκετοί Ιταλοί και Γιαπωνέζοι.
- Στην κατηγορία “Film” συναντώνται στο μεγαλύτερο ποσοστό Αμερικανοί καλλιτέχνες και στη συνέχεια, με μεγάλη διαφορά Ιταλοί.
- Στην κατηγορία “Photography” συναντώνται εκτός από Αμερικανούς καλλιτέχνες, αρκετοί Βραζιλιάνοι και Γερμανοί.
- Στην κατηγορία “Printed & Illustrated Books” το κυριότερο ποσοστό προέρχεται από Αμερικανούς καλλιτέχνες.

### 6.2.4 Αναλυτική Κάθοδος κατά Ηλικιακό γκρούπ

Από την εφαρμογή της αναλυτικής καθόδου στη διάσταση “age\_groups”, με μέτρο το συνολικό αριθμό έργων τέχνης, προκύπτει το διάγραμμα στο Σχήμα 6.11. Από το διάγραμμα γίνεται εμφανές ότι τα περισσότερα έργα έχουν προέλθει από καλλιτέχνες που ανήκουν στο ηλικιακό γκρούπ των “45-65”, ενώ ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός ότι το αμέσως επόμενο ηλικιακό γκρούπ με τα περισσότερα έργα είναι αυτό των καλλιτεχνών οι οποίοι δεν βρίσκονται πια εν ζωή (“not alive”). Σημαντικό επίσης ποσοστό έργων κατέχει και το ηλικιακό γκρούπ “over 65”. Τα ποσοστά όλων των υπολοίπων ηλικιακών groups που αντιστοιχούν στις ηλικίες κάτω των 45 ετών, συγκεντρώνουν μαζί ποσοστό χαμηλότερο του 20%.



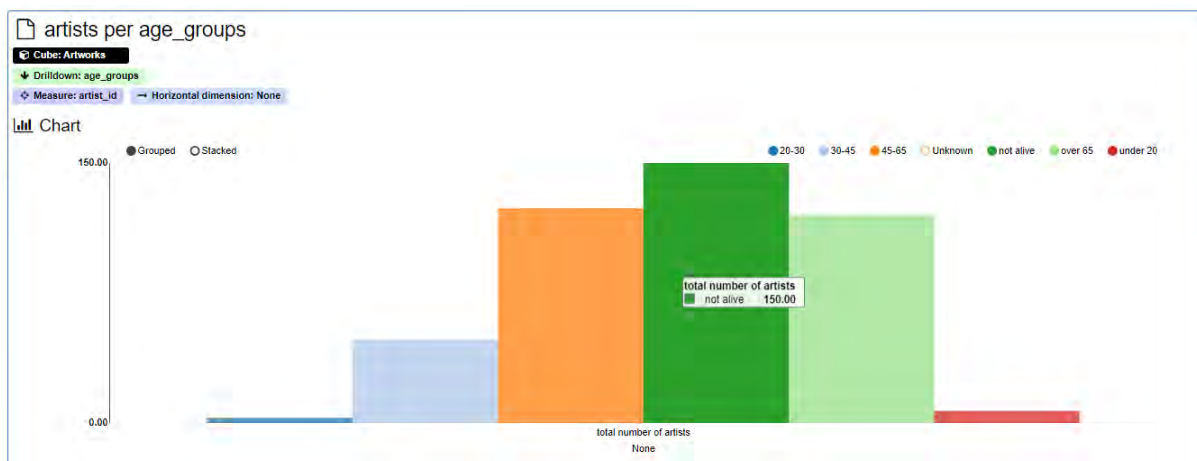
Σχήμα 6.11: Αριθμός έργων τέχνης ανά ηλικιακό γκρούπ.

