



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ
ΤΗΛΕΜΑΤΙΚΗ

ΚΑΤΕΥΘΥΝΣΗ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ

**Τίτλος Εργασίας: ΑΝΑΛΥΣΗ ΚΑΙ ΑΠΟΤΥΠΩΣΗ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ
ΟΙΚΟΝΟΜΙΚΩΝ ΔΕΙΚΤΩΝ**

Διπλωματική Εργασία

Όνομα φοιτητή: ΑΝΑΣΤΑΣΙΟΣ ΘΕΟΧΑΡΗΣ

ΑΘΗΝΑ, ΦΕΒΡΟΥΑΡΙΟΣ 2018



ΧΑΡΟΚΟΠΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

ΣΧΟΛΗ ΨΗΦΙΑΚΗΣ ΤΕΧΝΟΛΟΓΙΑΣ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΜΑΤΙΚΗΣ

**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΠΛΗΡΟΦΟΡΙΚΗ ΚΑΙ
ΤΗΛΕΜΑΤΙΚΗ**

ΚΑΤΕΥΘΥΝΣΗ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΙΣΤΟΥ

Τριμελής Εξεταστική Επιτροπή

Ηρακλής Βαρλάμης(Επιβλέπων)

**Επίκουρος Καθηγητής, Πληροφορική και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Γεώργιος Δημητρακόπουλος

**Επίκουρος Καθηγητής, Πληροφορική και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Χρήστος Μιχαλακέλης

**Επίκουρος Καθηγητής, Πληροφορική και Τηλεματικής, Χαροκόπειο
Πανεπιστήμιο**

Ο Αναστάσιος Θεοχάρης

δηλώνω υπεύθυνα ότι:

- 1) Είμαι ο κάτοχος των πνευματικών δικαιωμάτων της πρωτότυπης αυτής εργασίας και από όσο γνωρίζω η εργασία μου δε συκοφαντεί πρόσωπα, ούτε προσβάλλει τα πνευματικά δικαιώματα τρίτων.
- 2) Αποδέχομαι ότι η ΒΚΠ μπορεί, χωρίς να αλλάξει το περιεχόμενο της εργασίας μου, να τη διαθέσει σε ηλεκτρονική μορφή μέσα από τη ψηφιακή Βιβλιοθήκη της, να την αντιγράψει σε οποιοδήποτε μέσο ή/και σε οποιοδήποτε μορφότυπο καθώς και να κρατά περισσότερα από ένα αντίγραφα για λόγους συντήρησης και ασφάλειας.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω από καρδιάς τον επιβλέποντα καθηγητή Κύριο Ηρακλή Βαρλάμη για την υποστήριξη που μου έδωσε και τη δυνατότητα να εμβαθύνω τις γνώσεις μου στο τομέα του data analytics, την εξοικείωση με νέα εργαλεία, ιδιαίτερα δημοφιλή στην αγορά εργασίας και τον τρόπο προσέγγισης ερευνητικών ζητημάτων. Η ιδέα της διπλωματικής με βοήθησε να αποκτήσω μια πιο σφαιρική άποψη για τον τρόπο που προσεγγίζουμε ζητήματα ανάλυσης δεδομένων και η κατεύθυνση που μου έδωσε ήταν το έναυσμα της βαθιάς κατανόησης εννοιών σε όλη τη στοίβα της επιχειρηματικής ευφυΐας και της εξόρυξης γνώσης.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περίληψη στα Ελληνικά.....	σ.6
Abstract.....	σ.7
Κατάλογος Εικόνων.....	σ.8
Κατάλογος Πινάκων.....	σ.9
Κατάλογος Διαγραμμάτων	σ.10
Κεφ.1: Εισαγωγή.....	σ.11
1.1: Στόχος εργασίας.....	σ.12
1.2: Οφέλη εργασίας.....	σ.12
1.3: Σύντομη περιγραφή εργαλείων.....	σ.13
Κεφ.2: Σχετικές Έρευνες και το σύνολο δεδομένων που χρησιμοποιήθηκε.....	σ.14
2.1: Σχετικές Μελέτες.....	σ.14
2.2: Dataset.....	σ.17
Κεφ.3: Γενική αρχιτεκτονική και εργαλεία που χρησιμοποιήθηκαν.....	σ.20
3.1: Γενική αρχιτεκτονική.....	σ.20
3.2: Λογισμικά.....	σ.22
3.3: Συνδεσιμότητα όλων αυτών των εργαλείων.....	σ.28
Κεφ.4: Υλοποίηση.....	σ.31
4.1: Εξαγωγή, μετασχηματισμός και φόρτωση των δεδομένων.....	σ.31
4.2: Δημιουργία κύβων.....	σ.42
Κεφ.5: Αποτελέσματα.....	σ.47
Κεφ.6: Συμπεράσματα.....	σ.53
6.1: Μελλοντικά ζητήματα.....	σ.54

Περίληψη στα Ελληνικά

Σκοπός της εργασίας είναι να γίνει μια ανάλυση από δεδομένα της παγκόσμιας τράπεζας και να διαπιστωθούν κάποια συμπεράσματα από ένα συγκεκριμένο dataset, το οποίο έχει να κάνει με οικονομικούς δείκτες διάφορων χωρών στο κόσμο για συγκεκριμένα χρονικά διαστήματα, μέσα από ενδεικτικά γραφήματα/αναφορές.

Πιο συγκεκριμένα συλλέγονται δεδομένα για τις εξαγωγές και εισαγωγές αγαθών, τη παρούσα αξία του εξωτερικού χρέους ενώ μελετάται ακόμα και το κεντρικό δημόσιο χρέος για τις διάφορες χώρες και τα είδη οικονομιών που έχουν.

Χρησιμοποιούνται λογισμικά τόσο για την ανάλυση και την επεξεργασία του συνόλου δεδομένων όσο και για την οπτικοποίηση τους σε μια πληθώρα επιλογών διαγραμμάτων μέσω του εργαλείου tableau, με στόχο την ακριβή και λεπτομερή ανάλυση. Το μοντέλο υλοποίησης που ακολουθήθηκε είναι το πολυδιάστατο και πιο συγκεκριμένα το αστεροειδή σχήμα, δηλαδή υπάρχει ένας βασικός πίνακας που χαρακτηρίζεται fact table και πίνακες ξεχωριστοί για κάθε μια διάσταση, ενώ από αποτελέσματα μπορέσαμε και διαπιστώσαμε χώρες που έχουν καλή οικονομική ανάπτυξη, άλλες να μένουν στάσιμες και να χρειάζονται παρέμβαση για να επιβιώσουν ή άλλες να αναπτύσσονται με γρήγορους ρυθμούς δείχνοντας στάδια ανάπτυξης και άνθησης.

Λέξεις κλειδιά: επιχειρηματική ευφυΐα, οπτικοποίηση δεδομένων, προετοιμασία δεδομένων

Abstract ή Περίληψη στα Αγγλικά

The purpose of this master thesis is to analyze a dataset from the world bank and find some interesting topics, conclusion which has to do with economic indicators of various countries in the world for specific periods, through some indicative graphs/reports.

In particular, we collect data about exports and imports of goods, the present value of external debt and the central government debt for different countries and types of economies.

We use software to analyze and process the dataset and then to visualize it inside a variety of chart options via tableau, with aim of accurate and detailed analysis. The implementation model was the multidimensional and more specifically the star schema. In that case we have a basic table that is characterized as fact and then we have separate tables for each dimension.

The results show us that there are different kind of economies, some with good economic growth, others with economic stagnation, accompanied by high unemployment and need intervention to survive or others growing fast.

Keywords: Business Intelligence, data visualization, data preparation

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

Εικ.3.2.2: Join λειτουργία στο εργαλείο Datawatch	σ.24
Εικ.3.2.3: Ορισμός μια συγκεκριμένη γραμμής ως table header στο Datawatch.....	σ.25
Εικ.3.2.4: Μετατροπή μιας στήλης από συμβολοσειρά σε αριθμό στο Datawatch.....	σ.25
Εικ.3.2.5: Διαχείριση και εξαγωγή αποτελεσμάτων στο Datawatch.....	σ.26
Εικ.3.2.6: Κύριες Λειτουργίες του εργαλείου Alteryx.....	σ.26
Εικ.3.3.1: Εισαγωγή δεδομένων στο εργαλείο Alteryx για να τροφοδοτήσουν τις ροές..	σ.28
Εικ.3.3.2: Γράψιμο των αποτελεσμάτων των ροών σε Oracle βάση δεδομένων.....	σ.29
Εικ.3.3.3: Δημιουργία νέας πηγής δεδομένων και σύνδεση σε oracle server(tableau)..	σ.29
Εικ.3.3.4: Το μοντέλο δεδομένων group-αριθμένο ανά data source(tableau).....	σ.30
Εικ.4.1.1: Δημιουργία διάστασης Χώρα(alteryx).....	σ.31
Εικ.4.1.2: Η έξοδος αποτελεσμάτων από τη ροή της διάστασης της Χώρας(alteryx).....	σ.32
Εικ.4.1.3: Δημιουργία διάστασης της εισοδηματικής κατάστασης(alteryx).....	σ.32
Εικ.4.1.4: Η έξοδος των αποτελεσμάτων από τη ροή της διάστασης της εισοδηματικής κατάστασης(alteryx).....	σ.33
Εικ.4.1.5: Δημιουργία διάστασης χρονολογιών(alteryx).....	σ.33
Εικ.4.1.6: Η έξοδος των αποτελεσμάτων της ροής της διάστασης των χρονολογιών.....	σ.34
Εικ.4.1.7: Δημιουργία διάστασης οικονομικών δεικτών.....	σ.34
Εικ.4.1.8: Η έξοδος των αποτελεσμάτων της ροής της διάστασης των οικονομικών δεικτών(alteryx).....	σ.35
Εικ.4.1.9: Post create sql statement για τον πίνακα της διάστασης της χώρας.....	σ.35
Εικ.4.1.10: Η ροή για τον fact πίνακα του κεντρικού δημόσιου χρέους.....	σ.36
Εικ.4.1.11: Η ροή για τον fact πίνακα των εξαγωγών αγαθών και υπηρεσιών.....	σ.37
Εικ.4.1.12: Η ροή για τον fact πίνακα των εισαγωγών αγαθών και υπηρεσιών.....	σ.38
Εικ.4.1.13: Ο κώδικας που υλοποιεί το Materialized view.....	σ.39
Εικ.4.1.14: Η εικόνα της βάσης μετά το πέρας της εκτέλεσης των ροών.....	σ.40
Εικ.4.2.1: Υλοποίηση κύβου για τις εξαγωγές αγαθών.....	σ.42
Εικ.4.2.2: Δημιουργία του κύβου χρησιμοποιώντας όλες τις διαστάσεις και τα facts.....	σ.43
Εικ.4.2.4: Συνάρτηση get_isotimia(from dollars to euros).....	σ.45
Εικ.4.2.5: Κλήση συνάρτησης στο materialized view.....	σ.45
Εικ.5.2.1: Φίλτρο στο tableau που γίνεται επιλογή συγκεκριμένων χωρών.....	σ.48

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίν.2.2.1: Αρχικά δεδομένα με τις χώρες και την εισοδηματική τους κατάσταση.....	σ.18
Πίν.2.2.2: Αρχικά δεδομένα που σχετίζονται με οικονομικούς δείκτες	σ.18
Πίν.2.2.3: Όλες οι μετρικές για διάφορες χώρες και οικονομικούς δείκτες.....	σ.19
Πιν.4.1.15: Ο fact πίνακας των εξαγωγών αγαθών με τα πρωτεύοντα και ξένα κλειδιά.....	σ.41
Πιν.4.2.3: Πίνακας με όλες τις ισοτιμίες από δολάρια σε ευρώ από το 1999 μέχρι και το 2017.....	σ.44

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διαγ.2.1.4: Το καθαρό περιθώριο τόκων σε τόσο ισχυρές όσο και αδύναμες οικονομίες..	σ.16
Διαγ.2.1.5: Δείκτες της ασφαλιστικής αγοράς ανά είδος εισοδήματος.....	σ.16
Διαγ.3.1: Διάγραμμα του αστεροειδούς σχήματος που υλοποιήθηκε.....	σ.21
Διάγ.3.2.1: Pipeline διάγραμμα για olap.....	σ.22
Διάγ.4.2.6: Εμφάνιση ποσού τόσο σε ευρώ όσο και σε δολάρια(tableau).....	σ.46
Διαγ.5.2.2: Εξαγωγές αγαθών και υπηρεσιών στη Βουλγαρία, Μάλτα και Σλοβενία για τη δεκαετία του 1990.....	σ.49
Διαγ.5.2.3: Εξαγωγές αγαθών και υπηρεσιών σε δολάρια για τις χώρες Βουλγαρία, Φιλανδία, Ελλάδα, Πολωνία, Πορτογαλία, Τουρκία μεταξύ των χρονολογιών 1990 και 2004.....	σ.50
Διαγ.5.2.4: Εισαγωγές αγαθών και υπηρεσιών σε δολάρια για τις χώρες της Ανατολικής Ασίας για τα χρονολογίες από 1970 μέχρι και 2003.....	σ.51
Διαγ.5.2.5: Κεντρικό δημόσιο χρέος για τις χώρες Αυστραλία, Ισπανία, Ηνωμένο Βασίλειο και Ηνωμένες Πολιτείες της Αμερικής από το 1995 μέχρι και το 2004.....	σ.52

ΚΕΦ.1: ΕΙΣΑΓΩΓΗ

Η Επιχειρηματική Ευφυΐα βρίσκεται στο επίκεντρο του ενδιαφέροντος των σύγχρονων μεγάλων επιχειρήσεων. Η λήψη αποφάσεων σε συνθήκες αβεβαιότητας, οι προκλήσεις της παγκοσμιοποίησης, η οικονομική κρίση, η διαθεσιμότητα δεδομένων και οι νέες τεχνολογίες και μέθοδοι ανάλυσης αποτέλεσαν το έναυσμα της ενσωμάτωσης της μέσα σε οργανισμούς για τη πρόβλεψη και κατάστροψη σχεδίων ώστε να είναι ανταγωνιστική και βιώσιμη.

Ο όρος Επιχειρηματική Ευφυΐα δεν είναι πρόσφατος. Ο Devens το 1865 χρησιμοποιεί αυτό τον όρο για να αναφερθεί στον τρόπο με τον οποίο ο τραπεζίτης Sir Henry Furnese αξιοποιούσε πληροφορίες νωρίτερα από τους ανταγωνιστές του, έτσι ώστε να επιτύχει αύξηση των κερδών του (Κύρκος, 1865, σ.24). Η επόμενη εμφάνιση του όρου καταγράφεται το 1958 σε τίτλο άρθρου του Luhn σε περιοδικό της IBM, όπου ο αναγνώστης θα συναντήσει διαφοροποιημένους ορισμούς της Επιχειρηματικής Ευφυΐας (Κύρκος, 1865, σ.24). Στο συγκεκριμένο σύγγραμμα, ορίζεται ως ένα σύνολο από μεθόδους ανάλυσης, τεχνολογίες, ικανότητες και στρατηγικές, οι οποίες έχουν ως στόχο την επεξεργασία των διαθέσιμων δεδομένων και την εξαγωγή χρήσιμης πληροφορίας από αυτά, για την διαδικασία λήψης επιχειρηματικών αποφάσεων. Η επιχειρηματική ευφυΐα επιτρέπει σε έναν οργανισμό να μαθαίνει, να αντιλαμβάνεται καταστάσεις και συμβάντα, να σκέφτεται αφαιρετικά, να προβλέπει τάσεις και μελλοντικά συμβάντα, να σχεδιάζει και να καινοτομεί. Η παραγόμενη πληροφορία μετουσιώνεται σε γνώση που αξιοποιείται από τα διοικητικά στελέχη, ώστε να δρομολογήσουν κατάλληλες δράσεις, που θα οδηγήσουν στο καθορισμό και την επίτευξη επιχειρηματικών στόχων, με τρόπο αποτελεσματικό και αποδοτικό. Τα συστήματα Επιχειρηματικής Ευφυΐας είναι εξειδικευμένα πληροφοριακά συστήματα που βασίζονται σε συγκεντρωτικά δεδομένα, τα οποία συνδυάζονται με λογισμικό ικανό να διεξάγει κατάλληλες αναλύσεις. Τα εργαλεία αυτά επιτρέπουν επιπλέον την ταχύτερη πρόσβαση στην πληροφορία, τη σύνταξη αναφορών και την προχωρημένη ανάλυση των δεδομένων.

Ο εξέλιξη της τεχνολογίας και οι νέες ανάγκες οδήγησαν τη δημιουργία νέων σταθμών της επιχειρηματικής Ευφυΐας. Οι αποθήκες δεδομένων (Data Warehouse) δηλαδή να συγκεντρώνεται δεδομένα από διάσπαρτα διάφορες πηγές σε μια κοινή καθώς και οι τεχνικές OLAP, δηλαδή να μπορεί ο χρήστης να προβάλει και να αναλύσει δεδομένα σε διάφορα επίπεδα

γενίκευσης (π.χ. Πωλήσεις ανά μήνα ή ανά τρίμηνο) είχε ως αποτέλεσμα την βέλτιστη και πιο λεπτομερή ανάλυση καθώς και το συνδυασμό μεγαλύτερου όγκου δεδομένων από διαφορετικά συστήματα που αλληλεπιδρούν μεταξύ τους.

1.1 ΣΤΟΧΟΣ ΕΡΓΑΣΙΑΣ

Ο τελικός στόχος της εργασίας είναι να δώσει τη δυνατότητα στο τελικό χρήστη να φτιάξει τις δικές του αναλύσεις, να βάλει τα δικά του φίλτρα και να φέρει τη δική του πληροφορία που νομίζει ότι θα του φανεί χρήσιμη μέσα από ένα μοντέλο δεδομένων έτοιμο που θα έχει φτιαχτεί και θα περιλαμβάνει όλη την απαραίτητη πληροφορία του dataset.

Μια τέτοια διαδικασία γίνεται αρκετά συχνά, δηλαδή όταν χρήστες συστημάτων θέλουν να κάνουν μια ανάλυση, έναν έλεγχο πάνω σε κάποια δεδομένα που έχουν συλλεχθεί ή καταχωρηθεί με κάποιο τρόπο σε μια εφαρμογή, τότε τα συστήματα επιχειρηματικής ευφυίας είναι αυτά που φτιάχνονται με σκοπό να αξιοποιήσουν αυτή την πληροφορία και να αποτυπώσουν γνώση δημιουργώντας ανάλογα σχετικά γραφήματα και αναφορές.

Το μοντέλο υλοποίησης που ακολουθήθηκε ξεφεύγει από το σχεσιακό, δηλαδή δεν έχουμε πλέον δεδομένα σε γραμμές και στήλες, αλλά ερχόμαστε στο πολυδιάστατο δηλαδή ένας απλός τρόπος να προσεγγίσουμε μετρικές με πολλές διαστάσεις ανάλογα τη πληροφορία που θέλουμε να εμφανίσουμε.

1.2 ΟΦΕΛΗ ΕΡΓΑΣΙΑΣ

Η εργασία ήταν χρήσιμη γιατί με βοήθησε να εξοικειωθώ με εργαλεία που δεν γνώριζα και να αφομοιώσω καλύτερα τη διαδικασία ανάλυσης και την εξόρυξη επιχειρηματικής γνώσης μέσα από ένα πολύ ενδιαφέρον και πάντα επίκαιρο γνωστικό αντικείμενο, αυτό της οικονομικής ανάπτυξης. Πολλά πράγματα μέσα από το κάθε εργαλείο μου έδωσαν πολλές ιδέες για μελλοντικά ζητήματα που πιθανώς να συναντήσω στην εργασιακή μου πορεία και θα επιτυγχάνουν τη διαδικασία επίλυσης ζητημάτων με το βέλτιστο και γρηγορότερο τρόπο. Επίσης παρατήρησα την εξέλιξη των τεχνολογιών που υπάρχουν στο κόσμο της πληροφορικής μέσα από αυτά τα πολύ ευρέως διαδεδομένα λογισμικά που χρησιμοποιήθηκαν στα πλαίσια αυτής της εργασίας. Τόσα τα front όσο και τα back λογισμικά είχαν να μου προσφέρουν τα δικά τους οφέλη, το κάθε ένα από τη δική του οπτική γωνία. Για το μπροστά κομμάτι είδα μια πολύ

γρήγορη εκτέλεση ερωτημάτων μέσα από ένα ιδιαίτερο φιλικό περιβάλλον τόσο για τον προγραμματιστή όσο και για το τελικό χρήστη, ενώ για το πίσω κομμάτι έμαθα πόσο αυτοματοποιημένες διαδικασίες μπορούν να φτιαχτούν και να τροφοδοτήσουν αποθήκες δεδομένων με την ελάχιστη ανθρώπινη παρέμβαση.

