

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS

## **Ανάπτυξη και Ασφάλεια Πληροφοριακών Συστημάτων**

### **Συστήματα Ανάλυσης και Διαχείρισης Μεγάλων Δεδομένων**

#### **Project 1: Αποθήκες Δεδομένων**

Ονοματεπώνυμο: Ελένη Τράμπαρη Λάρδα

Αριθμός Μητρώου: f3312217

## Περιεχόμενα

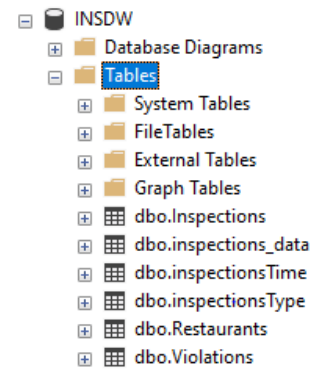
<b>Ζήτημα Πρώτο</b>	<b>3</b>
Ερώτημα 1	3
Ερώτημα 2	3
Ερώτημα 3	4
Ερώτημα 4	4
<b>Ζήτημα Δεύτερο</b>	<b>5</b>
Ερώτημα 1	5
Ερώτημα 2	5
Ερώτημα 3	5
<b>Ζήτημα Τρίτο</b>	<b>6</b>
Ερώτημα 1	6
1.1.	6
1.2.	7
1.3.	7
Ερώτημα 2	8

## Ζήτημα Πρώτο

### Ερώτημα 1

Να δημιουργήσετε μία βάση δεδομένων με όνομα INSDW (Inspections Data Warehouse). Στη συνέχεια να δημιουργήσετε τον πίνακα inspections\_data στον οποίο να φορτώσετε τα δεδομένα του αρχείου inspections\_data.txt.

```
CREATE TABLE inspections_data(  
    rid int,  
    lat float,  
    lon float,  
    insdate date,  
    insyear int,  
    insmonth int,  
    insday int,  
    insweekday int,  
    inscode int,  
    instype nvarchar(100),  
    criticalIssue int,  
    nonCriticalIssue int,  
    vcode int,  
    vdescription nvarchar(255),  
    vcategory nvarchar(255)  
);  
  
BULK INSERT inspections_data  
FROM 'C:\Users\User\Desktop\inspections_data.txt'  
WITH (DATAFILETYPE = 'widechar', FIRSTROW =2, FIELDTERMINATOR= '|',  
ROWTERMINATOR = '\n');
```



### Ερώτημα 2

Να υλοποιήσετε το λογικό σχήμα της αποθήκης δεδομένων το οποίο θα πρέπει να έχει την μορφή αστέρα (Star Schema).

```
create table inspectionsTime(  
    insdate date primary key,  
    insyear int,  
    insmonth int,  
    insday int,  
    insweekday int  
);  
  
create table inspectionsType(  
    inscode int primary key,  
    instype nvarchar(100),  
);  
  
create table Violations(  
    vcode int primary key,  
    vdescription nvarchar(255),  
    vcategory nvarchar(255)  
);  
  
create table Inspections(  
    rid int foreign key references Restaurants (rid),  
    insdate date foreign key references inspectionsTime (insdate),  
    inscode int foreign key references inspectionsType (inscode),  
    vcode int foreign key references Violations (vcode),  
    criticalIssue int,  
    nonCriticalIssue int  
    PRIMARY KEY (rid,insdate,inscode,vcode)  
);
```

### Ερώτημα 3

Να γράψετε κατάλληλες εντολές σε γλώσσα SQL, οι οποίες θα τροφοδοτούν το σχήμα της αποθήκης με τα απαραίτητα στοιχεία από τον πίνακα inspections\_data.

```
insert into Restaurants
select distinct rid,lat,lon
from inspections_data
```

```
insert into inspectionsTime
select distinct insdate, insyear,insmonth,insday,insweekday
from inspections_data
```

