SALES FORCASTING AND OPTIMIZATION

Presented by : Digital Egypt Pioneers
Amit Students

11 May 2025

CONTENT

Project Overview

1 INTRODUCTION

Think of every time you bought something at a store or online — that’s retail sales in action

💡 Key Takeaways:

- Sold directly to you, the end customer
- Every purchase = a recorded transaction
- Driven by promotions, holidays, seasons, location, and customer habits
- Used for everything from forecasting and inventory planning to economic health checks

Project Overview

AMIT

Digital Egypt Pioneers

1.1 MEET THE ROSSMANN STORE SALES DATASET

Imagine having access to 1,000+ stores' sales data across Germany — that's what this dataset offers!

- Daily sales per store
- Store-specific features like type, product variety, competition distance
- External signals like school & public holidays, and store closures
- Time Period: Jan 2013 – July 2015

Perfect for forecasting future trends, modeling demand spikes, and simulating real-world business decisions.

Project Overview

1.2

DATASET FILES OVERVIEW

train.csv

- Our playground
- Includes Store, Date, Sales, Customers, Open, Promo, StateHoliday, SchoolHoliday

test.csv

- The challenge
- Similar to train.csv but **without the Sales column**
- Used for prediction and submission

store.csv

- The secret sauce
- Includes:StoreType, Assortment, CompetitionDistance, Promo2, etc.
- Must be joined with train.csv and test.csv on the Store column

This dataset is more than numbers — it's a real-world business simulation, and your mission is to predict sales like a retail mastermind.
Ready to turn data into decisions?

Project Overview

AMIT

Digital Egypt Pioneers

1.3 TOOLS & LIBRARIES

NumPy

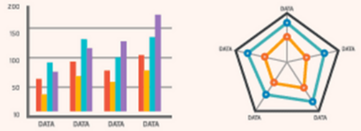
pandas

statsmodels

seaborn

plotly

matplotlib



EXPLORATORY DATA ANALYSIS

🔥 About preprocessing:

Preprocessing Step	Needed for Prophet?	Why?
Handling Missing Dates	✅	Prophet needs continuous time series.
Removing Stores that were closed (`Open == 0`)	✅	If stores were closed, sales are 0 = not helpful for learning real patterns.
Feature Engineering (Month, Promo2, etc.)	❌	Prophet already models trends, seasonality, holidays internally.
Scaling / Normalization	❌	Prophet doesn't need scaling, works with raw numbers.
Filling Missing Sales	✅	If there are NaNs, Prophet can fail — fill them smartly.

EXPLORATORY DATA ANALYSIS

Goal: Understand data structure, distributions, and relationships to inform modeling.

Focus on train.csv and store.csv datasets.

Insights will be grouped into:

- **Data Overview**
- **Feature Engineering**
- **Missing Values**
- **Store Type Insights**
- **Promotions & Sales Patterns**
- **Correlation Analysis**

TRAIN.CSV – OVERVIEW

Train Data Summary

Features:

Store, Date, Sales, Customers, Open, Promo, StateHoliday, SchoolHoliday

- Sales = Target variable
- No missing values
- Only StateHoliday is categorical
- Time series data ⇒ extract Year, Month, Day, WeekOfYear
- ~20% of data has Sales == 0
- Average spend per customer ≈ €9.50/day

Feature Engineering

FEATURE ENGINEERING FROM DATE & SALES

Extracted:

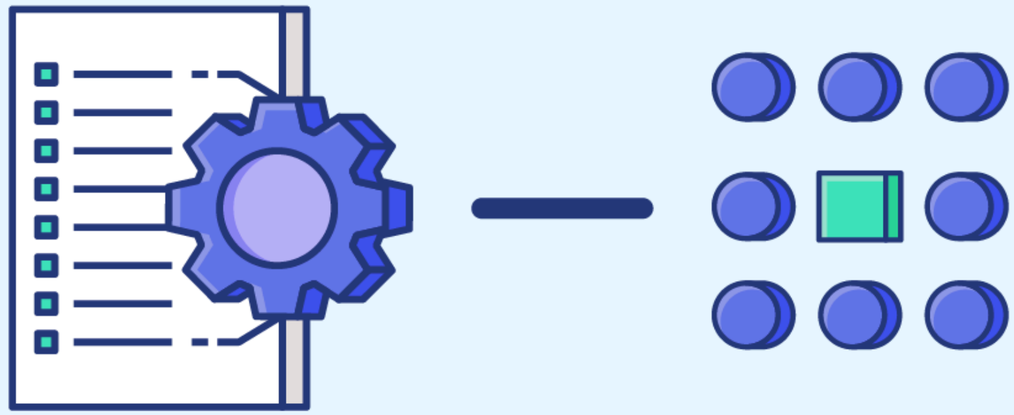
Year, Month, Day, WeekofYear

Created:

- SalesPerCustomer = Sales / Customers

ECDF used to examine distribution:

- ~80% of days had sales < €1000
- Outliers: Sales = 0 (needs further inspection)



Feature Engineering

HANDLING ZEROS & CLOSED STORES

Zero Sales & Closed Store Analysis

- ~172,817 records (≈10%) have Open = 0 → Drop for modeling
- Found open stores with Sales == 0 (only 54 cases)
- Possible external factors (strikes, delivery issues)

STORE.CSV – OVERVIEW

Store Data Summary

Key Features

StoreType, Assortment, CompetitionDistance, Promo2

Missing Values:

- CompetitionDistance: 3 → Impute median

- CompetitionOpenSinceMonth/Year, Promo2SinceWeek/Year,PromoInterval: 354–544 → Fill with 0 if Promo2 = 0

Merging Data

MERGING TRAIN & STORE DATA

- Inner join on Store
- Combined dataset: train_store
- Enables grouped analysis by StoreType, Promo, Assortment, etc.

