Small Warehouse, Big Profits:
Smart Inventory on a Budget

Problem Context & Business Challenge

- **Rising warehousing costs** and **supply chain disruptions** are increasing inventory complexity

- Events like **Red Sea crisis and Strait of Hormuz closure** cause longer lead times and higher working capital

- Combined with uncertain demand, companies risk:
  - **Overstock** (discounting losses)
  - **Stockouts** (lost sales)

- **Key business question:** How to **prioritise SKUs** when warehouse capacity is limited or being reduced?

www.savills.com.sg/research_articles/166122/223635-0
www.jpmorgan.com/insights/global-research/supply-chain/red-sea-shipping

Project Objective & Approach

1. Build **demand forecasting** model to improve visibility of future demand

2. Develop **SKU prioritisation** framework under capacity constraints

3. Simulate reduced **warehouse capacity scenarios**

4. **Goal: Minimise revenue loss while optimising space utilisation**

Assumption: Project to focus on scenario of local, single-store warehouse. Methodology can be scaled up to apply to larger or cross-border scenarios.

Dataset Overview & Forecasting Target

Data Source: Walmart POS dataset (2011–2016)

1. Sales

Daily unit sales (SKU–store level, wide format)

2. Calendar

Dates, events, holidays, SNAP indicators

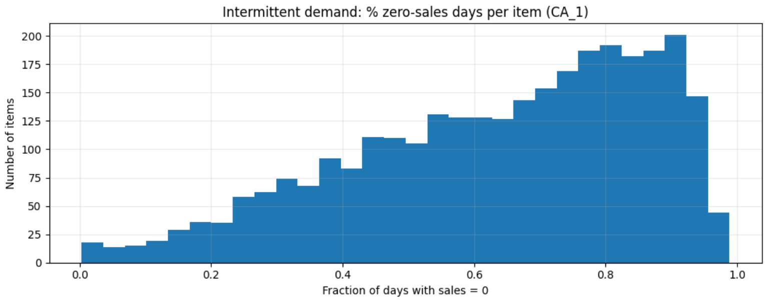
3. Prices

Weekly item prices by store

Forecasting Objective:

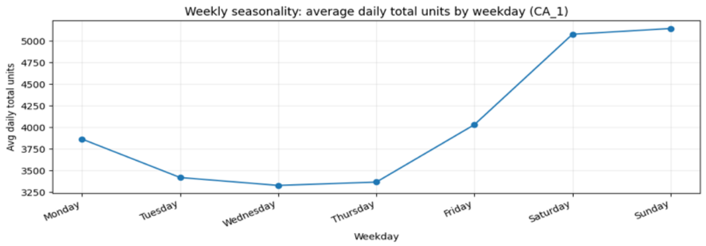
- **Predict daily demand per SKU (item_id)**
- **Predicted demand then used for prioritisation framework to support warehouse space utilisation**

Exploratory Data Analyses – Demand, Seasonality & Calendar Effects



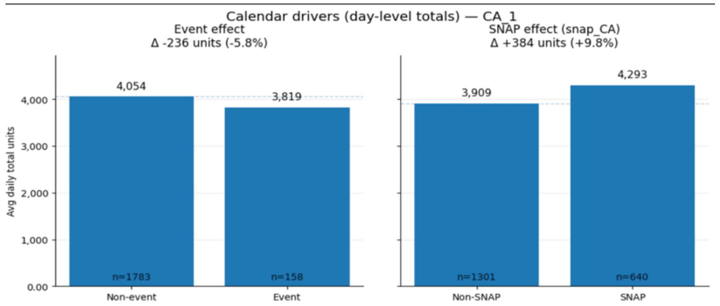
Insights:

- Many SKUs show a high proportion of zero-sales days
- This indicates intermittent demand, making daily forecasting more difficult
- To improve robustness, we used stabilising time-series features:
    - lagged demand
    - rolling statistics
    - weekly pattern features



Insights:

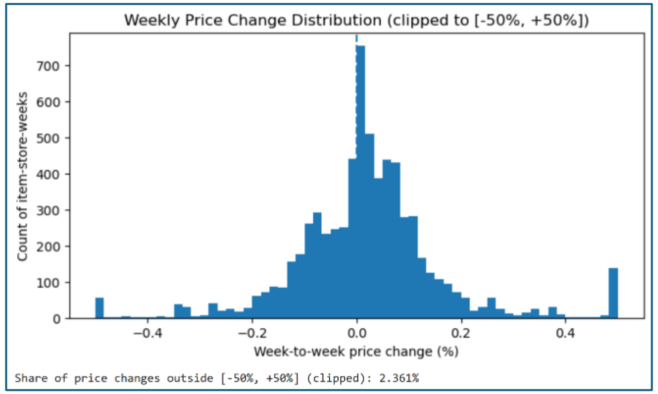
- At store level, demand shows a clear weekly seasonality pattern
- Sales are generally lower mid-week and peak on weekends
- Supports inclusion of:
    - weekday/weekend indicators
    - 7-day lag features



