Port of Baltimore: Project Overview

Project Abstract:

The Port of Baltimore, overseen by the Maryland Port Administration, is one of the biggest commercial shipping hubs in the United States. In 2019, they represented a total of $58 billion in total revenue from imports and exports. Last year, the Port of Baltimore recorded 8.2 billion tons in container cargo going through. This is a good port to study, because our main task involves optimizing the process of container shipping.

Our Goal: Bring the Port of Baltimore into the 21st century through a digital transformation. We aim for this task to serve as a stepping stone to get the entire United States’ shipping systems to utilize the most up-to-date practices.

Table of Contents

[**Datasets:**](#_3tjunrn2vque) **3**

[Shipment Data](#_j0wa1lvfxxfg) 3

[Ports Foreign](#_ikwtin6bm9jb) 3

[Ports Domestic](#_pe5cpqawkypc) 3

[Vendors](#_umb0yssl2222) 3

[Approved Vendors](#_w6vhizyms8jt) 4

[Approved Ports](#_kec9vnkids35) 4

[Time zone UTC offset](#_ooiy4bqg5njd) 4

[**Business Goals:**](#_8qw81d79mkv) **5**

[Answering Pertinent Business Questions Quickly](#_kb4ht9fnrr0i) 5

[Provide Value-Driven Analytics](#_y42tnalx839z) 5

[Public-Private Partnerships](#_u4bfar2hvfg3) 5

[Bringing Legacy Systems to the 21st Century](#_bohdelcw8uc3) 5

[Improving Things for ALL our Stakeholders](#_xu2hkqnzoy55) 5

[Lay the Groundwork for Semi-Autonomous Operations](#_ggwbzc5bc47w) 5

[**Data Modeling:**](#_jche1msw5iie) **6**

[Database Star Schema Design](#_2czamfjppjh3) 6

[ER Schema](#_h3txkg53riw2) 7

[DW Schema](#_p8vzf1o4i4ft) 8

[**ETL Implementation:**](#_w3hait5pncsk) **9**

[The ER Model Load:](#_xr9igt6ypkan) 9

[The Data Warehouse Model Load:](#_j6f5567lnuou) 9

[Takeaways from Implementation:](#_80b5l0fm97y1) 10

[Feedback from Presentation:](#_vx90kule9t9q) 10

[**Tableau Implementation:**](#_vuhau0b2vsp0) **11**

[Descriptive Part 1](#_41jp2lz8amft) 12

[Descriptive Part 2](#_fu9z1qnolx5h) 13

[Predictive Part 1](#_86qeea8m5pil) 14

[Predictive Part 2](#_3wi4ygu8qos9) 15

[General Tableau Takeaways](#_h0fqrill7y3h) 16

# Datasets:

For this project, we used a set of 7 disparate datasets to simulate the scenario of a shipping port attempting to consolidate its historical information and its outdated data architecture into a single Data Warehouse. The 7 dataset files generated for this project are displayed below as follows:

## Shipment Data

This dataset contains all the shipping details of the products. i.e. Date, Time, Origin, Destination. Below is the example of, how Shipment data file looks like:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Product\_Name | Product\_Type | Origin\_Port\_Code | Origin\_Name | Destination\_Port\_Code | Destination\_Name | Date\_Time\_Departed | Date\_Time\_Arrived |
| Jaguar | Automobile/Car | USBAL | Baltimore | GBLGP | London | 2020-02-22 8:30:29 | 2020-06-07 3:07:48 |

## Ports Foreign

This dataset contains the ports based in foreign countries. i.e. port\_code, state, region, timezone. Below is the example of, how Port foreign data file looks like:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Port\_Code | State | Region | Coordinates | Timezone\_ID | Timezone | Street\_Address | Postal\_Code |
| PTSIE | PT | Portugal | 37.93963,-8.85185 | **21** | Europe/Lisbon | 5 Bobwhite Lane | 2525-234 |