Σε συνέχεια της εφαρμογής της αναλυτικής καθόδου στη διάσταση “age\_groups” με μέτρο, αυτή τη φορά, το συνολικό αριθμό καλλιτεχνών, το διάγραμμα που προκύπτει φαίνεται στο Σχήμα 6.12. Από την ανάλυση του διαγράμματος μπορούμε να διακρίνουμε ότι ο μεγαλύτερος αριθμός καλλιτεχνών αντιστοιχεί στους καλλιτέχνες που δεν βρίσκονται εν ζωή. Τα αμέσως επόμενα groups των “45-65” και “over 65” κατέχουν ένα πολύ μεγάλο ποσοστό καλλιτεχνών, ενώ για ακόμα μία φορά τα υπόλοιπα groups δεν παρουσιάζουν ιδιαίτερα μεγάλα ποσοστά.

Ενδιαφέρον κρίνεται το γεγονός ότι, ενώ τα περισσότερα έργα αντιστοιχούν στο ηλικιακό γκρούπ “45-65”, οι περισσότεροι καλλιτέχνες ανήκουν στην κατηγορία “not alive”. Από αυτή την πληροφορία καταλαβαίνουμε ότι ενώ τα έργα τέχνης καλλιτεχνών που δεν βρίσκονται εν ζωή κατέχουν τεράστια θέση στο μουσείο, οι εκθέσεις που φιλοξενεί το μουσείο προέρχονται κυρίως από καλλιτέχνες των ηλικιών “45-65”. Αυτό είναι πολύ φυσιολογικό διότι εκθέσεις πραγματοποιούν οι καλλιτέχνες που είναι εν ενεργεία και βρίσκονται σε μία δημιουργική τους περίοδο.

Σε αυτό το σημείο αξίζει να σχολιαστεί ότι, σε ένα από τα μεγαλύτερα μουσεία Μοντέρνας τέχνης, τα έργα που προέρχονται από καλλιτέχνες που δεν βρίσκονται εν ζωή κατέχουν το δεύτερο μεγαλύτερο ποσοστό. Ουσιαστικά αποτελούν κυρίαρχο κομμάτι του πυρήνα του μουσείου. Κάτι

τέτοιο μας οδηγεί στο συμπέρασμα ότι μοντέρνο έργο δεν αποκαλείται μόνο ένα έργο το οποίο δημιουργείται τώρα, αλλά μπορεί να είναι κάτι που έχει δημιουργηθεί εδώ και πολλά χρόνια, και λόγω της αξίας και της ποιότητας του είναι διαχρονικό. Γι αυτό και έργα παλαιότερων χρόνων βρίσκουν θέση σε ένα μουσείο “μοντέρνας τέχνης”.

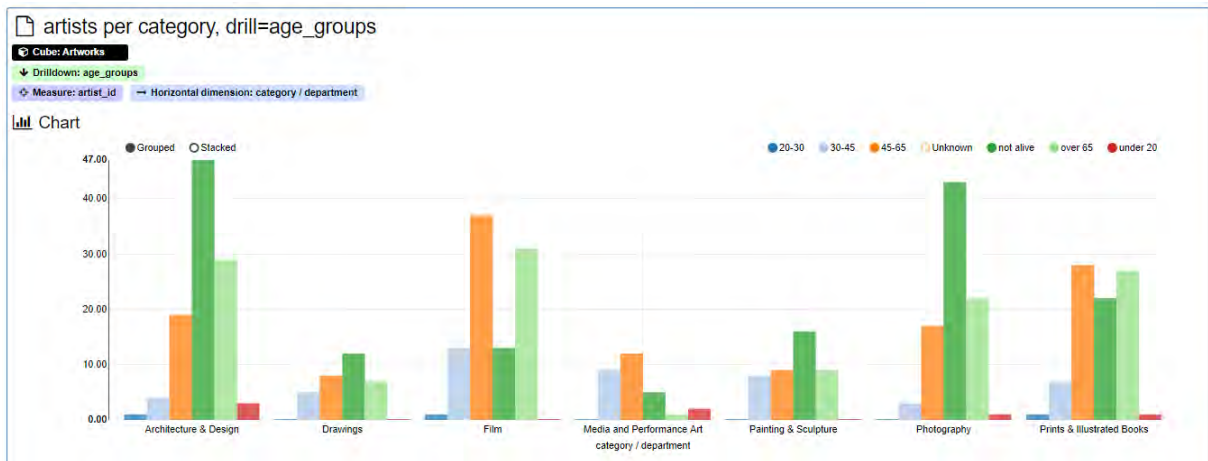


Σχήμα 6.12: Αριθμός καλλιτεχνών ανά ηλικιακό γκρούπ.

