FLOW

Predicting Average Container Terminal Dwell Times at the Port of Savannah, GA

Kelly Wang

June 13, 2024

U.S. Department of Transportation
Office of the Secretary of Transportation

Predicting Container Terminal Dwell Times using FLOW Data

Project Goals

- Explore how ports can leverage FLOW data for operational advantages

Methodology

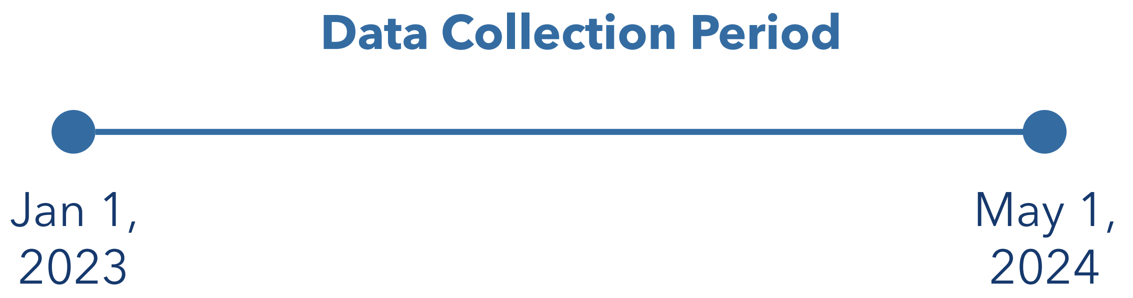
- Use FLOW ocean carrier bookings and SAV historical terminal dwell time data to develop a model to forecast daily average container dwell times at the port of SAV terminal

Benefits and Potential Applications

- Forecast average container terminal dwell times on a given day using forward-looking bookings data
- Aid in staffing and operational decision-making to enhance port efficiency

Methodology

1. Data Collection
2. Data Exploration
3. Model Development
   - Data splitting/preprocessing
   - Feature selection
   - Model selection
     - Multiple linear regression
     - Random forest
4. Model Evaluation and Validation
5. Model Prediction on FLOW Bookings Data
6. Results and Analysis



Data Collection

Ocean Carrier Bookings: Expected import TEUs being unloaded onto the SAV port each day

GPA Average Dwell Times: Average of the dwell times of all import containers sitting on the SAV terminal each day from Jan 1, 2023 to May 1, 2024

Marine Terminal Utilization: Total number of import, export, and empty container slots being used in the SAV terminal

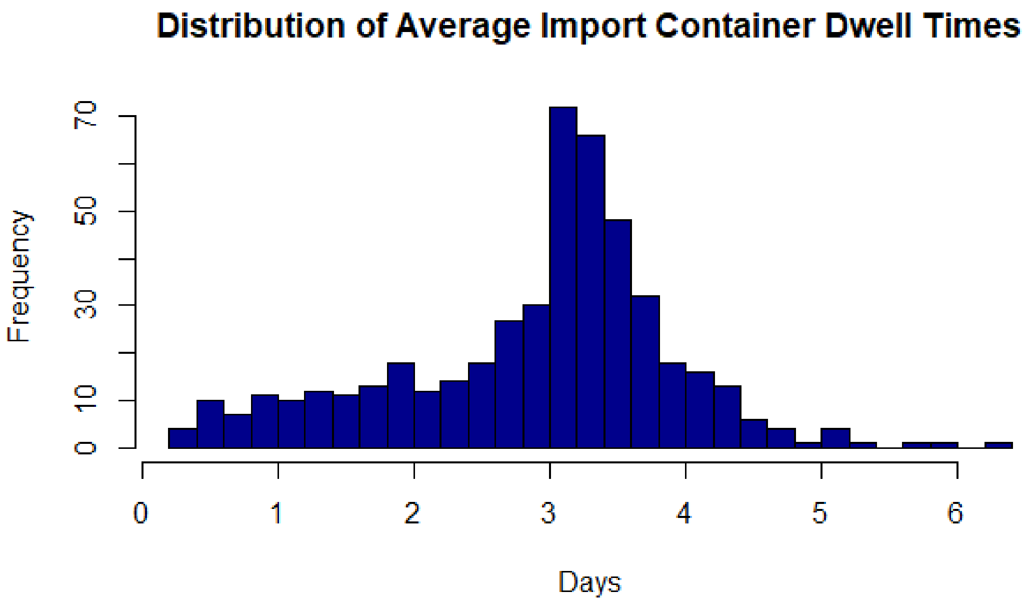
Gate Moves/Throughput: Number of containers loaded in and loaded out through the SAV gate each day

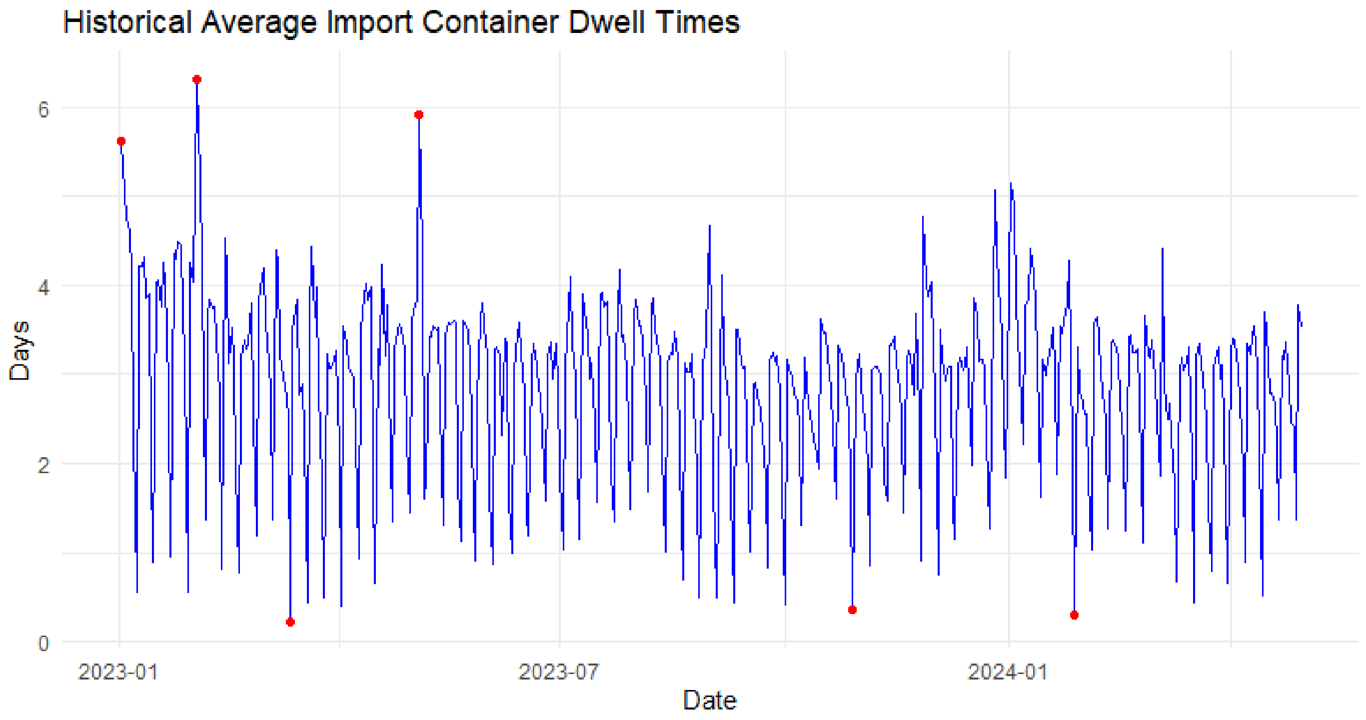
Data Exploration

Average Container Dwell Time Distributional Analysis

- Average dwell times range from 0.23 to 6.3
- Mean = 2.9 days
- Standard deviation = 1.02 days

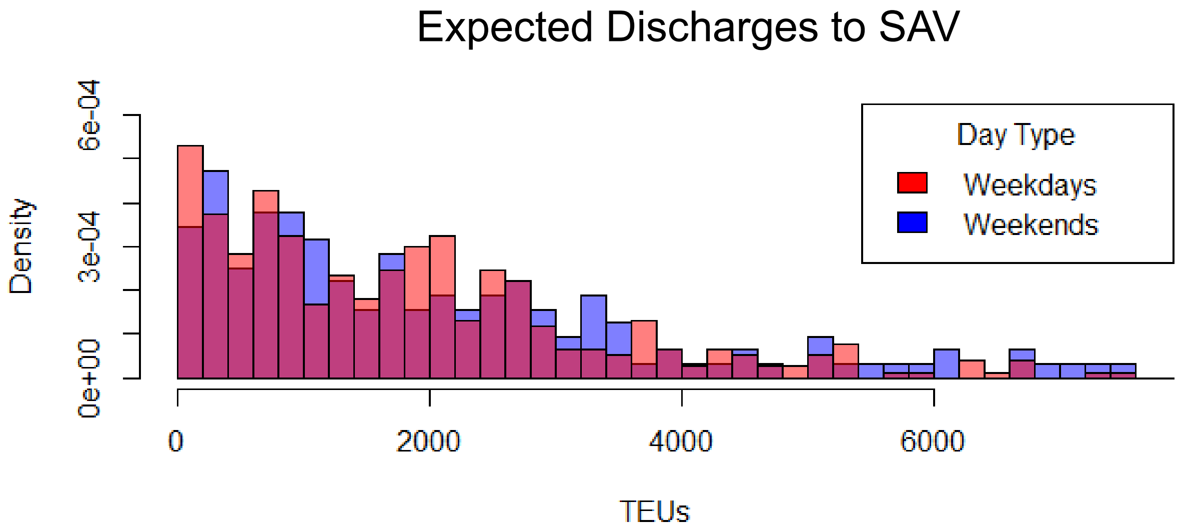
- 7 outliers (z score >= 2.5)
- 31 Days with avg dwell time < 1, 87% occur on Sundays
- 6 days with avg dwell time > 5, all occur at beginning of Jan, Feb, or May