## Ports Domestic

This dataset contains the ports based in the domestic country. i.e. port\_code, state, region, timezone. Below is the example of, how Port domestic data file looks like:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Port\_Code | City | State | Coordinates | Timezone\_ID | Street\_Name | Street\_Address | Postal\_Code |
| USPIT | Pittsburgh | PA | 40.44355,-79.9959 | 1 | Hoffman | 8 Clarendon Circle | 19191 |

## Vendors

This dataset contains the vendors associated with the ports shipping the products. i.e. vendor\_name, email, phone, ports. Below is the example of, how vendors data file looks like:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Vendor\_ID | Vendor\_Name | First\_Name | Last\_Name | Email | Phone | City | State | Street |
| 1 | Action One Logistics, Inc. | Belinda | Richardson | eclbee@aol.com | 410-452-5075 | Norfolk | VA | 1 Sommers Junction |

## Approved Vendors

This dataset contains the list of vendors which are verified and approved to ship the products. i.e.Vendor\_name, email. Below is the example of, how Approved vendors data file looks like:

|  |  |
| --- | --- |
| Vendor\_Name | Email |
| Action One Logistics, Inc. | eclbee@aol.com |

## Approved Ports

This dataset contains the list of ports approved to receive and ship the products. i.e. port\_code, port\_code, city, state. Below is the example of, how Approved Ports data file looks like:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Port\_Code | Port\_Name | City | State | Region |
| USBAL | Seaport Baltimore | Baltimore | MD | United States |

## Time zone UTC offset

This dataset contains the timezone of each of the countries/zones listed for ports and shipment. I.e. Timezone, UTC Offset. Below is the example of, how Time zone UTC Offset data file looks like:

|  |  |  |
| --- | --- | --- |
| Timezone\_ID | Timezone | UTC offset |
| 1 | America/Eastern | -5 |

# Business Goals:

## Answering Pertinent Business Questions Quickly

In this project, we aim to create a technological architecture that will make it easier for us to know what’s what in our organization. We want to pose questions regarding our logistics, and receive answers that are reliably backed up in a single repository of information. We want a repository that is updated in real-time, and allows for queries at a moment’s notice.

## Provide Value-Driven Analytics

We want data visualization tools to be a core aspect of our system architecture. We want these tools to be able to provide easily digestible reporting in order to provide insights into the current state of our organization. Furthermore, we want to utilize our vast swaths of legacy data in order to craft forecasts that will allow us to easily discard outdated business practices while adopting new ones.

## Public-Private Partnerships

We strive to utilize the best practices that technology has to offer. We believe we can work directly alongside partners in the private sector to make this transformation possible. Experts in the private sector can bring innovative solutions for public use.

## Bringing Legacy Systems to the 21st Century

We want a consolidation of our vast swaths of historical records. We aim to bring our legacy data into our new and improved architecture. This will serve as a vital part of our historical analysis.

## Improving Things for ALL our Stakeholders

It won’t matter if you are an employee of the Port, a logistics company, or any of the parties involved. When the ‘who’, ‘what’, ‘where’, and ‘why’ are reliably outlined in a Data Warehouse, every party involved will benefit financially from an easily accessible repository of information.

## Lay the Groundwork for Semi-Autonomous Operations

The core goal of this digital transformation is to add components of Data Warehousing and visual analytics tools into our outdated system architecture. But make no mistake, once achieved, these initiatives will serve as the basis of a true Automation Transformation; one where robotics and IoT can utilize our newly improved data repositories to perform tasks with minimal oversight.