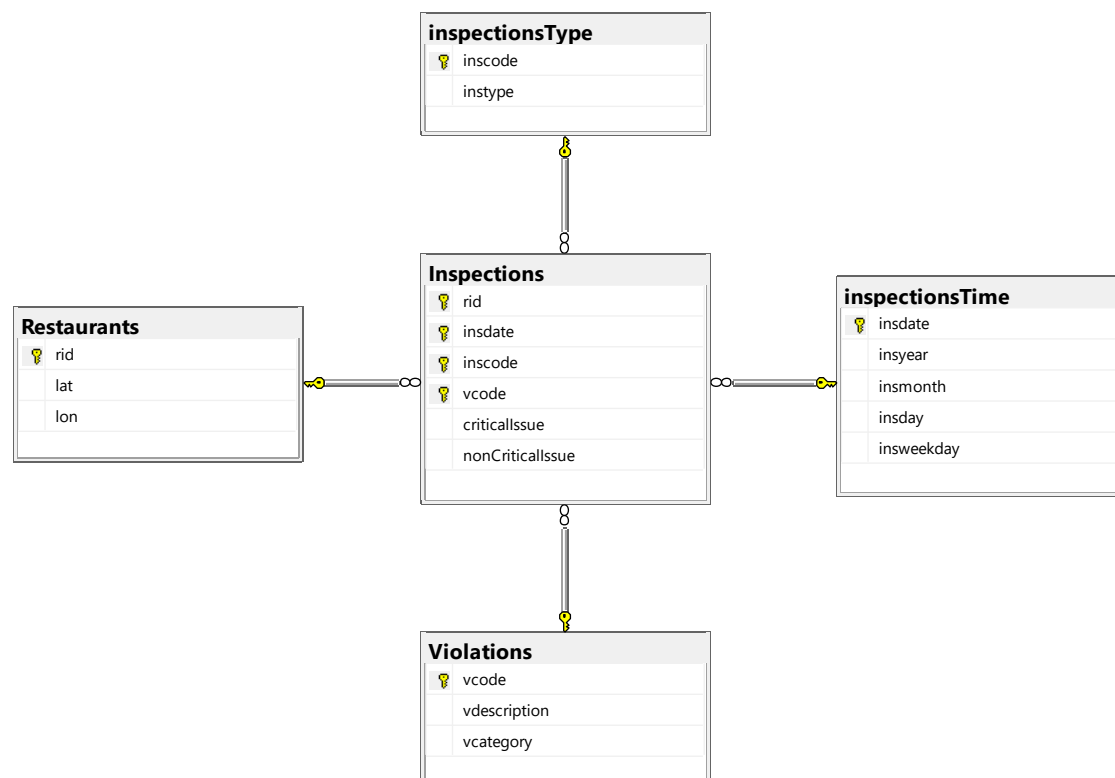
```
insert into inspectionsType
select distinct inscode,instype
from inspections_data
```

```
insert into Violations
select distinct vcode, vdescription, vcategory
from inspections_data
```

```
insert into Inspections
select rid,insdate,inscode,vcode,criticalIssue,nonCriticalIssue
from inspections_data
```

### Ερώτημα 4

Να αναπαραστήσετε διαγραμματικά το σχήμα της αποθήκης χρησιμοποιώντας την επιλογή "Database diagrams" του SQL Server Management Studio.



## Ζήτημα Δεύτερο

### Ερώτημα 1

Εμφανίστε έναν κατάλογο με τον αριθμό των επιθεωρήσεων ανά έτος και τύπο επιθεώρησης (instype). Ο κατάλογος πρέπει να είναι ταξινομημένος με βάση το έτος σε φθίνουσα διάταξη.

```
|select insyear, instype, count(Inspections.inscode) as TotalInspections  
from inspectionsTime, inspectionsType, Inspections  
where Inspections.insdate=inspectionsTime.insdate and Inspections.inscode=inspectionsType.inscode  
group by insyear, instype  
order by insyear DESC
```

### Ερώτημα 2

Εμφανίστε έναν κατάλογο με τα παρακάτω στοιχεία για κάθε εστιατόριο:

- Κωδικός εστιατορίου
- Συντεταγμένες
- Αριθμός μη κρίσιμων ζητημάτων που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.
- Αριθμός κρίσιμων ζητημάτων που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.
- Αριθμός ζητημάτων (κρίσιμων και μη κρίσιμων) που εντοπίστηκαν συνολικά από όλες τις επιθεωρήσεις που έγιναν στο συγκεκριμένο εστιατόριο.

```
select Restaurants.rid, lat, lon, sum(nonCriticalIssue) as nonCriticalIssues, sum(criticalIssue) as criticalIssues, sum(nonCriticalIssue+criticalIssue) as totalIssues  
from Restaurants, Inspections  
where Restaurants.rid=Inspections.rid  
group by Restaurants.rid, lat, lon  
order by totalIssues DESC
```