PROMOTIONS & WEEKLY SALES

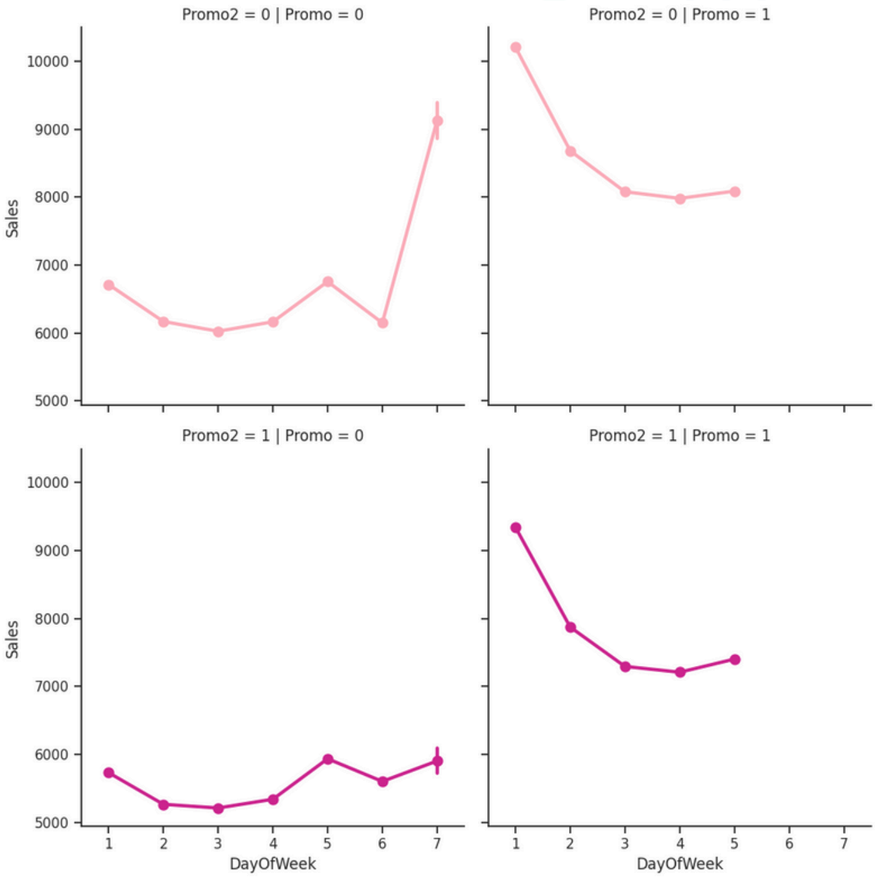
PROMOTION & DAY-OF-WEEK SALES BEHAVIOR

Without Promo: Peak on Sundays

With Promo: Peak on Mondays

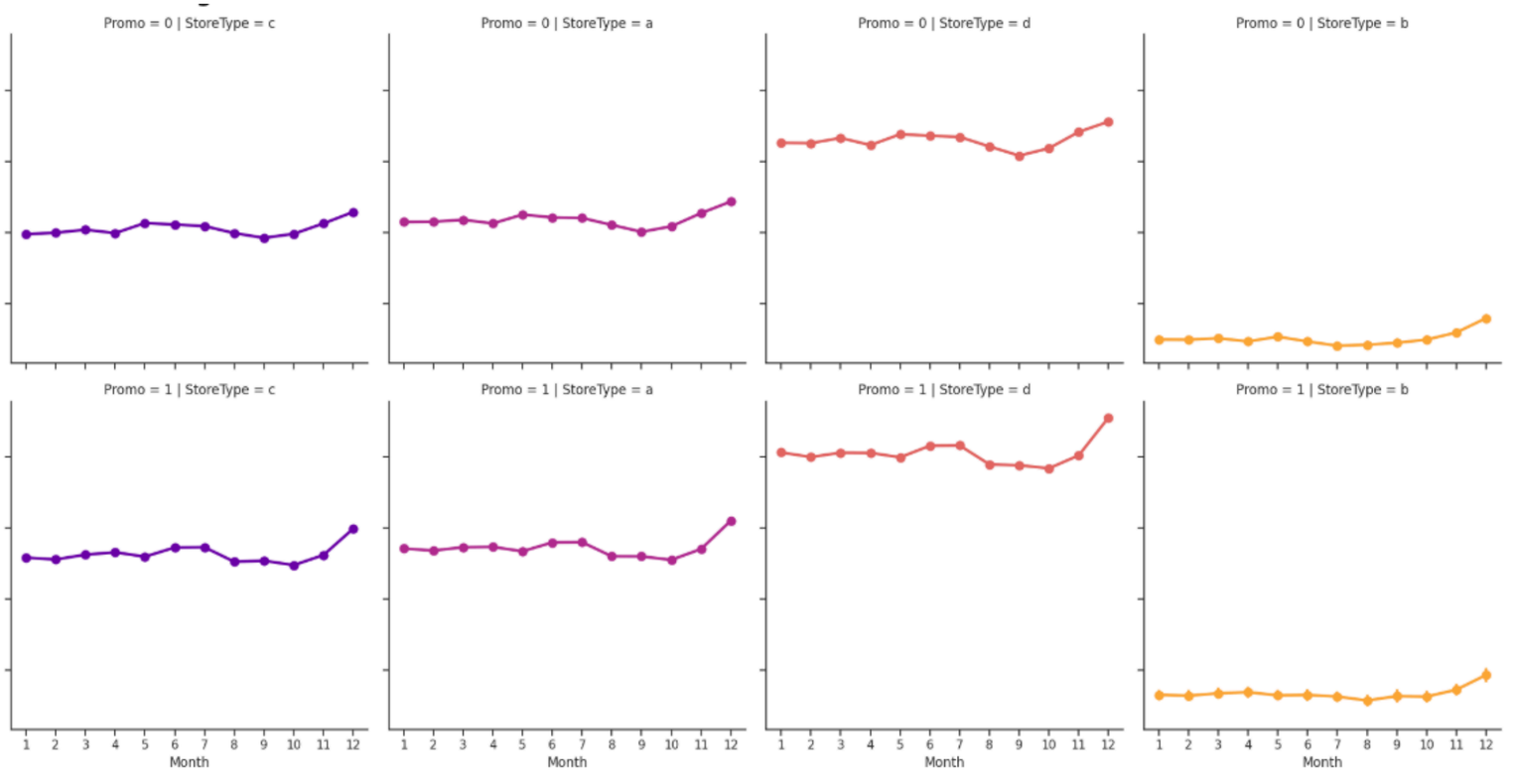
StoreType C: Always closed Sundays

StoreType D: Closed Sundays only Oct–Dec



