Results Summary and Business Recommendations

An analysis of Items 45, 5, and 9 reveals three distinct demand profiles with varying sensitivity to weather conditions. These items were selected due to their contrasting sales responses observed during exploratory analysis. Item 45 demonstrates the strongest relationship with weather variables, with an R^2 of approximately 40%, indicating a clear negative correlation with stable, high-pressure atmospheric systems. Item 5 shows moderate but non-linear sensitivity to weather ($R^2 = 0.173$), responding more to temperature (+8.45% coefficient) and seasonal daylight changes. Conversely, Item 9 exhibits negligible sensitivity to daily weather fluctuations (adjusted $R^2 < 4\%$), suggesting its demand is primarily driven by seasonal or non-weather-related factors.

Model Performance

Item 45

Item 45 exhibits the strongest weather dependency of the group, with the Random Forest model explaining 39.8% of the variance in sales—a significant improvement over the Linear Model's 28.2% R^2. The Random Forest model identifies 'Station Pressure' (stnpressure) as the single most dominant predictor, with its exclusion from the Linear model causing a massive 201.5% increase in Mean Squared Error (MSE). This indicates that Item 45 is likely storm driven; demand surges when low-pressure systems (associated with approaching storms) move in, while stable high pressure suppresses sales. This is evidenced by the significant negative coefficient of -34.62 in the linear model. Temperature and daylight (sunrise_mfm) contribute to steady linear growth, atmospheric pressure drives the most significant shifts in consumer demand. Furthermore, the Linear Model puts in perspective the impact of warm weather, showing units sold rising by 18.27 units for every unit increase in average temperature.

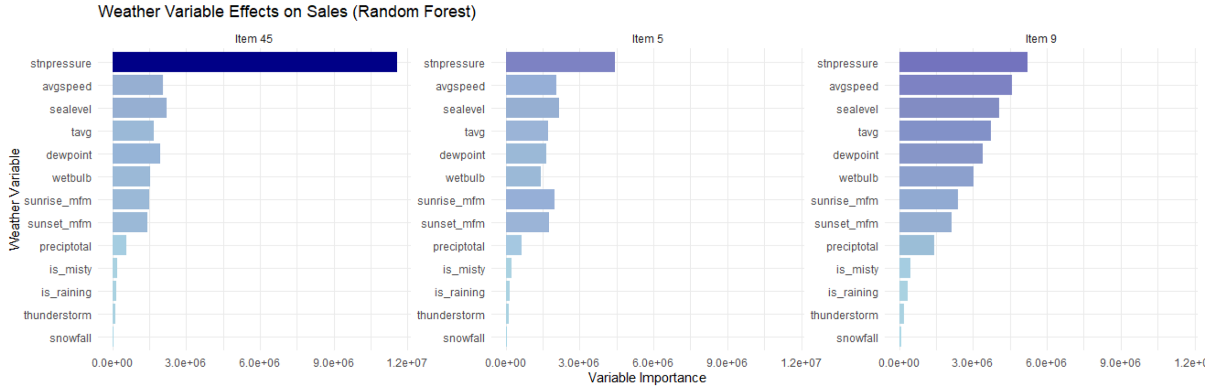
Item 5

Item 5 displays a volatile demand profile where station pressure and rain is a primary suppressor of sales. While the linear model captures only modest variance (Adj. R2 = 6.7%), the Random Forest reveals that sales are governed by non-linear atmospheric shifts, increasing the variance to 17.3%. The RandomForest identifies 'Station Pressure' and 'Daylight' (sunrise_mfm) ranking as critical variables (3.7 million and 2.2 million IncNodePurity, respectively).
The negative impact of precipitation is significant; rain events correlate with drops in sales volume, evidenced by the Linear Model's negative coefficient of -2.43%. This suggests Item 5 is likely an outdoor-use or impulse product where wet weather and shorter daylight hours physically deter purchase or usage, making precipitation forecasts a vital signal for inventory management.

