Australia, Land of Toilets: The Final Write-Up

By Pete Wells, Stuart Rimel

6/2/2022, CS-586: Intro to Database Management Systems

ח	F	C	\sim 1	5.	П	D.	ГΤ	<u> </u>	۱٨	L	n	E	E.	ТΙ	ĺ
U	Е	2	LI	٦.	L	Ρ	IΙ	.U	ı۱۱		U	'F		П	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 throughout 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 it 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 it 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 entire process of creating this database and inserting all data into was successfully automated with this ETL application. If we had more time, we would potentially re-organize how some of our data is normalized to prevent excessive joins and 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

CREATE TABLE STATEMENTS

```
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 BOOL,"\
                " 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 BOOL,"\
                " MLAKAfterHours BOOL,"\
                " OpeningHours VARCHAR(256),"\
                " OpeningHoursNote VARCHAR(400),"\
                " Accessible BOOL,"\
                " Parking BOOL,"\
                " ParkingAccessible BOOL,"\
                " 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 BOOL,"\
                " 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," \
                    Address1 VARCHAR(256)," \
                     Latitude FLOAT," \
                    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
______
(1 row)
--2. How many public toilets are there in Australia?
SELECT COUNT(*) AS "# Toilets Down Under" FROM toilets;
# Toilets Down Under
               22031
(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';
```


--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	45
Dubbo	45
Coffs Harbour	40
Hamilton	36
Manly	35
Muswellbrook	34
Orange	32
Mildura	31

(6326 rows)

For whole list see q4.txt

--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

```
(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
                                 4
(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
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 664 (1 row)

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

state	Average Number of toilets per city
VIC SA ACT QLD WA NSW	9.0436767917411157 9.0436767917411157 6.7060409924487594 7.1206030150753769 6.7565006075334143 6.8379866299646087 9.4648702031602709
TAS NT (8 rows)	6.2079207920792079 5.4755244755244755

--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 WHERE paymentrequired = true AND state = 'NSW';

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	
+				10000604	
VIC				18980684	
SA				6987726	
ACT				750031	
QLD				15494359	
WA				9573260	
NSW				26620447	
TAS				3426021	
NT				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
-----7673

(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
	Lloyd Street	VIC
270	Great Western	VIC
548	Mortlake Tea Tree Lake	VIC
557	Woorndoo	VIC

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; Number of toilets with water fountain and showers						
to 						
y"						

List of relations

Schema	Name	Type	Owner
spr2022bdb58 spr2022bdb58 spr2022bdb58	 access changing disposal	table table table	spr2022bdb58 spr2022bdb58 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"
T	ype	Collati

Column	Type	Collation Nullable Default
-	+	+
facilityid	integer	not null
keyrequired	boolean	
accessnote	character varying(400)	
paymentrequired	boolean	
mlak24	boolean	
mlakafterhours	boolean	
openinghours	character varying(256)	
openinghoursnote	character varying(400)	
accessible	boolean	
parking	boolean	
parkingaccessible	boolean	
parkingnote	character varying(400)	

Indexes:

"access_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:

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

- 17	II 20221 II FO	
IDNID	"spr2022bdb58	cnanging"
IUDIC	301 202204030	·CHalletile

Column	Type	Collation		l
facilityid	,		not null	

babychange	boolean		
babycareroom	boolean		
babychangenote	character varying(400)		
acshower	boolean		
adultchange	boolean		
adultchangenote	character varying(400)		
changingplaces	boolean	1	

Indexes:

"changing_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:

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

Table "spr2022bdb58.disposal"

Column	Type	Collation	Nullable	Default
facilityid sharpsdisposal sanitarydisposal menspaddisposal	integer boolean boolean boolean		not null	

Indexes:

"disposal_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:

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

Table "spr2022bdb58.dump points"

Column	Type	Collation	Nullable	Default
facilityid dpwashout dpafterhours dumppointnote	integer boolean boolean character varying(400)		not null 	

Indexes:

"dump_points_pkey" PRIMARY KEY, btree (facilityid)

Foreign-key constraints:

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

labie sprz	2022bab58.faci	iiity_rei	
Type	Collation	Nullable	Default
+	+	+	

	F			r
facilityid	integer		not null	١
typeid	integer		not null	

Indexes:

Column

"facility_rel_pkey" PRIMARY KEY, btree (facilityid, typeid) Foreign-key constraints: "facility rel fk" FOREIGN KEY (typeid) REFERENCES facility types(typeid) Table "spr2022bdb58.facility types" Type | Collation | Nullable | Default Column | typeid | integer | not null | name | character varying(128) | Indexes: "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" | Type | Collation | Nullable | Default Column -----| not null | facilityid | integer | byosling | boolean | ambulant | boolean | lhtransfer | boolean | rhtransfer | boolean | Indexes: "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	•		
locid	•	† 			

address1 character varying(256)
Table "spr2022bdb58.state_rel" Column Type Collation Nullable Default
locid integer not null stateid integer not null Indexes: "state_rel_pkey" PRIMARY KEY, btree (locid, stateid) Foreign-key constraints: "fk_loc_id" FOREIGN KEY (locid) REFERENCES locations(locid)

	Table "spr2022			
Column	Туре	Collation	Nullable	Default
		+		
stateid			not null	
state	character varying(16)			

Indexes:

"states_pkey" PRIMARY KEY, btree (stateid)

Table "spr2022bdb58.toilets"					
Column	Type	Collation	Nullable	Default	
facilityid url name male	+ integer character varying(256) character varying(128) boolean	+ 	+ not null 	 	
female unisex allgender toiletnote	boolean boolean boolean character varying(1024)	 	 		

drinkingwater boo					
Indexes:	RIMARY KEY, btree (facilityid`			
Referenced by:	KIMAKI KEI, DUTEE (iaciiicyiu,	,		
-	ONSTRAINT "access_f	k" FORFTGN	KFY (faci	litvid)	
REFERENCES toilets(fa	-	K TONEIGN	KET (Tuess	i i cy i a y	
•	CONSTRAINT "changi	ng fk" FORE	IGN KEY (facilityid)	
REFERENCES toilets(fa		0_	•	, ,	
TABLE "disposal"	CONSTRAINT "dispos	al_fk" FORE	EIGN KEY (facilityid)	
REFERENCES toilets(f	- · · · · · · · · · · · · · · · · · · ·				
	ts" CONSTRAINT "dum	p_points_f	c" FOREIGN	KEY (facilityio	d)
REFERENCES toilets(fa	- •	CL II . 500.			
-	CONSTRAINT "handic	ар_+к" гокы	EIGN KEY (1	racilityid)	
REFERENCES toilets(f	acility10)				
Table "sp	r2022bdb58.town rel	п			
Column Type	_		ī		
+-		+			
locid integer townid integer	not nul	.1			
	not nul	.1			
Indexes:					
	PRIMARY KEY, btree	(locid, tow	vnid)		
Foreign-key constrain		EDENCEC 1-		٠. ١	
TK_10C_10 FURE	IGN KEY (locid) REF	EKENCES 100	cations (100	210)	
	Table "spr2022bdb58	.towns"			

Table	"spr2022bdb58.towns"
-	1 6 11 1

Column	Type •	Collation		•
townid	integer character varying(128)		not null	

Indexes:
 "towns_pkey" PRIMARY KEY, btree (townid)