



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ Η/Υ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΕΡΓΑΣΙΑ ΕΞΑΜΗΝΟΥ

Προηγμένα Πληροφοριακά Συστήματα

Πετκίδης Δημήτριος 5628
Χασούρης-Χρηστίδης Κωνσταντίνος 5684
Καλούρης Γεράσιμος 5946
Στατήρη Μαρία 5654
Δάμπασης Ερωτόκριτος 5748

Υπεύθυνος Καθηγητής :
Γεώργιος Παυλίδης, Αθανάσιος Τσακαλίδης

Περιεχόμενα

1	Εισαγωγή	2
1.1	Σημασία και οφέλη DW	2
1.2	Σημασία και οφέλη OLAP	2
2	Μελέτη δεδομένων οδικής ασφάλειας	3
2.1	Αξιολόγηση δεδομένων και επιλεγμένα ερωτήματα .	3
2.2	Μετατροπή δεδομένων και δημιουργία DW	3
2.3	Δημιουργία OLAP κύβου	6
2.4	Οπτικοποίηση αποτελεσμάτων κύβου	6
3	Υλοποίηση bonus ερωτήματος	12
3.1	Δημιουργία DW και μετατροπή δεδομένων	12
3.2	Ενδεικτική οπτικοποίηση δεδομένων	14

Θέμα εργασίας: Δημιουργία DW και OLAP visualisation

1 Εισαγωγή

1.1 Σημασία και οφέλη DW

Τα Data Warehouse(DW), αποτελούν καθοριστικό κομμάτι του επιχειρηματικού κόσμου, και χρησιμοποιούνται ευρέως για την αναφορά, την καταγραφή και την εξόρυξη δεδομένων, που μετέπειτα θα τροποποιηθούν και θα μετατραπούν σε πληροφορία. Με τον όρο αυτό, δεν γίνεται απλή αναφορά σε μία βάση δεδομένων, η οποία περιέχει τα παραπάνω στοιχεία. Αντιθέτως, η υλοποίηση και η εφαρμογή των data warehouse, κρίθηκε απαραίτητη λόγω του μεγάλου όγκου ακατέργαστης πληροφορίας που αποθηκεύεται στις βάσεις δεδομένων. Με τη χρήση τους, δημιουργούνται βιβλιοθήκες κατεργασμένων και μορφοποιημένων δεδομένων, οι οποίες μπορούν να προσπελαστούν και να μελετηθούν για την εξαγωγή συγκεκριμένων συμπερασμάτων.

Όπως είναι κατανοητό, επιφέρουν πολλά οφέλη στις επιχειρήσεις και τους οργανισμούς που τα χρησιμοποιούν. Το πιο σημαντικό από αυτά είναι η ταχύτερη λήψη αποφάσεων. Άλλωστε, κάθε οργανισμός, προκειμένου να γίνει ανταγωνιστικότερος και αποδοτικότερος, καθώς και για να υλοποιήσει το όραμα και την αποστολή του, επιστρατεύει μία στρατηγική πορεία, η οποία καθορίζεται σε μεγάλο βαθμό από τις αποφάσεις που λαμβάνονται στα πλαίσιά του. Έτσι, η σημασία των data warehouses γίνεται αντιληπτή, αφού διευκολύνουν και επιταχύνουν τη διαδικασία αυτή. Επιπλέον, τα δεδομένα είναι πλέον εύκολα προσβάσιμα και τροποποιημένα στην επιθυμητή τους μορφή, με αποτέλεσμα να είναι ταχύτερη η προσέλαση και η μελέτη τους, ενώ ταυτόχρονα αυξάνεται η αξιοπιστία και η εγκυρότητά τους, αφού πριν εισαχθούν στο εκάστοτε data warehouse έχουν ελεγχθεί και μορφοποιηθεί κατάλληλα.

1.2 Σημασία και οφέλη OLAP

Το OLAP (online analytical processing) αποτελεί μία υπολογιστική διαδικασία, η οποία επιτρέπει στο χρήστη να αποσπάσει και να μελετήσει επιλεκτικά τα δεδομένα που επιθυμεί, από διαφορετικές οπτικές γωνίες. Έτσι, επιταχύνονται οι συγκρίσεις δεδομένων και πληροφοριών, καθώς και όλες οι παράγωγες διαδικασίες που προκύπτουν από τη μελέτη των επιθυμητών δεδομένων, όπως για παράδειγμα η λήψη αποφάσεων. Το OLAP υλοποιείται μέσω μίας πολυδιάστατης βάσης δεδομένων, η οποία θεωρεί κάθε χαρακτηριστικό(π.χ. χρόνος) ως ξεχωριστή διάσταση. Ουσιαστικά, έχει την ικανότητα να εντοπίζει τη διασταύρωση των διαστάσεων αυτών και να τις απεικονίζει. Χρησιμοποιείται κατά κύριο λόγο για data mining ή/και για την απεικόνιση των σχέσεων μεταξύ των οντοτήτων που περιέχονται, οι οποίες δεν είναι πάντα προφανείς.

