Faculty of
Computer & Informatics Eng.
Department of Software &
Information Systems Eng.

كلية
هندسة الحاسوب والمعلوماتية
قسم هندسة البرمجيات
ونظم المعلومات

# Building Business Intelligence Solution for The Syrian Ministry of Interior (Traffic Violations)

# بناء نظام ذكاء أعمال لـ وزارة الداخلية السورية (المخالفات المرورية)

**Prepared by**

**Ahmad Yones**

**Ibrahim Hafez**

**Oumama Nabaa**

**Supervised by**

**Dr. Mouhib Alnoukari**

**Academic Year**

**2020-2021**

# SUPERVISION CERTIFICATION

I certify that the preparation of this project entitled **Building Business Intelligence Solution for The Syrian Ministry of Interior (Traffic Violations)**, prepared by **Ahmad Yones**, **Ibrahim Hafez**, and **Oumama Nabaa** was made under my supervision at *Department of Software & Information System Eng. – Faculty of Computer & Informatics Engineering* in partial fulfillment of the Requirements for the Degree of *Bachelors of Software & Information System Engineering.*

X

Dr. Mouhib Alnoukari
Supervisor
Sunday, February 6, 2022

# Abstract:

Due to the rapid growth of new technologies, the Business Intelligence (BI) market is growing as well. Adoption of Business Intelligence system has become one of the most important technological and organizational innovations in modern organizations, and cornerstone of decision-making processes.

This project consists of implementing a business intelligence solution to the Ministry of Interior of Syria data, which increases decisions accuracy by providing accurate knowledge and easy-to-read reports to decision-makers.

This is accomplished by building a data warehouse relying on the extracted data, then analysis and business intelligence tools are applied to acquire the maximum benefits.

# ملخص:

نظراً للنمو السريع للتقنيات الجديدة، فإن سوق ذكاء الأعمال مستمر بالنمو أيضاً، فأصبح اعتماد نظام ذكاء الأعمال أحد أهم الابتكارات التكنولوجية والتنظيمية في المنظمات الحديثة، وحجر الأساس في عمليات صنع القرار.

يتكون هذا المشروع من تطبيق حلول ذكاء الأعمال لبيانات وزارة الداخلية السورية، مما يزيد من دقة القرارات من خلال توفير معرفة دقيقة وتقارير سهلة القراءة لصناع القرار.

يتم تحقيق ذلك من خلال بناء مستودع بيانات يعتمد على البيانات المستخرجة من الوزارة، ثم يتم تطبيق أدوات التحليل وذكاء الأعمال للحصول على الفائدة القصوى.

# Contents

**Contents** ..... V

**List Of Figures:** ..... VI

**Chapter 1 Theoretical Study** ..... 7

1.1. Introduction: ..... 8

1.2. The Importance of The Project: ..... 9

1.3. The Goal of The Project: ..... 10

1.4. Concepts: ..... 11

1.4.1. Business Intelligence: ..... 11

1.4.2. Business Intelligence: ..... 13

1.4.3. Data Warehouse (DW): ..... 16

1.4.4. Extract, Transform, and Load: ..... 17

1.4.5. Data Mining: ..... 22

**Chapter 2 Design and Analytical Study (Implementation)** ..... 26

2.1. Introduction: ..... 27

2.2. Identify End-User Requirements: ..... 28

2.3. Identify the Data sources: ..... 29

2.4. ETL: ..... 30

2.5. Design the Data Model: ..... 35

2.6. Visualization: ..... 38

**Conclusion:** ..... 46

**References** ..... 47

# List Of Figures:

Figure 1.1 BI Cycle (EMC Education Services (2014), Data Science and Big Data Analytics, p. 1)..... 11

Figure 1.2 BI environment (www.researchgate.net) ..... 12

Figure 1.3 Simpler BI environment (bimentalist.com) ..... 13

Figure 1.4 Data Warehouse Architecture (www.datawarehouse4u.info) ..... 16

Figure 1.5 ETL tool (thecustomizewindows.com) ..... 17

Figure 1.6 Data Integration Cycle (www.mobcoder.com) ..... 19

Figure 1.7 Loading changes to a dimension table (www.geeksfogeeks.com) ..... 21

Figure 1.8 Level of Analytics ..... 23

Figure 2.1 SQL loader environment ..... 37

# Chapter 1
# Theoretical Study

## 1.1. Introduction:

The world is becoming more and more data-driven, with endless amounts of data available to work with. Big companies like Google and Microsoft use data to make decisions, but they're not the only ones.

The importance of data analytics in any sector is compounded, creating enormous quantities of knowledge that can provide useful insights into the field. In the last ten years, this has led to a surge in the data market. In order to gain decision-making insights, the compilation of data can be supplemented by its analysis. Data analytics help organizations and businesses gain insight into the enormous amount of knowledge they need for further production and growth.

*"Information is the oil of the 21st century, and analytics is the combustion engine."* – Peter Sondergaard, senior vice president, Gartner Research.

Business intelligence (BI) software can help by combining online analytical processing (OLAP), location intelligence, enterprise reporting, and more. BI software offers enterprise businesses the opportunity to connect disparate data sources into one unified source, collate and structure the data, and offer an interface for end-users to extract reports and dashboards that can drive more informed business decisions.

## 1.2. The Importance of The Project:

The importance of the research lies in the following point work on important and relatively recent concepts, which are business intelligence systems in decision support:

1.2.1. A massive Shortage of decision support systems in Syria in general and in the governmental sectors in particular.

1.2.2. Understanding Your Government Customer" report, as people continue to demand a higher level of accountability and transparency from their government.

1.2.3. implementing a system to improve the level of services provided to beneficiaries in the governmental sector.

1.2.4. The need to develop analytical systems in order to take advantages of the large amount of data present in this organization to reach sound management decisions.

1.2.5. Effectively managing the allocation and utilization of funds requires in-depth insight into spending patterns.

## 1.3. The Goal of The Project:

This project ensures the importance of building a business intelligence system in the Ministry of Interior, providing accurate knowledge for decision makers. goals in a nutshell:

1.3.1. Monitor optimal use of public funds.

1.3.2. Study Big data.

1.3.3. Task and workforce management.

## 1.4. Concepts:

### 1.4.1. Business Intelligence:

Business is the act of doing something productive to serve someone's needs, and thus earn a living and make the world a better place. Business activities are recorded on paper or using electronic media, and then these records become data. There is more data from customers' responses and on the industry as a whole. All this data can be analyzed and mined using special tools and techniques to generate patterns and intelligence, which reflect how the business is functioning. These ideas can then be fed back into the business so that it can evolve to become more effective and efficient in serving customer needs. And the cycle continues.

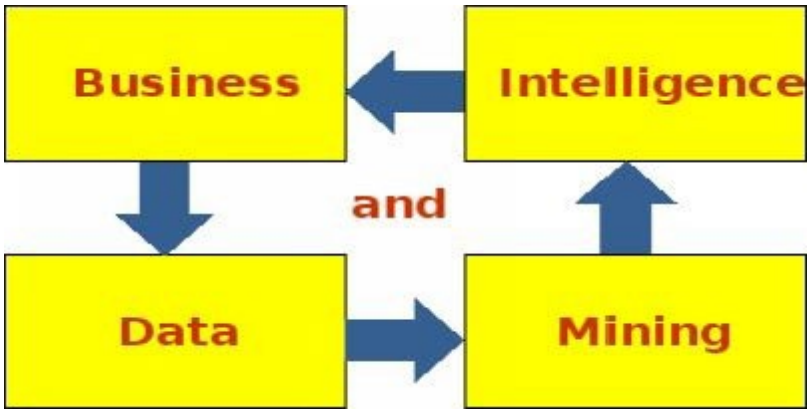



*Figure 1.1 BI Cycle[1] (EMC Education Services (2014), Data Science and Big Data Analytics, p. 1)*

BI is all about turning an organization's data into insights that can be used to inform business decisions. BI analysts will use BI tools, software or services to access and analyze datasets and translate their findings into reports, summaries, dashboards, graphs, charts or maps.

In recent years, the advent of modern data visualization and reporting tools has transformed the discipline, empowering businesses to use big data insights to identify, develop and create new business opportunities.

[1] EMC Education Services (2014), Data Science and Big Data Analytics

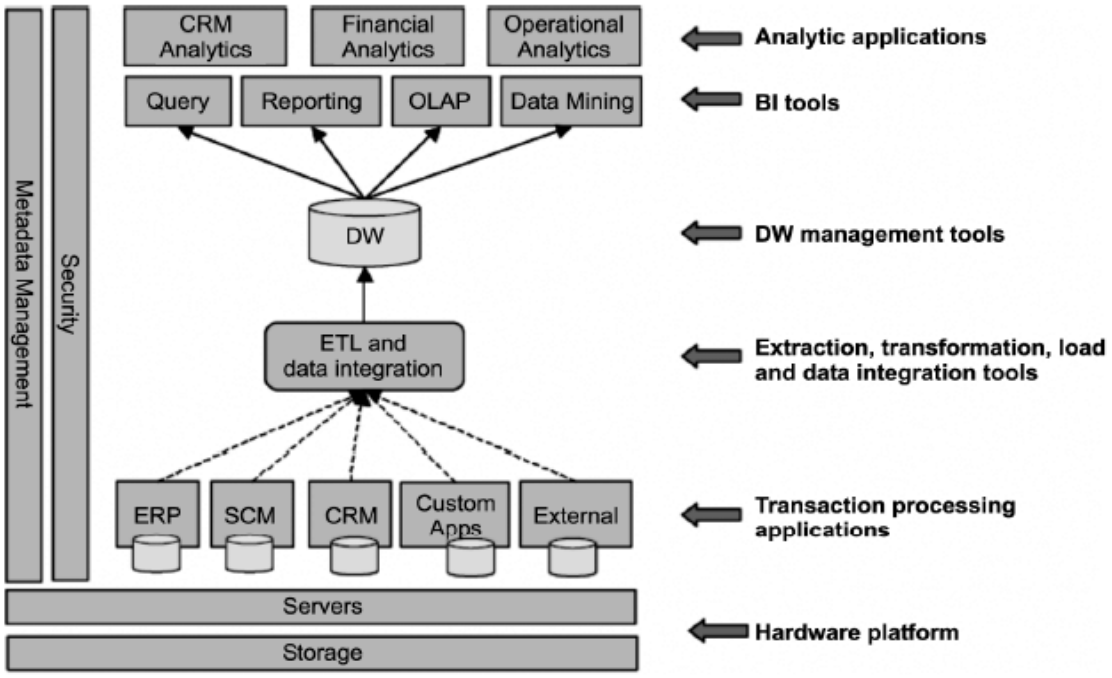



*Figure 1.2 BI environment (www.researchgate.net)*

The ultimate goal of BI initiatives is to drive better business decisions that enable organizations to increase revenue, improve operational efficiency and gain competitive advantages over business rivals. To achieve that goal, BI incorporates a combination of analytics, data management and reportingtools, plus various methodologies for managing and analyzing data.

BI data can include historical information and real-time data gathered from source systems as it's generated, enabling BI tools to support both strategic and tactical decision-making processes. Before it's used in BI applications, raw data from different source systems generally must be integrated, consolidated and cleansed using data integration and data quality management tools to ensure that BI teams and business users are analyzing accurate and consistent information.

### 1.4.2. Business Intelligence:

Business intelligence (BI) is all about converting large amounts of corporatedata into useful information, thereby triggering some profitable business action with the help of knowledge acquired through BI analysis.

Implements BI is a long process and it requires a lot of analysis and investment. A typical BI environment involves business models, data models, data sources, ETL, tools needed to transform and organize the data into useful information, target data warehouse, data marts, OLAP analysis and reporting tools.

Setting up a BI environment not only rely on tools, techniques and processes, it also requires skilled business people to carefully drive these in the right direction. Care should be taken in understanding the business requirements, setting up the targets, analyzing and defining the various processes associated with these, determining what kind of data needed to be analyzed, determining the source and target for that data, defining how to integrate that data for BI analysis and determining and gathering the tools to achieve this goal.

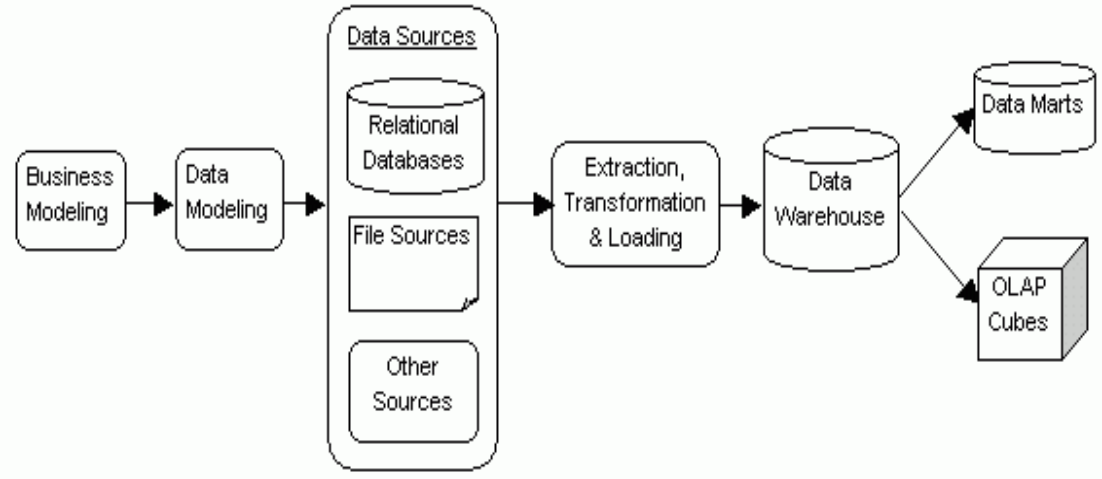



*Figure 1.3 Simpler BI environment (bimentalist.com)*

The importance of business intelligence continues to grow as businesses face an ever-increasing flow of raw data and the challenges of gaining insight from enormous volumes of information (big data). With the employment of business intelligence systems, businesses can

gain a comprehensive view of their organization's data and translate it into insights about their business processes, enabling improved and strategic business decisions. Business intelligence helps organizations analyze data with a historical context, optimize operations, track performance, accelerate and improve decision-making, identify and eliminate business problems and inefficiencies, identify market trends and patterns, drive new revenues and profitability, increase productivity and accelerate growth, analyze customer behavior, compare data with competitors, and ultimately gain a competitive advantage over rival businesses.

1- Faster analysis, intuitive dashboards.

2- Increased organizational efficiency.

3- Data-driven business decisions.

4- Improved customer experience.

5- Improved employee satisfaction.

6- Trusted and governed data.

Modern business intelligence systems prioritize self-service analysis, empowering businesses to gain insight into their market and improve performance with comprehensive data discovery tools, methods, processes, and platforms. Such business intelligence solutions include:

1- Ad hoc analytics: an analysis process designed to answer specific questions on the spot.

2- Online analytical processing (OLAP): a computing method that enables multi-dimensional analytical queries

3- Mobile BI: software that optimizes desktop business intelligence for mobile devices.

Real-time BI: a data analytics approach that delivers real-time information to users by feeding business transactions into a real-time data warehouse Operational BI: a data analysis approach that utilizes real-time business analytics to automatically integrate

real-time data into operational system for immediate use.

