AUGUST 2025

TEXAS
The University of Texas at Austin

ANALYSIS OF TARGET'S E-COMMERCE SALES DATA

Data Science Programming Project

GROUP 1
MSBA Students, The University of Texas at Austin

Team Introductions

John MacDonald

Vyshnavi Maringanti

Sarvesh Miskin

Sanchal Nachappa

Suryah Vadivel

Introduction

Business Problem Statement:

Target’s operations need accurate, timely, and actionable demand forecasts to avoid overstocking, stockouts, and inefficient workforce allocation. Current planning processes are fragmented and reactive, leading to lost sales and excess costs

Goal:

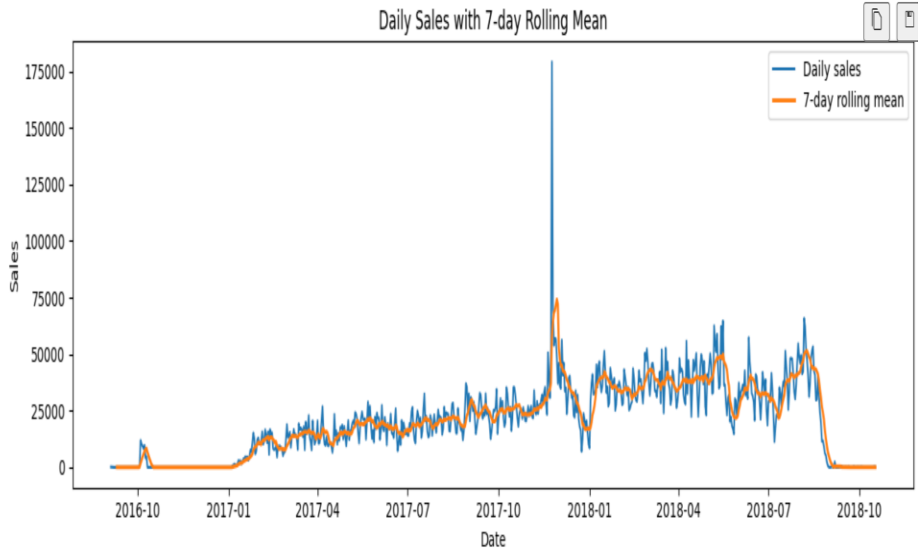
Forecast daily total sales and convert these forecasts into integrated operations plans covering inventory replenishment, staffing allocation, logistics scheduling, and promotional timing.

About Dataset

- This dataset focuses on Target’s operations in Brazil between 2016 and 2018
- This dataset has 7 files for each feature : customers, sellers, order items, geolocation, payments, orders and products
- The dataset in total covers 16M different sales
- Time Span:
  September 4, 2016 → October 17, 2018 (774 days)
- Granularity:
  Daily total sales aggregated from payments, based on order purchase date.

Business Value and Importance

- This dataset represents a realistic view of transactional data.
- Represents an analysis that is common in industry
- Insights gained from analysis can drive future profits
  - Optimize stocking by region
  - Where to digitally target customers



Exploratory Data Analysis

- The dataset in total covers 16M different sales
- Data Preparation:
- Joined orders and payments datasets.
- Filled missing calendar dates with zero sales to ensure continuous time series.
- Standardized formats for consistency in analysis and modeling.
- Rolling 7-Day mean shows short-term momentum
- Observed a spike in high-value days (right-tail) followed by a decline in late 2018

