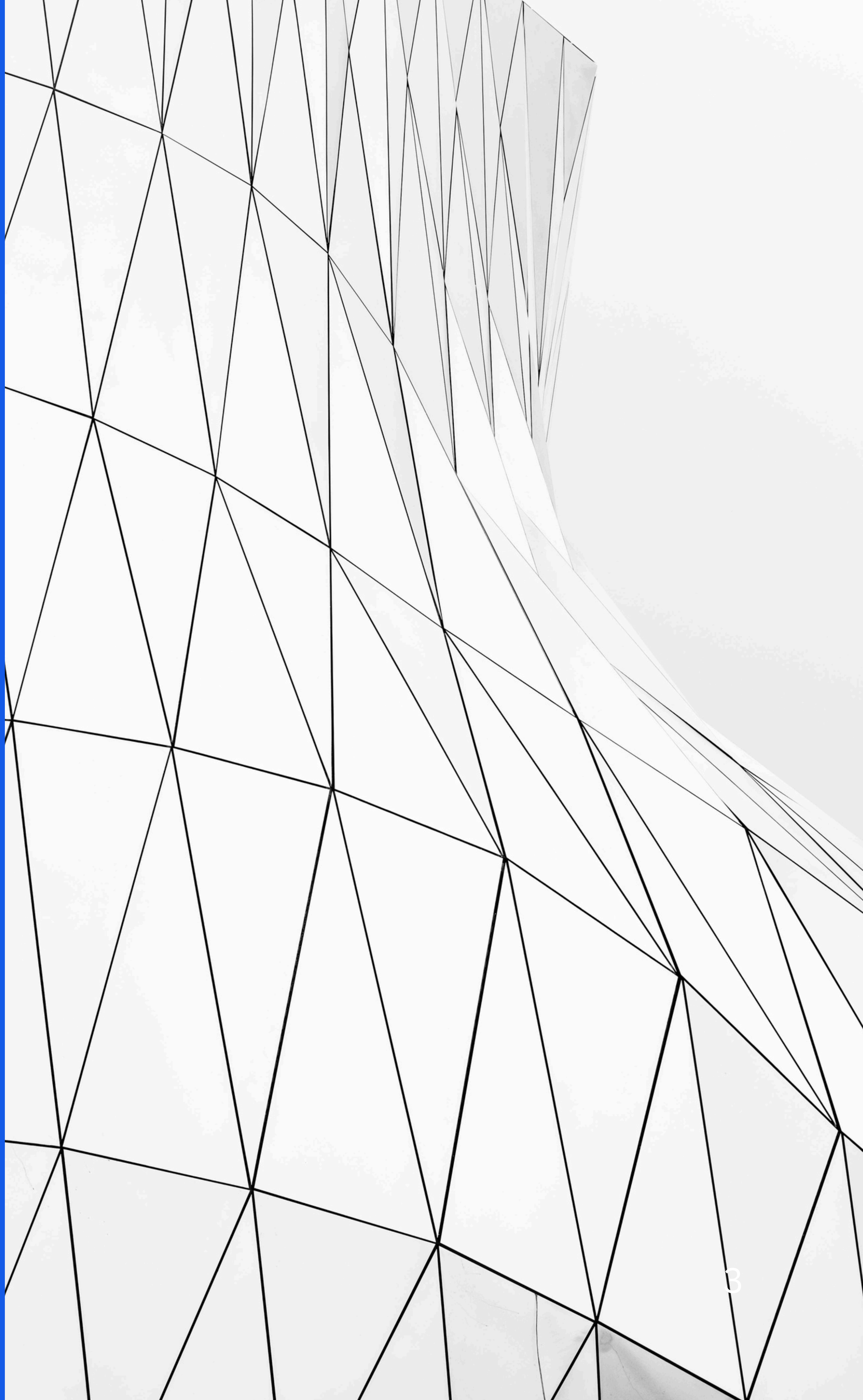
Team 7

Walmart
Final Presentation

Bei-Yu Xing
Bo-Syuan Peng
I-Tso Chen
Pin-Chen Chi
Shuo-An Wu

Executive Summary



Executive Summary

- Walmart aims to improve forecasting accuracy for store-level inbound cases to reduce stockouts, waste, and operational inefficiencies.
- Our project enhances Walmart's existing models by integrating external datasets (**demographics, income, GDP, unemployment, holiday and extreme weather**) with truck + store operations data.
- Engineered additional features such as:
    - inbound_cases, day/week/month seasonality, store demographics, and lag-based demand signals to improve predictive power.
- Evaluated multiple machine learning models including:
    - Ridge, Lasso, Random Forest, XGBoost, Histogram Gradient Boosting, and a LightGBM + Ridge stacked model.
- Best model performance achieved:
    - **Multi-Horizon model MAPE ↓ 0.02** compared to Baseline XGBoost model
- Results enable more reliable inventory planning, improved truck allocation, and enhanced operational efficiency across Walmart stores.

Our Team

Bei-Yu Xing

Bo-Syuan Peng

I-Tso Chen

Pin-Chen Chi

Shuo-An Wu

Executive Summary: Client Overview

01 Walmart U.S. Supply Chain – Inbound Logistics

- One of the largest retail logistics networks in the U.S.; supports **4,600+** stores
- Moves cases from vendors → DCs → stores through a high-volume inbound system

02 Operational Context

- Performance measured by forecast accuracy, service levels, inventory placement, transportation efficiency
- Sponsor team provided real operational datasets (truck arrivals, cases and store data)

Executive Summary: Business Problem

01 Walmart faces inconsistent inbound case forecasts, leading to:

- Stockouts on high-demand items
- Excess inventory + waste on slow movers
- Inefficient truck unloading and scheduling
- Higher operational costs

02 With 4,600+ stores and complex demand patterns, Walmart needs more **accurate, low-bias** predictions of inbound volume to support labor planning, replenishment, and supply chain efficiency.

Executive Summary: Project Objectives

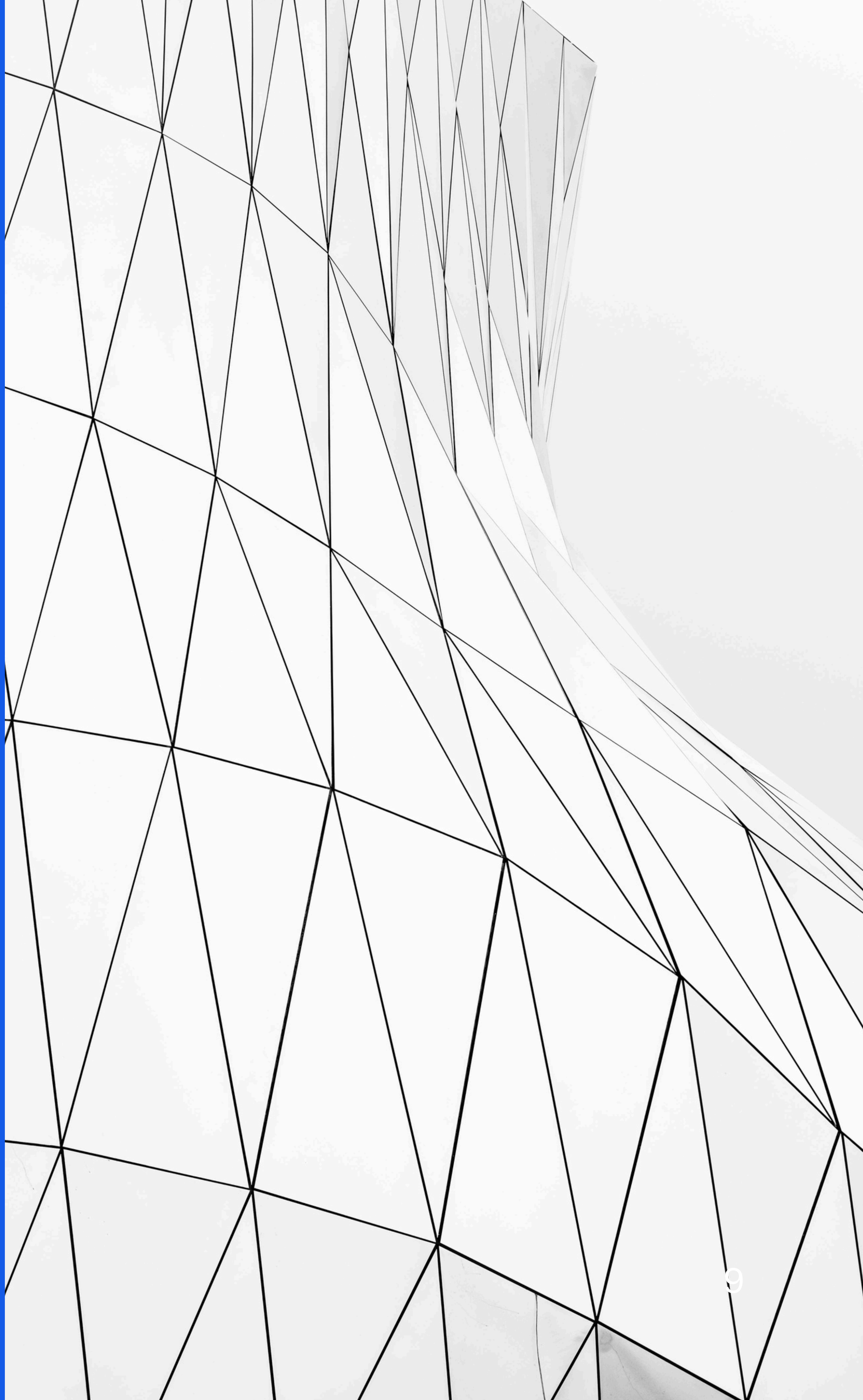
01 Reduce forecasting error by:

- **Lowering MAPE by ~10%** (vs baseline)
- **Reducing bias by ~10%** through feature engineering and model tuning.

02 Evaluate multiple ML models (Ridge, Lasso, Random Forest, XGBoost, Histogram GB, Stacked LightGBM+Ridge) to identify the best performer.

03 Deliver insights that support better truck planning, improved shelf availability, and reduced waste across Walmart stores

Data Highlights



Data Source - Primary

Datasets from Walmart:

- **inbound cases**
- **stores**
- **trucks**
- **departments**
  - 18: Seasonal (Toys & Seasonal)
  - 21: Book & Magazines
  - 34: Womens (Apparel)
  - 46: Beauty (Consumables)

```
merged_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 220000 entries, 0 to 219999
Data columns (total 8 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   dept_id          220000 non-null  int64
 1   store_id         220000 non-null  int64
 2   dt               220000 non-null  datetime64[ns]
 3   cases            208800 non-null  float64
 4   state_name       220000 non-null  object
 5   market_area_nbr  220000 non-null  int64
 6   region_nbr       220000 non-null  int64
 7   trucks           208800 non-null  float64
dtypes: datetime64[ns](1), float64(2), int64(4), object(1)
```

	dept_id	store_id	dt	cases	state_name	market_area_nbr	region_nbr	trucks
0	21	10004	2025-02-22	43.0	LA	66	13	2.0
1	46	10001	2025-02-02	68.0	MD	285	22	2.0
2	21	10001	2025-02-07	59.0	MD	285	22	3.0
3	34	10003	2025-02-26	60.0	NY	172	17	2.0
4	34	10002	2025-02-20	71.0	NC	296	26	2.0

Data Source - Additional

- **Demographic data from US Census Bureau (Mar, 2024-Sep, 2025)**
  - Total Population
  - Race Distribution
  - Age Group
  - Sex Ratio
  - Median Age
  - Median Household Income
- **Economy data from US Bureau of Labor Statistics**
  - Yearly and Monthly Salary
  - Unemployment Rate
  - State-quarter Level GDP Growth
  - Personal Income
- **Holidays**
- **Tax**
- **Severe Weather**

Data Cleansing and Preparation

Walmart Initial dataset was about **220,000 rows** and **8 columns**

External Data Preparation

- Imported yearly and quarterly data and aligned data based on state abbreviations and date.
- Cleaned data and replaced N/A values with mean values. Dropped columns that are unneeded.
- Predicted the unreleased 2025 monthly salary using 2020-2024 data.
- Predicted the unreleased Sep 2025 monthly unemployment rate using Mar 2024-Aug 2025 data
- Predicted the unreleased 2025 yearly population, race distribution, age group, sex ratio and median household income data using 2021-2024 data.
- Predicted GDP and Per Capita Personal Income data for 2025 Q3 and Q4 using 2022-2025 data.

Data Cleansing and Preparation

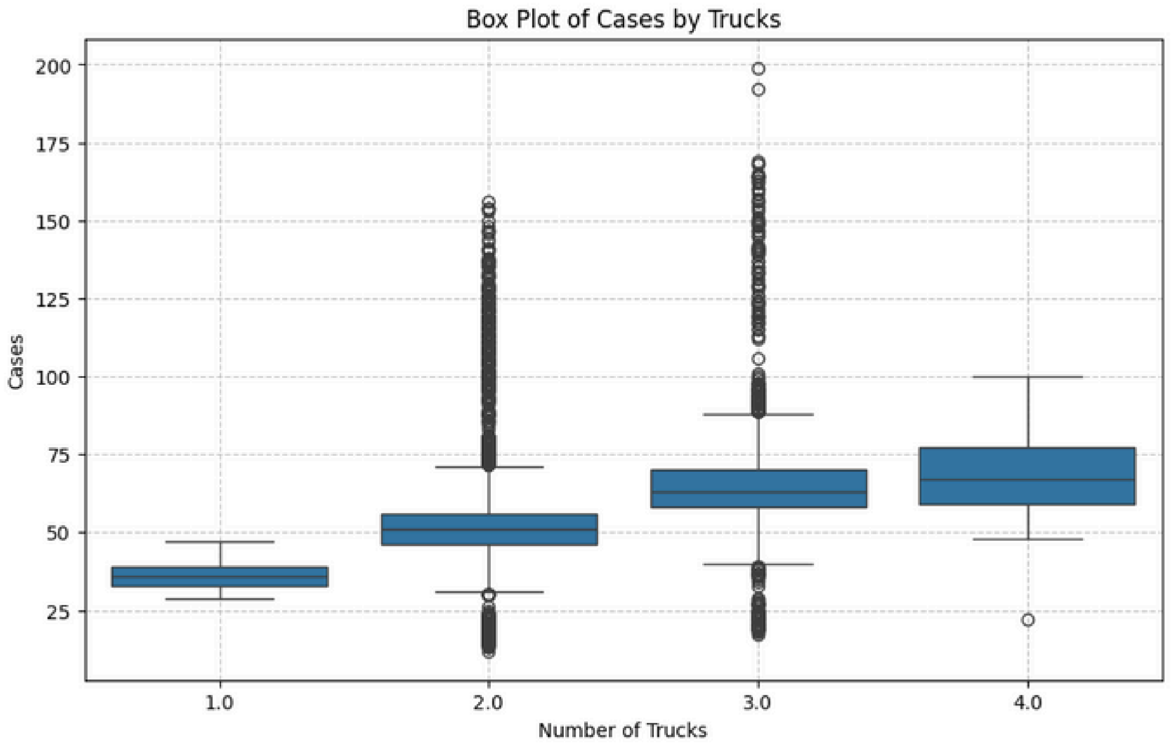
Data Processing Steps

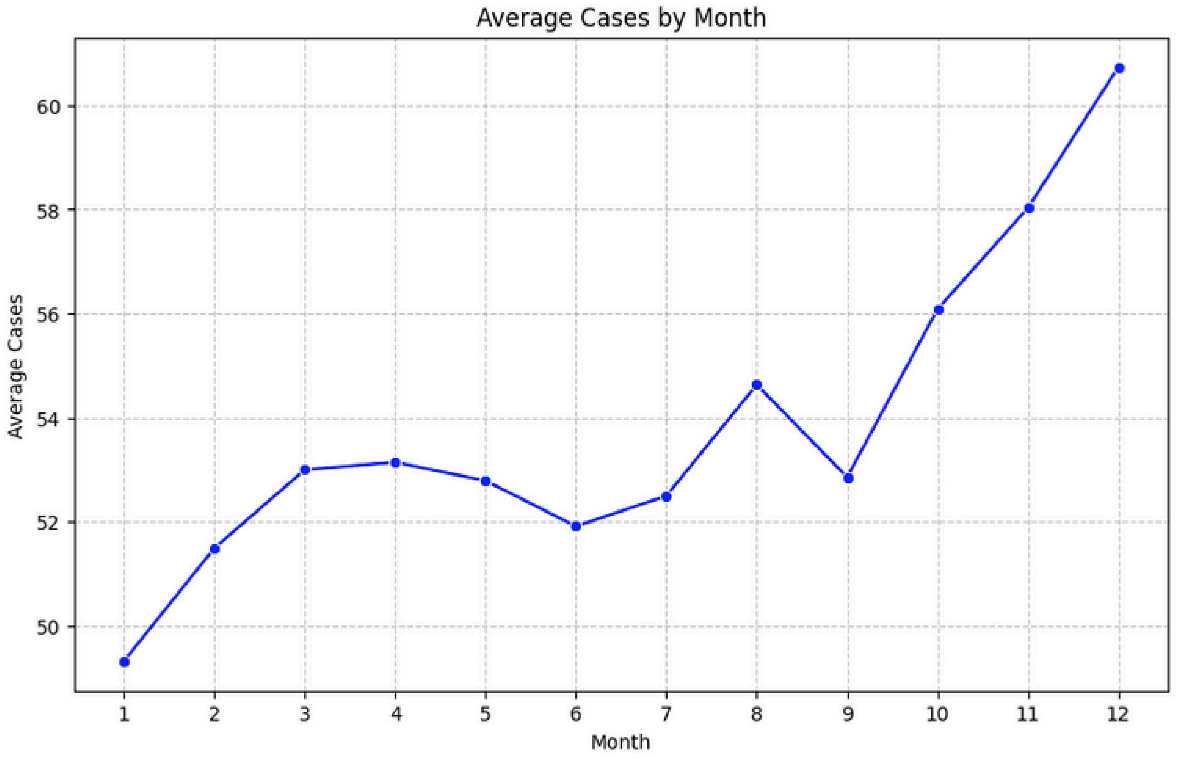
- Converted date fields to datetime.
- Aggregated inbound data to store–date level.
- Added store attributes using a left join.
- Calculated cases_per_truck efficiency metric

Future Data Cleansing Considerations

- Final dataset: ~220,000 rows.
- Each row = one store on one day with inbound volume, truck count, location info, efficiency metrics and other external data.