**Better Analytics Saves Money**

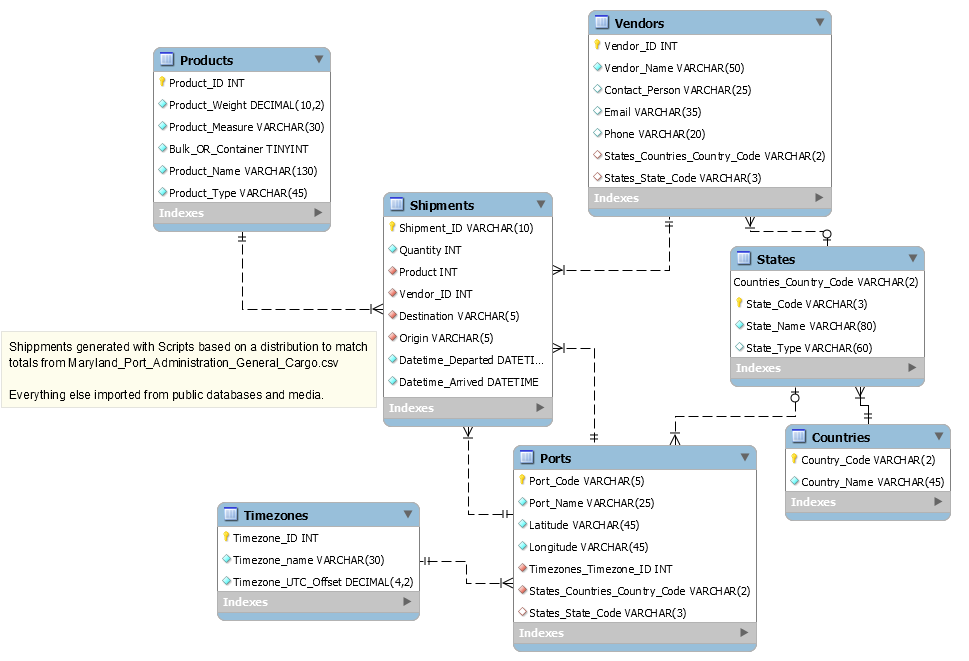
# Data Modeling:

Our proposed schema for our Data Warehouse is illustrated below through our design dictionary, as well as the diagrams for Data Warehouse and the classical ER model in the following pages:

## Database Star Schema Design

* + **Fact Table**
    - Shipment\_Fact - This table contains the fact and measures in the form of quantity, date and time
  + **Dimension Tables**
    - Vendor\_Dimension - This table contains all the details of the vendor which are associated with shipment fact
    - Date\_Dimension - This table contains a date measure which is an important dimension to know details about shipment.
    - Product\_Dimension - This table contains the information about product shipping across the ports.
    - Port\_Dimension - This table contains port details with their latitude and longitude.
    - Time\_Dimension - Time dimension contains information about the time arrived and departed in the form of hour and minute.
    - Timezone\_Dimension - Timezone dimension contains information about the UTC offset to identify the region.

