Database Programming

Data Warehouse for impact of alcohol, drugs and smoking during pregnancy on Births

Supervisor

Dr. Alaa El-Din Abdallah

Team Members

Mostafa Hamdy Abo El-Ela

Nada Mohamed Salah El-Din

Nourhan Reda Abd El-Hay

Yousef Ragab Abd El-Motaleb

1. Problem Definition

The health of a woman is an important factor in **pregnancy** – *healthy women have healthy babies.* There are many lifestyle factors affecting the physical and emotional well-being of a woman during pregnancy. Medical care is also important regular antenatal care and screening to monitor the health or mother and baby.

There are also factors associated with the risk to the health of the mother and baby – **smoking, alcohol, drug misuse, age, and deprivation**. Another important factor is the rising proportion of births to older women. Older maternal age may be associated with pre-existing ill health, low fertility, complications of pregnancy and an increased risk of adverse outcomes, including stillbirths and congenital anomalies.

Data warehouse works as foundation for **decision making** process, as for taking considering organization (data related), so, we preliminary focuses on the DW. It makes information easy to accessible as we can **generate reports**, like Operational & Enterprise report from the data warehouse. It is the most reliable technology used by organizations such as health centers in forecasting and management. Data storage is one of best technology for data integration.

2. Data Sources

We found it very difficult to access suitable data (Real) because it is very confidential, but we reached "**Public Health Scotland**" and we obtained data from **2015/16** to **2020/21** about **Births in Scotland**.

- ➢ Births Table Data ([Click Here](#))
- ➢ Smoking Table Data ([Click Here](#))

➢ DrugsMisuse Table Data ([Click Here](#))

➢ Alcohol Table Data ([Click Here](#))

3. Data Warehouse

We aimed to implement a data warehouse based on births in Scotland data sources. This section shown how we divided our work into steps in order to provide fully understanding to the procedure of implementing data warehouse. These steps shown in figure1.
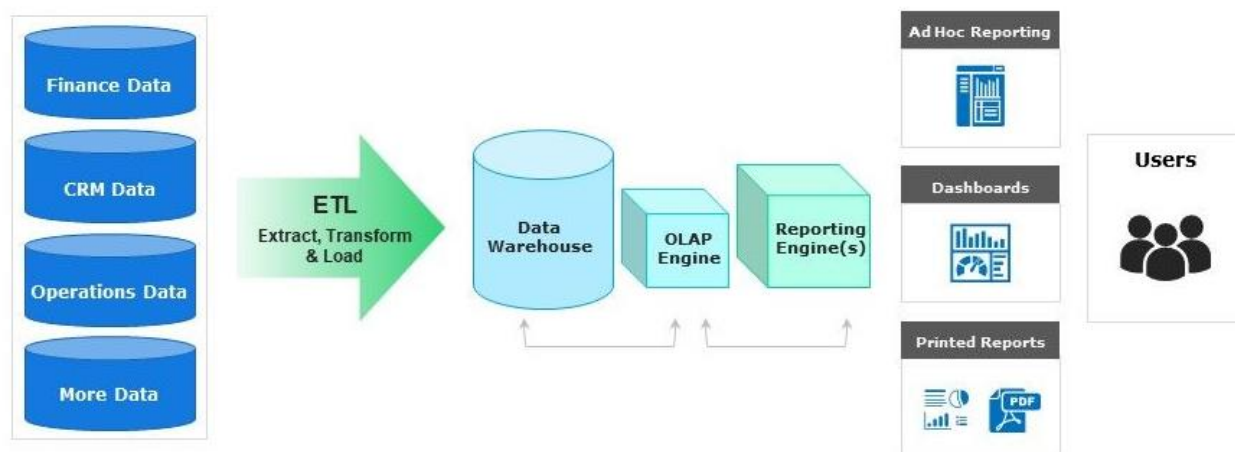

Figure 1: Data Warehouse Architecture

Data Acquisition

The process of extracting the data from different source (operational databases) systems, integrating the data, and transforming the data into a homogenous format and loading into the target warehouse database.