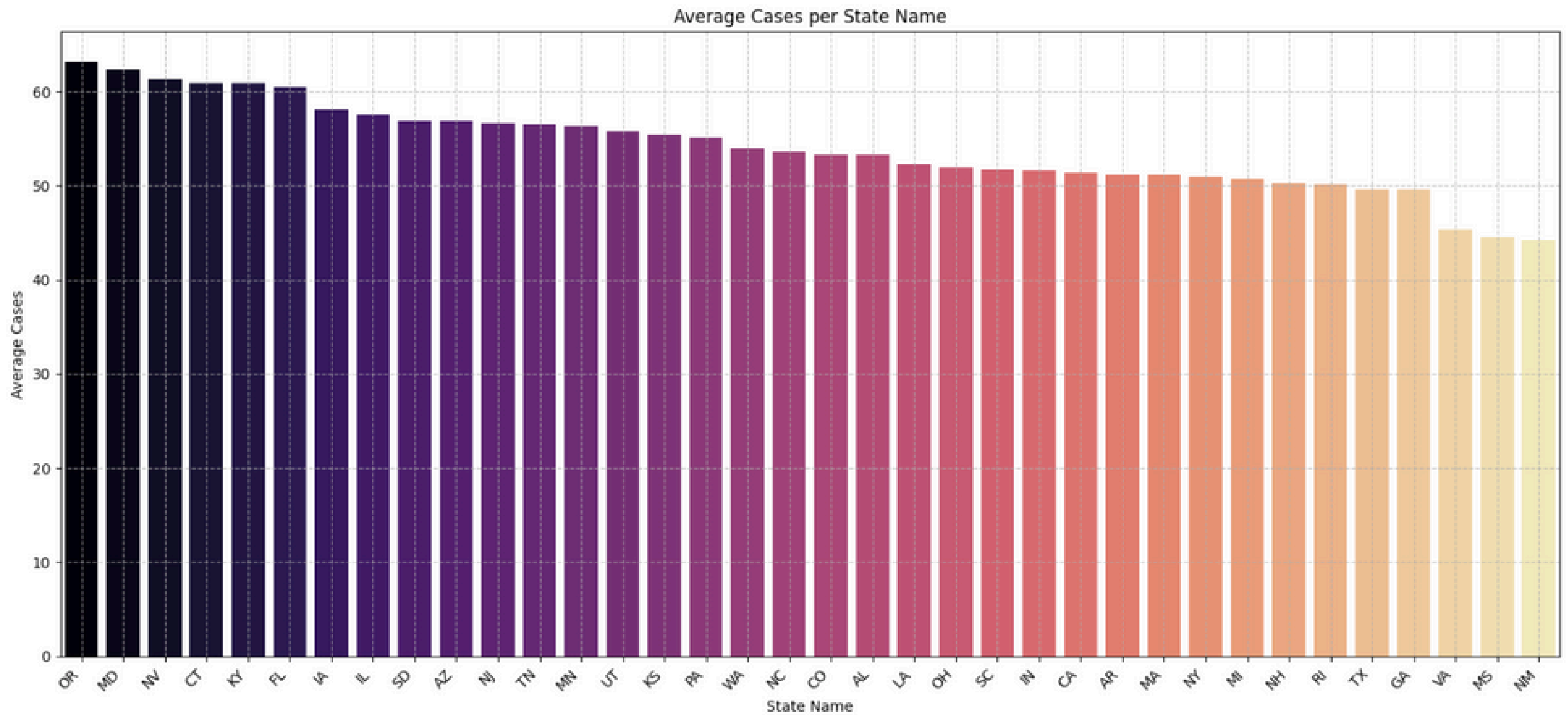
Preliminary Findings (EDA)

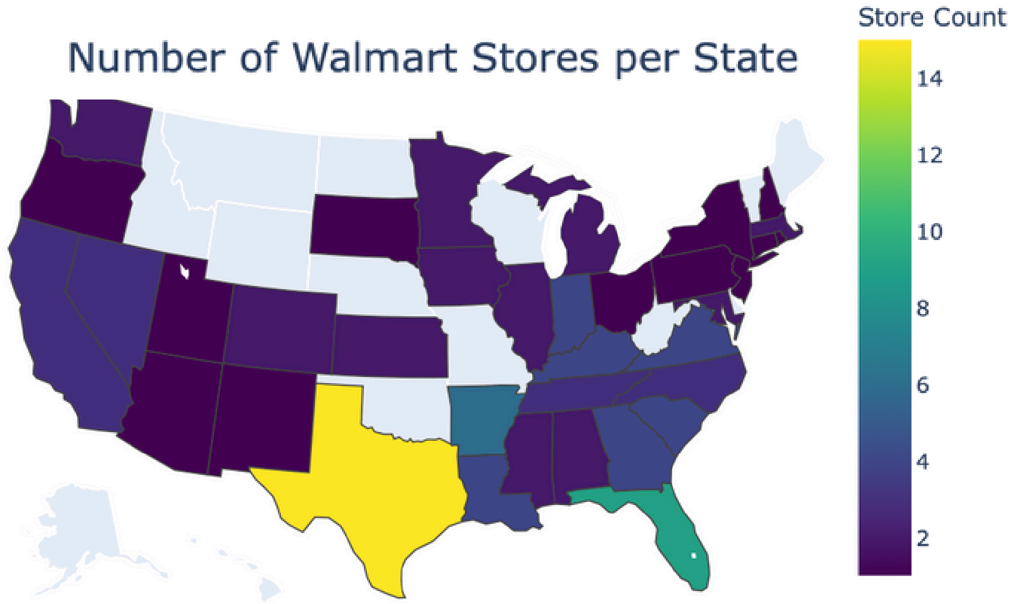




We explored the relationship between our target variable-cases, and other variables like trucks and time.

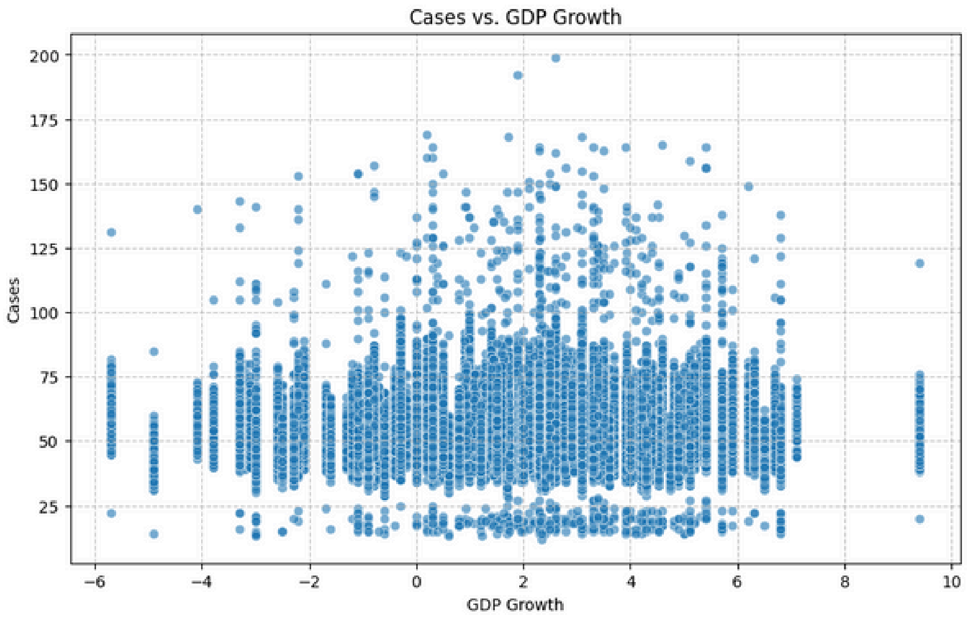
Number of cases increase as number of trucks, and has more outlier for 2 or 3 trucks. Average cases tend to increase towards the end of the year, with peaks during April and August, and the highest for December.

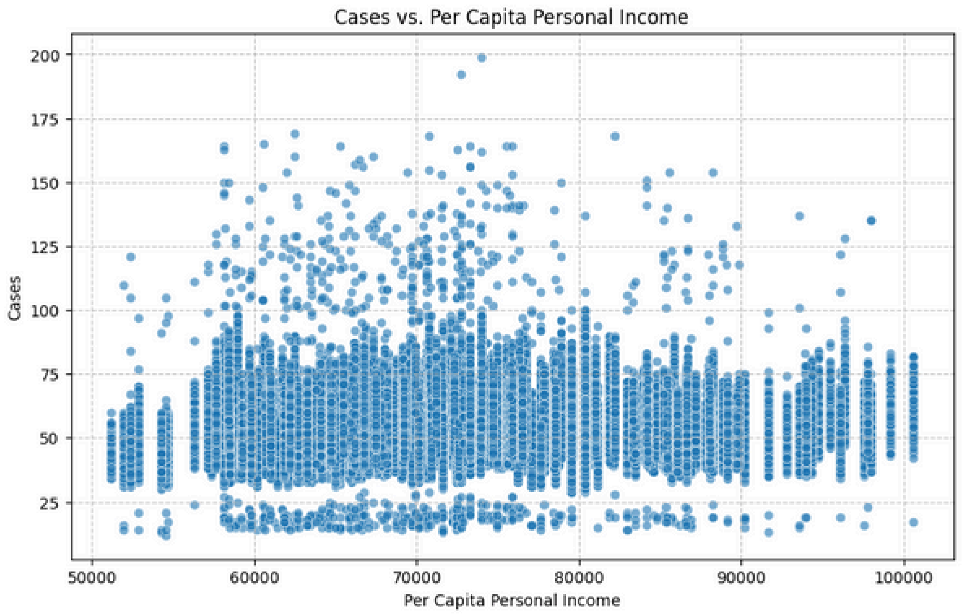


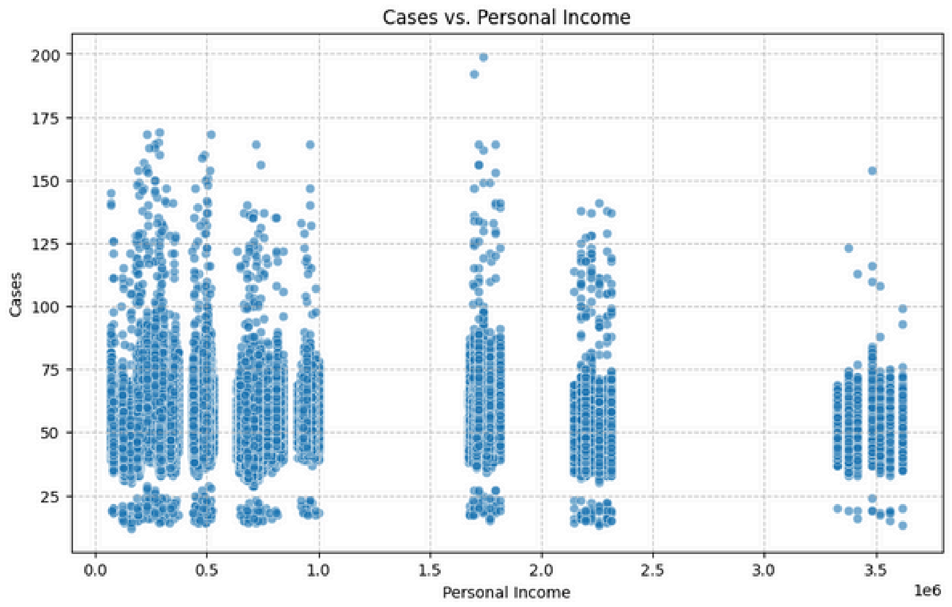


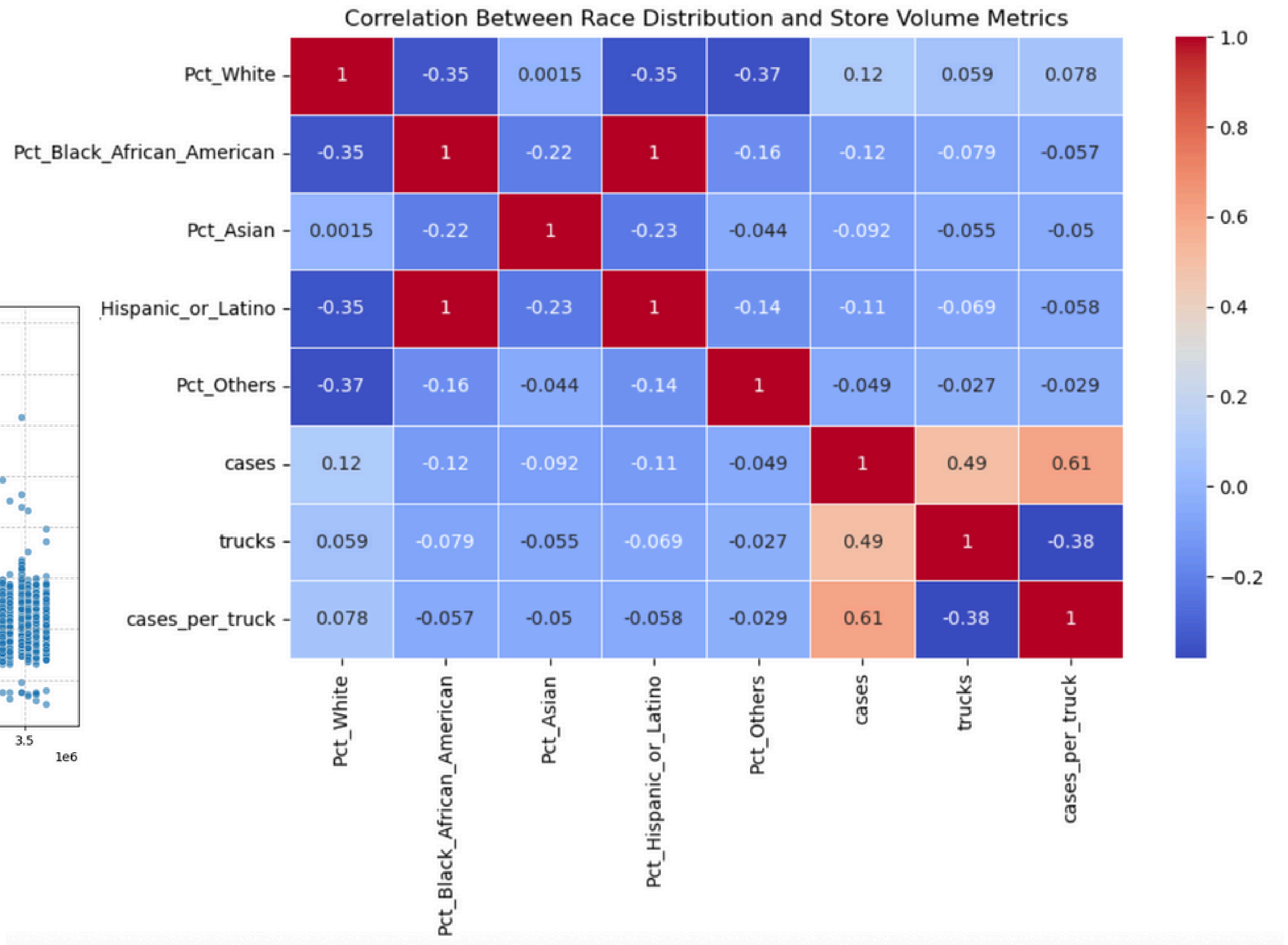
We also explored the average number of cases and store count in each state.
There are 36 states in our dataset, the highest cases states are Oregon and Maryland. Texas has the most number of stores.