2 Μελέτη δεδομένων οδικής ασφάλειας

2.1 Αξιολόγηση δεδομένων και επιλεγμένα ερωτήματα

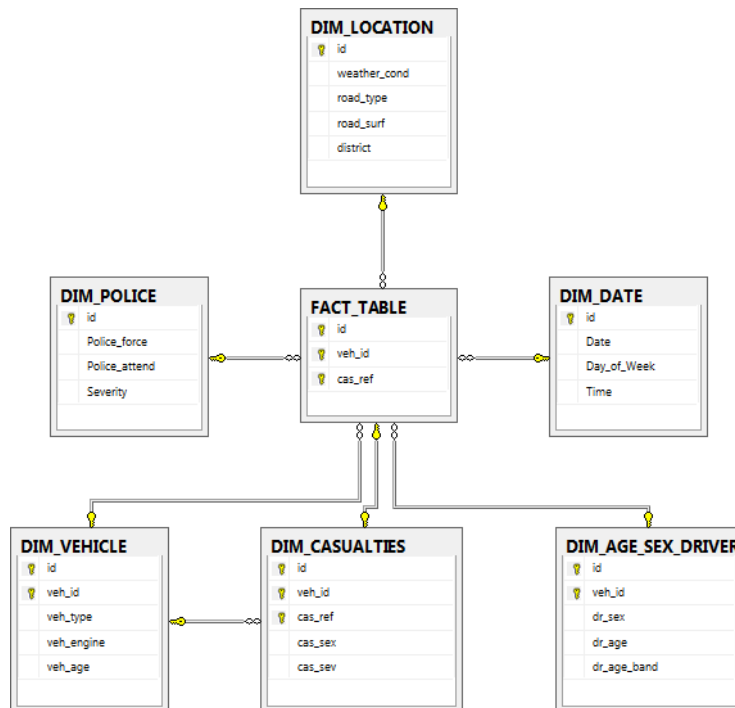
Τα δεδομένα της εργασίας αφορούν την οδική ασφάλεια και τα ατυχήματα των περιοχών της Μ.Βρετανίας. Κάθε χρονιά περιλαμβάνει δεδομένα σχετικά με τα ατυχήματα (accidents), τις απώλειες και τους τραυματισμούς (casualties), τον τύπο οχήματος/ών (vehicles) και τον τρόπο με τον οποίο προκλήθηκε το ατύχημα, καθώς και δεδομένα των κατασκευαστών των οχημάτων και δεδομένα που αφορούν παραβάσεις υπό την επήρεια αλκοόλ.

Διαπιστώνεται λοιπόν ότι λόγω της ευρείας ποικιλίας και του πλήθους των καταχωρημένων δεδομένων, η μελέτη και η επεξεργασία τους αποτελεί δύσκολο έργο, αφού σε αυτά περιέχονται λανθασμένες τιμές. Επιπλέον, αναλόγως με την οπτική γωνία και το σκοπό που τα μελετάει και τα επεξεργάζεται κάποιος, παρατηρείται πως όλα τα προαναφερθέντα και τα επιμέρους χαρακτηριστικά που περιέχονται σε αυτά, τα καθιστούν δύσκαμπτα. Έτσι, η εξέτασή τους γίνεται αργή και πιθανώς να προκύπτουν λάθη, καθώς τα δεδομένα δεν είναι μορφοποιημένα, και ορισμένα δεν είναι έγκυρα. Για το λόγο αυτό, δημιουργήθηκε ένα data warehouse, το οποίο περιέχει μόνο τα δεδομένα που είναι επιθυμητά για εξέταση. Φυσικά, τα περιεχόμενά του, καθορίζονται πλήρως από τα ερωτήματα, δηλαδή τις οπτικές γωνίες εξέτασης, καθιστώντας την επιλογή τους μία πολύ σημαντική διαδικασία.

Στα πλαίσια της άσκησης, επιλέχθηκαν ερωτήματα/περιπτώσεις μελέτης, οι οποίες δεν αποτελούν απλά στατιστικά δεδομένα, όπως για παράδειγμα η μείωση ή η αύξηση των ατυχημάτων σε πέρασμα εξαετίας, αλλά ερωτήματα, από τα οποία απορρέουν σημαντικές πληροφορίες, εκ των οποίων μπορούν να εξαχθούν προτάσεις για βελτίωση στους αρμόδιους οργανισμούς και φορείς. Πιο συγκεκριμένα, προτιμήθηκε έναντι της στατιστικής μελέτης, η μελέτη των συνθηκών των δρόμων σε ορισμένες περιοχές, ώστε όταν οι επικρατούσες διακινδυνεύουν την πρόκληση ατυχημάτων, να αναφέρονται στον αρμόδιο φορέα που είναι υπεύθυνος για την αποκατάστασή τους. Πιο αναλυτικά, ο τρόπος σκέψης και επιλογής για κάθε ερώτημα/περίπτωση μελέτης, θα παρουσιαστεί παρακάτω, μαζί με τον τρόπο εξαγωγής των απαραίτητων δεδομένων για αυτό/η.

2.2 Μετατροπή δεδομένων και δημιουργία DW

Αρχικά, σχεδιάστηκε το schema του data warehouse που χρησιμοποιήθηκε. Σε αυτό, παρουσιάζονται οι πίνακες-διαστάσεις που υλοποιήθηκαν, οι οποίοι περιέχουν το βασικό κλειδί που συνδέει κάθε μία από αυτούς με τον πίνακα fact table, ο οποίος αποτελεί ουσιαστικά τον τρόπο σύνδεσης των επιμέρους διαστάσεων και των γνωρισμάτων που περιέχονται σε αυτές. Το σχεδιάγραμμα του data warehouse παραθέτεται παρακάτω, με μία σύντομη εξήγηση για την επιλογή κάθε μίας από τις επιμέρους διαστάσεις του.