Insights:

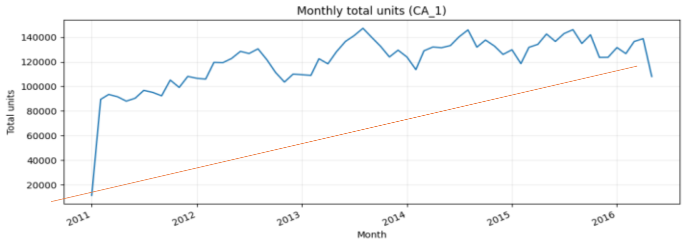
- SNAP days drive demand uplift ~+9.8% increase (≈ +384 daily units)
- Indicates that it is a consistent external demand driver
- Event effects are mixed ~−5.8% on average (≈ −236 daily units)
- Impact varies by event type and product category
- Use event-type indicators, not a single flag

Exploratory Data Analyses – Price Stability, Product Availability, SKU vs Revenue



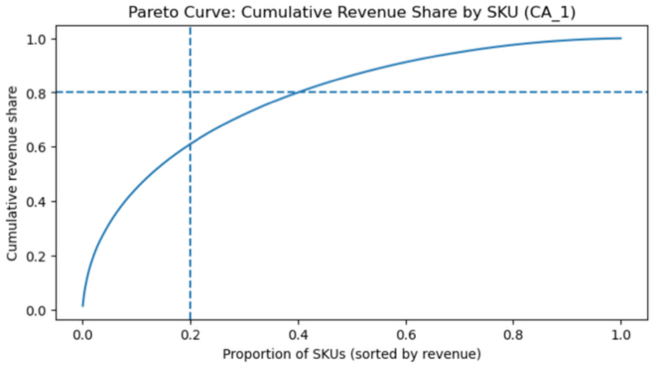
Insights:

- Prices are highly stable over time
  - Only ~1% of item–store–weeks show price changes
- Pricing follows a “stepwise” pattern
  - Long periods of constant price
  - Occasional discrete adjustments
- Most price changes are small and clustered near 0%
  - Few extreme changes (after clipping outliers)



Insights:

- Early-period demand appears lower (2011–2012)
- Driven by product availability, not true low demand
- ~28% of SKUs show no sales initially
  - Missing sales align with missing prices
  - Reflects inactive / pre-launch products
- Addressed via preprocessing: Active period definition using first price observation



Insights:

- Pareto curve shows highly concentrated revenue
- Top 20% of SKUs contribute about 60% of total revenue.
- Retain a large proportion of revenue even when selecting only a subset of SKUs under capacity constraints

Model Design, Validation & Testing Framework

Data Splitting , Training & Validation Design

1

Sliding Window (2011-2015)
Train 36 months, Validate 12 months, Step 6 months

- *No look-ahead bias*
- *Robust performance across time*

Model Training Integrity

Pipelines within each fold
Feature transformations fitted on training only

Final Model

2

Retrained on full 2011–2015 data with best hyperparameters

- *Robust – not dependent on single train-test split*

Testing Approach

3

Holdout Setup
2016 reserved as out-of-sample test set (real-world simulation)

Inputs & Model Scope

Exogeneous Variables Known in Advance
Calendar & events, and prices (fixed at 31/12/15)

- *Isolates demand model performance*
- *Models: Elastic Net, Random Forest, LightGBM, Prophet*

Recursive Forecasting (2016)

Predictions generated day-by-day
Uses previous predictions as inputs

- *Mimics real deployment (future unknown)*

Metrics For Evaluation

Accuracy: lowest average WAPE
Stability: consistent across folds
Practicality: runtime & complexity
Business relevance: strong performance on high-revenue SKUs

Data Preparation-Pricelist

Sell Price

	store_id	item_id	wm_yr_wk	sell_price
0	CA_1	HOBBIES_1_001	11325	9.58
1	CA_1	HOBBIES_1_001	11326	9.58
2	CA_1	HOBBIES_1_001	11327	8.26
3	CA_1	HOBBIES_1_001	11328	8.26
4	CA_1	HOBBIES_1_001	11329	8.26

