IST 718 Final Project Report

Team Members

Brandon Reyes, Leonidas Garcia, Myles Chalue, Jogesh Pugazhendhi

Overview:

Supply chain logistics create lots of data. There is collaboration between various departments in the logistics department to co-ordinate and ship products ordered by customer. There are various systems that are used in this process create details of a shipment and monitor the shipment at various times until it is delivered to customer.

Data:

Shipping data from Logistics system: The data is extracted from a logistics system for a period of 5 years from Jan 2015 to June 2021 with almost 0.5 million transactions. The data has following information from_location_zipcode, to _location_Zipcode, miles driven, cost of shipping, Dispatch date, delivered date and other metrics

The dataset has 2 dimensions (from_zipcode and to_zipcode) and rest are all shipping transaction metrics. Total number of attributes – 19

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	stops	weight_truck	truck_cube	pieces	pallets	miles_driven	estimated_ship_date	dispatch_date	delivered_date	cost	cost_after_adjustment	adjusted_amount	access_charge_amount	driving_duration	elapsed_duration	layover_duration	wait_duration	from_zipcode	to_zipcode
2	3	45677	19286	2520	28	430	2/7/2015	2/5/2015	2/7/2015	2178.1	2102.1	2009.7	92.4	375	379	5	1	24301	21224
3	3	46086	21250	1550	35	552	2/6/2015	2/5/2015	2/6/2015	2753.76	4380.96	4254.4	126.56	508	514	5	1	24301	8234
4	3	46423	17671	2607	20	256	2/6/2015	2/5/2015	2/5/2015	1262.08	1251	1210.44	40.56	185	187	5	1	98446	97062
5	3	46295	13969	2075	20	469	2/14/2015	2/12/2015	2/12/2015	1674.63	1674.63	1575	99.63	417	419	5	1	33563	29910
6	3	43764	12713	1921	19	469	2/13/2015	2/12/2015	2/12/2015	1674.63	1674.63	1575	99.63	417	419	5	1	33563	29910
7	3	47009	21062	3118	37	469	2/12/2015	2/12/2015	2/12/2015	1402.21	1387.08	1287.45	99.63	417	423	5	1	33563	29910
8	3	45792	13833	2340	24	253	2/14/2015	2/12/2015	2/12/2015	1477.36	1455.56	1414.25	41.31	181	184	5	1	24301	28273
9	3	45700	18750	3180	20	208	2/17/2015	2/12/2015	2/12/2015	1649.68	1619.68	1570	49.68	132	134	5	1	1057	3109
10	3	46491	18690	1540	24	114	2/20/2015	2/17/2015	2/17/2015	1158.5	1150.75	1147.25	3.5	36	39	5	1	98446	98421
11	3	46576	19317	850	23	176	2/20/2015	2/17/2015	2/17/2015	1351.12	1327.87	1308.87	19	97	100	5	1	98446	98270

There will be additional variables derived from this dataset to know number of days between estimated ship date and actual ship date. Also, the time it took to deliver dispatch date and delivered date

Another variable that created in the dataset would be the distance between from and to zipcode to understand the actual distance the miles driven and if this distance value affects any of the other metrics

Average Crude Prices: Since logistics has direct relation to the crude oil prices, the data for average crude prices by month/year will be joined to the above dataset on the shipping date to understand if there has been a linear progression between oil prices and cost of shipment

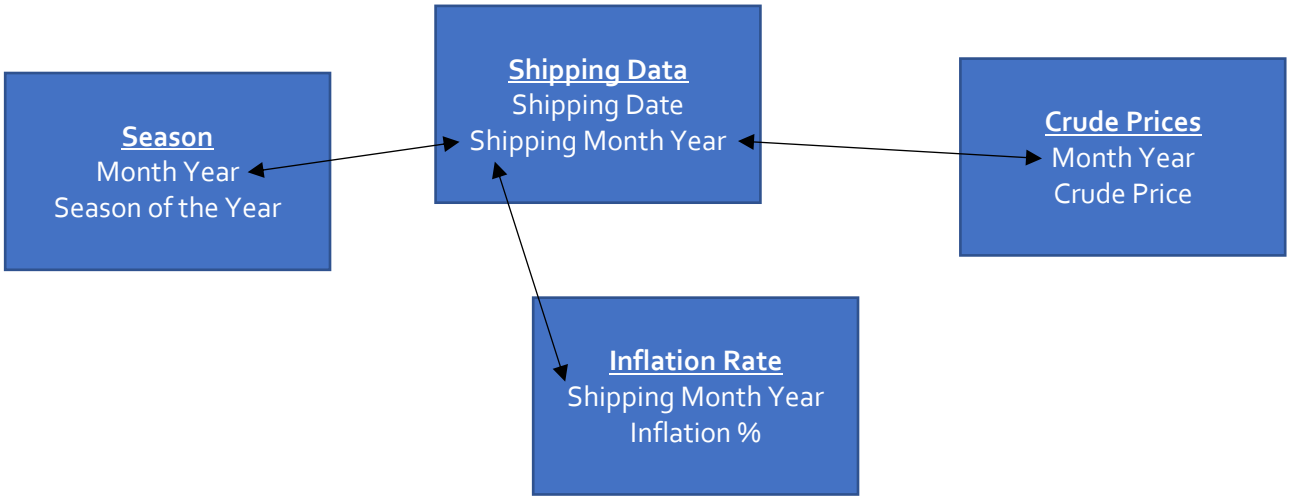
Data will be scraped from the below website
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=M

Season: This dataset will benefit from having day to day weather information for all zip codes. Since historical weather data by zip code was difficult to obtain, seasons of the year will be added based on the month of shipping

Inflation Rate: Adding inflation data by year and month from 2015 to 2021 to understand if the inflation rate has any effect in any of the other metrics

Inflation rate is extracted from the below website
https://www.usinflationcalculator.com/inflation/current-inflation-rates/

Data Relationships:



Analytics on Supply Chain Logistics:

The department who runs the supply chain logistics will benefit from predicting delays in shipment, and address challenges withing the supply chain to reduce number of days for certain shipments to be delivered. The business can also use other factors to calculate what the ideal price should be for a shipment based on external factors like oil price and season of the year

What problem are you attempting to solve?