StoreType Analysis

STORETYPE & SALES DISTRIBUTION



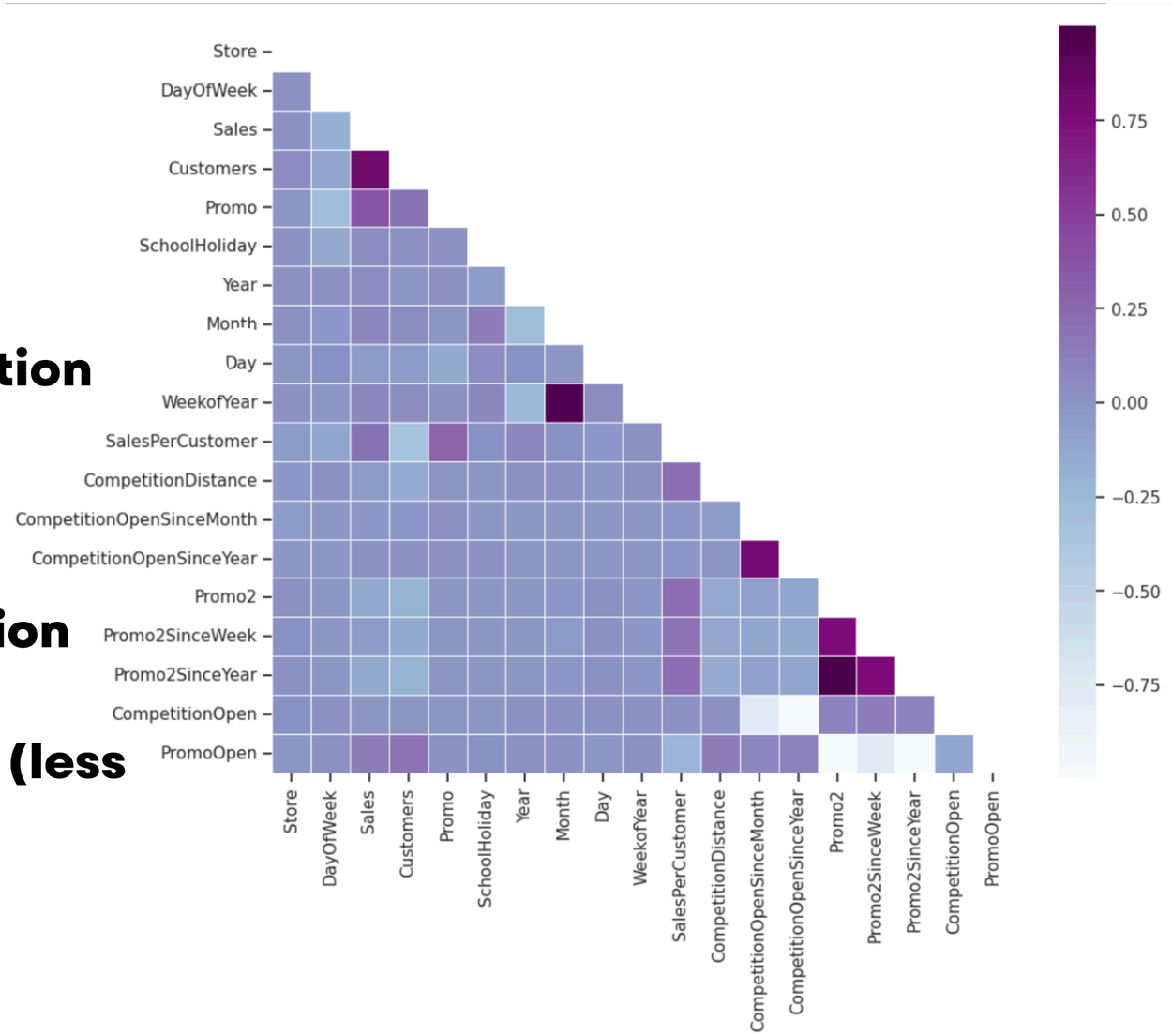
- StoreType B: Highest average sales but low count
- StoreType A: Highest total sales & customer volume
- StoreType D: Highest SalesPerCustomer (≈ €12 with Promo)