Software-as-a-service BI (SaaS BI): a cloud-hosted, subscription-based delivery model for business intelligence software solutions

Open-source BI (OSBI): business intelligence software solutions that do not require purchasing a software license

Collaborative BI: the merging of business intelligence software with collaboration tools in order to streamline the sharing process

Location intelligence (LI): software that is designed to relate geographic contexts to business data visualization software: facilitates the detection of patterns and correlations by providing visual context.

### 1.4.3. Data Warehouse (DW):

A Data Warehousing (DW) is process for collecting and managing data from varied sources to provide meaningful business insights. A Data warehouse is typically used to connect and analyze business data from heterogeneous sources. The data warehouse is the core of the BI system which is built for data analysis and reporting. It is a blend of technologies and components which aids the strategic use of data. It is electronic storage of a large amount of information by a business which is designed for query and analysis instead of transaction processing. It is a process of transforming data into information and making it available to users in a timely manner to make a difference.

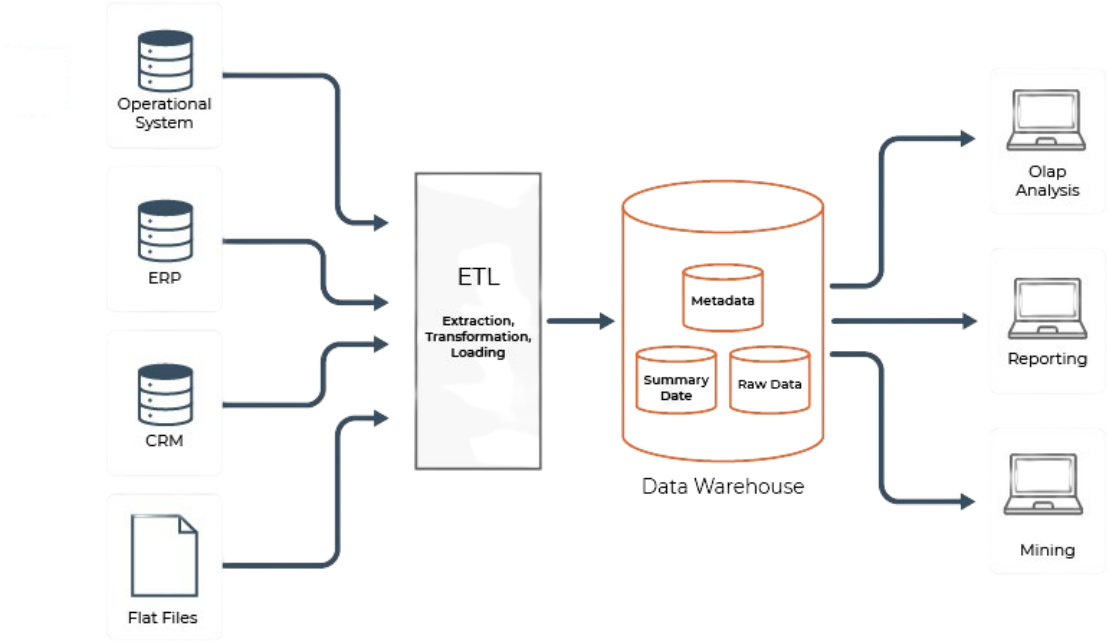



*Figure 1.4 Data Warehouse Architecture (www.datawarehouse4u.info)*

### 1.4.4. Extract, Transform, and Load:

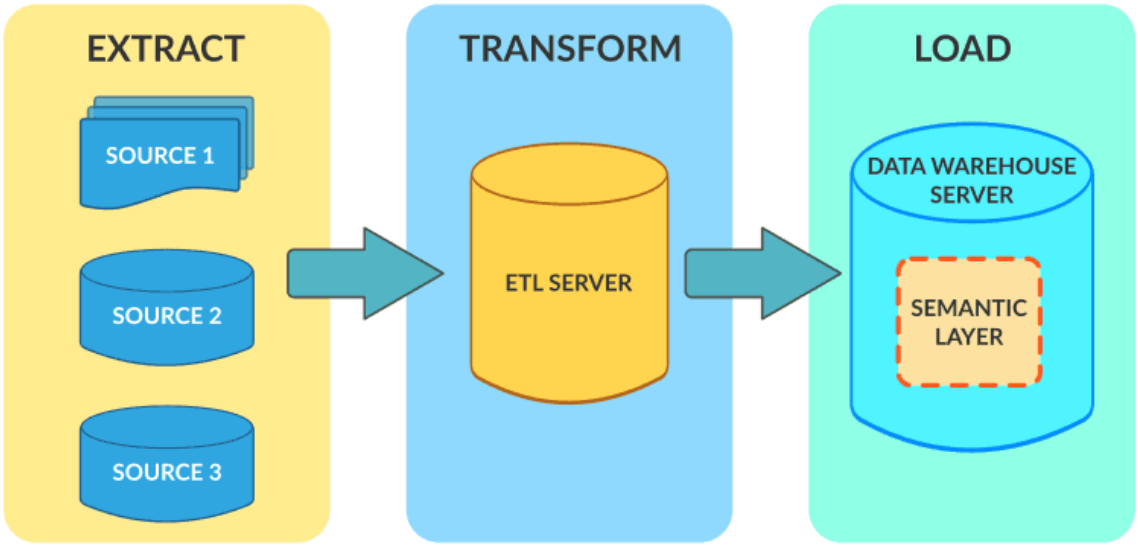



*Figure 1.5 ETL tool (thecustomizewindows.com)*

Thus, an ETL tool extracts data from disparate sources, transforms it to make itcompatible with the destination system, and then loads it into the destination system. This destination could be a data warehouse, data lake, database, or any other application system.

A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separatesources can be used together, and finally delivers data in a presentation-ready format so that application developers can build applications and end users can make decisions.

Since the data extraction takes time, it is common to execute the three phases in pipeline. While the data is being extracted, another transformation process executes while processing the data already received and prepares it for loading while the data loading begins without waiting for the completion of the previous phases.

#### 1.4.4.1. Extract:

In this step of ETL architecture, data is extracted from the source system intothe staging area. Transformations if any are done in staging area so that performance of source system in not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, rollback will be a challenge. Staging area gives an opportunity to validate extracted data before it moves into the Data warehouse.

Data warehouse needs to integrate systems that have different DBMS, Hardware, Operating Systems and Communication Protocols. Sources could include legacy applications like Mainframes, customized applications, point of contact devices like ATM, Call switches, text files, spreadsheets, ERP, data fromvendors, partners amongst others.

Hence one needs a logical data map before data is extracted and loaded physically. This data map describes the relationship between sources and targetdata. Three Data Extraction Methods:

1. Full Extraction.
2. Partial Extraction- without update notification.
3. Partial Extraction- with update notification.

Irrespective of the method used, extraction should not affect performance and response time of the source systems. These source systems are live production databases. Any slow down or locking could affect company's bottom line.

#### 1.4.4.2. Transform:

Data extracted from source server is raw and not usable in its original form. Therefore, it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful BI reports can be generated.

It is one of the important ETL concepts where you apply a set of functions on extracted data. Data that does not require any transformation is called as directmove or pass-through data.

In transformation step, you can perform customized operations on data. For instance, if the

user wants sum-of-sales revenue which is not in the database. Or if the first name and the last name in a table is in different columns. It is possible to concatenate them before loading.

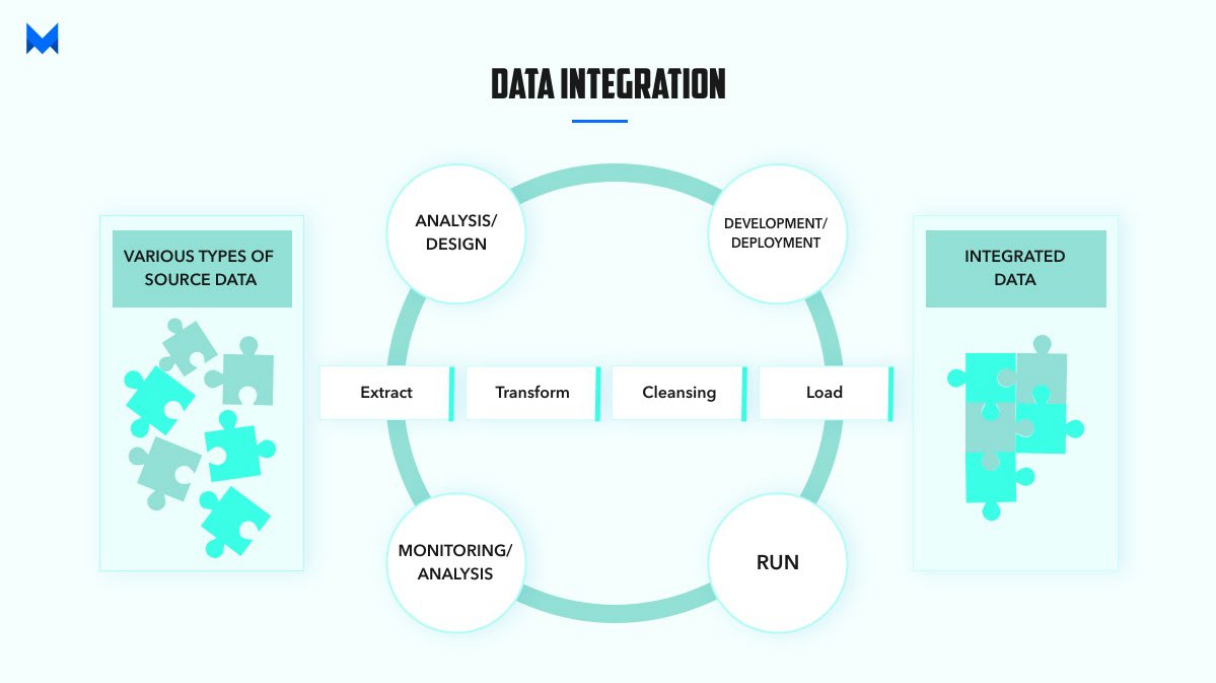



*Figure 1.6 Data Integration Cycle (www.mobcoder.com)*

Data integration is a general industry term referring to the requirement to combine data from multiple sources for a single unified view to process it further into insights. In separate words, it is a practice to consolidate data from multiple sources into a single dataset or unified view with the aim of providing users with consistent access across all data types.

**Here are the challenges and Data integration criteria:** There are three major data integration criteria to build a data warehouse are:

**Freshness**: The process of data integration architecture is executed at certain periods. In this, data freshness refers to the delay when data is tweaked on a source system to when the change is executed in the data warehouse.

**Granularity**: Data from the source system is channelized or aggregated in the very process of data integration. Here the data granularity refers to the level of detail with data stored in the data warehouse.

**Matching**: In the data integration process, data comes from disparate sources and needs to be specified in a specific category. Here matching defines how data from multiple sources should be matched can be complex.

### 1.4.4.3. Load:

Loading is the ultimate step in the ETL process. In this step, the extracted data and the transformed data are loaded into the target database. To make the data load efficient, it is necessary to index the database and disable the constraints before loading the data. All three steps in the ETL process can be run parallel. Data extraction takes time and therefore the second phase of the transformation process is executed simultaneously. This prepared the data for the third stage of loading. As soon as some data is ready, it is loaded without waiting for the previous steps to be completed.

The loading process is the physical movement of the data from the computer systems storing the source database(s) to that which will store the data warehouse database. The entire process of transferring data to a data warehouse repository is referred to in the following ways:

1- Initial Load: For the very first time loading all the data warehouse tables.

2- Incremental Load: Periodically applying ongoing changes as per the requirement. After the data is loaded into the data warehouse database, verify the referential integrity between the dimensions and the fact tables to ensure that all records belong to the appropriate records in the other tables. The DBA must verify that each record in the fact table is related to one record in each dimension table that will be used in combination with that fact table.

3- Full Refresh: Deleting the contents of a table and reloading it with fresh data.

**Refresh versus Update** After the initial load, the data warehouse needs to be maintained and updated and this can be done by the following two methods:

Update-application of incremental changes in the data sources.

Refresh-complete reloads at specified intervals.

**Data Loading:** Data is physically moved to the data warehouse. The loading takes place

within a “load window. The tendency is close to real-time updates for data warehouses as warehouses are growing used for operational applications.

**Loading the Dimension Tables**: Procedure for maintaining the dimension tables includes two functions, initial loading of the tables and thereafter applying the changes on an ongoing basis System geared keys are used in a data warehouse. The reeds in the source system have their own keys. Therefore, before an initial load or an ongoing load, the production keys must be co to system-generated keys in the data warehouse, another issue is related to the application of Type 1, Type 2, and Type 3 changes to the data warehouse. Fig. shows how to handle it.

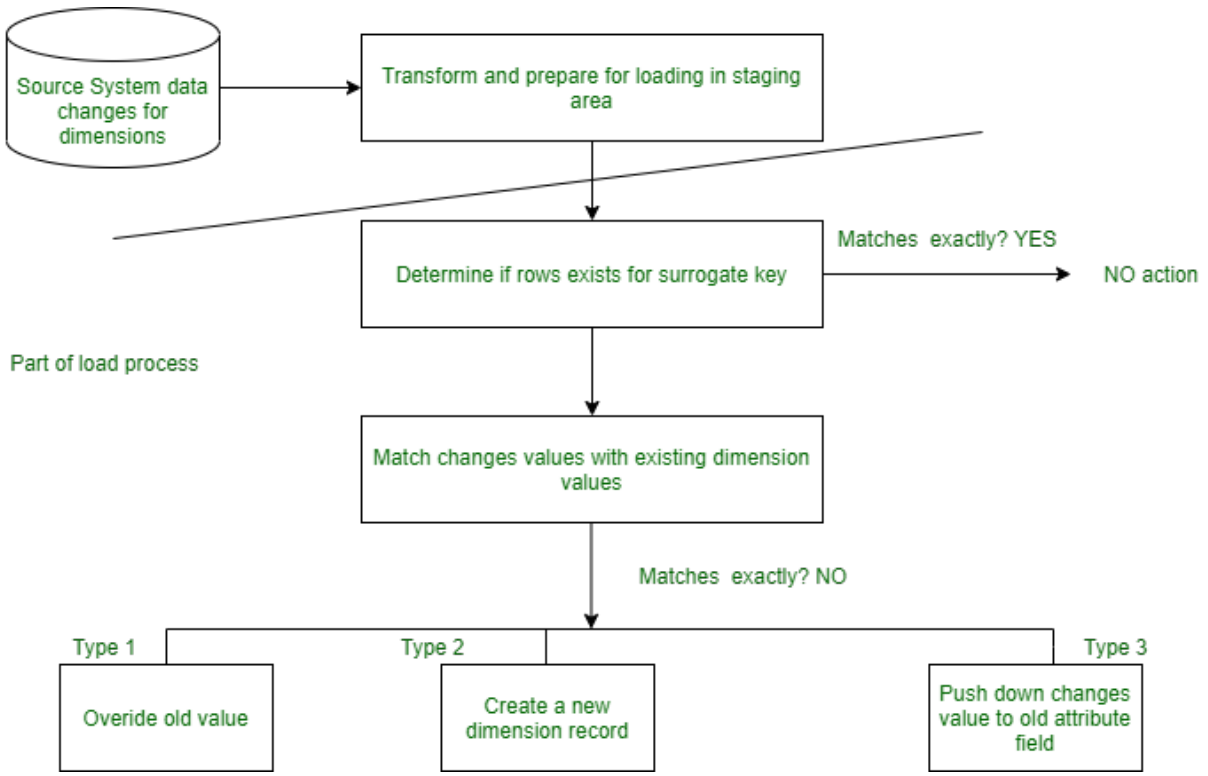



*Figure 1.7 Loading changes to a dimension table (www.geeksfogeeks.com)*

### 1.4.5. Data Mining:

#### 1.4.5.1. Introduction:

In general terms, “Mining” is the process of extraction of some valuable material from the earth e.g., coal mining, diamond mining, etc. In the context of computer science, “Data Mining” can be referred to as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. It is basically the process carried out for the extraction of useful information from a bulk of data or data warehouses. One can see that the term itself is a little confusing. In the case of coal or diamond mining, the result of the extraction process is coal or diamond. But in the case of Data Mining, the result of the extraction process is not data!! Instead, data mining results are the patterns and knowledge that we gain at the end of the extraction process. In that sense, we can think of Data Mining as a step in the process of Knowledge Discovery or Knowledge Extraction.