Was there any delay in shipping for certain transactions? If so, what factors might have contributed to delay in shipping and prevent delays in the future.
Predict number of delay days for a shipment
Predict cost of shipment based on average crude prices, season, and other metrics .

What have you observed thus far in the data?

We created a new column in the data to calculate the number of days a shipment arrives late by taking the dispatch date and subtracting from it the delivered date. We then created a new subset of the data containing only those shipments that arrived after the shipment date by setting all negative delays (early shipments) to zero, and then removing all rows where the days delayed equaled zero. Then, taking a closer look at the data through various scatterplots, using a pair plot, we were able to see that the number of shipments who were only delayed by one day was exponentially larger than shipments which where later than that. To get a more accurate picture of the rest of the data, another subset was created by removing shipments that were only late by one day, and the same pairplot was created, allowing for a much clearer picture of the rest of the data. One interesting thing to take from the scatterplots is that the duration of the shipment seems positively correlated with the number of days a shipment is delayed.

Added additional information to the dataset

Extracted Inflation data from the website
https://www.usinflationcalculator.com/inflation/current-inflation-rates/

Merged inflation information to the shipping data based on dispatch date

Extracted Crude oil prices data per month from the website
https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RWTC&f=M

Added Season Data based on dispatch date

Created additional column shipping days to know the number of days between dispatch date and delivered date for a shipment

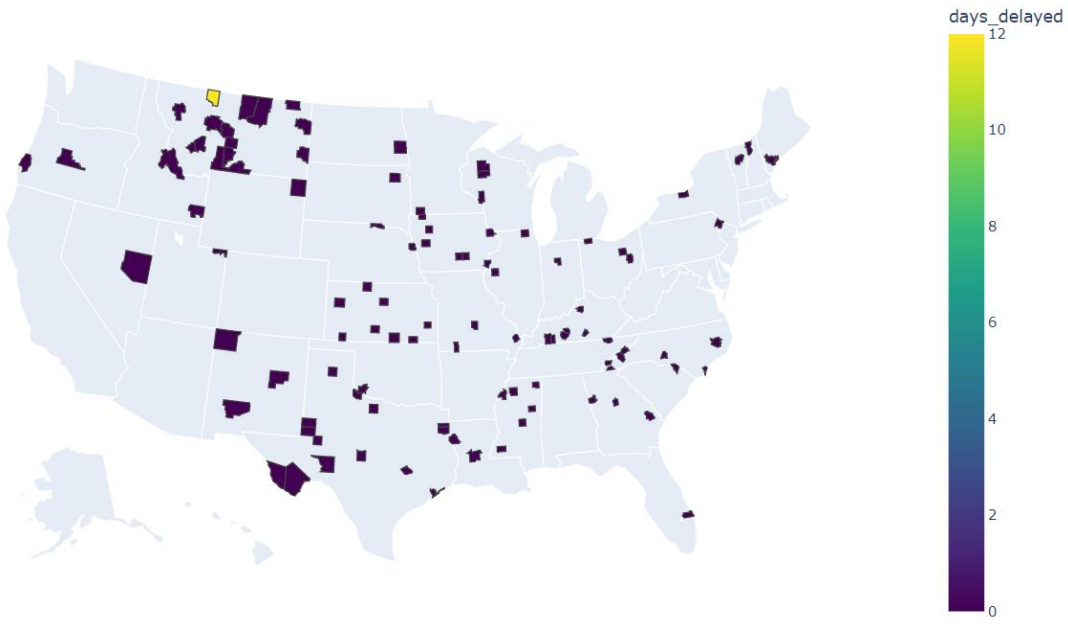
Combined all this external information to the shipping dataset