Calendar

	date	wm_yr_wk	weekday	wday	month	year	d	event_name_1	event_type_1	event_name_2	event_type_2	snap_CA	snap_TX	snap_WI	is_weekend
0	2011-01-29	11101	Saturday	1	1	2011	d_1	NaN	NaN	NaN	NaN	0	0	0	1
1	2011-01-30	11101	Sunday	2	1	2011	d_2	NaN	NaN	NaN	NaN	0	0	0	1
2	2011-01-31	11101	Monday	3	1	2011	d_3	NaN	NaN	NaN	NaN	0	0	0	0
3	2011-02-01	11101	Tuesday	4	2	2011	d_4	NaN	NaN	NaN	NaN	1	1	0	0
4	2011-02-02	11101	Wednesday	5	2	2011	d_5	NaN	NaN	NaN	NaN	1	0	1	0

Sales Transaction Data

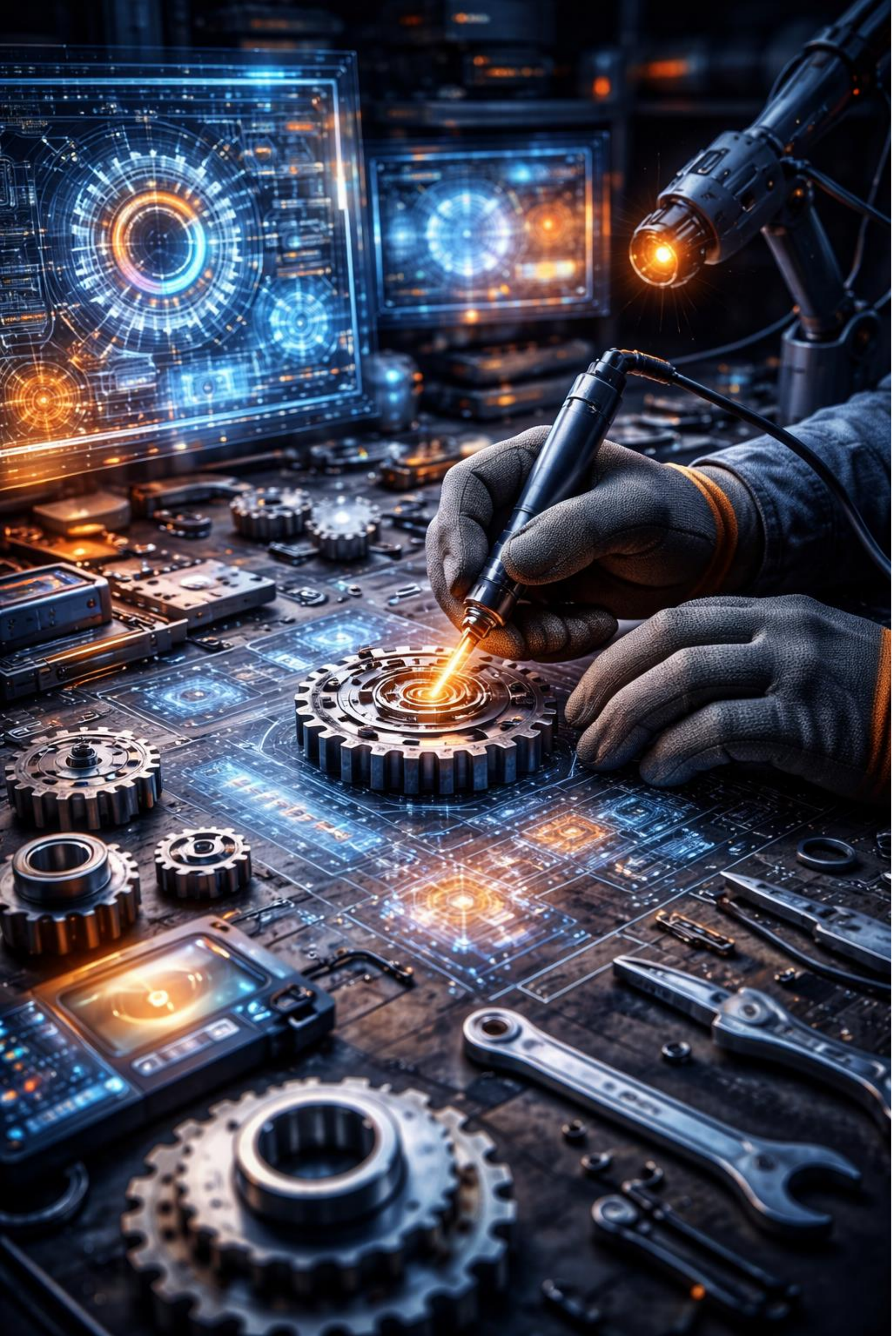
	id	item_id	dept_id	cat_id	store_id	state_id	d_1	d_2	d_3	d_4	...	d_1932	d_1933	d_1934
0	HOBBIES_1_001_CA_1_evaluation	HOBBIES_1_001	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	2	4	0
1	HOBBIES_1_002_CA_1_evaluation	HOBBIES_1_002	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	0	1	2
2	HOBBIES_1_003_CA_1_evaluation	HOBBIES_1_003	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	1	0	2
3	HOBBIES_1_004_CA_1_evaluation	HOBBIES_1_004	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	1	1	0
4	HOBBIES_1_005_CA_1_evaluation	HOBBIES_1_005	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
30485	FOODS_3_823_WI_3_evaluation	FOODS_3_823	FOODS_3	FOODS	WI_3	WI	0	0	2	2	...	1	0	3
30486	FOODS_3_824_WI_3_evaluation	FOODS_3_824	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	0	0	0
30487	FOODS_3_825_WI_3_evaluation	FOODS_3_825	FOODS_3	FOODS	WI_3	WI	0	6	0	2	...	0	0	1
30488	FOODS_3_826_WI_3_evaluation	FOODS_3_826	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	1	1	1
30489	FOODS_3_827_WI_3_evaluation	FOODS_3_827	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	1	2	0