### Ερώτημα 3

Γράψτε μια επερώτηση σε γλώσσα SQL το αποτέλεσμα της οποίας είναι η δημιουργία ενός κύβου (data cube), κάθε κελί του οποίου περιέχει τον συνολικό αριθμό των κρίσιμων ζητημάτων που εντοπίστηκαν από όλες τις επιθεωρήσεις ανα κατηγορία παράβασης (vcategory) και έτος επιθεώρησης (insyear).

```
select vcategory, insyear, sum(criticalIssue) as criticalIssues  
from Inspections, inspectionsTime, Violations  
where Inspections.insdate=inspectionsTime.insdate and Violations.vcode=Inspections.vcode  
group by cube (vcategory, insyear)
```

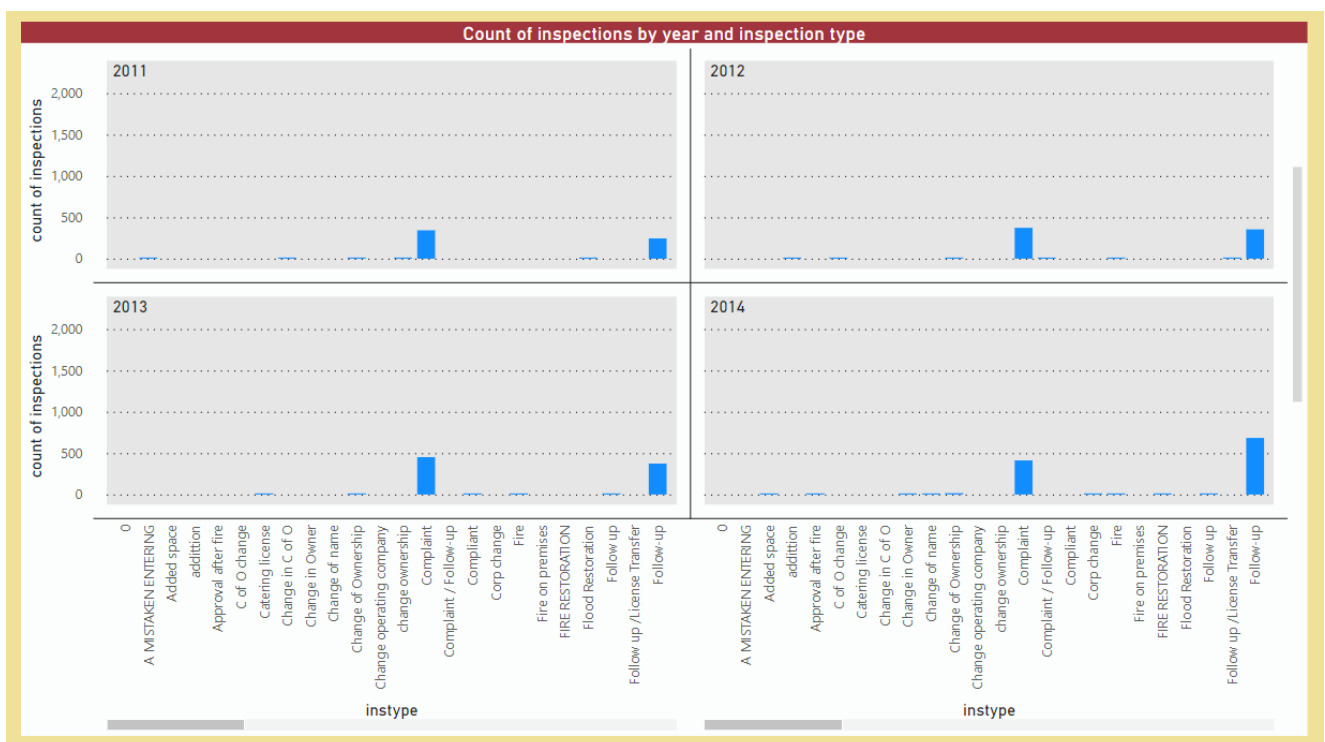
## Ζήτημα Τρίτο

### Ερώτημα 1

Δημιουργήστε μια αναφορά (report) με το power BI με τα παρακάτω:

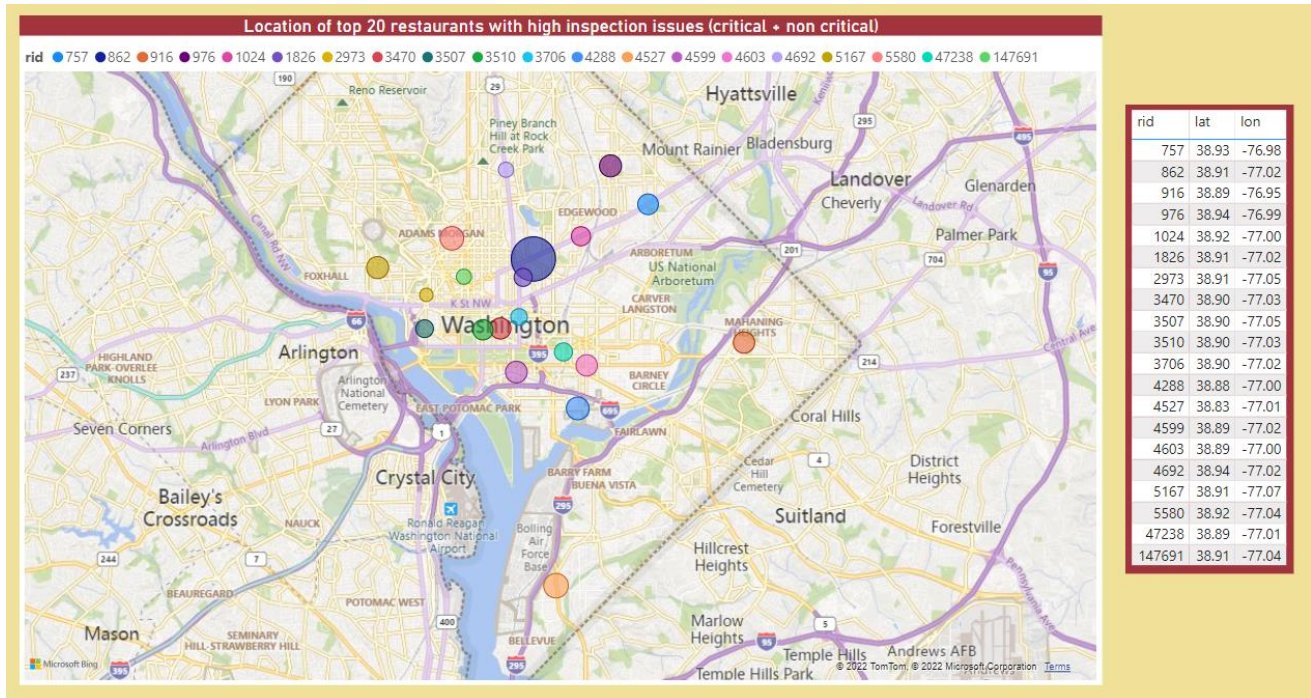
#### 1.1.

Κατάλληλο γράφημα στο οποίο θα απεικονίζονται τα αποτελέσματα της πρώτης επερώτησης που γράψατε στο προηγούμενο ζήτημα. Δηλαδή ο αριθμός των επιθεωρήσεων ανά έτος και τύπο επιθεώρησης.



## 1.2.

Απεικονίστε σε έναν χάρτη την τοποθεσία των είκοσι εστιατορίων στα οποία εντοπίστηκαν τα περισσότερα ζητήματα (κρίσιμα και μη) από όλες τις επιθεωρήσεις. Πρόκειται για τα εστιατόρια του καταλόγου που δημιουργήσατε στο δεύτερο ερώτημα του προηγούμενου ζητήματος. Ο χάρτης να συνοδεύεται από ένα τρίστηλο πίνακα κάθε γραμμή του οποίου θα περιέχει τον κωδικό και τις συντεταγμένες ενός εστιατορίου.



