Amruta Niteen Ranade (ANR200001) Individual Project Part II

Dallas Animal Shelter Data Analysis

```
Amruta Niteen Ranade (ANR200001) Individual Project Part II
Dallas Animal Shelter Data Analysis
Requirements
Data clean-up and columns considerations
Normalization
2. SQLs to create and populate tables
3. Loading of Intake & Outcome Tables
Intake Table-
Outcome-
Report 1 - Breeds survival report
Report1-snapshot
Report-2- Intake Type Versus Outcome
Report2-snapshot
List of normalized tables and counts of records per table
```

Requirements

The project purpose is to practice following skills obtained in the class:

- 1. Analyze raw data and design normalized ERD
- 2. Design and create physical tables
- 3. Load data in proper format to the target tables
- 4. Analyze data and write reports using SQL (no other tools should be used to manipulate data)

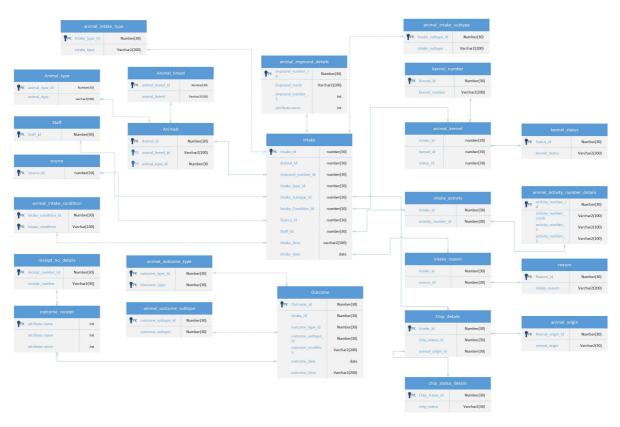
Data clean-up and columns considerations

```
/*Query to remove redundant data*/
Delete
FROM intake a
WHERE a.ROWID IN
(SELECT ROWID FROM
 (SELECT
  ROWID,
  ROW_NUMBER() OVER
    (PARTITION BY animal_id ORDER BY ROWID, intake_date, intake_time desc)
FROM intake)
WHERE dup > 1);
/*Splitting of Activity & Impound Number*/
UPDATE PROJECT2_DATA_temp
ACTIVITY_NUMBER_2 = regexp_substr(ACTIVITY_NUMBER, '[\land -][\land -]^*, 1, 1),
IMPOUND_NUMBER_1 = regexp_substr(IMPOUND_NUMBER, '^[^-]*[^-]', 1, 1),
IMPOUND_NUMBER_2 = regexp\_substr(IMPOUND_NUMBER, '[^ -][^-]*$', 1, 1);
```

```
/*Eliminating redundant columns */
alter table PROJECT2_DATA_temp drop column TAG_TYPE;
alter table PROJECT2_DATA_temp drop column ACTIVITY_SEQUENCE;
alter table PROJECT2_DATA_temp drop column INTAKE_TOTAL;
alter table PROJECT2_DATA_temp drop column SERVICE_REQUEST_NUMBER;
alter table PROJECT2_DATA_temp drop column DATE_MONTH;
alter table PROJECT2_DATA_temp drop column DATE_YEAR;
```

