

# CUNY MSDA - IS607 : DATA ACQUISITION AND MANAGEMENT

## Week 7 Project

A. Leopold HILLAH

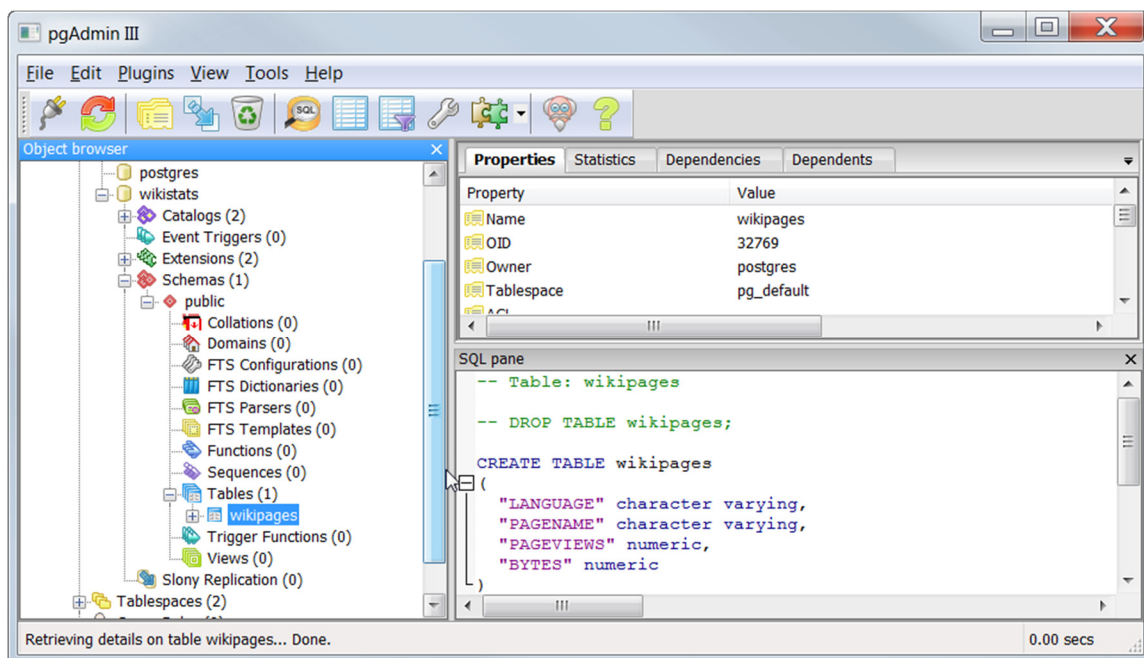
### ##### Creation of a new database #####

Creation of **wikistats** database.

```
CREATE DATABASE wikistats
WITH ENCODING='UTF8'
OWNER=postgres
TEMPLATE=postgres
LC_COLLATE='English_United Kingdom.1252'
LC_CTYPE='English_United Kingdom.1252'
CONNECTION LIMIT=-1
TABLESPACE=pg_default;

COMMENT ON DATABASE wikistats
IS 'Wikipedia Traffic Statistics';
```

```
CREATE TABLE wikipages
(
  "LANGUAGE" character varying,
  "PAGENAME" character varying,
  "PAGEVIEWS" numeric,
  "BYTES" numeric
)
WITH (
  OIDS=FALSE
);
ALTER TABLE wikipages
OWNER TO postgres;
COMMENT ON TABLE wikipages
IS 'Pageview statistics';
```



#### ##### Loading Data File #####

The relevant SQL statement is:

```
COPY WIKIPAGES FROM 'F:\R-Stats\20141001140000.txt' (DELIMITER(' '));
```

#### ##### The five Wikipedia pages that were visited most often #####

The relevant SQL statement is:

```
SELECT "PAGENAME", SUM("PAGEVIEWS") FROM WIKIPAGES  
GROUP BY "PAGENAME"  
ORDER BY 2 DESC  
LIMIT 5;
```

#### Results:

PAGENAME	sum
Database	257
Data:image/png	
base64,iVBORw0KGgoAAAANSUhEUgAAAAEAAAu	220
Data	203
Data_mining	196
Data_Protection_Act_1998	174

The screenshot shows a PostgreSQL query editor window titled "Query - wikistats on postgres@localhost:5432 \*". The SQL Editor tab is active, displaying the following query:

```
SELECT "PAGENAME", SUM("PAGEVIEWS") FROM WIKIPAGES  
GROUP BY "PAGENAME"  
ORDER BY 2 DESC  
LIMIT 5;
```

The Output pane at the bottom shows the results of the query in a table format:

	PAGENAME character varying	sum numeric
1	Database	257
2	Data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAAEAAAuCAIAAABmjeQ9AAAAARE1EQVR42mV0	220
3	Data	203
4	Data_mining	196
5	Data Protection Act 1998	174

The status bar at the bottom indicates "OK.", "Unix", "Ln 1, Col 1, Ch 1", "5 rows.", and "11 ms".