Australia, Land of Toilets: The Final Write-Up By Pete Wells, Stuart Rimel

n	ES	CR	Т	P٦	ГΤ	UΝ	L	n	F	F٦	П
IJ	כם	LR	L	Р 1	ш	U١		U	г	ᄗ	L

Our dataset includes over 22k public toilets throughout Australia, with each toilet containing 47 data points regarding all the attributes of the facilities (see TABLE CONTENT section). We have split up this data into 14 relational tables (see below CREATE TABLE STATEMENTS section), wich allows multiple perspectives on toilet data throught the land down under!

Our ETL application will load all 22k+ toilets split among 14 relational tables in about 45 minutes on the school's linux servers. The ETL application starts off by prompting for database credentials to login to our DB. Then, the ETL creates a connection by using python's psycopg2 library. This allows the application to create instances of our database such that we can run sql commands via python functions. The ETL then proceeds to create all the relational tables that will be utilized in our database (see SCHEMA DIAGRAM attachment). After each table is created (fully automated) a confirmation message is output. Once all the tables are created then the ETL invokes the data wrangling from our source csv file into dataframes using python's pandas library. This process allows our ETL to automate the organization of the source dataset into 14 dataframes that correspond to our 14 relational tables that were previously created by the application. Once the ETL wrangles the dataset into 14 dataframes, then the ETL will insert each row of each dataframe into it's respective relational table in sql via the to sql() function from the pandas library. With that, the entirety of the dataset is wrangled and inserted into newly created tables within our database.

Overall, the process of creating this database and inserting all data into was successfully automated with this ETL application. If we had more time, we would try to optimize the amount of time it takes to load the data into the database. There is definitely room for efficiency improvements!