Εικόνα 1: Σχεδιάγραμμα του Data Warehouse

Όπως διακρίνεται στο σχήμα, επιλέχθηκε μία διάσταση που περιλαμβάνει την τοποθεσία του ατυχήματος(συνθήκες καιρού, τύπος δρόμου, συνθήκες επιφάνειας δρόμου και τέλος διοικητική περιοχή στην οποία συνέβη το ατύχημα), ενώ δημιουργήθηκε μία ακόμη που περιλαμβάνει στοιχεία σχετικά με την ημερομηνία του(ημερομηνία, μέρα της εβδομάδας, ώρα ατυχήματος). Επιπλέον, υλοποιήθηκε μία διάσταση που εμπεριέχει στοιχεία σχετικά με τον οδηγό του οχήματος(ηλικία, φύλο, ηλικιακή ομάδα) καθώς και μία ακόμη που συμπεριλαμβάνει στοιχεία για την προσέλευση της αστυνομίας και τη σοβαρότητα του ατυχήματος(αρμόδια αστυνομική δύναμη, προσέλευση αστυνομίας, σοβαρότητα ατυχήματος). Οι παραπάνω διαστάσεις παράχθηκαν από τα δεδομένα Accidents, που επισυνάπτονται στην ιστοσελίδα της Βρετανικής κυβέρνησης. Από τα αντίστοιχα στοιχεία που σχετίζονται με τα οχήματα που αποτελούσαν μέρος του εκάστοτε ατυχήματος(Vehicles) προκύπτει η διάσταση που περιέχει τις πληροφορίες αυτές(τύπος οχήματος, κυβισμός οχήματος, ηλικία οχήματος). Τέλος, από τα δεδομένα που επισυνάπτονται στην προαναφερθείσα ιστοσελίδα σχετικά με τις απώλειες των ατυχημάτων(Casualties), προκύπτει η διάσταση που περιέχει στοιχεία για τις απώλειες-τραυματισμούς, καθώς και από τα ατυχήματα(σοβαρότητα τραυματισμού, φύλο τραυματία).

Προκειμένου να πραγματοποιηθούν τα προαναφερθέντα, χρειάστηκε να τροποποιηθούν τα δεδομένα, έτσι ώστε να εισαχθούν έγκυρα και μορφοποιημένα στο data warehouse. Με αυτό τον τρόπο, επιταχύνεται η προσπέλασή τους και κατά συνέπεια η εξαγωγή των αποτελεσμάτων. Για να συμβεί αυτό, αρχικά μελετήθηκε η φόρμα με την οποία καταγράφονται και αρχειοποιούνται τα ατυχήματα(σχετικό link στη βιβλιογραφία). Επιπλέον, εξετάστηκαν οι οδηγίες που δίνονται σχετικά με τη σημασία και την καταγραφή των δεδομένων που παραθέτονται από τον αρμόδιο φορέα(βρετανική κυβέρνηση). Έτσι, έγιναν γνωστές οι μη έγκυρες τιμές στα δεδομένα, και αφαιρέθηκαν οι καταγραφές αυτές (π.χ. ηλικία με τιμή -1), ώστε να είναι αξιόπιστα τα αποτελέσματα που προκύπτουν.

Πιο αναλυτικά, στα πλαίσια της υλοποίησης της βάσης δεδομένων, χρησιμοποιήθηκαν τα data tools του Visual Studio, τα οποία επιτρέπουν τη διαχείριση των δεδομένων ενός αρχείου. Αρχικά, δημιουργήθηκαν βοηθητικοί πίνακες(Lookups), οι οποίοι ελέγχουν τα δεδομένα ώστε να έχουν τις επιτρεπτές τιμές. Επιπλέον, οι κανονικοί πίνακες(Accidents, Vehicles, Casualties) έχουν ξένα κλειδιά ως προς τους βοηθητικούς πίνακες. Έτσι, ελέγχονται σε κάθε εισαγωγή οι τιμές, και σε περίπτωση που κάποια εισαγωγή περιλαμβάνει τιμές εκτός των προκαθορισμένων επιτρεπτών ορίων, τότε όλη εκείνη η "γραμμή" δεν εισάγεται στη βάση δεδομένων. Η εισαγωγή των δεδομένων ξεκινά με τον πίνακα Accidents, ενώ ακολουθούν οι πίνακες Vehicles και Casualties. Πιο συγκεκριμένα, για κάθε ένα από τα παραπάνω csv αρχεία, δημιουργήθηκε ένα data flow task με flat file source, το οποίο ακολουθείται από ένα μετατροπέα δεδομένων(data converter), ο οποίος μετατρέπει το κείμενο σε δεδομένα βάσης, τα οποία με τη σειρά τους εισάγονται σε ένα OLE DB destination. Τέλος, ανανεώνονται οι διαστάσεις μέσω του Update dimensions(SQL task), το οποίο εκτελεί store procedures στη βάση δεδομένων, μέσω των οποίων δημιουργούνται οι διαστάσεις και καθορίζεται το Fact table.

2.3 Δημιουργία OLAP κύβου

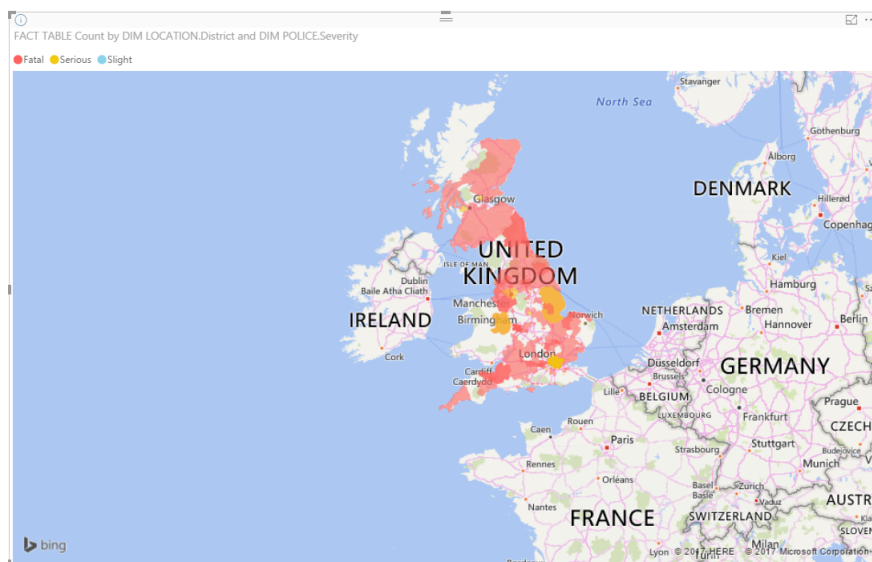
Για τη δημιουργία του κύβου OLAP, χρησιμοποιήθηκε το πρόγραμμα SQL server, και πιο συγκεκριμένα η λειτουργία του Analysis services project. Ως data source χρησιμοποιήθηκε η βάση δεδομένων που αναφέρθηκε παραπάνω, και εν συνεχεία δηλώθηκαν και δημιουργήθηκαν οι διαστάσεις της βάσης αυτής, στον κύβο.