Εάν στο προηγούμενο παράδειγμα που αναλύθηκε προστεθεί ως οριζόντια διάσταση η “department” μπορεί να υπάρξει συσχετισμός μεταξύ της ηλικίας των καλλιτεχνών και των κατηγοριών των έργων. Στο Σχήμα 6.13 που ακολουθεί είναι φανερό πως έργα της κατηγορίας “Architecture & Design” έχουν δημιουργηθεί κατά κύριο λόγο από καλλιτέχνες που δεν είναι εν ζωή και κατα δεύτερο λόγο από καλλιτέχνες άνω των 65 ετών. Τα ίδια περίπου, με μικρές διαφορές στα μεγέθη, ισχύουν και για τα έργα της κατηγορίας “Photography”. Οι δημιουργοί των έργων της κατηγορίας “Film” προέρχονται κυρίως από το ηλικιακά groups των “45-65” και “over 65”, ενώ της κατηγορίας “Media and Performance Art” από τα ηλικιακά groups των “45-65” και “30-45”.

Όπως φαίνεται και από τα παραπάνω, μικρότερες ηλικίες καλλιτεχνών παρουσιάζονται σε νεότερες και πιο σύγχρονες μορφές τέχνης, όπως για παράδειγμα στην κατηγορία “Media and Performance Art”. Στις πιο παλιές μορφές τέχνης, όπως η κατηγορία “Architecture”, κυριαρχούν μεγαλύτερης ηλικίας καλλιτέχνες ή ακόμα και αποθανόντες.

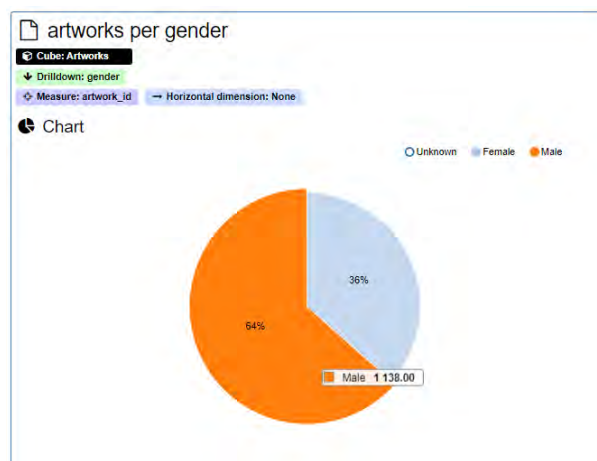




Σχήμα 6.13: Αριθμός καλλιτεχνών ανά κατηγορία, κατά ηλικιακό γκρούπ.

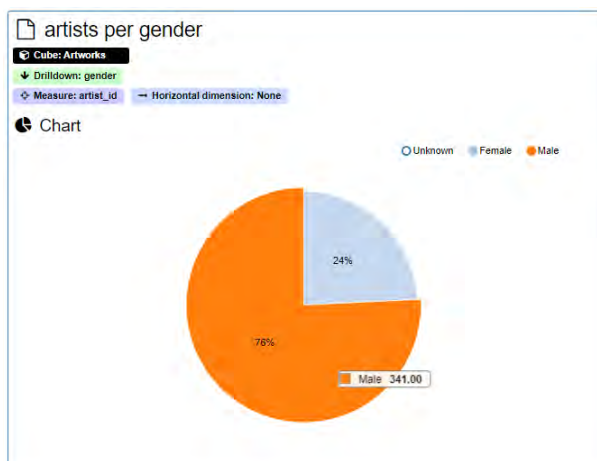
### 6.2.5 Αναλυτική Κάθοδος κατά Φύλο

Για την ανάκτηση περισσότερων πληροφοριών σχετικά με τους καλλιτέχνες και το προφίλ τους, ιδανική είναι η εφαρμογή της αναλυτικής καθόδου στη διάσταση “gender”. Όπως παρατηρείται ξεκάθαρα στο Σχήμα 6.14, εάν αφαιρέσουμε το ποσοστό των καλλιτεχνών με φύλο “Unknown”, το 64% του συνολικού αριθμού έργων προέρχεται από άντρες καλλιτέχνες, ενώ το 36% από γυναίκες. Παραπάνω από τα μισά έργα του μουσείου ανήκουν σε άντρες καλλιτέχνες.