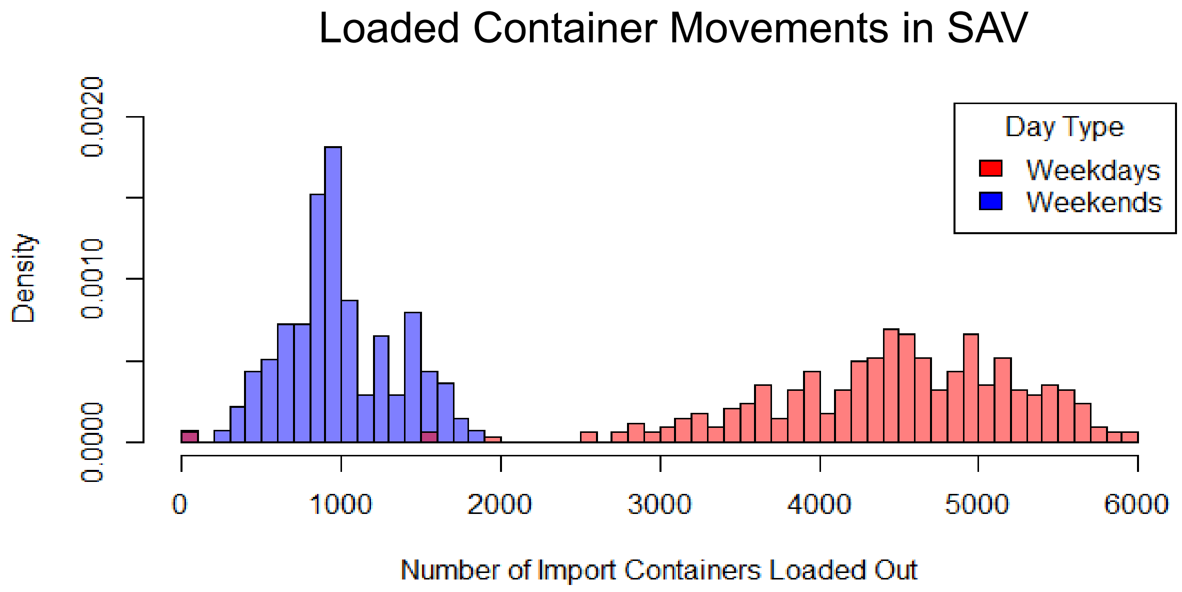




Weekday vs Weekend Effects

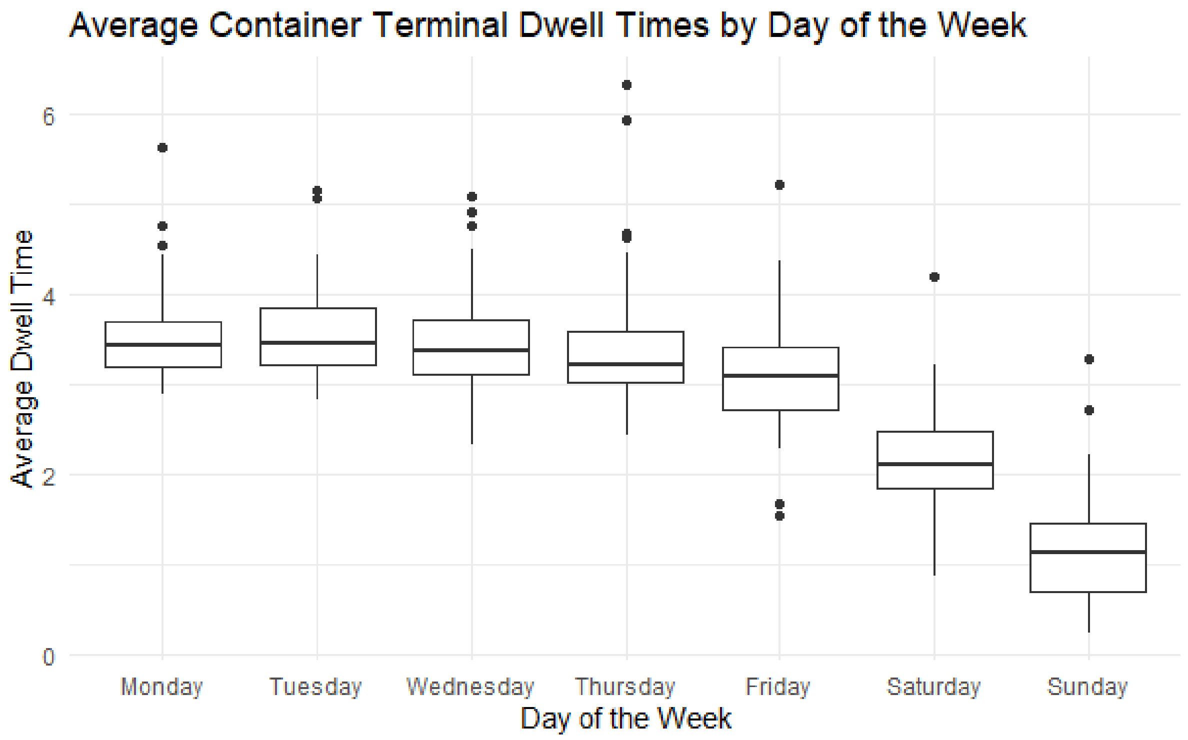
- No statistically significant difference between weekend and weekday distributions of TEUs discharged at the 5% significance level
- In contrast, there is a clear difference in gate move distributions, much higher on weekdays than weekends
- Looks as if dock-side, cranes are operating every day. Land-side, there are limited gates operating on the weekend

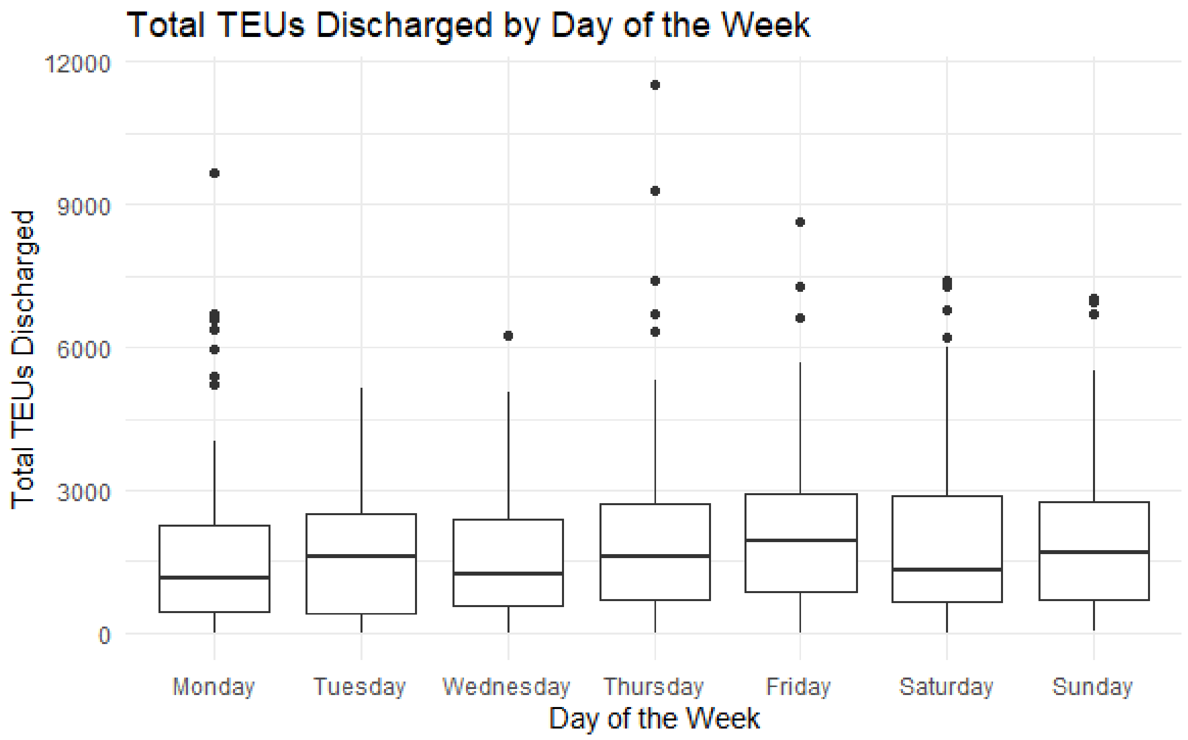




Day of Week Effects

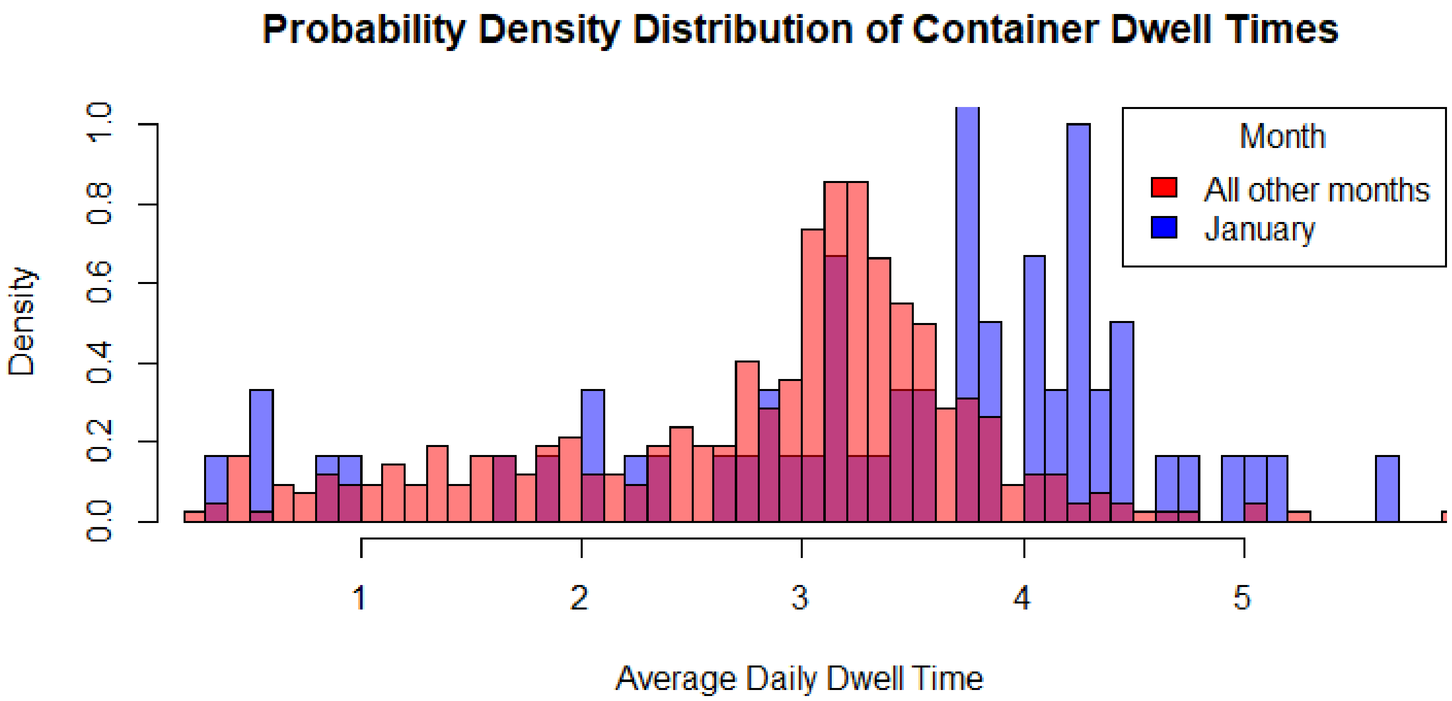
- Highest dwell times occur during the weekdays, with the lowest on weekends
- However, TEU discharge amounts are similar between weekdays and weekends
- This implies that weekday-discharged containers are typically cleared by Friday, resulting in shorter dwell times for weekend-discharged imports, often under two days.