2.4 Οπτικοποίηση αποτελεσμάτων κύβου

Προκειμένου να οπτικοποιηθούν τα αποτελέσματα, και να εξαχθούν σε μία πιο κατανοητή μορφή, οι διαστάσεις του κύβου που δημιουργήθηκε με την παραπάνω διαδικασία, 'φορτώθηκαν' στο εργαλείο Power BI της Microsoft, το οποίο επιτρέπει την επιλογή στοιχείων των διαστάσεων της βάσης δεδομένων και την οπτικοποίησή τους μέσω μίας πληθώρας διαγραμμάτων, και άλλων μέσων παρουσίασης. Παρακάτω παραθέτονται ενδεικτικά ερωτήματα που απαντήθηκαν με τη χρήση των παραπάνω προγραμμάτων και εργαλείων, τα οποία επιλέχθηκαν με βάση τα κριτήρια που καθορίστηκαν στην παράγραφο 2.1 της αναφοράς αυτής.

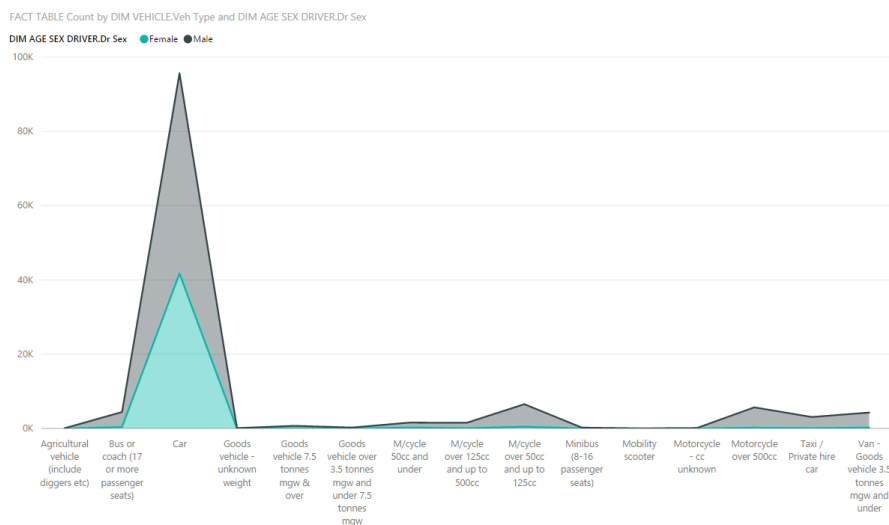
Ως πρώτο ενδεικτικό ερώτημα επιλέχθηκε η απεικόνιση όλων των κατηγοριών ατυχημάτων ως προς τη σοβαρότητά τους(μοιραία, σοβαρά, ελαφρά). Στο χάρτη που παραθέτεται, φαίνονται μόνο τα μοιραία ατυχήματα, ωστόσο στο πρόγραμμα Power BI, ο χάρτης είναι επεξεργάσιμος, και ανάλογα την επιλογή του χρήστη, παρουσιάζονται τα αντίστοιχα ατυχήματα που επιλέγονται.

Majority of accident types in each district



Εικόνα 2: Ατυχήματα ανά σοβαρότητα και περιοχή

Το επόμενο ερώτημα που επιλέχθηκε αφορά την παρουσίαση των ατυχημάτων ανά τύπο οχήματος και φύλο οδηγού. Στο διάγραμμα που ακολουθεί διακρίνονται τα ατυχήματα στα οποία συμμετείχαν τόσο άντρες όσο και γυναίκες, καθώς επίσης και ο τύπος οχήματος που υπήρξε μέρος των περισσότερων ατυχημάτων.



Εικόνα 3: Ατυχήματα ανά τύπο αυτοκινήτου και φύλο

Το επόμενο ερώτημα σχετίζεται με την παρουσία της αστυνομίας στα σοβαρά και θανάσιμα ατυχήματα της κάθε διοικητικής περιοχής. Πιο συγκεκριμένα, παρουσιάζεται ο αριθμός των μοιραίων και σοβαρών ατυχημάτων, στα οποία δεν παρευρέθηκε το εκάστοτε αστυνομικό παράρτημα. Επιλέχθηκαν τα 10 αστυνομικά παραρτήματα με τις περισσότερες φορές μη παρεύρεσης στον τόπο του ατυχήματος. Το ερώτημα αυτό στοχεύει στον καλύτερο συντονισμό και κινητοποίηση των αστυνομικών δυνάμεων, προκειμένου να διασφαλιστεί η οδική ασφάλεια.

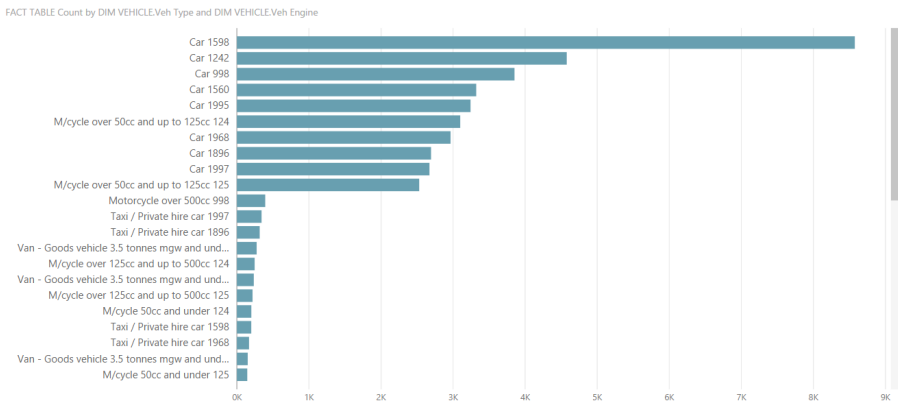
Ten police stations with the least attendance on serious and fatal accidents