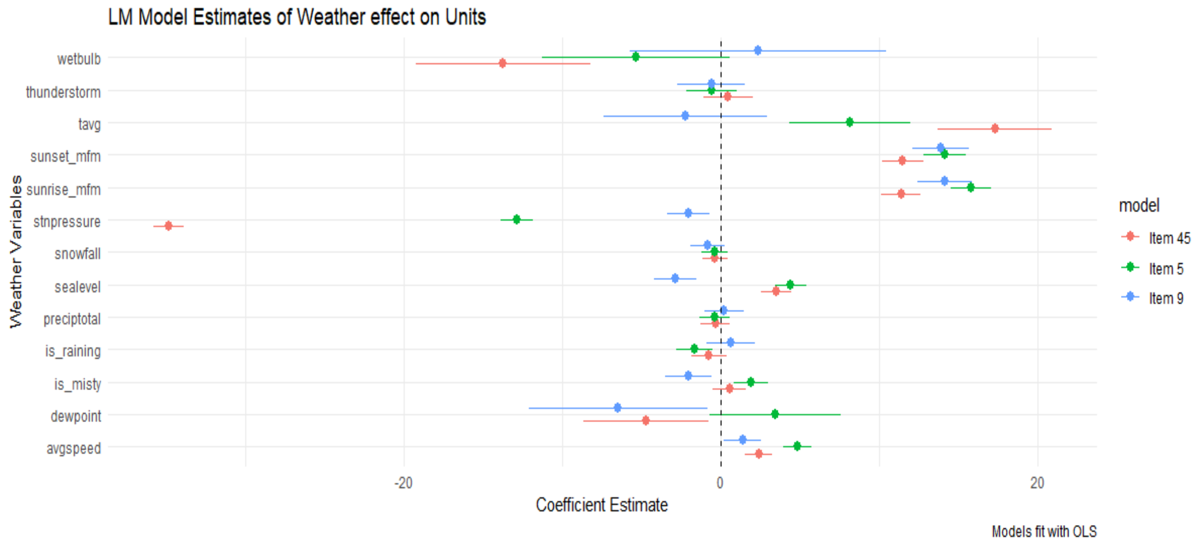
Item 9

Both models for this Item present it as a special case, as comparatively, Item 9 demonstrates - by a significant margin - the weakest link to daily meteorological conditions, with the RandomForest explaining 3.8% of the variance, which is very low in comparison to the other explored items. The high RMSE of 42.97 reflects unexplained variance. The Linear Model captures the items' lack of sensitivity to weather variables, for example "is_raining" and "tavg", that have small coefficients and insignificant p-values (0.82 and 0.18), showing no reliable impact on sales. The lack of strong daily correlation suggests that sales for Item 9 are driven almost entirely by broader seasonal cycles or non-weather factors (such as essential necessity or fixed consumption habits) rather than the immediate daily forecast.

