

Destination Cleaning

Vivian Zhou, ClipperData, LLC, Data Analyst Intern

I. TASK

There are over 3,000,000 raw destinations have not been cleaned in our database. Similar to Destination translation, I extract data from our database using PostgreSQL, and figured out the right country code and city code, then changed the "isclean" to "T".

II. CODE

```
select count(*) from as_alias_dest -----total number of raw destinations needed
to be cleaned: currently 3,665,878
```

```
-----group by country code-----
select uncountry, count(uncountry) from as_alias_dest where isclean = 'f'
group by 1
order by count(*) desc
```

```
-----Group by and list different location codes under a specific Country code-----
select unlocation, count(*) from as_alias_dest where isclean= 'f' and uncountry =
'FR'
group by unlocation
order by count(*) desc
```

```
----- count numbers of different transformed data under a specific location code,
and order them by a descending number-----
select transformed, count(transformed) from as_alias_dest where isclean= 'f' and
uncountry= 'FR' and unlocation= 'LEH'
group by 1
order by count(transformed) desc
```

```
--Search for the corresponding raw data based the largest number of transformed data-
select * from as_alias_dest where isclean = 'f' and uncountry = 'FR' and
unlocation='LEH' and transformed ilike '%LEH' and raw not ilike '%"%' and raw not
ilike '%"%'
and REVERSE(substring(REVERSE(raw),1,3)) not in (select lo_city_code from
lookup.port_un)
and REVERSE(substring(REVERSE(raw),1,2)) not in (select lo_country_code from
lookup.port_un)
order by length(raw)
```

```
select COUNT(*) from (
select uncountry, count(uncountry) from as_alias_dest where unlocation = 'OGS' group
by uncountry) as a -----search for a specific city code, and check if there
are more than one country have the same city code
```

```
select * from as_alias_dest where isclean = 'f' and unlocation = 'OGS';
```

```
select * from as_alias_dest where isclean = 'f' and uncountry = 'NL'
```

```
update as_alias_dest set uncountr y ='CN',unlocation ='SHA', isclean ='t' where raw  
='XXX';
```