Month Effects

- Average dwell times in January appear higher than all other months
- After conducting Mann-Whitney U Test, found a statistically significant difference between the distributions of average dwell times in January and all other months at the 5% significance level.



What Number of Lead Days is Most Correlated with Average Container Dwell Times?

Lead days = number of days before a specific date from which FLOW imports were reported to have been discharged

Nov 16, 2023

Avg dwell = 3.14 days

Date	TEUs Discharged	Lead days
Nov 16	2889.25	0
Nov 15	3686	1
Nov 14	5130	2
Nov 13	2732	3
Nov 12	6951.75	4
Nov 11	1771.25	5
Nov 10	5222.25	6

Results from Jan 16, 2023 to May 1st, 2024

- Lead days 1, 2, 7, 8, 9 stand out as having statistically significant correlation with average dwell times
- Excluding weekends significantly strengthens the correlation between lead days and average dwell

There is much stronger correlation between FLOW data and average import container dwell times when only looking at weekdays

	Including Weekends		Excluding Weekends	
# of lead days	correlation	P value	correlation	P value
0	-0.08128	0.089312	-0.19727	**0.000487**
1	-0.18618	**8.86E-05**	-0.26105	**3.30E-06**
2	-0.12165	**0.01083**	-0.18175	**0.001333**
3	0.004926	0.918116	-0.09613	0.091641
4	0.054164	0.25799	-0.06658	0.243244
5	0.003674	0.93889	-0.17435	**0.002099**
6	-0.04452	0.352654	-0.18252	**0.001271**
7	-0.11309	**0.017903**	-0.14324	**0.011709**
8	-0.14606	**0.00218**	-0.14897	**0.008723**
9	-0.09904	**0.038279**	-0.16944	**0.002808**
10	0.007378	0.877636	-0.11675	**0.04027**
11	0.034858	0.466821	-0.11916	**0.036297**
12	-0.0397	0.407154	-0.18844	**0.000872**

Model Development and Evaluation

Process

Data Preprocessing

- Clean data & remove outliers (n = 433)

Data Splitting

- Train and test data on random sample of all data by month
- Train model on 2023 data, test on 2024 data

Feature Selection

- Day of week
- TEUs discharged by lead day
- Month

Model Selection

- Multiple linear regression
- Random forest model

Model Development

Model 1a: Full Multiple Linear Regression Model

$$dwell = \beta_0 + \beta_1 TEUs_{day1} + \beta_2 TEUs_{day2} + \ldots + \beta_{13} TEUs_{day12} + \beta_6 Mon + \beta_7 Tues + \beta_8 Wed + \beta_9 Thurs + \beta_{10} Fri + \beta_{11} Sat$$

Model 1b: Subset of Multiple Linear Regression Model

$$dwell = \beta_0 + \beta_1 TEUs_{day1} + \beta_2 TEUs_{day2} + \beta_3 TEUs_{day6} + \beta_4 TEUs_{day7} + \beta_5 TEUs_{day8} + \beta_6 Mon + \beta_7 Tues + \beta_8 Wed + \beta_9 Thurs + \beta_{10} Fri + \beta_{11} Sat$$

Model 1c: Multiple Linear Regression, no lead days

$$dwell = \beta_0 + \beta_1 Mon + \beta_2 Tues + \beta_3 Wed + \beta_4 Thurs + \beta_5 Fri + \beta_6 Sat$$

Model 1d: Weighted Least Squares Regression

Assigned weights ω to all datapoints to underweight January by $\omega = \frac{1}{100\sigma^2}$ where σ is the variance of residuals in January

Model 1e: Subset of Multiple Linear Regression Model with January

$$dwell = \beta_0 + \beta_1 TEUs_{day1} + \beta_2 TEUs_{day2} + \beta_3 TEUs_{day6} + \beta_4 TEUs_{day7} + \beta_5 TEUs_{day8} + \beta_6 Mon + \beta_7 Tues + \beta_8 Wed + \beta_9 Thurs + \beta_{10} Fri + \beta_{11} Sat + \beta_{12} Jan$$

Model 2: Full Random Forest Regression

Including all covariates (TEUs discharged 1 – 12 days in advance and every day of the week)

Best Performing Model Results

Model Comparison: The multiple linear regression (MLR) model outperforms the random forest model, showing no sophisticated effects

Final Model (1e):

$$dwell = \beta_0 + \beta_1 TEUs_{day1} + \beta_2 TEUs_{day2} + \beta_3 TEUs_{day6} + \beta_4 TEUs_{day7} + \beta_5 TEUs_{day8} + \beta_6 Mon + \beta_7 Tues + \beta_8 Wed + \beta_9 Thurs + \beta_{10} Fri + \beta_{11} Sat + \beta_{12} Jan$$

Significant Predictors:

- All days of the week are most significant
- January second most significant
- TEUs discharged from lead days 1, 2, and 7 are third most significant

Model Performance: Model can effectively predict average dwell times, providing a useful measure of port congestion

- RMSE is less than half the dataset's standard deviation, indicating strong performance
- Predictions stabilize after training on 6 months worth of data

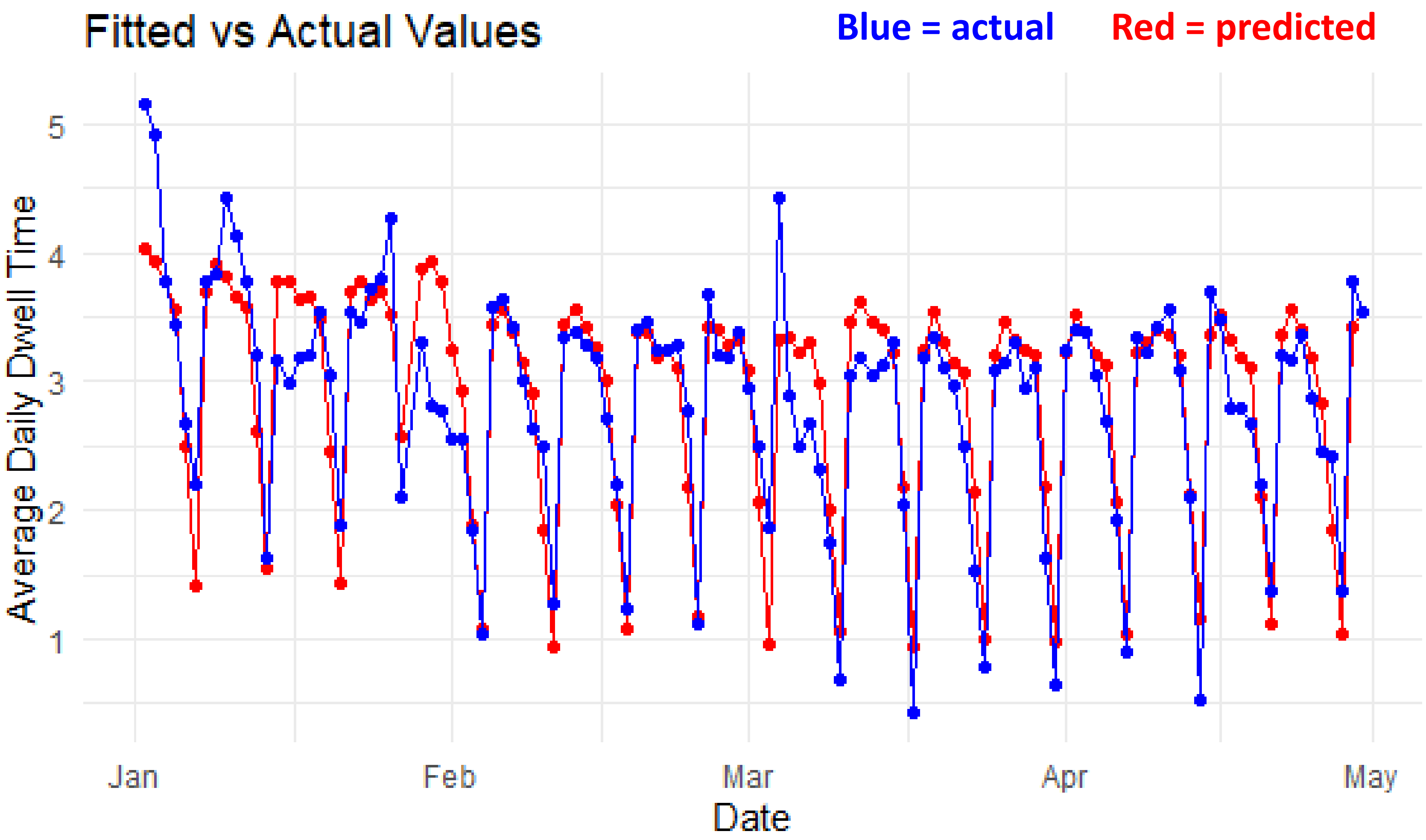
Model input:

- Day of the week
- Month
- FLOW ocean carrier bookings data

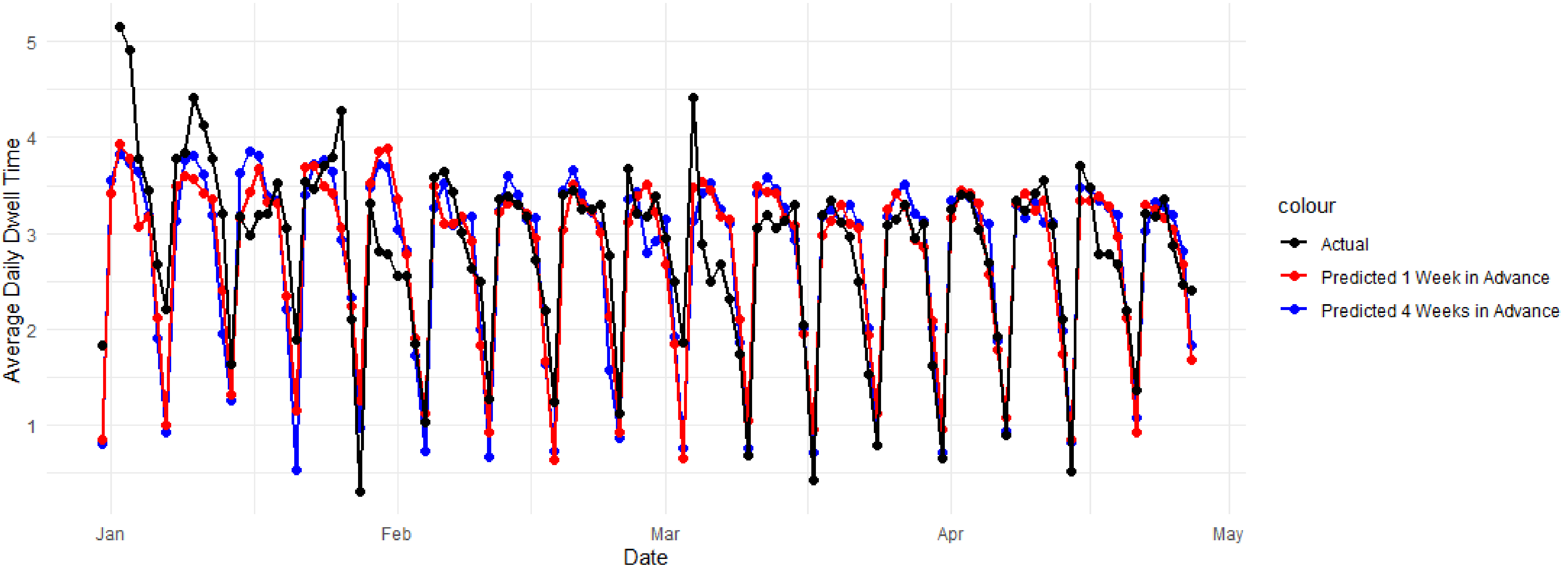
Model output:

- "on a given day, expect all import containers currently at the terminal to have been sitting on the terminal for an average of X days"

With this model, we can forecast the amount of container congestion at the terminal each day



How Far in Advance Can the Model Predict Average Dwell?



The model performs consistently well whether you predict average dwell times 1 or 4 weeks in advance, but it struggles to capture sharp fluctuations/peaks, especially in January

Key Benefits and Insights

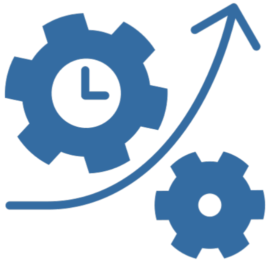
Summary of Key Insights

- Average dwell times primarily determined by day of week, with lowest average container dwell times occurring on Sunday and highest on Monday – Wednesday
- Average dwell times are significantly higher in January than in other months (by approximately 0.43 days)
- Most significant lead days to use for predicting dwell are 1, 2, and 7, but effects are very small
- Model can be used to predict average container dwell times up to 4 weeks in advance, but using FLOW ocean carrier bookings data is most beneficial when predicting 0-2 weeks in advance

Potential Model Benefits

Forecasting Terminal Congestion

- Optimize labor, equipment, and yard space
- Inform capital investments and dynamic pricing models for storage fees



Operational Improvements

- Track port performance
- Prepare for import surges with contingency plans
- Minimizing dwell times and maximizing throughput

Supply Chain Efficiency

- Align operations with trucking companies, BCOs, and ocean carriers
- Proactively alert customers about potential delays

Limitations and Future Research

Limitations

Seasonality: With only 1.5 years worth of data, adjusting for seasonality is challenging

Data Coverage: FLOW participants represent only a fraction of ocean carriers, limiting the model's accuracy in predicting total port container movements.

Historical Data: Early FLOW data is less comprehensive, restricting the model's ability to train on historical trends

Data Granularity: The model operates on daily level due to the lack of individual container-level details, which obscures individual container level details/trends

Future Assumptions: Training on 2023 data to test in 2024 assumes consistent patterns, which may not hold due to unpredictable external factors/events

Future Research

- Investigate container dwell time spikes and analyze the causes and effects
- Study on peak vs off peak trends by analyzing multiple years of data
- Study effects of additional factors such as weather conditions, economic indicators, capacity of downstream facilities, and commodities transported
- Develop similar models to predict throughput and terminal/container utilization
- Develop capacity and risk level indicators for GPA
- Apply models to other ports to compare performance with GPA



U.S. Department of Transportation

Thank you! Any Questions?

Appendix

Previous Research

- Existing studies on container dwell times predominantly rely on machine learning models with access to container-level data, enabling direct tracking throughout specific ports outside the US
- Researchers have identified several key determinants of container dwell time:
  - Container size, weight, and status
  - Discharge day/month
  - Vessel's origin port
  - Commodities transported
- However, none of these studies utilize forward-looking data on imports discharged to predict daily average container terminal dwell times

Data Collection

Ocean Carrier Bookings: Provided by ocean carriers, includes reported data on the volume of imports (in TEUs) discharged each date at the port of SAV by container size. This data tells us about the expected number of import containers being unloaded onto the SAV port each day.

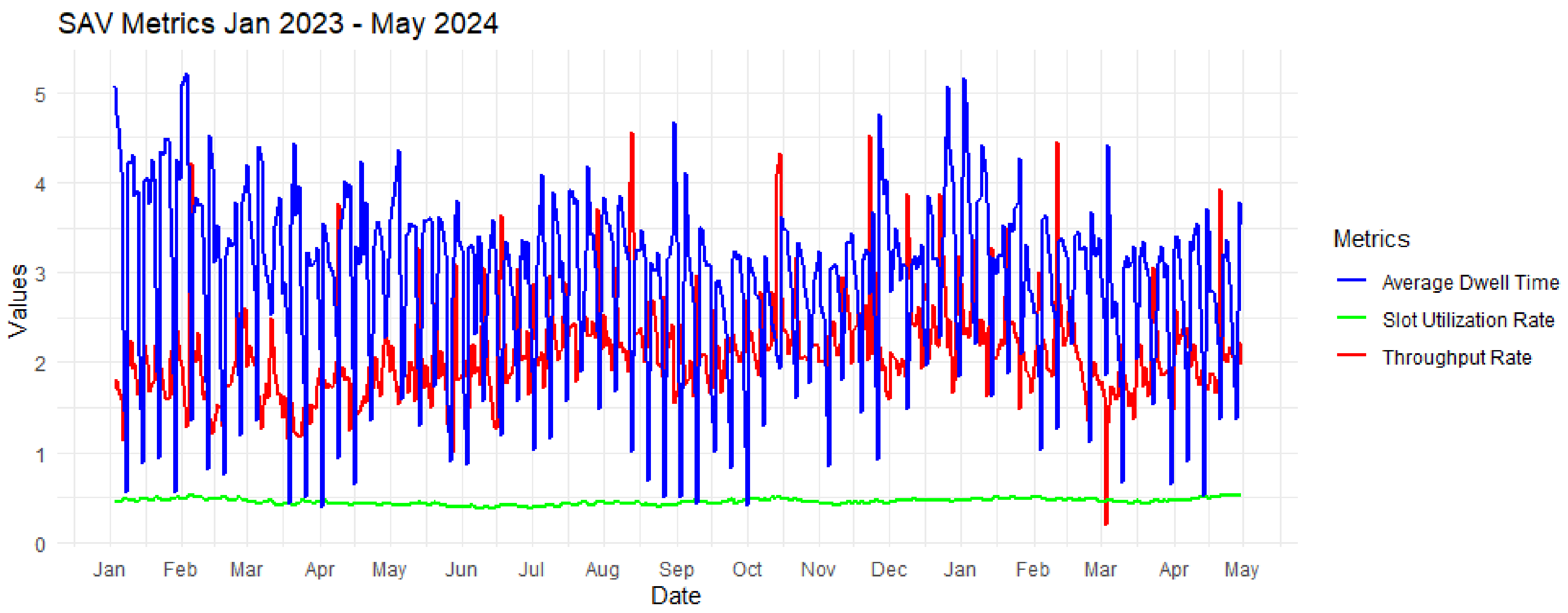
Marine Terminal Utilization: Data on total number of import, export, and empty container slots being used in the terminal, in TEUs, broken down by container type, size, and movement, as well as the terminal's total usable container slots.

Gate Moves/Throughput: Data on the number of containers loaded in (exports being brought onto the terminal to be shipped out) and loaded out (both loaded and empty containers leaving the terminal) each day at the port of SAV.

GPA Average Dwell Times: Data from Jan 1, 2023 to May 1, 2024 on the average daily import container dwell times, calculated as the mean of the dwell times of all import containers sitting on the Port of SAV terminal each day

Correlation with Slot Utilization and Throughput Rates

- Average dwell time, slot utilization, and throughput rate (containers loaded out/loaded in) show minimal correlation with each other
- Average dwell time has a weak negative correlation (r = -0.17) with throughput rate, as higher throughput is associated with more container movement and lower dwell times
- Average dwell time shows almost no correlation with slot utilization (r = -0.04), likely due to consistent slot utilization around 0.5