Preliminary Findings (EDA)





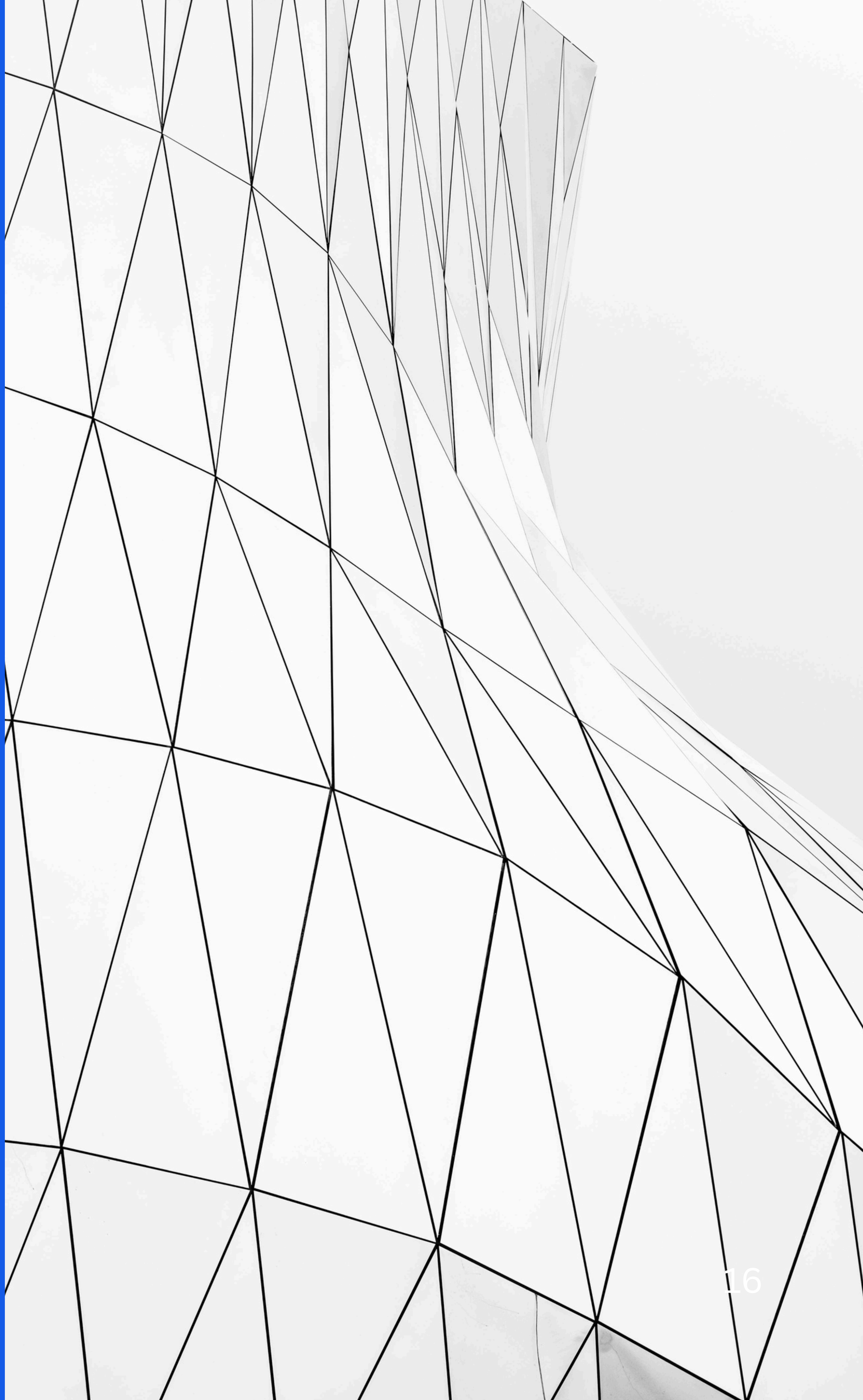




- EDA of our target variable with some external data was performed as well.
- For example: Cases vs GDP Growth, Cases vs Personal income, and Cases vs Per Capita personal income.

- Race composition of the neighborhood does not meaningfully predict Walmart inbound volume.
- Stores in different racial/ethnic neighborhoods receive roughly the same number of trucks.

Models & Findings



Baseline Model

Basic Information For Baseline Model

- X Columns = [store_id, dept_id, year, month, day, day_of_week, quarter]
- y Column = [cases]
- Train data time range: 2024-03-14 → 2025-07-20
- Test data time range: 2024-07-21 → 2025-08-17

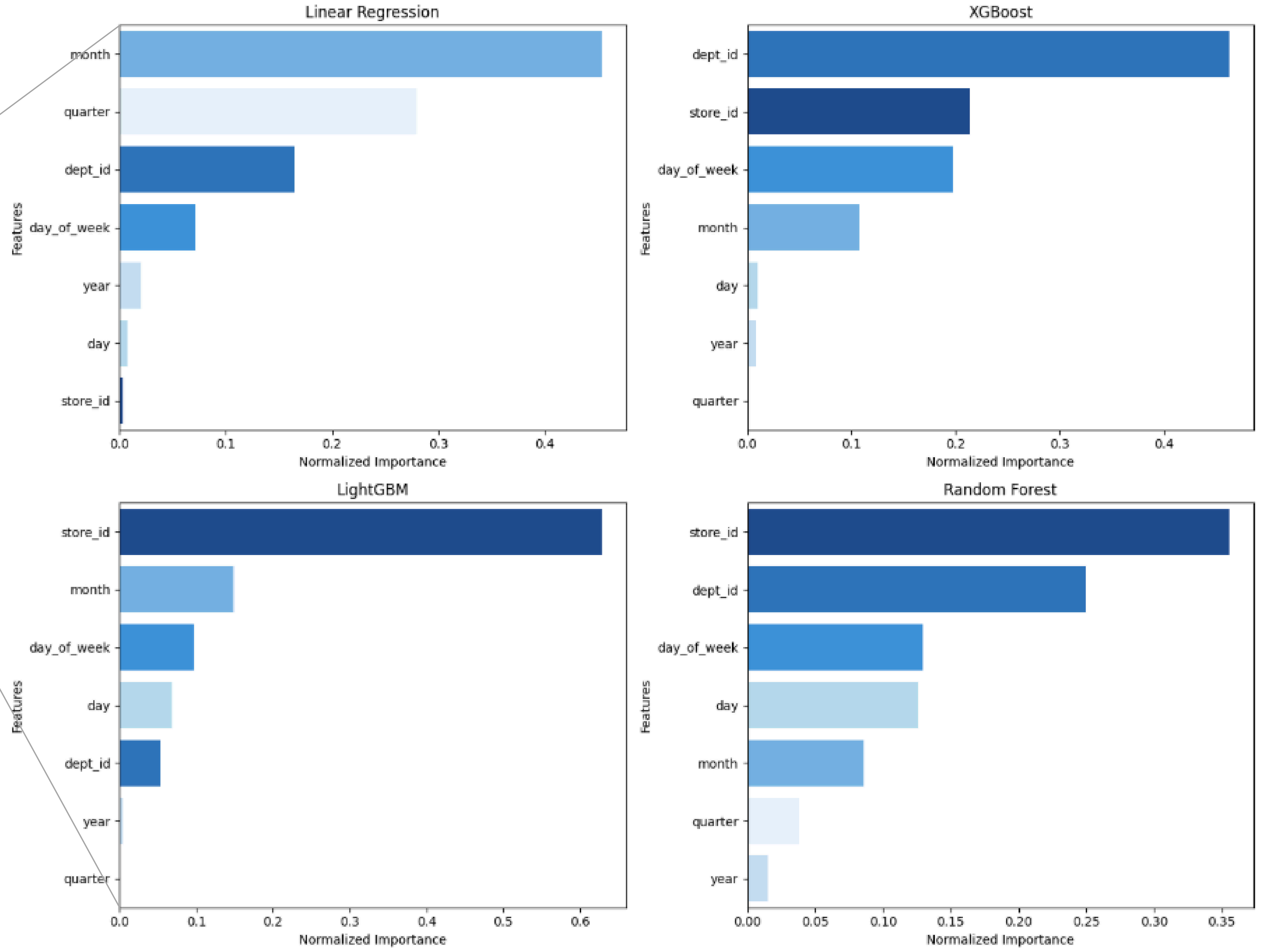
	MAPE	Bias
Linear Regression	12.901	0.363
Random Forest Regressor	7.342	-0.668
LightGBM	6.576	-0.609
XGBoost (n_est=500)	6.433	-0.236
Optimized Weighted Average Ensemble	6.422	-0.324

Optimized Ensemble Model Weights:

- XGBoost: 0.771
- LightGBM: 0.186
- Random Forest: 0.043
- Linear Regression: 0.000

Feature Importances Comparison Across Models

- store_id, dept_id, day_of_week



Baseline Model

Performance of Ensemble Model, By Department

	Department Name	+7d MAPE%	+14d MAPE%	+21d MAPE%	+28d MAPE%	+7d Bias	+14d Bias	+21d Bias	+28d Bias
0	Womens (Apparel)	6.16	6.65	6.48	6.46	-0.38	-0.92	-0.41	-0.54
1	Book & Magazines	6.40	6.72	6.14	6.64	-0.44	-1.60	-0.50	-0.74
2	Beauty (Consumables)	6.48	6.14	6.64	6.32	0.41	0.43	-0.14	0.04
3	Seasonal (Toys & Seasonal)	7.18	6.18	5.89	6.27	0.04	-0.18	-0.28	0.03

- For **+7 days** predictions, **Womens (Apparel)** is easier to predict.
- For **+7 days** predictions, **Seasonal (Toys & Seasonal)** is harder to predict.

- For **+28 days** predictions, **Seasonal (Toys & Seasonal)** is easier to predict.
- For **+28 days** predictions, **Book & Magazines** is harder to predict.

Performance of Ensemble Model, By Week

	week_start_date	Week	MAPE	Bias	N
0	2025-07-21 00:00:00	Week 1	6.555	-0.092	2800
1	2025-07-28 00:00:00	Week 2	6.421	-0.568	2800
2	2025-08-04 00:00:00	Week 3	6.288	-0.332	2800
3	2025-08-11 00:00:00	Week 4	6.424	-0.302	2800
4	Overall	Overall	6.422	-0.324	11200

Recursive Model

Predicts a time period ahead at a time, and then feeds that prediction back into the model to predict the next time period.

Add lag features

truck_lags = [1, 3, 5, 7, 14, 21, 28]

case_lags = [1, 3, 5, 7, 14, 21, 28]

no data before 2024-03-14

→ beginning NAs replaced with average

First date of our data set: 2024-3-14

	store_id	dept_id	dt	cases	trucks	lag_cases_1	lag_cases_3	lag_cases_5
0	10001	18	2024-03-14	69.0	3.0	NaN	NaN	NaN
550	10001	21	2024-03-14	21.0	3.0	NaN	NaN	NaN
1100	10001	34	2024-03-14	88.0	3.0	NaN	NaN	NaN
1650	10001	46	2024-03-14	69.0	3.0	NaN	NaN	NaN
2200	10002	18	2024-03-14	52.0	3.0	NaN	NaN	NaN

Why We Use Recursive Forecasting

- Matches real operational forecasting
- Allows multi-week forecasts (1–4 weeks ahead)
- Uses lag features effectively (lag_1, lag_7, lag_14, etc.)

Why It Helps Our Walmart Capstone

- Weekly inbound case volumes rely heavily on previous weeks' demand
- Useful for short-term labor planning and warehouse allocation

Recursive Model

- Predicted daily truck volumes using XGBoost (n_estimators = 300)
- Inserted the predicted truck lag features back into the main dataframe
- Ensures the recursive model has access to updated truck lags for future predictions

After predicted trucks (2025-09)

	dt	trucks	lag_trucks_1	lag_trucks_3	lag_trucks_7
219990	2025-09-05	1.985744	1.988868	2.151779	1.957911
219991	2025-09-06	1.988494	1.985744	2.179653	1.961241
219992	2025-09-07	2.068851	1.988494	1.988868	2.005359
219993	2025-09-08	1.902115	2.068851	1.985744	2.009998
219994	2025-09-09	2.188393	1.902115	1.988494	2.151779

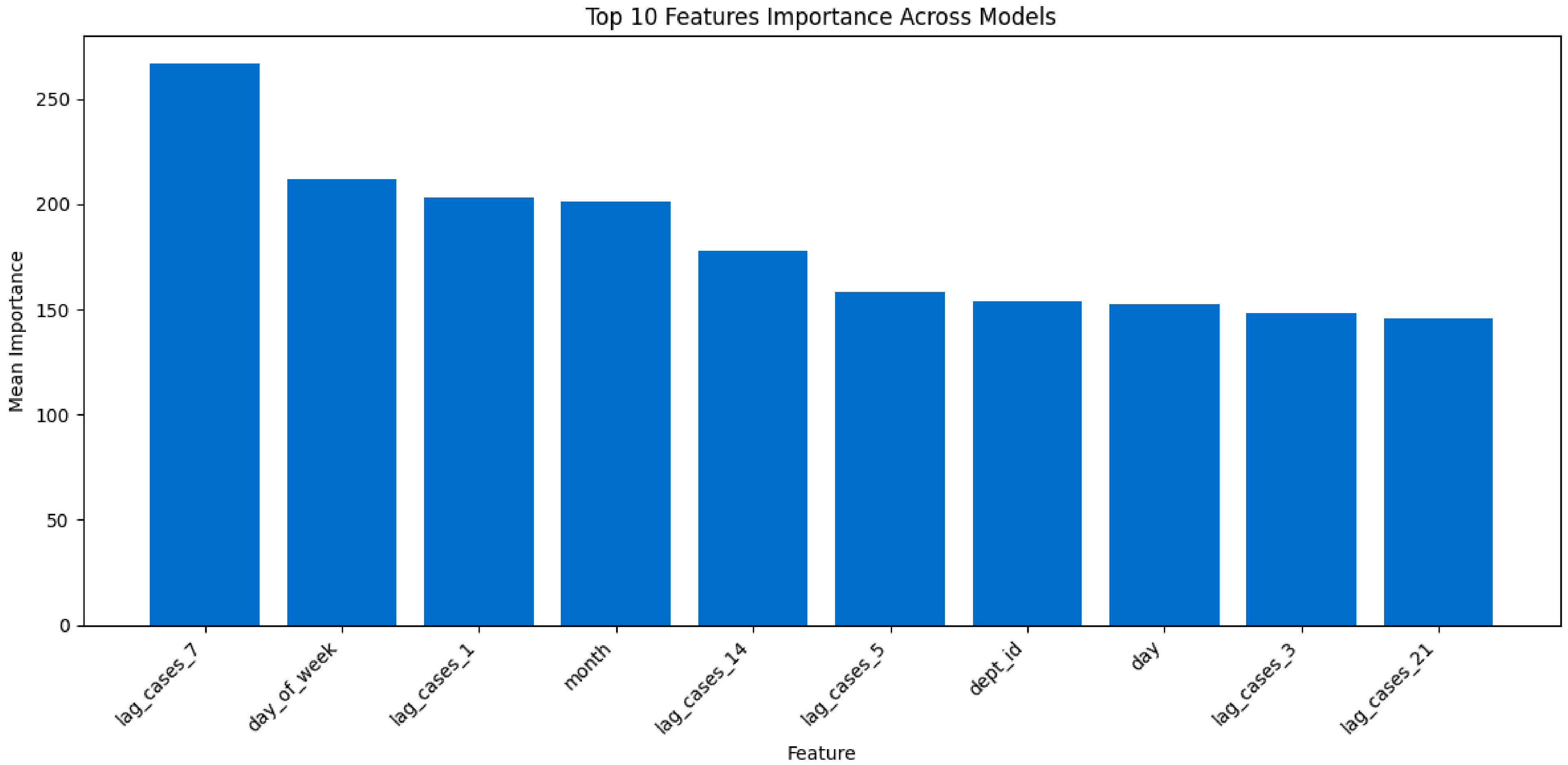
Add cases lag features (2025-08-19)

	store_id	dept_id	dt	cases	trucks	lag_cases_1	lag_cases_3	lag_cases_5
523	10001	18	2025-08-19	NaN	3.050264	NaN	60.0	66.0
1073	10001	21	2025-08-19	NaN	3.043105	NaN	60.0	70.0
1623	10001	34	2025-08-19	NaN	3.045612	NaN	75.0	89.0
2173	10001	46	2025-08-19	NaN	3.068719	NaN	69.0	73.0
2723	10002	18	2025-08-19	NaN	2.246606	NaN	51.0	48.0