Επίσης ιδιαίτερα χρήσιμη διαδικασία ήταν η διαδικασία ανάλυσης των δεδομένων πριν την οποιαδήποτε τεχνική υλοποίηση και την διάσπαση πολύπλοκων ζητημάτων σε μικρά και απλά κομμάτια. Έρχοντας στη θέση του τελικού χρήστη και θέτοντας ερωτήματα τι θα ήταν αυτό που θα ήθελε να δει και πως θα μπορούσα να δομήσω αυτή τη πληροφορία, κατέληξα σε μικρά σενάρια με υποθετικές οθόνες/mockups και ενέργειες που θα έκανε ο χρήστης για να βγάλει κάποια συμπεράσματα. Έτσι λοιπόν αυτή η διαδικασία με βοήθησε στην ανάλυση, σχεδιάζοντας ομαδοποιημένες κατηγορίες δεδομένων και στο πως αλληλοσχετίζονται μεταξύ τους.

1.3 ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ ΕΡΓΑΛΕΙΩΝ

Τα εργαλεία που χρησιμοποιούνται σε αυτή την εργασία αφορούν τη προετοιμασία δεδομένων έτσι ώστε να είναι σε μορφή έτοιμη για επεξεργασία, ύστερα τον μετασχηματισμό και μετά τη φόρτωση τους σε αποθηκευτικούς χώρους έτσι ώστε να είναι έτοιμα για οπτικοποίηση με γραφήματα και τη δημιουργία dashboard ή αναφορών. Η επιλογή αυτών των εργαλείων έγινε γιατί είναι πολύ δημοφιλή, προσφέρουν σταθερότητα και στη διάρκεια του compilation δείχνουν με σαφήνεια errors και warnings. Επίσης το ένα συμπληρώνει το άλλο και έχουν πλήρη συμβατότητα, αν εξαιρέσουμε την αδυναμία του tableau να διαβάσει κάποια oracle packages, πράγμα που η microsoft το είχε λύσει.

Ας δούμε λίγο πρώτα τη συνεργασία μεταξύ τους. Καταρχάς έχοντας αρχικά ένα μεγάλο excel με όλα τα δεδομένα, έπρεπε σε πρώτη φάση να φτιαχτούν σε μορφή σχεσιακών πινάκων με τύπο και ονόματα στηλών. Το datawatch ¹ ήταν ιδανικό για αυτή τη δουλειά δηλαδή να διαβάσει αρχεία και να τα εξάγει σε μορφή row data που αναγνωρίζει και μια βάση δεδομένων. Το alteryx ² από την άλλη διάβασε αυτά τα αρχεία και όλους τους τύπους των πεδίων, πλήρη συμβατό και με την oracle πήρε τα δεδομένα αυτά και ξεκίνησε να εφαρμόζει operations και να καταλήξει να φτιάχνει ένα data warehouse σχήμα το οποίο μπορούσε να το διαβάσει το tableau ³.

¹ <https://www.datawatch.com/>

² <https://www.alteryx.com/>

³ <https://www.tableau.com/>

ΚΕΦ.2: ΣΧΕΤΙΚΕΣ ΕΡΕΥΝΕΣ ΚΑΙ ΤΟ ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΕ

2.1. ΣΧΕΤΙΚΕΣ ΜΕΛΕΤΕΣ

Οι παράγοντες της οικονομικής ανάπτυξης της κάθε χώρας έχουν μελετηθεί ιδιαίτερα στην επιστήμη της οικονομίας. Από έρευνες έχει αποδειχθεί ότι το ανθρώπινο κεφάλαιο, οι επενδύσεις, η φυσική υποδομή και τα ιδρύματα αποτελούν κομβικό σημείο για τη ανάπτυξη κάποιας οικονομίας. Οι αλληλεπιδράσεις επίσης μεταξύ των χωρών συντελούν στην ανάπτυξη τόσο των εμπορευμάτων, στη τεχνολογική διάδοση και γενικά στις εισροές κεφαλαίων(FINLEY, 2005)

Μπορούμε να πούμε ακόμα ότι σε μια οικονομία της αγοράς, η προσοχή επικεντρώνεται στην έννοια των κυκλικών οικονομικών δραστηριοτήτων. Στη δεκαετία του 1960 ο καθηγητής Kenneth E. Boulding, πρωτοπόρος περιβαλλοντικός οικονομολόγος θεώρησε την οικονομία της γης ως ένα διαστημόπλοιο με μέτρο της οικονομικής επιτυχίας τη φύση, την έκταση, την ποιότητα και όχι τον οικονομικό πλούτο. Οικονομικοί πόροι και οικονομικά απόβλητα δεν μπορούν πλέον να θεωρούνται ανεξάρτητα. Αν οι άνθρωποι δεν ασχολούνται με την ανακύκλωση οικονομικών πόρων και τη διαχείριση αποβλήτων, τότε τα αποθέματα θα εξαφανιστούν σύντομα από τη γη(Donald, Brian, Chen, 2015).

Σε συγκεκριμένη αναφορά παρουσιάζονται πρόσφατες τάσεις στην οικονομική ανάπτυξη χρηματοπιστωτικών ιδρυμάτων και την κινητικότητα αγορών μεταξύ χωρών, θέμα το οποίο αναζωπυρώθηκε μετά την οικονομική κρίση του 2007.

Μελετούνται συγκεκριμένοι οικονομικοί δείκτες μέσα από δεδομένα της παγκόσμιας τράπεζας, όπου παρέχονται στατιστικά για το μέγεθος,τη δραστηριότητα, την αποδοτικότητα, την οικονομική υγεία των τραπεζών και των αγορών μέσα από ένα εύρος χρονολογιών για διάφορες χώρες. Κάνοντας κατηγοριοποίηση αυτών των δεικτών θα μπορούσαμε να πούμε ότι κάποιοι αναφέρονται στο τραπεζικό σύστημα(μέγεθος,σταθερότητα,αποδοτικότητα), άλλοι στις κεφαλαιαγορές και τον ασφαλιστικό τομέα και άλλοι γενικοί για την παγκόσμια οικονομία.

Για το τραπεζικό σύστημα αναφέρεται ότι αποτελεί αναμφισβήτητο το μεγαλύτερο κομμάτι του οικονομικού συστήματος για τις περισσότερες χώρες, κυρίως για τις αναδυόμενες και αναπτυσσόμενες αγορές. Τονίζεται επίσης ότι οι καταθέσεις δεν αποτελούν τη μόνη πηγή

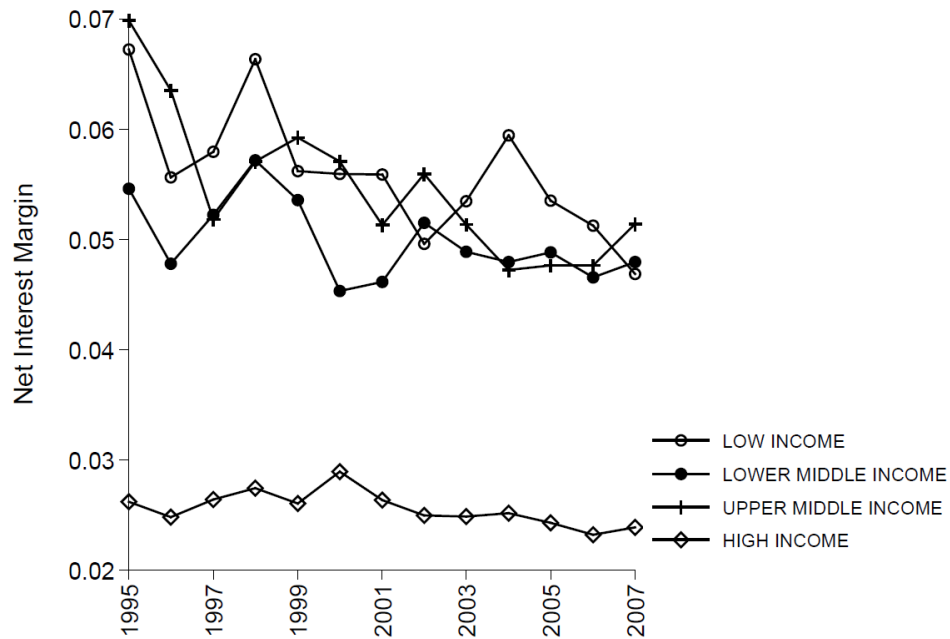
χρηματοδότησης των τραπεζών, ενώ πιο φτωχές χώρες (με εξαίρεση τα τελευταία χρόνια) έχουν μεγαλύτερα καθαρά περιθώρια επιτοκίου σε αντίθεση με πιο ισχυρές χώρες (διάγραμμα 2.1.4) που παρατηρείται να έχουν χαμηλά και σταθερά περιθώρια, γεγονός που συνετέλεσε στην οικονομική κρίση του 2007 και οδήγησε τις τράπεζες να στραφούν σε εναλλακτικές πηγές χρηματοδότησης.

Δείκτες για το σύνολο των χρηματιστηριακών συναλλαγών και του ασφαλιστικού τομέα οδηγούν στα δικά τους συμπεράσματα. Γίνεται αναφορά ότι ένα μικρό αλλά ενεργό χρηματιστήριο έχει συνήθως μεγάλη αναλογία κύκλου εργασιών, ενώ ένα μεγάλο αλλά όχι ιδιαίτερα ενεργό θα έχει μικρό σχετικά τζίρο. Επίσης μια σημαντική παρατήρηση συγκρίνοντας ένα δείγμα από χώρες είναι ότι η ρευστότητα μιας χρηματιστηριακής αγοράς είναι αυτό που έχει σημασία για την οικονομική ανάπτυξη και όχι το μέγεθος της. Ως δείκτες για το μέγεθος της εγχώριας αγοράς ομολόγων χρησιμοποιούμε το ιδιωτικό και δημόσιο ομόλογο. Ενώ η ιδιωτική κεφαλαιοποίηση των αγορών ομολόγων συσχετίζεται θετικά με το εισόδημα των χωρών, δεν υπάρχει αυτή η σαφής συσχέτιση για τις δημόσιες αγορές ομολόγων και από το πείραμα διαπιστώνεται ανοδική τάση τα τελευταία 12 χρόνια, αν και πολύ λιγότερο από ότι στις χρηματιστηριακές αγορές. Αυτό συμβαίνει γιατί τα ομόλογα διαπραγματεύονται συνήθως γύρω από την ονομαστική αξία, σε αντίθεση με τις μετοχές των οποίων οι τιμές μπορεί να είναι πολλαπλάσιο της αρχικής λογιστικής αξίας. Για του δείκτες που αφορούν το μέγεθος του ασφαλιστικού τομέα παρατηρείτε μια μεγάλη αύξηση με βάση το επίπεδο εισοδήματος μιας χώρας (διάγραμμα 2.1.5).

Όσο αναφορά τη διεθνή παγκόσμια οικονομία, η τάση της παγκοσμιοποίησης είχε γίνει ιδιαίτερα έντονη ως προς τα δάνεια και τα ομόλογα κυρίως σε χώρες με υψηλό εισόδημα, ενώ αντίθετα οι χώρες της χαμηλής και μεσαίας τάξης επωφελήθηκαν από τις αυξανόμενες ροές εμβασμάτων και στο μέλλον περιμένουμε να δούμε αν αυτές οι τάσεις θα συνεχιστούν (Thorsten, Asli, 2009).

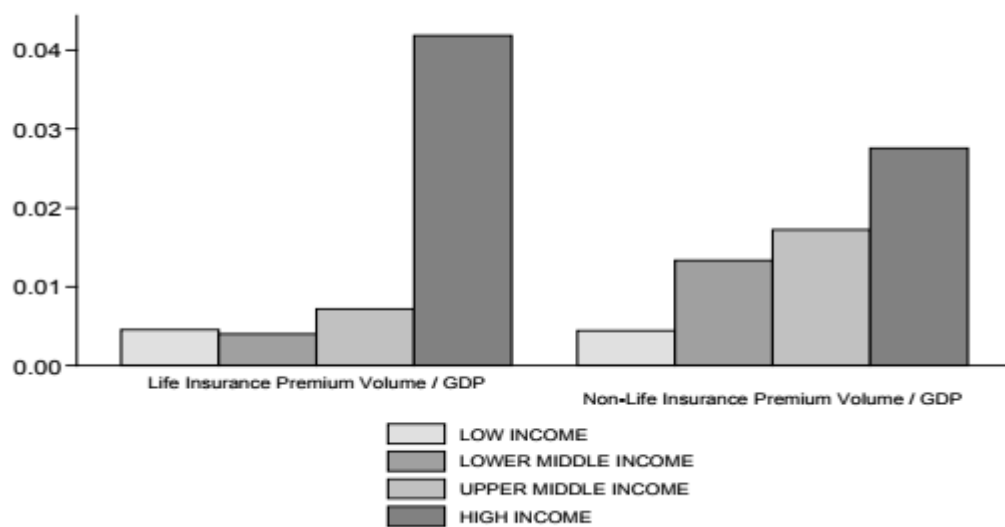
Διάγραμμα 2.1.4

Το καθαρό περιθώριο τόκων σε τόσο ισχυρές όσο και αδύναμες οικονομίες(Thorsten, Asli, 2009, εικόνα10)



Διάγραμμα 2.1.5

Δείκτες της ασφαλιστικής αγοράς ανά είδος εισοδήματος
(Thorsten, Asli, 2009, εικόνα 17)



Σε συγκεκριμένη μελέτη θέτονται ερωτήματα για τις έγκυρες οικονομικές προβλέψεις μετά την παγκόσμια οικονομική κρίση του 2018 που οδήγησε στο τρόπο που ερμηνεύουμε τα οικονομικά φαινόμενα που συνεχώς συμβαίνουν γύρω μας. Το πείραμα είχε να κάνει με τη δημιουργία μοντέλων πρόβλεψης της συμπεριφοράς διεθνών οίκων αξιολόγησης σε σχέση με τα οικονομικά δεδομένα που ανακοινώνονται για τις εθνικές οικονομίες. Με τα διαθέσιμα δεδομένα κατασκευάστηκε λοιπόν ένα δέντρο απόφασης που με ακρίβεια 70% δουλεύει σωστά, έδειξε ότι οι δείκτες που φαίνονται να έχουν την μεγαλύτερη βαρύτητα είναι time passed in years και έχουν gdp growth 2 έως 3(ρυθμό ανάπτυξης του ΑΕΠ στα έτη 2 και 3 στο μέσο του χρονικού διαστήματος των 5 ετών) με βαρύτητα συσχέτισης κοντά στο 0,1 και 0,02(Γεωργούσης Ηλίας, 2016).

2.2. DATASET

Το dataset που επιλέχτηκε έχει να κάνει όπως είπαμε με οικονομικούς δείκτες και την ανάπτυξη οικονομιών.

Το dataset περιλαμβάνει διάφορα δεδομένα, όπως κάποια από αυτά είναι δημογραφικά στοιχεία για χώρες και περιοχές των ηπείρων αυτών, εισοδηματική κατάσταση(πίνακας 2.2.1), άλλα για δείκτες οικονομικών στοιχείων και λεπτομέρειες για το τι ακριβώς περιγράφουν αυτά(πίνακας 2.2.2)και τις πηγές προέλευσης που έχουν συλλεχθεί. Τέλος έχουμε τα μετρήσιμα για τους διάφορους αυτούς δείκτες οικονομικών στοιχείων με ένα εύρος χρονολογιών από το 1960 μέχρι και το 2016(πίνακας 2.2.3).

Πίνακας 2.2.1

Αρχικά δεδομένα με τις χώρες και την εισοδηματική τους κατάσταση

	A	B	C	D	E	F	G	H
40	CHN	East Asia & Pacific	Upper middle income	On 1 July 1997 China	China			
41	CIV	Sub-Saharan Africa	Lower middle income		Côte d'Ivoire			
42	CMR	Sub-Saharan Africa	Lower middle income		Cameroon			
43	COD	Sub-Saharan Africa	Low income	WB-3 code changed	Congo, Dem. Rep.			
44	COG	Sub-Saharan Africa	Lower middle income		Congo, Rep.			
45	COL	Latin America & Car	Upper middle income		Colombia			
46	COM	Sub-Saharan Africa	Low income		Comoros			
47	CPV	Sub-Saharan Africa	Lower middle income	Cabo Verde is the n	Cabo Verde			
48	CRI	Latin America & Car	Upper middle income	National accounts a	Costa Rica			
49	CSS			Caribbean small sta	Caribbean small states			
50	CUB	Latin America & Car	Upper middle income	Based on official go	Cuba			
51	CUW	Latin America & Car	High income		Curaçao			
52	CYM	Latin America & Car	High income		Cayman Islands			
53	CYP	Europe & Central As	High income	A simple multiplier	Cyprus			
54	CZE	Europe & Central As	High income		Czech Republic			
55	DEU	Europe & Central As	High income	A simple multiplier	Germany			
56	DJI	Middle East & North	Lower middle income		Djibouti			
57	DMA	Latin America & Car	Upper middle income		Dominica			
58	DNK	Europe & Central As	High income		Denmark			
59	DOM	Latin America & Car	Upper middle income		Dominican Republic			
60	DZA	Middle East & North	Upper middle income	The new base year	Algeria			
61	EAP			East Asia and Pacific	East Asia & Pacific (excluding high income)			
62	FAR			Early-dividend coun	Early-demographic dividend			
		Data	Data (2)	Metadata - Countries	Metadata - Indicators			

Πίνακας 2.2.2

Αρχικά δεδομένα που σχετίζονται με οικονομικούς δείκτες

	A	B	C
1	INDICATOR_CODE	INDICATOR_NAME	SOURCE_NOTE
2	NE.EXP.GNFS.CN	Exports of goods and serv	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, fr
3	NE.EXP.GNFS.CD	Exports of goods and serv	Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, fr
4	NE.DAB.TOTL.ZS	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
5	NE.DAB.TOTL.KN	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
6	NE.DAB.TOTL.CD	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
7	NE.DAB.TOTL.CN	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
8	NE.DAB.TOTL.CD	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
9	NE.DAB.DEFL.ZS	Gross national expenditu	Gross national expenditure (formerly domestic absorption) is the sum of household final consumption expenditure (formerly private consumption), general governm
10	NE.CON.TOTL.KN	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general go
11	NE.CON.TOTL.KD	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general go
12	NE.CON.TOTL.CN	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (private consumption) and general governmen
13	NE.CON.TOTL.CD	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (private consumption) and general governmen
14	NE.CON.TETC.ZS	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (private consumption) and general governmen
15	NE.CON.TETC.KN	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general go
16	NE.CON.TETC.KD.ZG	Final consumption expen	Average annual growth of final consumption expenditure based on constant local currency. Aggregates are based on constant 2010 U.S. dollars. Final consumption ex
17	NE.CON.TETC.KD	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general go
18	NE.CON.TETC.CN	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (formerly private consumption) and general go
19	NE.CON.TETC.CD	Final consumption expen	Final consumption expenditure (formerly total consumption) is the sum of household final consumption expenditure (private consumption) and general governmen
20	NE.CON.PRVT.PP.KD	Household final consump	Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, was
21	NE.CON.PRVT.PP.CD	Household final consump	Household final consumption expenditure (formerly private consumption) is the market value of all goods and services, including durable products (such as cars, was
22	NE.CON.PRVT.PC.KD.ZI	Household final consump	Annual percentage growth of household final consumption expenditure per capita, which is calculated using household final consumption expenditure in constant 20
23	NE.CON.PRVT.PC.KD	Household final consump	Household final consumption expenditure per capita (private consumption per capita) is calculated using private consumption in constant 2010 prices and World Bank
	Data	Data (2)	Metadata - Countries
			Metadata - Indicators

Πίνακας 2.2.3

Όλες οι μετρικές για διάφορες χώρες και οικονομικούς δείκτες

	A	B	C	D	E	F
1	Data Source	World Development Indicators				
2	Last Updated Date	30/10/2017				
3						
4	Country Name	Country Code	Indicator Name	Indicator Code	1960	1961
5	Aruba	ABW	Exports of goods and services (current LCU)	NE.EXP.GNFS.CN		
6	Aruba	ABW	Exports of goods and services (current US\$)	NE.EXP.GNFS.CD		
7	Aruba	ABW	Gross national expenditure (% of GDP)	NE.DAB.TOTL.ZS		
8	Aruba	ABW	Gross national expenditure (constant LCU)	NE.DAB.TOTL.KN		
9	Aruba	ABW	Gross national expenditure (constant 2010 US\$)	NE.DAB.TOTL.KD		
10	Aruba	ABW	Gross national expenditure (current LCU)	NE.DAB.TOTL.CN		
11	Aruba	ABW	Gross national expenditure (current US\$)	NE.DAB.TOTL.CD		
12	Aruba	ABW	Gross national expenditure deflator (base year varies by country)	NE.DAB.DEFL.ZS		
13	Aruba	ABW	Final consumption expenditure (constant LCU)	NE.CON.TOTL.KN		
14	Aruba	ABW	Final consumption expenditure (constant 2010 US\$)	NE.CON.TOTL.KD		
15	Aruba	ABW	Final consumption expenditure (current LCU)	NE.CON.TOTL.CN		
16	Aruba	ABW	Final consumption expenditure (current US\$)	NE.CON.TOTL.CD		
17	Aruba	ABW	Final consumption expenditure, etc. (% of GDP)	NE.CON.TETC.ZS		
18	Aruba	ABW	Final consumption expenditure, etc. (constant LCU)	NE.CON.TETC.KN		
19	Aruba	ABW	Final consumption expenditure, etc. (annual % growth)	NE.CON.TETC.KD.ZG		
20	Aruba	ABW	Final consumption expenditure, etc. (constant 2010 US\$)	NE.CON.TETC.KD		
21	Aruba	ABW	Final consumption expenditure, etc. (current LCU)	NE.CON.TETC.CN		
22	Aruba	ABW	Final consumption expenditure, etc. (current US\$)	NE.CON.TETC.CD		
23	Aruba	ABW	Household final consumption expenditure PPP (constant 2011 international \$)	NE.CON.PRVT.PP.KD		

Εμείς θα εστιάσουμε σε συγκεκριμένους δείκτες και αυτοί έχουν να κάνουν με :

α) Εξαγωγές αγαθών και υπηρεσιών σε δολάρια - Exports of Goods and services in dollars

<https://data.worldbank.org/indicator/NE.EXP.GNFS.CD>

Οι εξαγωγές αναπαριστούν ένα σημαντικό παράγοντα για τα οικονομικά έσοδα μιας χώρας.