Generic Flow

Define release week for each item/ Sku

Create full weekly time line for all items → Ensure every item has all weeks (Cross Join) → Mark whether item is active each week → Map in known price into time line → Sort Data by item and time → Forward fill prices → Final Data set with complete timeline and Prices

Note: Forward-fill missing prices only after an item becomes active; inactive weeks are not filled.



Feature Engineering

Calendar Features

- is_weekend - captures intra-week demand patterns
- snap_bool - unified SNAP indicator (state-specific)
- Event encoding:
  - event_bool (any event)
  - Event-type dummies (Sporting, Cultural, etc.) - capture varied impacts

Product Availability

- is_active flag → distinguishes:
  - pre-launch / inactive SKUs
  - vs true zero demand (e.g., stockouts)

Price Features

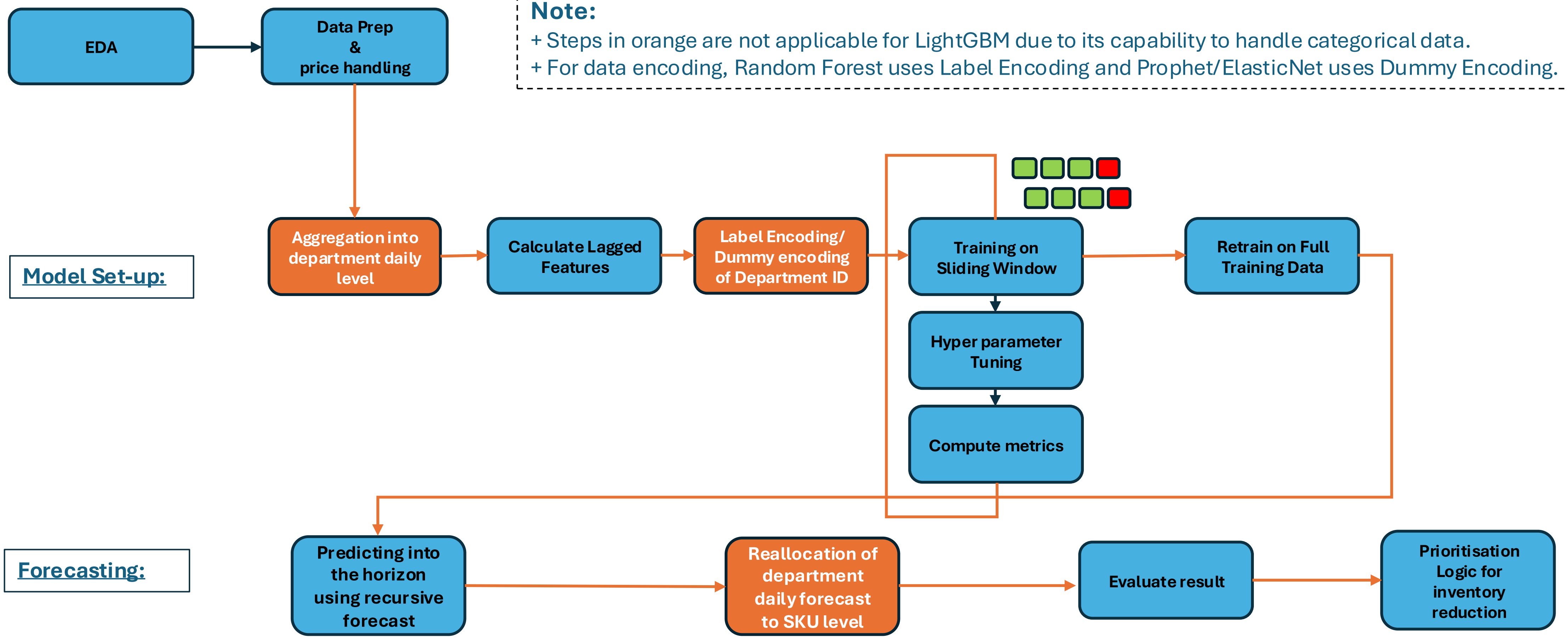
- Captures price level + promotional effects
- Rolling benchmarks (past-only):
  - 8-week median, 12-week max
- Ratio features:
  - identify if price is relatively high/low vs recent history