Εικόνα 4: Σοβαρά και μοιραία ατυχήματα με την έλλειψη αστυνομίας

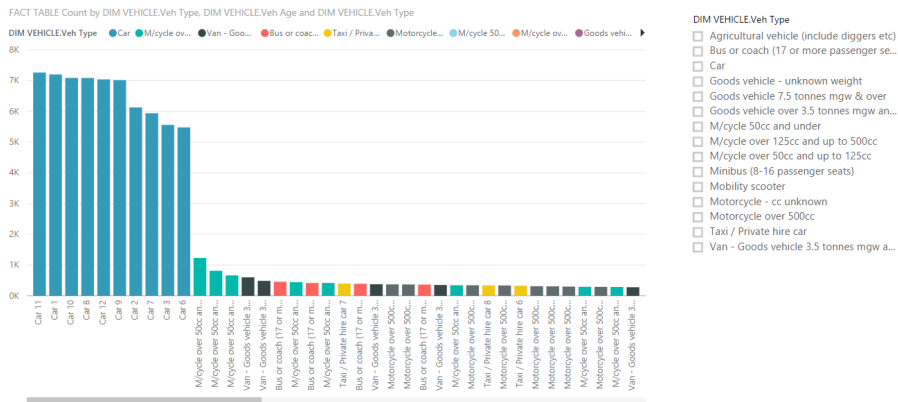
Το ερώτημα που ακολουθεί, αφορά τον κυβισμό των οχημάτων που συμμετείχαν στα ατυχήματα, και πιο συγκεκριμένα τους δέκα κυβισμούς μηχανής που εμφανίζονται συχνότερα σε αυτά. Οι ασφαλιστικές εταιρείες μπορούν να συνδυάσουν την οπτικοποίηση αυτή, με το φύλο και την ηλικία του οδηγού και τον τύπο οχήματος που εμφανίζεται συχνότερα στα ατυχήματα, καθώς και με τις ηλικίες των οχημάτων που συμμετέχουν σε αυτά, ώστε να προσαρμόσουν τα ασφαλιστικά τους κριτήρια και τις χρεώσεις τους αντίστοιχα. Παρακάτω παραθέτονται τρεις διαφορετικές οπτικοποιήσεις, μία που παρουσιάζει τις συχνότερες εμφανίσεις οχημάτων σε ατυχήματα ανά κυβισμό, μία ακόμη που παρουσιάζει τις συχνότερες εμφανίσεις ανά ηλικία οχήματος, και τέλος, μία που αφορά την ηλικιακή ομάδα και το φύλο του οδηγού, καθώς και τον τύπο οχήματος που είναι πιο πιθανό να εμπλακεί σε ατύχημα, σύμφωνα με τη συχνότητα εμφάνισής τους στα στοιχεία που χρησιμοποιήθηκαν.

Top 10 Accidents count by vehicle type and engine capacity (cc)



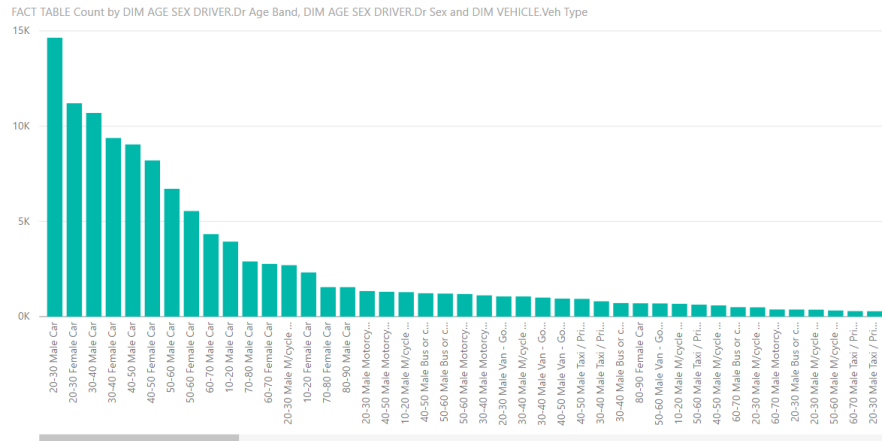
Εικόνα 5: Ατυχήματα ανά κυβισμό μηχανής

Top 10 vehicle ages with the most accidents per vehicle type



Εικόνα 6: Ατυχήματα ανά ηλικία οχήματος

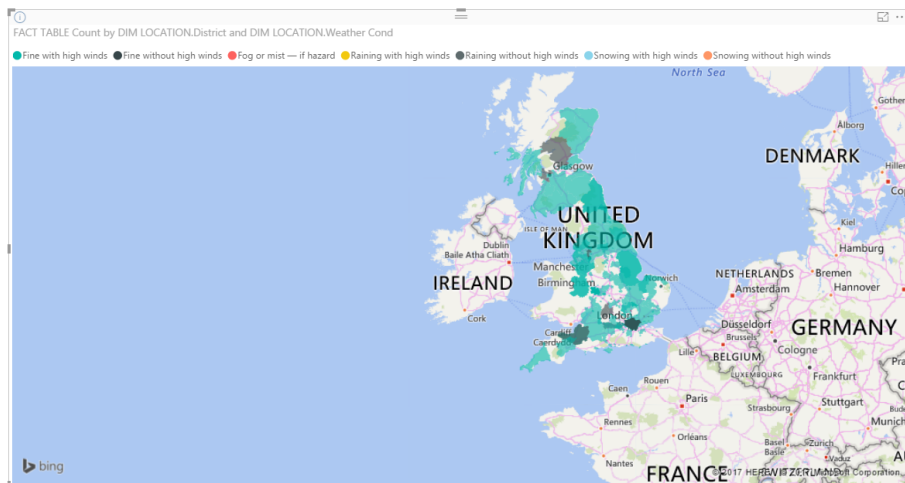
Driver age bands and gender with the most accidents per vehicle type