Αυτός ο δείκτης μετράει το συνολικό πλήθος σε δολάρια όλων των αγαθών και άλλων υπηρεσιών της αγοράς προς όλο τον υπόλοιπο κόσμο.

Οι εξαγωγές αυτές συμπεριλαμβάνουν την αξία των εμπορευμάτων, την ασφάλιση, την μεταφορά ανθρώπων, τα δικαιώματα αδείας, τα ταξίδια και άλλες υπηρεσίες όπως η επικοινωνία, η κατασκευή, οικονομικές υπηρεσίες, προσωπικές και κυβερνητικές υπηρεσίες.

Δεν συμπεριλαμβάνονται οι αποζημιώσεις των εργαζομένων, μεταφορές χρημάτων και εισοδήματα από επενδύσεις.

Μονάδα μέτρησης: Δολάρια

Όρια τιμών: 0-75 Trillion

β) Εισαγωγές αγαθών και υπηρεσιών σε δολάρια -Imports of Goods and services in dollars

<https://data.worldbank.org/indicator/NE.IMP.GNFS.CD>

Αυτός ο δείκτης μετράει το συνολικό ποσό σε δολάρια εισαγωγών αγαθών που περνάνε μέσα στη χώρα ετησίως . Οι εισαγωγές αυτές συμπεριλαμβάνουν την αξία των εμπορευμάτων, της ασφάλειας, των μεταφορών, των τελών άδειας, των ταξιδιών και άλλων υπηρεσιών όπως κυβερνητικές, χρηματοπιστωτικές, επιχειρησιακές και πληροφοριακές.

Δεν συμπεριλαμβάνονται οι αποζημιώσεις των εργαζομένων, μεταφορές χρημάτων και εισοδήματα από επενδύσεις.

Μονάδα μέτρησης: Δολάρια

Όρια τιμών: 0-79 Trillion

γ) Παρούσα αξία του εξωτερικού χρέους σε δολάρια

Το άθροισμα του βραχυπρόθεσμου εξωτερικού χρέους συν το προεξοφλημένο άθροισμα του συνολικού δημόσιου χρέους σε δολάρια.

Μονάδα μέτρησης: Δολάρια

Όρια τιμών: 0-1.2 Trillion

δ) Κεντρικό δημόσιο χρέος(% of GDP) -Central Government debt, total(% of GDP)

<https://data.worldbank.org/indicator/GC.DOD.TOTL.GD.ZS>

Το χρέος αυτό είναι ολόκληρο το κεφάλαιο σε χρηματικές μονάδες που χρωστάει ο ευρύτερος δημόσιος τομέας συμπεριλαμβάνοντας όλα τα επίπεδα δημόσιας διοίκησης ενός κράτους όπως κυβέρνηση, νομαρχία, δήμος κτλ. Το δημόσιο χρέος αυξάνεται από έτος σε έτος κατά το ποσοστό που ο ετήσιος κρατικός προϋπολογισμός παρουσιάζει έλλειμμα, ή αντιστρόφως μειώνεται κατά το ποσοστό που παρουσιάζει πλεόνασμα.

Επειδή το χρέος είναι ένα απόθεμα και όχι μια ροή, μετριέται ως προς μια συγκεκριμένη ημερομηνία, συνήθως την τελευταία ημέρα του εκάστοτε οικονομικού έτους.

Μονάδα μέτρησης: συνολικό % του ΑΕΠ

Όρια τιμών: 0 - 79,083%.

ΚΕΦ.3: ΓΕΝΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΚΑΙ ΕΡΓΑΛΕΙΑ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΗΘΗΚΑΝ

3.1. ΓΕΝΙΚΗ ΑΡΧΙΤΕΚΤΟΝΙΚΗ

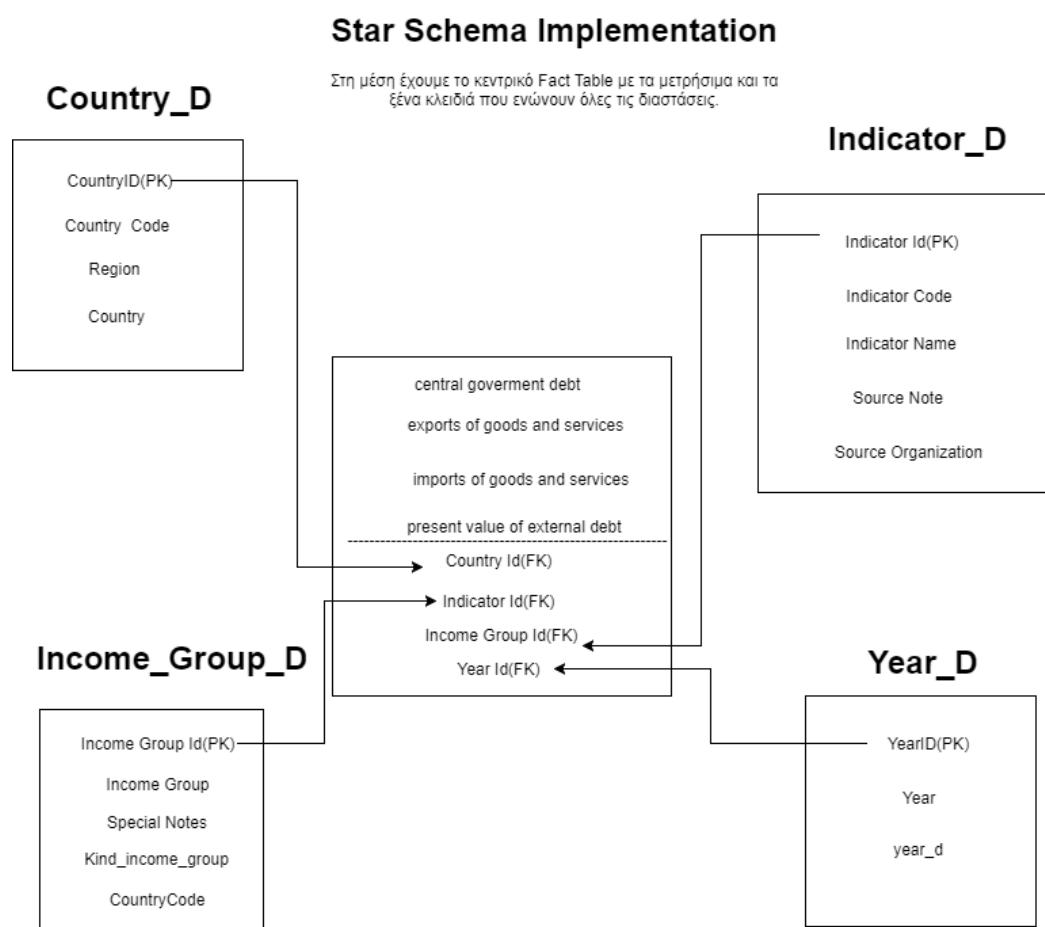
Η σχεδίαση και η οργάνωση των δεδομένων είναι διαφορετική από τις κλασσικές βάσεις δεδομένων. Τα διαγράμματα Οντοτήτων-Συσχετίσεων και οι τεχνικές κανονικοποίησης είναι οι κλασσικές μέθοδοι για τη σχεδίαση των βάσεων δεδομένων των συστημάτων επεξεργασίας δοσοληψιών (OLTP). Αυτές οι μέθοδοι αποδεικνύονται συχνά ακατάλληλες για τη σχεδίαση των Αποθηκών Δεδομένων, καθώς ο στόχος τους είναι να αντιμετωπίσουν προβλήματα, όπως ο πλεονασμός (redundancy) ή η ανανέωση των δεδομένων.

Η τεχνική που χρησιμοποιήσαμε είναι το αστεροειδή σχήμα. Εκεί υπάρχει ένας βασικός πίνακας που χαρακτηρίζεται fact table και πίνακες ξεχωριστοί για κάθε μια διάσταση. Κάθε εγγραφή του fact table αποτελείται από ένα δείκτη (ξένο κλειδί) σε μια εγγραφή κάθε ενός από τους πίνακες διαστάσεων. Κάθε πίνακας διάστασης (dimension table) περιλαμβάνει εγγραφές

που αντιστοιχούν σε τιμές των διαστάσεων(διάγραμμα 3.1.1). Η κύρια αδυναμία των αστεροειδών σχημάτων εντοπίζεται στον τρόπο με τον οποίο εκφράζουν τις ιεραρχίες των διαστάσεων, ενώ μια εναλλακτική καλύτερη μοντελοποίηση των ιεραρχιών γίνεται από τα σχήματα χιονονιφάδας (snowflake) όπου κάθε επίπεδο ιεραρχίας είναι και ένας διαφορετικός πίνακας σε κάθε διάσταση. Για να φτάσουμε στο σημείο να υλοποιήσουμε ένα αστεροειδή σχήμα έπρεπε να έχουμε αρχικά τα δεδομένα σε απλή μορφή, με το τύπο τους και διαχωρισμένα με λεπτομέρεια. Το εργαλείο που χρησιμοποιήσαμε για να μας κάνει εξαγωγή των δεδομένων στη μορφή που θέλουμε έγινε με το Datawatch , ενώ η έξοδος των δεδομένων αυτών έγινε parsing σε ένα άλλο εργαλείο που υποστηρίζει ETL και αυτό ήταν το Alteryx. Μέσα από αυτό το εργαλείο έγιναν οι κατάλληλοι μετασχητισμοί από διάφορες λειτουργίες και στο τέλος δημιουργήθηκαν όλα αυτά τα συστατικά στοιχεία(διαστάσεις και facts) για να υλοποιηθεί ένα πλήρης αστεροειδή σχήμα.

Διάγραμμα 3.1

Διάγραμμα του αστεροειδούς σχήματος

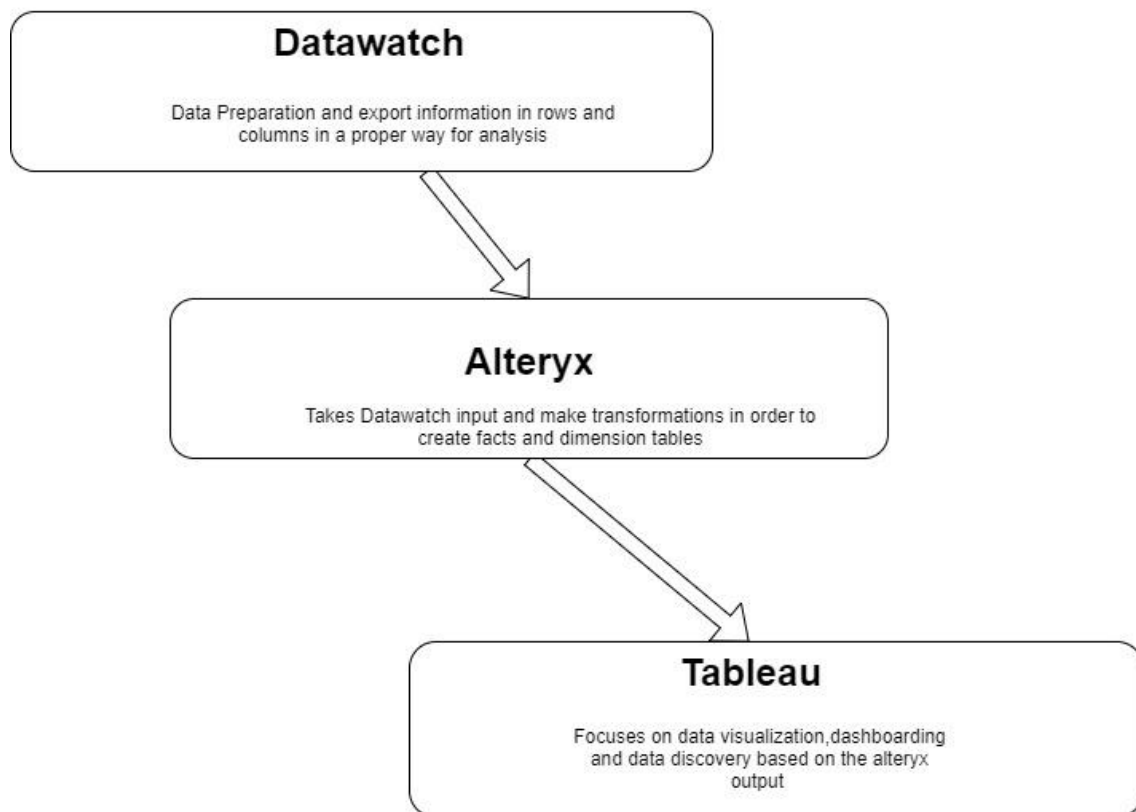


3.2. ΛΟΓΙΣΜΙΚΑ

Από κάτω(διάγραμμα 3.2.1) βλέπουμε το pipeline diagram που υλοποιείται για την OLAP σχεδίαση του συστήματος μας

Διάγραμμα 3.2.1

Pipeline διάγραμμα για olap



Πιο αναλυτικά τα εργαλεία που χρησιμοποιήθηκαν για την εργασία και αποτελούν την τελική υλοποίηση του συστήματος είναι τα :

(Datawatch, Alteryx, SQL Developer, Tableau)

1) Datawatch (<https://www.datawatch.com>)

2) Alteryx (<https://www.alteryx.com/>)

3) SQL Developer (<http://www.oracle.com/technetwork/developer-tools/sql-developer/overview/index.html>)

4) Tableau

(<https://www.tableau.com>)

Ας ξεκινήσουμε με το τι κάνει το κάθε ένα και ύστερα πως συνδέονται μεταξύ τους.

A) Το Datawatch είναι μια εφαρμογή που τρέχει σε windows μηχανήματα και έχει κάποιος τη δυνατότητα να κάνει drop αρχεία ή web pages , στη δική μας περίπτωση είχαμε excel file με διάφορα sheets και αυτόματα το εργαλείο εξάγει αυτή τη πληροφορία σε rows και columns σε μορφή τέτοια που αν κάνει κάποιος την κατάλληλη επεξεργασία τότε τα δεδομένα θα είναι έτοιμα για parsing και έτοιμα για ανάλυση.

Τα στάδια ανάλυσης των δεδομένων περιλαμβάνουν τόσο την επεξεργασία των στηλών δηλαδή πχ να γίνουν merge ή split κολόνες, να αλλαχτεί ο τύπο τους ή ακόμα να γίνουν clean κάποιοι αριθμητικοί χαρακτήρες σε μια συμβολοσειρά, αλλά ακόμα μπορούμε να εκτελέσουμε και κάποιους μετασχηματισμούς όπως

1) pivot

Να μετατρέψουμε τις γραμμές σε στήλες

2) unpivot

Να μετατρέψουμε τις στήλες σε γραμμές

3) transpose

Να κάνουμε swap τις γραμμές και τις στήλες

4) groupings και αθροίσματα

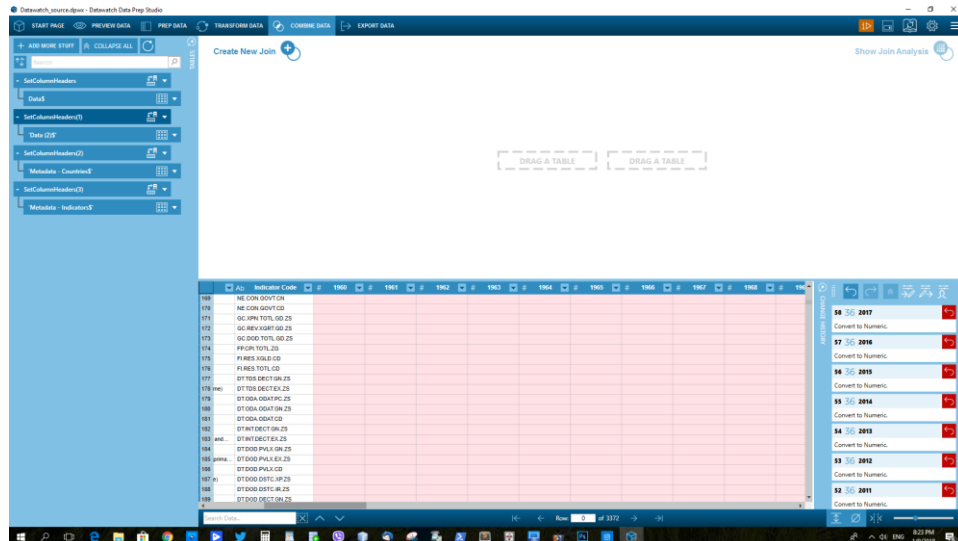
5) Να διαγράψουμε διπλότυπες εγγραφές

6) Να διαγράψουμε γραμμές με null τιμές, κενά και whitespace

Τέλος μπορείς να συνδυάσουμε πολλούς πίνακες μαζί αφού σου δίνει τη δυνατότητα του Join(εικόνα 3.2.2)

Εικόνα 3.2.2

Join λειτουργία στο εργαλείο Datawatch



Ένα πολύ χρήσιμο εργαλείο, κυρίως όταν θέλει κάποιος να χτίσει ένα datawarehouse από διαφορετικά data sources(διαφορετικά μεταξύ τους συστήματα) και θέλουμε ένα κοινό format στα δεδομένα είναι ιδιαίτερα χρήσιμο εργαλείο.

Οι κύριες ενέργειες που έγιναν στο συγκεκριμένο εργαλείο ήταν :

Ορισμός μια συγκεκριμένη γραμμής ως table header

[illegible]

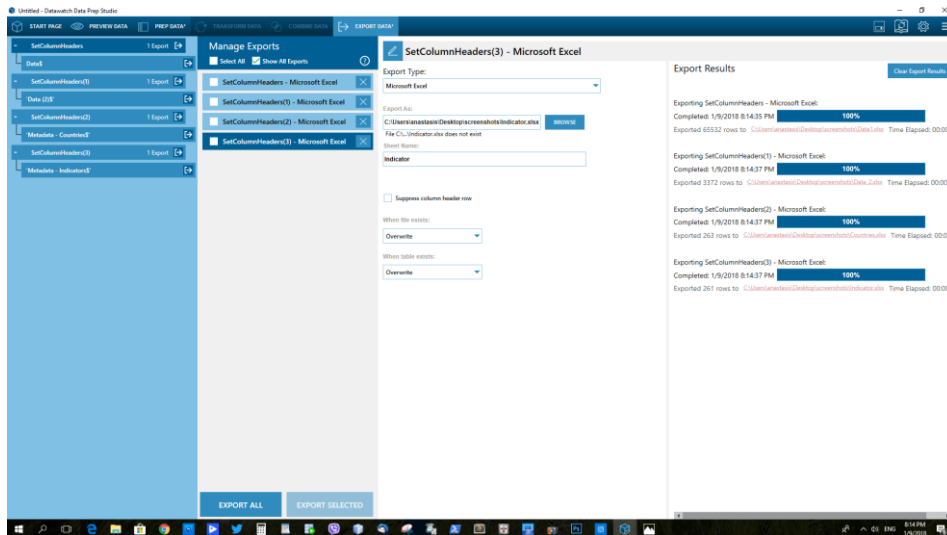
Μετατροπή μιας στήλης από συμβολοσειρά σε αριθμό

[illegible]

γ) Να γίνει ξεχωριστή εξαγωγή και ομαδοποίησης της κάθε ξεχωριστής οντότητας (εικόνα 3.2.5). Δηλαδή στο τέλος είχε φτιαχτεί τα τελικά αρχεία για επεξεργασία COUNTRIES, INDICATOR, DATA1, DATA2 που θα είναι έτοιμα για ανάλυση.