shipping_final2 - DataFrame

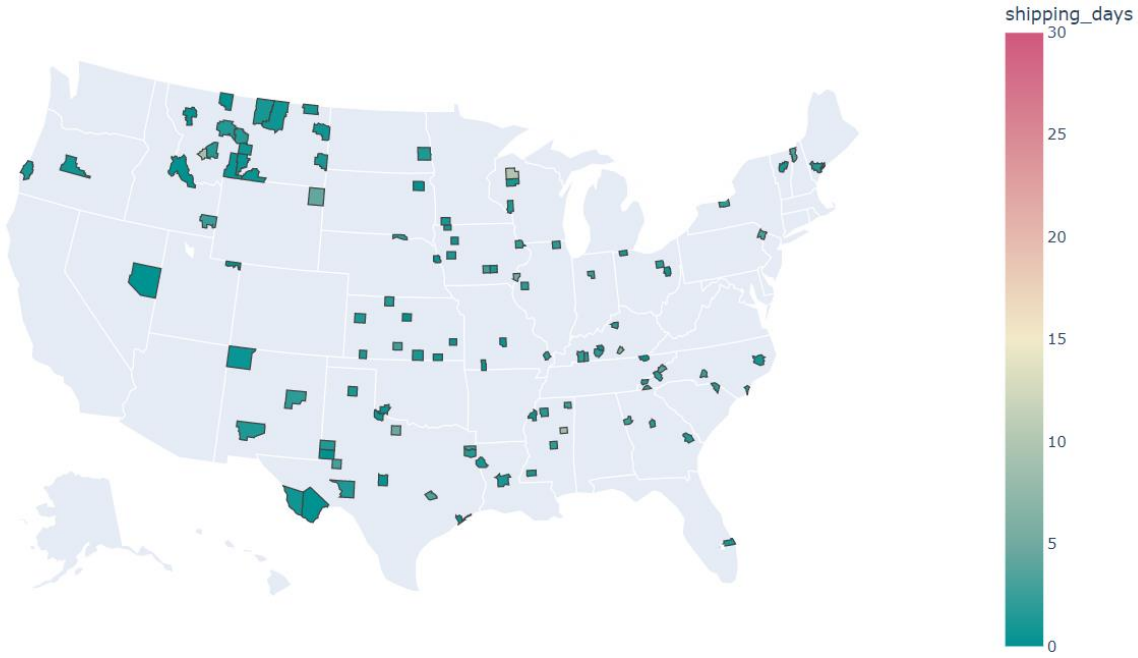
Index	estimated_ship_date	dispatch_date	delivered_date	cost	fter_adjus	sted_amc	_charge_a	ring_durat	sed_dura	ver_durat	ait_duratic	m_zipco	o_zipcod	patch_mo	dispatch_year	crude_price	inflation_rate
0	015-02-07 00:00:00.000000	2015-02-05 00:00:00.000000	2015-02-07 00:00:00.000000	2178.1	2102.1	2009.7	92.4	375	379	5	1	24301.0	21224	Feb	2015	50.58	0.0
1	015-02-06 00:00:00.000000	2015-02-05 00:00:00.000000	2015-02-06 00:00:00.000000	2753.76	4380.96	4254.4	126.56	508	514	5	1	24301.0	8234	Feb	2015	50.58	0.0
2	015-02-06 00:00:00.000000	2015-02-05 00:00:00.000000	2015-02-05 00:00:00.000000	1262.08	1251	1210.44	40.56	185	187	5	1	98446.0	97062	Feb	2015	50.58	0.0
3	015-02-14 00:00:00.000000	2015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	1674.63	1674.63	1575	99.63	417	419	5	1	33563.0	29910	Feb	2015	50.58	0.0
4	015-02-13 00:00:00.000000	2015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	1674.63	1674.63	1575	99.63	417	419	5	1	33563.0	29910	Feb	2015	50.58	0.0
5	015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	1402.21	1387.08	1287.45	99.63	417	423	5	1	33563.0	29910	Feb	2015	50.58	0.0
6	015-02-14 00:00:00.000000	2015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	1477.36	1455.56	1414.25	41.31	181	184	5	1	24301.0	28273	Feb	2015	50.58	0.0
7	015-02-17 00:00:00.000000	2015-02-12 00:00:00.000000	2015-02-12 00:00:00.000000	1649.68	1619.68	1570	49.68	132	134	5	1	1057.0	3109	Feb	2015	50.58	0.0
8	015-02-20 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1158.5	1150.75	1147.25	3.5	36	39	5	1	98446.0	98421	Feb	2015	50.58	0.0
9	015-02-20 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1351.12	1327.87	1308.87	19	97	100	5	1	98446.0	98270	Feb	2015	50.58	0.0
10	015-02-20 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1324	1308.47	1294.97	13.5	73	76	5	1	98446.0	98028	Feb	2015	50.58	0.0
11	015-02-21 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1324	1308.47	1294.97	13.5	73	77	5	1	98446.0	98028	Feb	2015	50.58	0.0
12	015-02-20 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1372	1354.18	1338.68	15.5	82	85	5	1	98446.0	98026	Feb	2015	50.58	0.0
13	015-02-21 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1372	1354.18	1338.68	15.5	82	85	5	1	98446.0	98026	Feb	2015	50.58	0.0
14	015-02-24 00:00:00.000000	2015-02-17 00:00:00.000000	2015-02-17 00:00:00.000000	1324	1308.47	1294.97	13.5	73	76	5	1	98446.0	98028	Feb	2015	50.58	0.0

Trends and observation:

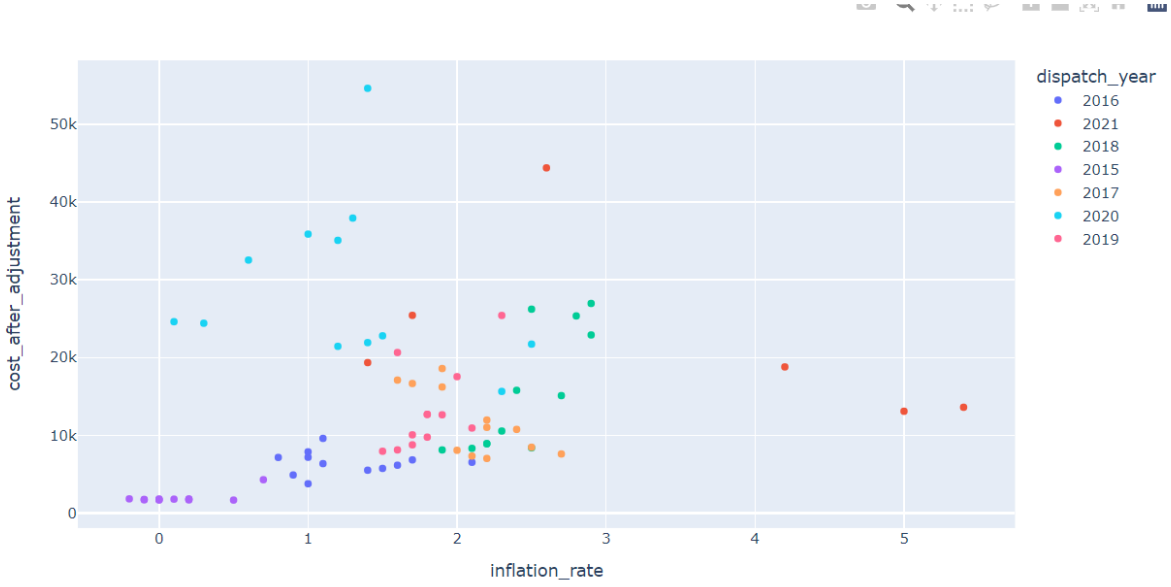
Delay Days by Zip Code



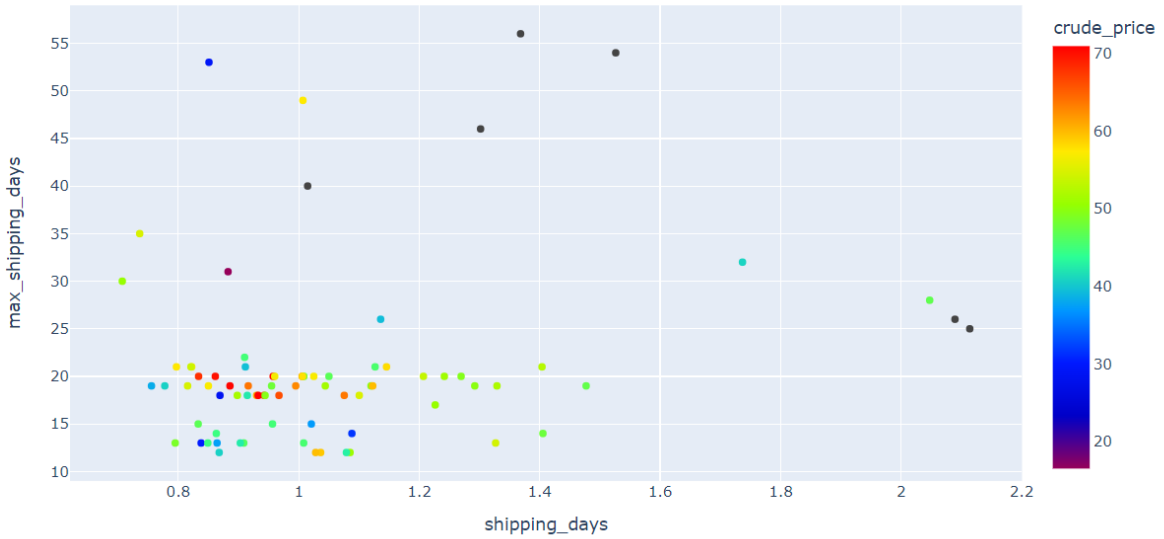
Average Shipping days by Zipcode



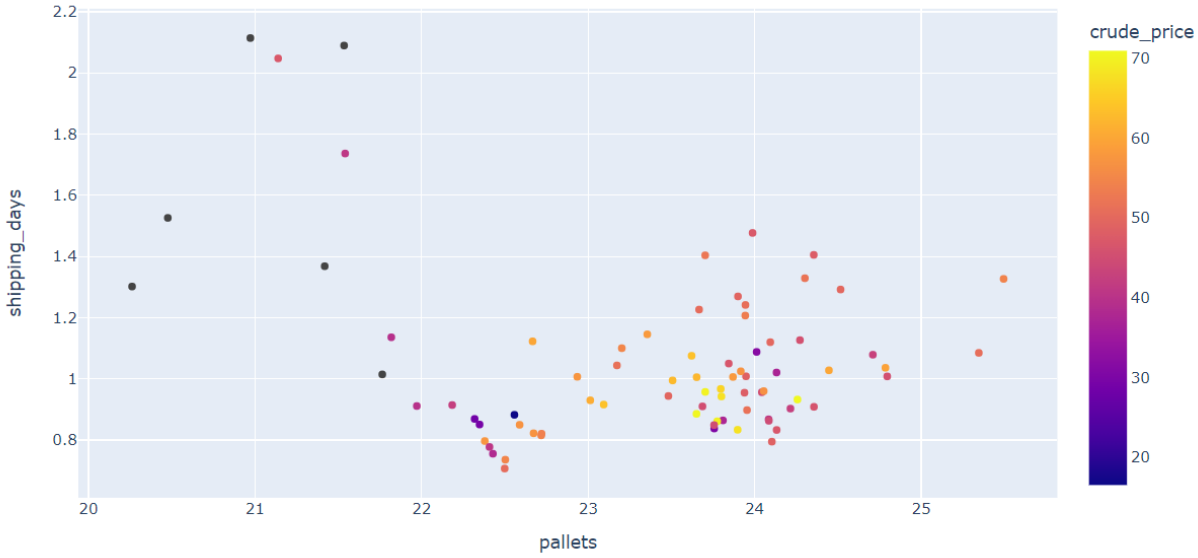
Inflation vs Cost



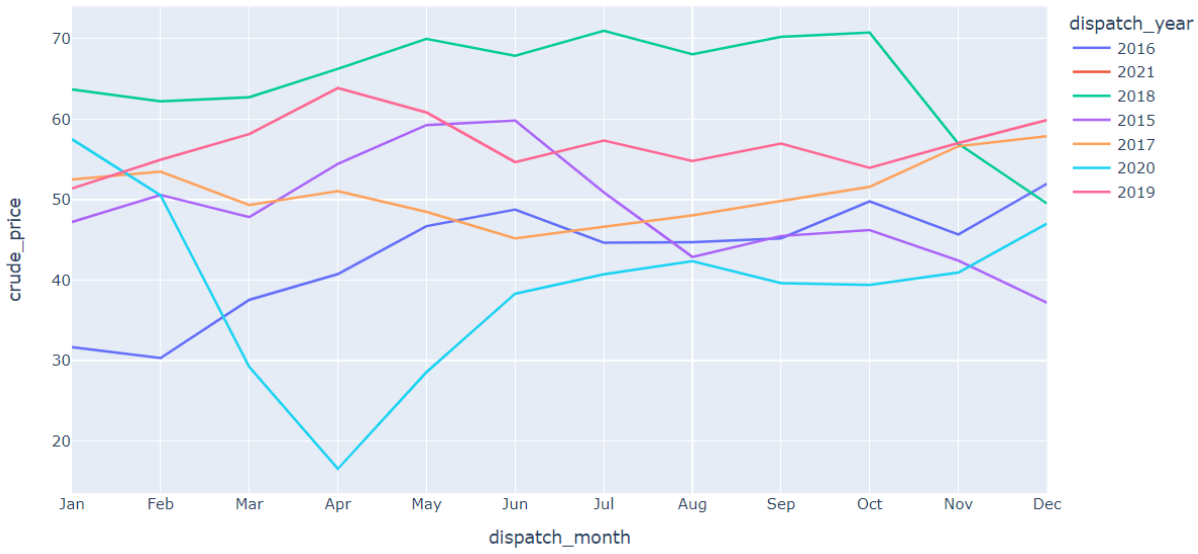
Shipping Days vs Maximum Shipping Days Vs Crude Price