Εικόνα 7: Ατυχήματα ανά φύλο, ηλικιακή ομάδα οδηγού και τύπο οχήματος

Το παρακάτω ερώτημα αφορά την πλειονότητα των ατυχημάτων που συνέβησαν σε κάθε διοικητική περιφέρεια του Ηνωμένου Βασιλείου, ανάλογα με τις καιρικές συνθήκες που επικρατούσαν, οι οποίες μπορούν να καθοριστούν από το χρήστη. Μέσω της οπτικοποίησης αυτής, μπορεί να γίνει αντιληπτό ποιές περιοχές έχουν πρόβλημα με κάποιο συγκεκριμένο τύπο καιρικών συνθηκών, ώστε να λάβουν τα αντίστοιχα μέτρα οδικής κυκλοφορίας την επόμενη φορά που θα επαναληφθεί το φαινόμενο αυτό.

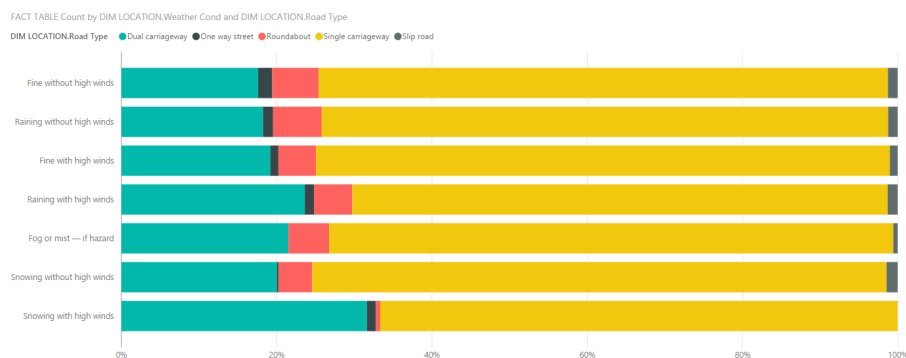
Majority of accidents in each district per weather type



Εικόνα 8: Ατυχήματα ανά περιφέρεια και καιρικές συνθήκες

Το τελευταίο ερώτημα που επιλέχθηκε αφορά την ποσοστιαία παρουσίαση των ατυχημάτων του εκάστοτε τύπου δρόμου και καιρικών συνθηκών που επικρατεί σε αυτόν. Η νοοτροπία πίσω από την επιλογή του ερωτήματος αυτού είναι περίπου ίδια με πριν, ωστόσο σε αυτή την περίπτωση συγκεκριμενοποιείται το είδος δρόμου και τα ατυχήματα που συνέβησαν σε αυτόν, και όχι η διοικητική περιφέρεια στην οποία έλαβαν χώρα διάφορα ατυχήματα.

Accidents percentages for every road/weather condition per road type



Εικόνα 9: Ποσοστιαία αναπαράσταση ατυχημάτων ανά τύπο δρόμου και καιρικές συνθήκες

Φυσικά, τα ερωτήματα αυτά είναι απλώς ενδεικτικά, αφού τόσο ο κύβος, όσο και η βάση δεδομένων που δημιουργήθηκε μπορεί να χρησιμοποιηθεί για την απάντηση τόσο σε πιο σύνθετα, όσο και σε απλούστερα ερωτήματα. Ουσιαστικά, κάθε χρήστης μπορεί να απαντήσει στα ερωτήματα που επιθυμεί, καθώς η βάση δεδομένων και ο κύβος που δημιουργήθηκε είναι ευέλικτα εργαλεία που μπορούν να χρησιμοποιηθούν, ανάλογα με τις ανάγκες κάθε χρήστη. **Στο τέλος της αναφοράς, επισυνάπτεται ένα link, το οποίο οδηγεί στην αναλυτικότερη παρουσίαση των δεδομένων, μέσω του Power BI**

3 Υλοποίηση bonus ερωτήματος

3.1 Δημιουργία DW και μετατροπή δεδομένων

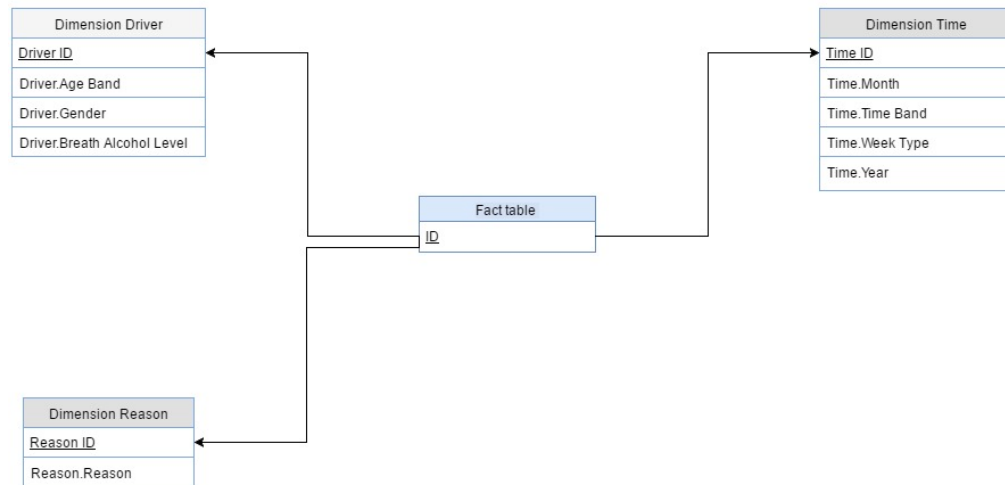
Για το ερώτημα αυτό, αρχικά έγινε προσπάθεια σύνδεσης των δεδομένων που αφορούν τα ποσοστά αλκοόλης στην αναπνοή των οδηγών με τα ατυχήματα που παρουσιάστηκαν παραπάνω. Ωστόσο, η σύνδεση αυτή δεν είναι δυνατή, αφού τα παραπάνω δεδομένα αναφέρονται σε καταγεγραμμένα ατυχήματα στο οδικό δίκτυο της Μεγάλης Βρετανίας, ενώ αυτά αναφέρονται στην προσότητα αλκοόλης στην αναπνοή των οδηγών, στους ελέγχους που έγιναν σε αυτούς, είτε λόγω παραβιάσεων του κώδικα οδικής κυκλοφορίας(K.O.K.), είτε λόγω κάποιου προυπάρχοντος αστυνομικού μπλόκου σε κάποια περιοχή του οδικού δικτύου. Τα δεδομένα δεν σχετίζονται μεταξύ τους και για το λόγο αυτό μελετήθηκαν ξεχωριστά, ενώ δημιουργήθηκε μία νέα βάση δεδομένων, αλλά και ένας νέος κύβος για αυτά.