Recursive Model

Train: 70 weeks
Test: 8 weeks (4 weeks prediction)

	MAE	MSE	RMSE	R2	MAPE	Bias
Optimized Weighted Average Ensemble	3.513393	29.228318	5.406322	0.680586	$\boxed{6.561374}$	-0.447156
XGBoost	3.517273	29.237499	5.407171	0.680486	6.574735	$\boxed{-0.395381}$
LightGBM	3.526360	29.354725	5.418000	0.679205	6.584547	-0.451305
CatBoost	3.545819	29.640938	5.444349	0.676077	6.608270	-0.556409
Random Forest	3.714474	31.813586	5.640353	0.652334	6.883396	-0.864256
Simple Average Ensemble	3.759053	32.220216	5.676285	0.647890	6.988707	-0.712466
Linear Regression	3.870304	33.963802	5.827847	0.628836	7.178794	-0.783107
SVR (Linear Kernel)	7.208613	93.654955	9.677549	-0.023483	13.484017	-1.224339



Optimized Ensemble Model Weights:

- XGBoost: 57.05%
- LightGBM: 16.54%
- CatBoost: 26.41%
- Others: 00.00%

Feature Importances Across Models:

1. "lag_cases_7"
2. "day_of_week"
3. "lag_cases_1"
4. "month"
5. "lag_cases_5"

Recursive Model

Train: 70 weeks
Test: 8 weeks (4 weeks prediction)

Weekly & Overall MAPE by Department (Optimized Ensemble Model)

dept_id	Department Name	1 week MAPE(%)	2 week MAPE(%)	3 week MAPE(%)	4 week MAPE(%)	Overall MAPE(%)
18	Seasonal (Toys & Seasonal)	7.205	6.643	6.261	6.051	6.605
21	Book & Magazines	6.583	6.874	6.575	6.814	6.677
34	Womens (Apparel)	6.188	6.723	6.348	6.851	6.540
46	Beauty (Consumables)	6.456	6.282	6.820	6.553	6.477

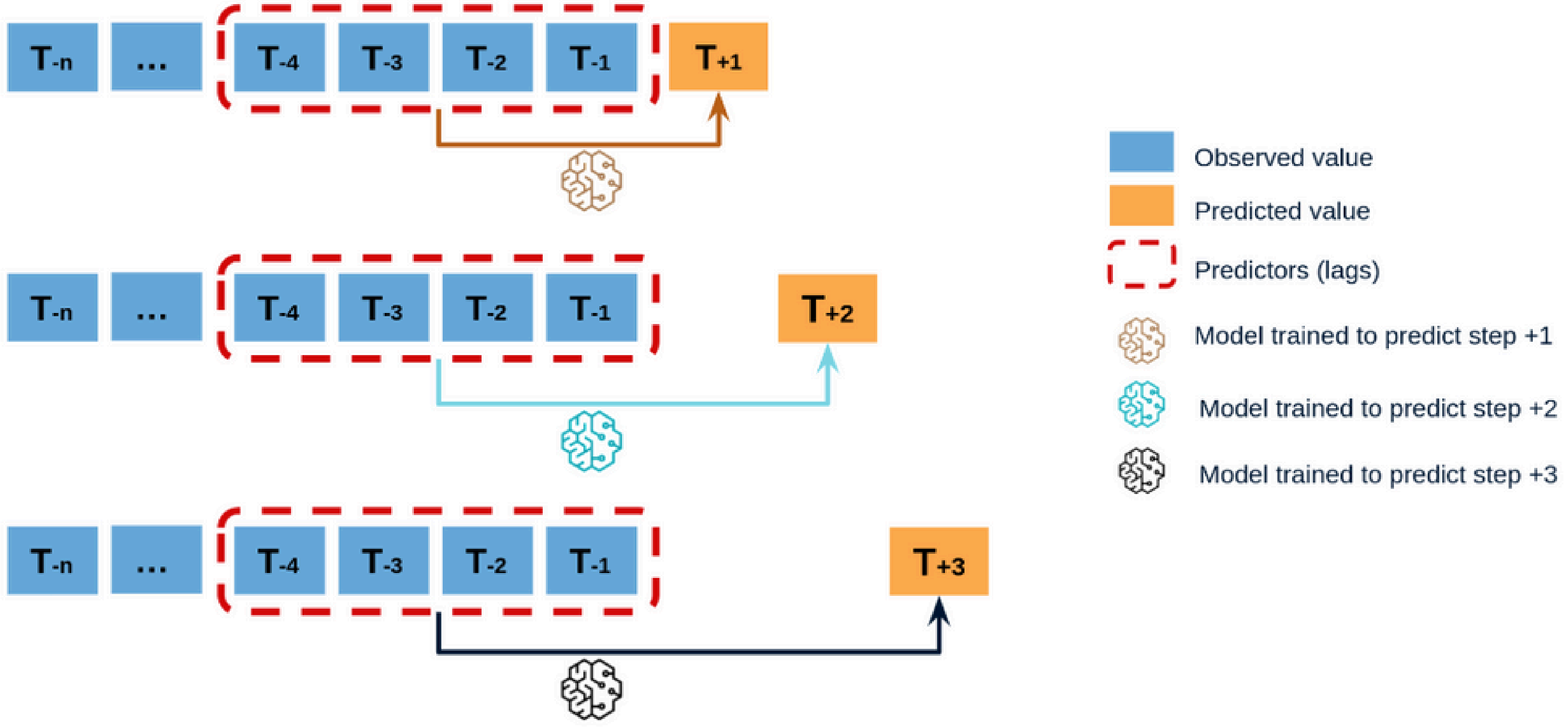
Weekly & Overall Bias by Department (Optimized Ensemble Model)

dept_id	Department Name	1 week Bias	2 week Bias	3 week Bias	4 week Bias	Overall Bias
18	Seasonal (Toys & Seasonal)	0.324	-0.013	-0.098	0.124	0.082
21	Book & Magazines	-1.033	-1.685	-0.537	-0.588	-0.943
34	Womens (Apparel)	-0.614	-1.283	-0.561	-0.518	-0.753
46	Beauty (Consumables)	0.232	0.185	-0.179	-0.077	0.033

- Beauty and Seasonal departments achieve the lowest average MAPE → Demand patterns is more stable.
- Book & Magazines has slightly higher 1–4 week MAPE → More volatility
- Only Book & Magazines shows notable negative bias across all weeks → model consistently under-predicts this department.

Direct Multi-Horizon Model

Concept of Multi-Horizon Models (from skforecast.org)



- Takes the last know date and features to forecast.
- Each forecast day uses one specifically trained model to predict that day.

Pros: Does not rely on predicted values to make prediction
Cons: Takes longer. Trains more models, one per forecast day

Basic Information

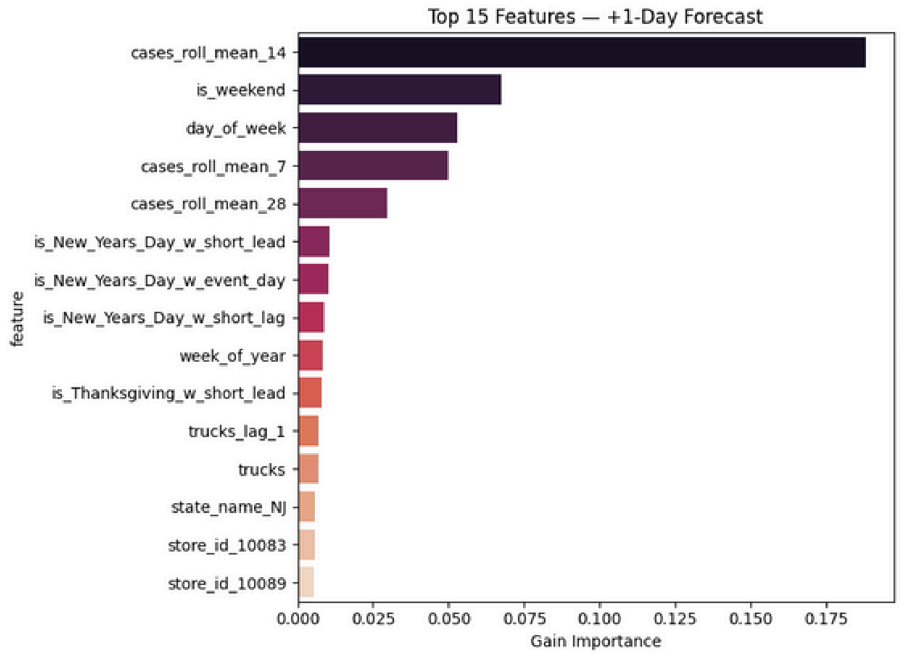
- Features: All external data, calendar,
  - cases Lag [1, 2, 3, 7, 14, 21, 28]
  - cases Rolling [7, 14, 28]*[mean, std]
- Train data time range: 2024-03-14 → 2025-07-20
- Test data time range: 2024-07-21 → 2025-08-17

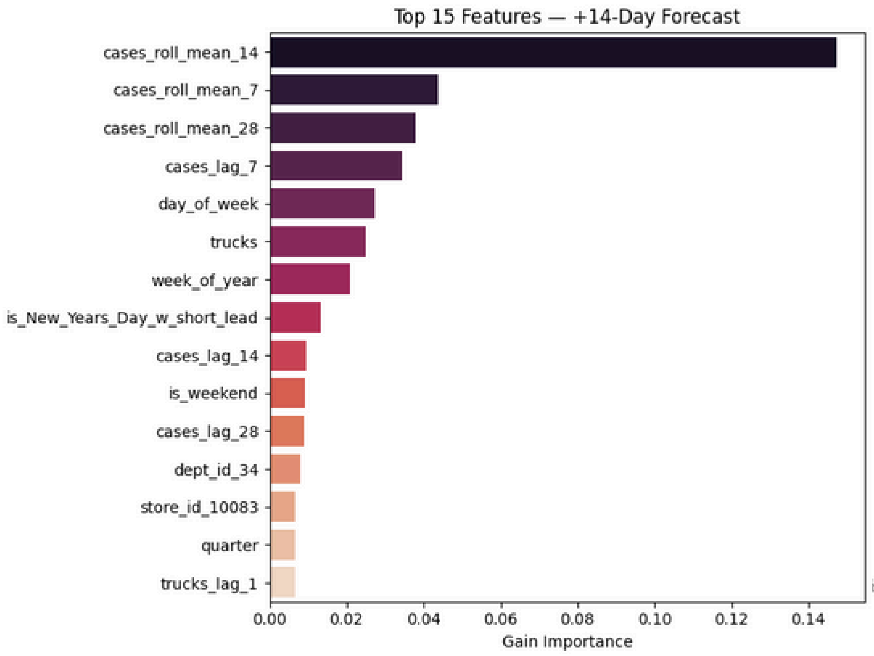
Evaluation Performance, by Week

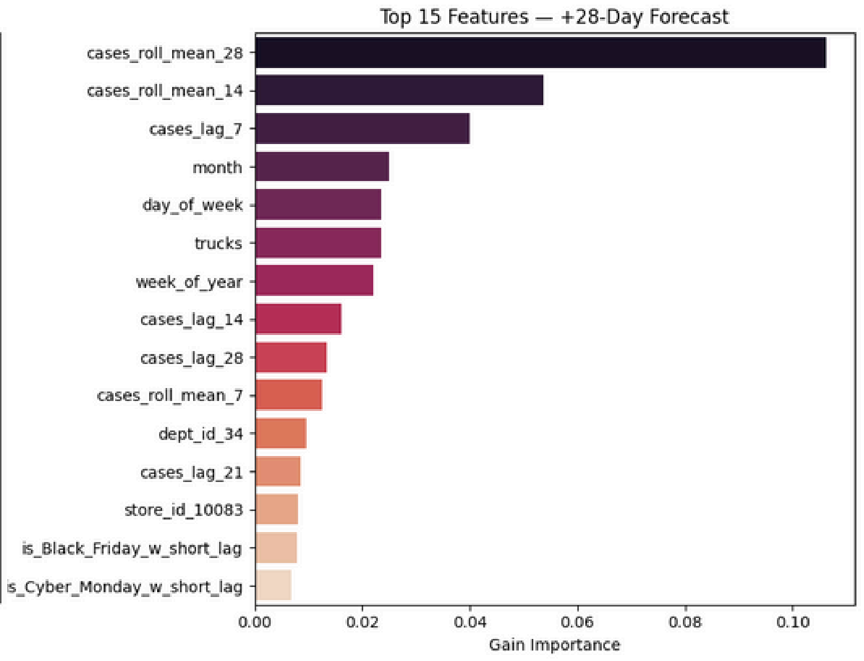
	Regular_MAPE	Bias
Week		
Week 1	6.54	-0.09
Week 2	6.42	-0.49
Week 3	6.27	-0.42
Week 4	6.40	-0.21
Overall (28 days)	6.41	-0.30

Direct Multi-Horizon Model

Feature Importance:







- cases_roll_mean_14,
- day_of_week,
- cases_roll_mean_7,
- cases_roll_mean_28,
- cases_lag_7