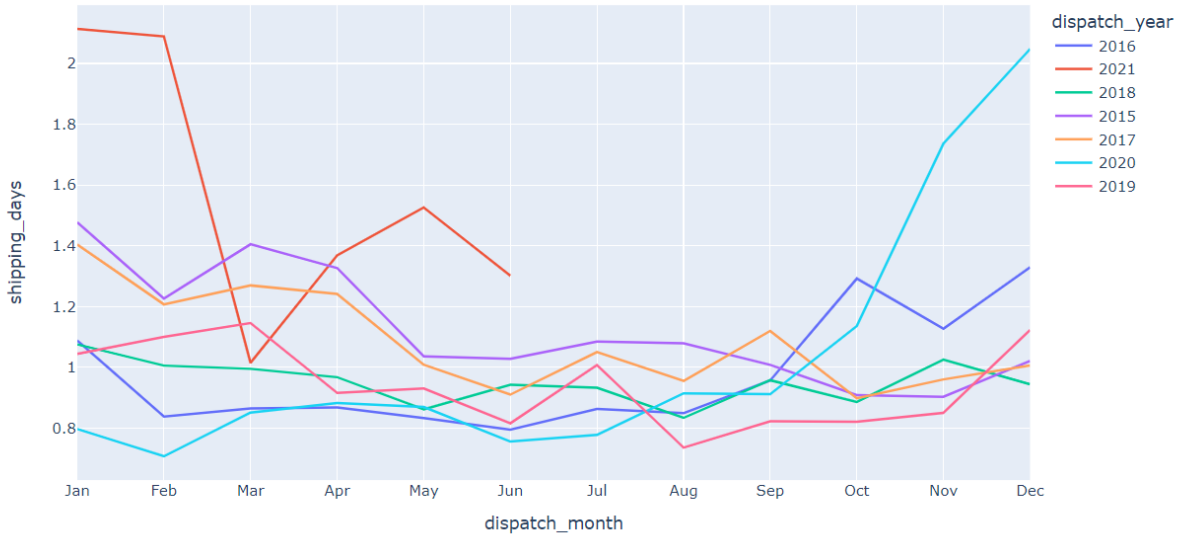
Number of Pallets vs Shipping days Vs Crude Price