Demand Signals

- **Lag features:**
  - 1-day, 7-day, 28-day → short-term + seasonal patterns
- Rolling statistics:
  - 28 / 56 / 84 days → capture trend + volatility

Methodology

Note:
+ Steps in orange are not applicable for LightGBM due to its capability to handle categorical data.
+ For data encoding, Random Forest uses Label Encoding and Prophet/ElasticNet uses Dummy Encoding.



Random Forest

Why

- Demand is driven by complex, non-linear interaction (price, seasonality, SNAP and holidays).
- A need for robust model with minimal assumptions.

Limitations

- Unable to extrapolate trends beyond observed data, which may lead to underestimation in uptrend.

Reallocation of department daily forecast to SKU level

- Improve model stability and reduce noise.
- Reduce computational complexity.

Label Encoding of Department ID

- Does not impose strict linear assumptions on category relationships

Aggregation into department daily level

- Forecast by department level loses SKU granularity.
- Need to reallocate using historical blended weights.

Dept X Mth Performance	WAPE	MSE	RMSE	MAE	R2
Training Validation	0.0103	2,339,541	1,517	1282	0.9969
Holdout Validation	0.0294	543,987	738	540	0.9984
Recursive Forecast*	0.1502	21,345,316	4, 620	2, 754	0.9380

* Random Forest and Prophet are forecasted at department level and reallocation based on past sales weightage by department daily level

Point 1:

- WAPE for Training validation is at **0.0103**
- WAPE for Holdout validation is at **0.0294**
- **R-square is at ~0.99 which means most of the variance are explained.**
- Monthly department recursive forecast and reallocation accuracy is at **0.1502**. This is surprisingly better than expected.

Point 2:

Performance by department:

	dept_id	total_abs_error	mae	mse	rmse	wape	r2
0	FOODS_1	2765.328832	553.065766	4.581630e+05	676.877391	0.061381	0.312855
1	FOODS_2	6082.684823	1216.536965	1.594357e+06	1262.678341	0.080347	0.343389
2	FOODS_3	51298.628880	10259.725776	1.207221e+08	10987.361901	0.168961	-4.655385
3	HOBBIES_1	21808.236187	4361.647237	2.159031e+07	4646.537761	0.311302	-9.003005
4	HOBBIES_2	1079.856946	215.971389	5.726543e+04	239.301973	0.186375	-10.675958
5	HOUSEHOLD_1	10397.917678	2079.583536	4.531866e+06	2128.817932	0.093228	0.388325
6	HOUSEHOLD_2	2976.583300	595.316660	4.631305e+05	680.536947	0.099425	-0.517284

Performance by forecast month:

	month	total_abs_error	mae	mse	rmse	wape	r2
0	2016-01	10196.521680	1456.645954	3.634455e+06	1906.424714	0.077482	0.988879
1	2016-02	15129.763386	2161.394769	1.188106e+07	3446.892030	0.119457	0.963192
2	2016-03	19385.670992	2769.381570	1.953748e+07	4420.121997	0.141960	0.948569
3	2016-04	27123.174860	3874.739266	3.656119e+07	6046.584593	0.195375	0.911845
4	2016-05	24574.105728	3510.586533	3.511240e+07	5925.571646	0.227426	0.867498

- **Monthly WAPE** is approx. **0.07 to 0.22**
- **Department WAPE** is ranging from 0.33 to 0.5.
- **Worst performing department is coming from Hobbies_1 (WAPE = 0.31).**
- **Multiple –ve R-square and higher WAPE indicates that the model might not be that suitable to forecast on these categories.**

R^2 becomes unstable when variance is low and sample size is small.