Σχήμα 6.14: Ποσοστό έργων τέχνης ανά φύλο.

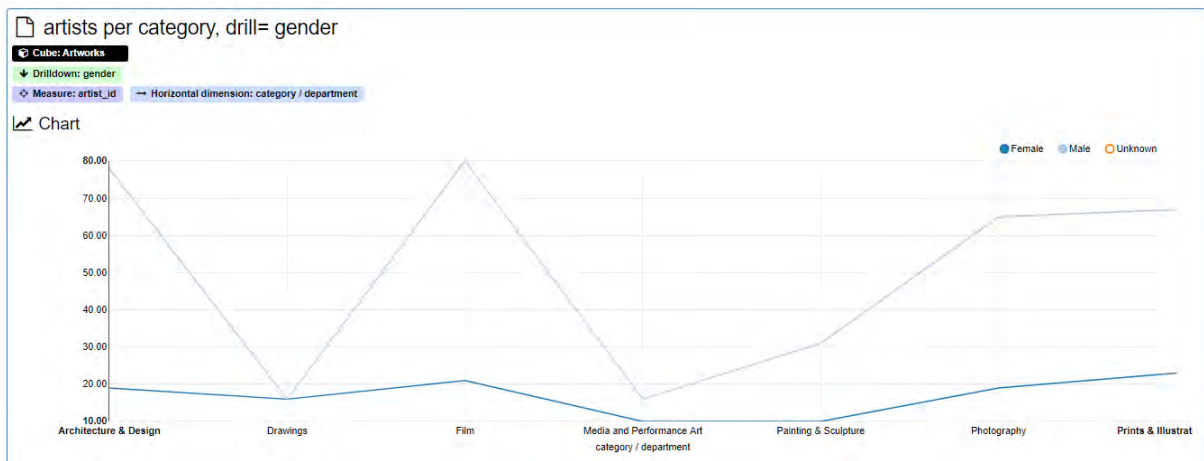
Εν συνεχεία, στο Σχήμα 6.15 ακολουθεί το ποσοστό των καλλιτεχνών ανά φύλο. Παρατηρούμε ότι το ποσοστό του συνολικού αριθμού των ανδρών καλλιτεχνών είναι 76%, ενώ των γυναικών 24%. Ο αριθμός των ανδρών έναντι των γυναικών είναι σχεδόν τριπλάσιος.



Σχήμα 6.15: Ποσοστό καλλιτεχνών ανά φύλο.

Ακόμα πιο σύνθετη πληροφορία προκύπτει αν στο προηγούμενο παράδειγμα προσθέσουμε ως οριζόντια διάσταση την “department”. Από το αντίστοιχο Σχήμα 6.16 μπορούμε να διακρίνουμε ότι τα υψηλότερα ποσοστά αντρών συναντώνται στις κατηγορίες “Architecture & Design”, “Film” και “Photography”, ενώ ιδιαίτερα χαμηλά και αρκετά κοντινά με τα ποσοστά των γυναικών είναι στις κατηγορίες “Drawings” και “Media and Performance Art”.