Data Staging

A staging area is mainly required in a Data Warehousing Architecture for timing reasons. In short, all required data must be available before data can be integrated into the Data Warehouse as it is temporary location where data from source systems is copied.

Extract, Transform and Load (ETL)

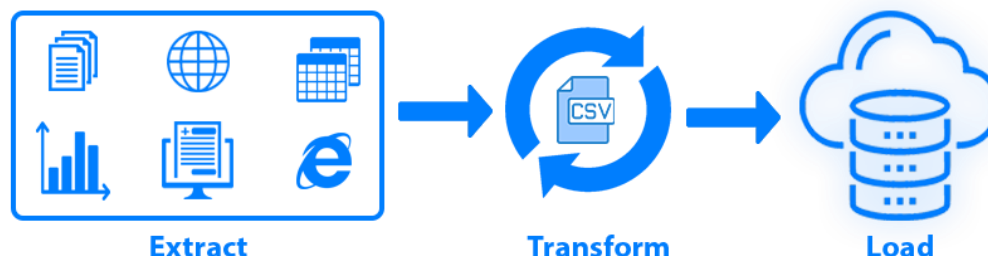The structure design of our ETL based on tools from **SQL Server** as shown in figure2.
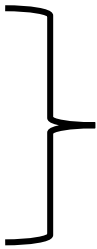

Figure 2: ETL Process

1. **Data Extraction:** is the process of reading the data from various types of source systems. As we mentioned in data sources, we get data from "Public Health in scoltland", we have taken **4 databases** from website: Births By Hospital - Smoking behaviour during pregancy - Alcohol Consumption in Pregnancy - Drug Misuse in Pregnancy. These data are formatted in csv file.

2. **Data Transformation:** is the process of transforming data and cleansing the data into the required business format. The staging database is created in the most denormalized form. Apart from the primary key, all other fields allow NULL value. Primary keys are filled up with incremental numbers before they are loaded onto database. Now, we have created **tables** in database in SQL Sever using SQL Quries (**CREATE TABLE**).

3. **Data Loading:** is the process of inserting the data into a target systems, load data from csv files to tables in database using SQL Quries (**INSERT**).

A new schema called "**Pregnancy**" is created. The **CREATE** statements of all the finalized tables in the staging schema will be captured, modified, and executed in the schema.

4. The Dimensions and Fact Tables

4.1 The Fact Table:

StillBirths Table consists of set of attributes:

- StillBirth_ID (int) as **primary key**

- Alcohol_SK (int)

- DrugMisuse_SK (int)

- Birth_SK (int)

- Smoking_SK (int)

    as **surrogate** or **Foreign** keys

- Year (YEAR)

- SmokerNum (int)

- AlcoholConsumers (int)

- DrugsConsumers (int)

- StillBirthsNum (int)

4.2 Four Dimension Tables:

- **Smoking Table:** Smoking behavior in pregnancy is collected at a woman's first antenatal booking appointment (**9435 records**).

- **Alcohol Table:** Alcohol consumption in pregnancy is collected at a woman's antenatal booking appointment (**551 records**).

- **Drug Misuse Table:** It stores the number of maternities where drug misuse in pregnancy has been recorded (**187 records**).

- **Births Table**: It stores the number of live and stillbirths by hospital of birth (**1812 records**).