Gregory Piatetsky-Shapiro coined the term “Knowledge Discovery in Databases” in 1989. However, the term ‘data mining’ became more popular in the business and press communities. Currently, Data Mining and Knowledge Discovery are used interchangeably. Nowadays, data mining is used in almost all places where a large amount of data is stored and processed. For example, banks typically use ‘data mining’ to find out their prospective customers who could be interested in credit cards, personal loans, or insurance as well. Since banks have the transaction details and detailed profiles of their customers, they analyze all this data and try to find out patterns that help them predict that certain customers could be interested in personal loans, etc.

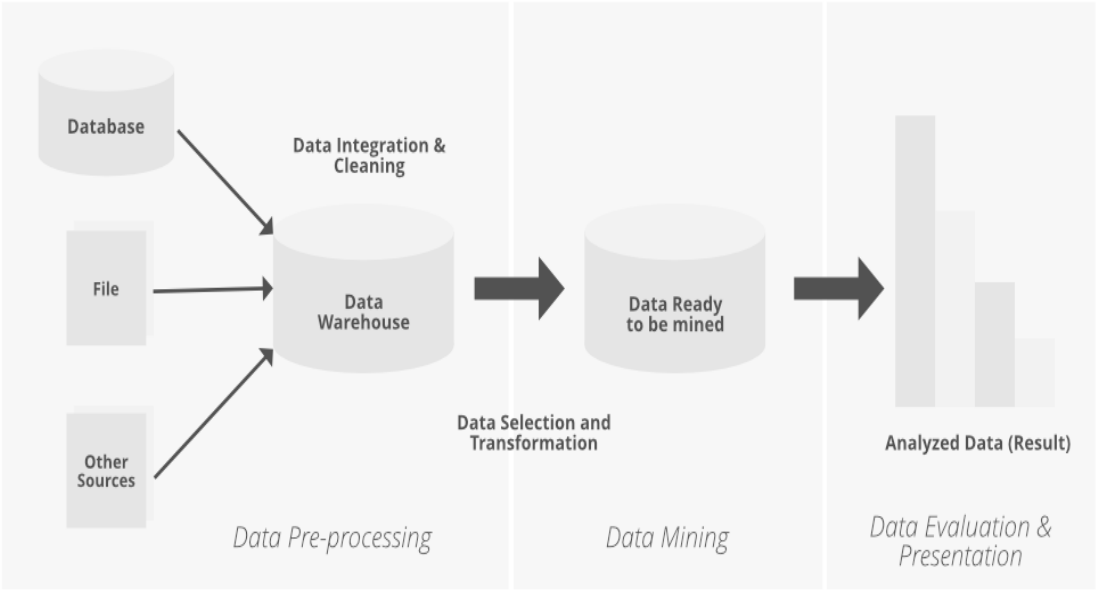



*Figure 1.8 Level of Analytics*

### 1.4.5.2. Methods and Techniques:

Data may be mined to help make more efficient decisions in the future. Or it may be used to explore the data to find interesting associative patterns. The right technique depends upon the kind of problem being solved. (Association, Classification, Predication, Clustering, Regression, …).

### a. Association:

Association analysis is the finding of association rules showing attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely used for a market basket or transaction data analysis. Association rule mining is a significant and exceptionally dynamic area of data mining research. One method of association-based classification, called associative classification, consists of two steps. In the main step, association instructions are generated using a modified version of the standard association rule mining algorithm known as Apriori. The second step constructs a classifier based on the association rules discovered.

### b. Classification:

Classification is the processing of finding a set of models (or functions) that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The determined model depends on the investigation of a set of training data information (i.e. data objects whose class label is known). The derived model may be represented in various forms, such as classification (if – then) rules, decision trees, and neural networks.

### c. Prediction:

Data Prediction is a two-step process, similar to that of data classification. Although, for prediction, we do not utilize the phrasing of "Class label attribute" because the attribute for which values are being predicted is consistently valued(ordered) instead of categorical (discrete-esteemed and unordered). The attribute can be referred to simply as the predicted attribute. Prediction can be viewed as the construction and use of a model to assess the class of an unlabeled object, or to assess the value or value ranges of an attribute that a given object is likely to have.

### d. Clustering:

Unlike classification and prediction, which analyze class-labeled data objects or attributes, clustering analyzes data objects without consulting an identified class label. In general, the class labels do not exist in the training data simply because they are not known to begin with. Clustering can be used to generate these labels. The objects are clustered based on the principle of maximizing the intra-class similarity and minimizing the interclass similarity. That is, clusters of objects are created so that objects inside a cluster have high similarity in contrast with each other, but are different objects in other clusters. Each Cluster that is generated can be seen as a class of objects, from which rules can be inferred. Clustering can also facilitate classification formation, that is, the organization of observations into a hierarchy of classes that group similar events together.

### e. Regression

Regression can be defined as a statistical modeling method in which previously obtained data is used to predicting a continuous quantity for new observations. This classifier is also known as the Continuous Value Classifier. There are two types of regression models: Linear regression and multiple linear regression models.

# Chapter 2
# Design and Analytical Study (Implementation)

## 2.1. Introduction:

Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information. BI systems leverage software and services beginning with collecting data, cleansing data, integration of data, and then using data visualization to transform data into actionable insights that might be helpful while decision making process. Implementing a business intelligence system requires careful planning to assure that it meets expectations. These are the basic steps:

1- Identify End-User Requirements

2- Identify the Data Sources (Data sources: which are reports in excel forms.)

3- ETL (Extract, Transform, and Load): in this area we managed to gather them, perform cleaning operations, and then visualize the data in a better form.

4- Design the Data Model

5- The Data Warehouse: Which contains all the relevant data, and can be described as an enterprise database.

6- The OLAP environment.

7- The BI application.

## 2.2. Identify End-User Requirements:

- ✓ What id the relation between number of violations and the quarters of year
- ✓ What is the relation between number of violations and License City?
- ✓ What is the relation between city of violation and city of license?
- ✓ How many numbers of violation paid after 1 month?
- ✓ How many numbers of violations paid in the same year?
- ✓ How many SP we get from a specific biller?
- ✓ What is the avg of discounted points?

## 2.3. Identify the Data sources:

We had the data directly from the Ministry of Interior provided by our supervisor Dr. Mouhib Alnoukari. It's an Excel file with 402540 records, 17 columns, and a size of **45.7MB**

| | No | Date | ViolationCity | ViolationType | IsPresented | TABI | CarType | CarCity | VAL | END_V | POIN | END_POIN | END_DATE | ECity | Biller |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 430166 | 6/5/2021 | Tartus | دفع المخالفة | 1 | 102430 | عامة | Tartus | 4,000 | 2,000 | 4 | 4 | 6/6/21 17:59 | Tartus | NULL |
| 3 | 430166 | 6/5/2021 | Tartus | دفع المخالفة | 1 | 102430 | عامة | Tartus | 7,000 | 3,500 | 6 | 6 | 6/6/21 17:59 | Tartus | NULL |
| 4 | 176791 | 8/25/2021 | Tartus | دفع المخالفة | 1 | 705531 | عامة | Tartus | 7,000 | 3,500 | 6 | 6 | 9/3/21 1:16 | Tartus | NULL |
| 5 | 176791 | 8/25/2021 | Tartus | دفع المخالفة | 1 | 705531 | عامة | Tartus | 500 | 250 | 0 | 0 | 9/3/21 1:16 | Tartus | NULL |
| 6 | 478416 | 9/21/2021 | Aleppo | دفع المخالفة | 1 | 715180 | خاصة | Tartus | 7,000 | 3,500 | 6 | 6 | 9/23/21 13:43 | Aleppo | NULL |
| 7 | 133253 | 9/27/2021 | Tartus | دفع المخالفة | 1 | 715644 | خاصة | Tartus | 2,000 | 1,000 | 0 | 0 | 9/28/21 16:42 | Tartus | NULL |
| 8 | 132259 | 9/15/2021 | Tartus | دفع المخالفة | 1 | 715825 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 9/21/21 14:50 | Tartus | NULL |
| 9 | 439758 | 6/19/2021 | Tartus | دفع المخالفة | 1 | 302822 | خاصة | Tartus | 2,000 | 1,000 | 0 | 0 | 6/22/21 13:41 | Tartus | NULL |
| 10 | 426190 | 8/29/2021 | Tartus | دفع المخالفة | 0 | 309614 | خاصة | Tartus | 5,000 | 2,500 | 4 | 4 | 9/26/21 12:34 | Homs | مرع مرور حمص |
| 11 | 426190 | 8/29/2021 | Tartus | دفع المخالفة | 0 | 309614 | خاصة | Tartus | 7,000 | 3,500 | 6 | 6 | 9/26/21 12:34 | Homs | مرع مرور حمص |
| 12 | 900902 | 8/18/2021 | Tartus | دفع المخالفة | 0 | 309614 | خاصة | Tartus | 5,000 | 2,500 | 4 | 4 | 9/26/21 12:34 | Homs | مرع مرور حمص |
| 13 | 900902 | 8/18/2021 | Tartus | دفع المخالفة | 0 | 309614 | خاصة | Tartus | 7,000 | 3,500 | 6 | 6 | 9/26/21 12:34 | Homs | مرع مرور حمص |
| 14 | 425893 | 6/3/2021 | Tartus | دفع المخالفة | 1 | 309614 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 6/7/21 14:01 | Tartus | NULL |
| 15 | 133075 | 9/20/2021 | Tartus | دفع المخالفة | 1 | 309614 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 9/29/21 15:08 | Tartus | NULL |
| 16 | 133075 | 9/20/2021 | Tartus | دفع المخالفة | 1 | 309614 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 9/29/21 15:08 | Tartus | NULL |
| 17 | 542881 | 5/4/2021 | Damascus | دفع المخالفة | 0 | 309194 | خاصة | Tartus | 7,000 | 3,500 | 6 | 6 | 8/11/21 10:37 | Latakia | مرع مرور اللاذقية |
| 18 | 623797 | 6/24/2021 | Hama | دفع المخالفة | 1 | 319006 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 6/30/21 13:21 | Hama | NULL |
| 19 | 885934 | 10/5/2021 | Homs | دفع المخالفة | 1 | 486067 | خاصة | Tartus | 4,000 | 2,000 | 4 | 4 | 10/10/21 14:14 | Homs | NULL |
| 20 | 436882 | 5/7/2021 | Tartus | دفع المخالفة | 1 | 327137 | خاصة | Tartus | 2,000 | 1,000 | 2 | 2 | 5/9/21 21:03 | Tartus | NULL |
| 21 | 437255 | 8/31/2021 | Tartus | دفع المخالفة | 1 | 308090 | خاصة | Tartus | 7,000 | 3,500 | 6 | 6 | 9/2/21 9:47 | Tartus | NULL |
| 22 | 436342 | 5/21/2021 | Tartus | دفع المخالفة | 1 | 308557 | خاصة | Tartus | 500 | 250 | 0 | 0 | 5/24/21 10:55 | Tartus | NULL |
| 23 | 378446 | 6/28/2021 | Latakia | [illegible] | 1 | 327248 | [illegible] | Tartus | 2,000 | 1,000 | 0 | 0 | 7/5/21 9:04 | Latakia | NULL |

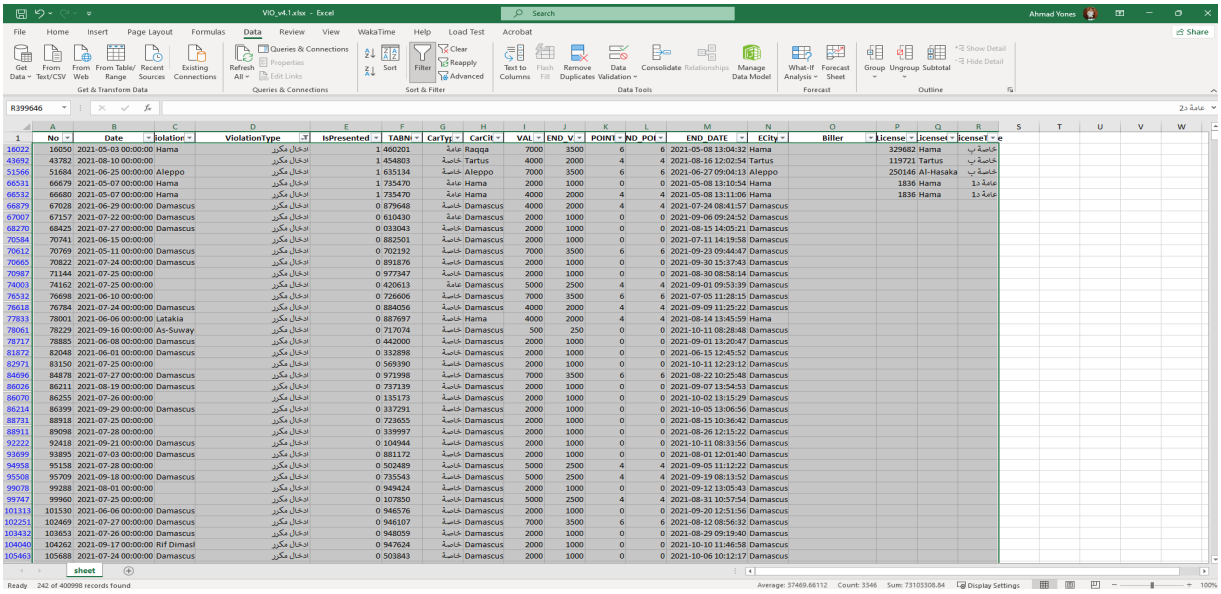
## 2.4. ETL:

Our biggest challenge was the data cleansing phase; we got into so many problems due to the low-quality data and the massive number of missing values, failed entries, and data ambiguity (We couldn't find someone to explain for us some attributes of the provided data).

We used Pandas and NumPy libraries for generating data quality reports, excel and power BI for dropping columns and filling null values.