PROMO2 IMPACT ON SALES

- **Weak or negative correlation with Sales**
- **Best results: Promo = 1 & Promo2 = 0**
- **Stores with both active promotions didn’t see substantial boost**

CORRELATION HEATMAP

- **Sales ↔ Customers: Strong positive correlation**
- **Promo ↔ Customers: Positive correlation**
- **Promo2 ↔ Sales: Weak or negative correlation**
- **Promo ↔ DayOfWeek: Negative correlation (less effective later in week)**



EDA SUMMARY

- StoreType A: Most customers & total sales
- StoreType D: Best sales per customer
- Promo campaigns drive Monday sales
- Promo2 has limited effectiveness
- Customer behavior varies with promotion and store type
- Seasonal effect: Sales spike near Christmas





TIME SERIES

What Makes Time Series Special?

- Observations depend on time, unlike typical regression.
- Violates independence assumption.
- Often includes trend and seasonality:
  e.g., Sales increase during Christmas holidays.

Goal: Understand patterns to make better forecasts.

Why Analyze by Store Type?

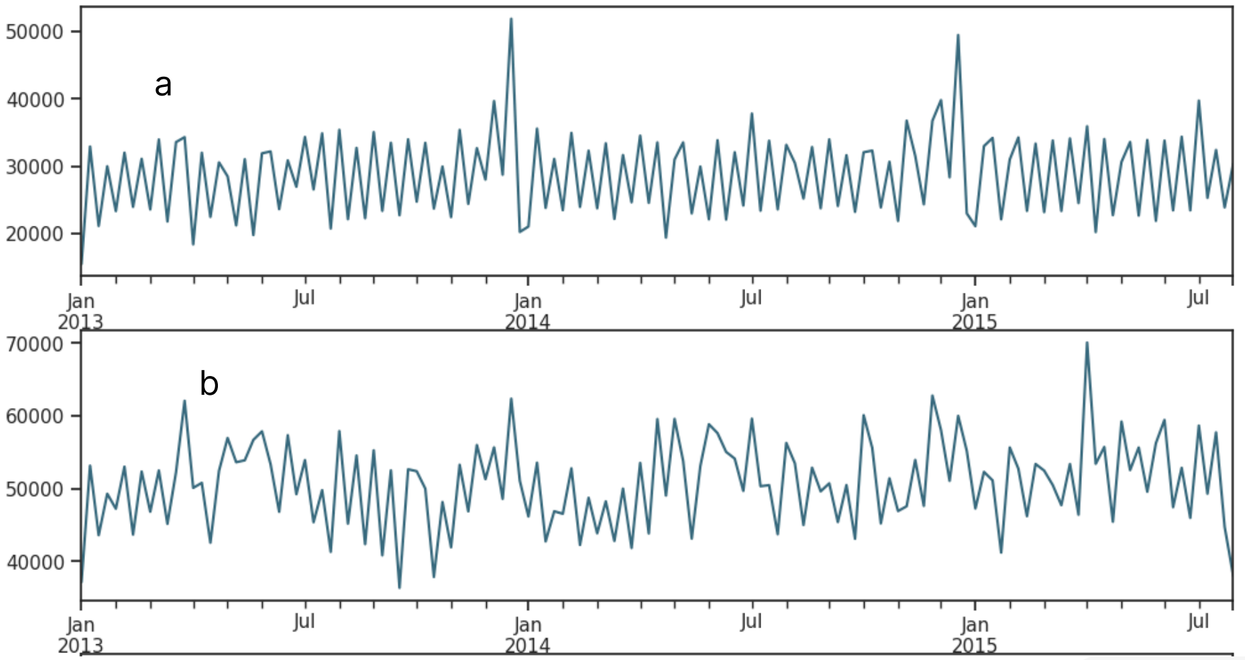
STORE TYPE-LEVEL TIME SERIES ANALYSIS

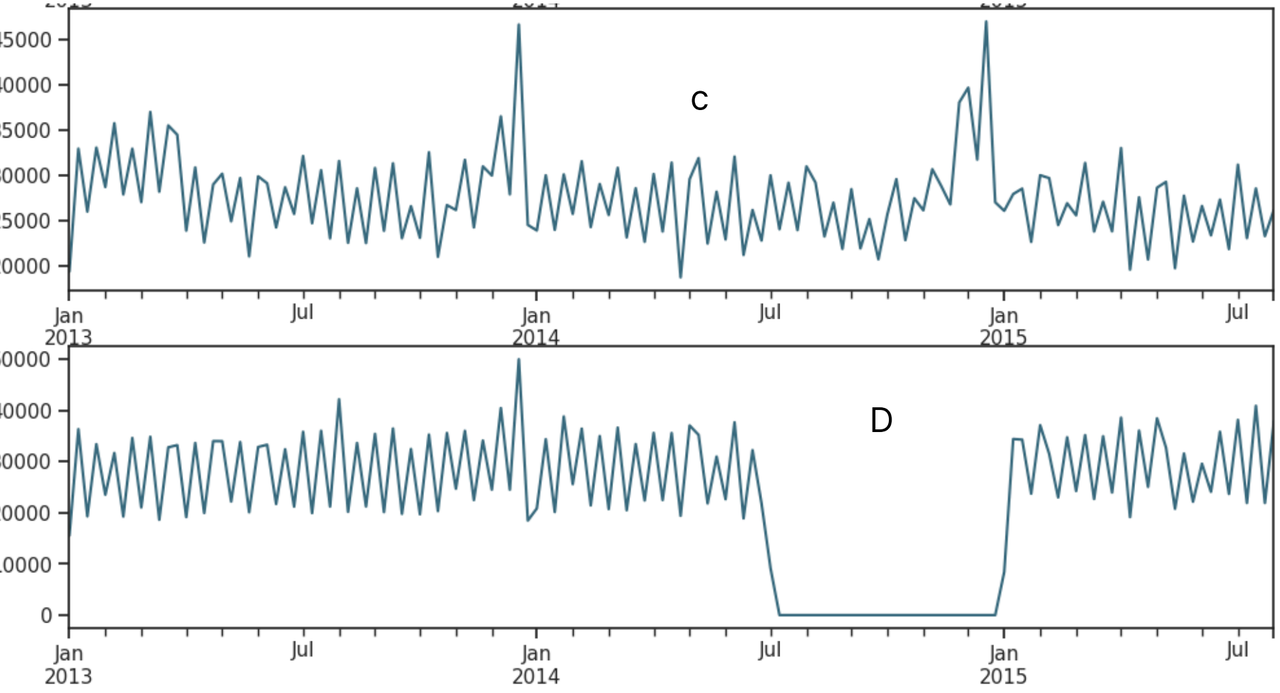
- Easier than analyzing each individual store.
- Reveals general trends & seasonalities.
- Stores analyzed:
- Store #2 → Type A
- Store #85 → Type B
- Store #1 → Type C
- Store #13 → Type D
- Data is resampled weekly for clarity.