Link to our ETL application repo --> https://github.com/srimel/ land_of_toilets

CREAT	Γ⊏	TARI		CTA	TEMEN	ITC
UKEAL	IE	LADL	- E	SIA		כוע

```
We created 14 relational tables for our dataset. See the following
create table statements that we used in our ETL:
# Create toilets table
def create_toilets_table(db_conn):
    cur = db conn.cursor()
    create stmt = "CREATE TABLE toilets(" \
                " FacilityID INT," \
                " URL VARCHAR(256)," \
                " Name VARCHAR(128)," \
                " Male B00L,"\
                " Female BOOL,"\
                " Unisex BOOL,"\
                " AllGender BOOL,"\
                " ToiletNote VARCHAR(1024),"\
                " DrinkingWater BOOL,"\
                " Shower BOOL,"\
                " PRIMARY KEY (FacilityID));"
    cur.execute(create_stmt)
# Create handicap table
def create handicap table(db conn):
    cur = db_conn.cursor()
    create_stmt = "CREATE TABLE handicap(" \
                " FacilityID INT," \
                " BYOSling BOOL,"\
                " Ambulant BOOL,"\
                " LHTransfer BOOL,"\
                " RHTransfer BOOL,"\
                " PRIMARY KEY (FacilityID),"\
                " CONSTRAINT handicap fk FOREIGN KEY(FacilityID) "\
                " REFERENCES toilets(FacilityID));"
    cur.execute(create stmt)
# Create changing table
def create_changing_table(db_conn):
    cur = db conn.cursor()
    create stmt = "CREATE TABLE changing(" \
                " FacilityID INT," \
                " BabyChange BOOL,"\
                " BabyCareRoom BOOL,"\
                " BabyChangeNote VARCHAR(400),"\
                " ACShower BOOL,"\
                " AdultChange BOOL,"\
                " AdultChangeNote VARCHAR(400),"\
                " ChangingPlaces BOOL,"\
                " PRIMARY KEY (FacilityID),"\
                " CONSTRAINT changing_fk FOREIGN KEY(FacilityID) "\
```

```
" REFERENCES toilets(FacilityID));"
    cur.execute(create_stmt)
# Create access table
def create access table(db conn):
    cur = db conn.cursor()
    create stmt = "CREATE TABLE access(" \
                " FacilityID INT," \
                " KeyRequired BOOL,"\
                " AccessNote VARCHAR(400)."\
                " PaymentRequired BOOL,"\
                " MLAK24 B00L,"\
                " MLAKAfterHours B00L,"\
                " OpeningHours VARCHAR(256),"\
                " OpeningHoursNote VARCHAR(400),"\
                " Accessible BOOL,"\
                " Parking BOOL,"\
                " ParkingAccessible B00L,"\
                " ParkingNote VARCHAR(400),"\
                " PRIMARY KEY (FacilityID),"\
                " CONSTRAINT access_fk FOREIGN KEY(FacilityID) "\
                " REFERENCES toilets(FacilityID));"
    cur.execute(create_stmt)
# Create disposal table
def create_disposal_table(db_conn):
    cur = db_conn.cursor()
    create stmt = "CREATE TABLE disposal(" \
                " FacilityID INT," \
                " SharpsDisposal BOOL,"\
                " SanitaryDisposal BOOL."\
                " MensPadDisposal B00L,"\
                " PRIMARY KEY (FacilityID),"\
                " CONSTRAINT disposal fk FOREIGN KEY(FacilityID) "\
                " REFERENCES toilets(FacilityID));"
    cur.execute(create_stmt)
# Create dump_points table
def create dump points table(db conn):
    cur = db conn.cursor()
    create_stmt = "CREATE TABLE dump_points(" \
                " FacilityID INT," \
" DPWashout BOOL,"\
                " DPAfterHours BOOL,"\
                " DumpPointNote VARCHAR(400),"\
                " PRIMARY KEY (FacilityID),"\
                " CONSTRAINT dump_points_fk FOREIGN KEY(FacilityID) "\
```

```
" REFERENCES toilets(FacilityID));"
    cur.execute(create_stmt)
# Create facility types table
def create_facility_types_table(db_conn):
    cur = db_conn.cursor()
    create stmt = "CREATE TABLE facility types(" \
                " TypeID INT," \
                " Name VARCHAR(128),"\
                " PRIMARY KEY (TypeID));"
    cur.execute(create_stmt)
# Create facility_rel table
def create_facility_rel(db_conn):
    cur = db conn.cursor()
    create_stmt = "CREATE TABLE facility_rel (" \
                     FacilityID INT," \
                            INT," \
                     TypeID
                     PRIMARY KEY (FacilityID, TypeID)," \
                     CONSTRAINT facility_rel_fk FOREIGN KEY(TypeID) "
\
                     REFERENCES facility_types(TypeID));"
    cur.execute(create_stmt)
# Create locations table
def create_locations(db_conn):
    cur = db conn.cursor()
    create_stmt = "CREATE TABLE locations (" \
                     LocID INT," \
                               VARCHAR(256)," \
                     Address1
                                FLOAT," \
                     Latitude
                     Longitude FLOAT," \
                     AddressNote
                                   VARCHAR(400)," \
                     PRIMARY KEY (LocID));"
    cur.execute(create_stmt)
# Create location rel table
def create location rel(db conn):
    cur = db conn.cursor()
    create stmt = "CREATE TABLE location rel (" \
                     FacilityID INT," \
                     LocID INT," \
                     PRIMARY KEY (FacilityID, LocID)," \
                     CONSTRAINT fk_loc_id FOREIGN KEY(LocID) " \
                     REFERENCES locations(LocID));"
    cur.execute(create_stmt)
```

```
# Create states table
def create_states(db conn):
    cur = db conn.cursor()
    create stmt = "CREATE TABLE states (" \
                     StateID INT," \
                     State VARCHAR(16)," \
                  " PRIMARY KEY (StateID));"
    cur.execute(create_stmt)
# Create state_rel table
def create_state_rel(db_conn):
    cur = db_conn.cursor()
    create_stmt = "CREATE TABLE state_rel (" \
                    LocID INT," \
                     StateID INT," \
                     PRIMARY KEY (LocID, StateID)," \
                     CONSTRAINT fk_loc_id FOREIGN KEY(LocID) " \
                     REFERENCES locations(LocID));"
    cur.execute(create_stmt)
# Create towns table
def create_towns(db_conn):
    cur = db_conn.cursor()
    create_stmt = "CREATE TABLE towns (" \
                    TownID INT," \
                     Town VARCHAR(128)," \
                     PRIMARY KEY (TownID));"
    cur.execute(create_stmt)
# Create town_rel table
def create town rel(db conn):
    cur = db_conn.cursor()
    create_stmt = "CREATE TABLE town_rel (" \
                    LocID INT," \
                     TownID INT," \
                     PRIMARY KEY (LocID, TownID)," \
                     CONSTRAINT fk loc id FOREIGN KEY(LocID) " \
                     REFERENCES locations(LocID));"
    cur.execute(create_stmt)
```

