Possible Transformations

INPUT:  
-Ports Table -- Keanu

-States Table(countries within it) -- Anderson

-Vendors Table -- Keanu

-Shipments Table (products within it) -- Mohit

Updates:

**1 - Approved Ports Table**

**2 - Approved Vendors Table**

**3 - Clean Ports Table (Get rid of coordinates1, add states to foreign)**

**4 - Split the Contact Name**

# Ports Table:

* Filter ports table to show ports only if they provide service to our port (e.g: if Port ID = USBAL, GBTIL, SGSIN, etc; see bottom of document)
* Splitting latitude/longitude diff columns
* Replacing State name with state ID from the state table

**1.) Get rid of 4 extraneous columns at the end (coordinates1, street, street\_name, postal code)**

**2.) Remove duplicate records**

**3.) Split coordinates attribute into individual latitude and longitude columns**

**4.) Perhaps clean up the state column in the foreign table. Maybe just signify as 'abroad' under the state column of foreign ports.**

**5.) Perhaps merge the foreign and domestic data when all is done and clean.**

**6.) Filter each table for only the ports we serve (located at the bottom of our Part 3 document)**

Initial Port Table:

Port ID

Port Code (Initial Date has buncha random ports: SGSIN-Singapore, USBAL-Baltimore, GBTIL- London, USLGB- Seaport Long Beach, sea ports below…)

Latitude/Longitude (written together, **split in transformation**)

State Name (lookup name and replace with id)

Timezone

Bs columns (drop)

# States Table:

* Maybe split state and country into two separate attributes or generate a new countries table

State ID

State Name

State/Country (lookup name and replace with id)

Or

Country (generate country table from it)

Bs columns?(drop)

==>

# Counties Table:

* Maybe split the countries table into a separate table from the States table… perhaps generated on transformation?
* Countries ID is generated in transformations (remove redundancies)

# Vendor Table:

* Merge or split employee names from input csv when importing as contact name
* Filter vendors that only serve the Port of Baltimore (Perhaps condition so that it only shows these vendors: <https://pobdirectory.com/directory/?category=Steamship%20and%20Barge%20Lines%20and%20Agents>)

1. **Split Contact Person Name**
2. **Remove vendors whose names don’t here match directory (above)**
3. **Remove City, State, and Street Columns**

Initial Table:

VendorID

VendorName

Contact Person (first and last name initially, merge or split in transformation, either one will do)

Email

Phone

Bs columns (drop)

Bs columns (drop)

Bs columns (drop)

# Shipments Table:

* Filter to show shipments that occurred only in a certain timeframe
* Filter to show shipments where origin OR destination ID is only Port of Baltimore
* If date departed is less than or greater, etc
* Filter certain products

ShipmentID

Quantity (remove quantity type eg: 50 tons becomes 50) **nice one**

Product description (lookup name and replace with id)

Origin Name (lookup name and replace with id)

Destination Name (lookup name and replace with id)

Vendor Name (lookup name and replace with id)

DateTime Departed

DateTime Arrived

Product Container? (true/false)

Product Description

Product Type

Product Measure (kg, tons, liters, units, boxes…)

Bs columns (drop)

Bs columns (drop)

Bs columns (drop)

Bs columns (this table could have lots of bs tables…)

# Products Table:

* How to generate? (normalize info from shipping?)

ProductID

# Ports we are servicing:

* USBAL - Baltimore, Maryland (that’s us, so obviously)
* GBLGP - Seaport London
* PTSIE - Seaport of Sines, Portugal
* NLRTM - Rotterdam, Netherlands
* USMIA - Seaport Miami, Florida
* USEZA - Seaport Elizabeth, New Jersey
* USSFO - Seaport San Francisco, CA
* CAMTR - Seaport Montreal, Canada
* USATL - Seaport Atlanta, Georgia
* USAER - Seaport Ayer, Massachusetts
* USMSY - Seaport New Orleans, Louisiana
* USNYC - Seaport New York City, NY
* USBPT - Seaport Beaumont, Texas
* USPIT - Seaport Pittsburgh, Pennsylvania
* MXESE - Seaport Ensenada, Mexico
* HNRTB - Seaport of Roatan, Honduras

# Ports we aren’t servicing:

* JPTKY - Port of Tokuyama, Japan
* AEJEA - Port of Jebel Ali, UAE
* ATVIE - Port of Wien, Austria
* AUSYD - Seaport of Sydney, Australia
* PUSAN - Seaport Busan, Korea
* INCCU - Seaport Calcutta, India
* USPDX - Seaport Portland, Oregon
* USSEA - Seaport Seattle, Washington