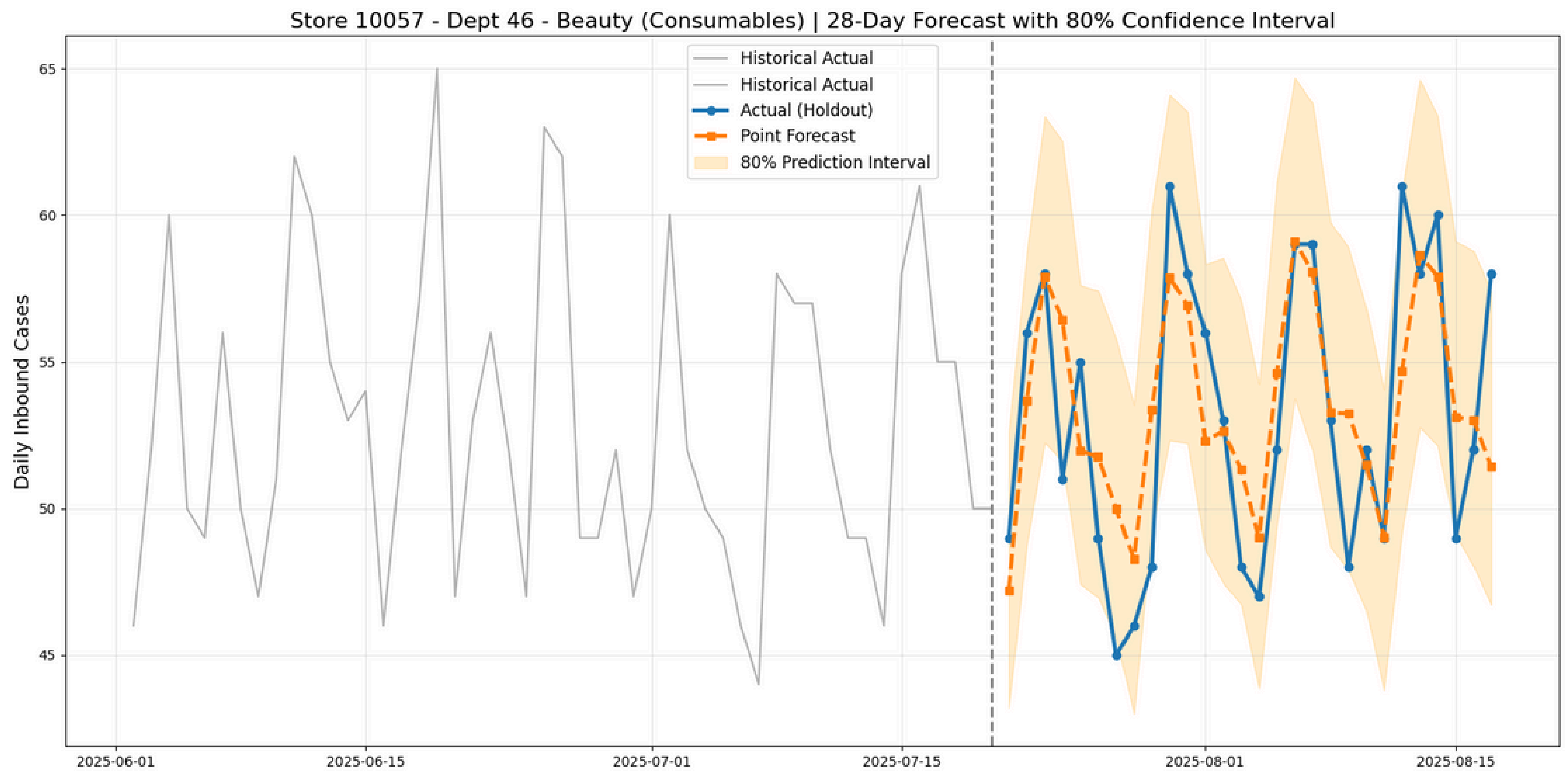
Evaluation Performance, by Department

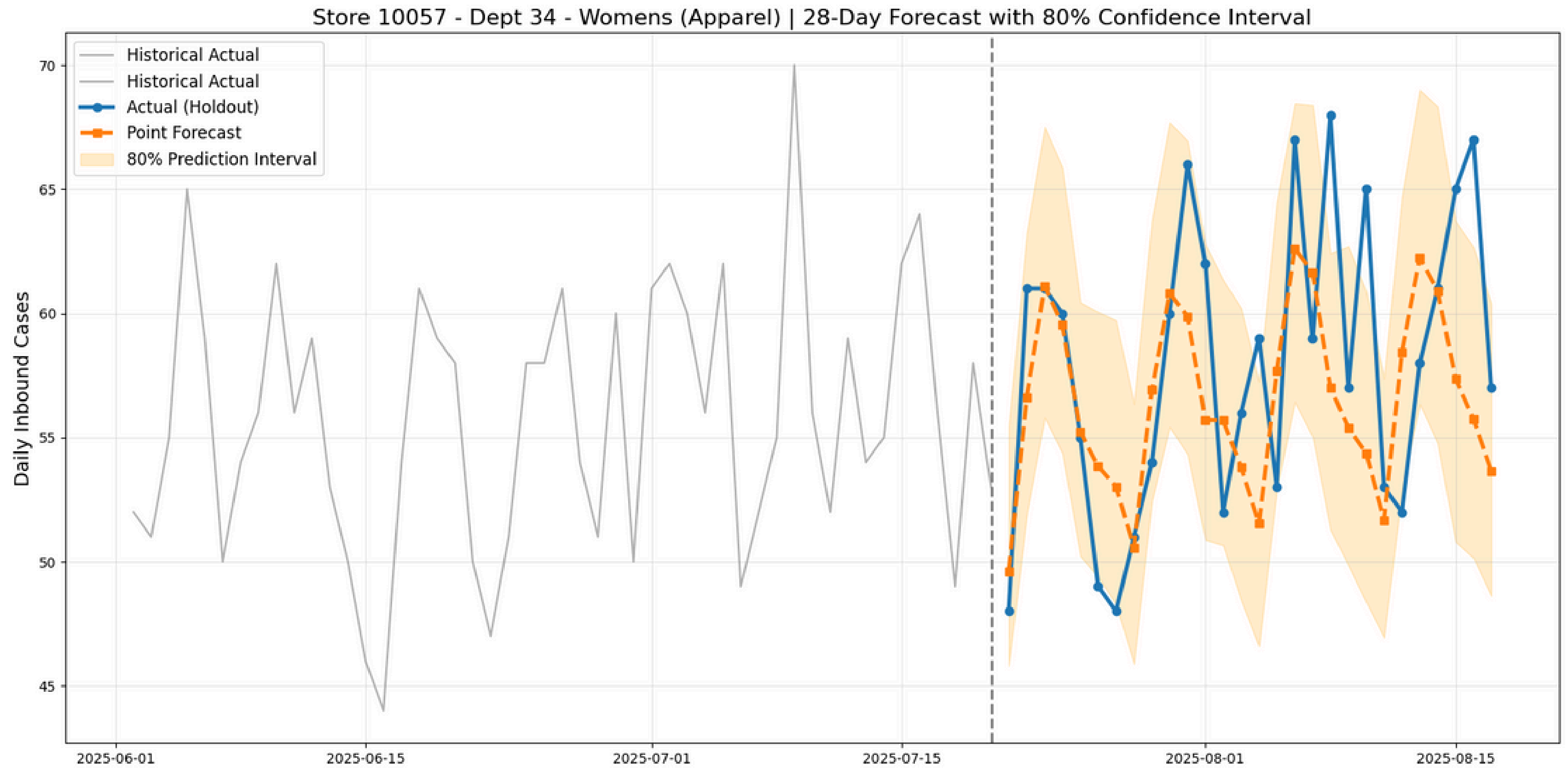
Dept_Id	Department Name	+7d MAPE%	+14d MAPE%	+21d MAPE%	+28d MAPE%
21	Book & Magazines	5.91	6.56	6.05	6.16
34	Womens (Apparel)	5.96	6.49	6.29	7.49
46	Beauty (Consumables)	6.82	6.56	6.47	5.25
18	Seasonal (Toys & Seasonal)	7.50	6.31	6.04	7.00

- For **+7 days** predictions, **Book & Magazines** is easier to predict.
- For **+7 days** predictions, **Seasonal (Toys & Seasonal)** is the hardest to predict.
- For **+28 days** predictions, **Beauty (Consumables)** is easier to predict.
- For **+28 days** predictions, **Womens (Apparel)** is the hardest to predict.

Direct Multi-Horizon Model

80% Confidence Interval Band to view Store-Dept Forecast





- Date: 2025-07-21 to 2025-08-17
- For **+28 days** predictions, **Beauty (Consumables)** is easier to predict.
- For **+28 days** predictions, **Womens (Apparel)** is the hardest to predict.

- Dept 46 - **Beauty (Consumables)**
- 28-Day MAPE for this series: **4.79%**
- Empirical coverage: **78.6%**

- Dept 34 - **Womens (Apparel)**
- 28-Day MAPE for this series: **6.85%**
- Empirical coverage: **71.4%**

In-Context Learning (ICL)

What Method Did We Use?

- In-Context Learning — No model training required
- AI directly learns patterns from the historical data
- Uses real past behaviors (store, department, weekday patterns) to extrapolate future inbound cases

How It Works?

- ICL Prediction:
  → Used 70 weeks of historical inbound data
  → Predicted the following 4 weeks (test)
- AI looks for regular patterns (e.g., weekday cycles, truck arrival habits)
- AI predicts them by asking: “When this situation happened before, what usually happened next?”
- This allows AI to fill in all future X-features and forecast cases (y)

In-Context Learning (ICL)

Patterns Identified by AI
and Why It Works for Walmart Inbound Data

(These patterns match our EDA)

- Strong repeating weekly patterns
- Truck arrivals are highly stable (mostly 2 or 3 per day)
- Cases depend on weekday + trucks
- Lag features (lag1, lag7, lag14) follow predictable cycles
- Walmart inbound behaves almost like a repeating weekly schedule

In-Context Learning (ICL)

Prediction Accuracy

Evaluation:

Actual vs. Predicted Cases (MAPE & Bias)

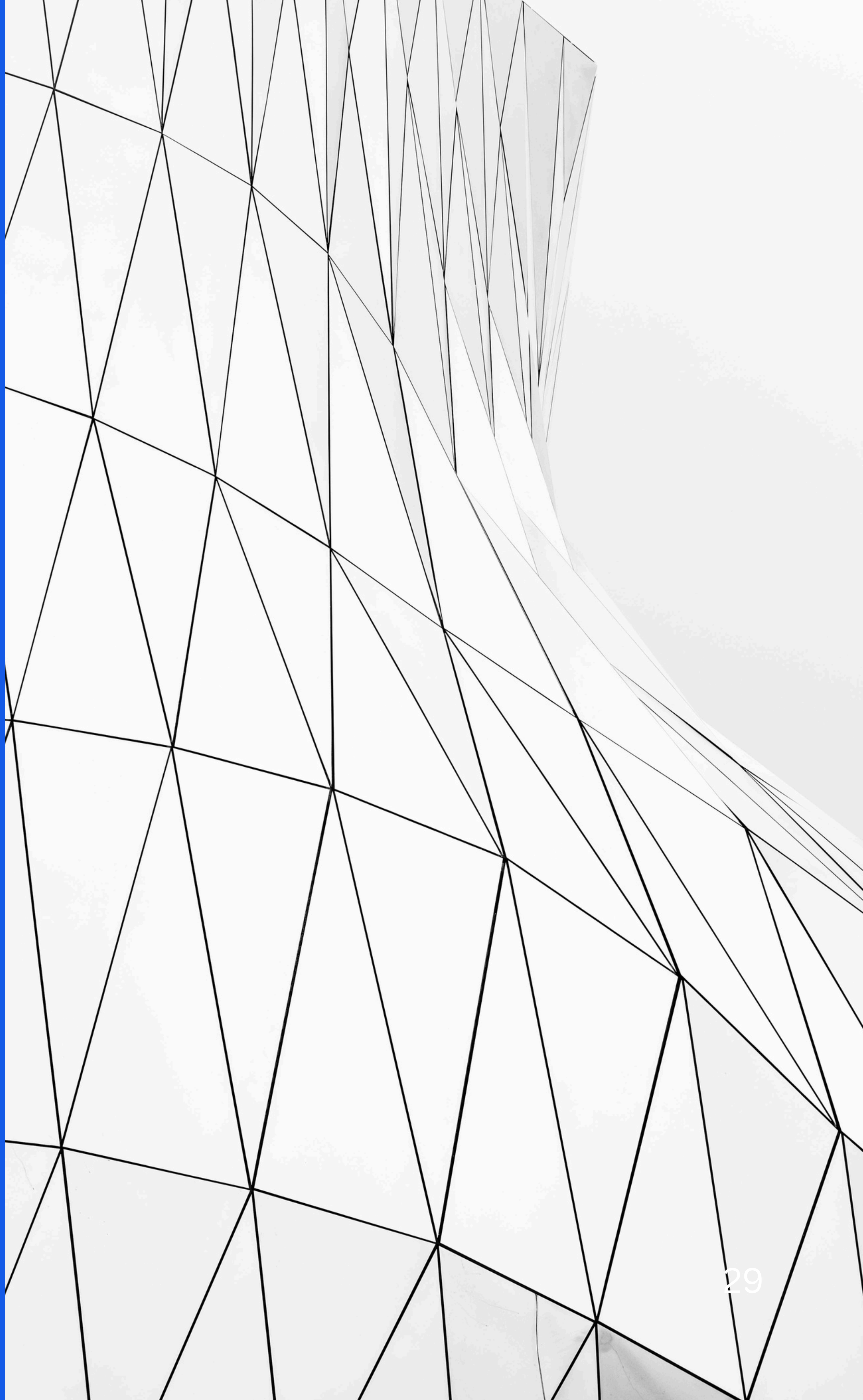
Prediction Time range:
2025-07-21 to 2025-08-17 (28 days)

Week	MAPE	Bias
Week 1	7.26	0.68
Week 2	6.79	-0.58
Week 3	6.80	-1.30
Week 4	6.85	-1.23
Overall (28 days)	6.93	-0.61

Summary

- **No predictive model required** — ICL lets AI extend historical patterns directly, offering a simple alternative to training a full model.
- **Overall performance is solid**, with MAPE consistently below 7%.
- **Extremely time-efficient** — ICL completes in about 1 minute, compared to 45 minutes for XGBoost training.

Models Comparison



Compare All Models

Comparison Across Models

- Prediction Time range: 2025-04-28 to 2025-05-25 (28 days)
- train 58w, test 16w, forcast 4w

Baseline Ensemble

Week	MAPE	Bias
Week 1	6.400	0.082
Week 2	6.622	0.155
Week 3	6.823	0.052
Week 4	6.734	0.398
Overall	6.645	0.172

Recursive Model

Week	MAPE	Bias
Week 1	6.870	0.594
Week 2	6.649	0.192
Week 3	7.048	0.035
Week 4	6.956	0.290
Overall	6.738	0.109

Direct Multi-Horizon

	Regular_MAPE	Bias
Week		
Week 1	6.55	0.50
Week 2	6.86	0.50
Week 3	6.92	0.39
Week 4	6.83	0.53
Overall (28 days)	6.79	0.48

In-Context Learning

Week	MAPE	Bias
Week 1	7.13	1.12
Week 2	7.50	1.26
Week 3	7.76	1.21
Week 4	7.55	1.33
Overall (28 days)	7.48	1.23

Compare All Models

Comparison Across Models

- Prediction Time range: 2025-07-21 to 2025-08-17 (28 days)
- train 70w, test 4w, forcast 4w

Baseline Ensemble

Week	MAPE	Bias
Week 1	6.555	-0.092
Week 2	6.421	-0.568
Week 3	6.288	-0.332
Week 4	6.424	-0.302
Overall	6.422	-0.324

Recursive Model

Week	MAPE	Bias
Week 1	6.606	-0.324
Week 2	6.606	-0.754
Week 3	6.488	-0.421
Week 4	6.555	-0.295
Overall	6.564	-0.448

Direct Multi-Horizon

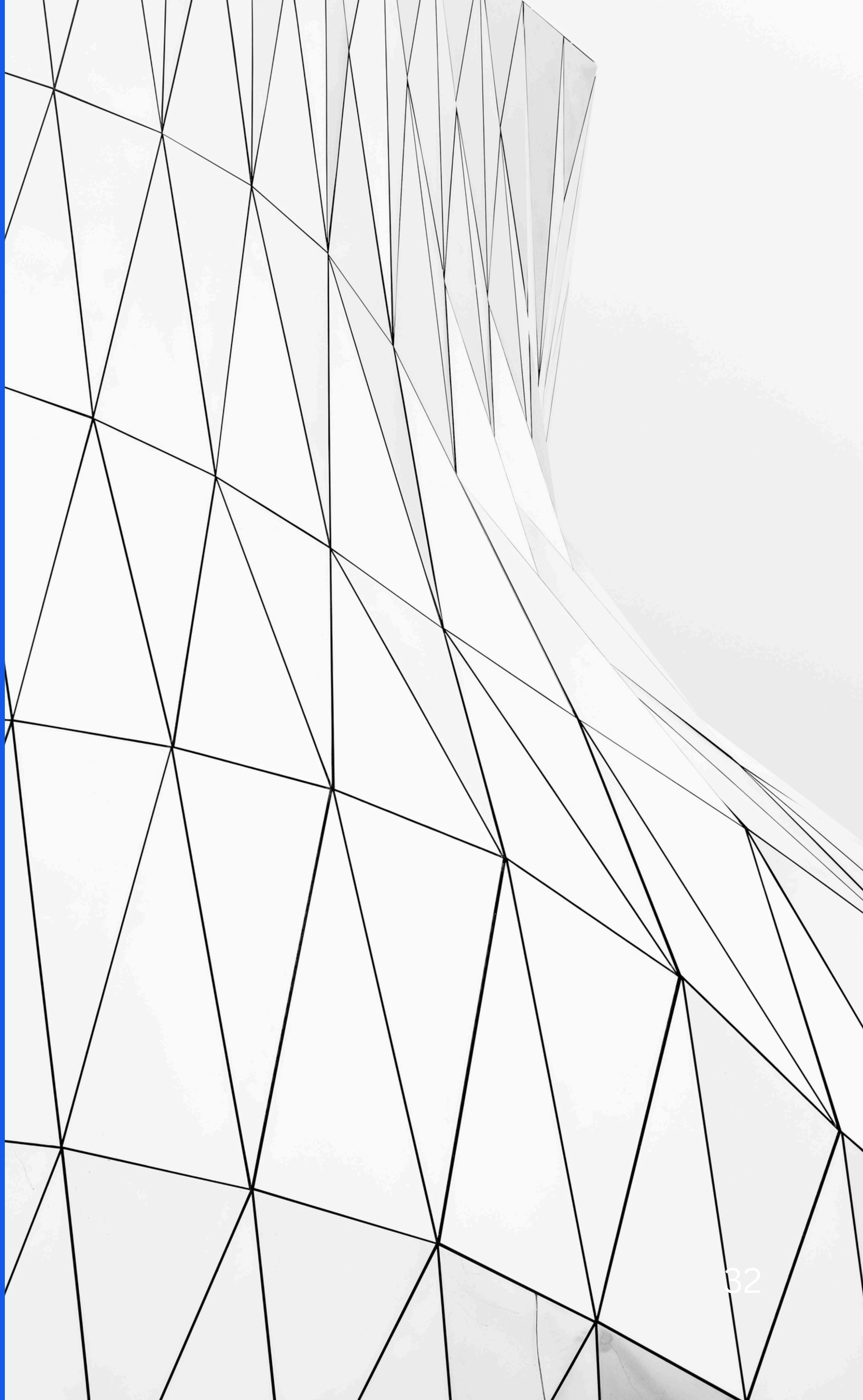
	Regular_MAPE	Bias
Week		
Week 1	6.54	-0.09
Week 2	6.42	-0.49
Week 3	6.27	-0.42
Week 4	6.40	-0.21
Overall (28 days)	6.41	-0.30

In-Context Learning

Week	MAPE	Bias
Week 1	7.26	0.68
Week 2	6.79	-0.58
Week 3	6.80	-1.30
Week 4	6.85	-1.23
Overall (28 days)	6.93	-0.61

- **The second prediction time range performs better**
- Model Performance: Multi-Horizon > Baseline > Recursive > ICL
- Overall, both models and the baseline perform well, with MAPE around 6.4–6.6%.
  The ICL approach is slightly less accurate, with a MAPE of 6.9%.
- We decided to use **Direct Multi-Horizon Model** to do the final prediction.