QUERIES AND RESULTS

--1. How many toilets in Cooma have parking?
SELECT COUNT(*) AS "# Toilets in Cooma with Parking"
FROM location_rel JOIN locations USING(locID) JOIN town_rel
USING(LocID)
JOIN towns USING(TownID) JOIN access USING(facilityID)
WHERE Town='Cooma' AND parking=True;

Toilets in Cooma with Parking

4

(1 row)

--2. How many public toilets are there in Australia?
SELECT COUNT(*) AS "# Toilets Down Under" FROM toilets;

(1 row)

--3. What are all the sporting facility toilets that are also dump
points? // REWORD: How many sporting facility toilets are also dump
points?
SELECT COUNT(*) AS "# Sporting Facilities That Double As Dumps"
FROM facility_rel JOIN facility_types FT USING(typeid)
JOIN dump points USING (facilityID)

WHERE FT.name='Sporting facility';

Sporting Facilities That Double As Dumps

1872

(1 row)

--4. Which city has the most public toilets?
SELECT town, COUNT(facilityID) AS "Toilet Count" FROM
location_rel JOIN locations USING(locID) JOIN town_rel USING(locID)
JOIN towns T USING(townID)
GROUP BY town ORDER BY town:

town	Toilet Count
Melbourne	 50
Sydney	49
Adelaide	j 45
Dubbo	j 45
Coffs Harbour	j 40
Hamilton	j 36
Manly	j 35

Muswellbrook 34 32 0range 31 Mildura (6326 rows) --5. What percentage of toilets are open 24 hours a day? SELECT 100*(SELECT COUNT(*) FROM toilets JOIN access USING(facilityID) WHERE openinghours='OPEN: 24 hours')/COUNT(*) AS "% Austrailian Toilets Open 24 Hours a Day" FROM toilets; % Austrailian Toilets Open 24 Hours a Day 46 (1 row) --6. How many toilets in WA require payment? SELECT COUNT(*) AS "# Toilets in WA That Require Money" FROM access JOIN location_rel USING(facilityID) JOIN locations USING(locID) JOIN state rel USING(locID) JOIN states USING(stateID) WHERE state='WA' AND paymentrequired=TRUE; # Toilets in WA That Require Money 9 (1 row) --7. How many unisex toilets are in a park or reserve? SELECT count(*) AS "# Unisex Toilets in Parks and Reserves" FROM toilets JOIN facility_rel USING(FacilityID) JOIN facility_types T USING(typeID) WHERE T.name='Park or reserve' AND unisex=True: # Unisex Toilets in Parks and Reserves 2077 (1 row) --8. How many toilets are on 1 Bay Street in Glebe? SELECT COUNT(*) AS "# Toilets on 1 Bay Street in Glebe" FROM location_rel JOIN locations USING(locID) JOIN town_rel USING(locID) JOIN towns USING(townid) WHERE town='Glebe' AND address1='1 Bay Street'; # Toilets on 1 Bay Street in Glebe (1 row)

--9. What percentage of public toilets are free?
SELECT 100*(SELECT COUNT(*) FROM access WHERE paymentrequired=False)/
COUNT(*) AS "Percentage of the Free" FROM toilets;

Percentage of the Free

99

(1 row)

--10. What percentage of public toilets with baby changing stations
are free?
SELECT 100*(SELECT COUNT(*) FROM access JOIN changing
USING(facilityID) WHERE paymentrequired=False and babychange=True)/
COUNT(*) AS "Percentage of the Free... With Babies"
FROM toilets JOIN changing USING(facilityID) WHERE babychange=True;

Percentage of the Free... With Babies

99

(1 row)

--11. What percentage of men's toilets contain baby changing stations?
SELECT 100*(SELECT COUNT(*) FROM toilets JOIN changing
USING(facilityID) WHERE male=True AND babychange=True)/COUNT(*) AS "%
Men's Restrooms with Baby Changing Stations"
FROM toilets JOIN changing USING (facilityID) WHERE male=True AND
babychange=False;