## ER Schema

****

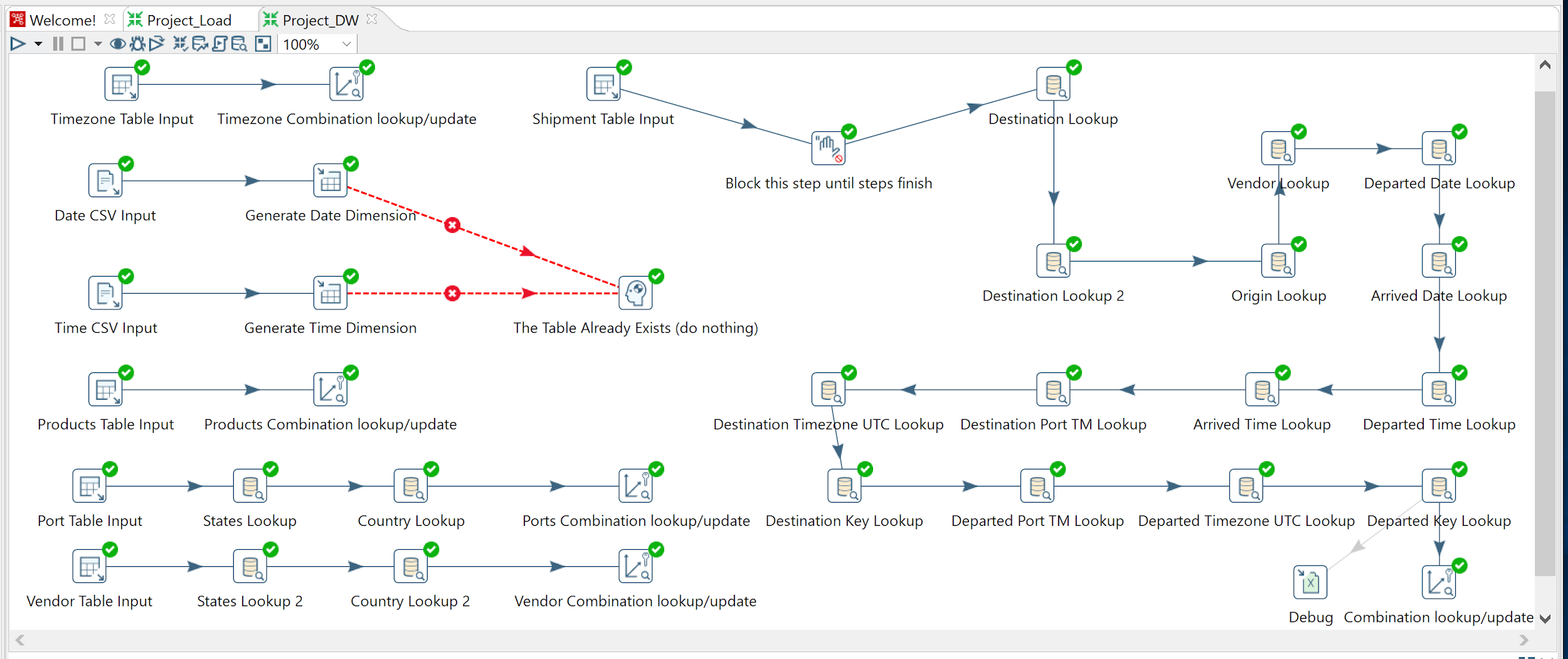