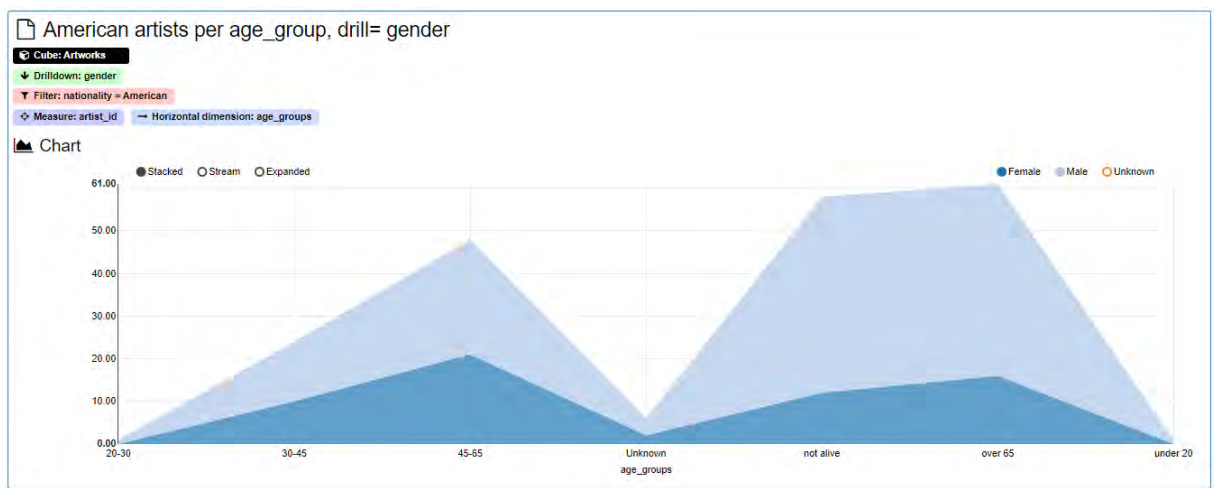
Τα μεγαλύτερα ποσοστά γυναικών συναντώνται στις κατηγορίες “Print and Illustrated Books”, “Film” και “Architecture & Design”. Σε καμία κατηγορία έργων τέχνης τα ποσοστά των γυναικών δεν ξεπερνάνε αυτά των ανδρών, εκτός από την κατηγορία “Drawings” που τα ποσοστά είναι ίδια. Σαφώς και αυτά τα συμπεράσματα είναι λογικά εφόσον ο συνολικός αριθμός των αντρών είναι πολύ μεγαλύτερος από αυτόν των γυναικών. Μπορεί όμως κάποιος να οδηγηθεί σε γενικά συμπεράσματα που έχουν σχέση για παράδειγμα, με ποιού είδους τέχνη τείνουν να ασχολούνται οι καλλιτέχνες με βάση το φύλο τους.



Σχήμα 6.16: Ποσοστό καλλιτεχνών κατά φύλο και ανά κατηγορία.

Τέλος, μπορούμε να μελετήσουμε συγκεκριμένα το γενικότερο “προφίλ” των Αμερικάνων καλλιτεχνών, “κοιτάζοντας” συνδυαστικά τα χαρακτηριστικά τους ως προς το φύλο σε σχέση με την ηλικία. Τέτοια στοιχεία είναι χρήσιμα για “cultural analysis” και πιθανώς να ενδιαφέρουν Υπουργεία Πολιτισμού και άλλους οργανισμούς που ασχολούνται με πολιτιστικά θέματα.

Για το Σχήμα 6.17 που ακολουθεί έχει εφαρμοστεί τομή στη διάσταση “nationality” ώστε να απομονώσουμε τους Αμερικάνους καλλιτέχνες και αναλυτική κάθοδος στη διάσταση “gender”. Ως οριζόντια διάσταση έχει οριστεί η “age\_groups”. Εστιάζουμε στους Αμερικάνους καλλιτέχνες διότι αποτελούν το μεγαλύτερο, με διαφορά, ποσοστό καλλιτεχνών του μουσείου.



Σχήμα 6.17: Αριθμός Αμερικάνων καλλιτεχνών κατά φύλο και ανά ηλικιακά groups.

Από το διάγραμμα μπορούμε να παρατηρήσουμε ότι στο ηλικιακό γκρούπ “30-45” ο αριθμός των αντρών καλλιτεχνών είναι περίπου ίσος με των γυναικών, ενώ και στο γκρούπ “45-65” οι αριθμοί είναι πολύ κοντά. Όσο αυξάνεται η ηλικία, στα ηλικιακά groups “over 65” και “not alive”, ο αριθμός των αντρών καλλιτεχνών γίνεται τρεις φορές μεγαλύτερος από αυτόν των γυναικών.