Prophet

What

A **time-series forecasting model** that decomposes demand into trend, seasonality, holiday effects and external regressors.

$$y(t) = g(t) + s(t) + h(t) + X(t)^T \beta + \epsilon_t$$

g(t) represents the trend component, capturing long-term demand growth/decline.
s(t) represents seasonality, such as weekly patterns in store sales.
h(t) captures the effect of events or holidays.
X(t)^T.β represents external regressors, such as SNAP participation or event indicators.
ε_t the error term.

Why

- It **captures recurring seasonal patterns** and can adapt to changes in trend over time.
- It is interpretable and easy to understand, allowing us to see whether each regressor has a positive or negative association with the target variable.

Limitation

- Its **performance may be weak** when **external regressors have little relationship** to the time-series pattern or do not show stable effects over time.
- It may not perform well for highly irregular intermittent demand.
- It has **limited ability to capture complex nonlinear relationships** compared with more advanced machine learning models.

Insight

- Validation performance: 3% WAPE and 99% R2.
- Recursive forecast performance: 5% WAPE and 99% R2.
- For Forecasting, the performance is **better at the department level** (WAPE 0.03-0.05 WAPE and 0.99 R2) than at the SKU level, and results are more stable when historical demand is **aggregated monthly.**
- Department performance: **WAPE for FOOD depts performs well at 3-4%** while and WAPE for **HOBBIES_2** is still worse at 8%. HOBBIES_2 department likely has low demand that the model cannot learn well.
- The correlation analysis shows that **lag and rolling features are strongly positively correlated with sales**, indicating that historical sales patterns are highly informative for forecasting.

Aggregated performance	WAPE	MSE	RMSE	MAE	R2
Validation	0.03	861,593.53	927.30	625.14	0.99
Recursive forecasting	0.05	2,378,069.73	1,542.09	922.91	0.99

*Aggregated per month department_id level.

Performance by dept_id:

	dept_id	MSE	RMSE	MAE	R2	WAPE
0	FOODS_1	266,706.4524	516.4363	421.4662	0.6000	0.0468
1	FOODS_2	245,559.3575	495.5395	455.6219	0.8989	0.0301
2	FOODS_3	10,745,777.0970	3,278.0752	2,747.9425	0.4966	0.0453
3	HOBBIES_1	558,590.4755	747.3891	620.2791	0.7412	0.0443
4	HOBBIES_2	14,715.0996	121.3058	97.9535	-2.0003	0.0845
5	HOUSEHOLD_1	4,719,844.5359	2,172.5203	1,855.7779	0.3630	0.0832
6	HOUSEHOLD_2	95,295.1460	308.6991	261.3085	0.6878	0.0436

Elastic Net (Baseline)

What

Regularised linear model for daily SKU demand forecasting, using lag, rolling, calendar, price, and categorical features. It provides a transparent and auditable view of demand drivers.

Why

- **Combines L1 + L2 regularisation** to **manage correlated engineered features**, shrink unstable coefficients and stay computationally efficient. Useful when explainability matters as much as raw predictive power.
- Useful when **explainability and coefficient-based interpretation are important** alongside predictive performance.

Limitation

- Its linear structure smooths peaks and troughs, so it **cannot fully capture nonlinear relationships**, intermittent demand, or sharp spikes.
- It may **perform less well for sparse, intermittent**, or highly volatile demand.
- In **recursive forecasting, errors can accumulate over longer horizons**.

Insights

- Validation performance: 3% WAPE and 99% R2.
- Recursive forecast performance: 19% WAPE and 91% R2.
- **Validation is strong after aggregation**, but **2016 recursive holdout** results **weaken materially** once the forecast rolls forward month by month
  => Its strength is interpretability and governance, making it **a useful baseline** rather than the leading forecasting model.
- **Stronger fit** in smoother departments such as **FOODS_2 / FOODS_3** during validation.
- **Error** concentrates in sparse / volatile categories such as **HOBBIES_2** and **HOUSEHOLD_2**.

Aggregated performance	WAPE	MSE	RMSE	MAE	R2
Validation	0.03	695,404	833.7	450.8	0.99
Recursive forecasting	0.19	28.1 million	5,303.9	3,444.8	0.91

*Aggregated per month department_id level.