## DW Schema

# 

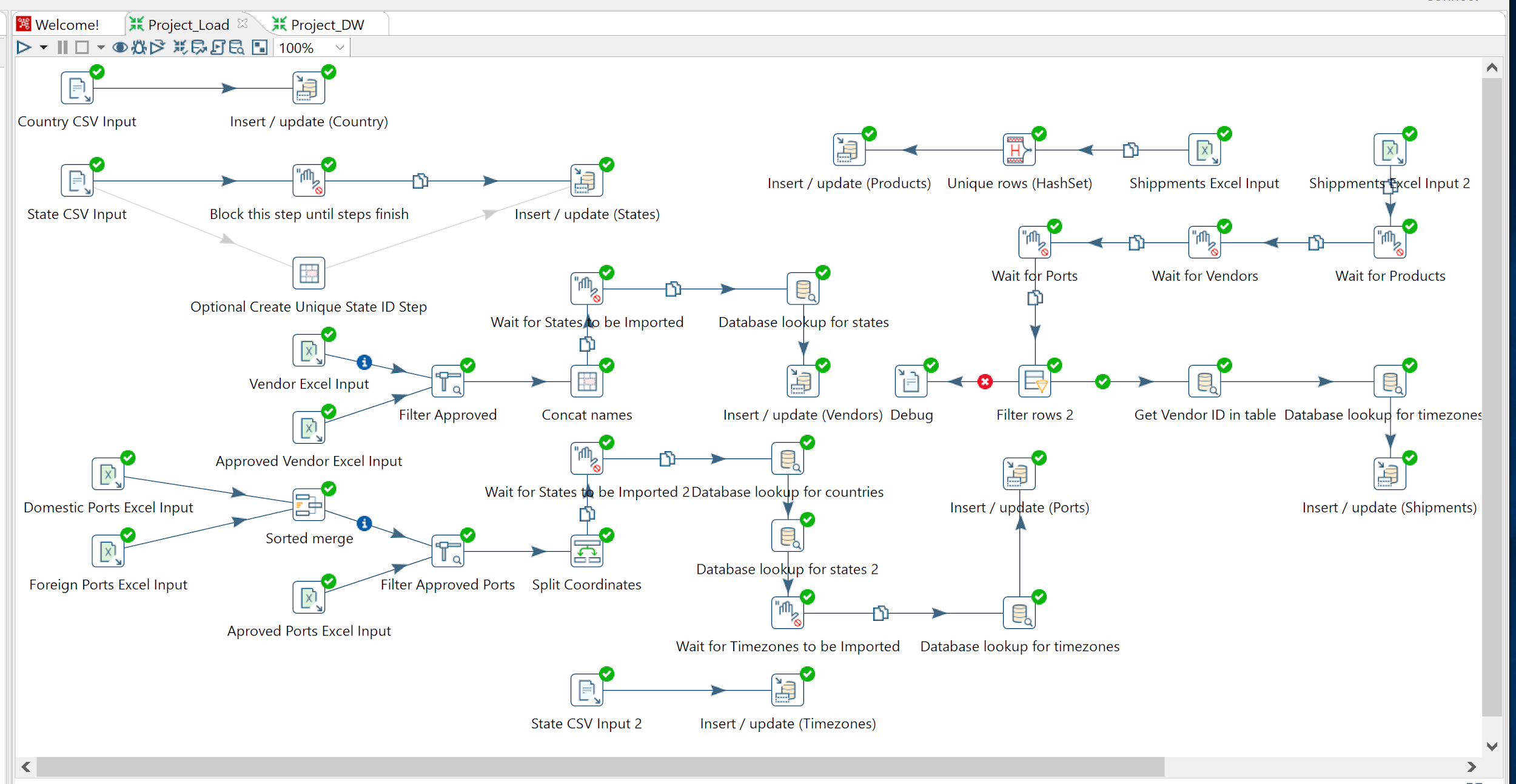
# ETL Implementation:

For the ETL of this project we used Pentaho. We had the source data for this project mocked up through different online tools and Excel beforehand, and then loaded into an ER model. The DW model loaded mostly from the ER model.

## The ER Model Load:



## The Data Warehouse Model Load:



## Takeaways from Implementation:

During the implementation of this ETL I encountered many issues regarding missing data. And it was only later in the project that I realized Pentaho allowed me to catch errors and divert them to another step, this made debugging the final stages of the transformation much easier.

Another takeaway from this ETL implementation was how working with the highly normalized tables from the ER model for the DW model import resulted in multiple levels of lookup when it came to denormalizing it. To the point that in a larger model using SQL joins would have been much quicker to implement. Another feature that Pentaho lucky supports.

## Feedback from Presentation:

During the presentation one of the relevant things to ETL that was pointed out was that our system was not real time and wasn’t thought out to interact with data flows from Google Analytics or any other ERP system. That is something that we could implement, even in Pentaho still using triggered jobs, but also through other systems.

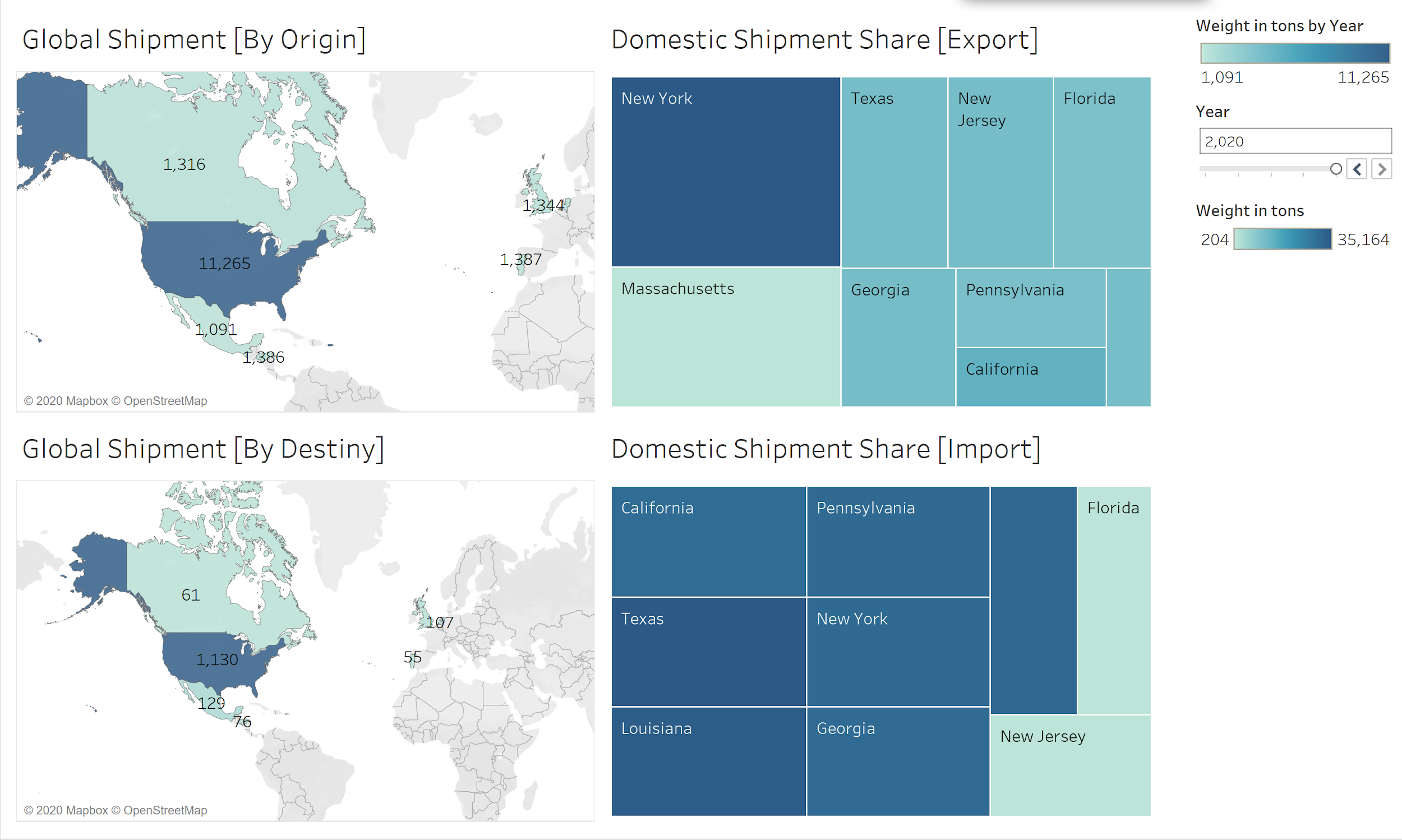
Another point on ETL that was raised during the presentation was regarding whether it would be hard, or if it would even be possible, to connect new fact tables to the model and change the granularity of the data. While we currently think that wouldn’t be much of a challenge, it is not something we planned extensively for, and in the future we will aim to have our solutions be more modular and expandable in both schema design and Pentaho Job sequences.

# Tableau Implementation:

After brainstorming and filtering which charts we would like to include, we decided to split our dashboards into 4 parts. 2 filled with descriptive data, and two filled with prescriptive data, each with their own theme.