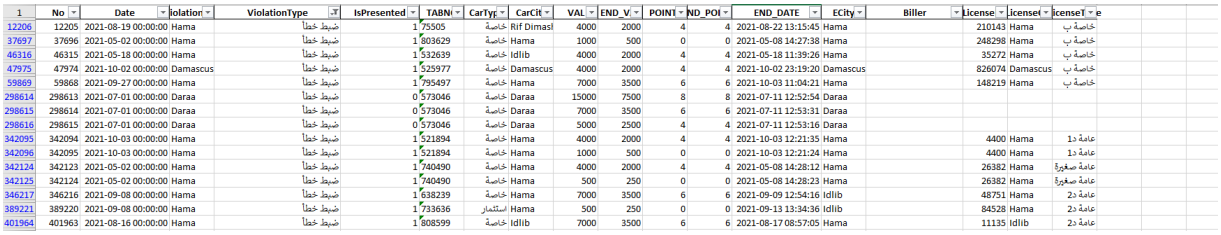
**Data cleansing phase:**

- Dropping records that consist of duplicated entries, wrong entries, conflicts, and ambiguous meanings. Duplicated Entries



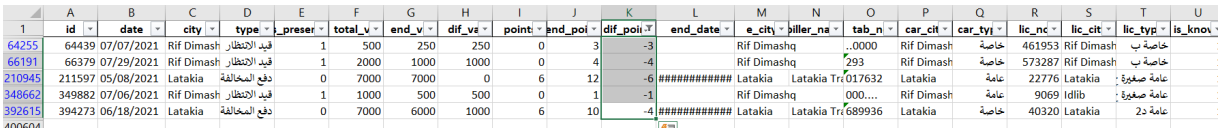
| | No | Date | olation | ViolationType | IsPresented | TABNI | CarTyp | CarCit | VAL | END_V | POINT | ND_POI | END_DATE | ECity | Biller | License | icenseC | icenseT | e |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 16022 | 16050 | 2021-05-03 00:00:00 | Hama | ادخال مكرر | 1 | 460201 | عامة | Raqqa | 7000 | 3500 | 6 | 6 | 2021-05-08 13:04:32 | Hama | | 329682 | Hama | خاصة ب | |
| 43692 | 43782 | 2021-08-10 00:00:00 | | ادخال مكرر | 1 | 454803 | خاصة | Tartus | 4000 | 2000 | 4 | 4 | 2021-08-16 12:02:54 | Tartus | | 119721 | Tartus | خاصة ب | |
| 51566 | 51684 | 2021-06-25 00:00:00 | Aleppo | ادخال مكرر | 1 | 635134 | خاصة | Aleppo | 7000 | 3500 | 6 | 6 | 2021-06-27 09:04:13 | Aleppo | | 250146 | Al-Hasaka | خاصة ب | |
| 66531 | 66679 | 2021-05-07 00:00:00 | Hama | ادخال مكرر | 1 | 735470 | عامة | Hama | 2000 | 1000 | 0 | 0 | 2021-05-08 13:10:54 | Hama | | 1836 | Hama | عامة د1 | |
| 66532 | 66680 | 2021-05-07 00:00:00 | Hama | ادخال مكرر | 1 | 735470 | عامة | Hama | 4000 | 2000 | 4 | 4 | 2021-05-08 13:11:06 | Hama | | 1836 | Hama | عامة د1 | |
| 66879 | 67028 | 2021-06-29 00:00:00 | Damascus | ادخال مكرر | 0 | 879648 | خاصة | Damascus | 4000 | 2000 | 4 | 4 | 2021-07-24 08:41:57 | Damascus | | | | | |
| 67007 | 67157 | 2021-07-22 00:00:00 | Damascus | ادخال مكرر | 0 | 610430 | عامة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-06 09:24:52 | Damascus | | | | | |
| 68270 | 68425 | 2021-07-27 00:00:00 | Damascus | ادخال مكرر | 0 | 033043 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-15 14:05:21 | Damascus | | | | | |
| 70584 | 70741 | 2021-06-15 00:00:00 | | ادخال مكرر | 0 | 882501 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-07-11 14:19:58 | Damascus | | | | | |
| 70612 | 70769 | 2021-05-11 00:00:00 | Damascus | ادخال مكرر | 0 | 702192 | خاصة | Damascus | 7000 | 3500 | 6 | 6 | 2021-09-23 09:44:47 | Damascus | | | | | |
| 70665 | 70822 | 2021-07-24 00:00:00 | Damascus | ادخال مكرر | 0 | 891876 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-30 15:37:43 | Damascus | | | | | |
| 70987 | 71144 | 2021-07-25 00:00:00 | | ادخال مكرر | 0 | 977347 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-30 08:58:14 | Damascus | | | | | |
| 74003 | 74162 | 2021-07-25 00:00:00 | | ادخال مكرر | 0 | 420613 | عامة | Damascus | 5000 | 2500 | 4 | 4 | 2021-09-01 09:53:39 | Damascus | | | | | |
| 76532 | 76698 | 2021-06-10 00:00:00 | | ادخال مكرر | 0 | 726606 | خاصة | Damascus | 7000 | 3500 | 6 | 6 | 2021-07-05 11:28:15 | Damascus | | | | | |
| 76618 | 76784 | 2021-07-24 00:00:00 | Damascus | ادخال مكرر | 0 | 884096 | خاصة | Damascus | 4000 | 2000 | 4 | 4 | 2021-09-09 11:25:22 | Damascus | | | | | |
| 77833 | 78001 | 2021-06-06 00:00:00 | Latakia | ادخال مكرر | 0 | 887697 | خاصة | Hama | 4000 | 2000 | 4 | 4 | 2021-08-14 13:45:59 | Hama | | | | | |
| 78061 | 78229 | 2021-09-16 00:00:00 | As-Suway | ادخال مكرر | 0 | 717074 | خاصة | Damascus | 500 | 250 | 0 | 0 | 2021-10-11 08:28:48 | Damascus | | | | | |
| 78717 | 78885 | 2021-06-08 00:00:00 | Damascus | ادخال مكرر | 0 | 442000 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-01 13:20:47 | Damascus | | | | | |
| 81872 | 82048 | 2021-06-01 00:00:00 | Damascus | ادخال مكرر | 0 | 332898 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-06-15 12:45:52 | Damascus | | | | | |
| 82971 | 83150 | 2021-07-25 00:00:00 | | ادخال مكرر | 0 | 569390 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-11 12:23:12 | Damascus | | | | | |
| 84696 | 84878 | 2021-07-27 00:00:00 | Damascus | ادخال مكرر | 0 | 971998 | خاصة | Damascus | 7000 | 3500 | 6 | 6 | 2021-08-22 10:25:48 | Damascus | | | | | |
| 86026 | 86211 | 2021-08-19 00:00:00 | Damascus | ادخال مكرر | 0 | 737139 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-07 13:54:53 | Damascus | | | | | |
| 86070 | 86255 | 2021-07-26 00:00:00 | | ادخال مكرر | 0 | 135173 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-02 13:15:29 | Damascus | | | | | |
| 86214 | 86399 | 2021-09-29 00:00:00 | Damascus | ادخال مكرر | 0 | 337291 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-05 13:06:56 | Damascus | | | | | |
| 88731 | 88918 | 2021-07-25 00:00:00 | | ادخال مكرر | 0 | 723655 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-15 10:36:42 | Damascus | | | | | |
| 88911 | 89098 | 2021-07-28 00:00:00 | | ادخال مكرر | 0 | 339997 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-26 12:15:22 | Damascus | | | | | |
| 92222 | 92418 | 2021-09-21 00:00:00 | Damascus | ادخال مكرر | 0 | 104944 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-11 08:33:56 | Damascus | | | | | |
| 93699 | 93895 | 2021-07-03 00:00:00 | Damascus | ادخال مكرر | 0 | 881172 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-01 12:01:40 | Damascus | | | | | |
| 94958 | 95158 | 2021-07-28 00:00:00 | | ادخال مكرر | 0 | 502489 | خاصة | Damascus | 5000 | 2500 | 4 | 4 | 2021-09-05 11:12:22 | Damascus | | | | | |
| 95508 | 95709 | 2021-09-18 00:00:00 | Damascus | ادخال مكرر | 0 | 735543 | خاصة | Damascus | 5000 | 2500 | 4 | 4 | 2021-09-19 08:13:52 | Damascus | | | | | |
| 99078 | 99288 | 2021-08-01 00:00:00 | | ادخال مكرر | 0 | 949424 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-12 13:05:43 | Damascus | | | | | |
| 99747 | 99960 | 2021-07-25 00:00:00 | | ادخال مكرر | 0 | 107850 | خاصة | Damascus | 5000 | 2500 | 4 | 4 | 2021-08-31 10:57:54 | Damascus | | | | | |
| 101313 | 101530 | 2021-06-06 00:00:00 | Damascus | ادخال مكرر | 0 | 946576 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-09-20 12:51:56 | Damascus | | | | | |
| 102251 | 102469 | 2021-07-27 00:00:00 | Damascus | ادخال مكرر | 0 | 946107 | خاصة | Damascus | 7000 | 3500 | 6 | 6 | 2021-08-12 08:56:32 | Damascus | | | | | |
| 103432 | 103653 | 2021-07-26 00:00:00 | Damascus | ادخال مكرر | 0 | 948059 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-08-29 09:19:40 | Damascus | | | | | |
| 104040 | 104262 | 2021-09-17 00:00:00 | Rif Dimasl | ادخال مكرر | 0 | 947624 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-10 11:46:58 | Damascus | | | | | |
| 105463 | 105688 | 2021-07-24 00:00:00 | Damascus | ادخال مكرر | 0 | 503843 | خاصة | Damascus | 2000 | 1000 | 0 | 0 | 2021-10-06 10:12:17 | Damascus | | | | | |

- Wrong Entities



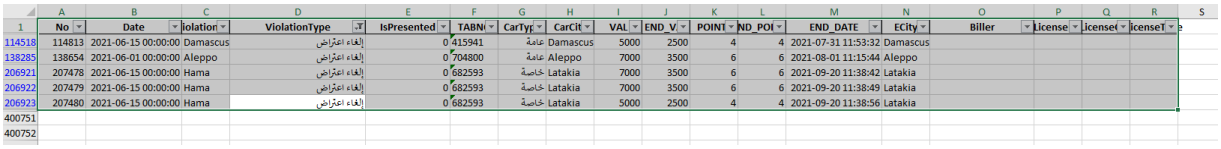
| | No | Date | olation | ViolationType | IsPresented | TABNI | CarTyp | CarCit | VAL | END_V | POINT | ND_POI | END_DATE | ECity | Biller | License | icenseC | icenseT | e |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12206 | 12205 | 2021-08-19 00:00:00 | Hama | ضبط خطأ | 1 | 75505 | خاصة | Rif Dimasl | 4000 | 2000 | 4 | 4 | 2021-08-22 13:15:45 | Hama | | 210143 | Hama | خاصة ب | |
| 37697 | 37696 | 2021-05-02 00:00:00 | Hama | ضبط خطأ | 1 | 803629 | خاصة | Hama | 1000 | 500 | 0 | 0 | 2021-05-08 14:27:38 | Hama | | 248298 | Hama | خاصة ب | |
| 46316 | 46315 | 2021-05-18 00:00:00 | Hama | ضبط خطأ | 1 | 532639 | خاصة | Idlib | 4000 | 2000 | 4 | 4 | 2021-05-18 11:39:26 | Hama | | 35272 | Hama | خاصة ب | |
| 47975 | 47974 | 2021-10-02 00:00:00 | Damascus | ضبط خطأ | 1 | 525977 | خاصة | Damascus | 4000 | 2000 | 4 | 4 | 2021-10-02 23:19:20 | Damascus | | 826074 | Damascus | خاصة ب | |
| 59869 | 59868 | 2021-09-27 00:00:00 | Hama | ضبط خطأ | 1 | 795497 | خاصة | Hama | 7000 | 3500 | 6 | 6 | 2021-10-03 11:04:21 | Hama | | 148219 | Hama | خاصة ب | |
| 298614 | 298613 | 2021-07-01 00:00:00 | Daraa | ضبط خطأ | 0 | 573046 | خاصة | Daraa | 15000 | 7500 | 8 | 8 | 2021-07-11 12:52:54 | Daraa | | | | | |
| 298615 | 298614 | 2021-07-01 00:00:00 | Daraa | ضبط خطأ | 0 | 573046 | خاصة | Daraa | 7000 | 3500 | 6 | 6 | 2021-07-11 12:53:31 | Daraa | | | | | |
| 298616 | 298615 | 2021-07-01 00:00:00 | Daraa | ضبط خطأ | 0 | 573046 | خاصة | Daraa | 5000 | 2500 | 4 | 4 | 2021-07-11 12:53:16 | Daraa | | | | | |
| 342095 | 342094 | 2021-10-03 00:00:00 | Hama | ضبط خطأ | 1 | 521894 | خاصة | Hama | 4000 | 2000 | 4 | 4 | 2021-10-03 12:21:35 | Hama | | 4400 | Hama | عامة د1 | |
| 342096 | 342095 | 2021-10-03 00:00:00 | Hama | ضبط خطأ | 1 | 521894 | خاصة | Hama | 1000 | 500 | 0 | 0 | 2021-10-03 12:21:24 | Hama | | 4400 | Hama | عامة د1 | |
| 342124 | 342123 | 2021-05-02 00:00:00 | Hama | ضبط خطأ | 1 | 740490 | خاصة | Hama | 4000 | 2000 | 4 | 4 | 2021-05-08 14:28:12 | Hama | | 26382 | Hama | عامة صغيرة | |
| 342125 | 342124 | 2021-05-02 00:00:00 | Hama | ضبط خطأ | 1 | 740490 | خاصة | Hama | 500 | 250 | 0 | 0 | 2021-05-08 14:28:23 | Hama | | 26382 | Hama | عامة صغيرة | |
| 346217 | 346216 | 2021-09-08 00:00:00 | Hama | ضبط خطأ | 1 | 638239 | خاصة | Hama | 7000 | 3500 | 6 | 6 | 2021-09-09 12:54:16 | Idlib | | 48751 | Hama | عامة د2 | |
| 389221 | 389220 | 2021-09-08 00:00:00 | Hama | ضبط خطأ | 1 | 733636 | استثمار | Hama | 500 | 250 | 0 | 0 | 2021-09-13 13:34:36 | Idlib | | 84528 | Hama | عامة د2 | |
| 401964 | 401963 | 2021-08-16 00:00:00 | Hama | ضبط خطأ | 1 | 808599 | خاصة | Idlib | 7000 | 3500 | 6 | 6 | 2021-08-17 08:57:05 | Hama | | 11135 | Idlib | عامة د2 | |