Εικόνα 3.2.5

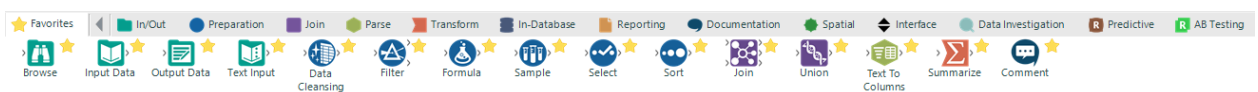
Διαχείριση και εξαγωγή αποτελεσμάτων



B) Το Alteryx είναι ένα ETL εργαλείο συμβατό με το tableau και έχει σκοπό τη δημιουργία των workflows και την επεξεργασία δεδομένων από μια αρχική σε μια τελική μορφή. Έχει διάφορες λειτουργίες που μπορεί να χρησιμοποιήσει κάποιος χρήστης με τις παρακάτω πιο δημοφιλείς(εικόνα 3.2.6) :

Εικόνα 3.2.6

Κύριες Λειτουργίες



Δηλαδή η γενική εικόνα ήταν να έχουμε τα δεδομένα εισόδου που έγιναν εξαγωγή από το datawatch, να γίνει ύστερα η κατάλληλη προετοιμασία μετασχηματισμών των δεδομένων με στόχο την υλοποίηση της OLAP αρχιτεκτονικής. Για έξοδο είχαμε την εγγραφή σε πίνακες στη βάση έτσι ώστε μετά από κάποιο εργαλείο στο Presentation/Visualization Layer να εξάγω τα δεδομένα.

Πιο συγκεκριμένα με το Alteryx φτιάξαμε τις ροές προκειμένου να δημιουργηθεί ένα data warehouse schema.

Δηλαδή δημιουργήσαμε όλες τις διαστάσεις και τους περισσότερους fact tables μέσα από κάποια operation που μας παρείχε το εργαλείο.

Γ) Ο SQL Developer είναι ένα εργαλείο της Oracle που επιτρέπει τη διαχείριση βάση δεδομένων μεταξύ πινάκων. Η έξοδος των workflows , έχει τη δημιουργία ή το επαναγράψιμο πινάκων που απεικονίζονται στο συγκεκριμένο εργαλείο. Από εκεί και πέρα , σε επίπεδο database , μπορούν να φτιαχτούν επιπλέον function, views, materialized view και άλλα διάφορα που επιτρέπουν επιπλέον λειτουργικότητα.

Στη δική μας περίπτωση στήσαμε μια τοπική Oracle DB 11g , στο μηχάνημα.

Δ) Το εργαλείο Tableau δίνει τη δυνατότητα να υλοποιήσει κάποιος τα data marts, δηλαδή να ενώσει τους fact πίνακες με τις διαστάσεις και συνεπώς να απεικονίσει όλα τα δεδομένα που έχουν γίνει retrieve από το data warehouse. Προσφέρει λειτουργίες ένωσης πινάκων με όλα τα ήδη των joins που υπάρχουν ή ακόμα και union. Βέβαια στο Presentation Layer ενδεχομένως να μην θέλουμε να εμφανίσουμε όλα τα δεδομένα οπότε μπορούμε να κάνουμε hide κάποια πεδία όπως για παράδειγμα Primary Keys ή ακόμα να δημιουργήσουμε τους δικούς μας φακέλους και να οργανώσουμε τη δική μας δομή εμφάνισης είτε με βάση το data source είτε με βάση τους φακέλους.

Επιπρόσθετα μπορούμε να ορίσουμε ιεραρχίες και έτσι να μπορούμε να κάνουμε drill in και drill out στα dimension ανάλογα τα επίπεδα που έχουμε υλοποιήσει.

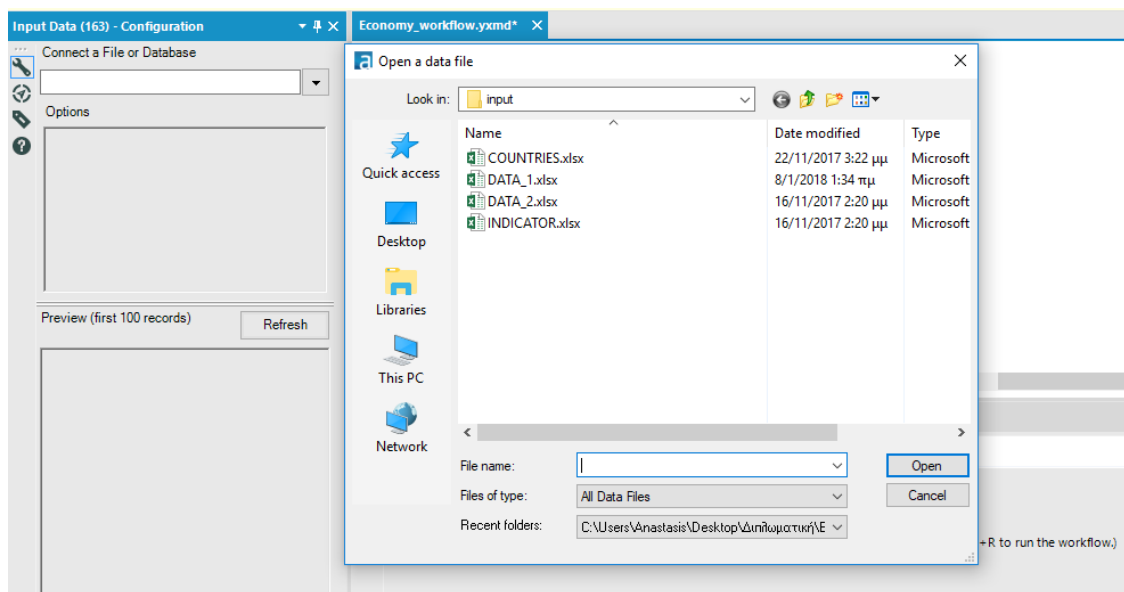
Συνεπώς το tableau είναι το εργαλείο που πλέον έχει όλη τη πληροφορία έτοιμη και αυτό με τη σειρά του είναι σε θέση να απεικονίσει χρήσιμη πληροφορία στο χρήστη, μέσω γραφημάτων και dashboards.

3.3. ΣΥΝΔΕΣΙΜΟΤΗΤΑ ΟΛΩΝ ΑΥΤΩΝ ΤΩΝ ΕΡΓΑΛΕΙΩΝ

Όπως είπαμε το εργαλείο Datawatch δημιουργεί το σωστά δομημένο input, αυτό με τη σειρά του γίνεται είσοδος στο Alteryx ETL tool , μέσω του Input Data σαν action(εικόνα 3.3.1)

Εικόνα 3.3.1

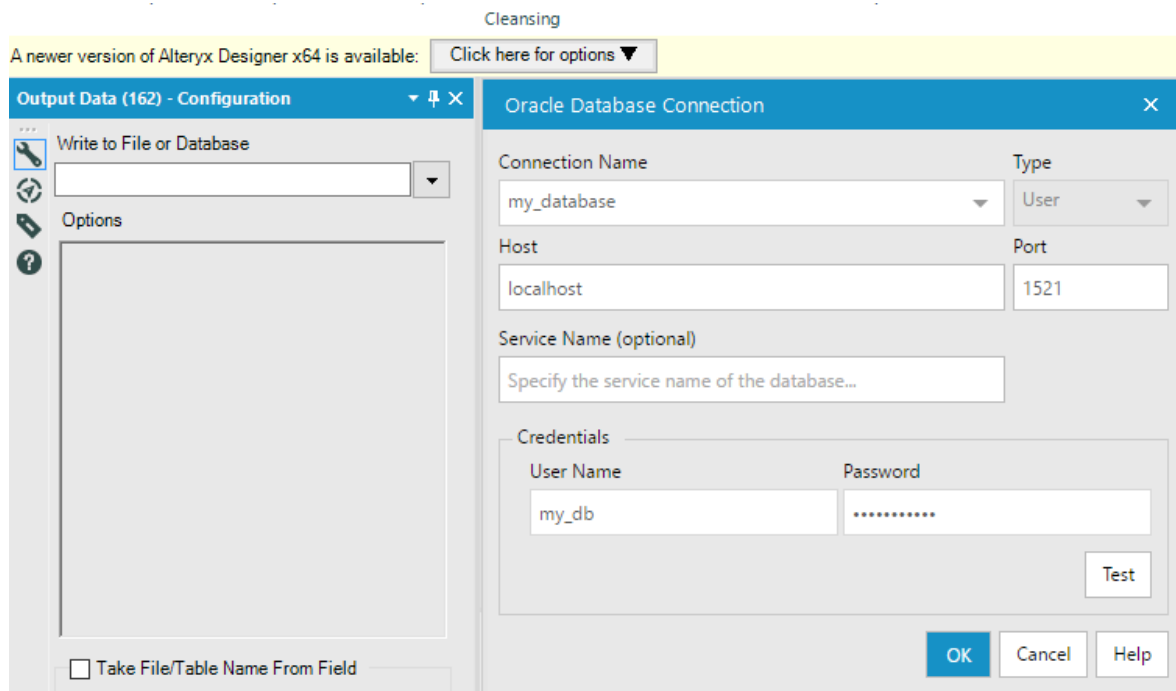
Εισαγωγή δεδομένων στο εργαλείο Alteryx για να τροφοδοτήσουν τις ροές



και ύστερα μέσα από εκεί αφού δημιουργήσουμε τις τελικές ροές επιλέγουμε το Output data σαν τελευταίο Action(εικόνα 3.3.2)

Εικόνα 3.3.2

Γράψιμο των αποτελεσμάτων των ροών σε Oracle βάση δεδομένων

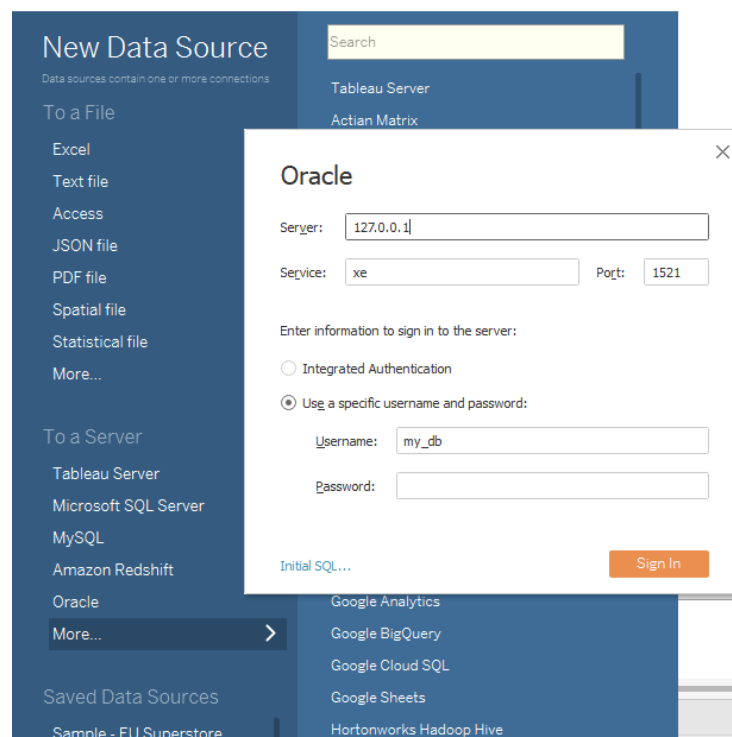


Και γράφουμε στην Oracle κάνοντας τα drop εκεί.

Ύστερα το tableau μετά δίνει τη δυνατότητα τη δημιουργία ενός νέου data source και τη σύνδεση σε Oracle server(εικόνα 3.3.3)

Εικόνα 3.3.3

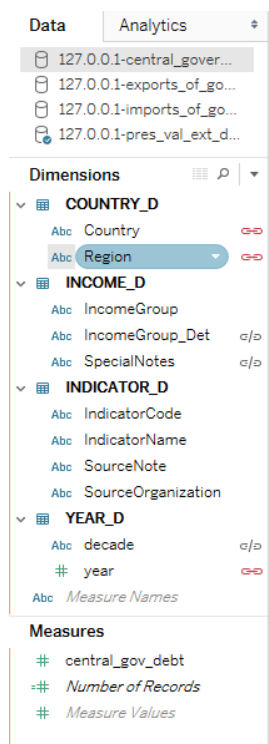
Δημιουργία νέας πηγής δεδομένων και σύνδεση σε oracle server



Αν το μοντέλο δεδομένο είναι group-αρισμένο ανά data source table, τότε θα φαίνεται κάπως έτσι(εικόνα 3.3.4)

Εικόνα 3.3.4

Το μοντέλο δεδομένων group-αρισμένο ανά data source



ΚΕΦ.4: ΥΛΟΠΟΙΗΣΗ

Υπάρχουν δύο συνήθη data warehouse σχήματα και αυτά είναι το star και το snowflake. Στην υλοποίηση του δικού μας συστήματος υλοποιήθηκε το star schema , όπου στο κέντρο έχουμε ένα fact table με τα μετρήσιμα και γύρω του , όλες τις διαστάσεις.

Αξίζει να σημειώσουμε ότι δοκιμάσαμε 2 προσεγγίσεις για την κατασκευή των fact πινάκων, όπου στη μια περίπτωση χρησιμοποιήσαμε κλασσικό etl , ενώ στην άλλη είχαμε Materilized view. Στη περίπτωση των διαστάσεων χρησιμοποιήσαμε κανονικά ροές.

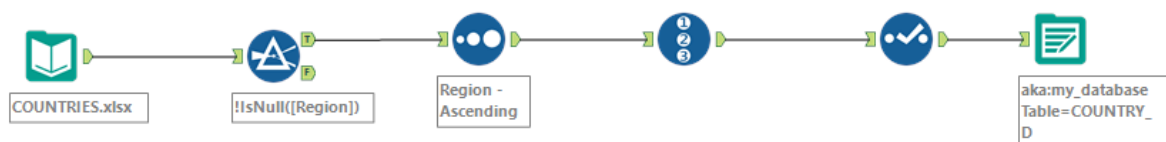
4.1. ΕΞΑΓΩΓΗ, ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΚΑΙ ΦΟΡΤΩΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Ας δούμε πρώτα τη δημιουργία των διαστάσεων από το εργαλείο Alteryx:

1) Δημιουργία διάστασης COUNTRY_D(εικόνα 4.1.1) και τα αποτελέσματα της(εικόνα 4.1.2) , βλέπουμε όλες τις χώρες αναλυτικά και τη περιοχή που ανήκει η κάθε μία

Εικόνα 4.1.1

Δημιουργία διάστασης Χώρα



Εικόνα 4.1.2

Η έξοδος των αποτελεσμάτων από τη ροή της διάστασης της Χώρας

Record #	CountryId	CountryCode	Region	TableName
1	1	ASM	East Asia & Pacific	American Samoa
2	2	AUS	East Asia & Pacific	Australia
3	3	BRN	East Asia & Pacific	Brunei Darussalam
4	4	CHN	East Asia & Pacific	China
5	5	FJI	East Asia & Pacific	Fiji
6	6	FSM	East Asia & Pacific	Micronesia, Fed. Sts.
7	7	GUM	East Asia & Pacific	Guam
8	8	HKG	East Asia & Pacific	Hong Kong SAR, China
9	9	IDN	East Asia & Pacific	Indonesia
10	10	JPN	East Asia & Pacific	Japan
11	11	KHM	East Asia & Pacific	Cambodia
12	12	KIR	East Asia & Pacific	Kiribati
13	13	KOR	East Asia & Pacific	Korea, Rep.
14	14	LAO	East Asia & Pacific	Lao PDR
15	15	MAC	East Asia & Pacific	Macao SAR, China
16	16	MHL	East Asia & Pacific	Marshall Islands
17	17	MMR	East Asia & Pacific	Myanmar
18	18	MNG	East Asia & Pacific	Mongolia
19	19	MNP	East Asia & Pacific	Northern Mariana Islands
20	20	MYS	East Asia & Pacific	Malaysia
21	21	NCL	East Asia & Pacific	New Caledonia
22	22	NRU	East Asia & Pacific	Nauru

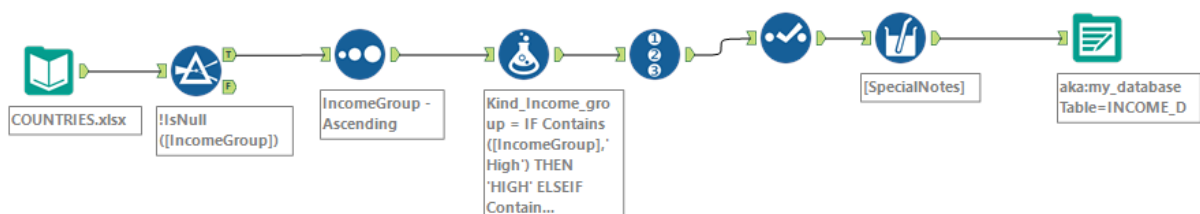
2) Δημιουργία διάστασης INCOME_D(εικόνα 4.1.3) και τα αποτελέσματα της(εικόνα 4.1.4)

Εδώ πρόκειται για τις ομάδες εισοδήματος που υπάρχουν για διάφορες χώρες.