Weekly Sales Trends

WEEKLY RETAIL SALES TRENDS

- Type A & C: Clear Christmas sales spikes.
- Type D: Data missing from Jul 2014–Jan 2015 (closed stores).
- Post-holiday drop observed across types.

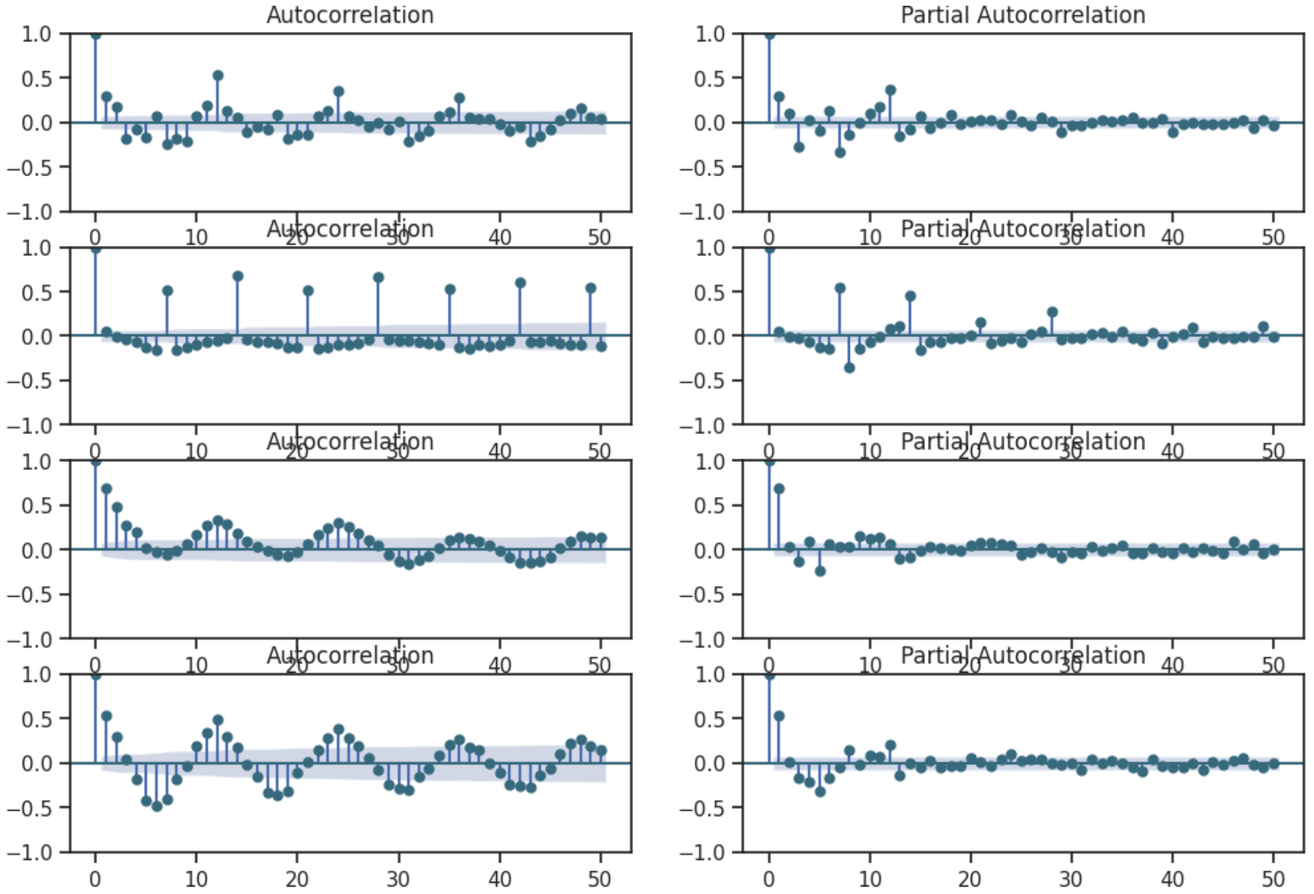




Seasonal Patterns

SEASONALITY & LAG ANALYSIS

- Horizontal plots (ACF & PACF) by store type.
- Common patterns:
- Lag-1 autocorrelation is high → Suggests differencing.
- Type A: Seasonal spikes every 12/24 lags (monthly).
- Type B: Weekly spikes at lag 7, 14, 21, 28.
- Type C & D: Complex correlations with adjacent lags.