# Κεφάλαιο 7

## Επίλογος

### 7.1 Σύνοψη και συμπεράσματα

Σκοπός της παρούσας διπλωματικής εργασίας ήταν η εξαγωγή πληροφορίας από δεδομένα έργων τέχνης με την χρήση αναλυτικών μεθόδων επεξεργασίας. Για την υλοποίηση της αναλυτικής επεξεργασίας χρησιμοποιήθηκε το εργαλείο Cubes, της προγραμματιστικής γλώσσας Python. Στο Κεφάλαιο 3 αναφέρθηκαν με λεπτομέρεια η βασική αρχιτεκτονική του εργαλείου και ο τρόπος χρήσης του.

Το σετ δεδομένων που χρησιμοποιήθηκε στη παρούσα εργασία, παρουσιάζεται στο Κεφάλαιο 4. Η διαθέσιμη αυτή βάση δεδομένων του μουσείου Moma, ενός απ' τα μεγαλύτερα μουσεία του κόσμου, επιλέχθηκε λόγω της μεγάλης ποικιλίας της πληροφορίας που μπορεί να εξαχθεί από αυτή. Στη συνέχεια του κεφαλαίου αναλύεται η βάση αυτής ως υπερκύβος με πέντε διαστάσεις και δύο μέτρα, ώστε να προκύπτουν σύνθετα ερωτήματα τόσο για τα έργα τέχνης όσο και για τους καλλιτέχνες.

Το εργαλείο CubesViewer χρησιμοποιήθηκε για την επεξεργασία των δεδομένων από τον χρήστη αλλά και την οπτικοποίηση των αποτελεσμάτων σε γραφήματα. Οι ποικίλες δυνατότητες του παραπάνω εργαλείου στην εξαγωγή διάφορων τύπων γραφημάτων συντέλεσαν στην καλύτερη διεξαγωγή συμπερασμάτων. Ο εύχρηστος τρόπος λειτουργίας του CubesViewer παρουσιάζεται στο Κεφάλαιο 5.

Την ερμηνεία των βασικών συμπερασμάτων που προέκυψαν από την αναλυτική επεξεργασία των δεδομένων καλύπτει το Κεφάλαιο 6. Έχουν παρουσιαστεί γενικά συμπεράσματα αλλά και πιο ειδικά, που προέκυψαν από την εφαρμογή των αναλυτικών καθόδων σε όλες τις διαστάσεις που ορίστηκαν. Η εξαγωγή της πληροφορίας καθορίζεται από την επιθυμία του χρήστη και μπορεί να διαφέρει ανάλογα με το μοντέλο που έχει οριστεί. Οι τρόποι για την εξαγωγή της έχουν παρουσιαστεί εκτενώς.

## 7.2 Μελλοντικές επεκτάσεις

Ως μελλοντική επέκταση της παρούσας διπλωματικής εργασίας, θα μπορούσε να θεωρηθεί η αύξηση του μεγέθους των δεδομένων του μουσείου που χρησιμοποιούνται. Θα μπορούσαν να προστεθούν δεδομένα και από άλλα έτη λειτουργίας του μουσείου, ώστε να προκύψουν ακόμη πιο σημαντικά και μεγαλύτερου εύρους συμπεράσματα. Ακόμη μια ενδιαφέρουσα ιδέα θα ήταν να δημιουργηθούν και άλλοι κύβοι εκτός αυτού που χρησιμοποιείται, με δεδομένα από σημαντικά μουσεία μοντέρνας τέχνης, της εμβέλειας του MoMa. Σε αυτή τη περίπτωση θα μπορούσε να γίνει σύγκριση ανάμεσα σε διαφορετικά μουσεία αλλά και καλλιτέχνες. Οι πληροφορίες που θα προέκυπταν από τέτοιου είδους εκτενή ανάλυση θα οδηγούσαν στην διεξαγωγή περίπλοκων και σύνθετων συμπερασμάτων.