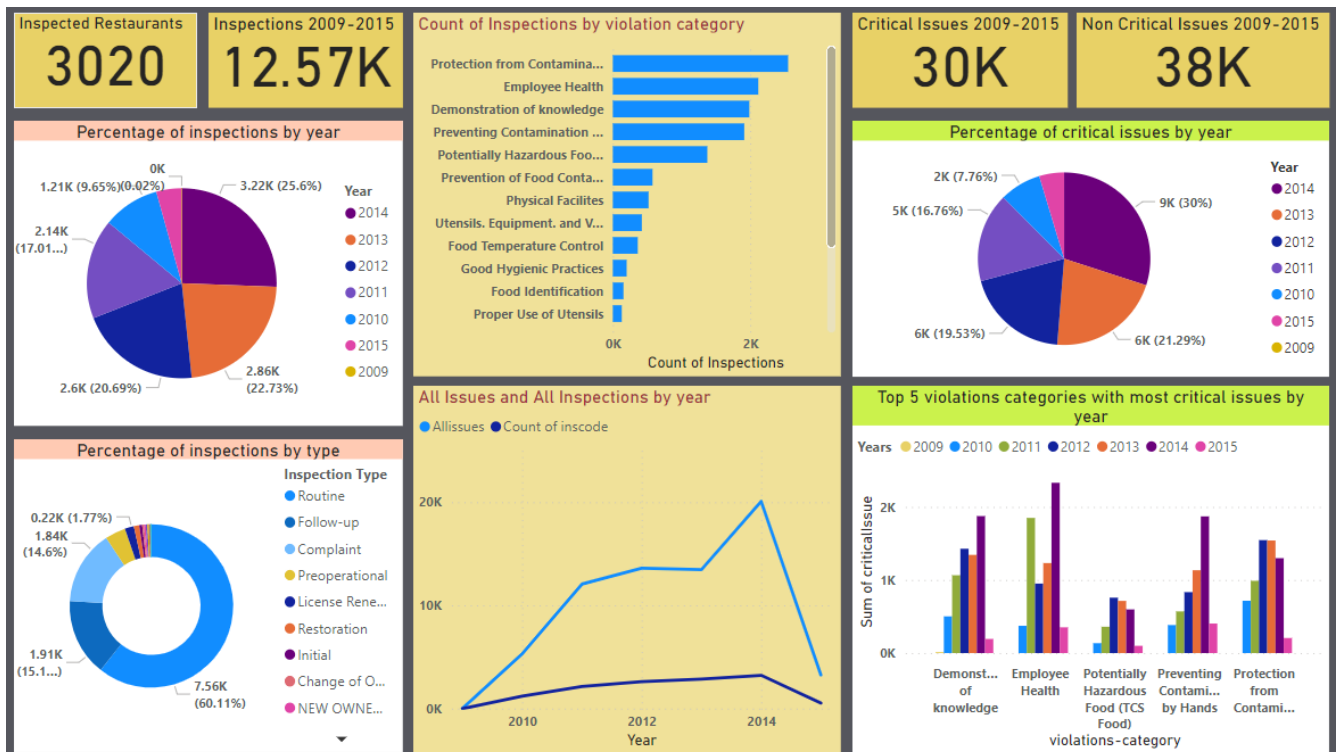
## 1.3.

Κατάλληλο γράφημα ή/και πίνακα για την αναπαράσταση των δεδομένων του κύβου που ζητείται στο προηγούμενο ζήτημα.