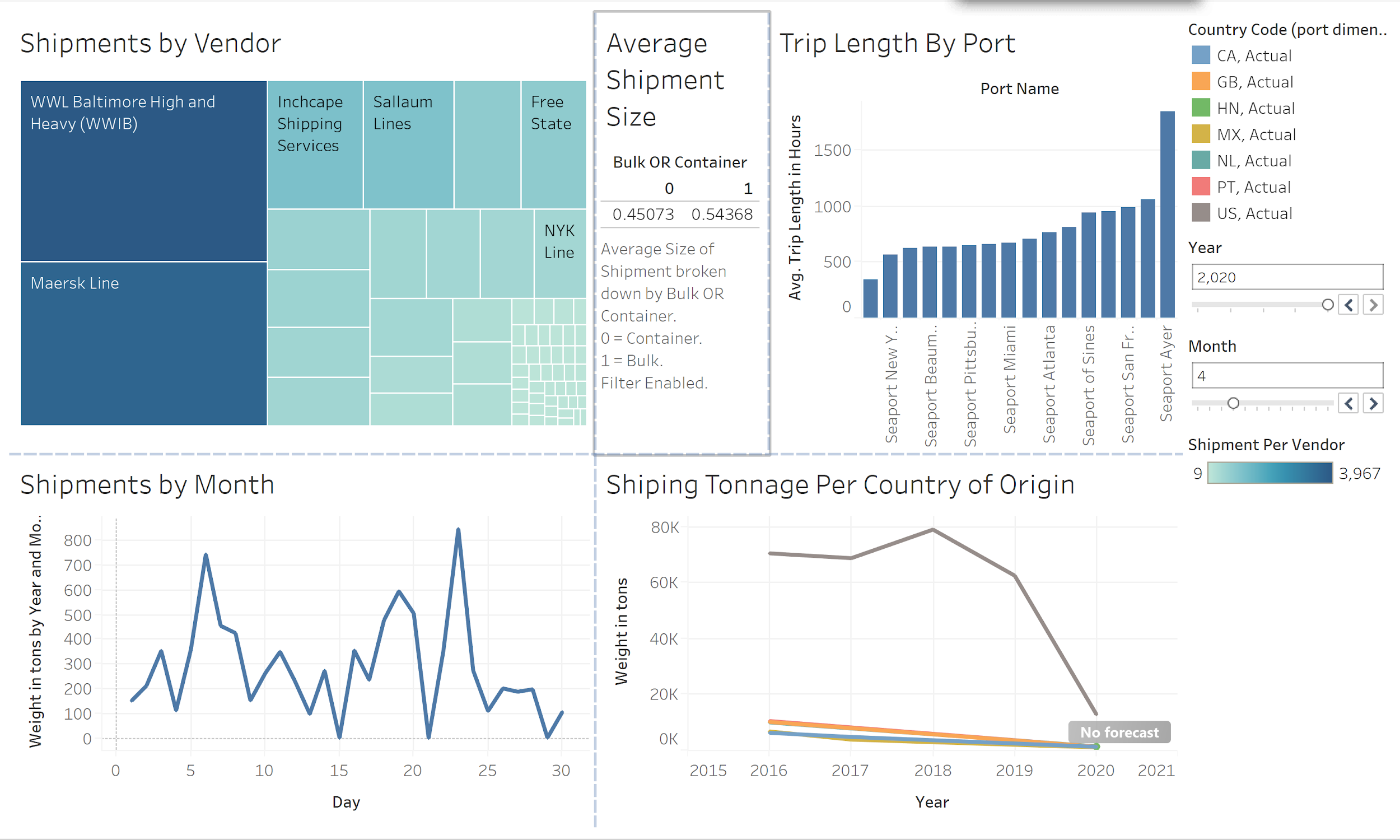
## Descriptive Part 1

This first dashboard was about shipping by geographic location. It was filterable by year and either state or country. There were no complaints on this dashboard in specific during the presentation nor any particular issue with it.