These Figures will describe the above steps (**Dimension Tables**):

ID	Year	Consumption	Maternities
1	2015/16	No	525
2	2015/16	Unknown	11
3	2015/16	Yes	87
4	2015/16	No	112
5	2015/16	Unknown	18
6	2015/16	Yes	3
7	2015/16	No	532
8	2015/16	Unknown	NULL
9	2015/16	Yes	NULL
10	2015/16	No	1104
11	2015/16	Unknown	11
12	2015/16	Yes	141
13	2015/16	No	860
14	2015/16	Unknown	13
15	2015/16	Yes	399
16	2015/16	No	801
17	2015/16	Unknown	11
18	2015/16	Yes	213
19	2015/16	No	718
20	2015/16	Unknown	112

Figure 3: Alcohol Dimension table (Summarized)

ID	Year	AgeGroup	SmokingStatus	Maternities
1	2015/16	25-34	Current smoker	7
2	2015/16	25-34	Former smoker	16
3	2015/16	25-34	Never smoked	43
4	2015/16	25-34	Unknown	16
5	2015/16	35 and over	Former smoker	3
6	2015/16	35 and over	Never smoked	12
7	2015/16	35 and over	Unknown	4
8	2015/16	Under 25	Current smoker	5
9	2015/16	Under 25	Former smoker	8
10	2015/16	Under 25	Never smoked	16
11	2015/16	Under 25	Unknown	3
12	2015/16	25-34	Current smoker	21
13	2015/16	25-34	Former smoker	4
14	2015/16	25-34	Never smoked	51
15	2015/16	25-34	Current smoker	14
16	2015/16	25-34	Former smoker	11
17	2015/16	25-34	Never smoked	68
18	2015/16	25-34	Current smoker	4
19	2015/16	25-34	Former smoker	5
20	2015/16	25-34	Never smoked	44

Figure 4: Smoking Dimension Table (Summarized)

ID	Year	Num	Maternities
1	2015/16	15	1905
2	2015/16	5	451
3	2015/16	17	1600
4	2015/16	65	3728
5	2015/16	45	3772
6	2015/16	89	2943
7	2015/16	38	4902
8	2015/16	64	6598
9	2015/16	26	2135
10	2015/16	95	2997
11	2015/16	28	2707
12	2015/16	59	3764
13	2015/16	20	2895
14	2015/16	30	3002
15	2015/16	60	9600
16	2015/16	30	2425
17	2015/16	160	8006
18	2015/16	86	7867
19	2015/16	8	2008
20	2015/16	452	16129

Figure 5: DrugsMisuse Dimension Table (Summarized)

ID	Year	Hospital	Outcome	Num
1	2015/16	A111H	Live	1
2	2015/16	B120H	Live	72
3	2015/16	C418H	Live	2
4	2015/16	F705H	Live	2
5	2015/16	G405H	Live	8
6	2015/16	H202H	Live	3
7	2015/16	H202H	Still	7
8	2015/16	L308H	Live	6
9	2015/16	N161H	Live	6
10	2015/16	N411H	Live	1
11	2015/16	S308H	Live	3
12	2015/16	S314H	Live	17
13	2015/16	T101H	Live	2
14	2015/16	V217H	Live	1
15	2015/16	Y104H	Live	12
16	2015/16	F705H	Live	3
17	2015/16	G108H	Live	2
18	2015/16	G405H	Live	4
19	2015/16	T101H	Live	1
20	2015/16	V217H	Live	535

Figure 6: Births Dimension Table (Summarized)