Normalization

1. **ERD**



2. SQLs to create and populate tables

```
/*Creating multiple tables to normalize data*/

/*Kennel_number*/
create table kennel_number(
kennel_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by 1),
kennel_number varchar2(200),
CONSTRAINT kennel_id_pk PRIMARY KEY(kennel_id));

insert into kennel_number(kennel_number)
select distinct(kennel_number) from project2_data_temp;

commit;

/*Kennel_status*/
create table kennel_status(
status_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by 1),
kennel_status varchar2(200),
CONSTRAINT kennel_status_pk PRIMARY KEY (status_id));
```

```
insert into kennel_status(kennel_status)
select distinct(kennel_status) from project2_data_temp;
commit;
/*Impound Number*/
create table animal_impound_details(
impound_number_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT
by 1),
impound_comb varchar2(100),
impound_number_1 varchar2(100),
impound_number_2 varchar2(100),
CONSTRAINT animal_impound_details_pk PRIMARY KEY (impound_number_id));
insert into animal_impound_details
(impound_comb,impound_number_1,impound_number_2)
(select distinct impound_number, impound_number_1, impound_number_2 from
project2_data_temp);
commit;
/*Activity_Number*/
create table animal_activity_number_details(
activity_number_id number(30) GENERATED ALWAYS as IDENTITY(START with 1
INCREMENT by 1),
activity_number_comb varchar2(100),
activity_number_1 varchar2(100),
activity_number_2 varchar2(100),
animal_activity_number_details_pk PRIMARY KEY (activity_number_id));
insert into animal_activity_number_details
(activity_number_comb,activity_number_1,activity_number_2)
(select activity_number, activity_number_1, activity_number_2 from
project2_data_temp);
commit;
/*Intake_activity*/
create table intake_activity (
intake_id number(30) REFERENCES Intake(intake_id),
activity_number_id number(30) REFERENCES
animal_activity_number_details(activity_number_id)
);
Insert into Intake_Activity (Intake_Id,activity_number_id)
select i.Intake_Id, a.activity_number_id
from project2_data_temp d
JOIN Intake i ON i.animal_id = d.animal_id
JOIN animal_activity_number_details a ON a.activity_number_comb =
d.activity_number;
commit;
/*Animal_kennel*/
create table animal_kennel (
status_id number(30) references kennel_status (status_id)
intake_id number(30) REFERENCES Intake(intake_id),
```

```
kennel_id number(30) REFERENCES kennel_number(kennel_id)
);
INSERT INTO Animal_Kennel(Status_Id,Kennel_Id,intake_id)
(SELECT S.Status_Id,K.Kennel_Id,i.intake_id
FROM project2_data_temp dl
JOIN intake i ON i.animal_id = dl.animal_id
JOIN Kennel_Status S ON S.Kennel_Status = dl.Kennel_Status
JOIN Kennel_Number K ON dl.Kennel_Number = K.Kennel_Number);
commit;
/*Animal_type*/
create table animal_type (
animal_type varchar2(100),
animal_type_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by
CONSTRAINT animal_type_id_pk PRIMARY KEY (animal_type_id)); --6
insert into animal_type(animal_type)
select distinct(animal_type) from project2_data_temp;
commit;
/*Animal_breed*/
create table animal_breed(
animal_breed_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT
by 1),
animal_breed varchar2(100),
CONSTRAINT animal_breed_id_pk PRIMARY KEY (animal_breed_id),
);
insert into animal_breed(animal_breed)
select distinct(animal_breed) from project2_data_temp; --353
commit;
/*Animals*/
create table Animals(
Animal_Id varchar2(100) PRIMARY KEY,
animal_breed_id number(30) REFERENCES animal_Breed(animal_breed_id),
animal_type_id number(30) REFERENCES animal_type (animal_type_id) );
insert into Animals (Animal_Id, animal_breed_id, animal_type_id)
select distinct d.Animal_Id,b.animal_breed_id,a.animal_type_id
from project2_data_temp d, animal_breed b, animal_type a
where b.animal_breed=d.animal_breed
and a.animal_type=d.animal_type;
/*Animal_intake_type*/
create table animal_intake_type(
intake_type_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by
1),
intake_type varchar2(200),
CONSTRAINT ItypeId_PK PRIMARY KEY (intake_type_id));
insert into animal_intake_type (intake_type)
select distinct intake_subtype from project2_data;
```

```
commit;
/*Animal_intake_subtype*/
create table animal_intake_subtype(
intake_subtype_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT
by 1),
intake_subtype varchar2(100),
CONSTRAINT intype_pk PRIMARY KEY (intake_subtype_id)); --51
insert into animal_intake_subtype (intake_subtype)
select distinct intake_subtype from project2_data;
commit;
/*Animal_origin*/
create table animal_origin (
animal_origin varchar2(30),
animal_origin_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT
by 1),
CONSTRAINT animal_origin_id PRIMARY KEY (animal_origin_id));
insert into animal_origin(animal_origin)
select distinct(animal_origin) from project2_data_temp; --10
commit;
/*Chip_status*/
create table chip_status_details (
chip_status varchar2(30),
chip_status_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by
CONSTRAINT chip_status_id_pk PRIMARY KEY (chip_status_id));
insert into chip_status_details(chip_status)
select distinct(chip_status) from project2_data_temp; --6
commit;
/*Chip_details*/
create table chip_details (
intake_id REFERENCES Intake(intake_id),
chip_status_id REFERENCES chip_status_details(chip_status_id),
animal_origin_id REFERENCES animal_origin(animal_origin_id),
);
insert into chip_and_origin (intake_id,chip_status_id,animal_origin_id)
select i.intake_id, c.chip_status_id,o.animal_origin_id from
final_temp1 t join intake i ON i.animal_id = t.animal_id
join Chip_Status c on c.chip_status=t.chip_status
join animal_origin o on o.animal_origin=t.animal_origin;
commit;
/*Reason*/
create table reason(
reason_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by 1),
intake_reason varchar2(200),
CONSTRAINT reason_id_pk PRIMARY KEY (reason_id));
```

```
insert into reason(intake_reason)
select distinct(reason) from project2_data_temp;
commit;
/*Intake_reason*/
create table intake_reason (
intake_id number(30) REFERENCES Intake(intake_id),
reason_id number(30) REFERENCES reason(reason_id)
Insert into Intake_Reason (Intake_Id,Reason_Id)
select i.Intake_Id, r.Reason_Id
from project2_data_temp d
JOIN Intake i ON d.animal_id = i.animal_id
JOIN Reason r ON d.reason = r.intake_reason;
/*Source*/
create table source (
source_id varchar2(200),
CONSTRAINT source_id_pk PRIMARY KEY (source_id)
);
insert into source (source_id)
select distinct source_id from project2_data_temp;
commit;
/*Staff*/
create table staff_details (
staff_id varchar2(30),
CONSTRAINT staff_details_id_pk PRIMARY KEY (staff_id));
insert into staff_details (staff_id)
select distinct staff_id from project2_data_temp;
commit;
/*Animal_intake_condition*/
create table animal_intake_condition(
intake_condition_id number(30) GENERATED ALWAYS as IDENTITY(START with 1
INCREMENT by 1),
intake_condition varchar2(100),
CONSTRAINT intake_pk PRIMARY KEY (intake_condition_id));
insert into animal_intake_condition(intake_condition)
select distinct(intake_condition) from project2_data_temp;
commit:
/*receipt_number*/
create table receipt_no_details (
receipt_number varchar2(30),
receipt_number_id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT
by 1),
CONSTRAINT receipt_number_id_pk PRIMARY KEY (intake_date_time_id));
```