Εδώ αξίζει να σημειώσουμε τη δημιουργία ιεραρχίας, σε ομάδες εισοδήματος, δηλαδή έχουμε το 1ο επίπεδο σε χαμηλό, μεσαίο, μεγάλο ενώ σε 2ο επίπεδο το μεσαίο μπορεί να δώσει επιπλέον λεπτομέρεια(βλέπε εικόνα από κάτω)

Εικόνα 4.1.3

Δημιουργία διάστασης της εισοδηματικής κατάστασης



Εικόνα 4.1.4

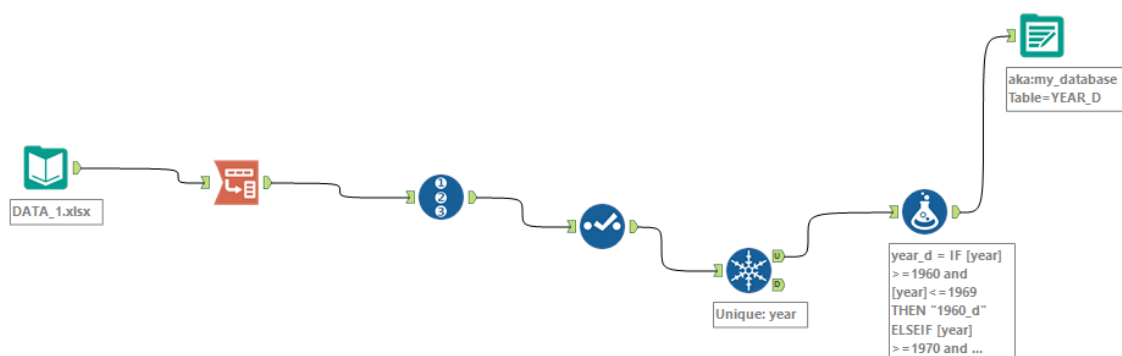
Η έξοδος των αποτελεσμάτων από τη ροή της διάστασης της εισοδηματικής κατάστασης

Record #	IncomeGroupId	IncomeGroup	SpecialNotes	Kind_Income_group	CountryCode
106	26782	Low income	Tanzania reports using a blend of SNA 1993 an...	LOW	TZA
107	26783	Low income	Fiscal year end: June 30; reporting period for n...	LOW	UGA
108	26784	Low income	Fiscal year end June 30 reporting period for na...	LOW	ZWE
109	26785	Lower middle in...	[Null]	MIDDLE	AGO
110	26786	Lower middle in...	[Null]	MIDDLE	ARM
111	26787	Lower middle in...	Fiscal year end: June 30; reporting period for n...	MIDDLE	BGD
112	26788	Lower middle in...	[Null]	MIDDLE	BOL
113	26789	Lower middle in...	[Null]	MIDDLE	BTN
114	26790	Lower middle in...	[Null]	MIDDLE	CIV
115	26791	Lower middle in...	[Null]	MIDDLE	CMR
116	26792	Lower middle in...	[Null]	MIDDLE	COG
117	26793	Lower middle in...	Cabo Verde is the name for the country previo...	MIDDLE	CPV
118	26794	Lower middle in...	[Null]	MIDDLE	DJI
119	26795	Lower middle in...	Fiscal year end: June 30; reporting period for n...	MIDDLE	EGY
120	26796	Lower middle in...	Fiscal year ends on September 30; reporting p...	MIDDLE	FSM
121	26797	Lower middle in...	Value added data in constant prices before 20...	MIDDLE	GEO
122	26798	Lower middle in...	[Null]	MIDDLE	GHA
123	26799	Lower middle in...	[Null]	MIDDLE	GTM
124	26800	Lower middle in...	[Null]	MIDDLE	HND
125	26801	Lower middle in...	Fiscal year end: March 31; reporting period for...	MIDDLE	IDN
126	26802	Lower middle in...	Fiscal year end: March 31; reporting period for...	MIDDLE	IND
127	26803	Lower middle in...	[Null]	MIDDLE	JOR

3) Δημιουργία διάστασης των χρονολογιών TIME_D(εικόνα 4.1.5) και τα αποτελέσματα της(εικόνα 4.1.6)

Εικόνα 4.1.5

Δημιουργία διάστασης χρονολογιών



Εικόνα 4.1.6

Η έξοδος των αποτελεσμάτων της ροής της διάστασης των χρονολογιών

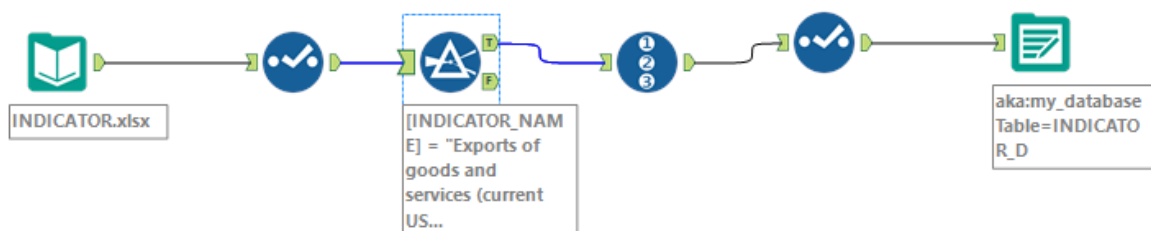
Record #	YearID	year	year_d
8	26627	1967	1960_d
9	26628	1968	1960_d
10	26629	1969	1960_d
11	26630	1970	1970_d
12	26631	1971	1970_d
13	26632	1972	1970_d
14	26633	1973	1970_d
15	26634	1974	1970_d
16	26635	1975	1970_d
17	26636	1976	1970_d
18	26637	1977	1970_d

Εδώ έχουμε όλες τις χρονολογίες ενώ αξίζει να επισημάνουμε τη δημιουργία ιεραρχίας σε 1ο επίπεδο τις δεκαετίες και σε 2ο αναλυτικά το κάθε έτος ξεχωριστά.

4) Δημιουργία διάστασης INDICATOR_D(εικόνα4.1.7) και τα αποτελέσματα της (εικόνα 4.1.8), δηλαδή πληροφορίες για του συγκεκριμένους οικονομικούς δείκτες που θα μελετήσουμε

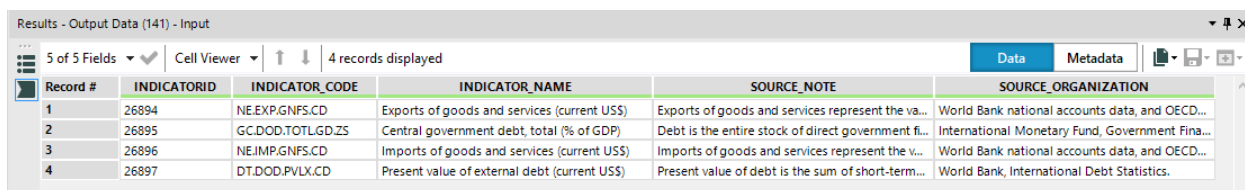
Εικόνα 4.1.7

Δημιουργία διάστασης οικονομικών δεικτών



Εικόνα 4.1.8

Η έξοδος των αποτελεσμάτων της ροής της διάστασης των οικονομικών δεικτών



Results - Output Data (141) - Input

5 of 5 Fields | Cell Viewer | 4 records displayed

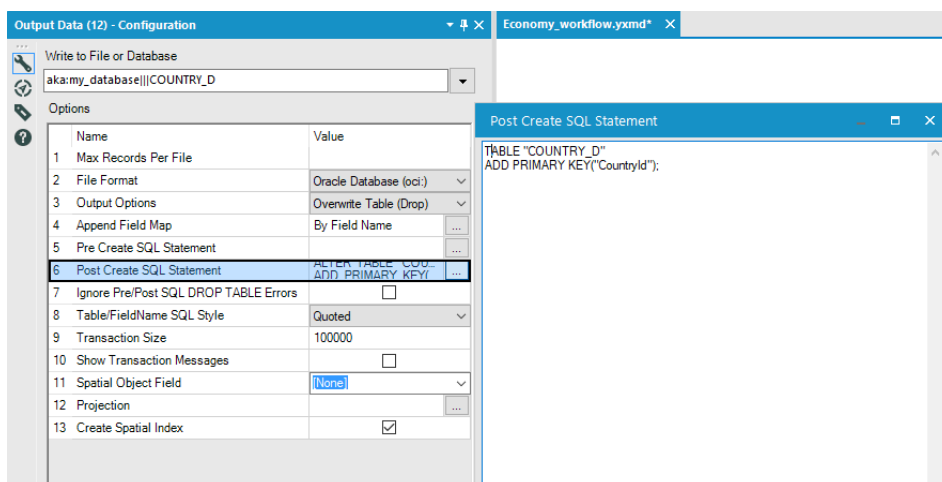
Record #	INDICATORID	INDICATOR_CODE	INDICATOR_NAME	SOURCE_NOTE	SOURCE_ORGANIZATION
1	26894	NE.EXP.GNFS.CD	Exports of goods and services (current US\$)	Exports of goods and services represent the va...	World Bank national accounts data, and OECD...
2	26895	GC.DOD.TOTLGD.ZS	Central government debt, total (% of GDP)	Debt is the entire stock of direct government fi...	International Monetary Fund, Government Fina...
3	26896	NE.IMP.GNFS.CD	Imports of goods and services (current US\$)	Imports of goods and services represent the v...	World Bank national accounts data, and OECD...
4	26897	DT.DOD.PVLX.CD	Present value of external debt (current US\$)	Present value of debt is the sum of short-term...	World Bank, International Debt Statistics.

Σε κάθε δημιουργία διάστασης έχει μπει ένα Post Create sql statement(εικόνα 4.1.9) για να μπαίνουν τα Primary Keys αυτόματα.

Πχ για τη διάσταση της χώρας

Εικόνα 4.1.9

Ένα Post create sql statement για τον πίνακα της διάστασης της χώρας



Output Data (12) - Configuration | Economy_workflow.yxmd*

Write to File or Database
aka:my_database|||COUNTRY_D

Options

Name	Value
1 Max Records Per File	
2 File Format	Oracle Database (oci)
3 Output Options	Overwrite Table (Drop)
4 Append Field Map	By Field Name
5 Pre Create SQL Statement	
6 Post Create SQL Statement	ALTER TABLE COUNTRY_D ADD PRIMARY KEY(CountryId);
7 Ignore Pre/Post SQL DROP TABLE Errors	<input type="checkbox"/>
8 Table/FieldName SQL Style	Quoted
9 Transaction Size	100000
10 Show Transaction Messages	<input type="checkbox"/>
11 Spatial Object Field	None
12 Projection	
13 Create Spatial Index	<input checked="" type="checkbox"/>

Post Create SQL Statement

```
TABLE "COUNTRY_D"  
ADD PRIMARY KEY("CountryId");
```

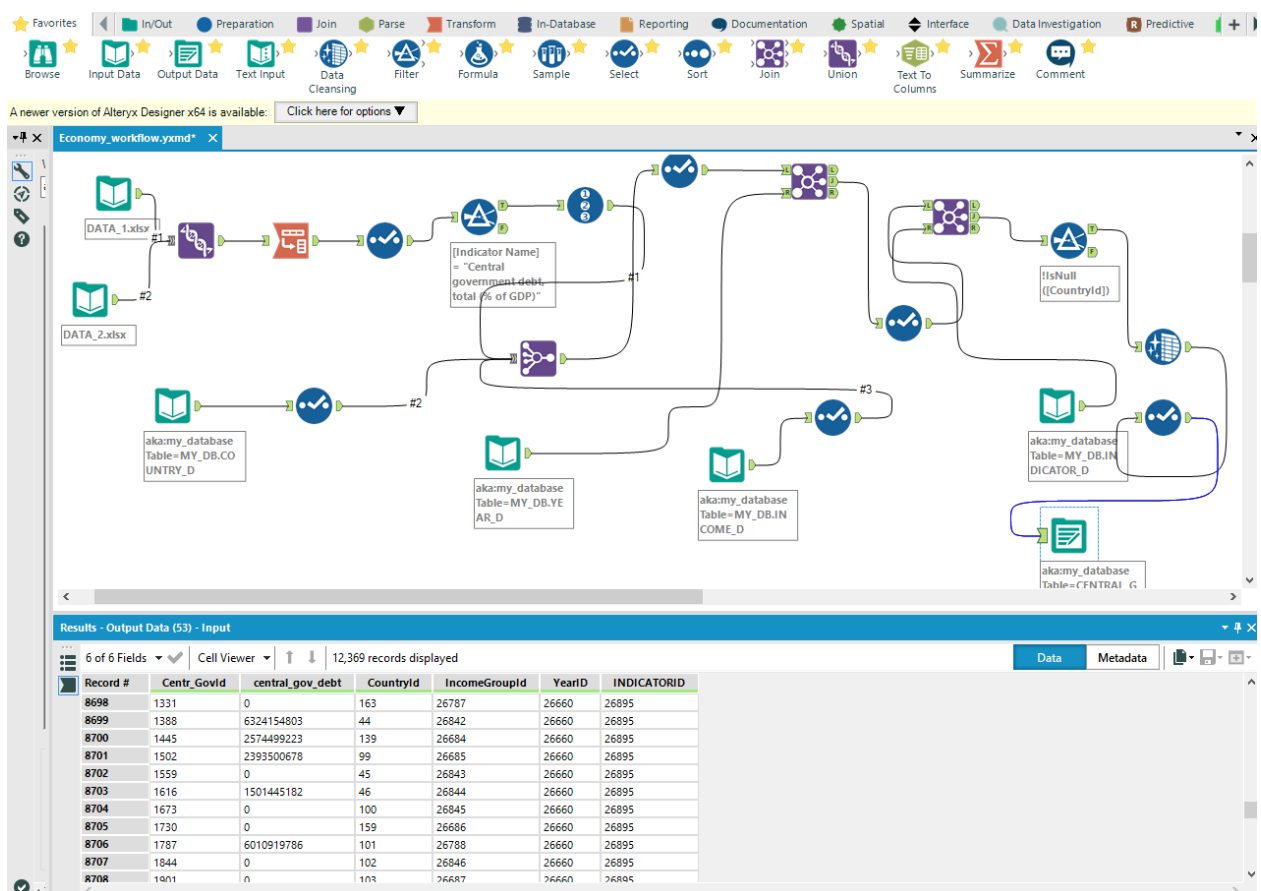
Ας δούμε τώρα τη δημιουργία των fact tables(εργαλείο Alteryx)

-Μελέτη περίπτωσης Α(με τη δημιουργία κλασικού ETL)

1) Δημιουργία fact για το κεντρικό δημόσιο χρέος(εικόνα 4.1.10)

Εικόνα 4.1.10

Η ροή για τον fact πίνακα του κεντρικού δημόσιου χρέους



Κάνοντας τους κατάλληλους μετασχηματισμούς και συμπεριλαμβάνοντας τα primary keys των διαστάσεων καταλήγουμε να έχουμε τις παραπάνω εγγραφές στο πίνακα μας.

Και σε αυτή τη περίπτωση έχουμε post create sql statements για να μπαίνουν τα primary και τα foreign keys στο πίνακα αυτόματα, χωρίς να χρειάζεται να μούνε custom από συστήματα διαχείριση βάσεων δεδομένων.

Επίσης να αναφέρουμε ότι στο αρχικό building , δηλαδή στο union της εισόδου(data1.xlsx, data2.xlsx) στη παραπάνω εικόνα, θα μπορούσαμε να βάλουμε όσα excel θέλαμε και να

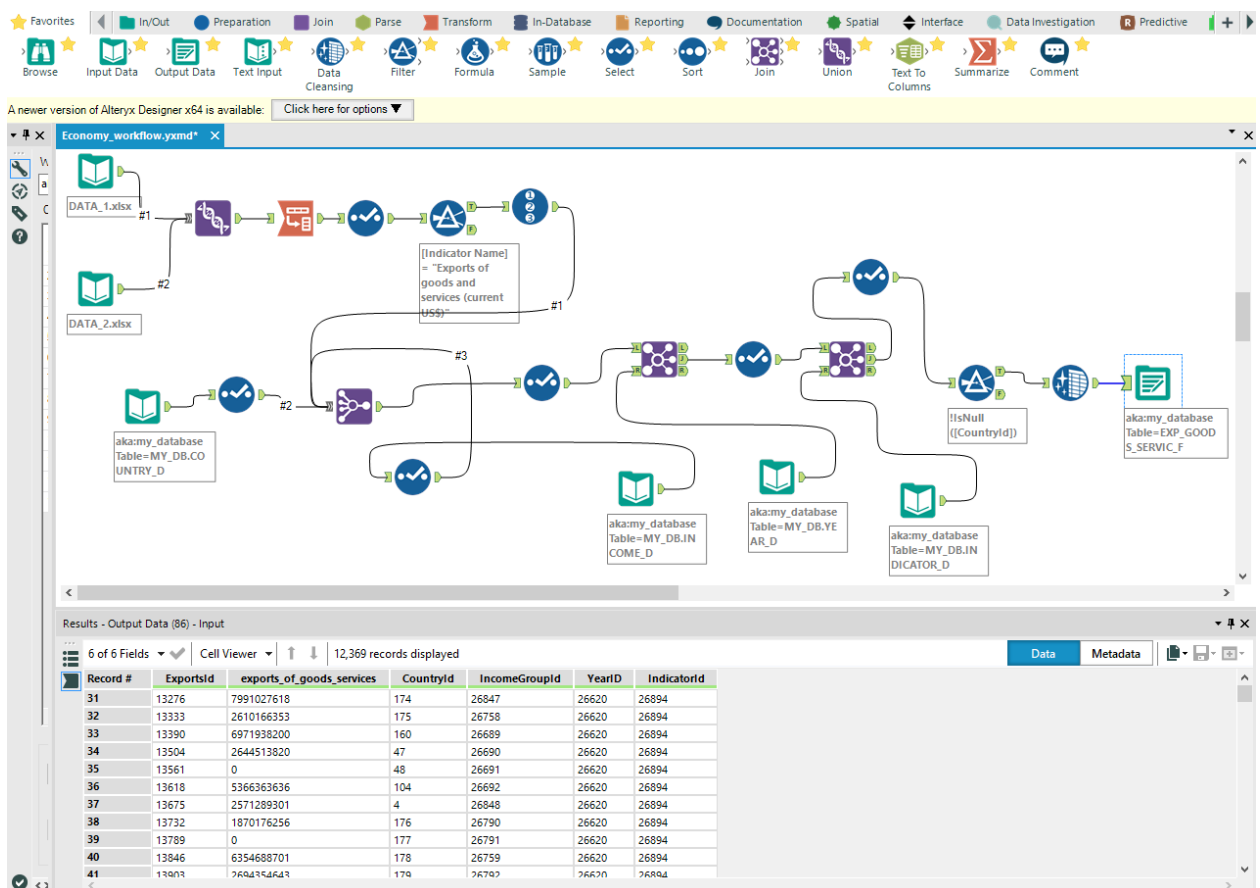
βγάλουμε αποτελέσματα και για νέα έτη στο μέλλον χωρίς να χρειάζεται κάποια παρέμβαση το visualization επίπεδο.

Παρόμοια δουλειά έχει γίνει και για τους υπόλοιπους fact, δηλαδή για τα :

2) Εξαγωγές αγαθών και υπηρεσιών σε δολάρια (εικόνα 4.1.11)

Εικόνα 4.1.11

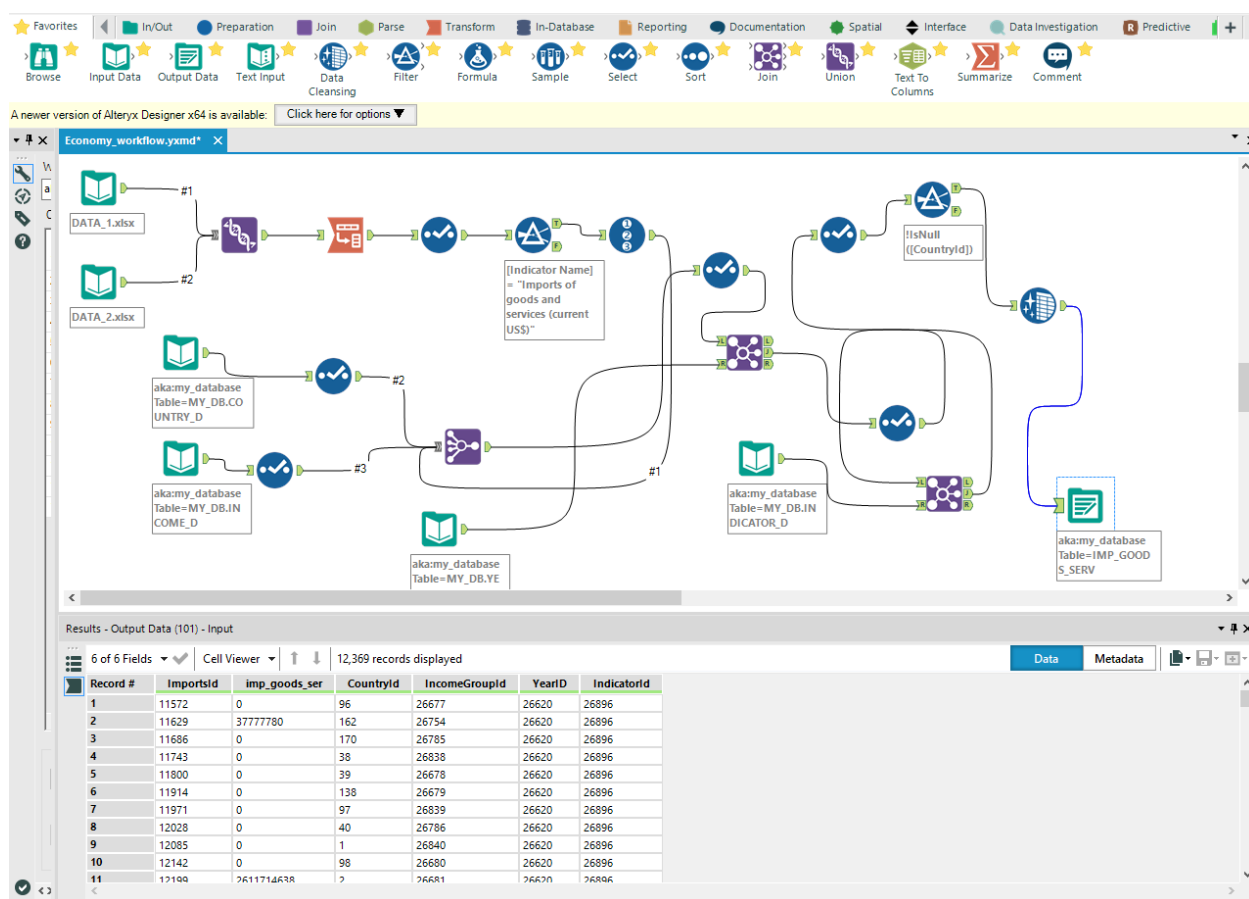
Η ροή για τον fact πίνακα των εξαγωγών αγαθών και υπηρεσιών