% Men's Restrooms with Baby Changing Stations

15

(1 row)

--11.5 What percentage of women's toilets contain baby changing
stations?
SELECT 100*(SELECT COUNT(*) FROM toilets JOIN changing
USING(facilityID) WHERE female=True AND babychange=True)/COUNT(*) AS
"% Women's Restrooms with Baby Changing Stations"
FROM toilets JOIN changing USING (facilityID) WHERE female=True AND
babychange=False;

% Women's Restrooms with Baby Changing Stations

15

(1 row)

--11.75 What percentage of unisex toilets contain baby changing
stations?
SELECT 100*(SELECT COUNT(*) FROM toilets JOIN changing
USING(facilityID) WHERE unisex=True AND babychange=True)/COUNT(*) AS

"% Unisex Restrooms with Baby Changing Stations"
FROM toilets JOIN changing USING (facilityID) WHERE unisex=True AND babychange=False;

% Unisex Restrooms with Baby Changing Stations

26

(1 row)

--12. How many carpark dump points are there?
SELECT COUNT(*) as "Number of Carpark dumpoints"
FROM toilets t join dump_points using(facilityid) join facility_rel
using(facilityid) join facility_types ft using(typeid)
WHERE ft.name = 'Car park';

Number of Carpark dumpoints

66

664

(1 row)

--13. What is the average number of toilets per city within each state?

SELECT state, AVG(c.count) as "Average Number of toilets per city" FROM (

SELECT town, COUNT(facilityid)

FROM toilets natural join location_rel natural join town_rel natural join towns

GROUP BY (town)

) as c join towns using(town) join town_rel using(townid) join location_rel using(locid) join state_rel using(locid) join states using(stateid) GROUP BY(state);

state	Average Number of toilets per city
VIC SA ACT QLD WA NSW	+
TAS NT	6.207920792079 5.4755244755244755
to rows)

--14. Which public toilets with showers also have a fee in New South Wales?

SELECT facilityid, name, state

FROM toilets natural join location_rel natural join state_rel natural join states natural join access

facilityid	name	state
3188	Tourist & Travellers Centre	NSW
10963	Hyde Park – North 1	NSW
33491	Deniliquin Dump Point	NSW
33559	Tenterfield Showground Dump Point	NSW
33567	Wagga Wagga Showgrounds	NSW
50335	St Ives Shopping Village 1	NSW
55813	Stocko	NSW
56023	Packsaddle Roadhouse	NSW
56508	Glendora Campground	NSW
56810	Brackens Hut	NSW
57531	Bombala Caravan Park Dump Point	NSW
57538	Cessnock Showground Dump Point	NSW
57567	Gundagai Cabins & Tourist Park	NSW
57583	Kyogle Showground Dump Point	NSW
57586	Lismore Showgrounds Dump Point	NSW
57597	Mittagong Caravan Park Dump Point	NSW
57645	Camp Blackman Dump Point	NSW
59195	Clemton Park Village Shopping Centre	NSW
(18 rows)		

--15. How many toilets with sharp disposals are in every state?
SELECT state, COUNT(*) as "Toilets with sharp disposal"
FROM toilets natural join disposal natural join state_rel join states using(stateid)
WHERE sharpsdisposal = true
GROUP BY(state);

state	Toilets	with	sharp	disposal
VIC SA ACT QLD WA NSW TAS NT	+			18980684 6987726 750031 15494359 9573260 26620447 3426021 1077934
(8 rows))			

--16. How many toilets that are accessible have parking?
SELECT COUNT(*) as "Number of accessible toilets with parking"
FROM toilets natural join access
WHERE accessible = true AND parking = true;

Number of accessible toilets with parking

(1 row)

--17. Which toilets in VIC are ambulant?
SELECT facilityid, name, state
FROM toilets natural join location_rel natural join state_rel natural
join states natural join handicap
WHERE ambulant = true AND state = 'VIC';

facilityid	name	state
253 270 548 557 562 563 572 1087 1239	Lloyd Street Lloyd Street Great Western Mortlake Tea Tree Lake Woorndoo Dudley W Cornell Park May Park Horsham Town Hall Central Park Centenary Park	VIC VIC VIC VIC VIC VIC VIC VIC
1338	Princes Park 1	ATC

(210 rows)

--18. How many restrooms have an accessible toilet?
SELECT COUNT(*) as "Numbe of toilets that are accessible"
FROM toilets natural join access
WHERE accessible = true;

Numbe of toilets that are accessible

11779

(1 row)

--19. How many toilets have drinking water and showers?
SELECT COUNT(*) as "Number of toilets with water fountain and showers"
FROM toilets
WHERE drinkingwater = true AND shower = true;

400

(1 row)

--20. How many toilets require the master locksmith's access key (MLAK) to enter?