Challenges & Workarounds



Challenges & Workarounds

Unavailable Recent External Data (e.g., Demographics, Income, GDP)

Several external datasets did not include 2024–2025 values, requiring us to forecast missing data using historical trends.

Workaround & Fix:

- Forecasted 2024–2025 values using state-level data
- Generated predicted values for population, age distribution, sex ratio, etc
- Ensured complete and consistent feature availability across all states and dates

Challenges & Workarounds

Row Duplication When Merging Holiday Data

While integrating holiday/event windows, merging on dates created multiple rows per date. Because a single date can fall into several holiday lead/lag windows (e.g., −14, −7, −3 days).
This produced duplicate operational records differing only in holiday fields.

Workaround & Fix:

- Converted each holiday into multiple window-based binary features (e.g., is_Thanksgiving_w_long_lead, is_Black_Friday_w_event_day)
- Handled overlapping windows using multiple 1's instead of duplicate rows

Challenges & Workarounds

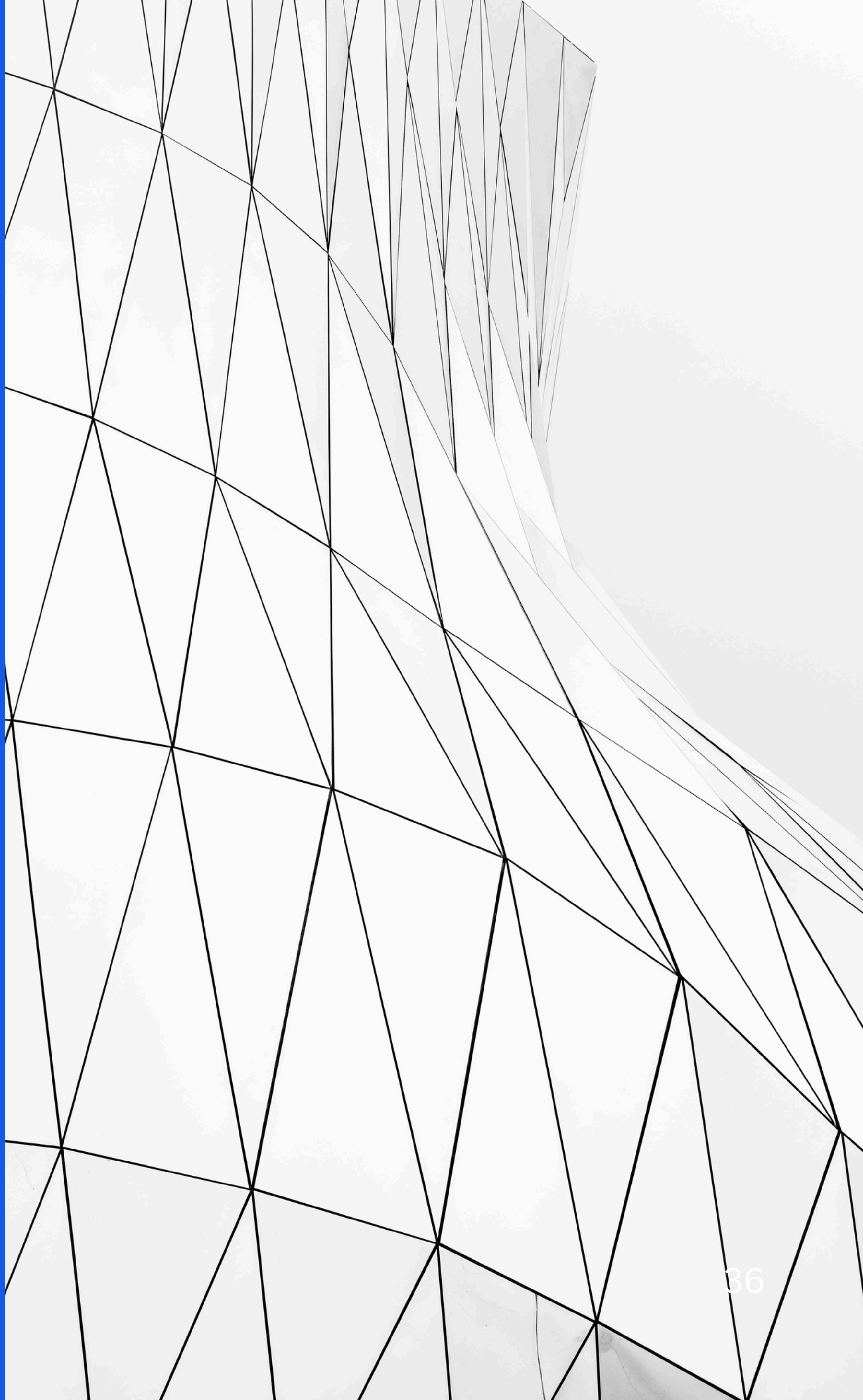
Data Leakage During Model Training (Lag Feature)

Case data was unintentionally included in training set, causing inflated performance and unrealistically low errors.

Workaround & Fix (Three Modeling Approaches):

- **Recursive Model** – Predict T+1 using valid history; feed prediction back to forecast T+2, T+3.
- **Direct Multi-Horizon Model** – Train separate models for T+1, T+2, T+3 as shown, fully avoiding future leakage.
- **In-Context Learning** – AI directly extrapolates future inbound cases from historical patterns (store, dept, weekday cycles).

Recommendations & Opportunities



Recommendations

Recommendation → Walmart

- Maintain lightweight, reliable forecasting models
- Use forecasts to create standardized labor templates
- Add anomaly detection for unexpected spikes
- Treat these departments as stable “baseline labor” so planners can focus on volatile categories

What We Want to Do for Next Steps

- Integrate richer external signals (weather, fuel prices, mobility)
- Explore AI agent–based forecasting workflow
- Build deployable Visual Studio forecasting prototype
- Provide dashboards for planners (Power BI / custom UI)

Opportunities

“At Walmart scale, small gains compound into massive operational savings.”

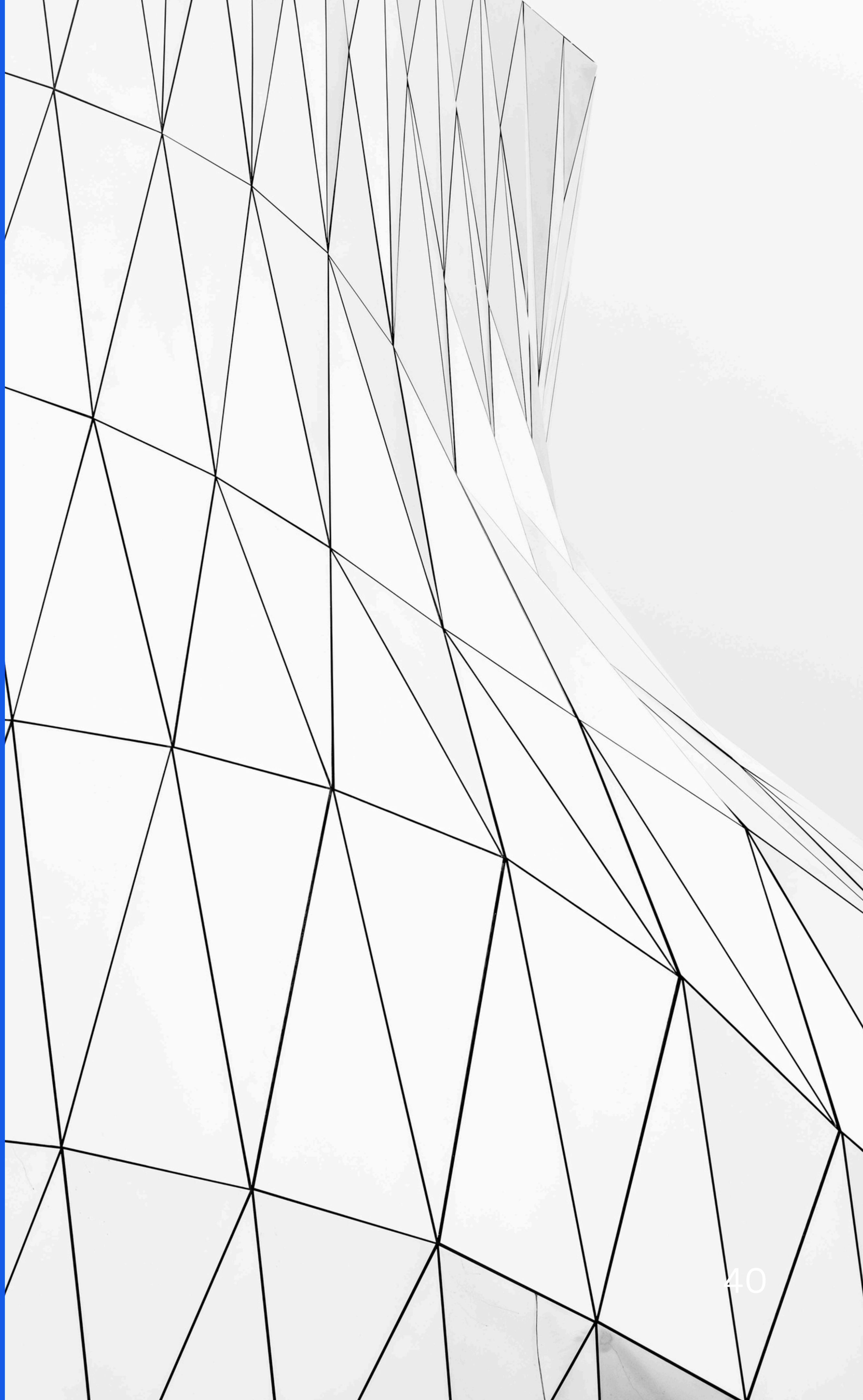
Business Impact of Current Results

- Across 4,600 stores × 40 departments, even 1% less forecast error reduces:
    - **20k–45k** weekly labor hours of overstaffing/understaffing
    - **$500k–$1.2M** per week in labor inefficiency
- Better workload matching improves associate satisfaction, retention, and operational stability
- Directly enhances customer experience via improved product availability
- **For stable departments:**
    - Lower forecasting cost by 30–50%
    - Save 3–5 hours/week of manager scheduling effort

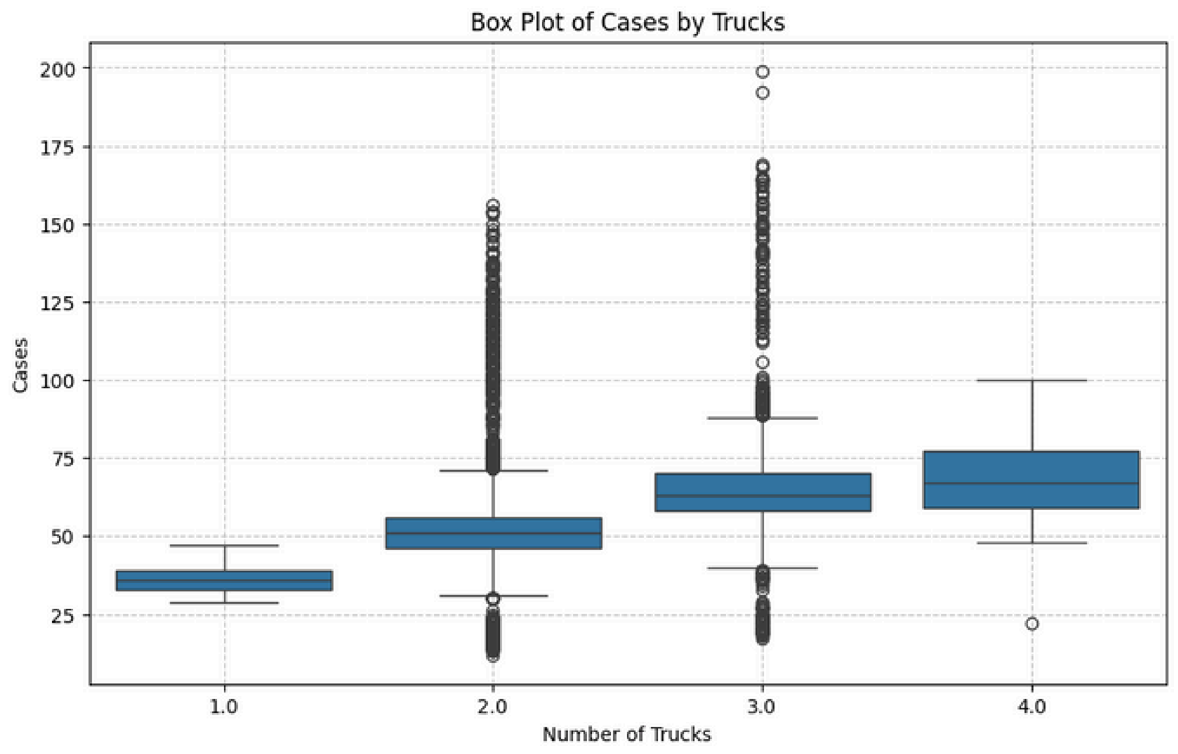
Thank You

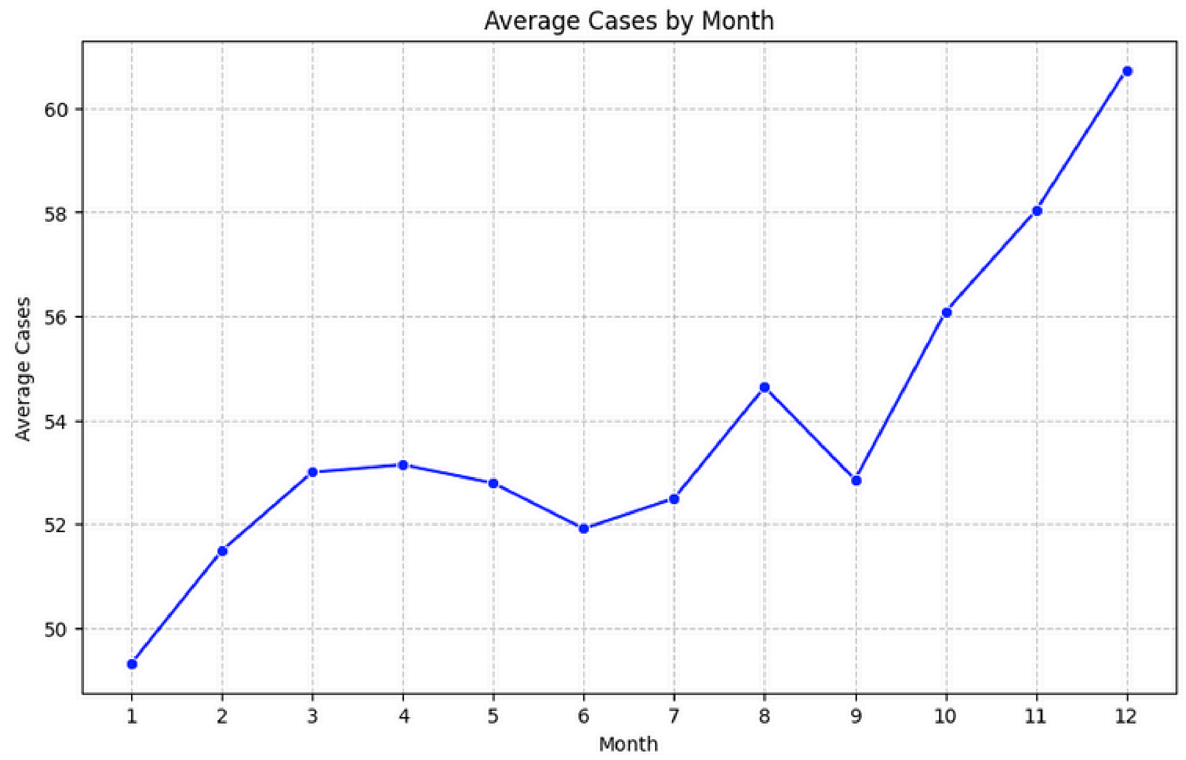
We are ready to assist you!