3) Εισαγωγές αγαθών και υπηρεσιών(εικόνα 4.1.12)

Εικόνα 4.1.12

Η ροή για τον fact πίνακα των εισαγωγών αγαθών και υπηρεσιών



-Μελέτη περίπτωσης Β(με materialized view στην oracle)

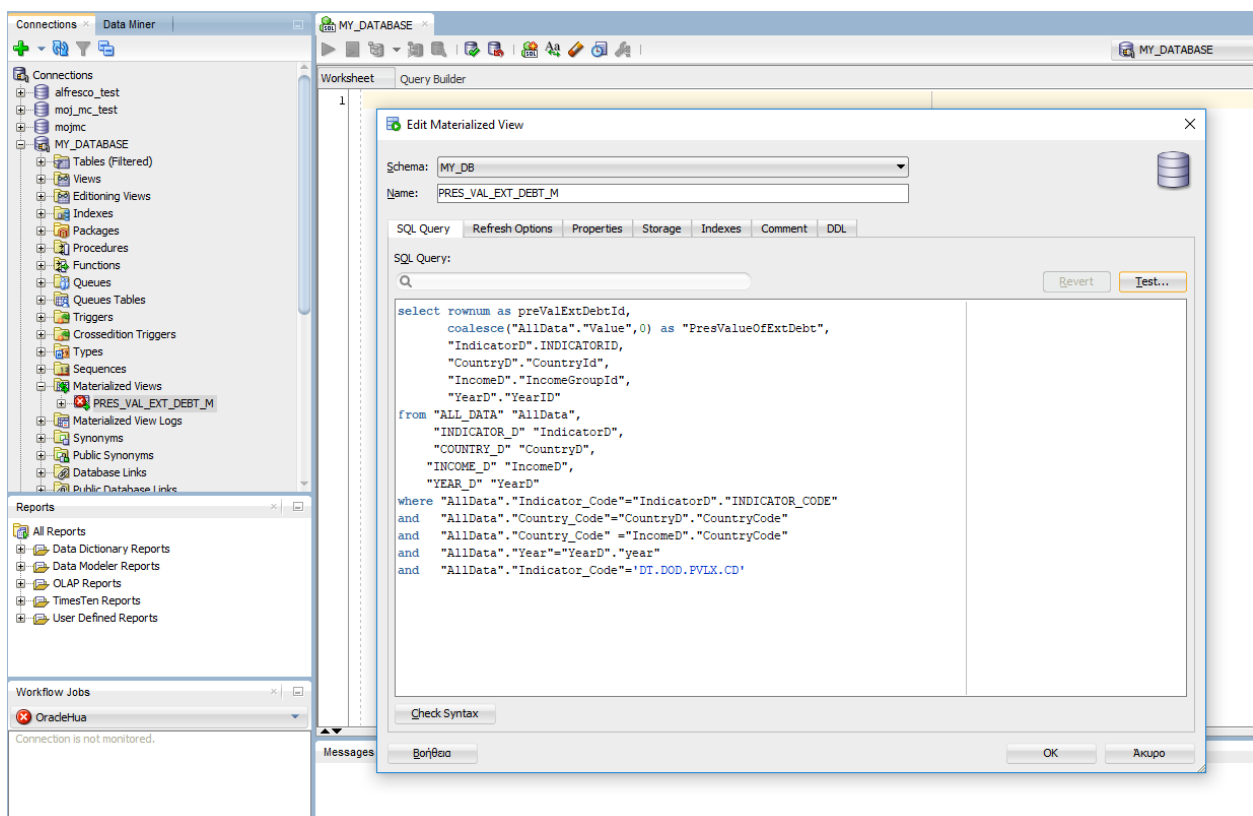
Χρησιμοποιήσαμε και εναλλακτικά αυτή την επιλογή. Το materialized view είναι ένας πίνακας στο δίσκο και περιλαμβάνει τα αποτελέσματα του ερωτήματος που έχουν κανονιστεί να τρέξουν

σε ένα συγκεκριμένο χρονικό διάστημα. Ο σκοπός του είναι να αυξήσουμε τη ταχύτητα που τρέχει ένα query και είναι πολύ πιο γρήγορο από ένα view (ακόμα και αν έχει indexes) αφού τη στιγμή που θα ζητήσουμε τα δεδομένα μας σε επίπεδο warehousing τα δεδομένα έχουν ήδη γίνει fetched.

Για τη δική μας περίπτωση του fact table που αντιστοιχεί στη τωρινή τιμή εξωτερικού χρέους(σε δολάρια) υλοποιήσαμε το παρακάτω materialized view(εικόνα 4.1.13) με refresh option sysdate +1, δηλαδή να τρέχει κάθε φορά στο τέλος της ημέρας.

Εικόνα 4.1.13

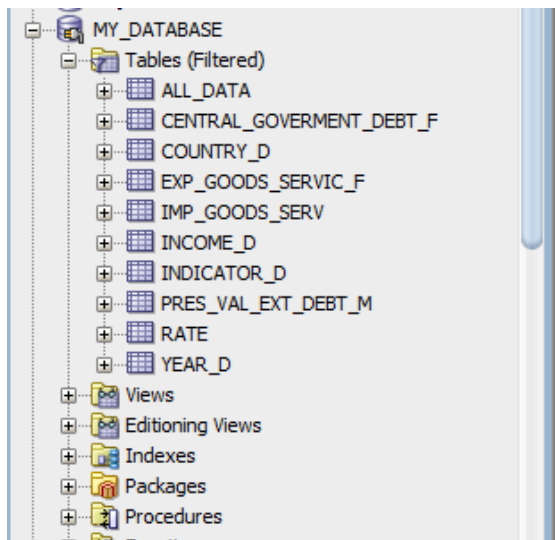
Ο κώδικας που υλοποιεί το Materialized view



Αφού ολοκληρώσουμε όλη τη διαδικασία της κατασκευής του data warehouse , θα δείξουμε την τελική εικόνα της εξόδου από τα workflows(εικόνα 4.1.14) αλλά και από το materialized view στη βάση :

Εικόνα 4.1.14

Η εικόνα της βάσης μετά το πέρας της εκτέλεσης των ροών



Όπως βλέπουμε οι πίνακες που τελειώνουν σε _D είναι όλες οι διαστάσεις μας, οι πίνακες που τελειώνουν σε _F είναι όλοι οι fact πίνακες από τις ροές μας στο alteryx , ενώ το PRES_VAL_EXT_DEBT_M είναι ο πίνακας που έβγαλε το materialized view.

Ένα ενδεικτικό παράδειγμα ενός fact πίνακα(εξαγωγές αγαθών και υπηρεσιών) δημιουργημένο από μια ροή είναι από κάτω(πίνακας 4.1.15) .

Βλέπουμε ότι πρώτα έχουμε το Primary Key του πίνακα(ExportId), μετά το μετρήσιμο(exports_of_goods_services) και μετά όλα τα ξένα κλειδιά των διαστάσεων(CountryId, IncomeGroupId, YearId, IndicatorId)

Πίνακας 4.1.15

Ο fact πίνακας των εξαγωγών αγαθών με τα πρωτεύοντα και ξένα κλειδιά

	ExportsId	exports_of_goods_services	CountryId	IncomeGroupId	YearID	IndicatorId
1371	16018	20082873495	58	26704	26626	26894
1372	16075	0,0	59	26797	26626	26894
1373	16132	3109609679	185	26798	26626	26894
1374	16189	0,0	60	26705	26626	26894
1375	16246	0,0	186	26763	26626	26894
1376	16303	1599187994	187	26764	26626	26894
1377	16360	0,0	188	26765	26626	26894
1378	16417	3028333333	189	26858	26626	26894
1379	16474	8877505772	61	26706	26626	26894
1380	16531	0,0	113	26859	26626	26894
1381	16588	0,0	62	26707	26626	26894
1382	16645	263400000	114	26799	26626	26894
1383	16702	0,0	7	26708	26626	26894
1384	16759	1257647059	115	26860	26626	26894
1385	16873	1889960077	8	26709	26626	26894
1386	16930	157400000	116	26800	26626	26894
1387	17044	0,0	63	26861	26626	26894
1388	17101	0,0	117	26766	26626	26894
1389	17158	0,0	64	26710	26626	26894
1390	17443	0,0	9	26801	26626	26894
1391	17557	0,0	65	26711	26626	26894
1392	17614	1894349461	165	26802	26626	26894
1393	17728	1072019668	66	26712	26626	26894
1394	17785	1106379248	143	26862	26626	26894
1395	17842	0,0	144	26863	26626	26894
1396	17899	2065871779	67	26713	26626	26894
1397	17956	7936666667	145	26714	26626	26894
1398	18013	1000000000	68	26715	26626	26894

Αν θέλαμε να ομαδοποιήσουμε τα επίπεδα στο οποία δουλεύαμε, θα λέγαμε ότι μέχρι τώρα έχουμε φτάσει μέχρι το physical layer. Στην επόμενη φάση θα πρέπει να δημιουργήσουμε ξεχωριστές θεματικές περιοχές/κύβους που θα αναπαραχθούν από το εργαλείο tableau.

Δηλαδή αυτή τη στιγμή έχουμε όλο το datawarehouse, δηλαδή όλο το back end κομμάτι του συστήματος αλλά από εκεί και πέρα μας μένει να ορίσουμε όλα τα data marts, τους κύβους δηλαδή.

Κάθε κύβος θα είναι ένα ξεχωριστό Data Source γιατί το εργαλείο δεν επιτρέπει να έχεις πίνακες χωρίς να τους έχεις ενώσει. Συνεπώς έχουμε φτιάξει 4 πηγές δεδομένων και κάθε μία κάνει join τον fact με τις διαστάσεις. Οι διαστάσεις είναι κοινές, οπότε αυτό που αλλάζει είναι ο fact που φέρνουμε κάθε φορά.

4.2. ΔΗΜΙΟΥΡΓΙΑ ΚΥΒΩΝ

Στη παρακάτω εικόνα (εικόνα 4.2.1) βλέπουμε αναλυτικά την υλοποίηση του κύβου για τις εξαγωγές αγαθών και υπηρεσιών

Εικόνα 4.2.1

Υλοποίηση κύβου για τις εξαγωγές αγαθών

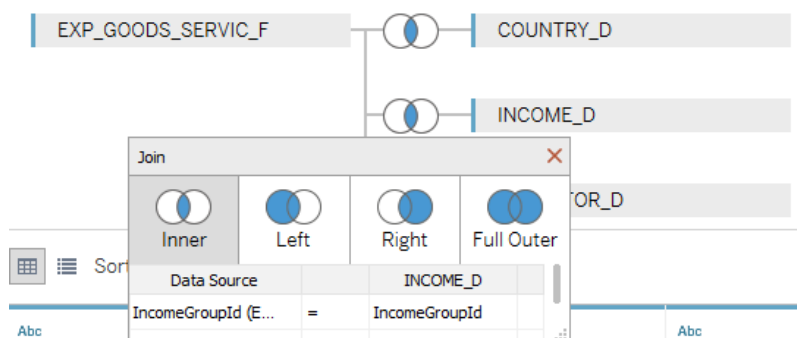
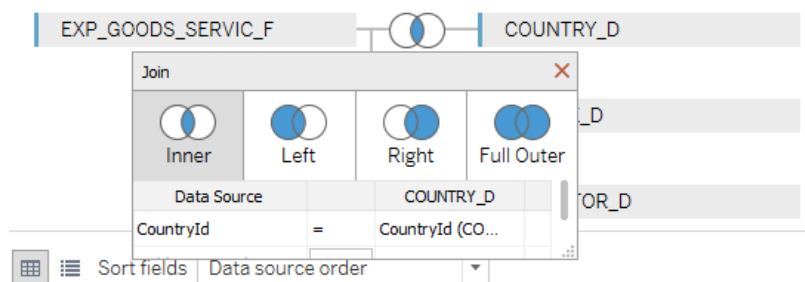
The screenshot displays a data tool interface for creating a cube. The top bar shows the connection '127.0.0.1-exports_of_goods_and_services' with 'Live' and 'Extract' options. The left sidebar lists connections and a schema 'MY_DB' containing tables like 'ALL_DATA', 'CENTRAL_GO...ENT_DEBT_F', 'COUNTRY_D', 'EXP_GOODS_SERVIC_F', 'IMP_GOODS_SERV', 'INCOME_D', 'INDICATOR_D', 'PRES_VAL_EXT_DEBT_M', 'RATE', 'YEAR_D', 'New Custom SQL', and 'New Union'. The central diagram illustrates the cube structure: a fact table 'EXP_GOODS_SERVIC_F' is joined with three dimension tables: 'COUNTRY_D', 'INCOME_D', and 'INDICATOR_D'. The bottom table view shows the resulting data with 12,369 rows. The table has columns: 'COUNTRY_D', 'COUNTRY_D', '#', 'INCOME_D', 'SpecialNotes', 'INCOME_D', 'INDICATOR_D', and 'INDICATOR_D'. The data rows show various regions and countries with their respective export values and income groups.

Abc COUNTRY_D	Abc COUNTRY_D	#	Abc INCOME_D	Abc SpecialNotes	Abc INCOME_D	Abc INDICATOR_D	Abc INDICATOR_D
Region	Country	exports_of_goods...	IncomeGroup_Det		IncomeGroup	INDICATOR_CODE	INDICATOR_CODE
Sub-Saharan Africa	Botswana	7.991.027.618,00	Upper middle income	Fiscal year end: Marc...	MIDDLE	NE.EXP.GNFS.CD	Exp...
Sub-Saharan Africa	Central African Repub...	2.610.166.353,00	Low income	The source of nationa...	LOW	NE.EXP.GNFS.CD	Exp...
North America	Canada	6.971.938.200,00	High income	Fiscal year end: Marc...	HIGH	NE.EXP.GNFS.CD	Exp...
Europe & Central Asia	Switzerland	2.644.513.820,00	High income	null	HIGH	NE.EXP.GNFS.CD	Exp...
Europe & Central Asia	Channel Islands	0,00	High income	null	HIGH	NE.EXP.GNFS.CD	Exp...
Latin America & Carib...	Chile	5.366.363.636,00	High income	null	HIGH	NE.EXP.GNFS.CD	Exp...
East Asia & Pacific	China	2.571.289.301,00	Upper middle income	On 1 July 1997 China r...	MIDDLE	NE.EXP.GNFS.CD	Exp...
Sub-Saharan Africa	Cote d'Ivoire	1.870.176.256,00	Lower middle income	null	MIDDLE	NE.EXP.GNFS.CD	Exp...

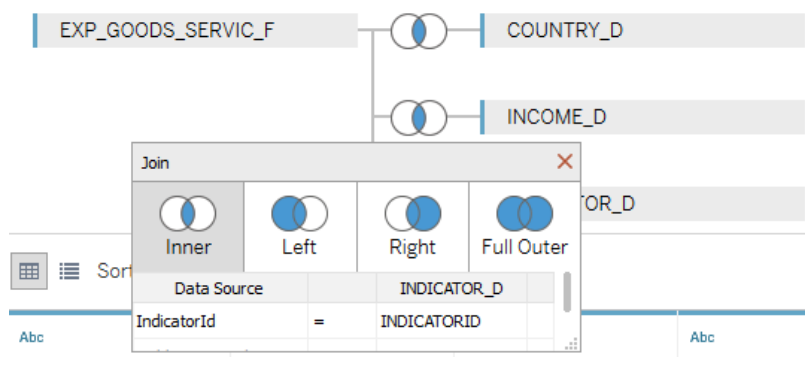
Όπως βλέπουμε κάνουμε join τον Fact(EXP_GOODS_SERFICE_F) με τις διαστάσεις (COUNTRY_D, INCOME_D, INDICATOR_D, YEAR_D), όπου το κάθε primary key των διαστάσεων ενώνεται με τα foreign keys του fact(εικόνα 4.2.2) .

Εικόνα 4.2.2

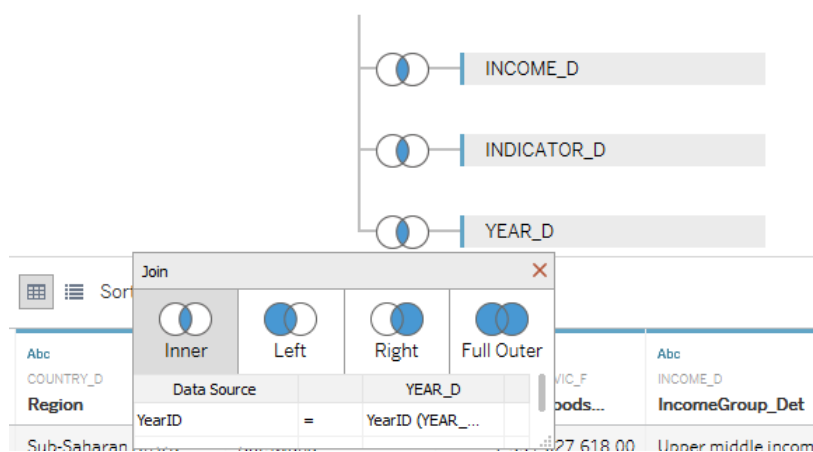
Δημιουργία του κύβου χρησιμοποιώντας όλες τις διαστάσεις και τα facts



127.0.0.1-exports_of_goods_and_services



127.0.0.1-exports_of_goods_and_services



Τα αποτελέσματα είναι αυτά που θέλουμε δηλαδή να έχουμε όλα τα μετρήσιμα του fact και τα πεδία των διαστάσεων. Βεβαία στο μοντέλο δεδομένων φτιάχνουμε την εμφάνιση όπως θέλουμε, δηλαδή κάνουμε hide πεδία όπως πρωτεύοντα κλειδιά και κάνουμε μετονομασία σε πιο ρεαλιστικά ονόματα, για να μπορεί ο χρήστης να διαβάσει με ευκολία το μοντέλο δεδομένων.

Η ίδια διαδικασία γίνεται και για τους υπόλοιπους κύβους φτιάχνοντας ξεχωριστό data source για τη κάθε περίπτωση. Μια επιπλέον λειτουργικότητα που έχει φτιαχτεί είναι η μετατροπή ποσών από δολάρια(έτσι είναι δοσμένα από τη παγκόσμια τράπεζα) σε ευρώ.

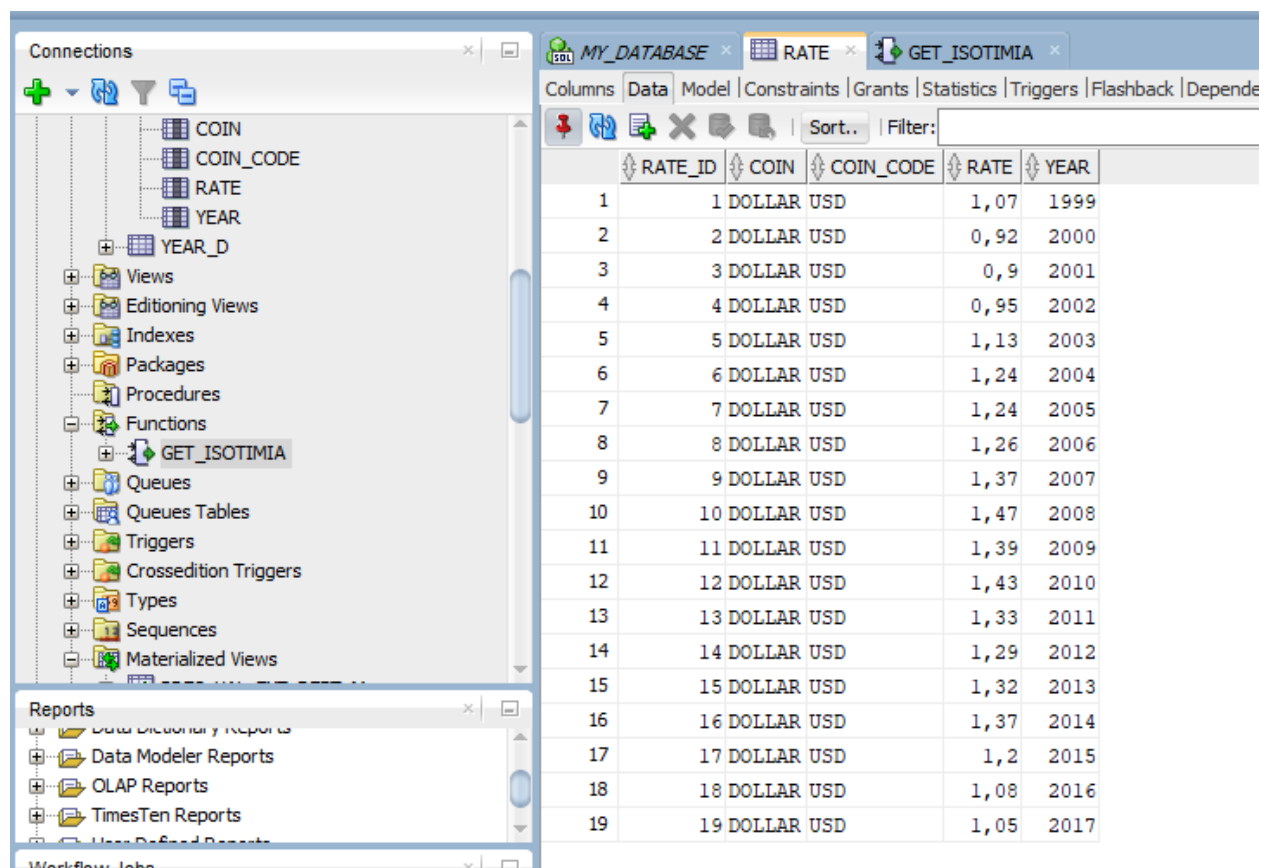
Έτσι λοιπόν δημιουργήθηκε μια συνάρτηση στην Oracle, η οποία παίρνει παράμετρο το έτος και το ποσό και επιστρέφει την αξία σε ευρώ.