Η δημιουργία του data warehouse του ερωτήματος αυτού έγινε με παρόμοιο τρόπο με αυτόν που χρησιμοποιήθηκε παραπάνω. Τα δεδομένα τοποθετήθηκαν σε διαστάσεις, αναλόγως με το σκοπό και το περιεχόμενο που εξυπηρετούν. Έτσι, δημιουργήθηκε μία διάσταση που περιλαμβάνει στοιχεία σχετικά με τον οδηγό του οχήματος. Πιο συγκεκριμένα, στη διάσταση αυτή τοποθετήθηκαν στοιχεία που διευκρινίζουν την ηλικιακή ομάδα του οδηγού, το φύλο του και την ποσότητα αλκοόλ που βρέθηκε στην αναπνοή του κατά την ώρα του ελέγχου. Επιπλέον, δημιουργήθηκε μία διάσταση που αναφέρεται στο λόγο που κλήθηκε ο εκάστοτε οδηγός για έλεγχο, και τέλος μία διάσταση που σχετίζεται με το χρόνο του ελέγχου, στην οποία περιλαμβάνονται στοιχεία που αφορούν το έτος, το μήνα, την ζώνη ώρας ελέγχου και τέλος τον τύπο ημέρας που παίρνει δυαδικές τιμές(Καθημερινή, Σαββατοκύριακο). Όπως ακριβώς και προηγουμένως, οι διαστάσεις αυτές συνδέονται μεταξύ τους μέσω του ID που περιέχεται στον πίνακα fact table.

Όσον αφορά τη μετατροπή των δεδομένων και τον καθορισμό των επιτρεπτών τιμών τους, χρησιμοποιήθηκαν τα ίδια εργαλεία με το προηγούμενο ερώτημα. Οι επιτρεπτές τιμές έγιναν γνωστές από τον οδηγό δεδομένων που συμπεριλαμβάνει μαζί με το αρχείο τους σε μορφή csv η Βρετανική Κυβέρνηση στην ιστοσελίδα της.

Το schema του data warehouse των δεδομένων αυτών παρατίθεται παρακάτω.

Digital Breath DW schema



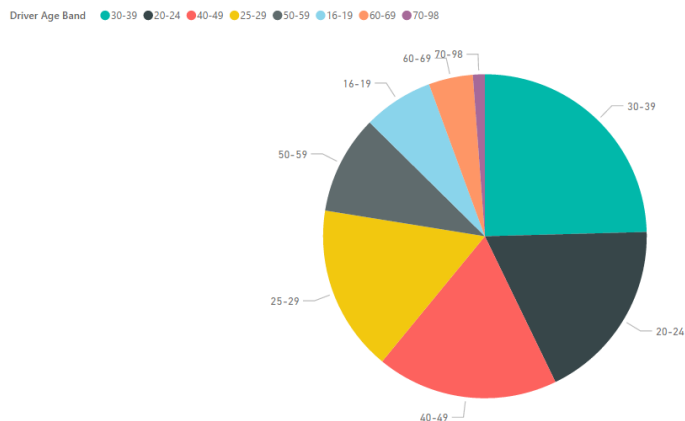
Εικόνα 10: Σχήμα του DW των δεδομένων ποσότητας αλκοόλ στην αναπνοή του οδηγού

3.2 Ενδεικτική οπτικοποίηση δεδομένων

Στην περίπτωση μελέτης των δεδομένων αυτών επιλέχθηκαν οπτικοποιήσεις δεδομένων που αφορούν κυρίως στατιστικά και γενικότερα χαρακτηριστικά, αφού δεν επιτρέπεται η εις βάθος συμπερασματική μελέτη τους από την ίδια τη φύση των δεδομένων που παρέχονται.

Πιο συγκεκριμένα, αρχικά επιλέχθηκε η οπτικοποίηση σε διάγραμμα πίτας των ηλικιακών ομάδων που εμφανίζονται στις παραβιάσεις και τους ελέγχους της αστυνομίας.

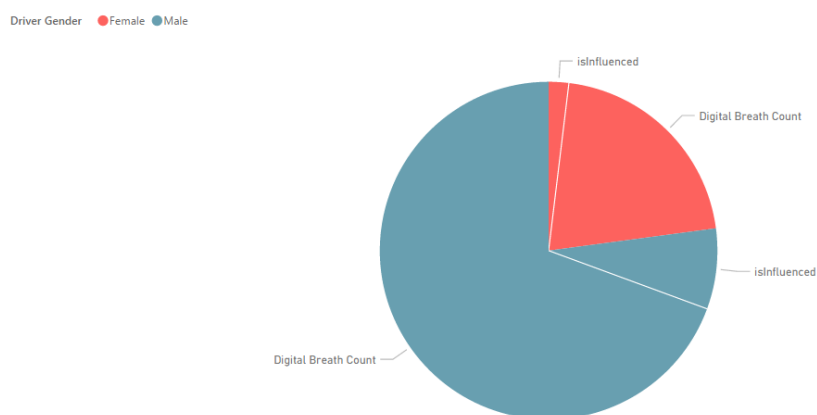
Alcohol influenced drivers per driver's age band



Εικόνα 11: Αναπαράσταση των ηλικιακών ομάδων που παρευρέθηκαν στους ελέγχους αλκοόλ

Επιπλέον, οπτικοποιήθηκαν δεδομένα, και πάλι σε διάγραμμα πίτας, που αναπαριστούν το φύλο και το πλήθος των οδηγών που πραγματοποίησαν μέρος των ελέγχων, και το πλήθος που ήταν πάνω από το επιτρεπτό όριο (35mg/αναπνοή). Παρατηρείται ότι περίπου το 9.2% των γυναικών και το 11.1% των ανδρών οδηγούσαν αφού είχαν καταναλώσει περισσότερη από την επιτρεπτή ποσότητα αλκοόλ.