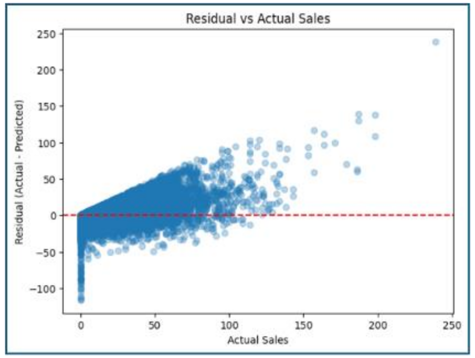
dept_id	wape	mse	rmse	mae
HOUSEHOLD_2	0.379752	6.430137e+06	2535.771509	2273.801410
HOBBIES_2	0.272989	1.291909e+05	359.431351	316.339437
FOODS_2	0.206463	1.066301e+07	3265.427096	3126.060882
FOODS_3	0.196970	1.599068e+08	12645.427855	11960.534565
HOUSEHOLD_1	0.169219	1.486158e+07	3855.071917	3774.663074
HOBBIES_1	0.127070	4.000571e+06	2000.142766	1780.372024
FOODS_1	0.097839	9.291430e+05	963.920621	881.567195

LightGBM

What - Gradient boosting tree model designed for **fast, large-scale, high-dimensional data**

Why

- Capable of handling high-dimensional categorical without one-hot encoding
- Direct modelling at the SKU–day level, preserving granularity and capturing calendar effects
- Category-specific interactions



Limitations

- Smooths intermittent demand
- For low or zero actual sales: model overpredicts demand.
- For high sales/spike: model systematically underpredicts high-demand periods or spikes.
- Leads to weaker performance for **HOBBIES_2 (sparse, volatile demand).**

Insights

- For both validation and testing, model is **more reliable at aggregated monthly department level**. Thus, **suitable for higher level planning, e.g. warehouse capacity.**
- For validation, **performance is consistent across time** (i.e. folds).
- For testing, **performance deteriorates over time driven by error accumulation in recursive forecasting.**
- Performance **varies across departments**, reflecting differences in demand patterns. Most have an acceptable WAPE at ~5 – 17%, except for HOBBIES_2 (highest WAPE at 56%).

Model	Description	WAPE	MSE	RMSE	MAE	R2
LightGBM	Validation	0.03	751K	861.85	543.6	0.99
LightGBM	Testing (Recursive Forecasting)	0.14	20 million	4,483	2,599.3	0.94

*Aggregated per month department_id level.

Results Across Models & Decision Implication

Observations

- RMSE, MSE, and MAE are **scale-dependent, resulting in large absolute values** driven by high-volume SKUs.

- Thus, WAPE is used as the primary metric as it:
  1. **Reflect relative forecast error (% demand)**
  2. **Comparable** across SKUs with different demand scales
  3. More aligned with planning decisions

- Daily SKU-level demand is noisy and intermittent. Aggregated views (of monthly and dept id) provides more stable signal.

Insights

- Accuracy varies across models.
- **Each model captures different aspects of demand patterns, indicating that no single model is universally optimal.**

Model	WAPE	Note
Prophet	**0.05**	Time series model capturing trend and seasonality
LightGBM	**0.14**	Boosting-based model that learns from errors
Random Forest	**0.15**	Captures nonlinear relationships and interactions; more robust to noise
ElasticNet (Baseline)	**0.19**	Interpretable linear model with regularisation

Key Decision for Prioritisation:

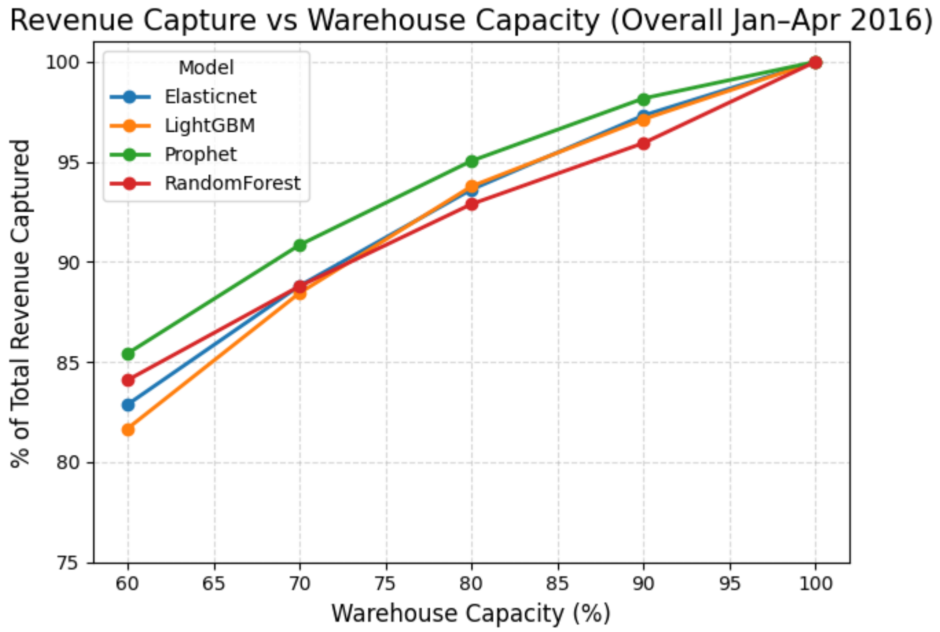
Given each model's strengths and limitations, we use all **four models to obtain the maximum of revenue loss (%) across** to represent a worst-case scenario. This enables more robust and risk-aware decision making process.

