

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО”
ФАКУЛЬТЕТ ІНФОРМАТИКИ ТА ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ

Кафедра інформатики та програмної інженерії

Звіт до лабораторної роботи №1

з курсу

«АД»

студента 2 курсу
групи IT-02
Макарова Іллі Сергійовича

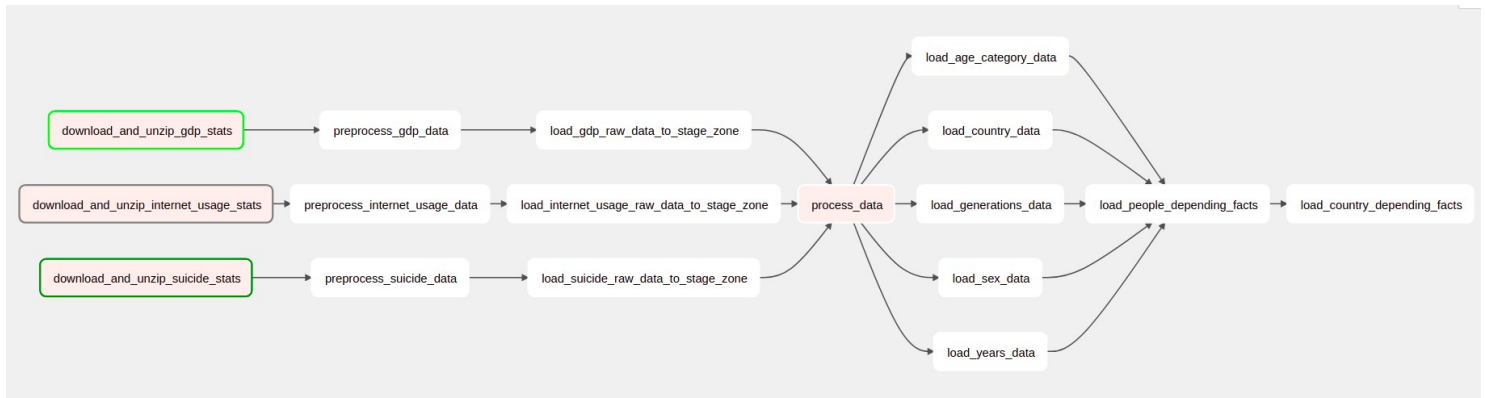
Викладач:

Київ – 2022

Для виконання роботи я обрав три датасети з Kaggle, перший зі статистикою ВВП на душу населення країн світу, статистику суїциду по країнам та статистику розповсюдження інтернету.

Весь код можна знайти тут: <https://github.com/kinfi4/ELT-for-gathering-suicide-and-internet-usage-stats>

Структура пайплайну:



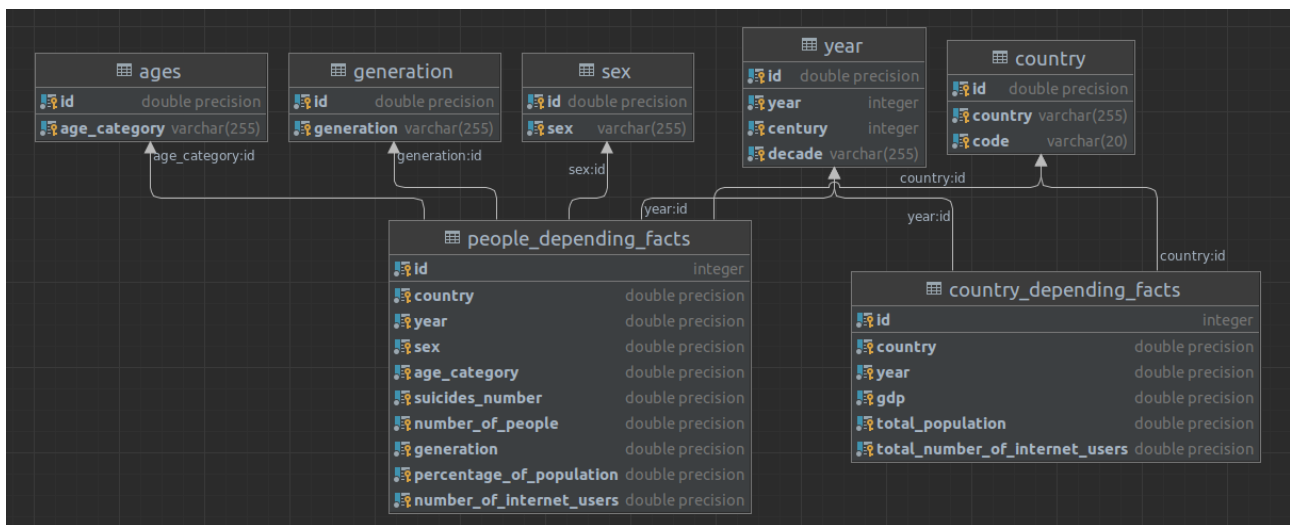
Stage zone:

suicide_data_raw	
id	integer
country	varchar(255)
year	double precision
sex	varchar(255)
age	varchar(255)
suicides_number	double precision
population	double precision
generation	varchar(255)

internet_usage_data_raw	
id	integer
country	varchar(255)
code	varchar(255)
year	double precision
number_of_internet_users	double precision

country_data_raw	
id	integer
country	varchar(255)
year	double precision
gdp	double precision

Star schema:



Python code for pipeline can be found here: <https://github.com/kinfi4/ELT-for-gathering-suicide-and-internet-usage-stats>

SQL scripts for creating tables:

```
CREATE TABLE sex (  
  id DOUBLE PRECISION PRIMARY KEY,  
  Sex VARCHAR(255)  
);
```

```
CREATE TABLE generation (  
  id DOUBLE PRECISION PRIMARY KEY,  
  Generation VARCHAR(255)  
);
```

```
CREATE TABLE year (  
  id DOUBLE PRECISION PRIMARY KEY,  
  Year INT NULL,  
  Century INT NULL,  
  Decade VARCHAR(255)  
);
```

```
CREATE TABLE country (  
  id DOUBLE PRECISION PRIMARY KEY,  
  Country VARCHAR(255) NULL,  
  Code VARCHAR(20) NULL  
);
```

```
CREATE TABLE ages (  
  id DOUBLE PRECISION PRIMARY KEY,  
  Age_category VARCHAR(255) NULL  
);
```

```
CREATE TABLE people_dependent_facts (  
  id SERIAL PRIMARY KEY,  
  Country DOUBLE PRECISION NULL,
```

```

Year DOUBLE PRECISION NULL,
Sex DOUBLE PRECISION NULL,
Age_category DOUBLE PRECISION NULL,
Suicides_number DOUBLE PRECISION NULL,
Number_of_people DOUBLE PRECISION NULL,
Generation DOUBLE PRECISION NULL,
Percentage_of_population DOUBLE PRECISION NULL,
Number_of_internet_users DOUBLE PRECISION NULL,

FOREIGN KEY (Country) REFERENCES country(id),
FOREIGN KEY (Year) REFERENCES year(id),
FOREIGN KEY (Sex) REFERENCES sex(id),
FOREIGN KEY (Age_category) REFERENCES ages(id),
FOREIGN KEY (Generation) REFERENCES generation(id)
);

CREATE TABLE country_dependent_facts (
  id SERIAL PRIMARY KEY,
  Country DOUBLE PRECISION NULL,
  Year DOUBLE PRECISION NULL,
  GDP DOUBLE PRECISION NULL,
  Total_population DOUBLE PRECISION NULL,
  Total_number_of_internet_users DOUBLE PRECISION NULL,

  FOREIGN KEY (Country) REFERENCES country(id),
  FOREIGN KEY (Year) REFERENCES year(id)
);

CREATE TABLE country_data_raw(
  id SERIAL PRIMARY KEY,
  Country VARCHAR(255) NULL,
  Year DOUBLE PRECISION NULL,
  GDP DOUBLE PRECISION NULL
);

CREATE TABLE suicide_data_raw(
  id SERIAL PRIMARY KEY,
  Country VARCHAR(255) NULL,
  Year DOUBLE PRECISION NULL,
  Sex VARCHAR(255) NULL,
  Age VARCHAR(255) NULL,
  Suicides_number DOUBLE PRECISION NULL,
  Population DOUBLE PRECISION NULL,
  Generation VARCHAR(255)
);

CREATE TABLE internet_usage_data_raw(
  id SERIAL PRIMARY KEY,
  Country VARCHAR(255) NULL,
  Code VARCHAR(255) NULL,
  Year DOUBLE PRECISION NULL,
  Number_of_internet_users DOUBLE PRECISION NULL
);

```