5. Data Warehouse Schema

We designed the data warehouse schema based on our births data fields. The proposed schema is **star schema** which it is the most appropriate schema for our project because it can be understood by the professionals and users. Figure (7) shows our proposed schema for the proposed Data Warehouse.
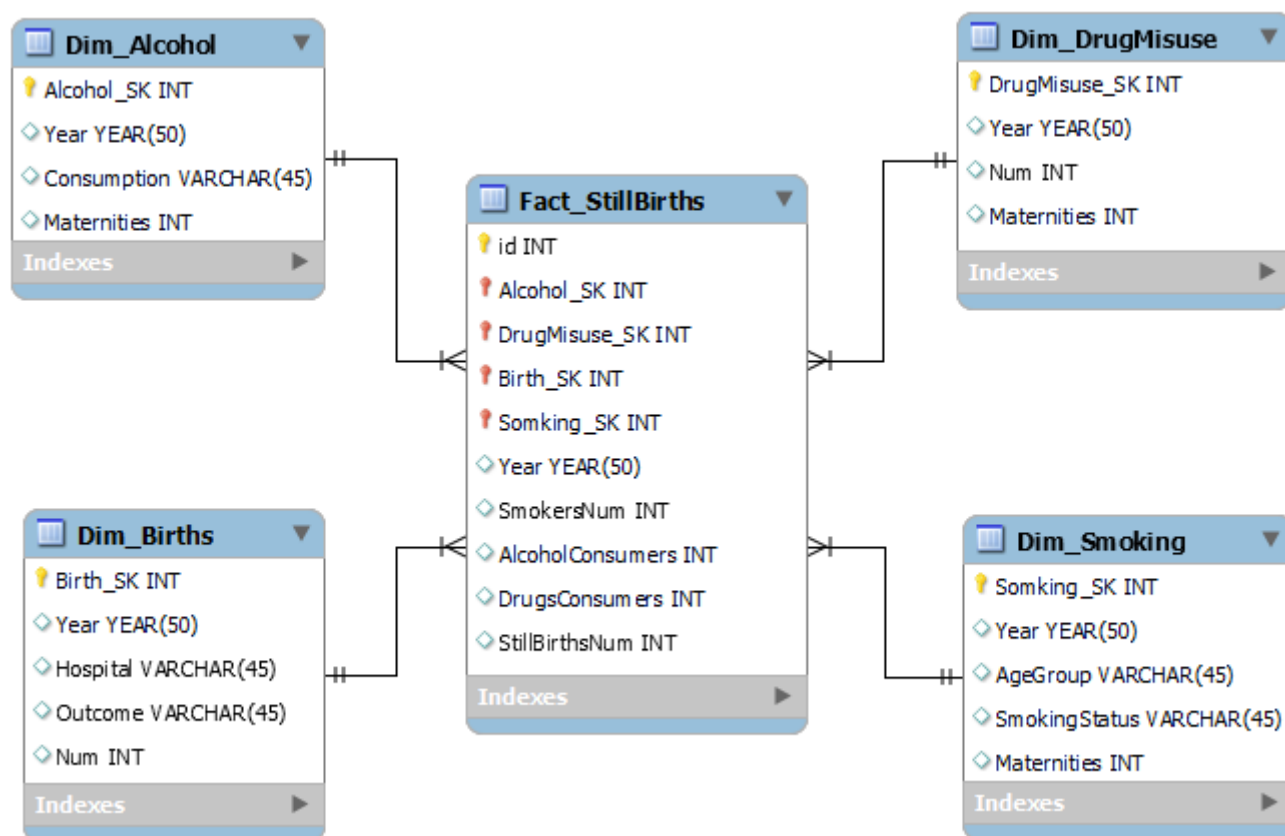