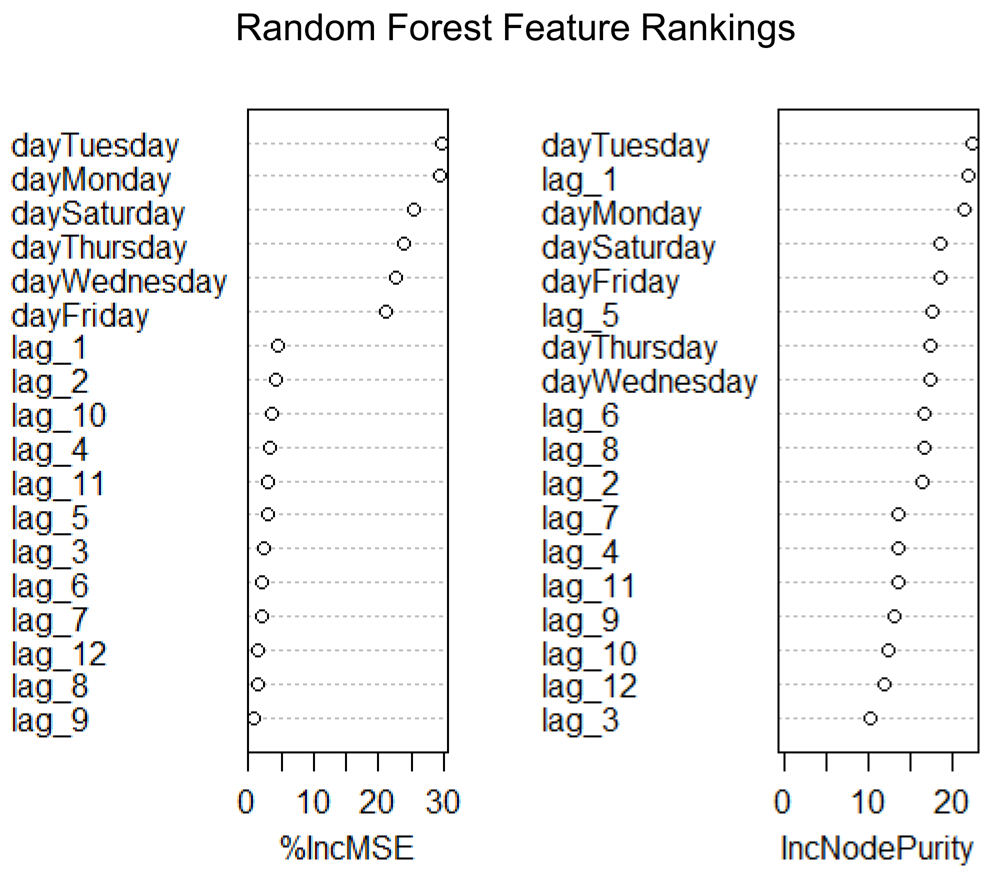
Splitting the Data

- I examined all data points from Jan 1, 2023 to May 1, 2024
  - 441 total entries
  - 433 entries in cleaned dataset after removing outliers (z >= 2.5)

- After grouping by month to account for seasonality effects, I randomly split the dataset into training and test sets
  - 346 observations in training data
  - 87 observations in test data

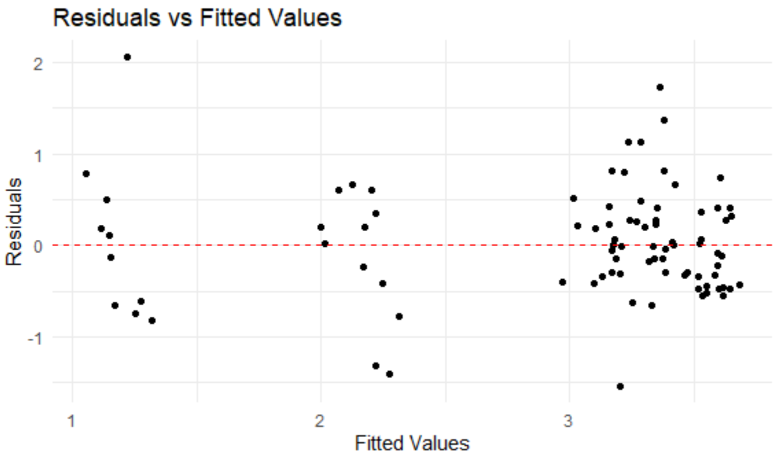
Feature Selection

- Correlation analysis
  - Imports discharged **1, 2**, and **6-9 days** before tend to be significantly correlated with average dwell times
- Lasso regression
  - **Every weekday deemed significant**
  - Every lead day except lead 4 was deemed relevant by the lasso regression, but most important lead days were **1, 2, 6, 7, 8, 10, and 12**
- Random Forest
  - Every weekday deemed significant, with **Tuesday, Monday, and Saturday** being the most predictive
  - Most important days are **1, 2, 5, 6, 7**

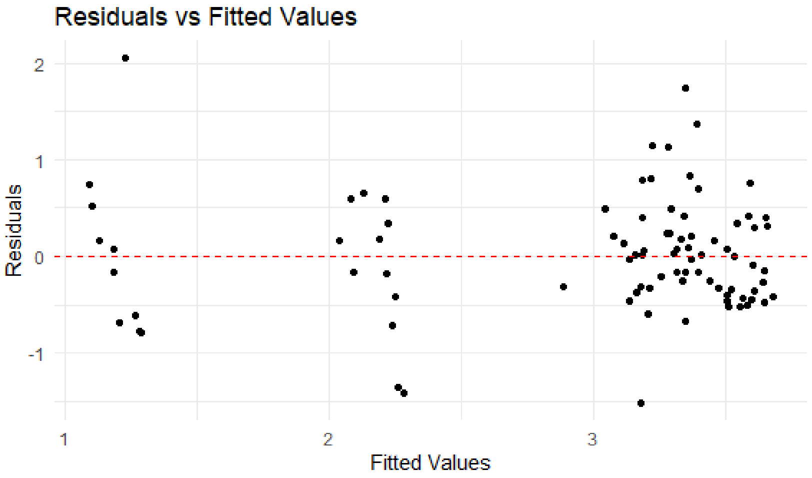


Multiple Linear Regression Models 1a-1c Performance

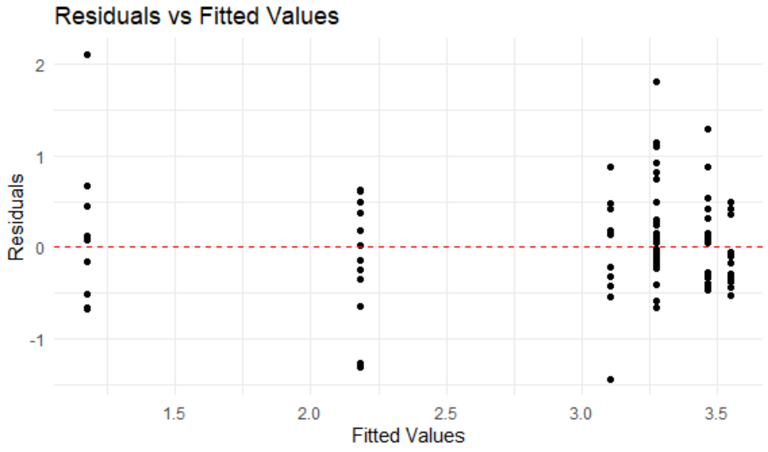
Full Multiple Linear Regression (1A)



Subset of Multiple Linear Regression (1B)



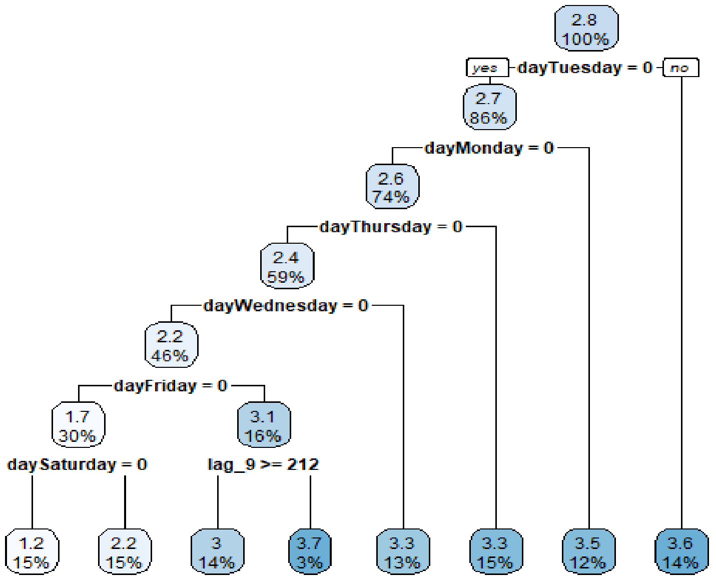
Multiple Linear Regression, No Lead Days (1C)

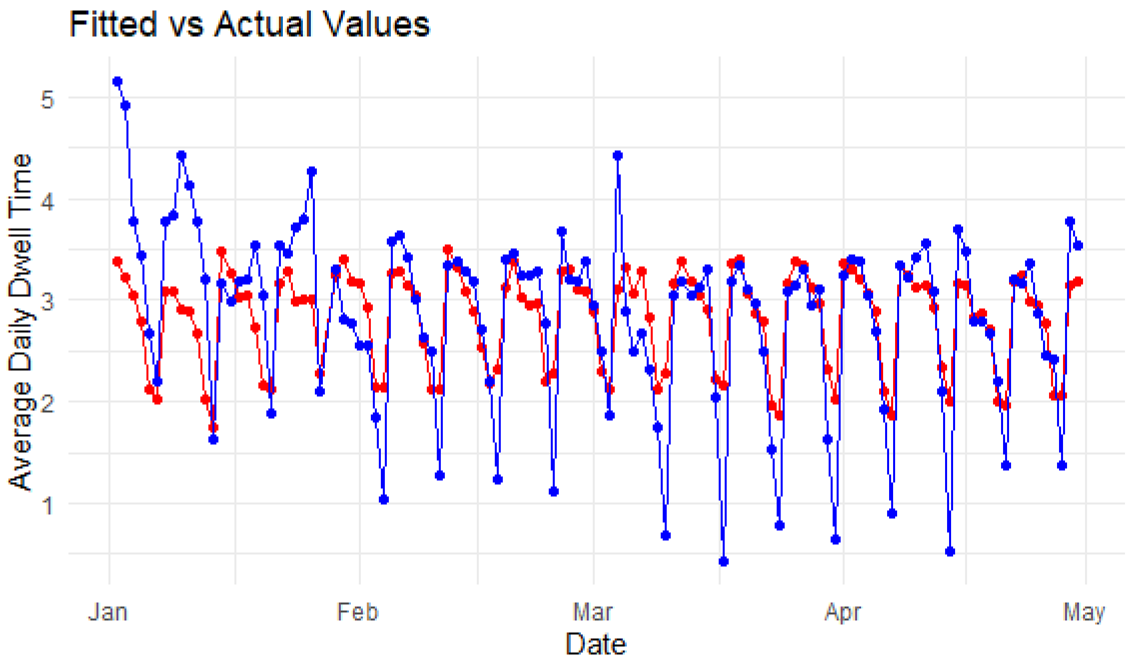


- All MLR models performed very similar to each other, with R squared around 0.75, p value < 2.2e10-16, and test RMSE around 0.6
- Every day of the week was deemed statistically significant by all models
- Only TEUs from lead day 1 were deemed as statistically significant in models 1a and 1b
- Model results not affected by removing data on import TEUs discharged by lead day