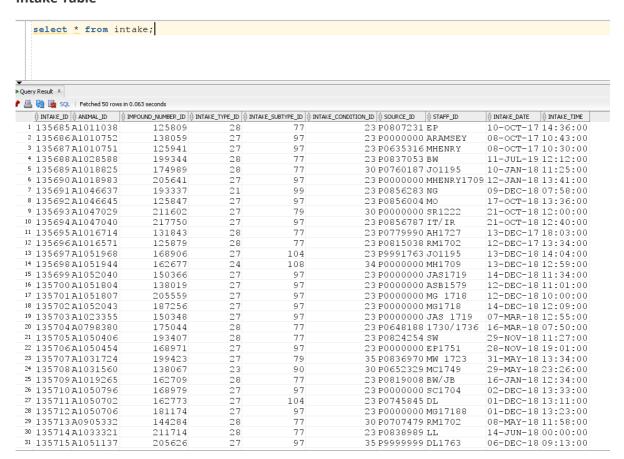
```
insert into receipt_no_details(receipt_number)
select distinct(receipt_number) from project2_data_temp;
/*outcome_receipt*/
create table outcome_receipt (
outcome_id number(30) REFERENCES animal_outcome_type(outcome_type_id),
outcome_type_id number(30) REFERENCES animal_outcome_type(outcome_type_id));
insert into outcome_receipt(receipt_number)
select distinct(receipt_number) from project2_data_temp;
/*Animal_outcome_type*/
CREATE TABLE animal_outcome_type(
outcome_type_id NUMBER(30) GENERATED ALWAYS AS IDENTITY(START WITH 1 INCREMENT
outcome_type VARCHAR2(200),
CONSTRAINT outcome_type_id_pk PRIMARY KEY (outcome_type_id));
INSERT INTO animal_outcome_type(outcome_type)
(SELECT DISTINCT(outcome_type) FROM project2_data_temp);
commit;
/*Animal_outcome_subtype*/
CREATE TABLE animal_outcome_subtype(
outcome_subtype_id NUMBER(38) GENERATED ALWAYS as IDENTITY(START with 1
INCREMENT by 1),
outcome_subtype VARCHAR2(200),
CONSTRAINT outcome_subtypeid_pk PRIMARY KEY (outcome_subtype_id));
INSERT INTO animal_outcome_subtype(outcome_subtype)
SELECT DISTINCT(outcome_subtype) FROM project2_data_temp;
/*Animal_outcome_condition*/
CREATE TABLE animal_outcome_condition(
outcome_condition_id NUMBER(38) GENERATED ALWAYS AS IDENTITY(START WITH 1
INCREMENT BY 1),
outcome_condition VARCHAR2(200),
CONSTRAINT outcome_condition_id_pk PRIMARY KEY (outcome_condition_id));
INSERT INTO animal_outcome_condition(outcome_condition)
SELECT DISTINCT(outcome_condition) FROM project2_data_temp;
commit;
```