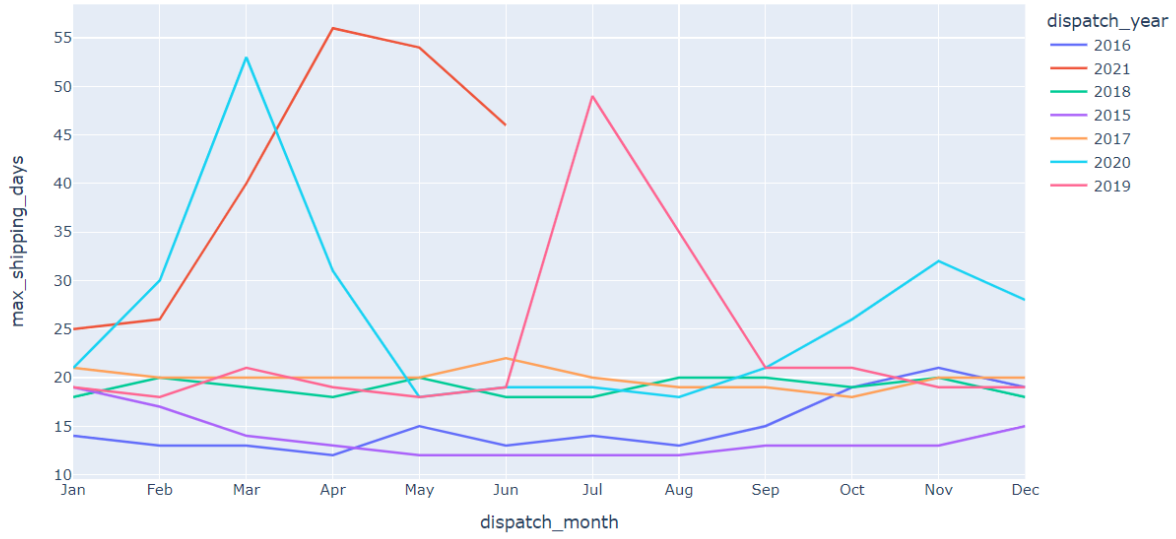
Crude Prices Trend



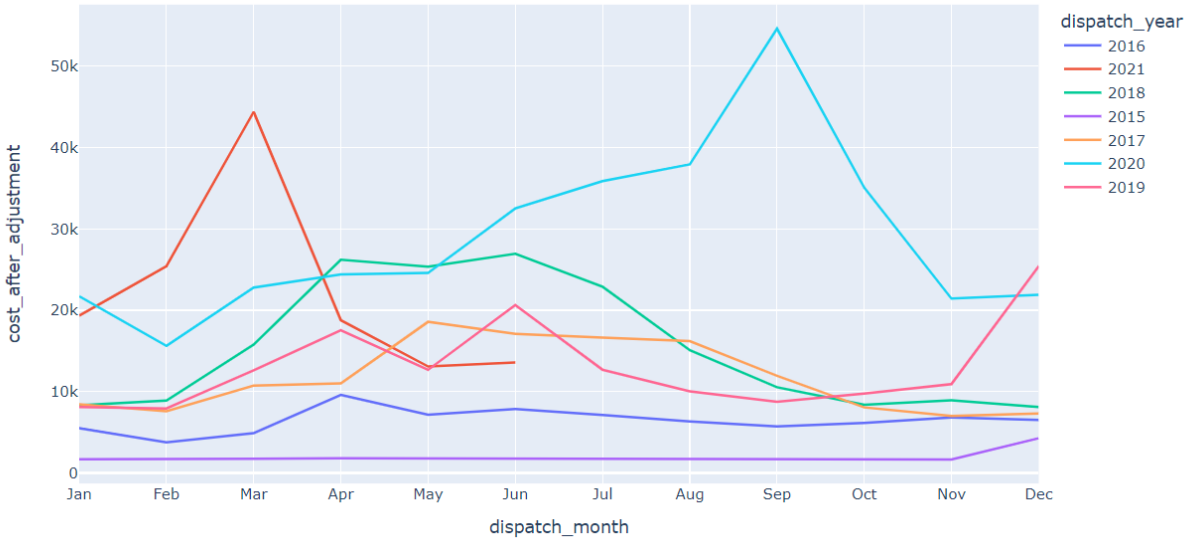
Shipping Days Trend



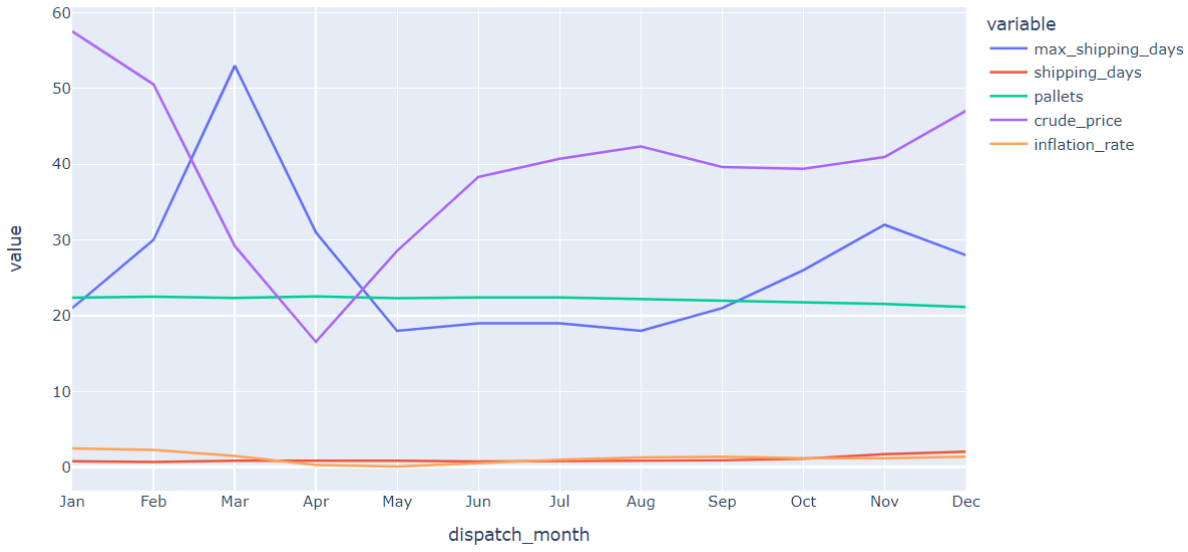
Maximum Shipping Days



Cost Trend



Comparing metrics for 2020



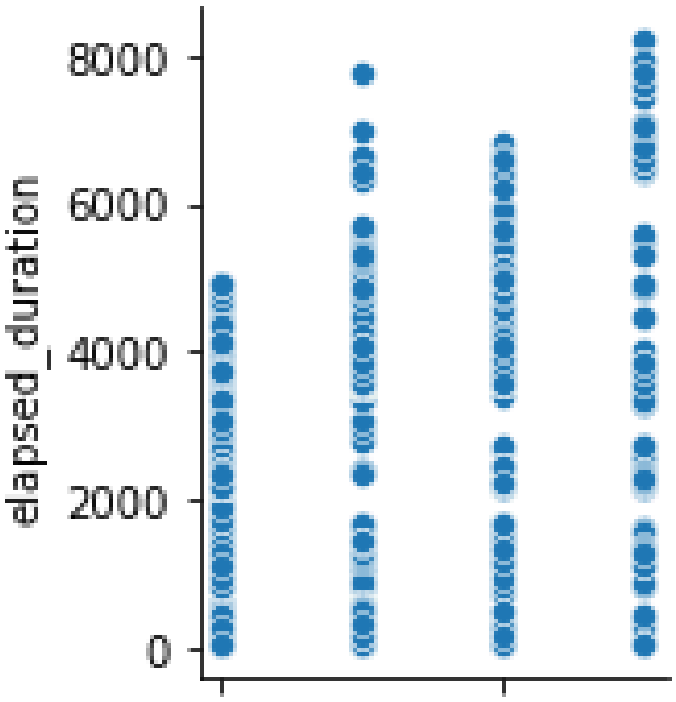
Questions answered:

Were there shipping delays?

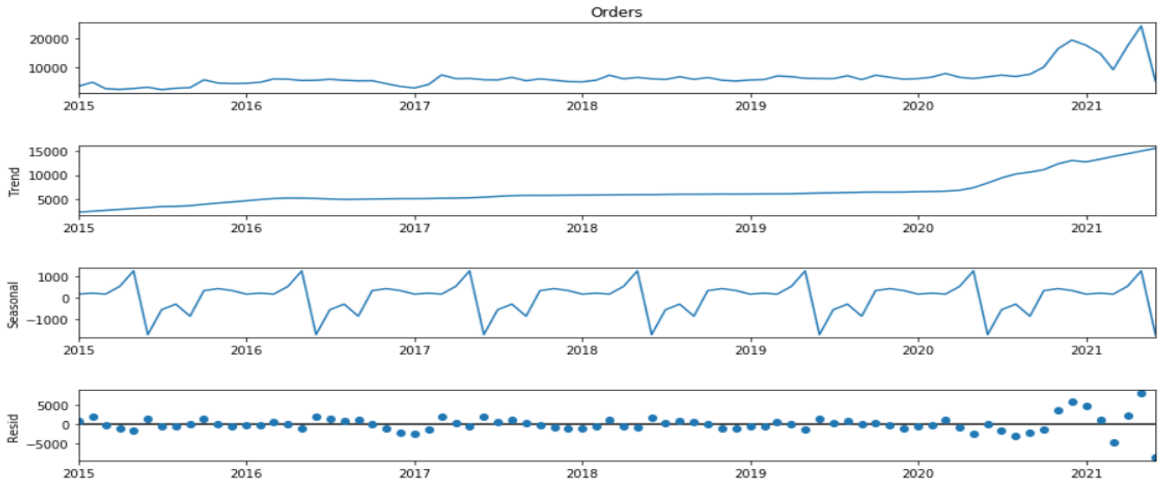
- Majority of shipments are on time (89.3%)
- Of those delayed, most are by one day (7.8%)
- Occasional delays of 2-5 days (2.8%)
- Very rare delays of > 5 days (0.1%)

If Delayed, Why ?