Figure 7 Pregnancy Data Warehouse Star Schema

Here, we combine births data, flat file and data stored in **SQL Server**. The relationship between fact table and dimensions is constructed by using surrogate keys (**SK**).
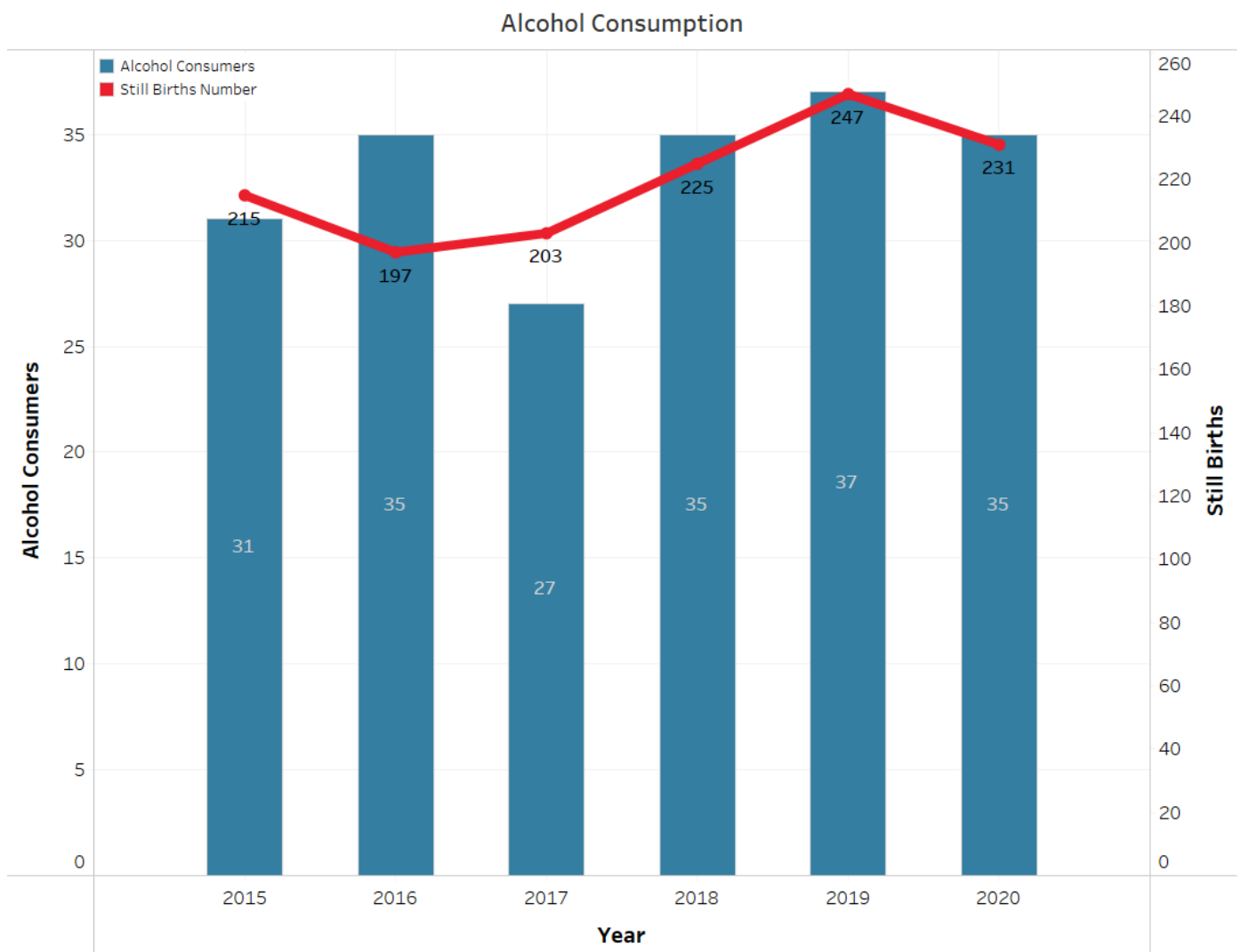
6. Reports

Our Goal is to create statistics about the impact of (**Smoking - Alcohol - Drugs**) during pregnancy on Births in the period from 2015 to 2020:

6.1 Smoking Report:
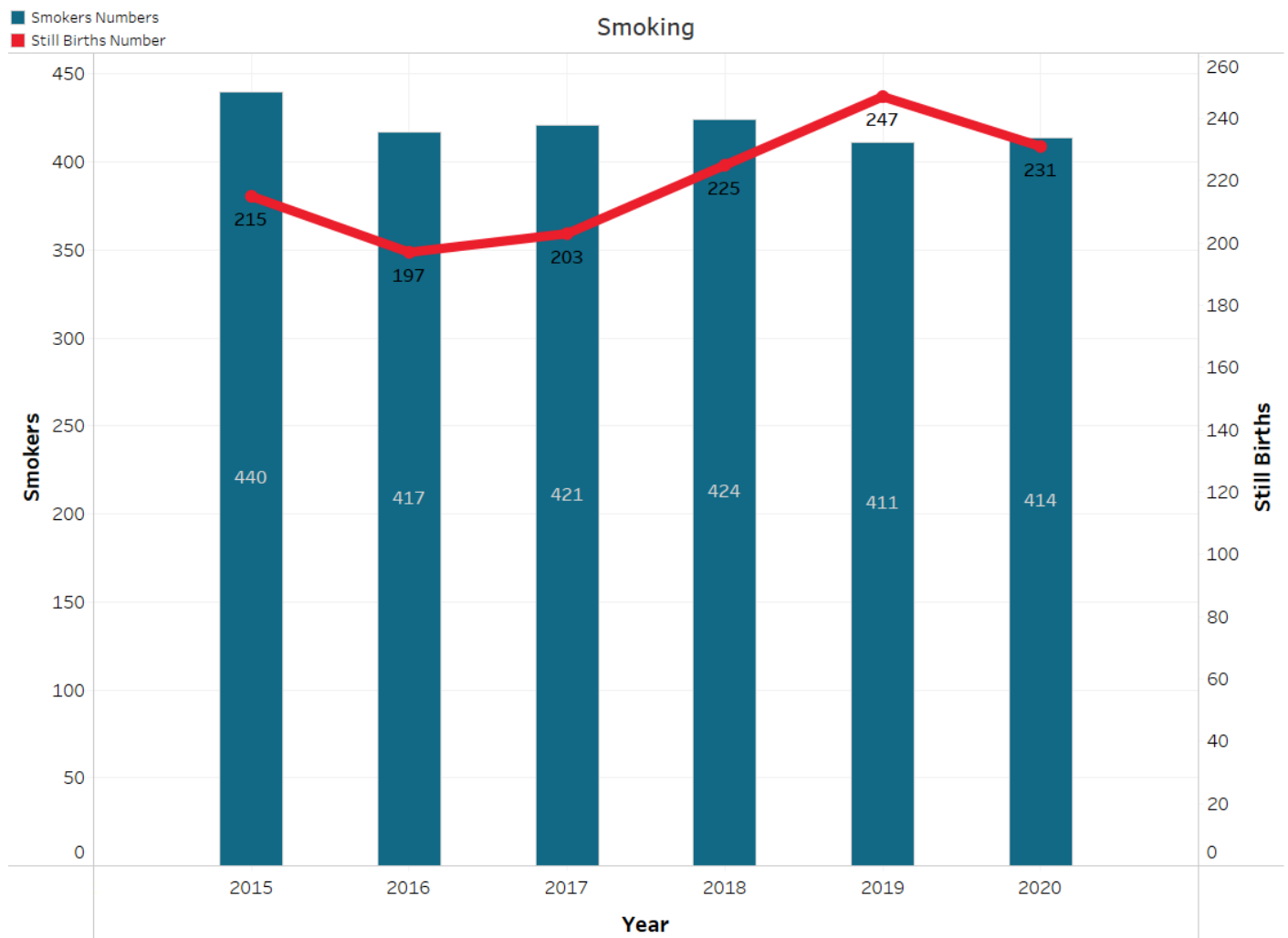
> ➤ This chart describes the relationship between **the number of Maternities where smokers are recorded** and **Still Births** in the duration between 2015 and 2020.

6.2 Alcohol Report:

> ➤ This chart describes the relationship between **the number of Maternities where Alcohol Consumers are recorded** and **Still Births** in the duration between 2015 and 2020.