Prioritisation

Warehouse capacity	Overall % Rev Loss	Jan 2015 Rev Loss %	Feb 2016 Rev Loss %	Mar 2016 Rev Loss (%)	Apr 2016 Rev Loss (%)
90%	4.06	7.22	2.04	5.55	3.2
80%	7.11	10.32	5.74	9.55	6.8
70%	11.55	15	10.1	14.38	11.61
60%	18.32	20.85	16.34	19.46	18.57

Framework & Result

- Prioritisation combines **expected revenue per unit and seasonality**, ensuring capacity is given to the most impactful SKUs first. Thus, revenue remains controlled even with some dept having a higher WAPE.
- **Acceleration of loss** as capacity tighten:
  1. Dropping from 90% to 80% "costs" ~3.05% in revenue.
  2. Dropping from 80% to 70% "costs" ~4.44% in revenue.
- The revenue loss **varies month-to-month**, reflecting different demand patterns and seasonality effects. But, **sharper losses observed as capacity drops below 80%.**
- All **four models converge at the 80% scenario**, showing it is a “stable zone”.



Recommendation

- **Operate at ~80% capacity as a balanced strategy:**
  1. Overall loss is 7.11%, with highest in Jan is ~10%.
  2. Achieves a meaningful reduction in capacity while preserving most revenue
  3. Provides a critical buffer
- However, the final decision should consider **trade-off between revenue loss and warehouse cost savings,** and depend on the business’ risk appetite.

Challenges, Future Work, Recommendation

Challenges & Future Work

Key challenges

- **Intermittent / sparse demand** and **category heterogeneity** reduce SKU-level accuracy, especially in volatile groups.
- **Department-to-SKU reallocation** can support top-down consistency but lowers item-level precision.
- **Feature refresh is still limited** for price, promotion, event, and calendar-related drivers.

Future work

- **Different models for different department or ensemble models** by demand pattern rather than forcing one model to fit every SKU.
- **Test direct multi-horizon (sales lag) forecasting** and add richer business drivers such as promotions, stock-outs, and price signals.
- **Scale to more stores only after stability** in the data pipeline, monitoring layer, and unified planning dashboard.

Recommendation: Planner-Centric

1. Build a **planner-facing dashboard** for comparison view so planners can assess model outputs, ranges, and business trade-offs in one place. E.g. planner can set different scenario bounds or benchmark from various models.
2. Work closely with planner on weight decisions toward **monthly / department forecasts** for capacity planning; so that daily SKU forecasts should remain **directional** for volatile items.

Operating logic

Multiple model outputs	Dashboard	Planner Judgement	Informed Decision

Bottom line: The goal is not to choose the best model, **but equip planners with transparent view** of model output and uncertainty, enabling more robust decision making.

Annex - References

Baum, C., Hauer, M., Joglekar, A., & Turco, A. (2023, May 8). *Thinking Beyond Markdowns to Tackle Retail's Inventory Glut*. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/thinking-beyond-markdowns-to-tackle-retails-inventory-glut

Brooks, S., Schuham, R. (2025, Jul 11). *Warehousing Costs: A Global Perspective*. Savills Singapore. https://www.savills.com.sg/research_articles/166122/223635-0

JPMorgan Chase & Co. (2024, Feb 8). What are the Impacts of the Red Sea Shipping Crisis https://www.jpmorgan.com/insights/global-research/supply-chain/red-sea-shipping

Steyerberg, E., Vickers, A., Cook, N., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M., & Kattan, M. (2010). Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology, 21, 128-138. https://doi.org/10.1097/ede.0b013e3181c30fb2

Dehghani, M., Arnab, A., Beyer, L., Vaswani, A., & Tay, Y. (2021). The Efficiency Misnomer. ArXiv, abs/2110.12894. https://arxiv.org/pdf/2110.12894