Τέλος, ενδιαφέρον θα είχε εάν συνδυάζονταν τα υπάρχοντα δεδομένα με οικονομικά στοιχεία του μουσείου καθώς και με στοιχεία επισκεψιμότητας (αριθμός επισκεπτών, φύλο κτλ), εφόσον είναι εφικτό να βρεθούν τα δεδομένα αυτά. Κατ' αυτόν τον τρόπο επεκτείνεται η διεξαγωγή συμπερασμάτων και σε οικονομικό επίπεδο, κάτι το οποίο αποδεικνύεται ιδιαίτερα χρήσιμο για την διαδικασία λήψης αποφάσεων ενός οργανισμού.

# Βιβλιογραφία

- [1] Commission proposes ambitious New European Agenda for Culture. [https://ec.europa.eu/culture/news/new-european-agenda-culture\\_en](https://ec.europa.eu/culture/news/new-european-agenda-culture_en). Ημερομηνία πρόσβασης: 28-09-2019.
- [2] What is a decision support system? <https://waterpartnership.org.au/what-is-a-decision-support-system-dss/>. Ημερομηνία πρόσβασης: 28-09-2019.
- [3] H. Schütze C.D. Manning και P. Raghavan. *Introduction to information retrieval*. 2008.
- [4] R. Elmarsi και S.B.Navathe. *Θεμελιώδεις Αρχές Συστημάτων Βάσεων Δεδομένων*. Εκδόσεις Διάλογος, Αθήνα, 2016.
- [5] European Commision eurostat. *Cultural statistics*. 2011.
- [6] I.Papaioannou G.Ravasopoulos και V.Boutsinas. Conceptual partitioning: an efficient method for continuous nearest neighbor monitoring. 2010.
- [7] Jose Juan Montes. Cubes Viewer - Data exploration and visualization tool. <http://www.cubesviewer.com>, 2012-2016. Last accessed 24/09/2019.
- [8] M. Steinbach P. Tan και V. Kumar. *Introduction to Data Mining*. Pearson Addison-Wesley, 2006.
- [9] R.Ramakrishnan και J.Gehrke. *Database Management Systems*. McGraw-Hill, 3ή έκδοση, 2003.
- [10] S. Urbanek. Data a Brewery. <http://cubes.databrewery.org>. Last accessed 24/09/2019.
- [11] S. Urbanek. Cubes - Lightweight Python OLAP. EuroPython. <https://www.slideshare.net/Stiivi/cubes-lightweight-python-olap>, 2012. Last accessed 24/09/2019.
- [12] S. Urbanek. Cubes Documentation: Release 1.1. <http://cubes.readthedocs.io/en/v1.1/>, 2016. Last accessed on 12/09/2019.
- [13] Rakesh Verma. THE DATA MINING HYPERTEXTBOOK. [http://www.hypertextbookshop.com/dataminingbook/public\\_version/contents/](http://www.hypertextbookshop.com/dataminingbook/public_version/contents/)

chapters/chapter003/section004/blue/page004.html, 2009. Last accessed 29/09/2019.

- [14] "K. Wendt". "Traffic Monitor: Data Display for Traffic Visualisation at Airports". "Diploma Thesis", "2007".
- [15] W.H.Inmon. The data warehouse and data mining. *Communications of the ACM*, 39(11):49–50, 1996.
- [16] Ευστάθιος.Γ.Κύρκος. *Επιχειρηματική Ευφύια και Εξόρυξη Δεδομένων*. Σύνδεσμος Ελληνικών Ακαδημαϊκών Βιβλιοθηκών, Αθήνα, 2015.



# Συντομογραφίες

κ.λπ.	και λοιπά
κ.α	και άλλα
π.χ.	παραδείγματος χάριν
OLAP	Online Analytical Processing
BI	Business Intelligence
JSON	JavaScript Object Notation
csv	comma-separated values



# Ορολογία - Γλωσσάρι

## Ελληνικός όρος

Επιχειρηματική Ευφυΐα

κύβος

ομάδα

ανάλυση

προκαθορισμένο

ανοιχτού κώδικα

εργαλείο

δομή

εφαρμογή

δεδομένα

συνάθροιση

σύνδεση

φιλτράρισμα

διακομιστής

## Αγγλικός όρος

Business Intelligence

cube

group

analysis

default

open source

tool

framework

application

data

aggregation

join

filtering

server