Οι ισοτιμίες καταχωρήθηκαν σε ένα πίνακα με στήλες το νόμισμα, κωδικό νομίσματος, ισοτιμία και η χρονολογία της πρώτης ημέρας του χρόνου(πίνακας 4.2.3) από το παρακάτω σύνδεσμο για τις χρονολογίες 1999 μέχρι 2017

<http://www.x-rates.com/historical/?from=USD&amount=1&date=2017-01-01> .

Πίνακας 4.2.3

Πίνακας με όλες τις ισοτιμίες από δολάρια σε ευρώ από το 1999 μέχρι και το 2017

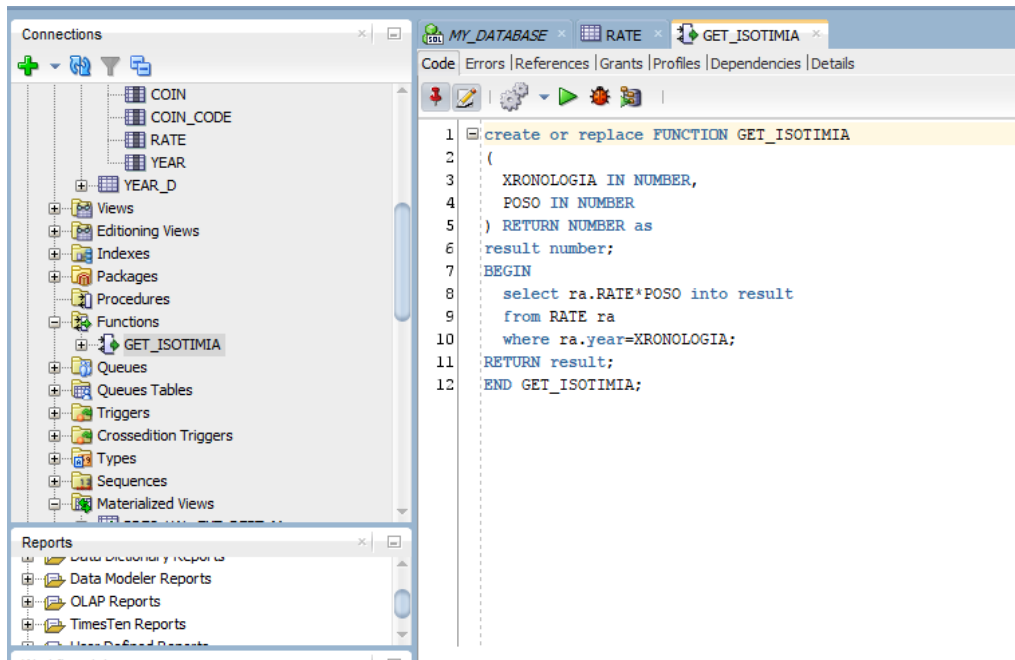


RATE_ID	COIN	COIN_CODE	RATE	YEAR
1	1 DOLLAR	USD	1,07	1999
2	2 DOLLAR	USD	0,92	2000
3	3 DOLLAR	USD	0,9	2001
4	4 DOLLAR	USD	0,95	2002
5	5 DOLLAR	USD	1,13	2003
6	6 DOLLAR	USD	1,24	2004
7	7 DOLLAR	USD	1,24	2005
8	8 DOLLAR	USD	1,26	2006
9	9 DOLLAR	USD	1,37	2007
10	10 DOLLAR	USD	1,47	2008
11	11 DOLLAR	USD	1,39	2009
12	12 DOLLAR	USD	1,43	2010
13	13 DOLLAR	USD	1,33	2011
14	14 DOLLAR	USD	1,29	2012
15	15 DOLLAR	USD	1,32	2013
16	16 DOLLAR	USD	1,37	2014
17	17 DOLLAR	USD	1,2	2015
18	18 DOLLAR	USD	1,08	2016
19	19 DOLLAR	USD	1,05	2017

Ο κώδικας της συνάρτησης που υλοποιεί την ζητούμενη μετατροπή είναι ο παρακάτω(εικόνα 4.2.4) :

Εικόνα 4.2.4

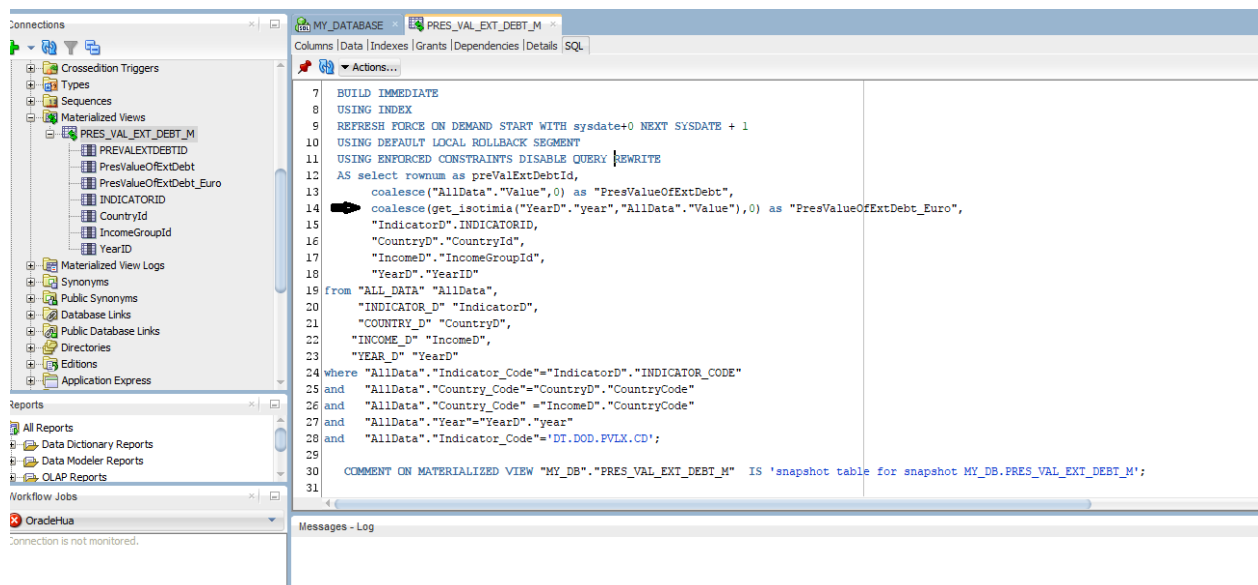
Συνάρτηση που παίρνει ως είσοδο τιμή, χρονολογία και επιστρέφει το ποσό σε ευρώ



Και η κλήση της στο materialized view (εικόνα 4.2.5)

Εικόνα 4.2.5

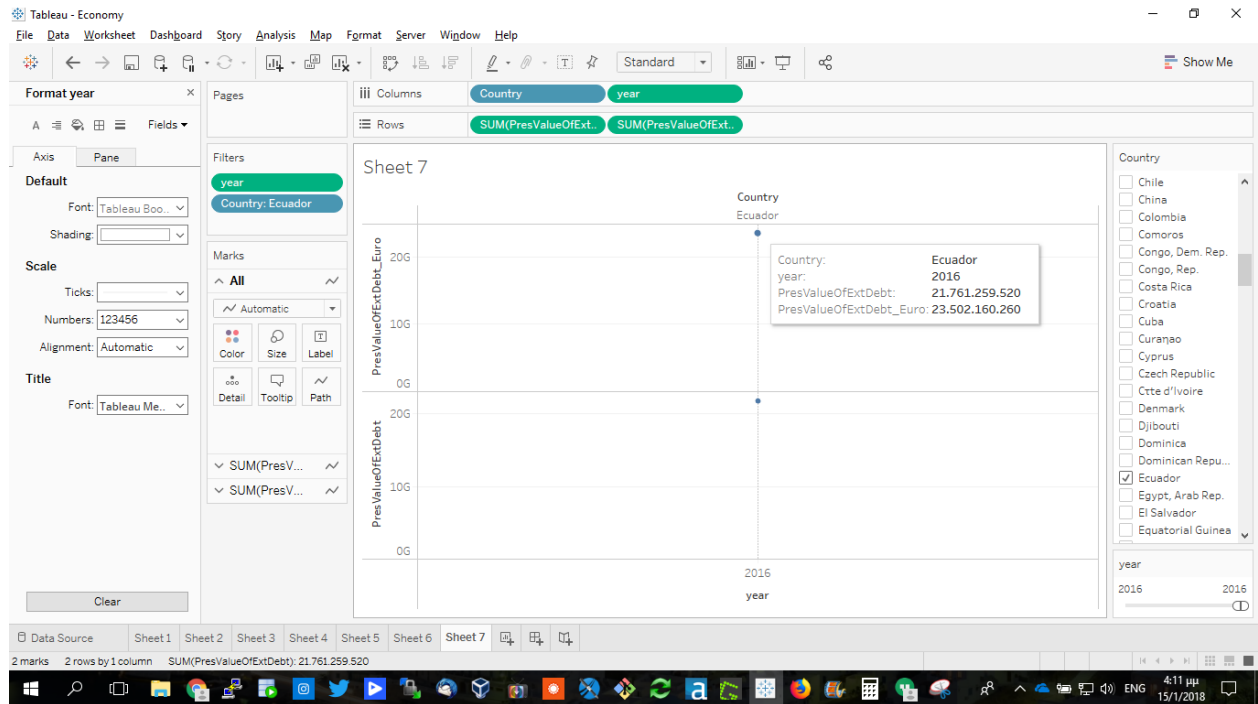
Κλήση συνάρτησης στο materialized view



Ένα παράδειγμα στο πως απεικονίζεται στο tableau, η πρώτη κουκκίδα αντιστοιχεί σε αξία ευρώ ενώ η από κάτω σε δολάρια (διάγραμμα 4.2.6) .

Διάγραμμα 4.2.6

Εμφάνιση ποσού τόσο σε ευρώ όσο και σε δολάρια



Βλέπουμε ότι για το έτος 2016 για το Ecuador η παρούσα αξία του εξωτερικού χρέους σε δολάρια είναι 21.761 δισεκατομμύρια ενώ η αξία αυτή σε ευρώ αντιστοιχεί σε 23.502 δισεκατομμύρια($21.761 \cdot 1,08$ που είναι το rate)

ΚΕΦ.5 : ΑΠΟΤΕΛΕΣΜΑΤΑ

Έχοντας έτοιμο το μοντέλο δεδομένων, μπορούμε να ξεκινήσουμε να δημιουργούμε αναλύσεις αφού έχουμε στο μυαλό μας κάποια στατιστικά που θέλουμε να προβάλλουμε.

Συνεπώς πριν ξεκινήσουμε να οπτικοποιούμε συγκεκριμένα σενάρια, πρέπει να γνωρίζουμε τη δομή δεδομένων που έχουμε. Έτσι λοιπόν το πρώτο πράγμα που πρέπει να κάνουμε είναι να επιλέξουμε θεματική περιοχή, δηλαδή τι ακριβώς θέλουμε να μετρήσουμε. Έστερα αφού καταλήξουμε τι είναι αυτό που θέλουμε να δούμε, πρέπει να γνωρίζουμε τα πεδία εκείνα που θα μας φανούν χρήσιμα για τα συγκεκριμένα metrics. Έτσι λοιπόν έχουμε 4 διαστάσεις και μέσα σε κάθε διάσταση έχουμε τα γνωρίσματα που σχετίζονται με τη συγκεκριμένη κατηγοριοποίηση της διάστασης.

Αναλυτικά οι διαστάσεις που έχουν δημιουργηθεί είναι οι εξής:

α) Χώρα

Γνωρίσματα: χώρα, ήπειρος

β) Εισοδηματική κατάσταση χώρας

Γνωρίσματα: Εισοδηματική ένδειξη περιγραφικά, Εισοδηματική ένδειξη με λεπτομέρεια, Περιγραφή εισοδηματικής ένδειξης

γ) Οικονομικοί δείκτες

Γνωρίσματα: Κωδικός οικονομικής ένδειξης, όνομα οικονομικού δείκτη, περιγραφή οικονομικού δείκτη, πηγή προέλευσης δεδομένων για το συγκεκριμένο δείκτη

δ) Χρόνος

Γνωρίσματα: Χρόνος, Δεκαετία

Ενώ για Θεματικές περιοχές/κύβους έχουμε τα παρακάτω μετρήσιμα:

α) Εξαγωγές αγαθών και υπηρεσιών (σε δολάρια)

β) Εισαγωγές αγαθών και υπηρεσιών(σε δολάρια)

γ) Παρούσα αξία του εξωτερικού χρέους(σε δολάρια)

δ) Κεντρικό δημόσιο χρέος(% of GDP)

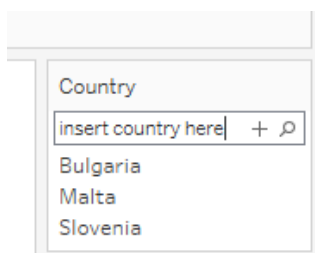
Ένα σενάριο θα ήταν για παράδειγμα να βρούμε το ποσό σε δολάρια για εξαγωγές αγαθών στις χώρες Βουλγαρία, Μάλτα και Σλοβενία για τη δεκαετία του 1990.

Πρέπει να σκεφτούμε αρχικά πιο data source θα επιλέξουμε, δηλαδή να δούμε τι θέλουμε να μετρήσουμε και ύστερα πια είναι αυτά τα πεδία από τις διαστάσεις που έχουν να κάνουν με το ερώτημα. Οπότε αφού πρόκειται για χώρες, πρέπει να δούμε τα διαθέσιμα γνωρίσματα από τη διάσταση χώρα και ύστερα από τη διάσταση του χρόνου.

Οπότε αρχικά επιλέγουμε το κύβο με τις εξαγωγές αγαθών για να βγάλει το ανάλογο μετρήσιμο και ύστερα σκεφτόμαστε ότι αρχικά πρέπει να πάρουμε από τη διάσταση χώρα το πεδίο Country και να το τοποθετήσουμε στη περιοχή που μπαίνουν οι κολόνες. Αυτό αν το κάνουμε θα μας φέρει όλες τις χώρες στον άξονα X, οπότε πάμε και βάζουμε και ένα φίλτρο (εικόνα 5.2.1) και επιλέγουμε μόνο τις τρεις ζητούμενες χώρες

Εικόνα 5.2.1

Φίλτρο στο tableau που γίνεται επιλογή συγκεκριμένων χωρών

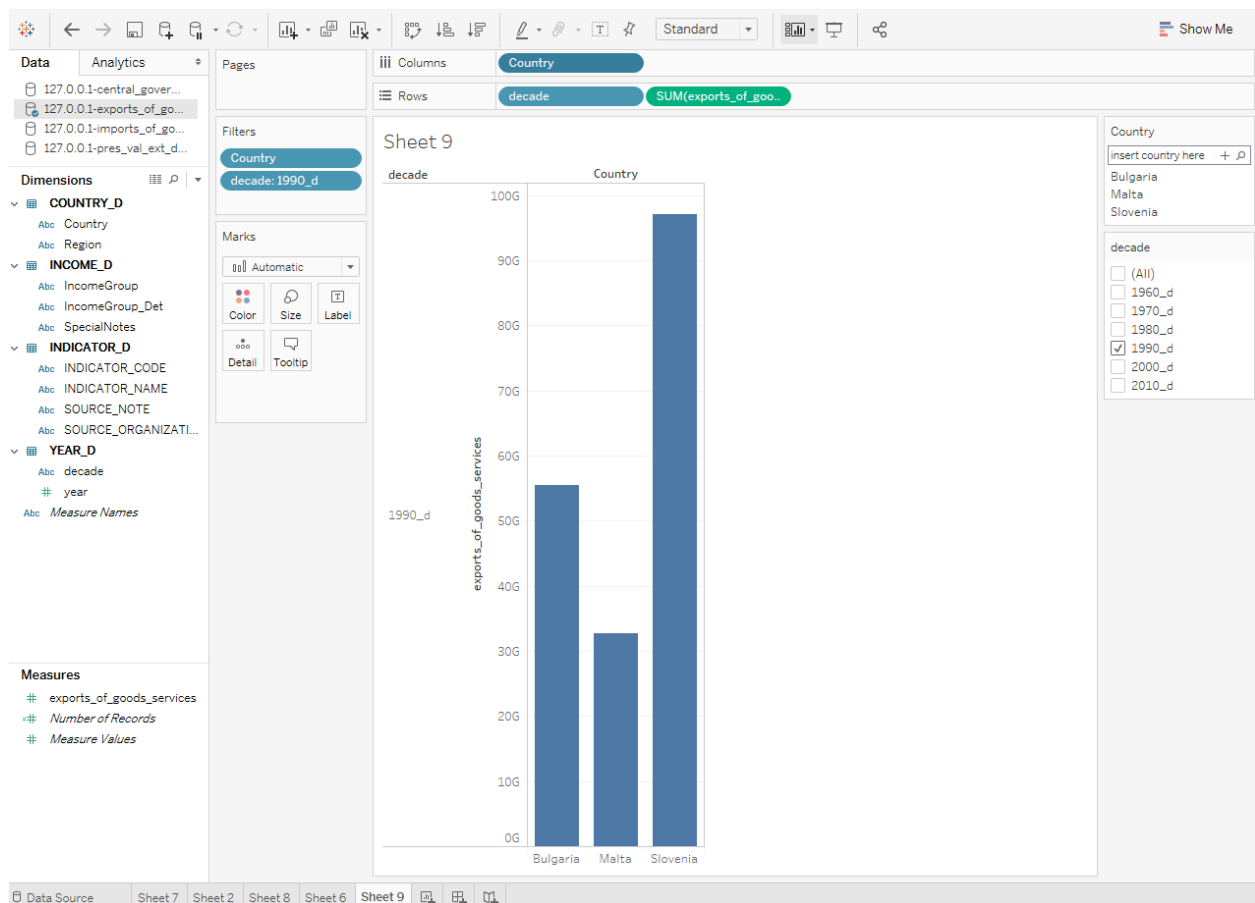


Ύστερα τοποθετούμε στη περιοχή των γραμμών, από τη διάσταση του χρόνου, το πεδίο δεκαετία φιλτραρισμένο με μόνο τη δεκαετία του 1990 και το ζητούμενο μετρήσιμο από το measures data tab .

Αυτό περιμένουμε να βγάλει ένα γράφημα με τον άξονα X να περιλαμβάνει τις 3 χώρες που επιλέξαμε και στον άξονα Y τη δεκαετία και το εύρος τιμών σε δολάρια ανάλογα με τη παραγωγή εξαγωγών των συγκεκριμένων χωρών (διάγραμμα 5.2.2) .

Διάγραμμα 5.2.2

Εξαγωγές αγαθών και υπηρεσιών στη Βουλγαρία, Μάλτα και Σλοβενία για τη δεκαετία του 1990



Τώρα θα δούμε κάποιες άλλες αναλύσεις με πιο ρεαλιστικά παραδείγματα.

A)

Στη παρακάτω ανάλυση (Διάγραμμα 5.2.3) βλέπουμε εξαγωγές αγαθών και υπηρεσιών σε δολάρια για τις χώρες Βουλγαρία, Φιλανδία, Ελλάδα, Πολωνία, Πορτογαλία και Τουρκία για τα έτη 1990 μέχρι το 2004.