3. Loading of Intake & Outcome Tables

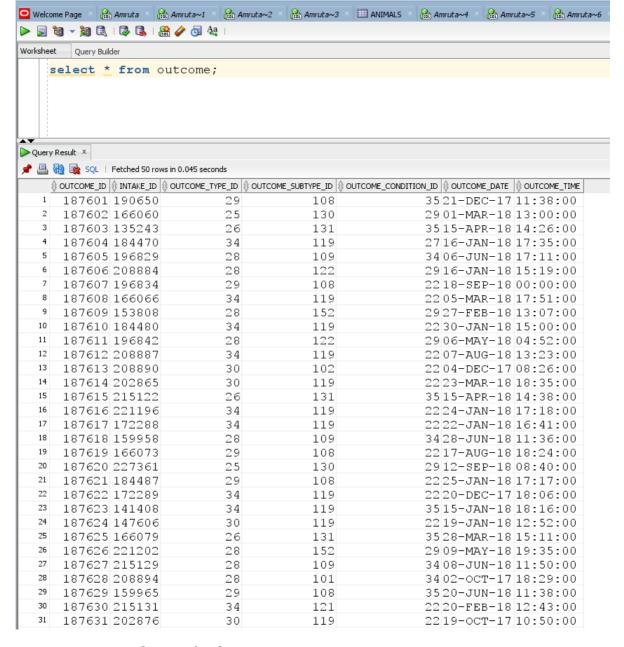
```
/*Intake */
Create table Intake(
Intake_Id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by 1),
Animal_Id varchar2(200) REFERENCES Animals(Animal_Id),
impound_number_id number(30) REFERENCES
animal_impound_details(impound_number_id),
intake_type_id number(30) REFERENCES Animal_Intake_Type(Intake_Type_Id),
```

```
Intake_Subtype_Id number(30) REFERENCES
Animal_Intake_Subtype(Intake_Subtype_Id),
Intake_Condition_Id number(30) REFERENCES
Animal_Intake_Condition(Intake_Condition_Id),
Source_Id varchar2(200) REFERENCES Source(SOURCE_ID),
staff_id varchar2(30) REFERENCES staff_details(staff_id),
Intake_date Date,
CONSTRAINT Intake_table_PK PRIMARY KEY (Intake_Id));
Insert into
Intake(Animal_Id,impound_number_id,Intake_Type_Id,Intake_Subtype_Id,Intake_Condi
tion_Id,Source_Id,Staff_Id,Intake_date,Intake_time)
select a.Animal_Id, ai.impound_number_id, a_t.Intake_Type_Id,
a_s.Intake_Subtype_Id, a_c.Intake_Condition_Id,
s.Source_Id,s1.Staff_Id,to_date(SUBSTR(Intake_date, 1,10)
,'MM/DD/YYYY'),intake_time
from project2_data_temp d
join Animals a on d.Animal_Id = a.Animal_Id
join animal_impound_details ai on d.Impound_number = ai.impound_comb
join Animal_Intake_Type a_t on d.Intake_Type = a_t.Intake_Type
join Animal_Intake_Subtype a_s on d.Intake_Subtype = a_s.Intake_Subtype
join Animal_Intake_Condition a_c on d.Intake_Condition = a_c.Intake_Condition
join Source s on d.Source_Id = s.Source_Id
join staff_details s1 on d.Staff_Id = s1.Staff_Id;
commit;
/*Outcome*/
Create table Outcome(
Outcome_Id number(30) GENERATED ALWAYS as IDENTITY(START with 1 INCREMENT by 1),
Intake_id number(30) REFERENCES Intake(intake_id),
Outcome_Type_Id number(30) REFERENCES Animal_Outcome_Type(Outcome_Type_Id),
Outcome_Subtype_Id number(30) REFERENCES
Animal_Outcome_Subtype(Outcome_Subtype_Id),
Outcome_Condition_Id number(30) REFERENCES
Animal_Outcome_Condition(Outcome_Condition_Id),
Outcome_date Date,
Outcome_time varchar2(100),
CONSTRAINT Outcome_PK PRIMARY KEY (Outcome_Id));
INSERT INTO
Outcome(Intake_Id,Outcome_Type_Id,Outcome_Subtype_Id,Outcome_Condition_Id,Outcom
e_date,Outcome_time)
(SELECT i.Intake_Id,t.Outcome_Type_Id, s.Outcome_Subtype_Id,
c.Outcome_Condition_Id,
to_date(SUBSTR(Outcome_date, 1,10) ,'MM/DD/YYYY'),Outcome_time
from project2_data_load d
JOIN Intake i ON d.animal_id = i.animal_id
JOIN Animal_Outcome_Type t on d.Outcome_Type = t.Outcome_Type
JOIN Animal_Outcome_Subtype s on d.Outcome_Subtype = s.Outcome_Subtype
JOIN Animal_Outcome_Condition c on d.Outcome_Condition = c.Outcome_Condition);
commit;
```