Random Forest Model Performance

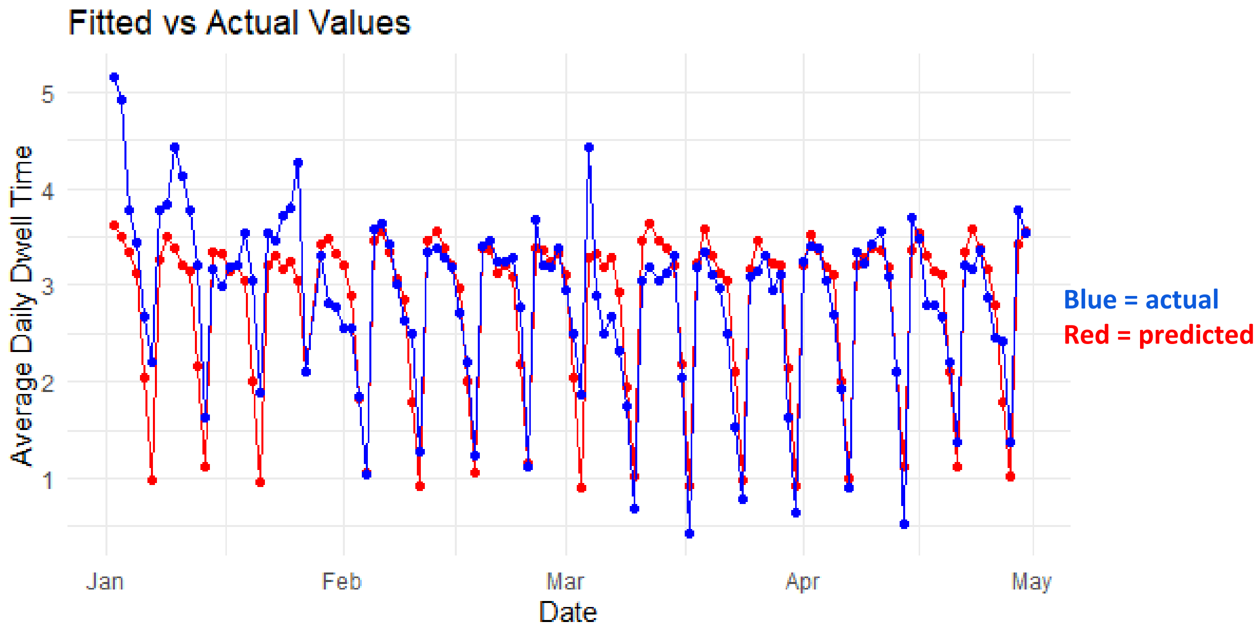
- Random forest model performed much worse than MLR models, likely due to overfitting
  - Tends to overestimate values from 2 - 2 and underestimate values above 2.5
  - Tends to underpredict weekdays (particularly Wednesdays) and over predicting weekends (particularly Sundays)
  - Variance of predictions (0.264) much lower than actual variance (0.978)
  - When training on 2023 data and testing on 2024 data, RMSE = 1.263

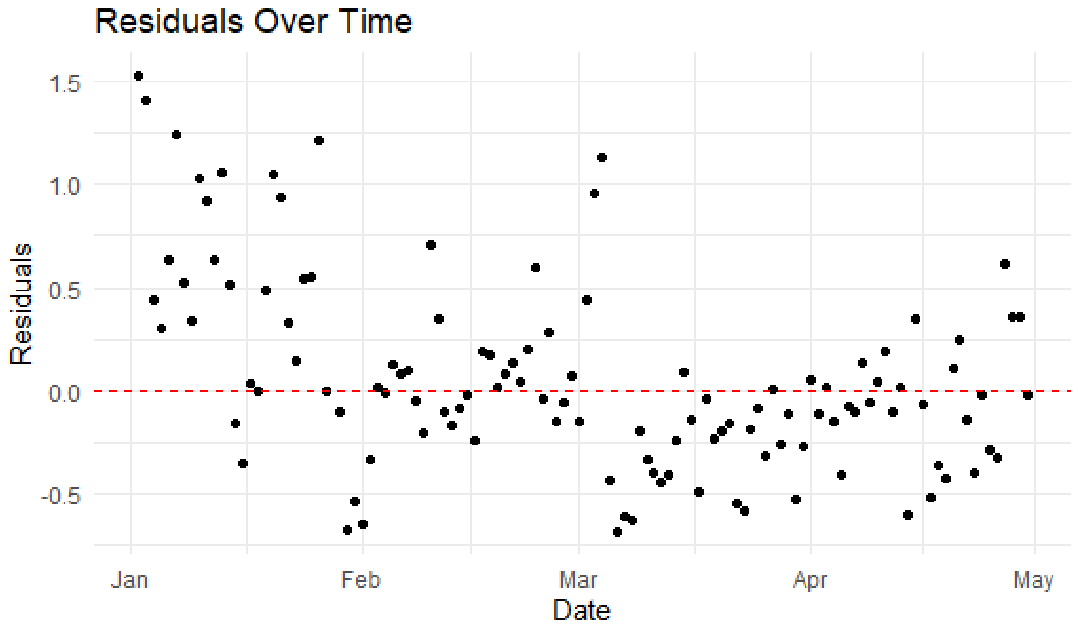




Addressing Heteroskedasticity

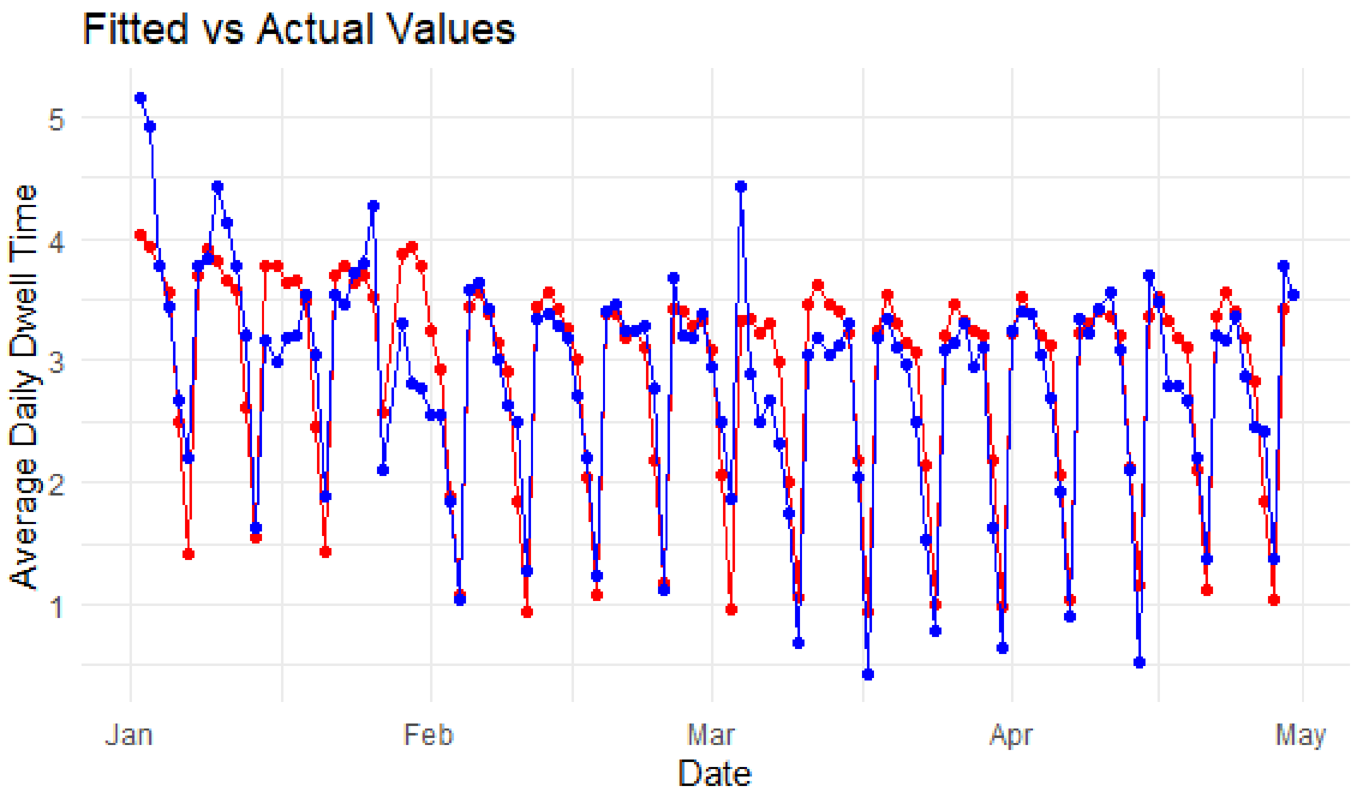
- When training on 2023 data and testing on 2024 data, model 1b outperformed all the others (test RMSE = 0.472)
- However, heteroskedasticity appeared present in residuals, as model tends to underpredict > overpredict, especially for the month of January
- Weighted Least Squares regression did not improve accuracy of predictions nor fix heteroskedasticity
- January is most likely a different beast





Best Performing Model

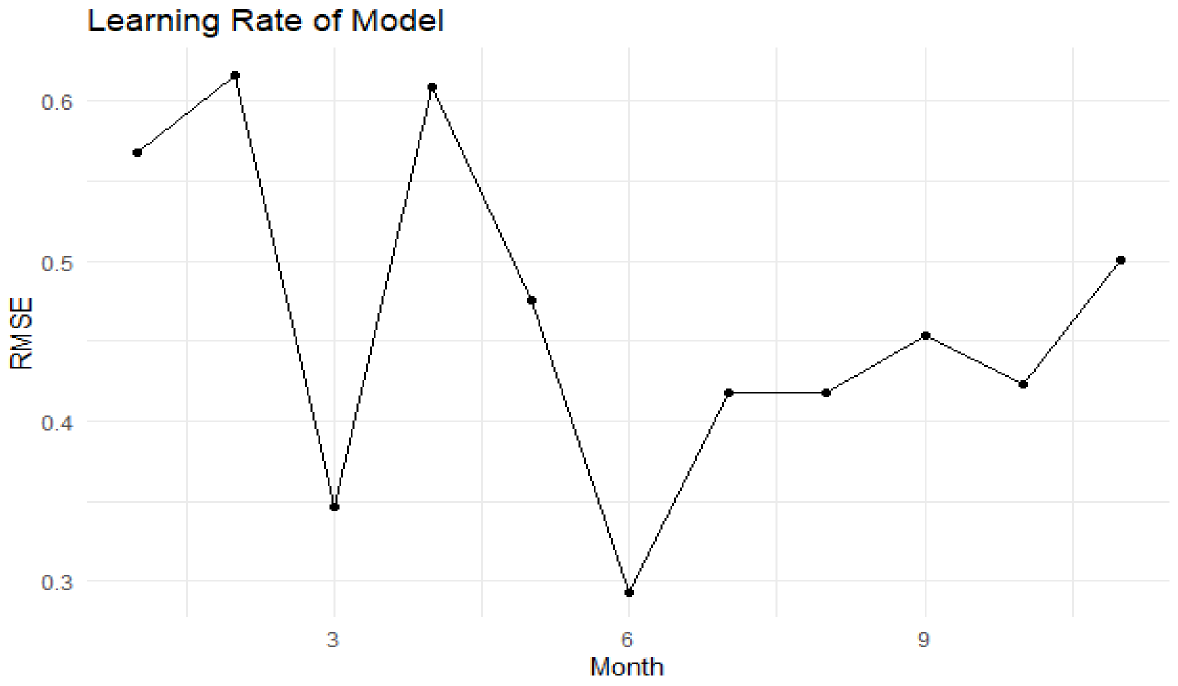
- By adding January, model performance slightly improved
  - Effect of January is statistically significant
  - All days of the week remain significant
- Model does fairly good job at predicting average dwell times of all containers sitting in terminal each day
  - Test RMSE is < ½ standard deviation of the full dataset!