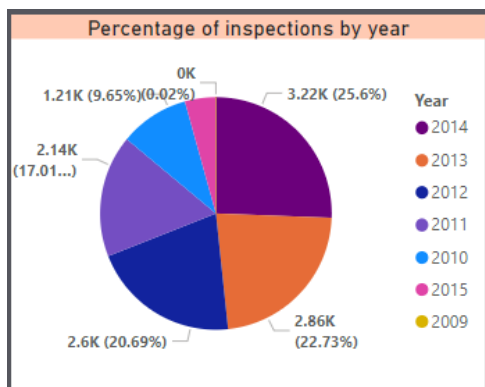
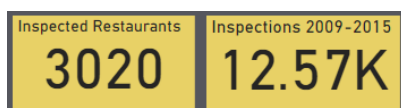
Total of critical issues by violation category and inspection year								
vcategory	2009	2010	2011	2012	2013	2014	2015	Total
Approved Source		56	42	56	92	143	26	415
Chemical		3	1	7	13	30		54
Conformance with Approved Procedures		2	3	7	9	20	0	41
Consumer Advisory		20	10	12	9	11		62
Demonstration of knowledge	1	502	1065	1428	1344	1875	193	6408
Employee Health		373	1850	951	1232	2331	352	7089
Food Identification		5	5	11	11	15	0	47
Food Temperature Control		19	31	30	18	33	5	136
Good Hygienic Practices		58	43	117	131	536	76	961
Physical Facilities		26	28	46	52	131	24	307
Potentially Hazardous Food (TCS Food)		135	360	758	714	597	99	2663
Preventing Contamination by Hands		383	570	834	1134	1871	404	5196
Prevention of Food Contamination		23	22	45	73	118	11	292
Proper Use of Utensils		3	4	5	3	7	3	25
Protection from Contamination		716	989	1546	1541	1299	206	6297
Safe Food and Water		0	1	0	2	1		4
Utensils, Equipment, and Vending	2	14	26	30	35	19	4	130
<b>Total</b>	<b>3</b>	<b>2338</b>	<b>5050</b>	<b>5883</b>	<b>6413</b>	<b>9037</b>	<b>1403</b>	<b>30127</b>

## Ερώτημα 2

Εξερευνήστε τα δεδομένα της αποθήκης ώστε να εντοπίσετε επιπλέον πληροφορίες οι οποίες κατά την γνώμη σας είναι χρήσιμες για την υπηρεσία δημόσιας υγιεινής. Δημιουργήστε ένα καλαίσθητο dashboard με κατάλληλα γραφήματα και άλλα στοιχεία παρουσίασης (π.χ. χάρτες, πίνακες κ.λπ.) στο οποίο να απεικονίζονται οι πλέον χρήσιμες κατά την γνώμη σας πληροφορίες.



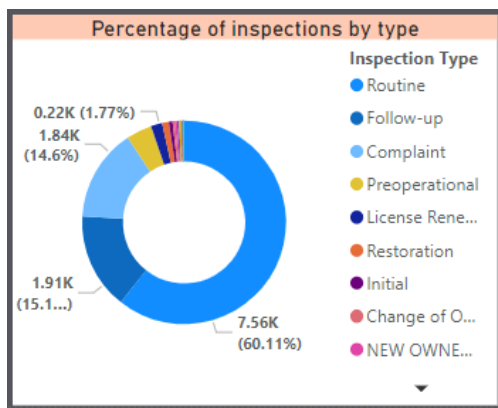
### Διαγράμματα – επεξήγηση :



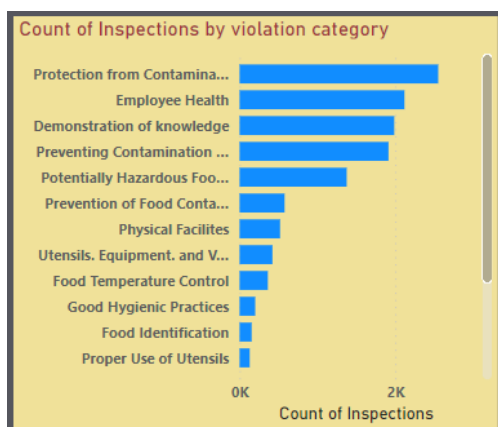
- Τα εστιατόρια που έχουν ελεγχθεί στη διάρκεια 7 ετών καθώς και το πλήθος των ελέγχων αυτή τη περίοδο. Έτσι, αντιλαμβανόμαστε τη τάξη μεγέθους του δείγματός μας.

- Ποσοστό των ελέγχων ανά έτος, από όπου διαπιστώνουμε ότι στα έτη 2009 και 2015 έχουμε μικρό πλήθος ελέγχων στο δείγμα (γιατί πιθανόν δεν αναφέρεται σε όλους τους μήνες του έτους). Δηλαδή το 95% των ελέγχων έχει γίνει στα έτη 2010-2014.