Intake Table-



Outcome-



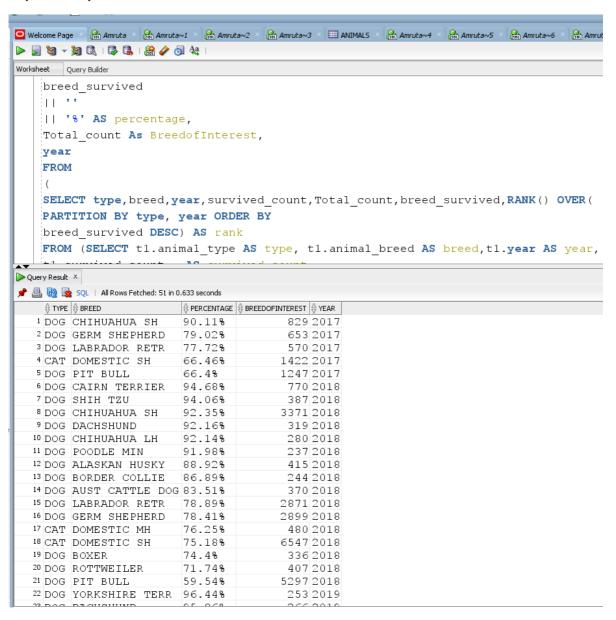
Report 1 - Breeds survival report

```
SELECT
type,
breed.
breed_survived
11 ''
|| '%' AS percentage,
Total_count As BreedofInterest,
year
FROM
SELECT type,breed,year,survived_count,Total_count,breed_survived,RANK() OVER(
PARTITION BY type, year ORDER BY
breed_survived DESC) AS rank
FROM (SELECT t1.animal_type AS type, t1.animal_breed AS breed,t1.year AS year,
t1.survived_count AS survived_count,
t2.total_count
                    as Total_count,
round((t1.survived_count / t2.total_count) * 100, 2) AS breed_survived
FROM (SELECT ab.animal_breed_id,at.animal_type AS animal_type,
ab.animal_breed AS animal_breed,
```

```
COUNT(an.animal_id) AS survived_count,
EXTRACT(YEAR FROM aid.intake_date) AS year
FROM animals an, animal_breed ab, animal_type at, intake aid, outcome aod,
animal_outcome_type aot
WHERE an.animal_id = aid.animal_id
AND an.animal_breed_id = ab.animal_breed_id
AND an.animal_type_id = at.animal_type_id
AND at.animal_type IN (
'CAT',
'DOG'
)
AND to_char(aid.intake_date, 'YYYY') > '2016'
AND aod.intake_id = aid.intake_id
AND aot.outcome_type_id = aod.outcome_type_id
AND aot.outcome_type NOT IN (
'TREATMENT',
'DIED',
'DEAD ON ARRIVAL',
'EUTHANIZED'
GROUP BY
ab.animal_breed_id,
at.animal_type,
ab.animal_breed,
EXTRACT(YEAR FROM aid.intake_date)
HAVING COUNT(an.animal_id) > 200) t1,(
SELECT
ab.animal_breed_id,
at.animal_type AS animal_type,
ab.animal_breed AS animal_breed,
COUNT(an.animal_id) AS total_count,
EXTRACT(YEAR FROM aid.intake_date) AS year
FROM animals an, animal_breed ab, animal_type at,
intake aid, outcome aod,
animal_outcome_type aot
an.animal_id = aid.animal_id
AND an.animal_breed_id = ab.animal_breed_id
AND an.animal_type_id = at.animal_type_id
AND at.animal_type IN (
'CAT',
'DOG'
)
AND to_char(aid.intake_date, 'YYYY') > '2016'
AND aod.intake_id = aid.intake_id
AND aot.outcome_type_id = aod.outcome_type_id
GROUP BY
ab.animal_breed_id,
at.animal_type,
ab.animal_breed,
EXTRACT(YEAR FROM aid.intake_date)
COUNT(an.animal_id) > 200
) t2
WHERE
t1.animal_breed_id = t2.animal_breed_id
AND t1.year = t2.year
ORDER BY t1.year
```

```
ORDER BY
rank,
year)
ORDER BY
year,
breed_survived DESC;
```