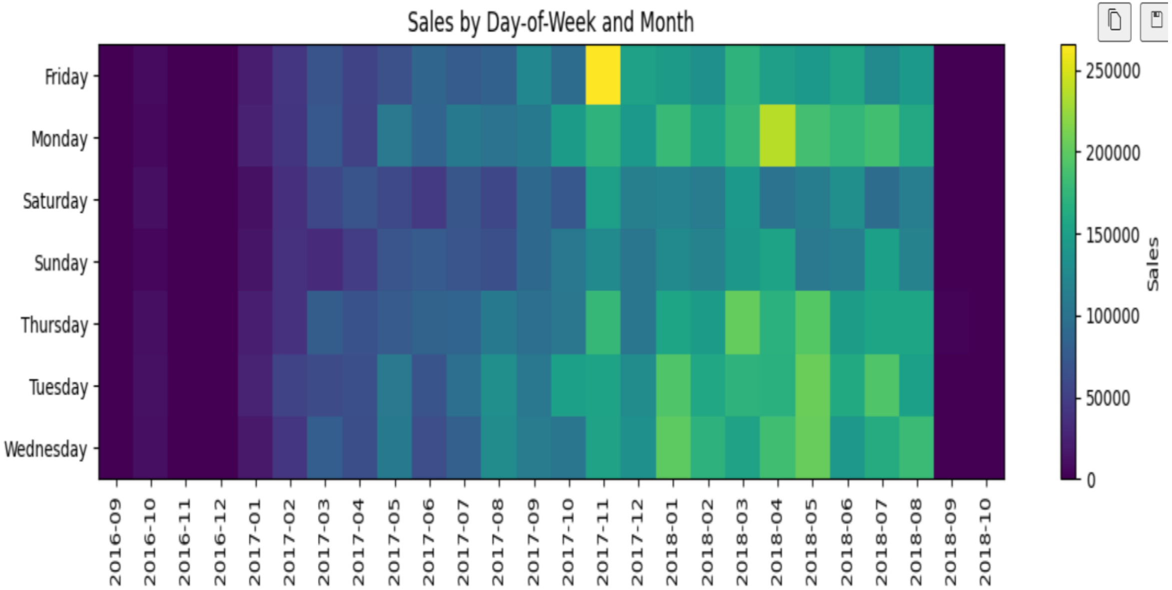
The sales data has some common statistics

Mean = 20.7k

Maximum = 179.2k

Median = 20.5k

Exploratory Data Analysis



Insights from the Heatmap:

- **Strong weekly rhythm** — sales patterns repeat consistently each week.
- **Notable November surge** — aligns with seasonal promotions/events.
- **Spike anomalies** visible as dark cells outside normal seasonal trends.

Takeaway for Modeling:

- Focus evaluation on the **last 30 "normal" days** before the dataset's zero tail to ensure models aren't biased by abnormal periods.

Model Features

Training & Evaluation:

- Train: 2016-09-04 → 2018-08-04
- Test: 2018-08-05 → 2018-09-03

Models Tested:

- Naïve (lag-1)
- Seasonal-Naïve (lag-7)
- Linear Regression
- Gradient Boosting

Feature Engineering:

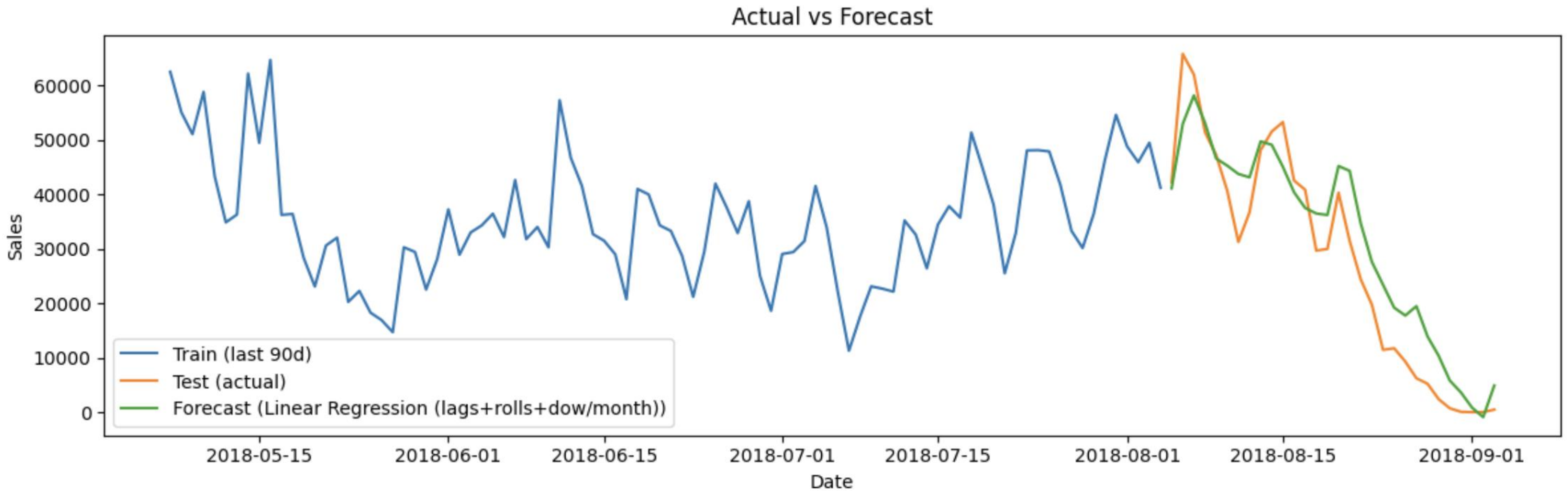
- Time Lags: 1, 2, 3, 7, 14 days
- Rolling Stats: 7-day & 14-day rolling mean and std

Evaluation Metrics:

MAE, RMSE, SMAPE,
WAPE (robust to zeros)

Modeling Results

- Best model was Linear Regression



- MAE = $6,102, RMSE = $7,265, SMAPE= 56.7, WAPE ≈ 21.9% on test data
- Simple, well-chosen features significantly outperform naive and seasonal baselines by a wide margin.

Model Analysis and Insights

- Feature importance (from Gradient Boosting) identifies key drivers:
  - 7-day rolling mean and standard deviation
  - Lag-1 value
  - Day of week
- Indicates strong **weekly seasonality** and recent momentum are dominant influences on target variable.
- Daily forecasts to support inventory, staffing, logistics, and promotional planning
- Transparent, easy to update model
- Deliverables include:
  - Model performance comparison CSV
  - Feature importance rankings CSV
  - Visual actual-vs-forecast plots for interpretability and validation.

Conclusions and Recommendations

- Planning outputs show in Excel
- Operational planning outputs centralized in **ops_planning_pack.xlsx** for easy stakeholder access.
- Critical inventory metrics generated: daily demand lead time (DLT), safety stock, reorder points.
- Staffing and logistics planning supported with forecasted orders, labor hours, headcount.
- Assumptions are tunable, enabling operations teams to customize and optimize plans.

Conclusions and Recommendations

Metric	Value	How it's computed
Forecast sales	58,211	model output
Orders	582	sales ÷ AOV (100)
Labor hours	58.2 h	orders × 6 min ÷ 60
Headcount	8 people	ceil(labor hours ÷ 8)
DLT (5 days)	1,582 orders	avg daily orders × 5
Safety stock	511 orders	$z \times \sigma \times \sqrt{LT}$ (z=1.28)
Reorder point (ROP)	2,094 orders	DLT + safety

How teams use the outputs

Model output	Who uses it	How they use it	Example decision
Forecast sales → Orders	Planning / Finance	Set the baseline demand for the day/week	Approve targets for the week based on expected sales volume
Orders → Labor hours → Headcount	Warehouse / Workforce	Build shift rosters and overtime plans	Schedule 8 associates for 2018-08-07; add 2 more if promo runs
ROP (DLT + Safety)	Inventory / Supply	Trigger replenishment when inventory position drops below ROP	If net available < 2,094 orders → place PO today
“What-if” uplift (e.g., +25%)	Marketing + Ops	Check readiness before promos; adjust staffing/logistics	Uplift → 728 orders → 10 associates, 2 routes