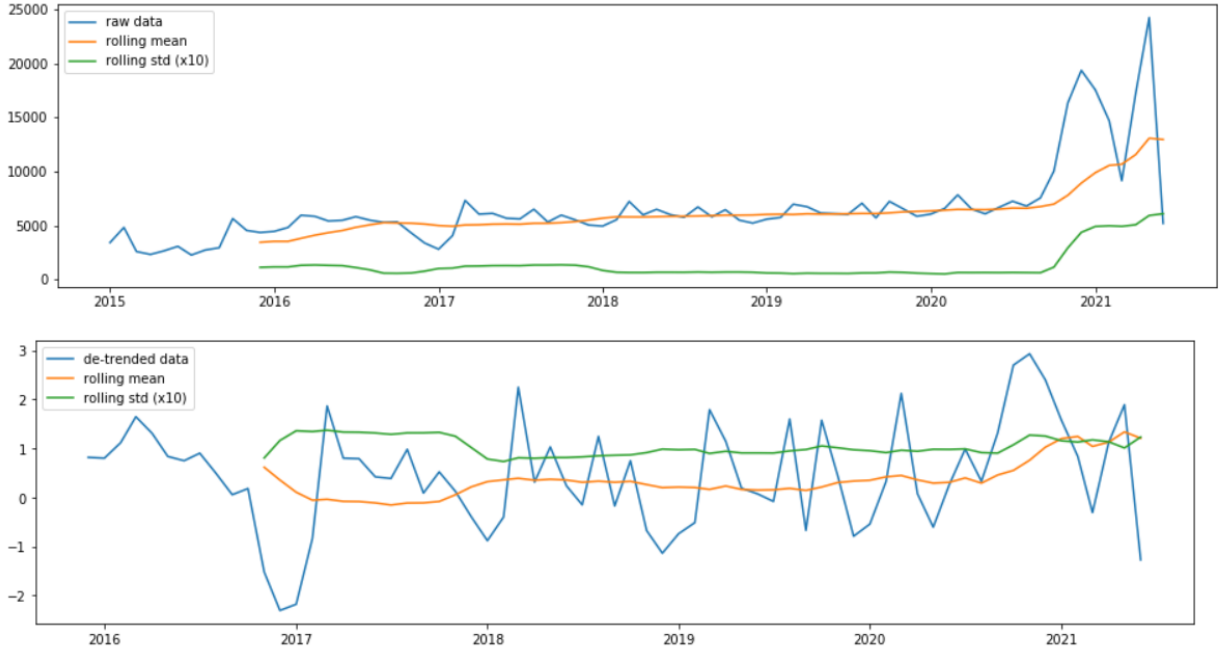
- Possible factors?
  - Number of stops
  - Truck Weight
  - Number of Pallets
  - Duration of Shipment



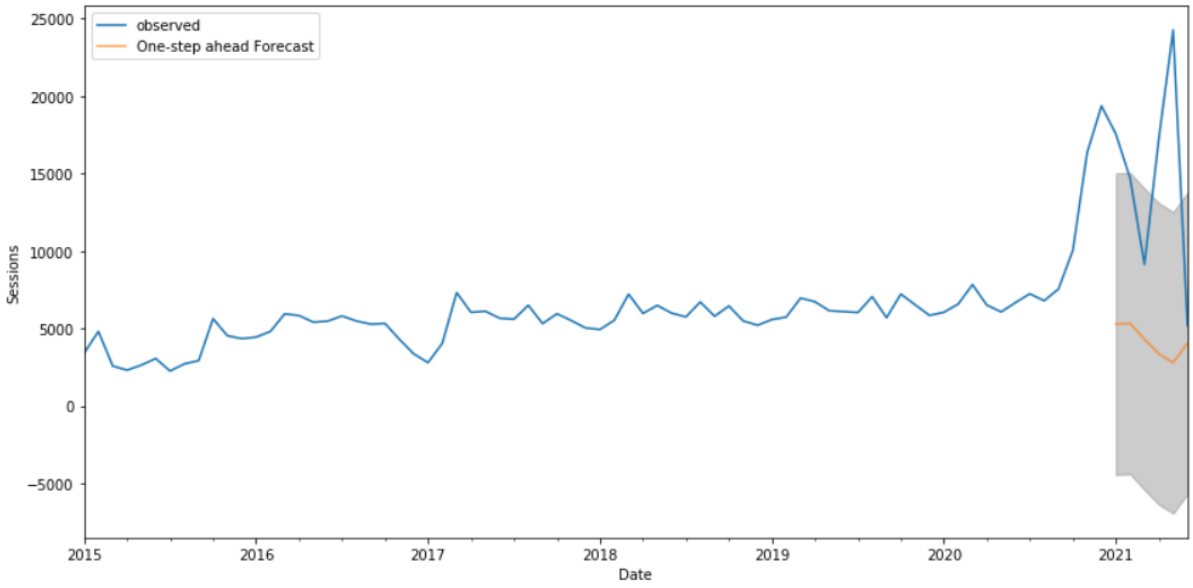
Time Series - Predicting Number of Loads per month (Seasonal Decompose)