Εξαγωγές αγαθών και υπηρεσιών σε δολάρια για τις χώρες Βουλγαρία, Φιλανδία, Ελλάδα, Πολωνία, Πορτογαλία, Τουρκία μεταξύ των χρονολογιών 1990 και 2004

INDICATOR_NAME / SOURCE_NOTE

Exports of goods and services (current US\$)

Exports of goods and services represent the value of all goods and other market services provided to the rest of the world. They include the value of merchandise, freight, insurance, transport, travel, royalties, license fees, and other services, such as communication, construction, financial, information, business, personal, and government services. They exclude compensation of employees and investment income (formerly called factor services) and transfer payments. Data are in current U.S. dollars.

Country

year

2005

2000

1995

1990

Bulgaria

2005

2000

1995

1990

Finland

2005

2000

1995

1990

Greece

2005

2000

1995

1990

Poland

2005

2000

1995

1990

Portugal

2005

2000

1995

1990

Turkey

0G

10G

20G

30G

40G

50G

60G

70G

80G

90G

exports_of_goods_services

Finland

Greece

Poland

Portugal

Turkey

Παρατηρούμε ότι μεγάλη ανοδική πορεία έχουν οι Τουρκία και ύστερα η Πολωνία.

Αξίζει να επισημάνουμε την ανοδική πορεία της Τουρκίας που το 1990 είχε 20.198 δισεκατομμύρια, το 2004 έφτασε τα 92.088 δισεκατομμύρια δολάρια εξαγωγές.

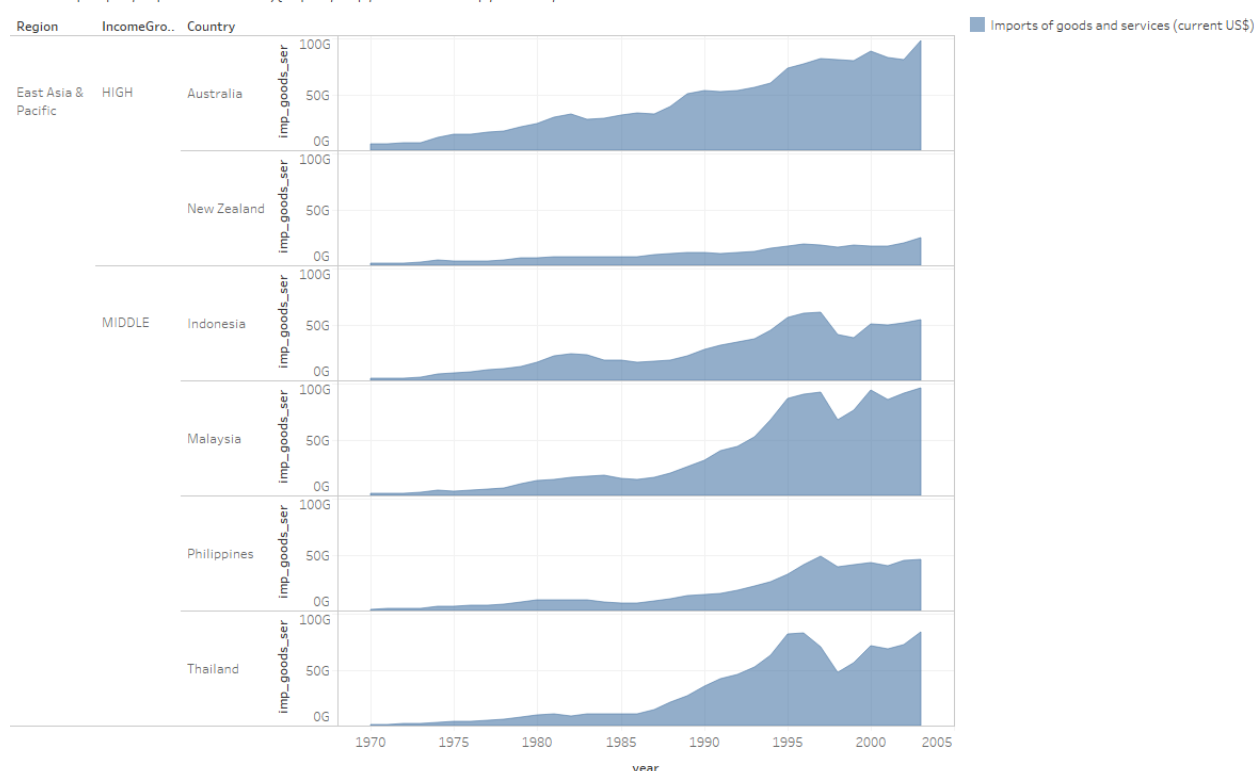
Η Ελλάδα αντίστοιχα ξεκίνησε το 1990 με 14.5 δισεκατομμύρια, το 2000 έφτασε τα 30.8 , ενώ το 2004 έφτασε κοντά τα 50 δισεκατομμύρια δολάρια

B) Στη παρακάτω ανάλυση (διάγραμμα 5.2.4) θα μελετήσουμε εισαγωγές αγαθών για κάποιες χώρες του Ανατολικής Ασίας και του Ειρηνικού, δηλαδή το κόστος σε δολάρια για την απόκτηση και μεταφορών προϊόντων που προέρχονται από τρίτες χώρες για χρονολογίες από το 1970 μέχρι και το 2003.

Διάγραμμα 5.2.4

Εισαγωγές αγαθών και υπηρεσιών σε δολάρια για τις χώρες της Ανατολικής Ασίας για τα χρονολογίες από 1970 μέχρι και 2003

<Εισαγωγές αγαθών από χώρες της Ανατολικής Ασίας >



The plot of sum of imp_goods_ser for year broken down by Region, IncomeGroup and Country. Color shows details about INDICATOR_NAME. The view is filtered on Country and year. The Country filter keeps 6 of 217 members. The year filter ranges from 1970 to 2003.

Βλέπουμε ότι τις μεγαλύτερες εισαγωγές τις κάνει η Αυστραλία με το 2003 να φτάνει τα 98, 1 δισεκατομμύρια δολάρια. Αξιοσημείωτη είναι και οι εισαγωγές της Μαλαισίας όπου το 2003 έφτασε τα 96 δισεκατομμύρια δολάρια. Η Νέα Ζηλανδία παρατηρούμε να κρατάει τις χαμηλότερες τιμές όπου για κανένα έτος δεν ξεπερνά τα 25 δισεκατομμύρια δολάρια, ενώ αμέσως μετά ακολουθούν οι Φιλιππίνες που δεν ξεπερνούν τα 45 δισεκατομμύρια στα διάφορα χρονικά έτη. Για την Ταϊλάνδη παρατηρείται μια πτώση στη κλίμακα τιμών από το 1995(81 δισ) μέχρι και το 1998(48 δισ) ενώ μετά έχει πάλι ανοδική πορεία. Επίσης να αναφέρουμε ότι χώρες όπως η Μαλαισία και η Ταϊλάνδη είναι μεσαίες εισοδηματικής προέλευσης σε σχέση με

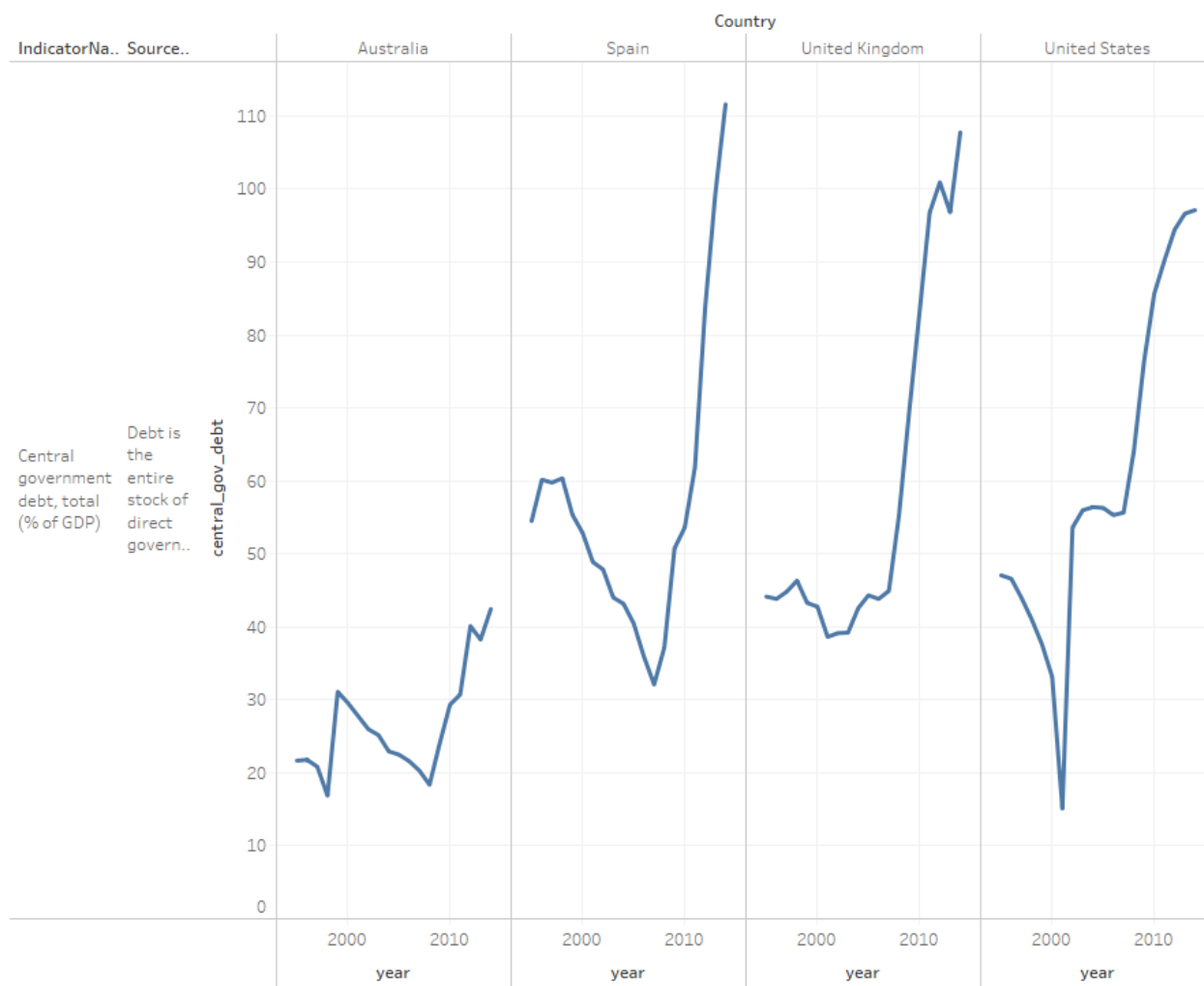
την Αυστραλία και τη Νέα Ζηλανδία και όμως έχουν ένα μεγάλο ποσό σε δολάρια που γίνονται εισαγωγή ετησίως.

Γ) Παρακάτω θα μελετήσουμε (διάγραμμα 5.2.5) στατιστικά στοιχεία για το κεντρικό δημόσιο χρέος για τις χώρες Αυστραλία, Ισπανία, Ηνωμένο Βασίλειο και Ηνωμένες Πολιτείες της Αμερικής για τα χρονικά έτη από το 1995 μέχρι και το 2004

Διάγραμμα 5.2.5

Κεντρικό δημόσιο χρέος για τις χώρες Αυστραλία, Ισπανία, Ηνωμένο Βασίλειο και Ηνωμένες Πολιτείες της Αμερικής από το 1995 μέχρι και το 2004

<Κεντρικό δημόσιο χρέος>



The trend of sum of central_gov_debt for year broken down by Country vs. IndicatorName and SourceNote. The view is filtered on year and Country. The year filter ranges from 1995 to 2014. The Country filter keeps Australia, Spain, United Kingdom and United States.

Παρατηρούμε ότι Ισπανία είναι κυρίαρχη στο κεντρικό δημόσιο χρέος που έχει. Από το 2007 και μετά βλέπουμε να αυξάνεται δραματικά από 32 σε 111 % of GDP ενώ από το 1998 μέχρι και το 2007 μειωνόταν. Από την άλλη πλευρά βλέπουμε ότι η κατάσταση της οικονομίας της Αυστραλίας είναι σε καλή κατάσταση και κρατάει χαμηλά επίπεδα χρέους που έχει. Ηνωμένο Βασίλειο και Ηνωμένες Πολιτείες της Αμερικής έχουν και αυτές φάσεις καθόδου στην οικονομία τους από το 2007 και μετά, δηλαδή αυξάνεται το δημόσιο χρέος που έχουν αισθητά πάνω στο χρόνο. Αξίζει επίσης να σημειωθεί η φάση της άνθησης για τις Ηνωμένες Πολιτείες από το 1995 μέχρι και το 2001, ενώ από εκεί και πέρα ξεκινάει η περίοδος της ύφεσης.

ΚΕΦ.6: ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην Επιστήμη των υπολογιστών καθημερινά σχεδιάζονται εφαρμογές με στόχο να καλύψουν ανάγκες, δηλαδή συλλέγονται περιοδικά δεδομένα μέσα από φόρμες καταχώρησης και ύστερα καταχωρούνται σε συνήθως σχεσιακές βάσεις δεδομένων. Μπορεί επίσης να έχουμε πολλαπλά συστήματα που αλληλεπιδρούν μεταξύ τους, παρέχοντας έτσι διαλειτουργικότητα. Η γενική ιδέα είναι να μαζέψουμε όλα τα δεδομένα από διαφορετικές πηγές προέλευσης, να μπορούμε να προσθέσουμε και επιπλέον πληροφορία αν θέλουμε είτε αυτό είναι μεμονωμένα αρχεία είτε το οτιδήποτε άλλο και να τα βάλουμε όλα σε μια κοινή βάση σε όποια απλή και κοινή μορφή γίνεται. Επίσης μπορούμε να ορίσουμε εμείς πότε να τρέχουν τέτοιες διαδικασίες συλλογής δεδομένων μέσα από ETL εργαλεία όπως αυτό που χρησιμοποιήσαμε στην εργασία, το Alteryx. Ύστερα αφού ολοκληρωθεί αυτή η διαδικασία μπορούμε να μετασχηματίσουμε αυτά τα δεδομένα και να φτιάξουμε facts και διαστάσεις και αυτά να απεικονιστούν σε επίπεδο visualization σε φιλική για το χρήστη μορφή. Στη δική μας περίπτωση πήραμε έτοιμα τα δεδομένα μας από την παγκόσμια τράπεζα, τα οποία ήταν excel αρχεία, τα φορτώσαμε, τα μετασχηματίσαμε και αφού τα φέραμε σε μορφή για να κάνει κάποιος ανάλυση τα περάσαμε στο εργαλείο tableau, όπου εκεί ο χρήστης θα μπορεί να τα επεξεργαστεί μόνος του και να φτιάξει τις δικές του αναλύσεις. Μελετώντας τα γραφήματα στα δικά μας παραδείγματα, θα μπορεί κάποιος να διαπιστώσει οικονομίες που χρειάζονται παρέμβαση ή και χώρες που έχουν καλή οικονομία και στάδια ανάπτυξης και άνθησης. Βλέποντας για παράδειγμα την Αυστραλία να έχει αρκετά μικρό κεντρικό δημόσιο χρέος σε σχέση με άλλες αρκετά μικρότερες χώρες της Ευρώπης όπως η Ισπανία που έχει αρκετά υψηλό και αυξάνεται με γρήγορους ρυθμούς. Επίσης μπορέσαμε να διαπιστώσουμε χώρες όπως η Πολωνία με ένα μεγάλο ποσοστό σε εξαγωγές

αγαθών και υπηρεσιών να αναπτύσσεται με γρήγορους ρυθμούς και αντίθετα την Βουλγαρία να μένει στάσιμη και να χρειάζεται παρέμβαση για να επιβιώσει.

6.1. ΜΕΛΛΟΝΤΙΚΑ ΖΗΤΗΜΑΤΑ

Τώρα για διάφορα μελλοντικά ζητήματα, μπορούμε να πούμε ότι υπάρχουν διάφορες τάσεις στην σύγχρονη επιχειρηματική ευφυΐα που στοχεύουν στην ανάπτυξη νέων καινοτόμων λειτουργιών. Η ενσωματωμένη επιχειρηματική ευφυΐα (Embedded BI) για παράδειγμα, δηλαδή να μπορούν οι χρήστες μέσα από υπάρχουσες εφαρμογές να προστέσουν features όπως dashboard reporting, data visualization και analytics tools είτε μέσα από το ίδιο το software, είτε από ένα εξωτερικό framework που ενσωματώνεται μέσα στο κύριο, υλοποιώντας και στις δύο περιπτώσεις κάποιες βιβλιοθήκες που με κατάλληλες μεθόδους (GET) ανακτούνται τα δεδομένα από τη βάση και γίνεται επεξεργασία και ανάλυση μετά πάνω σε αυτά, θα μπορούσε να θεωρηθεί μεγάλη πρόκληση.

Δεδομένου ότι το πλήθος δεδομένων αυξάνεται δραματικά τα τελευταία χρόνια, ένα άλλο επίκαιρο ζήτημα είναι η ανάλυση μεγάλου όγκου δεδομένων. Το να πάρουμε τα δεδομένα για παράδειγμα από μια κατανεμημένη βάση που υποστηρίζει την αποθήκευση δεδομένων πολλών πινάκων όπως είναι το HBase⁴ ή από μια data warehouse υποδομή που κάνει αυτόματα αθροίσματα στα δεδομένα και παρέχει τη δυνατότητα Ad-hoc ερωτημάτων όπως είναι το Hive⁵ μας κατευθύνουν στο ότι πλέον τα δεδομένα εισόδου σε ροές που τρέχουν για να φτιάξουν διαστάσεις και facts θα είναι τέτοιας μορφής και όχι πλέον μιας απλής σχεσιακής βάσης.

Μια άλλη πρόκληση είναι οι αναλύσεις σε δεδομένα πραγματικού χρόνου. Σε αυτή τη περίπτωση, όταν υπάρχουν εφαρμογές και καταχωρούνται δεδομένα, θέλουμε κάποιες φορές εκείνη τη χρονική στιγμή να τα αντλήσουμε χωρίς καμία καθυστέρηση. Αυτό που είδαμε μέχρι τώρα είναι να έχουμε ροές και materialized view που γίνονται schedule σε συγκεκριμένα

⁴ <http://hbase.apache.org/>

⁵ <http://hive.apache.org/>

χρονικά διαστήματα, όμως στη περίπτωση που τα δεδομένα αλλάζουν από λεπτό σε λεπτό ή ακόμα από δευτερόλεπτο σε δευτερόλεπτο μια etl διαδικασία θα αποτύχει να κάνει retrieve αυτή τη πληροφορία.

Τέλος, θα θέλαμε να δούμε ιδανικά το cloud computing με την επιχειρηματική ευφυΐα ως ένα άριστο συνδυασμό. Δηλαδή να χρησιμοποιούμε Business Intelligence εφαρμογές προσβάσιμες από πολλές συσκευές (σταθερές και κινητές) και διάφορους browsers, ανά πάσα στιγμή σε άτομα που έχουν τα κατάλληλα δικαιώματα με μόνο τη χρήση του διαδικτύου. Με αυτό το τρόπο θα πετυχαίναμε μεγαλύτερο deployment speed αφού δεν θα χρειαζόμασταν επιπλέον software ή hardware installations, ευκολία στη χρήση καθώς και μεγαλύτερο scalability, δηλαδή την ικανότητα του συστήματος να αντέξει υψηλό φόρτο από requests κάνοντας το υλικό καλύτερο (scale up) ή προσθέτοντας περισσότερους κόμβους (scale out).

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Donald A. R. George, Brian Chi-ang Lin, Yunmin Chen (2015). A circular economy model of economic growth, 60-63.doi: 10.1016/j.envsoft.2015.06.014
- Murray, D. G. (2013). Tableau your data!:fast and easy visual analysis with tableau software. John Wiley & Sons
- Jones, B. (2014). Communicating Data with Tableau: Designing, Developing and Delivering Data Visualizations. “O'Really Media,Inc.”.
- Κύρκος Ευστάθιος (2015). Επιχειρηματική Ευφυΐα και Εξόρυξη Δεδομένων. Ανοικτά συγγράμματα Κάλλιπος.Διαθέσιμο στο <https://repository.kallipos.gr/handle/11419/1226>
- Γεωργούσης Ηλίας (2016). Knowledge mining from financial data and credit evaluation data from rating organisations. Διπλωματική εργασία. Χαροκόπειο Πανεπιστήμιο
- Thorsten Beck, Asli Demirkuk-Kunt (2009). Financial Institutions and Markets across Countries and over Time,1-16
- L.A. Finley (2005). Economic Growth Issues,24-25