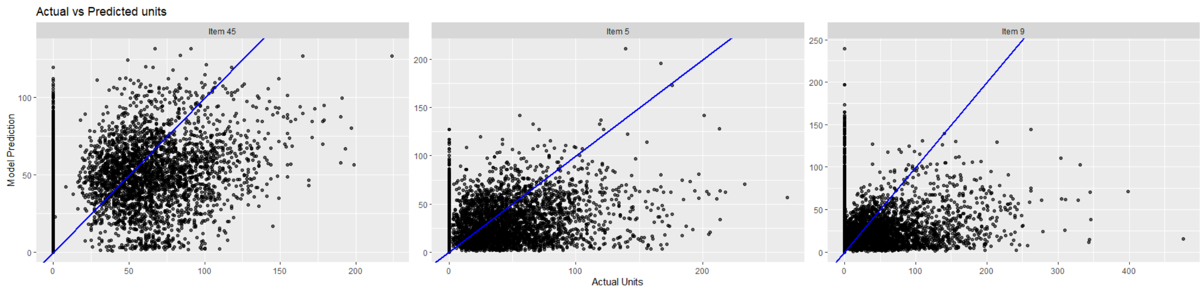
Model Visualisations



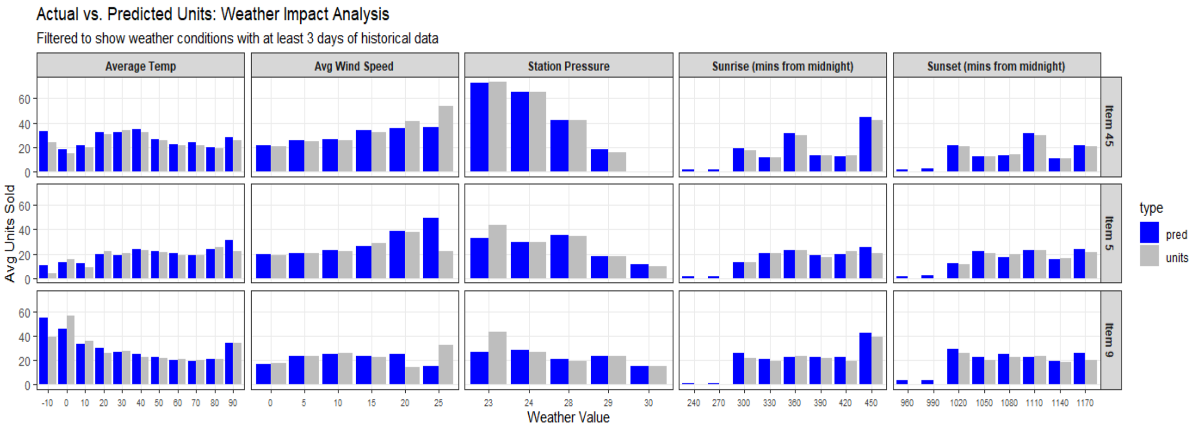
This Visualisation shows the importance of the weather variables for each item. This was used during the model testing phase to see initial results from the Random Forest Models. Station Pressure for each item consistently remains the most important with all models and items yielding a near identical order of importance for the variables.



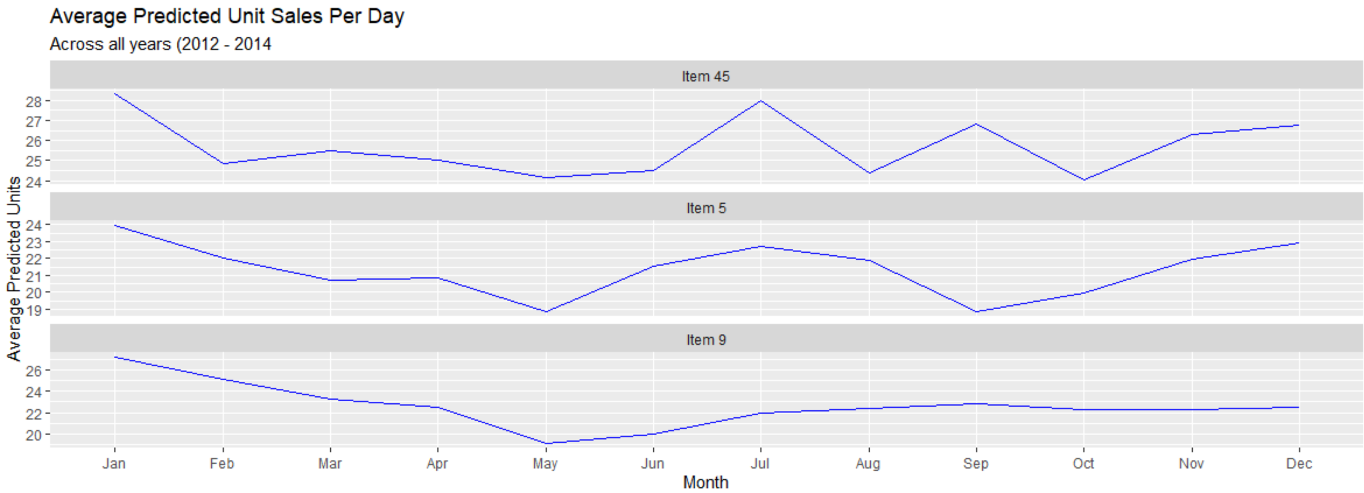
The above image visualises the Coefficient Estimates for each weather variable, this time using the models fitted with OLS (Ordinary Least Squares). Item 45 shows the strongest weather sensitivity with station pressure and snowfall having meaningful effects, indicating shifts in demand in extreme conditions. Item 5 has several moderate coefficients but is noticeably less predictable than item 45. Item 9 in contrast, has its coefficients clustered around zero, reinforcing the observation that daily weather has no impact on sales.