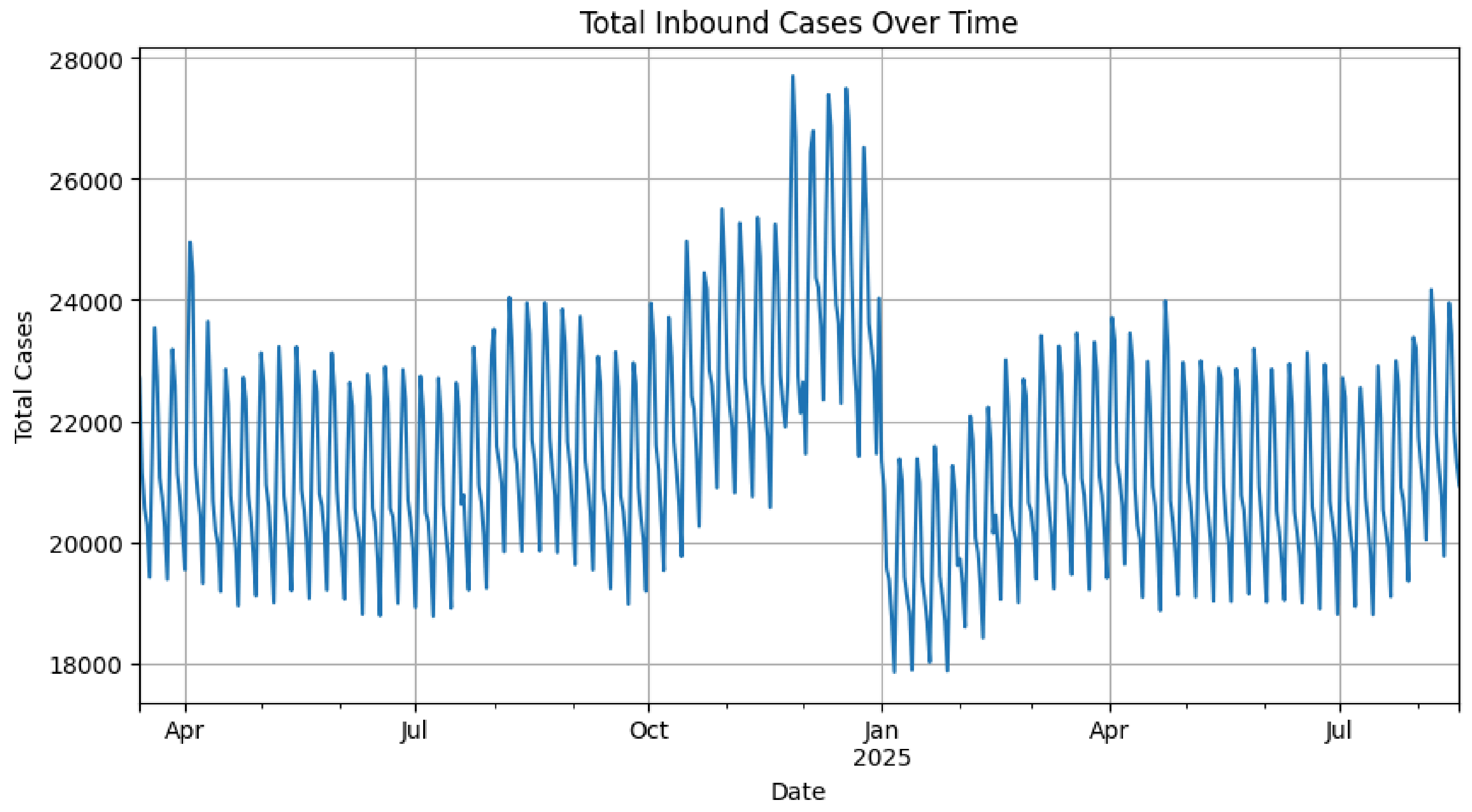
Appendix

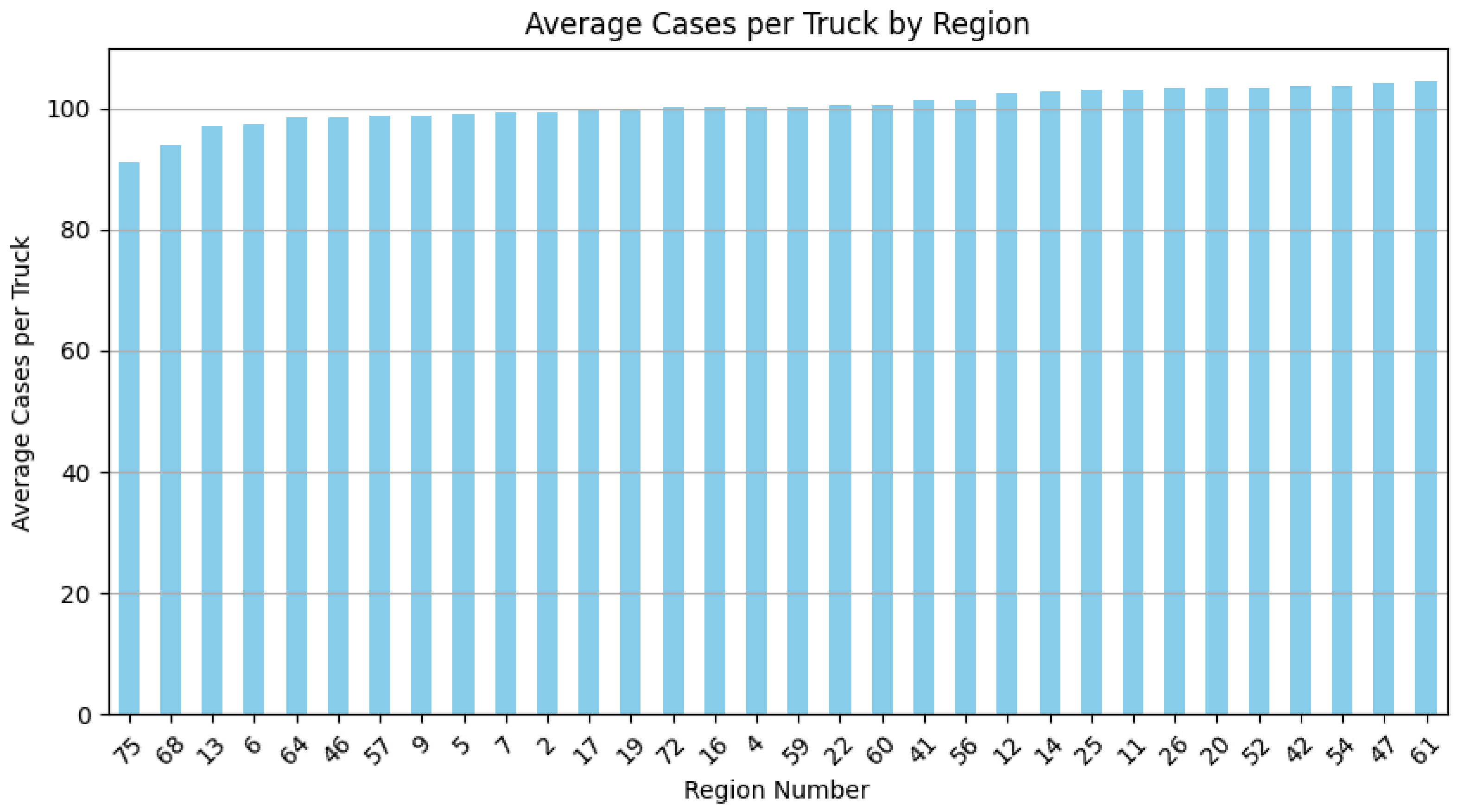


Preliminary Findings (EDA)









Baseline Model (train end date 2025-04-27)

Performance of All Baseline Model, train end date 2025-04-27

	MAE	MSE	RMSE	R2	MAPE	Bias
Linear Regression	6.86	79.36	8.91	0.08	13.88	2.22
Random Forest Regressor	3.80	33.84	5.82	0.61	7.41	0.27
LightGBM	3.38	27.39	5.23	0.68	6.61	0.04
XGBoost (n_est=500)	3.36	27.13	5.21	0.69	6.56	0.07
Optimized Weighted Average Ensemble	3.35	26.97	5.19	0.69	6.53	0.06

Performance of Ensemble Model, By Week

	week_start_date	Week	MAPE	Bias	N
0	2025-04-28 00:00:00	Week 1	6.400	0.082	2800
1	2025-05-05 00:00:00	Week 2	6.622	0.155	2800
2	2025-05-12 00:00:00	Week 3	6.823	0.052	2800
3	2025-05-19 00:00:00	Week 4	6.734	0.398	2800
4	Overall	Overall	6.645	0.172	11200

Performance of Ensemble Model, By Department

	Department Name	+7d MAPE%	+14d MAPE%	+21d MAPE%	+28d MAPE%	+7d Bias	+14d Bias	+21d Bias	+28d Bias
0	Womens (Apparel)	5.89	6.96	6.58	6.06	0.15	0.33	-0.25	0.34
1	Book & Magazines	6.33	6.36	6.77	6.57	0.01	0.03	0.38	0.47
2	Seasonal (Toys & Seasonal)	6.57	6.73	7.02	6.97	0.12	-0.14	-0.08	0.23
3	Beauty (Consumables)	6.81	6.43	6.92	7.34	0.05	0.40	0.16	0.55

Baseline Model (train end date 2025-07-20)

Performance of All Baseline Model, train end date 2025-07-20

	MAE	MSE	RMSE	R2	MAPE	Bias
Linear Regression	6.797	81.587	9.033	0.114	12.901	0.363
Random Forest Regressor	3.961	37.205	6.100	0.596	7.342	-0.668
LightGBM	3.544	30.046	5.481	0.674	6.576	-0.609
XGBoost (n_est=500)	3.444	28.835	5.370	0.687	6.433	-0.236
Optimized Weighted Average Ensemble	3.443	28.863	5.372	0.687	6.422	-0.324

Performance of Ensemble Model, By Week

	week_start_date	Week	MAPE	Bias	N
0	2025-07-21 00:00:00	Week 1	6.555	-0.092	2800
1	2025-07-28 00:00:00	Week 2	6.421	-0.568	2800
2	2025-08-04 00:00:00	Week 3	6.288	-0.332	2800
3	2025-08-11 00:00:00	Week 4	6.424	-0.302	2800
4	Overall	Overall	6.422	-0.324	11200

Performance of Ensemble Model, By Department

	Department Name	+7d MAPE%	+14d MAPE%	+21d MAPE%	+28d MAPE%	+7d Bias	+14d Bias	+21d Bias	+28d Bias
0	Womens (Apparel)	6.16	6.65	6.48	6.46	-0.38	-0.92	-0.41	-0.54
1	Book & Magazines	6.40	6.72	6.14	6.64	-0.44	-1.60	-0.50	-0.74
2	Beauty (Consumables)	6.48	6.14	6.64	6.32	0.41	0.43	-0.14	0.04
3	Seasonal (Toys & Seasonal)	7.18	6.18	5.89	6.27	0.04	-0.18	-0.28	0.03

Recursive Model (train end date 2025-04-27)

Performance of All Recursive Model, train end date 2025-04-27

	Model	MAE	MSE	RMSE	R2	MAPE	Bias
0	LightGBM	3.446156	28.074031	5.298493	0.675541	6.735946	0.104705
1	XGB	3.448812	28.093547	5.300335	0.675315	6.754296	0.202881
2	CatBoost	3.466156	28.333847	5.322955	0.672538	6.774281	0.078935
3	RandomForest	3.613775	29.920209	5.469937	0.654204	7.070510	0.167358
4	LinearReg	3.753721	31.817716	5.640719	0.632274	7.336371	0.112466
5	Ensemble_mean	3.826433	32.188319	5.673475	0.627991	7.564012	0.748545
6	SVR_linear	8.776923	149.203015	12.214869	-0.724380	17.403086	3.824925

Performance of LightGBM Model, By Department

	dept_id	Department Name	1 week MAPE(%)	2 week MAPE(%)	3 week MAPE(%)	4 week MAPE(%)	Overall MAPE(%)	1 week Bias	2 week Bias	3 week Bias	4 week Bias	Overall Bias
0	18	Seasonal (Toys & Seasonal)	6.954	6.799	7.273	7.204	7.058	0.565	-0.163	0.069	-0.033	0.110
1	21	Book & Magazines	6.569	6.553	6.984	6.768	6.719	0.507	0.141	0.362	0.402	0.353
2	34	Womens (Apparel)	6.442	6.793	7.069	6.446	6.687	0.506	0.144	-0.320	0.249	0.145
3	46	Beauty (Consumables)	7.515	6.449	6.867	7.407	7.060	0.798	0.645	0.029	0.542	0.503

Recursive Model (train end date 2025-07-20)

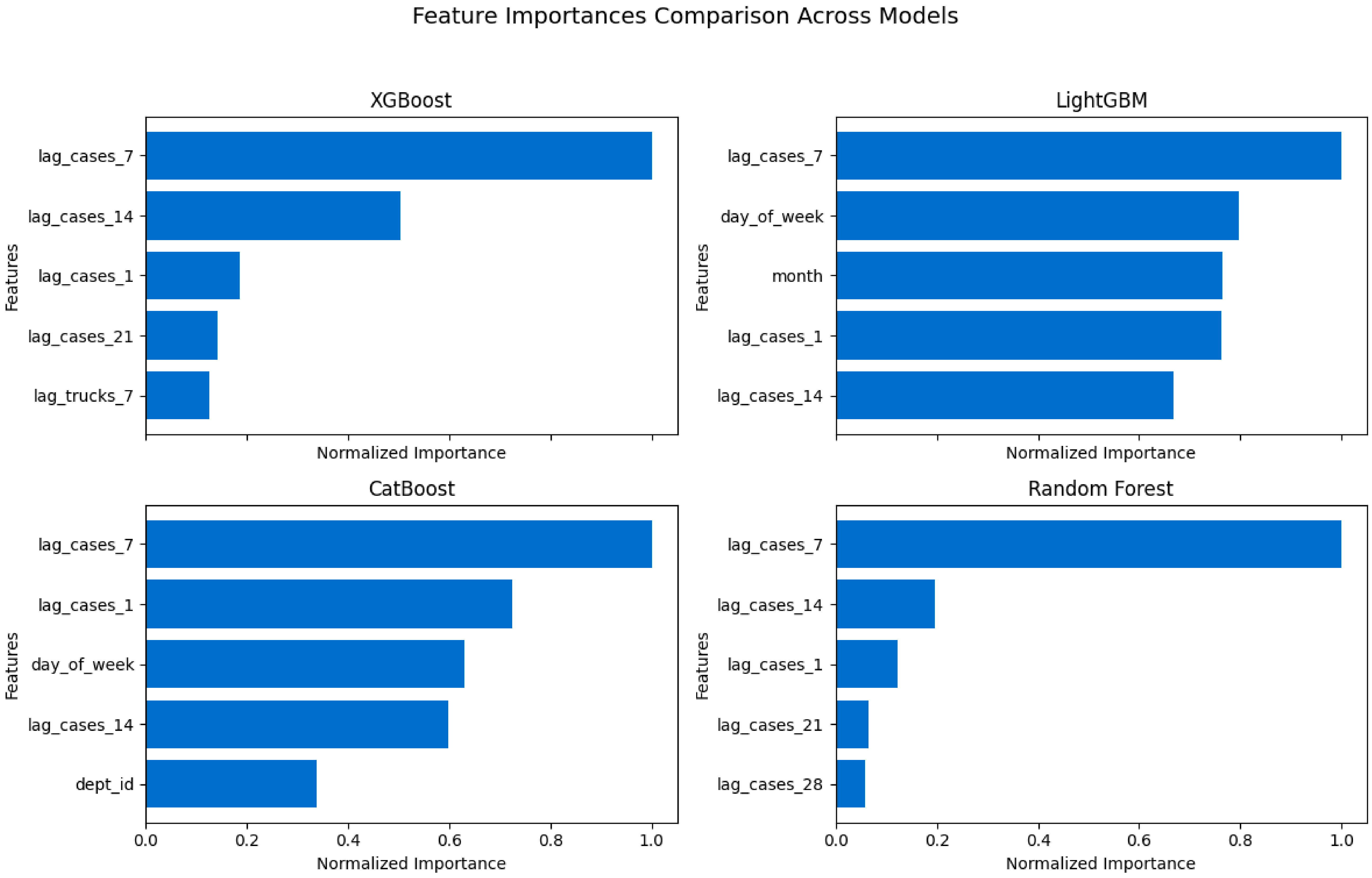
Performance of All Recursive Model, train end date 2025-07-20

	Model	MAE	MSE	RMSE	R2	MAPE	Bias
0	XGB	3.517273	29.237499	5.407171	0.680486	6.574735	-0.395381
1	LightGBM	3.526360	29.354725	5.418000	0.679205	6.584547	-0.451305
2	CatBoost	3.545819	29.640938	5.444349	0.676077	6.608270	-0.556409
3	RandomForest	3.714474	31.813586	5.640353	0.652334	6.883396	-0.864256
4	Ensemble_mean	3.741507	31.959818	5.653302	0.650736	6.957516	-0.764910
5	LinearReg	3.870304	33.963802	5.827847	0.628836	7.178794	-0.783107
6	SVR_linear	6.890293	87.918699	9.376497	0.039204	12.979136	-1.539004