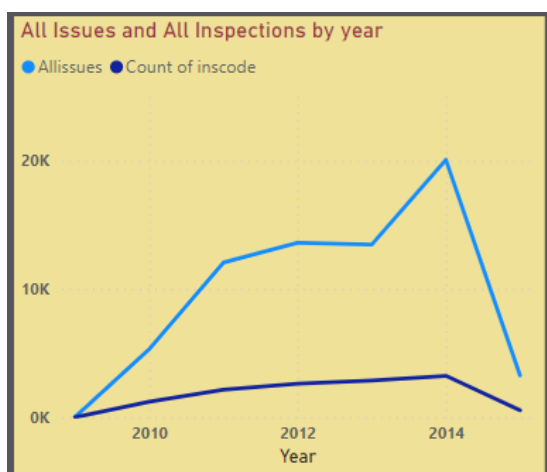




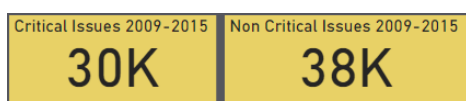
- Ποσοστό των ελέγχων ανά τύπο επιθεωρήσεων, όπου φαίνεται ποιες είναι οι πιο συχνές κατηγορίες ελέγχων στη χρονική διάρκεια του δείγματος. Διαπιστώνουμε ότι περίπου το 95% των ελέγχων ανήκει στις κατηγορίες Routine, Follow-up, Complain & Preoperational.



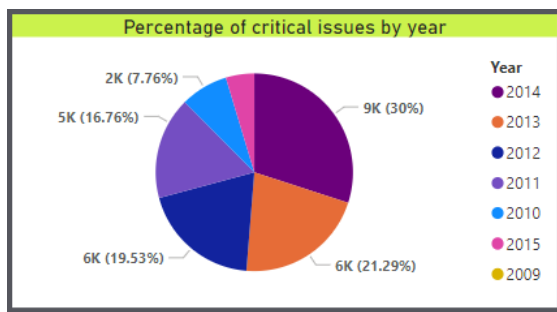
- Πλήθος των ελέγχων/επιθεωρήσεων ανά κατηγορία παράβασης. Διαπιστώνουμε ότι οι 5 πρώτες κατηγορίες του διαγράμματος αφορούν περίπου το 80% (9.946 από 12.571) των ελέγχων .



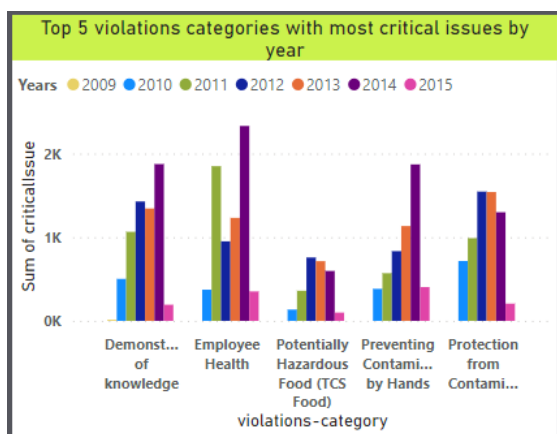
- Καμπύλη πορείας του πλήθους των κρίσιμων και μη κρίσιμων ζητημάτων και ταυτόχρονη καμπύλη πορείας του πλήθους των επιθεωρήσεων στη διάρκεια των ετών. Διαπιστώνουμε ότι ο ρυθμός αύξησης και μείωσης των ζητημάτων είναι αισθητά μεγαλύτερος από το ρυθμό αύξησης των επιθεωρήσεων.



- Πλήθος των κρίσιμων και μη κρίσιμων ζητημάτων στο σύνολο των ετών, όπου φαίνεται ο πολύ μεγάλος όγκος τους.



- Ποσοστό των κρίσιμων ζητημάτων ανά έτος. Διαπιστώνουμε να αυξάνεται το μερίδιο για τα έτη 2010-2014. Για τα έτη 2009, 2015 δε χρειάζεται να βγάλουμε συμπεράσματα λόγω μικρού δείγματος. Το 2011 παρουσιάζεται υπερδιπλασιασμός του μεριδίου, τα έτη 2012, 2013 παρουσιάζουν μικρή αύξηση, ενώ του 2014 αυξάνονται περίπου κατά 50%.



- Ραβδόγραμμα των 5 πρώτων κατηγοριών παραβάσεων που έχουν το μεγαλύτερο πλήθος κρίσιμων ζητημάτων και πως αυτό μεταβάλλεται ανά έτος για τη κάθε κατηγορία.