- Conflicts



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | id | date | city | type | _preser | total_v | end_v | dif_va | point | end_poi | dif_poi | end_date | e_city | biller_na | tab_n | car_cit | car_ty | lic_nc | lic_cit | lic_typ | is_knov |
| 64255 | 64439 | 07/07/2021 | Rif Dimash | قيد الانتظار | 1 | 500 | 250 | 250 | 0 | 3 | -3 | | Rif Dimashq | | ..0000 | Rif Dimash | خاصة | 461953 | Rif Dimash | خاصة ب | |
| 66191 | 66379 | 07/29/2021 | Rif Dimash | قيد الانتظار | 1 | 2000 | 1000 | 1000 | 0 | 4 | -4 | | Rif Dimashq | | 293 | Rif Dimash | خاصة | 573287 | Rif Dimash | خاصة ب | |
| 210945 | 211597 | 05/08/2021 | Latakia | دفع المخالفة | 0 | 7000 | 7000 | 0 | 6 | 12 | -6 | ########## | Latakia | Latakia Tr | 017632 | Latakia | عامة | 22776 | Latakia | عامة صغيرة | |
| 348662 | 349882 | 07/06/2021 | Rif Dimash | قيد الانتظار | 1 | 1000 | 500 | 500 | 0 | 1 | -1 | | Rif Dimashq | | 000.... | Rif Dimash | عامة | 9069 | Idlib | عامة صغيرة | |
| 392615 | 394273 | 06/18/2021 | Latakia | دفع المخالفة | 0 | 7000 | 6000 | 1000 | 6 | 10 | -4 | ########## | Latakia | Latakia Tr | 689936 | Latakia | خاصة | 40320 | Latakia | عامة د2 | |
| 400604 | | | | | | | | | | | | | | | | | | | | | |

- Ambiguous Meaning



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | No | Date | iolation | ViolationType | IsPresented | TABN | CarTyp | CarCit | VAL | END_V | POINT | ND_POI | END_DATE | ECity | Biller | License | license( | icenseT | e |
| 114518 | 114813 | 2021-06-15 00:00:00 | Damascus | إلغاء اعتراض | 0 | 415941 | عامة | Damascus | 5000 | 2500 | 4 | 4 | 2021-07-31 11:53:32 | Damascus | | | | | |
| 138285 | 138654 | 2021-06-01 00:00:00 | Aleppo | إلغاء اعتراض | 0 | 704800 | عامة | Aleppo | 7000 | 3500 | 6 | 6 | 2021-08-01 11:15:44 | Aleppo | | | | | |
| 206921 | 207478 | 2021-06-15 00:00:00 | Hama | إلغاء اعتراض | 0 | 682593 | خاصة | Latakia | 7000 | 3500 | 6 | 6 | 2021-09-20 11:38:42 | Latakia | | | | | |
| 206922 | 207479 | 2021-06-15 00:00:00 | Hama | إلغاء اعتراض | 0 | 682593 | خاصة | Latakia | 7000 | 3500 | 6 | 6 | 2021-09-20 11:38:49 | Latakia | | | | | |
| 206923 | 207480 | 2021-06-15 00:00:00 | Hama | إلغاء اعتراض | 0 | 682593 | خاصة | Latakia | 5000 | 2500 | 4 | 4 | 2021-09-20 11:38:56 | Latakia | | | | | |
| 400751 | | | | | | | | | | | | | | | | | | | |
| 400752 | | | | | | | | | | | | | | | | | | | |

We added a new column to map between null values that we found in primary keys in LicenseNO column and the new column (new primary key). (Every null value takes a specific value).

We needed to loop over 150000 x 85000 elements, takes O(n2) and it's an impossible solution so we applied the dictionary data structure property. (Pandas library takes the Data Frame and covert it to a dictionary) in order to use the hashing ability that dictionary provides so we can assign a new primary key to the License Dimension distinct values in **O(n)**.

```
newDf["LicenseNO"] = newDf["LicenseNO"].astype("str")
newDf.info()
```

```
count = 1
for ind in newDf.index:
  newDf.iloc[ind, 1] = count
  count += 1
```

```
sliced_text = slice(0, -2)
for ind in newDf.index:
  newDf.iloc[ind, 0] = newDf.iloc[ind, 0][sliced_text]
```

```
dic =  newDf.set_index('LicenseNO').to_dict()
```

```
dic
```

- Table Modification

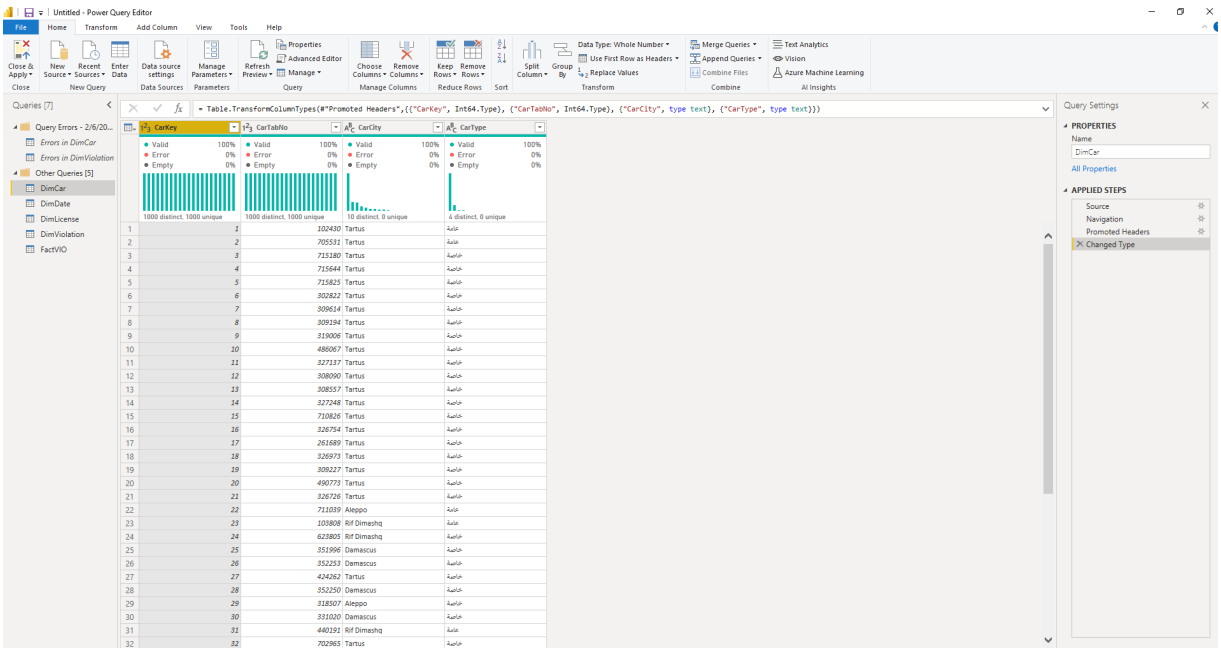
Change Data Types:

| Pandas dtype | Python type | NumPy type | Usage |
|---|---|---|---|
| object | str or mixed | string_, unicode_, mixed types | Text or mixed numeric and non-numeric values |
| int64 | int | int_, int8, int16, int32, int64, uint8, uint16, uint32, uint64 | Integer numbers |
| float64 | float | float_, float16, float32, float64 | Floating point numbers |
| bool | bool | bool_ | True/False values |
| datetime64 | NA | datetime64[ns] | Date and time values |
| timedelta[ns] | NA | NA | Differences between two datetimes |
| category | NA | NA | Finite list of text values |

```python
df["id"] = df["id"].astype("str")
df["date_key"] = df["date_key"].astype("str")
df["city"] = df["city"].astype("category")
df["type"] = df["type"].astype("category")
df["is_presented"] = df["is_presented"].astype("bool")
df["total_val"] = df["total_val"].astype("int64")
df["end_val"] = df["end_val"].astype("int64")
df["points"] = df["points"].astype("int64")
df["end_points"] = df["end_points"].astype("int64")
df["e_city"] = df["e_city"].astype("category")
df["biller_name"] = df["biller_name"].astype("category")
df["tab_no"] = df["tab_no"].astype("str")
df["car_city"] = df["car_city"].astype("category")
df["car_type"] = df["car_type"].astype("category")
df["lic_id"] = df["lic_id"].astype("str")
df["lic_city"] = df["lic_city"].astype("category")
df["lic_type"] = df["lic_type"].astype("category")

df.info()
```

- Data integration while loading to PowrBI

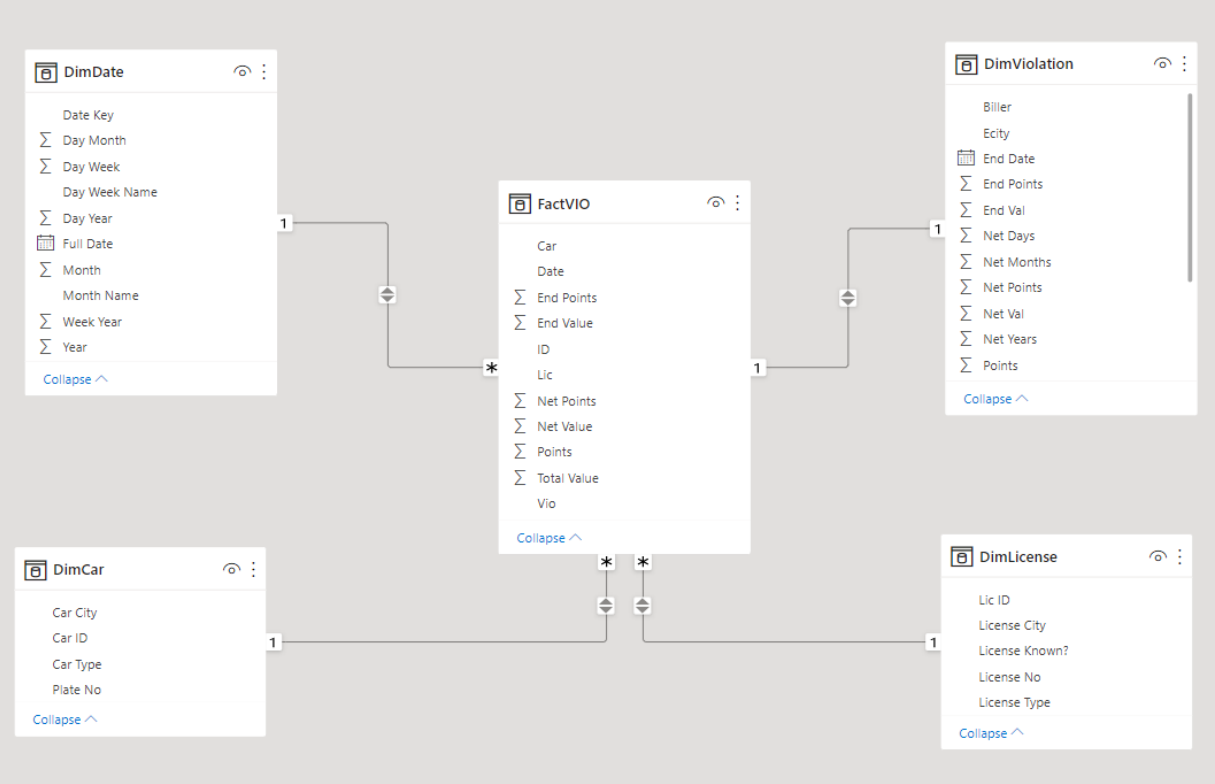


| | Car ID | Plate No | Car City | Car Type |
|---|---|---|---|---|
| | Valid 100% | Valid 100% | Valid 100% | Valid 100% |
| | Error 0% | Error 0% | Error 0% | Error 0% |
| | Empty 0% | Empty 0% | Empty 0% | Empty 0% |
| 1 | 1 | 102430 | Tartus | عامة |
| 2 | 2 | 705531 | Tartus | عامة |
| 3 | 3 | 715180 | Tartus | خاصة |
| 4 | 4 | 715644 | Tartus | خاصة |
| 5 | 5 | 715825 | Tartus | خاصة |
| 6 | 6 | 302822 | Tartus | خاصة |
| 7 | 7 | 309614 | Tartus | خاصة |
| 8 | 8 | 309194 | Tartus | خاصة |
| 9 | 9 | 319006 | Tartus | خاصة |
| 10 | 10 | 486067 | Tartus | خاصة |
| 11 | 11 | 327137 | Tartus | خاصة |
| 12 | 12 | 308090 | Tartus | خاصة |
| 13 | 13 | 308557 | Tartus | خاصة |
| 14 | 14 | 327248 | Tartus | خاصة |

| | License ID | License No | License City | License Type | License Is Known |
|---|---|---|---|---|---|
| | Valid 100% | Valid 100% | Valid 100% | Valid 100% | Valid 100% |
| | Error 0% | Error 0% | Error 0% | Error 0% | Error 0% |
| | Empty 0% | Empty 0% | Empty 0% | Empty 0% | Empty 0% |
| 1 | 1 | 247717 | Hama | خاصة ب | TRUE |
| 2 | 2 | 236555 | Tartus | خاصة ب | TRUE |
| 3 | 3 | 165958 | Aleppo | خاصة ب | TRUE |
| 4 | 4 | 83358 | Tartus | خاصة ب | TRUE |
| 5 | 5 | 208524 | Hama | خاصة ب | TRUE |
| 6 | 6 | 147262 | Tartus | خاصة ب | TRUE |
| 7 | 7 | 69648 | Homs | خاصة ب | TRUE |
| 8 | 8 | 192626 | Tartus | خاصة ب | TRUE |
| 9 | 9 | 54403 | Latakia | خاصة ب | TRUE |
| 10 | 10 | 64284 | Hama | خاصة ب | TRUE |
| 11 | 11 | 238550 | Tartus | خاصة ب | TRUE |
| 12 | 12 | 233607 | Tartus | خاصة ب | TRUE |
| 13 | 13 | 248671 | Tartus | خاصة ب | TRUE |
| 14 | 14 | 5005 | Tartus | خاصة ب | TRUE |
| 15 | 15 | 225290 | Latakia | خاصة ب | TRUE |
| 16 | 16 | 243241 | Tartus | خاصة ب | TRUE |
| 17 | 17 | 229600 | Latakia | خاصة ب | TRUE |
| 18 | 18 | 243087 | Tartus | خاصة ب | TRUE |
| 19 | 19 | 313623 | Aleppo | خاصة ب | TRUE |
| 20 | 20 | 217378 | Tartus | خاصة ب | TRUE |
| 21 | 21 | 117598 | Latakia | خاصة ب | TRUE |
| 22 | 22 | 250179 | Latakia | خاصة ب | TRUE |
| 23 | 23 | 616755 | Aleppo | خاصة ب | TRUE |
| 24 | 24 | 213541 | Tartus | خاصة ب | TRUE |
| 25 | 25 | 431495 | Rif Dimashq | خاصة ب | TRUE |
| 26 | 26 | 382127 | Rif Dimashq | خاصة ب | TRUE |

## 2.5. Design the Data Model:

The Data warehouse was built using SQL Server 2019 and we used the star schema to model the DW. Star Schema in data warehouse, in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star. The Star Schema data model is the simplest type of Data Warehouse schema. It is also known as Star Join Schema and is optimized for querying large data sets. The design one fact table, which is connected further with multiple dimension tables. The fact table contains all the facts, while the dimension table contains the objects.




### ➢ Why Star Schema?

A star schema is a database organizational structure optimized for use in a data warehouse or business intelligence that uses a single large fact table to store transactional or measured data, and one or more smaller dimensional tables that store attributes about the data. It is called a star schema because the fact table sits at the center of the logical diagram, and the small dimensional

tables branch off to form the points of the star.

A fact table sits at the center of a star schema database, and each star schema database only has a single fact table. The fact table contains the specific measurable (or quantifiable) primary data to be analyzed, such as sales records, logged performance data or financial data. It may be transactional -- in that, rows are added as events happen -- or it may be a snapshot of historical data up to a point in time.

- **How Star Schema Works?**

The fact table stores two types of information: numeric values and dimension attribute values. Using a sales database as an example:

**Numeric value cells** are unique to each row or data point and do not correlate or relate to data stored in other rows. These might be facts about a transaction, such as an order ID, total amount, net profit, order quantity or exact time.