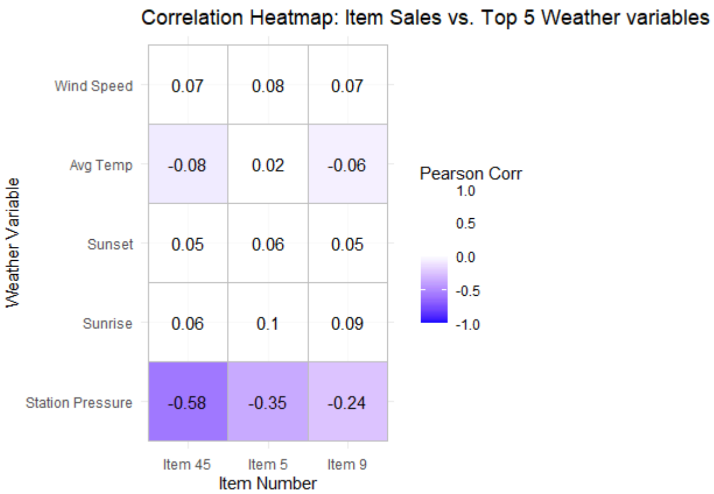
This Visualisation is a scatter graph of the ‘Actual vs Predicted units’, using the Random Forest models. Showing if weather has a positive or negative relationship with unit sales. Item 45 shows the strongest alignment, indicating that weather variables explain a large portion of its sales, additionally Item 5 shows weaker and more dispersed predictions and finally Item 9 shows little alignment at all, further indicating weather’s lack of predictive power for its demand.



The Multiple bar charts above illustrate the relationship between weather thresholds and average sales by grouping continuous data into discrete bins, with blue bins as the predictions, and grey as the actual units sold. Station pressure for Item 45 and Item 5 show a negative correlation with higher pressure seemingly lower sales, with Item 9 showing a similar result but with a brief spike in predicted and actual sales in the middle. Moreover, Item 9’s data points further show a lack of predictive power, with the predictions stuck in a narrow range. Item 5 also shows sensitivity to high speed winds, where sales outperformed predictions during high winds. Item 9 remains largely unaffected by weather variables.



The above plot is a line chart, depicting each items sales predictions for each month. The aim of the visualisation is to view each items' sales trends throughout the year as predicted by the weather variables. Item 45 - along with all three items - has an extremely strong average daily units sold in January (28), with a sharp drop to February and not peaking again until summertime whereafter the sales fluctuate drastically. Item 5's predictions are slightly more gradual with a notably sustained high average sales record maintained during peak summertime, with other peaks appearing at the beginning and end of the year. This could imply that the item is related to term time, perhaps school related, as its peak and sustained time frames take place near/during school holidays. Item 9 differs, the y-axis fluctuates between roughly 19 and 25 units, which is a much narrower range of movement in comparison to the others. Implying that weather does impact slightly but it's less easily swayed by hot days and pressure shifts.



A heatmap, showing the correlation between weather variables and item numbers. The weather variables were chosen as a result of the data modelling and importance plots.

The overwhelmingly dominant variable remains "Station Pressure", with negative correlations for all items (-0.58 Item 45, -0.35 Item 5 and -0.24 for item 9). Notably, item 45 is twice as sensitive to pressure changes as item 9. Weak drivers are outline, with "Wind Speed", "Sunrise" and "Sunset" having very low coefficients. This could imply that time of the year may be a factor, however specific daylight hours don't drive sales as much as a atmospheric pressure does. A key limitation of this heatmap is that it only looks at one variable at a time, the heat map is unable to see any work between the variables but the models and model visualisation can.

Business Insights

Item 45:

The binned impact analysis and coefficient plots show a significant surge in sales during periods of low atmospheric pressure, this suggests item 45 serves a critical need during unstable weather.

- ➢ The business should align their deliveries of inventory with short-term forecasts of station-pressure. Ensuring full shelf availability during these volatile windows will allow the retailer to meet the documented spikes in consumer demand that occur immediately before and during atmospheric shifts.

Item 5:

This Item shows a positive correlation with wind speed and daylight hours. The bar charts show sales nearly doubling during peak wind events compared to calmer days.

- ➢ One key recommendation would be to increase store-level stock during high wind, to capture the surge in demand.
- ➢ Secondly, due to its sensitivity to daylight hours, the item should be placed in high traffic zones during the summer due to the longer daylight hours, but conversely, scale back during winter to avoid waste.
- ➢ Finally, due to its sensitivity to wet weather, consider scaling back during periods of wet weather or when rain is forecast to lessen wasted stock and/or space.

Item 9:

Both the scatter plots, binned charts as well as performance insights from the initial testing phase show almost no variation in sales across different weather conditions, which is in stark contrast to the other two items discussed above.

- ➢ Due to the lack of evidence that weather plays a pivotal role in this items units sold, going forward I’d suggest removing Item 9 from any weather based replenishment logic to save operational time and effort wasted.
- ➢ Use seasonal standard baseline for ordering as the demand is likely to be driven by necessity or long-term habits rather than daily forecast.