SELECT COUNT(mlak24) as "Toilets requiring master locksmith's access key"

FROM toilets join access using(facilityid)
WHERE mlak24 = false;

Toilets requiring master locksmith's access key 21639

(1 row)

TABLE CONTENTS

List of relations Schema Name Type Owner						
	Name 					
spr2022bdb58	access	table	spr2022bdb58			
spr2022bdb58	changing	table	spr2022bdb58			
spr2022bdb58	disposal	table	spr2022bdb58			
spr2022bdb58	dump_points	table	spr2022bdb58			
spr2022bdb58	facility_rel	table	spr2022bdb58			
spr2022bdb58	facility_types	table	spr2022bdb58			
spr2022bdb58	handicap	table	spr2022bdb58			
spr2022bdb58	location_rel	table	spr2022bdb58			
spr2022bdb58	locations	table	spr2022bdb58			
spr2022bdb58	state_rel	table	spr2022bdb58			
spr2022bdb58	states	table	spr2022bdb58			
spr2022bdb58	toilets	table	spr2022bdb58			
spr2022bdb58	town_rel	table	spr2022bdb58			
spr2022bdb58	towns	table	spr2022bdb58			

	Table "spr2022bdb58	.access"	
Column Default	Type	Collation	Nullable
+		+	+
facilityid keyrequired accessnote paymentrequired mlak24 mlakafterhours openinghours openinghoursnote accessible parking parkingaccessible	integer boolean character varying(400) boolean boolean character varying(256) character varying(400) boolean boolean	 	not null
_parkingnote	character varying(400)		

Indexes:

[&]quot;access_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:
 "access_fk" FOREIGN KEY (facilityid) REFERENCES
toilets(facilityid)

Column Default	Table "spr2022bdb58 Type		Nullable
+	+	+	+
facilityid	integer		not null
babychange babycareroom	integer boolean boolean	 	
babychangenote	character varying(400)	<u> </u>	
acshower	boolean		
adultchange	boolean		
adultchangenote	character varying(400)		
changingplaces	boolean	l	1

Indexes:

"changing_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:
 "changing_fk" FOREIGN KEY (facilityid) REFERENCES
toilets(facilityid)

Column	•	22bdb58.dispo Collation		Default
facilityid sharpsdisposal sanitarydisposal menspaddisposal	integer boolean boolean boolean		not null	

"disposal_pkey" PRIMARY KEY, btree (facilityid)
Foreign-key constraints:

"disposal_fk" FOREIGN KEY (facilityid) REFERENCES
toilets(facilityid)

Column	Table "spr2	022bdb58.dump_points"	Nullable
Default	Type	Collation	
facilityid dpwashout dpafterhours dumppointnote Indexes:	integer boolean boolean character vary	 ing(400)	not null

"dump_points_pkey" PRIMARY KEY, btree (facilityid)
Foreign-key constraints:

"dump_points_fk" FOREIGN KEY (facilityid) REFERENCES
toilets(facilityid)

Table "spr2022bdb58.facility_rel"					
Column	Type	Collation	Nullable	Default	
		+	+	- 	
facilityid	integer		not null		
typeid	integer		not null		
Indovect	-	-	•		

"facility_rel_pkey" PRIMARY KEY, btree (facilityid, typeid) Foreign-key constraints:

"facility_rel_fk" FOREIGN KEY (typeid) REFERENCES
facility_types(typeid)

Column	Table "spr2022bdb58 Type	3.facility_ty Collation	' !	Default
	integer character varying(128)	!	not null	

"facility_types_pkey" PRIMARY KEY, btree (typeid)
Referenced by:

TABLE "facility_rel" CONSTRAINT "facility_rel_fk" FOREIGN KEY (typeid) REFERENCES facility_types(typeid)