**The dimension attribute values** do not directly store data, but they store the foreign key value for a row in a related dimensional table. Many rows in the fact table will reference this type of information. So, for example, it might store the sales employee ID, a date value, a product ID or a branch office ID.

**Dimension tables** store supporting information to the fact table. Each star schema database has at least one dimension table, but will often have many. Each dimension table will relate to a column in the fact table with a dimension value, and will store additional information about that value.

Constructing a start schema should be carefully done. Each table should have either fact data or dimension data, and avoid mixing the two. Consider the total number of dimension tables to maximize performance. Also, consider the granularity of the data captured to optimize for the types of queries that will be run.

Optimized for querying large data sets, data warehouses and data marts, star schemas support online analytical processing (OLAP) cubes, analytic application, ad hoc queries and business intelligence (BI). They also support count, sum, average and other rapid aggregations

of many fact records. Users can filter and group (sliced and diced) these aggregations by dimensions.

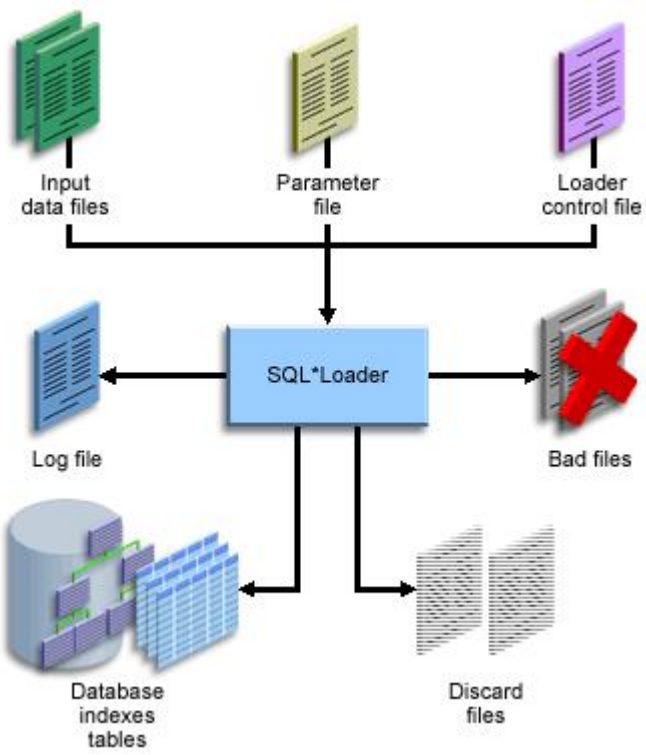



*Figure 2.1 SQL loader environment*

After the data was loaded into the data warehouse, we needed to connect the SQL server DW to the power BI environment to build the dashboards.

## 2.6. Visualization:

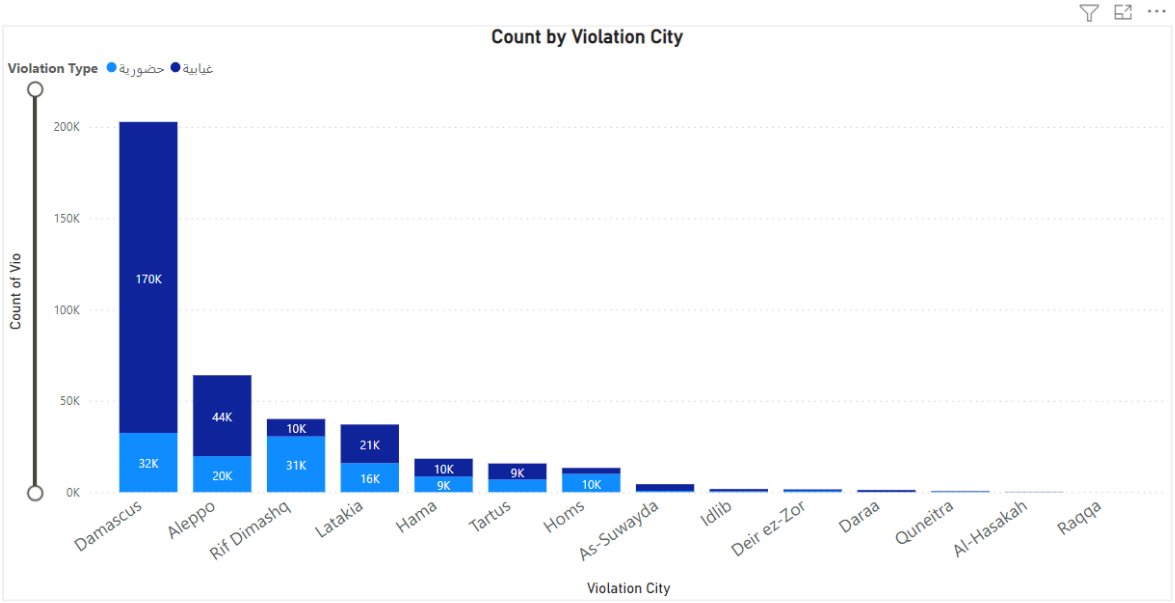
Data Visualization is a process of taking raw data and transforming it into graphical or pictorial representations such as charts, graphs, diagrams, pictures, and videos which explain the data and allow you to gain insights from it. So, users can quickly analyze the data and prepare reports to make business decisions effectively.

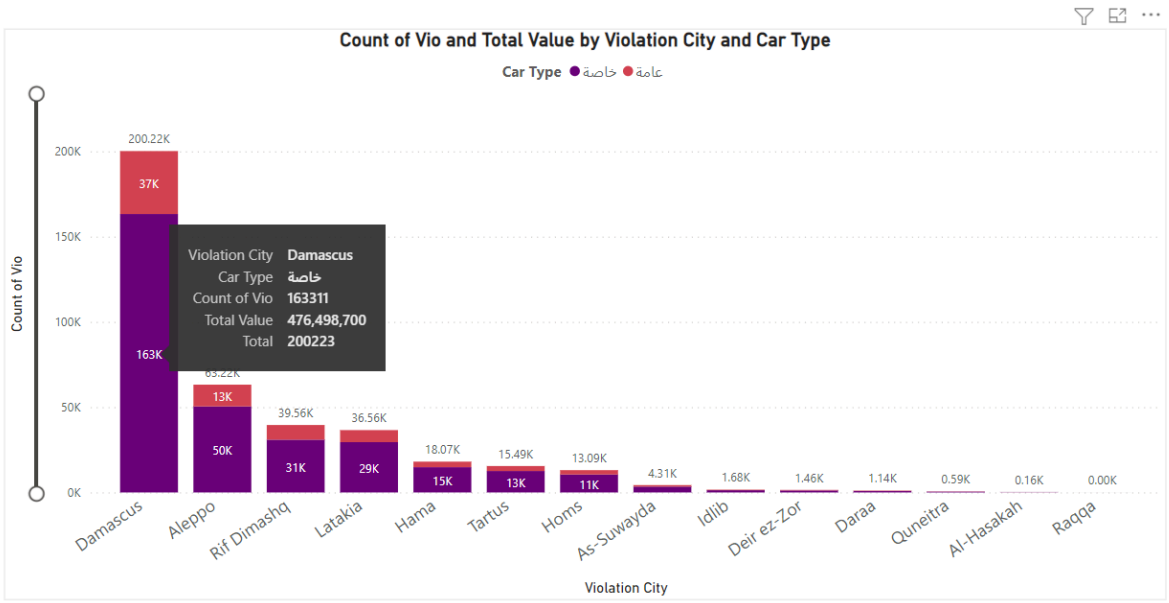
- **What is the importance of Data Visualization?**

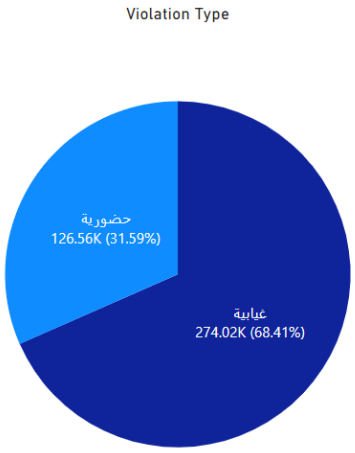
We are inherently in the visual world where pictures or images speak more than words. So, it is easy to visualize a large amount of data using graphs and charts than depending on reports or spreadsheets.

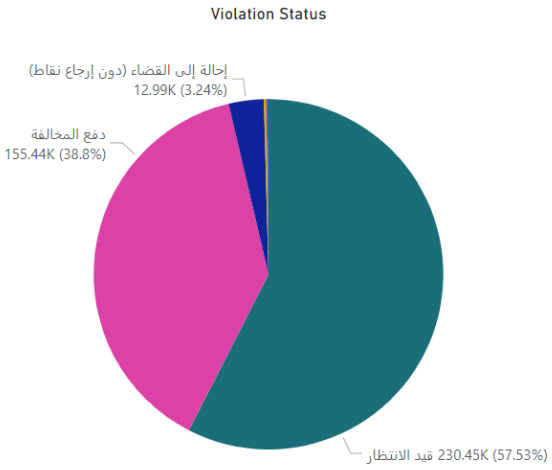
Data visualization is a quick and easy way to convey concepts to the end-users, and you can do experiments with different scenarios by making slight changes.

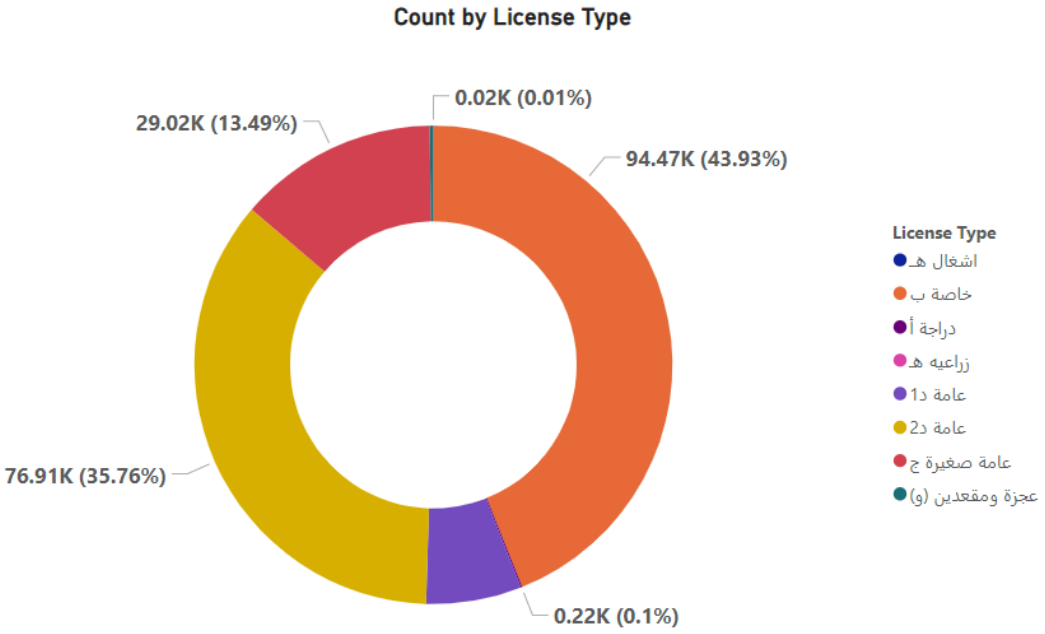
- **Samples of Our Analysis and Visualizing:**

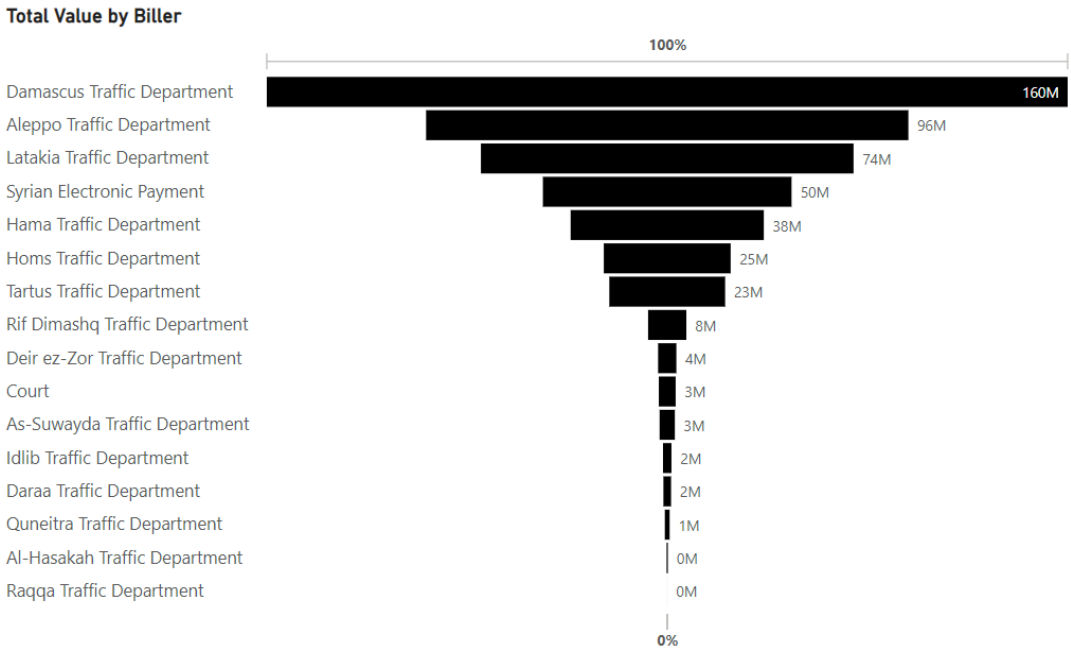

Count by Violation City
Violation Type حضورية غيابية
Count of Vio
200K
150K
100K
50K
0K
170K
32K
44K
20K
10K
31K
21K
16K
10K
9K
9K
10K
Damascus
Aleppo
Rif Dimashq
Latakia
Hama
Tartus
Homs
As-Suwayda
Idlib
Deir ez-Zor
Daraa
Quneitra
Al-Hasakah
Raqqa
Violation City