Report1-snapshot

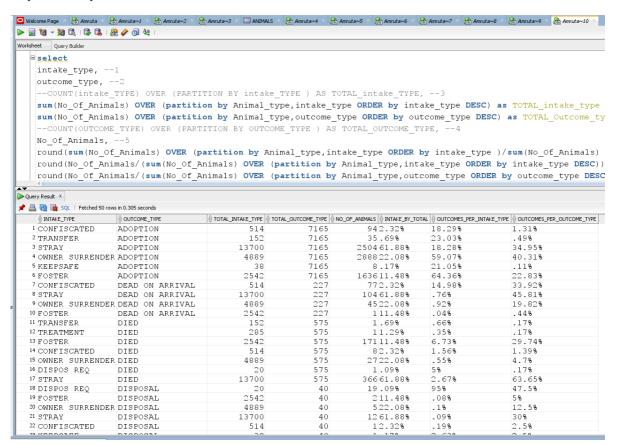


Report-2- Intake Type Versus Outcome

```
select
intake_type, --1
outcome_type, --2
--COUNT(intake_TYPE) OVER (PARTITION BY intake_TYPE ) AS TOTAL_intake_TYPE, --3
sum(No_Of_Animals) OVER (partition by Animal_type,intake_type ORDER by
intake_type DESC) as TOTAL_intake_type , --3
sum(No_Of_Animals) OVER (partition by Animal_type,outcome_type ORDER by
outcome_type DESC) as TOTAL_Outcome_type , --4
--COUNT(OUTCOME_TYPE) OVER (PARTITION BY OUTCOME_TYPE ) AS TOTAL_OUTCOME_TYPE, --4
```

```
No_Of_Animals, --5
round(sum(No_Of_Animals) OVER (partition by Animal_type,intake_type ORDER by
intake_type )/sum(No_Of_Animals) OVER (PARTITION BY ANIMAL_TYPE
)*100,2)||''||'%' as intake_by_total, --6
round(No_Of_Animals/(sum(No_Of_Animals) OVER (partition by
Animal_type,intake_type ORDER by intake_type DESC))*100,2)||''||'%' as
Outcomes_Per_Intake_Type,
round(No_Of_Animals/(sum(No_Of_Animals) OVER (partition by
Animal_type,outcome_type ORDER by outcome_type DESC))*100,2)||''||'%' as
Outcomes_Per_Outcome_Type
from
(
select at.Animal_type, ait.intake_type, aot.Outcome_Type, count(an.animal_id) as
No_Of_Animals
from Animals an, Animal_Breed ab, Animal_Type at, Animal_Intake_Type ait,
intake aid,
Animal_Outcome_Type aot, outcome aod
where an.animal_id = aid.animal_id
and an.animal_breed_id = ab.animal_breed_id
and at.animal_Type_ID=an.animal_Type_ID
and at.Animal_type in ('CAT', 'DOG')
and aod.intake_id = aid.intake_id
and aot.Outcome_Type_Id = aod.Outcome_Type_Id
and ait.Intake_Type_ID = aid.Intake_Type_ID
group by at.Animal_type, ait.intake_type, aot.Outcome_Type
order by at.Animal_type, ait.intake_type, aot.Outcome_Type
);
```

Report2-snapshot



List of normalized tables and counts of records per table

Table	Record Count
Kennel_number	1149
kennel_status	13
animal_impound_details	98156
animal_activity_number_details	42592
intake_activity	57213
animal_kennel	98156
animal_type	6
animal_breed	353
animals	98156
animal_intake_type	9
animal_intake_subtype	51
animal_origin	10
chip_status_details	6
chip_details	93879
reason	79
intake_reason	52920
source	41863
staff_details	2086
animal_intake_condition	17
receipt_no_details	37354
outcome_receipt	37437
animal_outcome_type	14
animal_outcome_subtype	80
animal_outcome_condition	17
Intake	98156
outcome	93156
Report1-query	51
Report2-query	103

Amruta Niteen Ranade

anr200001