Blue = actual
Red = predicted

Model Validation

Month	RMSE
1	0.7420904
2	0.6018447
3	0.4396970
4	0.3335381
5	0.5841114
6	0.4143524
7	0.2789263
8	0.4302466
9	0.4160440
10	0.4686363
11	0.4026001
12	0.5011722



Leave One Out Cross Validation: Leave one month out at a time, train model on remaining data, test on the left-out month

- Average error = 0.47

Learning Rate of Model: Train model on first n months, evaluate on n+1 month, repeat for all months

- From month 6 onwards, the model starts to stabilize

Comparing Lead Times of 1-4 Weeks for the Week of March 17 – 24, 2024

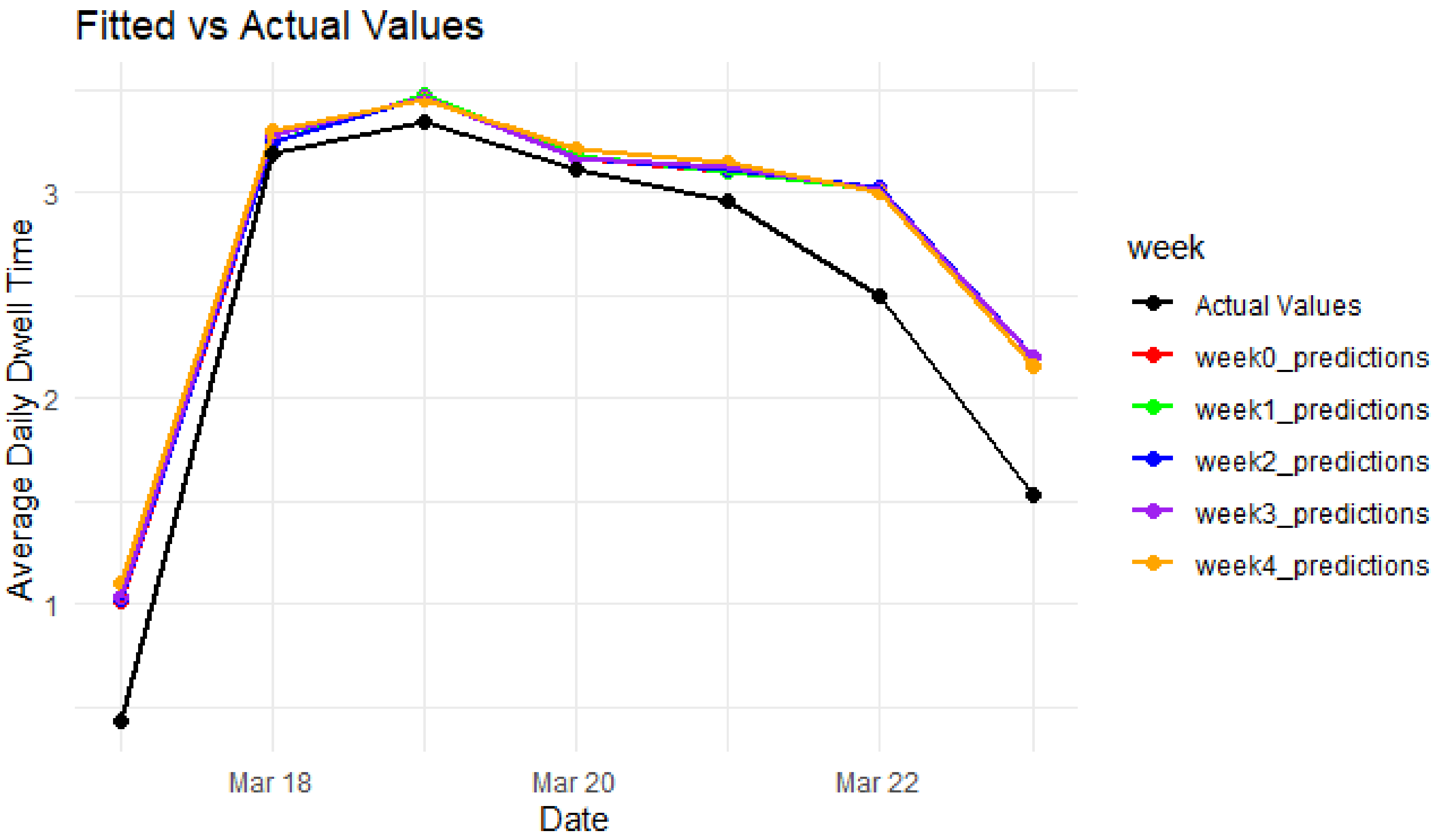
Lead Time	Description	Statistically significant lead days	Test RMSE
0 Weeks Before	As of March 17, how well does model 1e, trained on bookings and dwell data from 1/13/23 to 3/17/24, predict average dwell times for the following week from the forward-looking bookings data for that week?	1, 2, 7	0.397
1 Week Before	As of March 10, how well does model 1e, trained on bookings and dwell data from 1/13/23 to 3/10/24, predict average dwell times for the week of 3/17 to 3/24 from the forward-looking bookings data for that week?	1, 2, 7	0.400
2 Weeks Before	As of March 3, how well does model 1e, trained on bookings and dwell data from 1/13/23 to 3/03/24, predict average dwell times for the week of 3/17 to 3/24 from the forward-looking bookings data for that week?	1, 2, 7	0.40027
3 Weeks Before	As of Feb 25, how well does model 1e, trained on bookings and dwell data from 1/13/23 to 2/25/24, predict average dwell times for the week of 3/17 to 3/24 from the forward-looking bookings data for that week?	1	0.4038
4 Weeks Before	As of Feb 18, how well does model 1e, trained on bookings and dwell data from 1/13/23 to 2/18/24, predict average dwell times for the week of 3/17 to 3/24 from the forward-looking bookings data for that week?	none	0.4087

Reference date

	Sun	Mon	Tues	Wed	Thurs	Fri	Sat
Predict 4 weeks in advance	2/18	2/19	2/20	2/21	2/22	2/23	2/24
Predict 3 weeks in advance	2/25	2/26	2/27	2/28	2/29	3/1	3/2
Predict 2 weeks in advance	3/3	3/4	3/5	3/6	3/7	3/8	3/9
Predict 1 week in advance	3/10	3/11	3/12	3/13	3/14	3/15	3/16
Predict 0 weeks in advance	3/17	3/18	3/19	3/20	3/21	3/22	3/23

Week for which you are predicting average dwell times

- Model's predictions on the week of march 17 to 24 are very similar, no matter which time/week of FLOW data you train the model on
- The model tends to overpredict average dwell times, especially in later days of the week