Table "spr2022bdb58.handicap"					
Column	Type	Collation	Nullable	Default	
	+	+	+	+	
facilityid	integer		not null]	
byosling	boolean	ĺ	İ	ĺ	
ambulant	boolean	ĺ	İ	İ	
lhtransfer	boolean	ĺ	İ	İ	
rhtransfer	boolean	j	İ	İ	
Indovoc.	•	•	•	•	

"handicap_pkey" PRIMARY KEY, btree (facilityid)
Foreign-key constraints:

"handicap_fk" FOREIGN KEY (facilityid) REFERENCES
toilets(facilityid)

Table "spr2022bdb58.location_rel"					
Column	Type	Collation	Nullable	Default	
	+	+	+	+	
facilityid	integer		not null		
locid	integer		not null		

```
Indexes:
    "location rel pkey" PRIMARY KEY, btree (facilityid, locid)
Foreign-key constraints:
    "fk_loc_id" FOREIGN KEY (locid) REFERENCES locations(locid)
                    Table "spr2022bdb58.locations"
   Column
                    Type | Collation | Nullable | Default
locid | integer
address1 | character varying(256) |
latitude | double precision |
longitude | double precision |
                                                   I not null I
 addressnote | character varying(400) |
Indexes:
    "locations_pkey" PRIMARY KEY, btree (locid)
Referenced by:
    TABLE "location_rel" CONSTRAINT "fk_loc_id" FOREIGN KEY (locid)
REFERENCES locations(locid)
    TABLE "state_rel" CONSTRAINT "fk_loc_id" FOREIGN KEY (locid)
REFERENCES locations(locid)
   TABLE "town_rel" CONSTRAINT "fk_loc_id" FOREIGN KEY (locid)
REFERENCES locations(locid)
           Table "spr2022bdb58.state_rel"
 Column | Type | Collation | Nullable | Default
 Indexes:
   "state_rel_pkey" PRIMARY KEY, btree (locid, stateid)
Foreign-key constraints:
    "fk loc id" FOREIGN KEY (locid) REFERENCES locations(locid)
                   Table "spr2022bdb58.states"
                   Type | Collation | Nullable | Default
 Column |
 stateid | integer
                                             | not null |
 state | character varying(16) |
Indexes:
```

"states_pkey" PRIMARY KEY, btree (stateid)

+			
facilityid	integer		not null
url	character varying(256)	İ	i i
name	character varying(128)	İ	i i
male	boolean	İ	i i
female	boolean	İ	i i
unisex	boolean	İ	i i
allgender	boolean	İ	İ İ
toiletnote	character varying(1024)	İ	İ İ
drinkingwater	boolean	İ	İ İ
shower	boolean		
<pre>Indexes:</pre>			
"toilets_pk	ey" PRIMARY KEY , btree (fa	cilityid)	
Referenced by:			
TABLE "acces	ss" CONSTRAINT "access_fk"	FOREIGN KEY	(facilityid)
REFERENCES toil	ets(facilityid)		
TABLE "chand	ging" CONSTRAINT "changing	fk" FOREIGN	KEY (facility

TABLE "changing" CONSTRAINT "changing_fk" FOREIGN KEY (facilityid)
REFERENCES toilets(facilityid)

TABLE "disposal" CONSTRAINT "disposal_fk" FOREIGN KEY (facilityid) REFERENCES toilets(facilityid)

TABLE "dump_points" CONSTRAINT "dump_points_fk" FOREIGN KEY (facilityid) REFERENCES toilets(facilityid)

TABLE "handicap" CONSTRAINT "handicap_fk" FOREIGN KEY (facilityid) REFERENCES toilets(facilityid)

	Type	spr2022bdb58. Collation +	Nullable	•
locid	integer integer	Ī	not null not null	

"town_rel_pkey" PRIMARY KEY, btree (locid, townid)

Foreign-key constraints:

"fk_loc_id" FOREIGN KEY (locid) REFERENCES locations(locid)

Table	"spr202	22b	db	58.	t	owns	3"
			_			_	_

Column		Collation	Nullable	Default
	integer character varying(128)		not null 	

Indexes:

"towns_pkey" PRIMARY KEY, btree (townid)