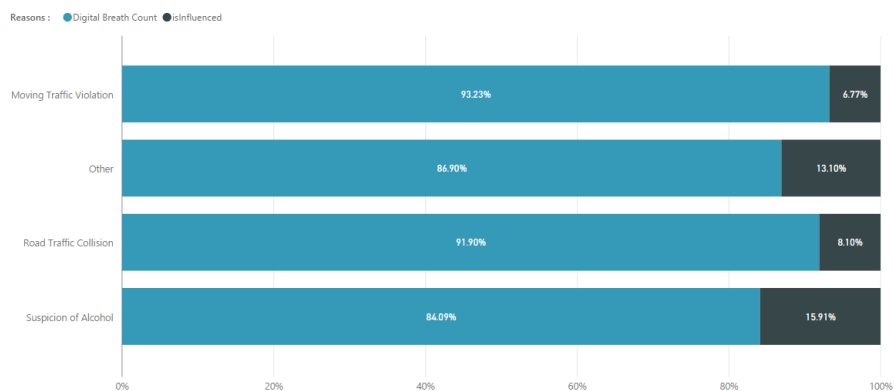
Male vs Female driver alcohol influence (over 35mg/breath)



Εικόνα 12: Αποτελέσματα γυναικών και ανδρών στους ελέγχους που πραγματοποιήθηκαν

Η επόμενη οπτικοποίηση απεικονίζει το ποσοστό των οδηγών που ήταν υπό την επήρεια αλκοόλ, ανάλογα με το λόγο για τον οποίο ελέγχθηκαν. Με σκούρο μπλέ χρώμα απεικονίζεται το ποσοστό των οδηγών που ήταν πάνω από το επιτρεπτό όριο, ενώ με ανοιχτό μπλε χρώμα απεικονίζεται το ποσοστό των οδηγών που δεν ήταν πάνω από την επιτρεπτή τιμή.

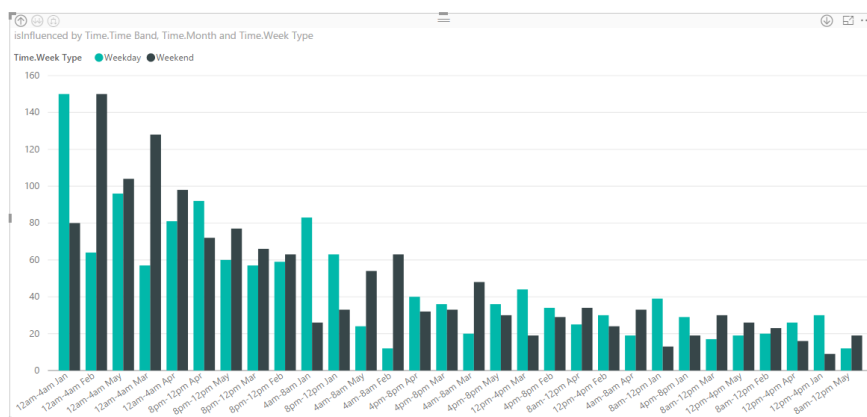
Police checks vs alcohol influence percentages (over 35mg/breath)



Εικόνα 13: Ποσοστιαία αναπαράσταση οδηγών υπό την επήρεια αλκοόλ, ανάλογα με το λόγο ελέγχου

Τέλος, οπτικοποιούνται οι πέντε μήνες, και οι ζώνες ώρας τους με το μεγαλύτερο πλήθος οδηγιών πάνω από το επιτρεπτό όριο.

Top 5 months with the most incidents of alcohol influence per time band



Εικόνα 14: Αναπαράσταση πέντε μηνών και ζωνών ώρας με το μεγαλύτερο πλήθος οδηγιών πάνω από το επιτρεπτό όριο κατανάλωσης αλκοόλ

Ομοίως με παραπάνω, οι οπτικοποιήσεις αυτές που παρατέθηκαν, αποτελούν ενδεικτικές απεικονίσεις των δεδομένων που επεξεργάστηκαν. Οι πληροφορίες και τα δεδομένα που παρουσιάζονται μπορούν, όπως και παραπάνω, να τροποποιηθούν προκειμένου να ανταποκρίνονται στις ανάγκες του χρήστη. Στο τέλος της αναφοράς, επισυνάπτεται ένα link, το οποίο οδηγεί στην αναλυτικότερη παρουσίαση των δεδομένων αυτών, μέσω του Power BI

Αναφορές

- [1] Φόρμα συμπλήρωσης πληροφοριών ατυχημάτων στο Ηνωμένο Βασίλειο
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/230590/stats19.pdf
- [2] Δεδομένα ατυχημάτων στο Ηνωμένο Βασίλειο
<https://data.gov.uk/dataset/road-accidents-safety-data>
- [3] Πληροφορίες σχετικά με τα όρια κατανάλωσης αλκοόλ στο Ηνωμένο Βασίλειο
<https://www.gov.uk/drink-drive-limit>
- [4] Παρουσίαση δεδομένων ατυχημάτων στο Ηνωμένο Βασίλειο μέσω **Power BI**
- [5] Αναλυτική παρουσίαση δεδομένων ελέγχων αλκοόλ στο Ηνωμένο Βασίλειο μέσω **Power BI**