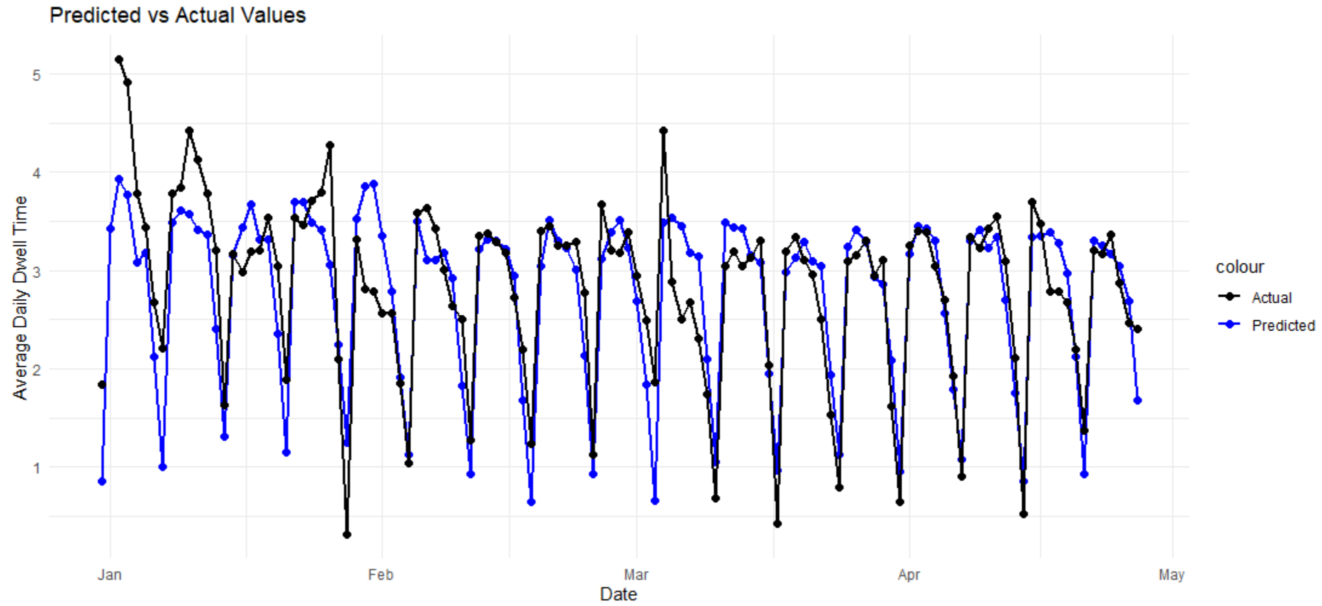


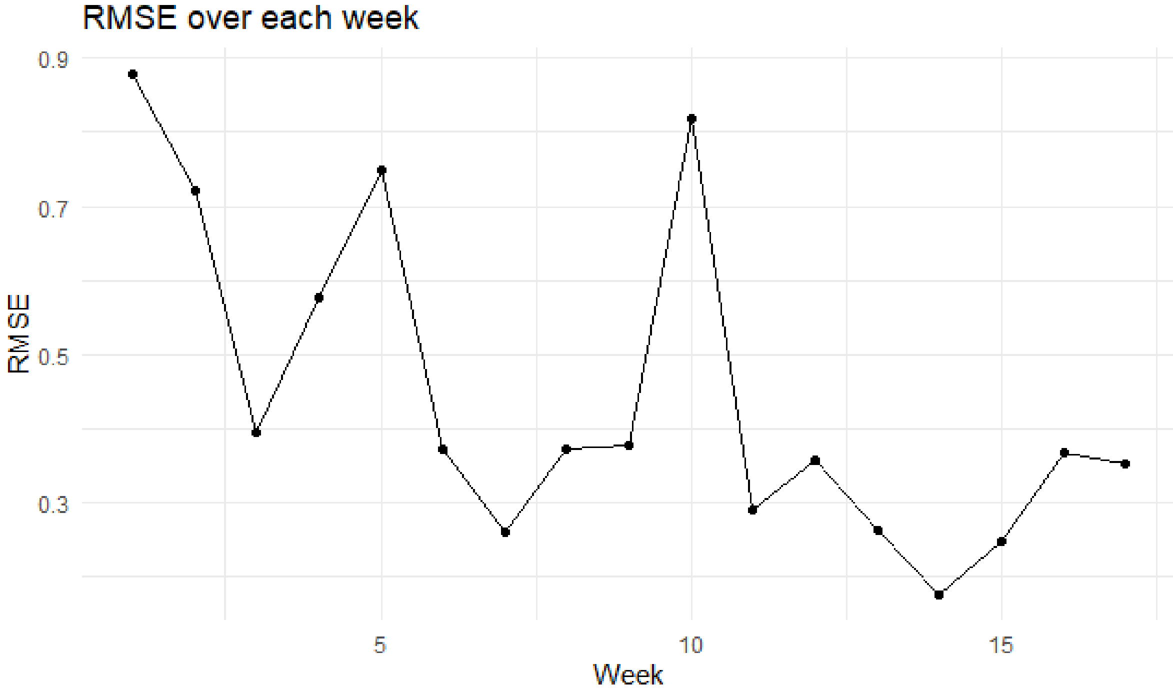
All weeks from Jan 1, 2024 to April 27, 2024

For each week from January 1 to April 27, 2024, predict on Sunday (reference date) by training model on all data up to that date, and forecast the average dwell times from the prediction start date to prediction end date. Repeat this process using FLOW data from 1 week ahead and 4 weeks ahead.

Week #	Prediction Start Date	Prediction end date	Reference date (1 week before)	Reference date (4 weeks before)
1	12/31/23	1/6/2024	12/24/2023	12/3/2023
2	1/7/24	1/13/2024	12/31/2023	12/10/2023
3	1/14/24	1/20/2024	1/7/2024	12/17/2023
4	1/21/2024	1/27/2024	1/14/2024	12/24/2023
5	1/28/2024	2/3/2024	1/21/2024	12/31/2023
6	2/4/2024	2/10/2024	1/28/2024	1/7/2024
7	2/11/2024	2/17/2024	2/4/2024	1/14/2024
8	2/18/2024	2/24/2024	2/11/2024	1/21/2024
9	2/25/2024	3/2/2024	2/18/2024	1/28/2024
10	3/3/2024	3/9/2024	2/25/2024	2/4/2024
11	3/10/2024	3/16/2024	3/3/2024	2/11/2024
12	3/17/2024	3/23/2024	3/10/2024	2/18/2024
13	3/24/2024	3/30/2024	3/17/2024	2/25/2024
14	3/31/24	4/6/2024	3/24/2024	3/3/2024
15	4/7/24	4/13/2024	3/31/2024	3/10/2024
16	4/14/24	4/20/2024	4/7/2024	3/17/2024
17	4/21/24	4/27/2024	4/14/2024	3/24/2024

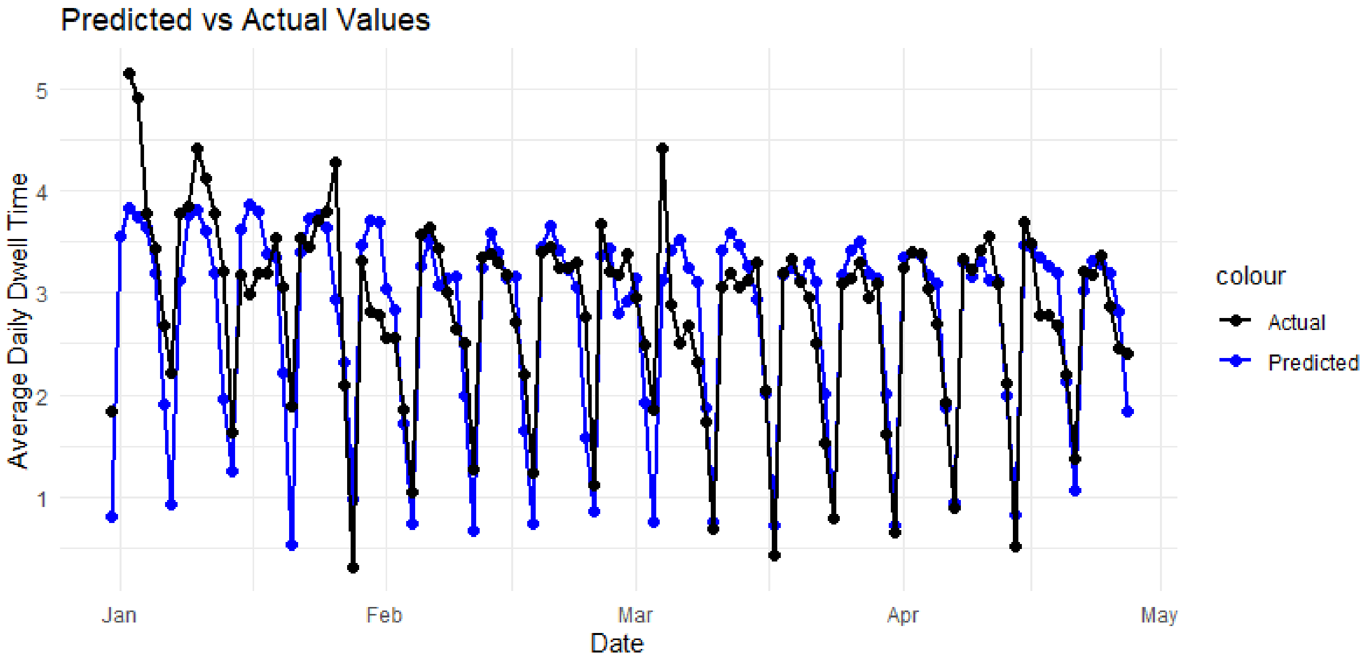
Results 1 Week in Advance

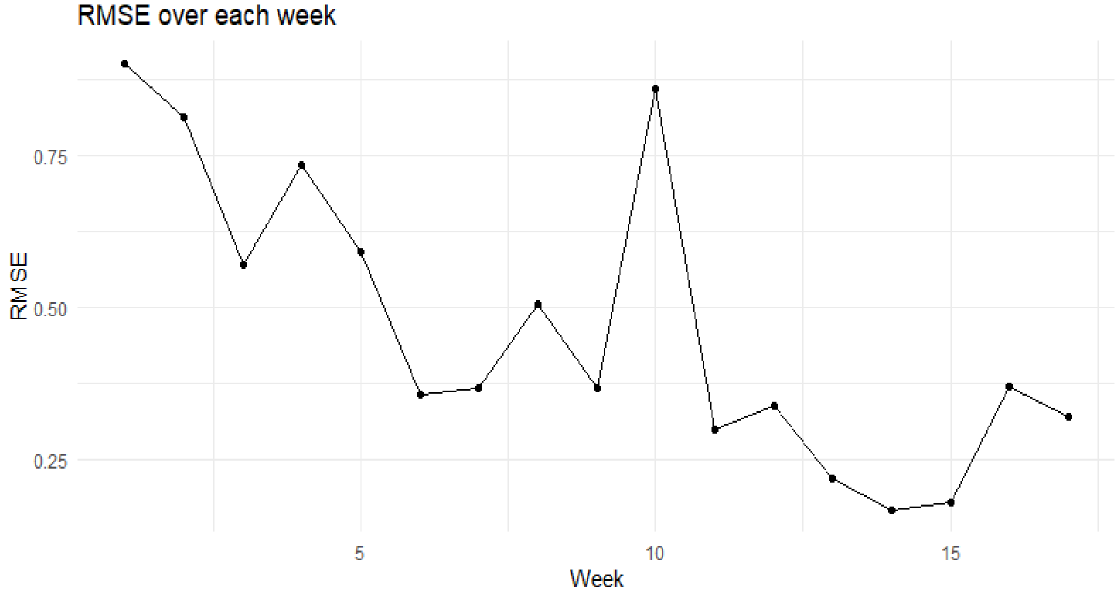




- Average RMSE = 0.445
- Model stabilizes after 10 weeks
- Model underestimates dwell times 45.8% of the time, overestimates 54.2% of the time

Results 4 Weeks in Advance





- Average RMSE = 0.468, slightly higher
- Learning rate of model appears the same
- Model slightly more likely to overestimate
  - underestimates dwell times 44.9% of the time, overestimates 55% of the time

Day of Week Effects

- Average dwell times primarily determined by day of week effects
- Lowest average container dwell times typically occur on the weekends, as the terminal clears out during weekdays
- Compared to Sunday, every other day typically increases average dwell times by around 2 days, with Monday – Wednesday having the highest effect

Month Effects

- We observed strong month effects for January
  - Average dwell times are significantly higher in January than in other months ($p < 0.05$)
  - January is associated with an increase in average dwell time of approximately 0.43 days
- High average dwell times in early January are likely due to many people taking 1-2 weeks off during the holiday season, leading to delays when the new year begins

Lead Day Effects

- Most significant lead days to use for predicting dwell are 1, 2, and 7, but effects are very small (in order of 10^{-5})
    - Interestingly, TEUs discharged 1 or 2 days prior are linked to slightly reduced dwell times, while those discharged 6-8 days beforehand tend to extend dwell times slightly
- Overall, the hypothesis that FLOW participant TEUs discharged by lead days can predict dwell times does not appear very promising

Model's Forward-looking Prediction Power

- Minimal difference in prediction accuracy when forecasting a week's worth of average dwell times from 0 to 4 weeks in advance
- Using FLOW ocean carrier bookings data is most beneficial when predicting 0-2 weeks in advance
- As it accumulates more data each week, the model takes about 10 weeks to stabilize its predictions
- Model tends to overpredict average dwell times (55%), especially in later days of the week
- Model struggles to capture sharp fluctuations/peaks, especially in January