Performance of XGBoost Model, By Department

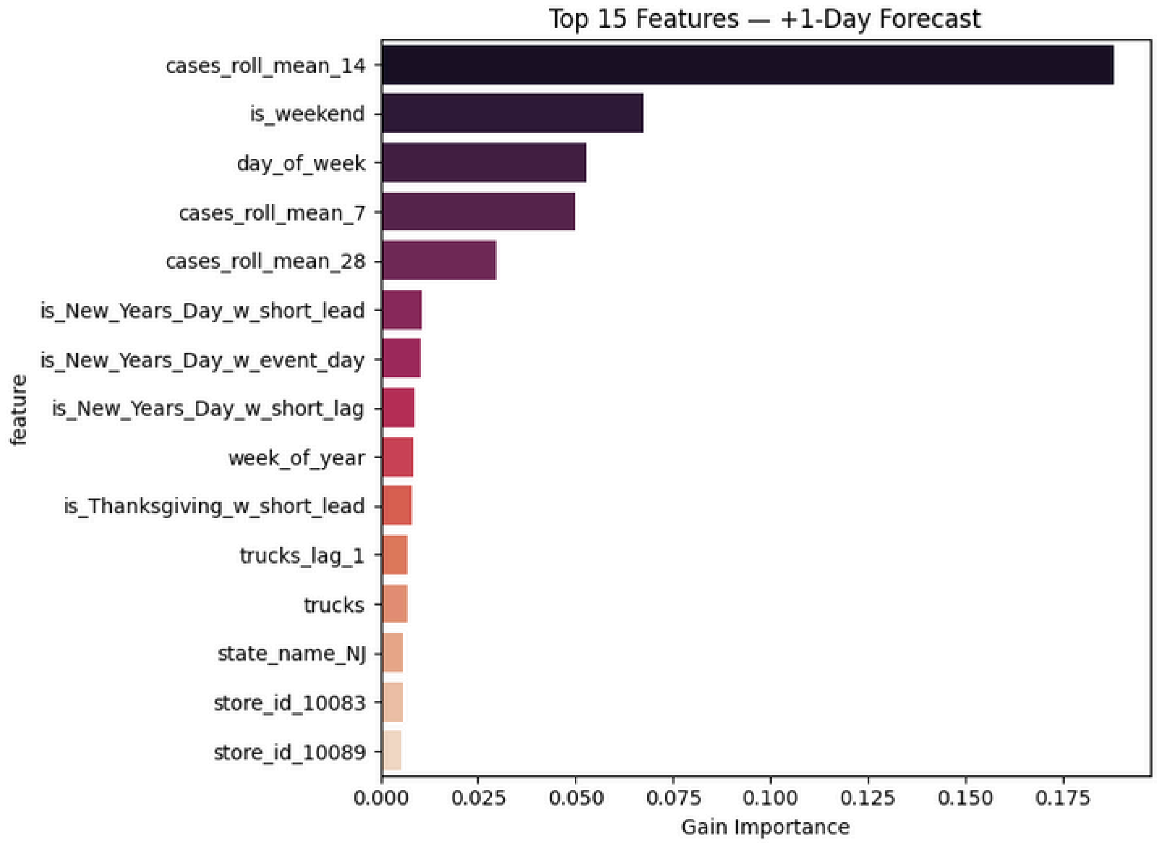
	dept_id	Department Name	1 week MAPE(%)	2 week MAPE(%)	3 week MAPE(%)	4 week MAPE(%)	Overall MAPE(%)	1 week Bias	2 week Bias	3 week Bias	4 week Bias	Overall Bias
0	18	Seasonal (Toys & Seasonal)	7.205	6.643	6.261	6.051	6.540	0.324	-0.013	-0.098	0.124	0.084
1	21	Book & Magazines	6.583	6.874	6.575	6.814	6.712	-1.033	-1.685	-0.537	-0.588	-0.961
2	34	Womens (Apparel)	6.188	6.723	6.348	6.851	6.528	-0.614	-1.283	-0.561	-0.518	-0.744
3	46	Beauty (Consumables)	6.456	6.282	6.820	6.553	6.528	0.232	0.185	-0.179	-0.077	0.040

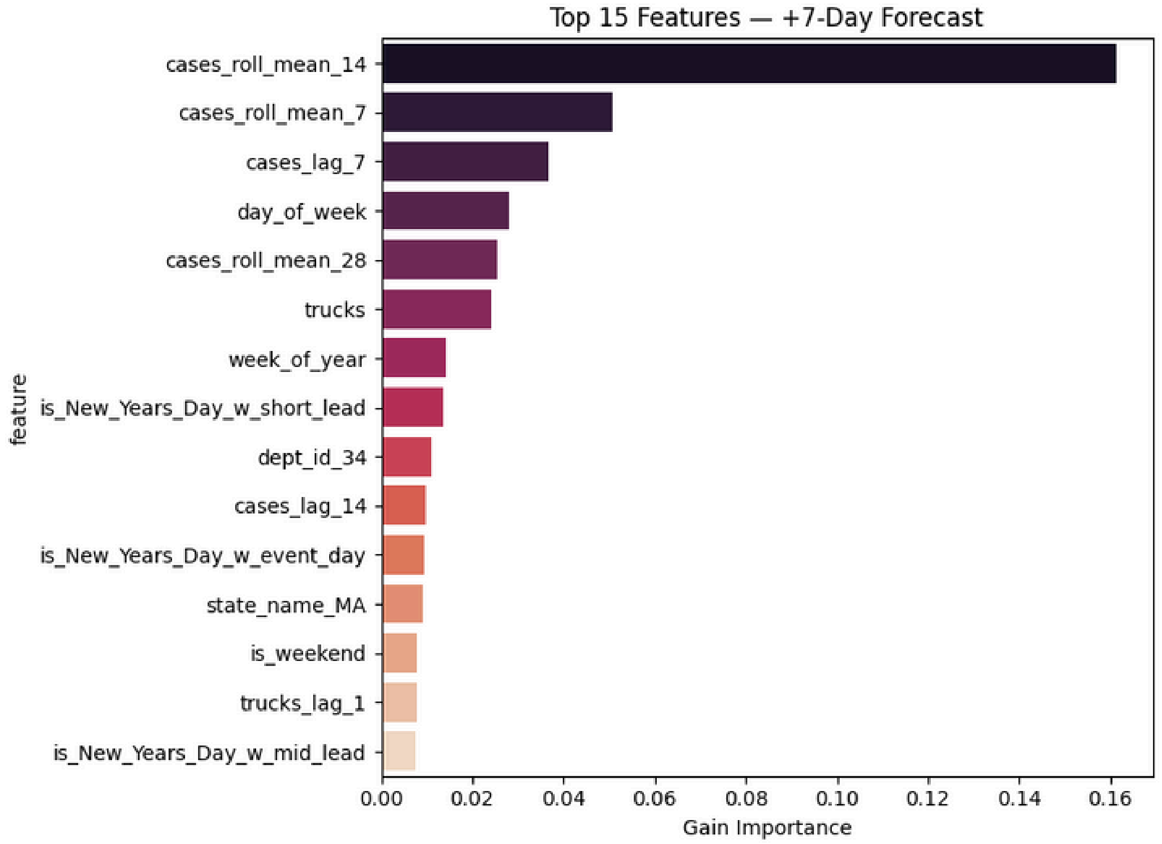
Recursive Model

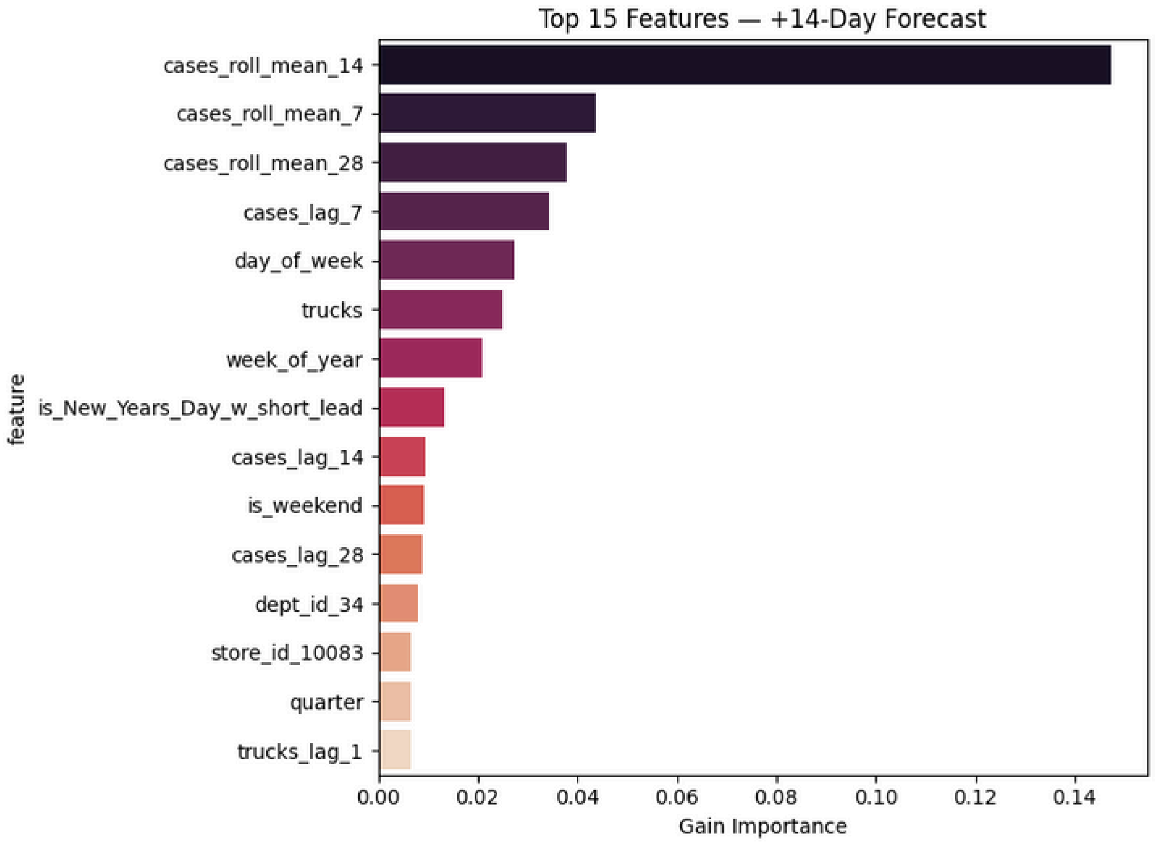


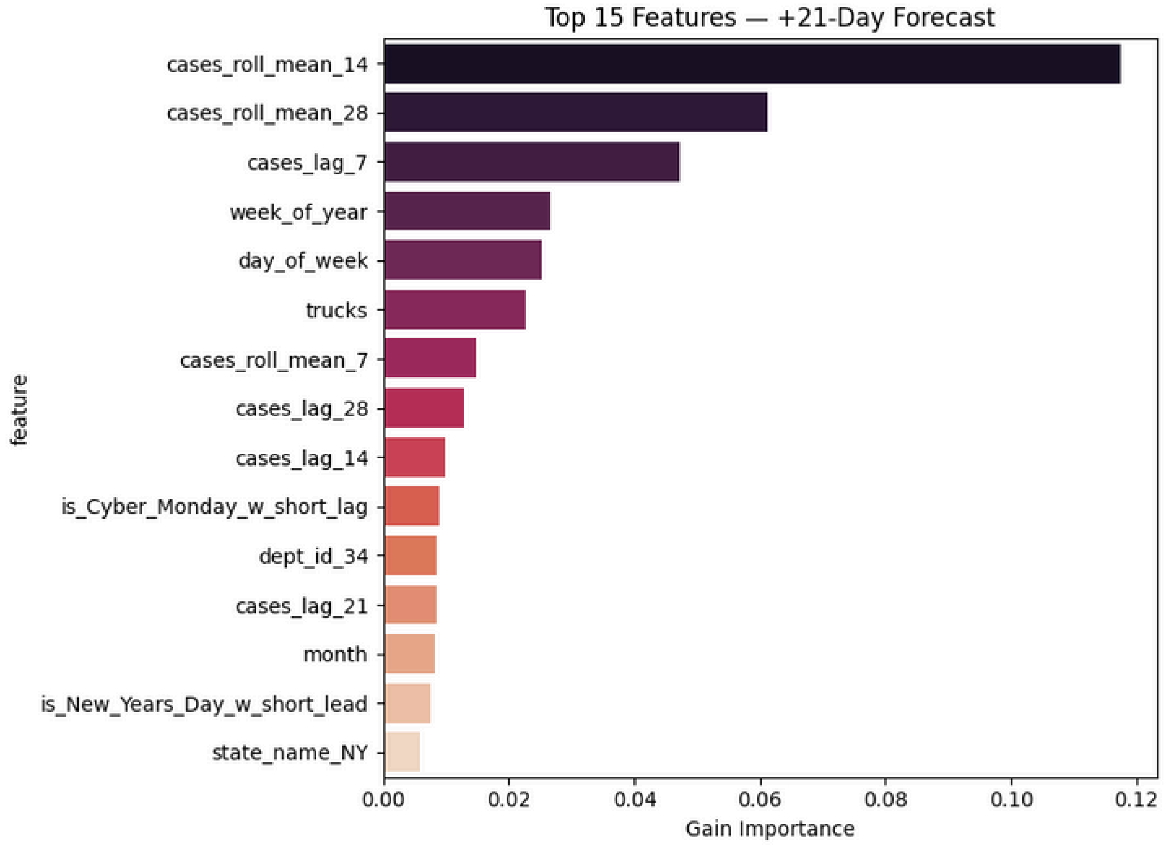
Direct Multi-Horizon Model

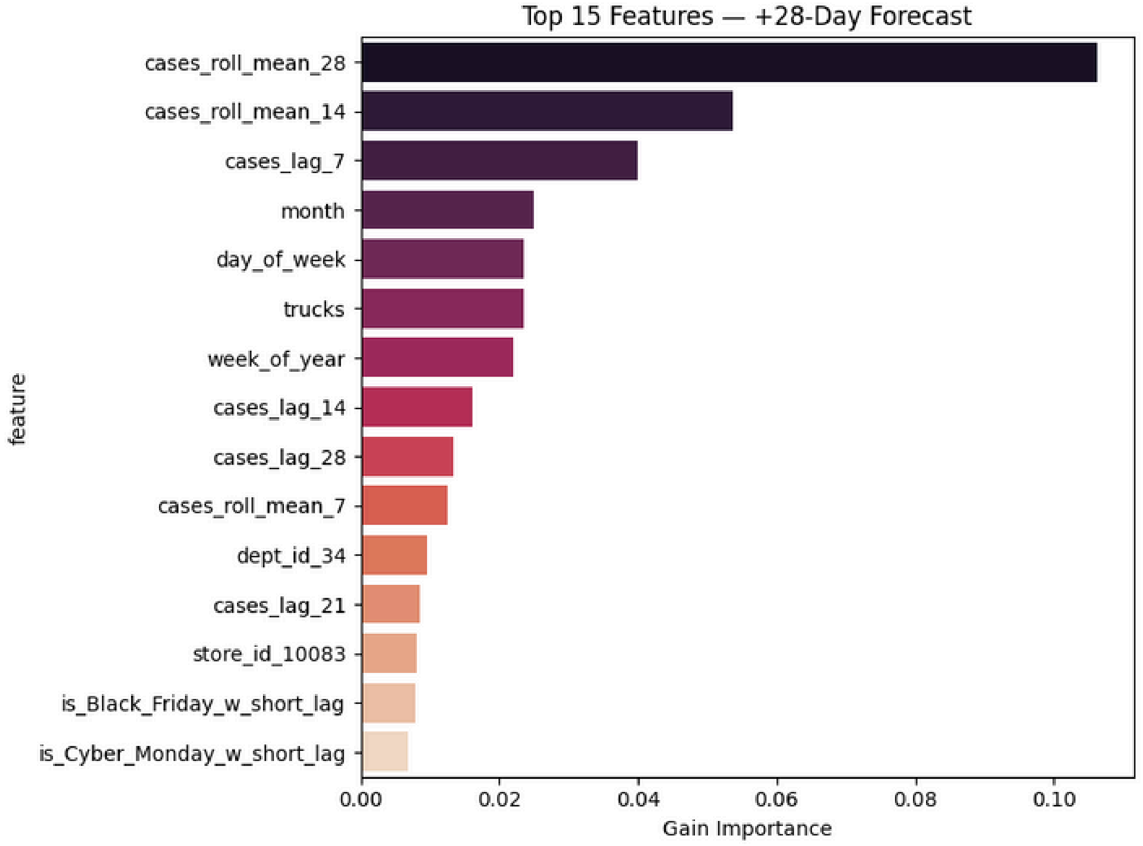
Feature Importance












WALMART