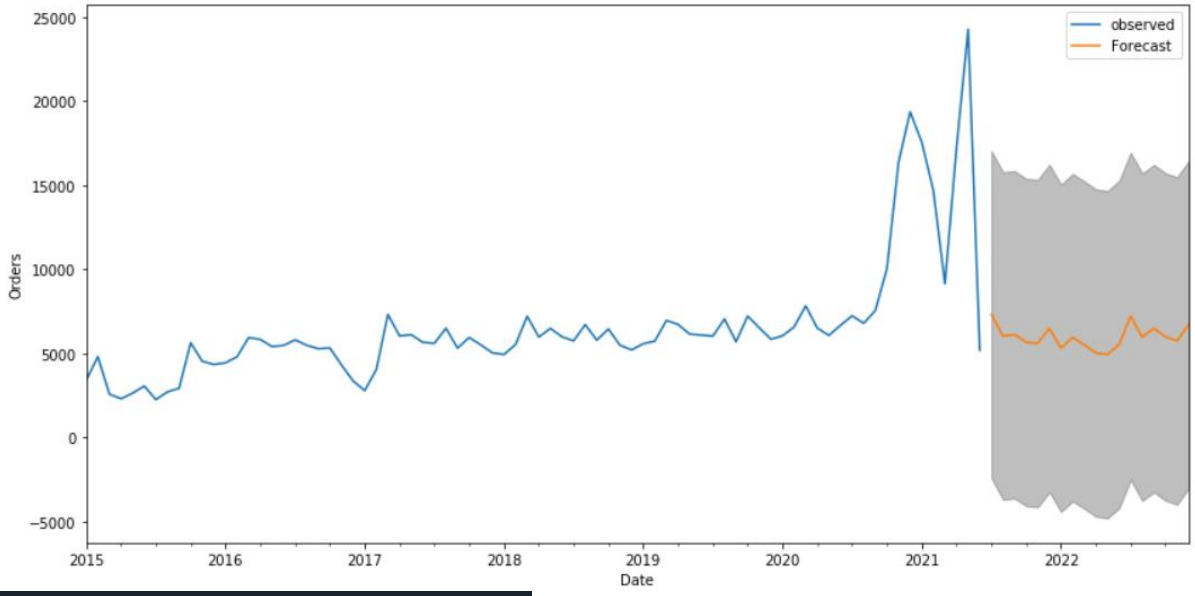


Loads Per month – ADF Test



Loads per month (Compare Train Vs Test and Forecast)

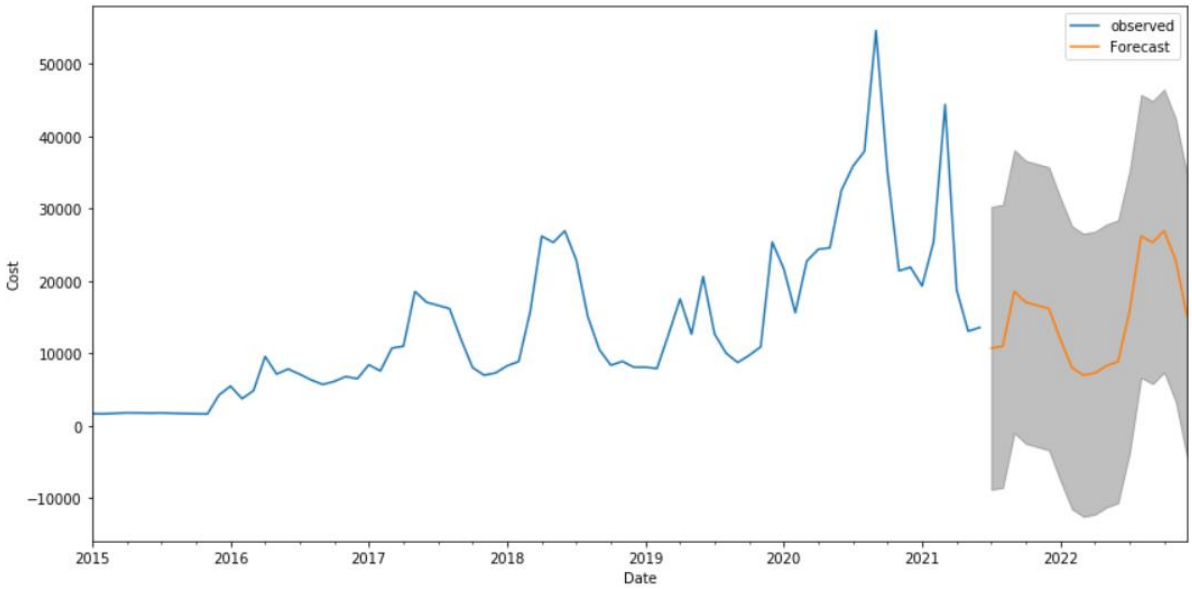





```
In [32]: final_table
Out[32]:
         Date  Predicted_Mean
0  2021-07-01   7319.00000000
1  2021-08-01   6044.00000000
2  2021-09-01   6117.00000000
3  2021-10-01   5664.00000000
4  2021-11-01   5601.00000000
5  2021-12-01   6499.00000000
6  2022-01-01   5316.00000000
7  2022-02-01   5951.00000000
8  2022-03-01   5512.00000000
9  2022-04-01   5034.00000000
10 2022-05-01   4937.00000000
11 2022-06-01   5530.00000000
12 2022-07-01   7215.00000000
13 2022-08-01   5976.00000000
14 2022-09-01   6490.00000000
15 2022-10-01   5992.00000000
16 2022-11-01   5749.00000000
17 2022-12-01   6712.00000000
```

Above table shows the prediction of loads as per the time series model

Cost Prediction:




```
In [37]: final_table
Out[37]:
         Date  Predicted_Mean
0  2021-07-01  10745.58224211
1  2021-08-01  11010.03230146
2  2021-09-01  18575.93344613
3  2021-10-01  17086.91500000
4  2021-11-01  16648.16489913
5  2021-12-01  16200.34897215
6  2022-01-01  11951.70894846
7  2022-02-01   8057.66138632
8  2022-03-01   7008.87130261
9  2022-04-01   7317.73120977
10 2022-05-01   8307.91350820
11 2022-06-01   8891.19314647
12 2022-07-01  15773.27535828
13 2022-08-01  26210.32755689
14 2022-09-01  25343.11220031
15 2022-10-01  26947.35122830
16 2022-11-01  22894.64075317
17 2022-12-01  15092.43885727
```

The data for 2021 varies every month and its doesn't follow similar patterns as the previous years. Hence, 2021 data is not ideal to be used for predicting forecast

Takeaway and Suggestions:

1. Though the forecasting looks inline with other years except for 2021, further tweaks are needed to refine the timeseries model
2. The forecast is using lower value of predicted mean as values less than 0. In our case, number of loads and cost won't be less than 0 at any point. Hence, need to try out other forecast models with a condition that the forecast cannot be less than 0
3. External factors like inflation and crude prices didn't have direct co-relation to the cost and number of loads over the years