Destination Translation
Vivian Zhou, ClipperData, LLC, Data Analyst Intern

I.   Purpose

Manually check all translations of raw destination data and ensure their translations are correct.

II.   Background

Every day, raw destination data from position signals for each vessel and each voyage is generated. In order to better maintain and use these data in our models, we will translate the raw destinations to machine readable data.

For example, raw destination = 'Manhattan', raw destination = 'New York', raw destination = 'NYC'. We will change all of them to "US NYC" (which is a country + city code combination stored in our database).

After machine translation, isclean = 'f' is set for every record, which means it is waiting for human-eye double check. We will prioritize double checking those destinations which are currently being used in our models. There are around 100 on Monday, and around 30-40 records on other weekdays.

III.   Steps

1. Every morning At 10:50AM, our data will be refreshed. Open the link and manually check all translations in the given table.

✏️	is_Clean	is_Junk	Class	Vessel	IMO	RAW	un_country	un_location	transformed	matched_clean
Edit	No	No	VLCC	87670	9384590	CN FOR ORDER	CN	.	CN	
Edit	No	No	Aframax	563085100	9800738	SG BUKOM	SG	SIN	SG BUKOM	

score	source	country_name	port_name	bargo
		CHINA	CHINA	Bargo link
		SINGAPORE	Singapore	Bargo link

Fig1: Table sample

2. The actual country code and city code can be edited by clicking on blue 'Edit' button, then input the actual 'un_country' and 'un_location' values, check the 'Is Clean' box, the transformation will be saved and the is_clean column will change from red 'No' to green 'Yes'.

un_country:

CN

un_location:

.

☐ Is Clean?

Update   Cancel

Fig2: 'Edit' Page

3. Column "RAW" are the original destinations we need to look at. The default 'un_country' and 'un_location' values are our initial translations. Some of the translations are obviously correct, you can go ahead change is_clean column to "yes". Others will require double check.

4. Some translations have "transformed", "matched_clean" and "score" values. "Transformed" are the cleaned raw destination. "Matched_clean" are the machine-matched raw destination stored in our database. In the initial translating process, we compare the similarity between the transformed and matched_clean value, and use the most similar one's translation as the raw data translation. The higher the "score", the more similar they are. Thus, when the score is high enough, and the "transformed" value is also reasonable, we can consider the initial translations for country and location value most likely correct. However, double check in our database is always necessary.

5. The SQL queries that applied for destination translation are attached below. The codes here mean that, the vessel arrives the destination is declared to the same or similar raw destination in our database. We substitute the 'xxxxxx' to the raw destination that want to

check. Sometimes, we need to reduce the raw destination content to search. For example, '%sodnfebek NYC%' may not be found in our database, but if changed to '%NYC%', we can find it in our database.

Code appended as follow:

```sql
select
    a.poi,
    a.destination_arrive,
    b.lo_country_code,
    b.lo_city_code,
    b.cmdty
from
    asvt_arrival a
left join
    as_poi b on a.poi=b.poi
where
    a. destination_arrive ilike '%XXXXXX%'
order by
    a.date_arrive desc;
```

6. If the raw destination contains two locations, for example, "CN SHA US NYC". This record means the vessel travels from Shanghai, China to New York, US. When looking for the potential destination, we will only take the last location for consideration.

7. When the raw destination cannot be found in our database, either searching it directly or looking for its IMO in google, and check if any information can be found. Google sometimes can automatically correct the spelling mistakes of the raw destination; you can replace the misspelled destination and re-search for it in our database with the SQL queries above. For IMO searching, we mainly look at the "MarineTraffic" website which normally provides reliable information on vessels' voyages.

8. The Bargo link can also be used when more information is needed to verify the destinations. After opening the Bargo map, voyage can be clearly showed. Considering the estimated arrival time of the vessel from IMO search, we can look for the nearby locations if there are any location names are similar to the given raw destination. If so, check the country and city code of the POI, then replace the initial "un_country" and "un_location" value.

9. some raw data are meaningless, such as "for order", "security check", "arm on board" or some irregular letters and so on, which are obvious not real destinations, we can check the "is_junk" box.

10. Sometimes the raw destination will appear as an ocean name, such as "Atlantic", "Pacific", we will use the following code to search for the name, and then look at the rows which has "A1" as "lo_country_code", since all ocean names are listed under "A1". For example, "Pacific Ocean" will be transferred to "A1" and "PACI" as lo_country_code and "lo_city_code"

```
select * from lookup.port_un where name ilike '%Pacific%';
```

Resources

```
select
    a.poi,
    a.destination_arrive,
    b.lo_country_code,
    b.lo_city_code,
    b.cmdty
from
    asvt_arrival a
left join
    as_poi b on a.poi=b.poi
where
    a. destination_arrive ilike '%XXXXXX%'
order by
    a.date_arrive desc;


select * from lookup.port_un where name ilike '%Pacific%';
```

```
select * from as_alias_dest limit 10;          --- destination translation table
 select * from as_dest_x limit 10;             ---junk raw destination table
 Select * from lookup.port_un limit 10;        --- all country+city combinations
 select * from as_poi limit 10;                ---poi table
 select * from country limit 10;               --- country table
```