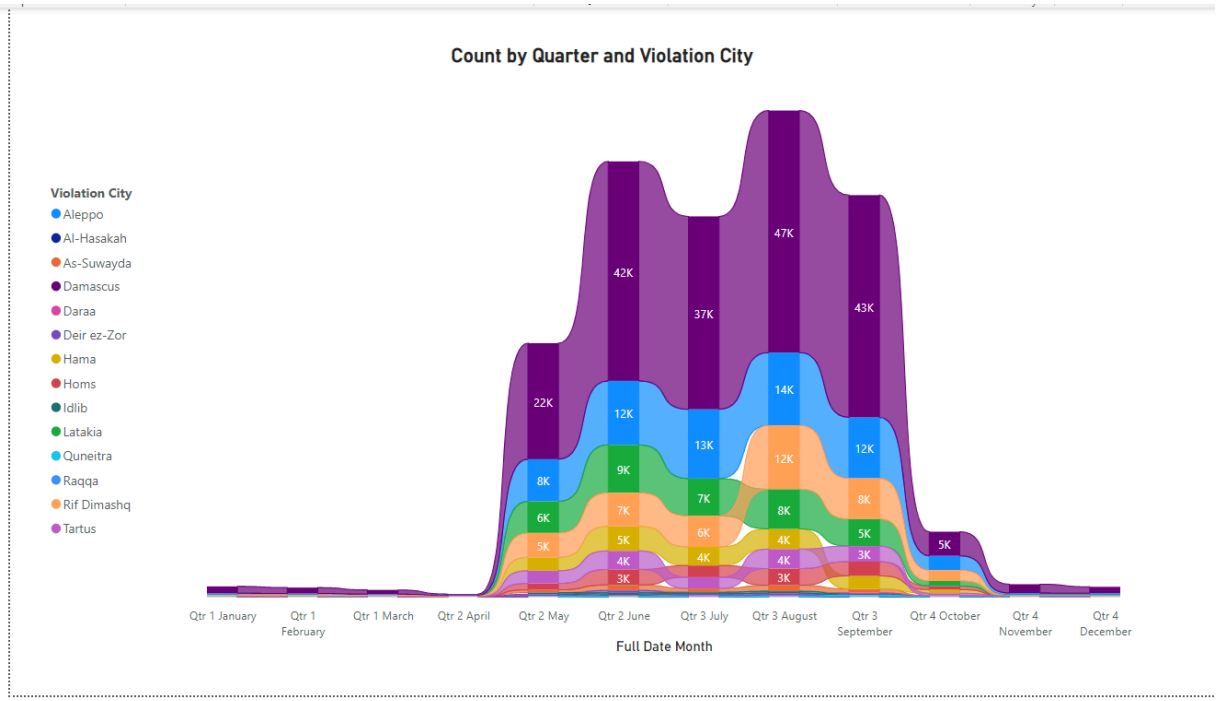

Count of Vio and Total Value by Violation City and Car Type
Car Type خاصة عامة
Count of Vio
200K
150K
100K
50K
0K
200.22K
37K
163K
Violation City Damascus
Car Type خاصة
Count of Vio 163311
Total Value 476,498,700
Total 200223
63.22K
13K
50K
39.56K
31K
36.56K
29K
18.07K
15K
15.49K
13K
13.09K
11K
4.31K
1.68K
1.46K
1.14K
0.59K
0.16K
0.00K
Damascus
Aleppo
Rif Dimashq
Latakia
Hama
Tartus
Homs
As-Suwayda
Idlib
Deir ez-Zor
Daraa
Quneitra
Al-Hasakah
Raqqa
Violation City

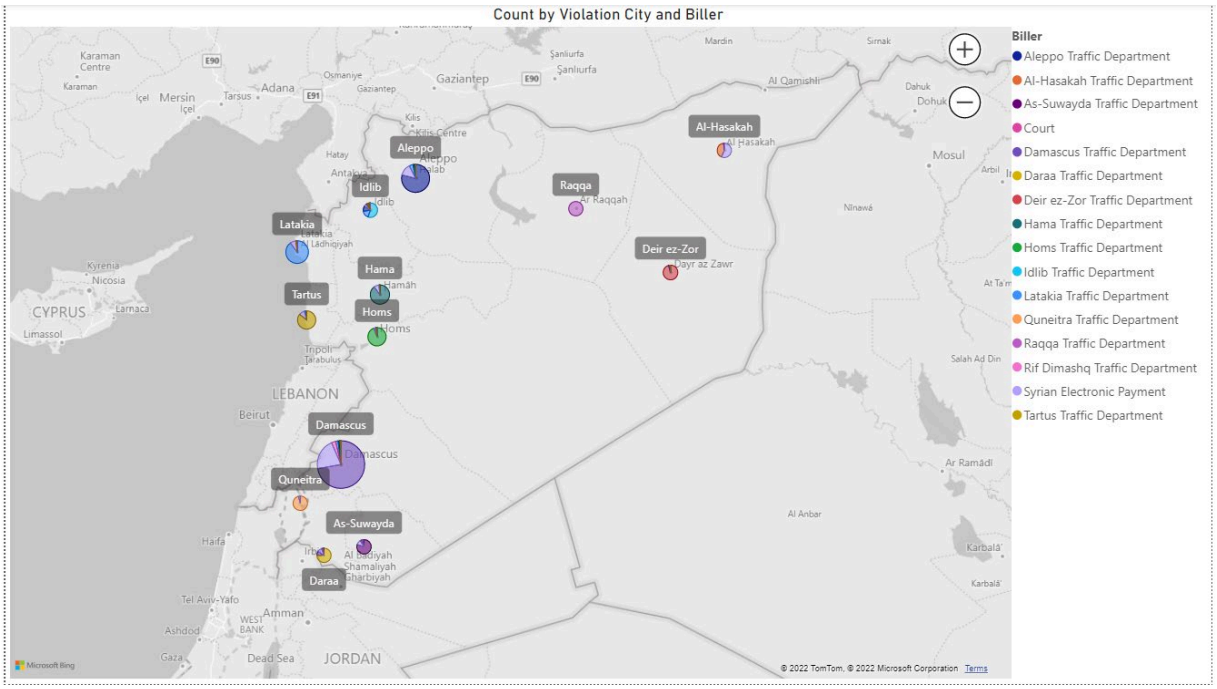

Violation Type
حضورية
126.56K (31.59%)
غيابية
274.02K (68.41%)

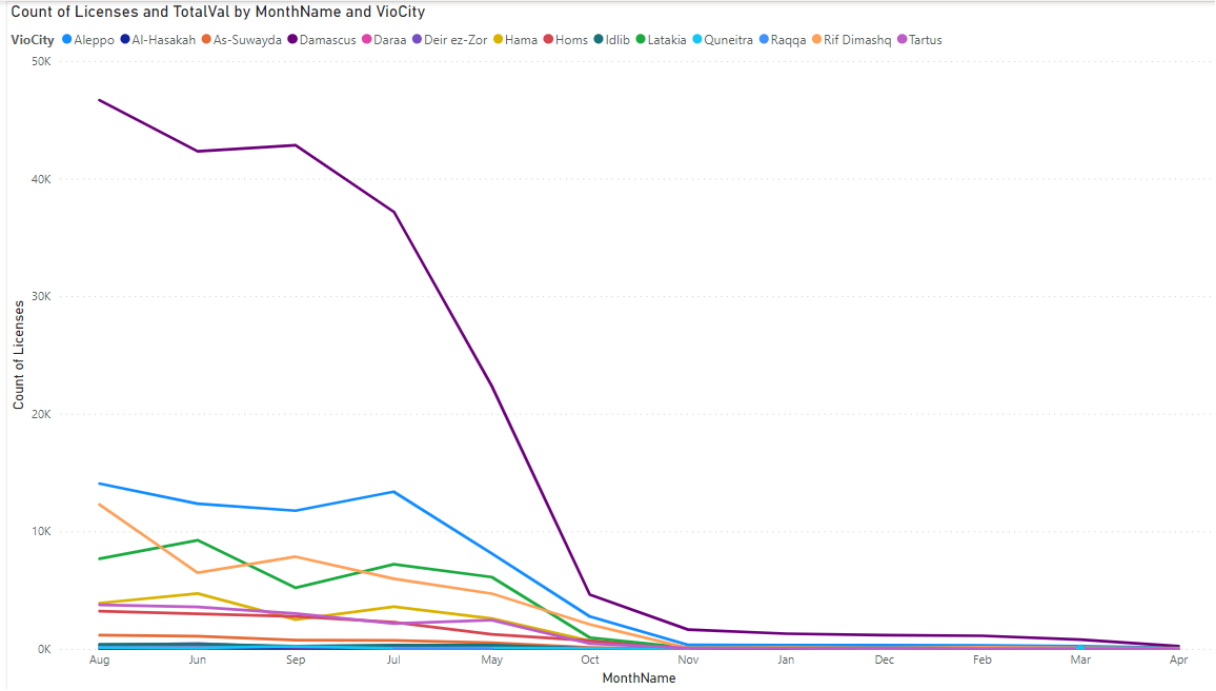

Violation Status
إحالة إلى القضاء (دون إرجاع نقاط)
12.99K (3.24%)
دفع المخالفة
155.44K (38.8%)
قيد الانتظار 230.45K (57.53%)

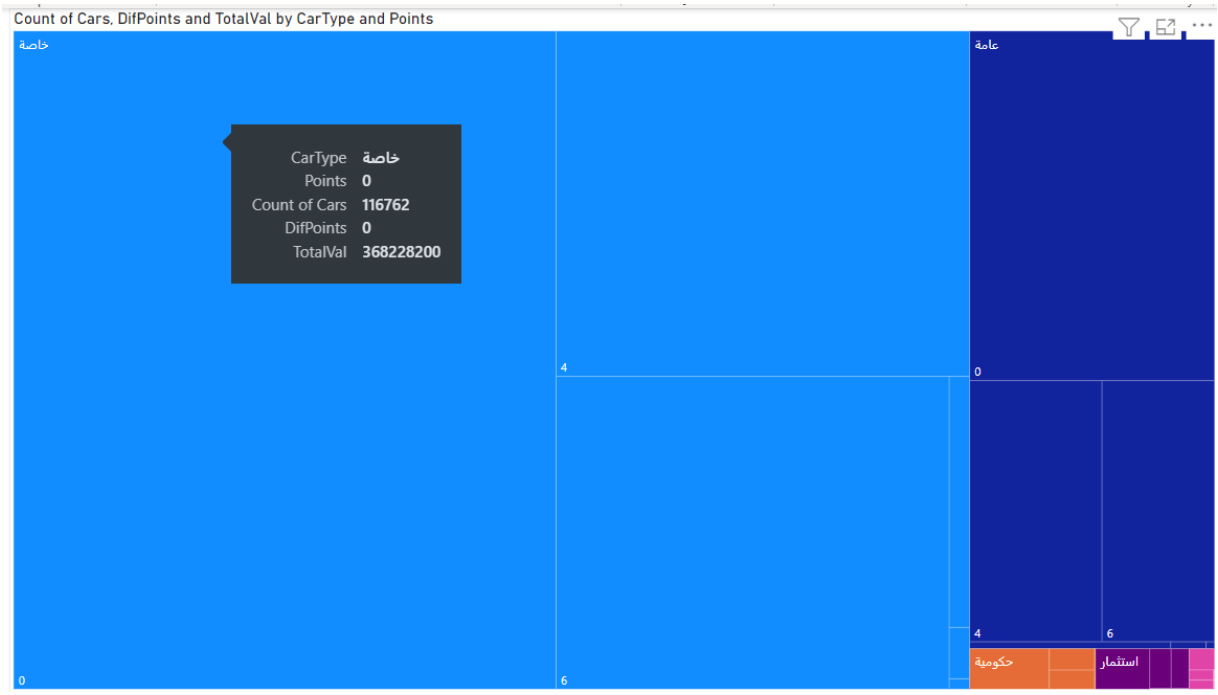

Count by License Type
0.02K (0.01%)
29.02K (13.49%)
94.47K (43.93%)
76.91K (35.76%)
0.22K (0.1%)
License Type
اشغال هـ
خاصة ب
دراجة أ
زراعيه هـ
عامة د1
عامة د2
عامة صغيرة ج
عجزة ومقعدين (و)

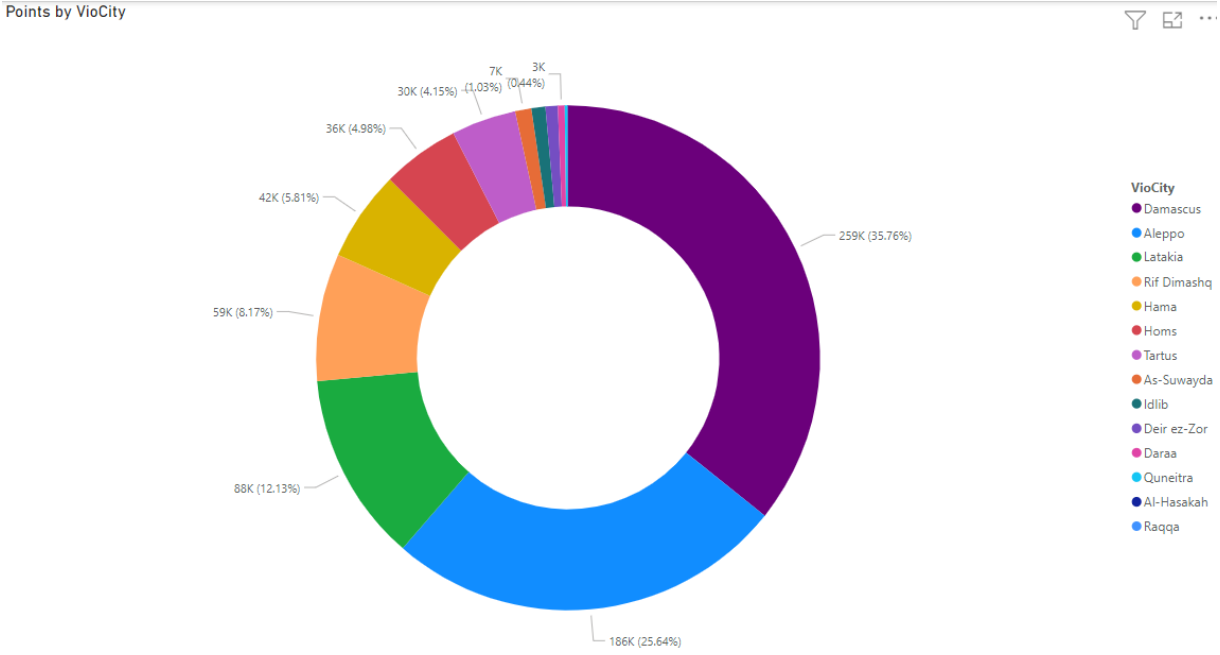

Total Value by Biller
100%
Damascus Traffic Department
160M
Aleppo Traffic Department
96M
Latakia Traffic Department
74M
Syrian Electronic Payment
50M
Hama Traffic Department
38M
Homs Traffic Department
25M
Tartus Traffic Department
23M
Rif Dimashq Traffic Department
8M
Deir ez-Zor Traffic Department
4M
Court
3M
As-Suwayda Traffic Department
3M
Idlib Traffic Department
2M
Daraa Traffic Department
2M
Quneitra Traffic Department
1M
Al-Hasakah Traffic Department
0M
Raqqa Traffic Department
0M
0%