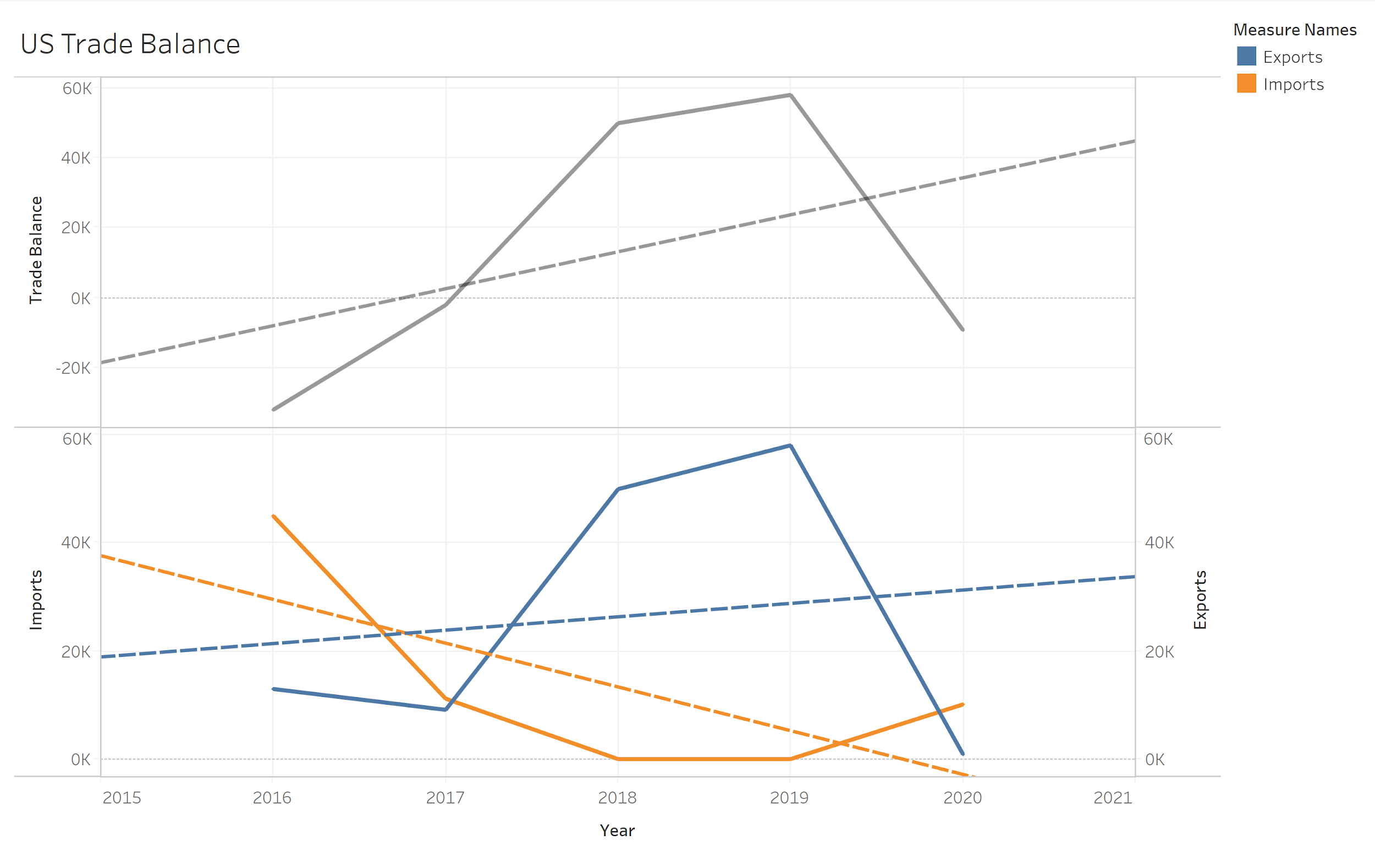
## Descriptive Part 2

This dashboard was about shipment volume (our fact) as related to dimensions other than purely their location, that included ports, product type, route length(in hours), vendors, shipment amount over time within a given month, and countries over years. For this one our main takeaway ended up being the fact we should have integrated more forecasts into this section. But due to the way we didn’t precalculate fields, Tableau couldn’t generate a forecast.



## Predictive Part 1

This dashboard we made about a very relevant lag measure for shipment amounts: trade balance forecast. Assuming there should be a return to the mean in trade, this chart generated making use of Tableau’s trend function is useful to anticipate an upcoming bounceback in trade in 2021.



## Predictive Part 2

This dashboard was made to explore an additional scenario with 2021, concerning what the market would look like with different degrees of consolidation. Considering WWIB and Maersk are almost half of our trade volume, it was interesting to see how their growth at the cost of smaller shippers might impact our trade to other ports. This dashboard also received no criticism nor warranted comments from our team.



## General Tableau Takeaways

While creating graphs in Tableau three things that quickly became apparent were that: datetime fields would work much better than custom integer fields in time dimensions, patchy data makes generating forecasts much more difficult, and finally it was the point in the project that gave us the clearest vision of what data we would want most for our reports. It never occurred to us to have price data until we got to Tableau, but once we were putting together the dashboards it became painfully clear how helpful it would have been.