6.3 Drugs Misuse Report:

Drugs Misuse
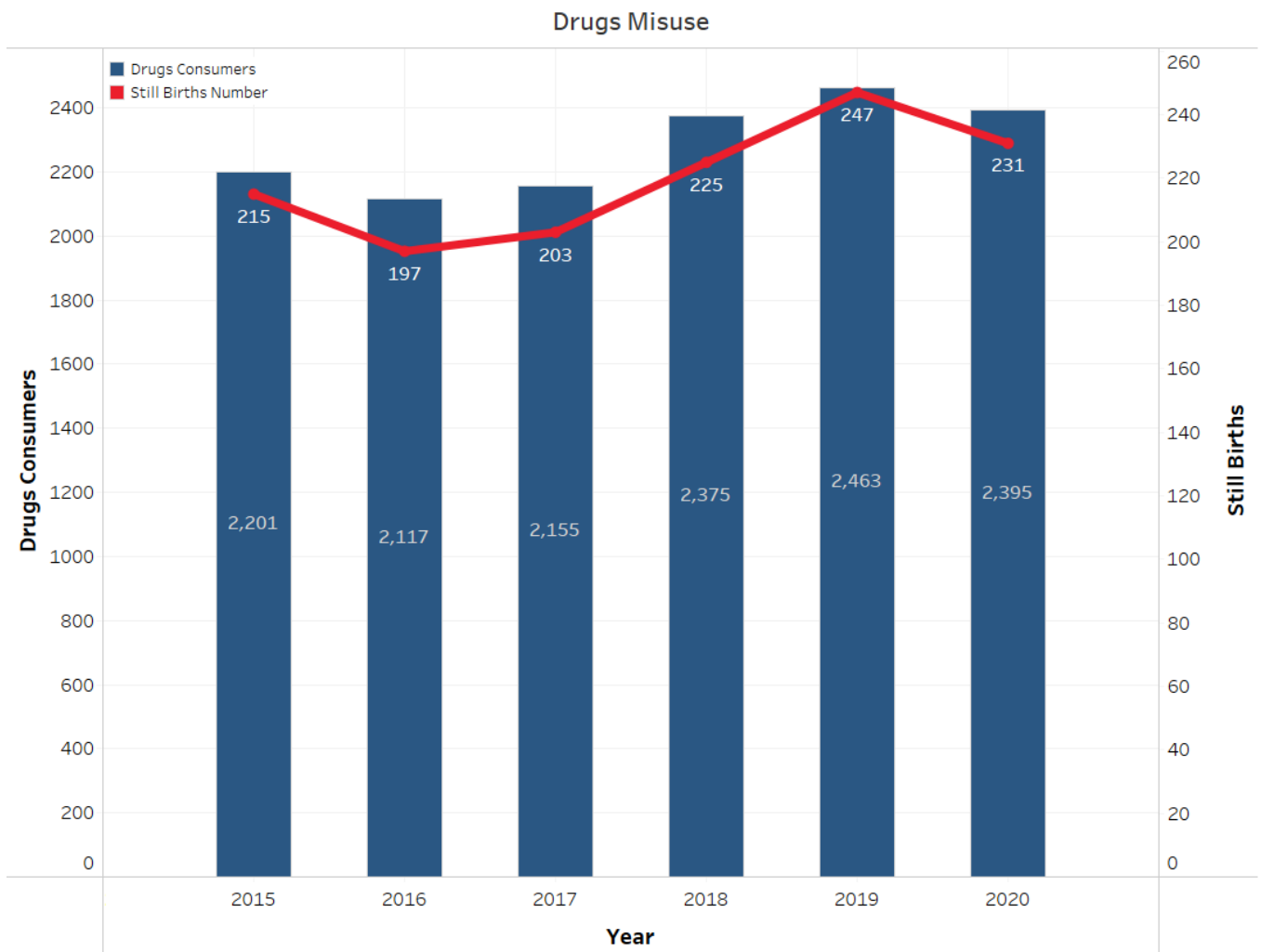
➢ This chart describes the relationship between **Drugs Consumers** and **Still Births** in the duration between 2015 and 2020.