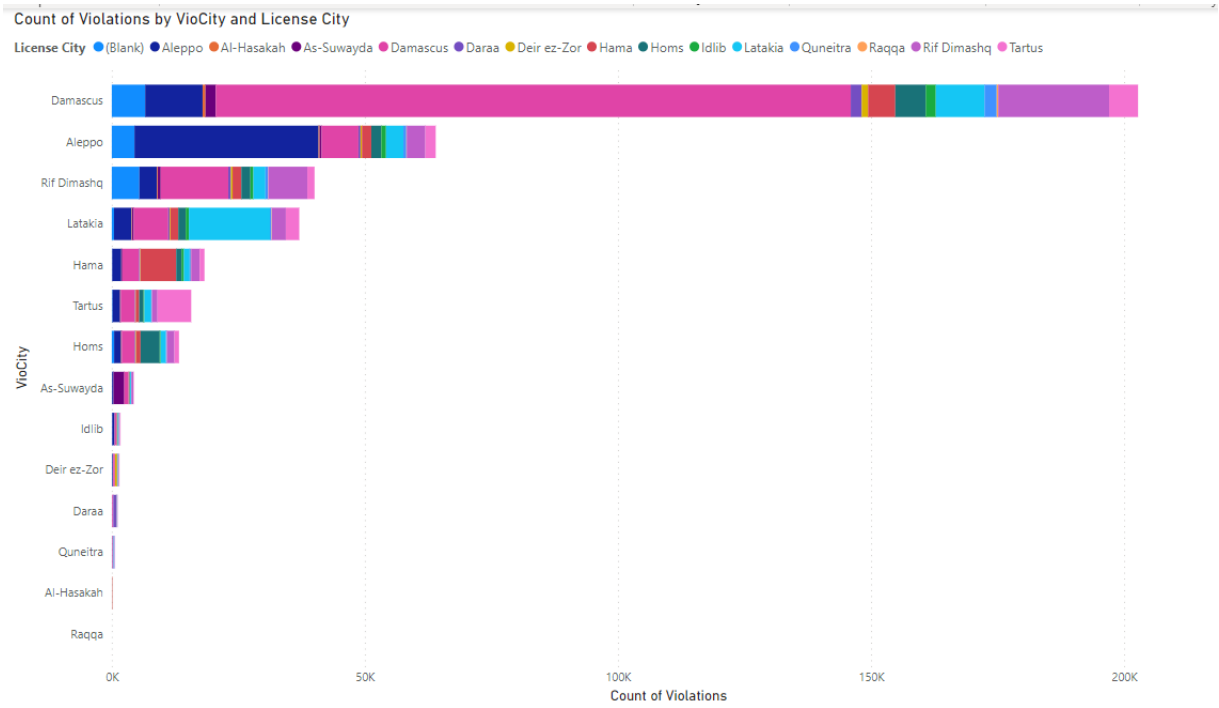

Count by Quarter and Violation City
Violation City
Aleppo
Al-Hasakah
As-Suwayda
Damascus
Daraa
Deir ez-Zor
Hama
Homs
Idlib
Latakia
Quneitra
Raqqa
Rif Dimashq
Tartus
Qtr 1 January
Qtr 1 February
Qtr 1 March
Qtr 2 April
Qtr 2 May
Qtr 2 June
Qtr 3 July
Qtr 3 August
Qtr 3 September
Qtr 4 October
Qtr 4 November
Qtr 4 December
Full Date Month

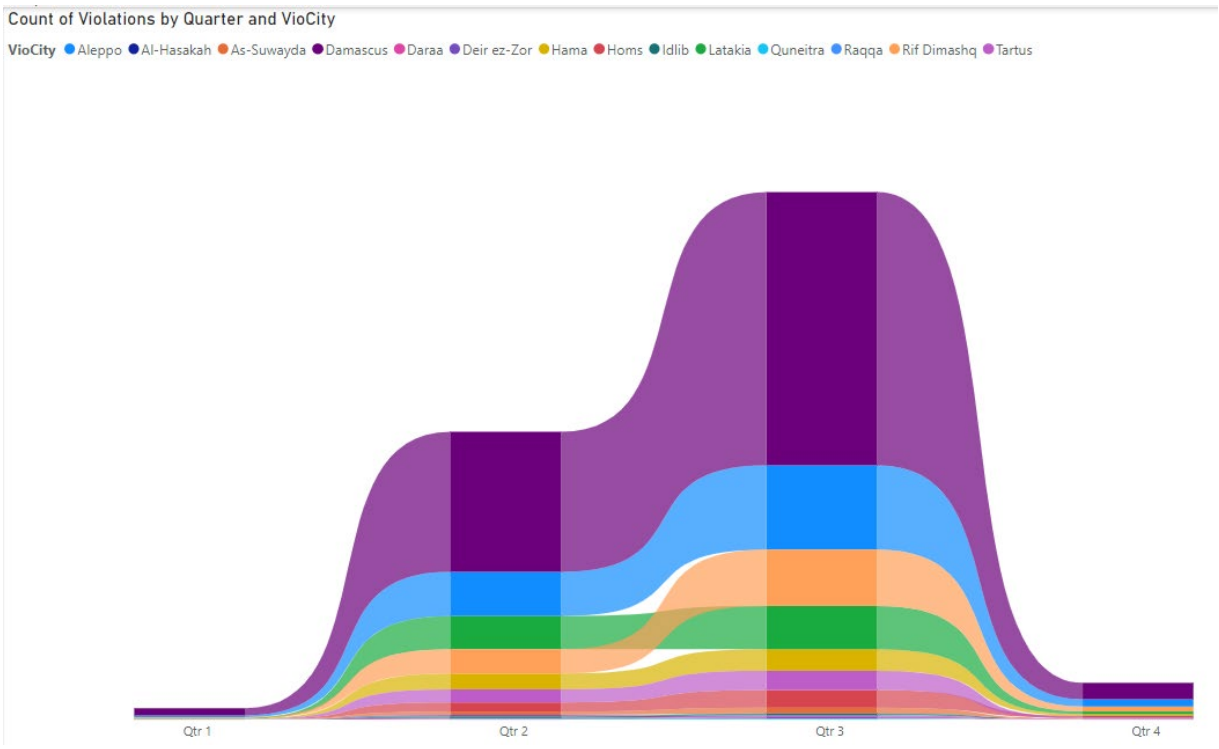

Count by Violation City and Biller
Biller
Aleppo Traffic Department
Al-Hasakah Traffic Department
As-Suwayda Traffic Department
Court
Damascus Traffic Department
Daraa Traffic Department
Deir ez-Zor Traffic Department
Hama Traffic Department
Homs Traffic Department
Idlib Traffic Department
Latakia Traffic Department
Quneitra Traffic Department
Raqqa Traffic Department
Rif Dimashq Traffic Department
Syrian Electronic Payment
Tartus Traffic Department
Aleppo
Idlib
Latakia
Hama
Tartus
Homs
Damascus
Quneitra
As-Suwayda
Daraa
Al-Hasakah
Raqqa
Deir ez-Zor
LEBANON
JORDAN
CYPRUS
© 2022 TomTom, © 2022 Microsoft Corporation Terms

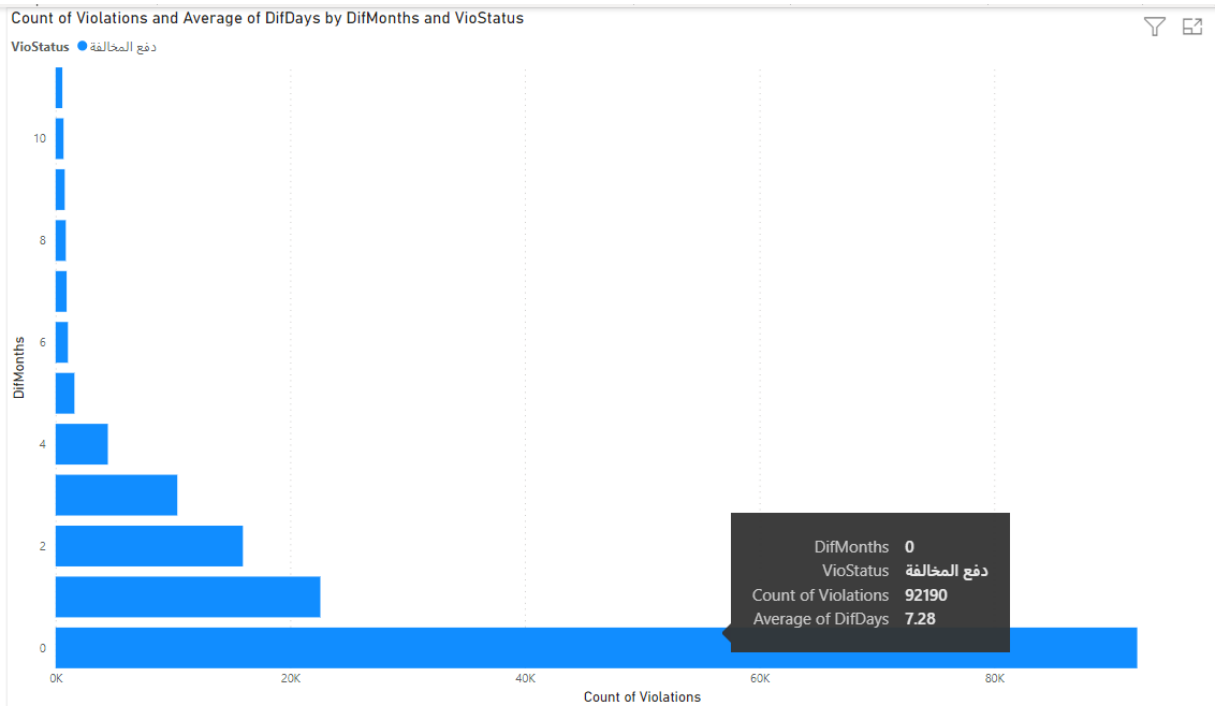

Count of Licenses and TotalVal by MonthName and VioCity
VioCity Aleppo Al-Hasakah As-Suwayda Damascus Daraa Deir ez-Zor Hama Homs Idlib Latakia Quneitra Raqqa Rif Dimashq Tartus
50K
40K
30K
20K
10K
0K
Count of Licenses
Aug Jun Sep Jul May Oct Nov Jan Dec Feb Mar Apr
MonthName

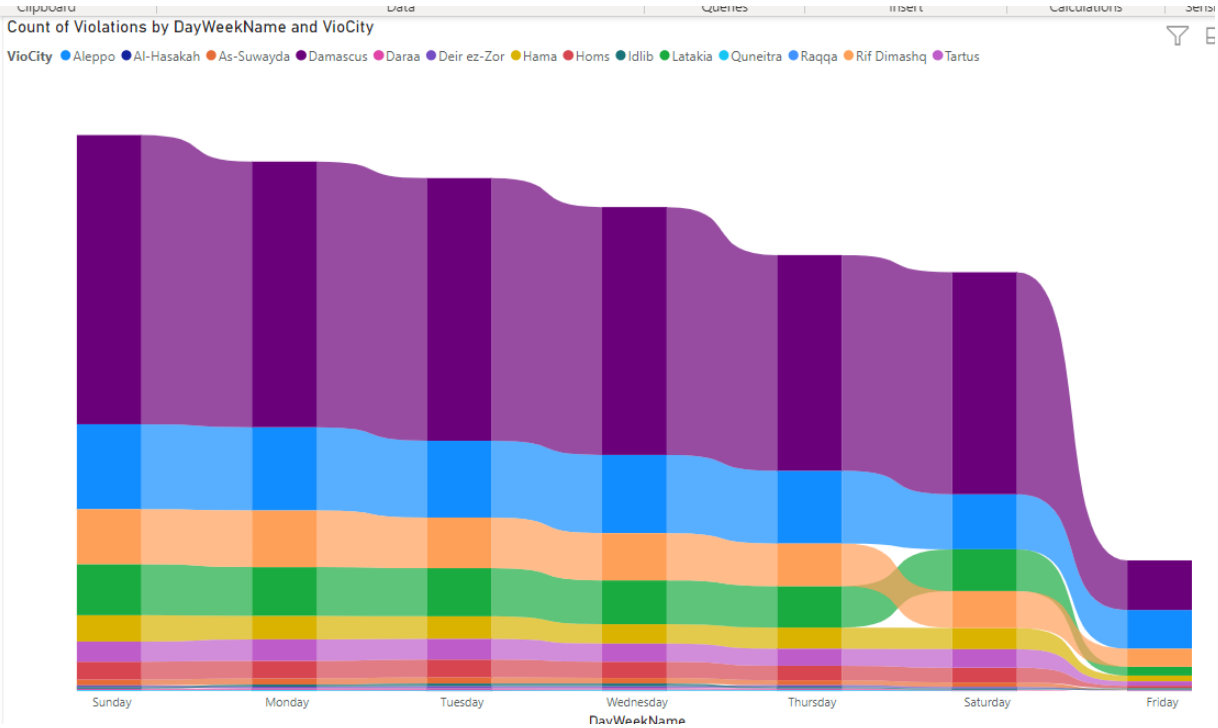

Count of Cars, DifPoints and TotalVal by CarType and Points
خاصة
عامة
CarType خاصة
Points 0
Count of Cars 116762
DifPoints 0
TotalVal 368228200
0
4
6
0
4
6
حكومية
استثمار


Points by VioCity
259K (35.76%)
186K (25.64%)
88K (12.13%)
59K (8.17%)
42K (5.81%)
36K (4.98%)
30K (4.15%)
7K (1.03%)
3K (0.44%)
VioCity
Damascus
Aleppo
Latakia
Rif Dimashq
Hama
Homs
Tartus
As-Suwayda
Idlib
Deir ez-Zor
Daraa
Quneitra
Al-Hasakah
Raqqa


Count of Violations by VioCity and License City
License City (Blank) Aleppo Al-Hasakah As-Suwayda Damascus Daraa Deir ez-Zor Hama Homs Idlib Latakia Quneitra Raqqa Rif Dimashq Tartus
VioCity
Damascus
Aleppo
Rif Dimashq
Latakia
Hama
Tartus
Homs
As-Suwayda
Idlib
Deir ez-Zor
Daraa
Quneitra
Al-Hasakah
Raqqa
0K
50K
100K
150K
200K
Count of Violations


Count of Violations by Quarter and VioCity
VioCity Aleppo Al-Hasakah As-Suwayda Damascus Daraa Deir ez-Zor Hama Homs Idlib Latakia Quneitra Raqqa Rif Dimashq Tartus
Qtr 1
Qtr 2
Qtr 3
Qtr 4


Count of Violations and Average of DifDays by DifMonths and VioStatus
VioStatus دفع المخالفة
DifMonths
10
8
6
4
2
0
0K
20K
40K
60K
80K
Count of Violations
DifMonths 0
VioStatus دفع المخالفة
Count of Violations 92190
Average of DifDays 7.28


Count of Violations by DayWeekName and VioCity
VioCity Aleppo Al-Hasakah As-Suwayda Damascus Daraa Deir ez-Zor Hama Homs Idlib Latakia Quneitra Raqqa Rif Dimashq Tartus
Sunday
Monday
Tuesday
Wednesday
Thursday
Saturday
Friday
DayWeekName

## Conclusion:

There are many reasons why companies and ministerial departments adopt BI. Many use it to support functions as diverse as hiring, compliance, production, and marketing. BI is a core business value; it is difficult to find a business area that does not benefit from better information to work with.

Some of the many benefits companies can experience after adopting BI into their business models include faster, more accurate reporting and analysis, improved data quality, better employee satisfaction, reduced costs, and increased revenues, and the ability to make better business decisions. The main objective of this project to show how to apply Business intelligence solutions on the ministry. Performance of the ministry requires continuous monitoring. Therefore, implementing decision support systems and business intelligence technologies, which could provide ministerial officials with summarized data, is considered highly important.

# References

bimentalist.com. (n.d.).

EMC Education Services (2014), Data Science and Big Data Analytics. (n.d.).

thecustomizewindows.com. (n.d.).

www.datawarehouse4u.info. (n.d.).

www.geeksfogeeks.com. (n.d.).

www.mobcoder.com. (n.d.).

www.researchgate.net. (n.d.).