Forecasting with Prophet

FORECASTING SALES WITH PROPHET

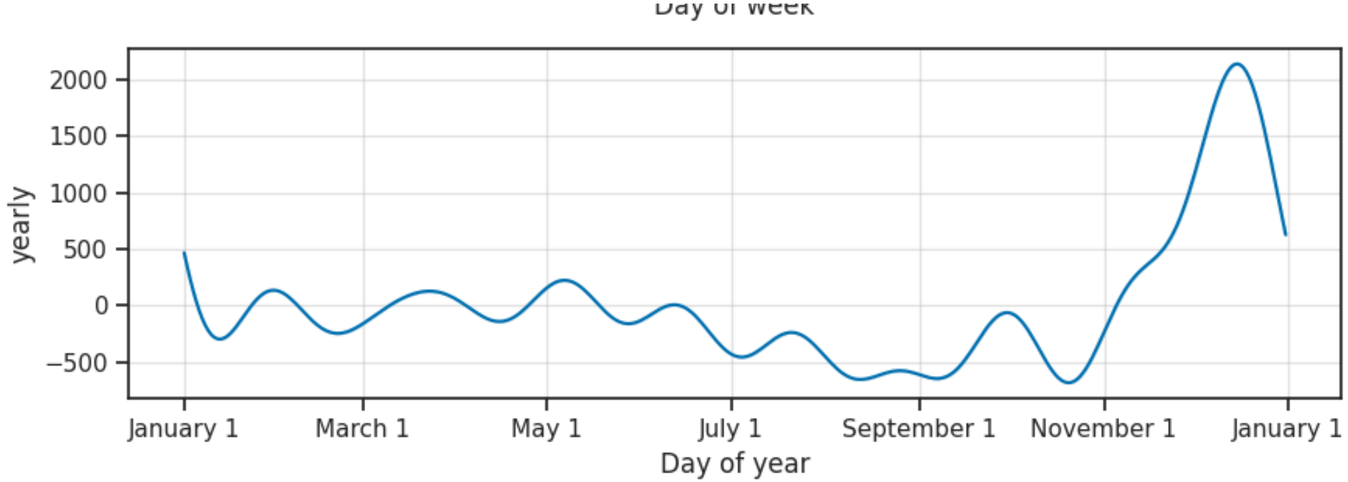
- Prophet allows:

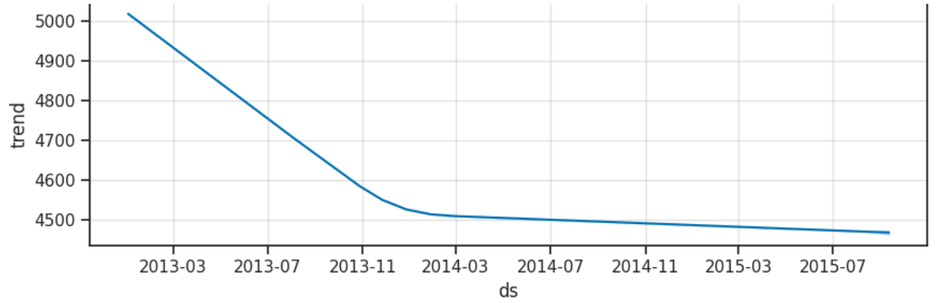
- Trend & seasonality modeling
- Holiday effects (state & school holidays)
- Forecast horizon: 6 weeks (42 days) into the future
- Prophet forecasts include:
- Trend, weekly effect, holiday impact

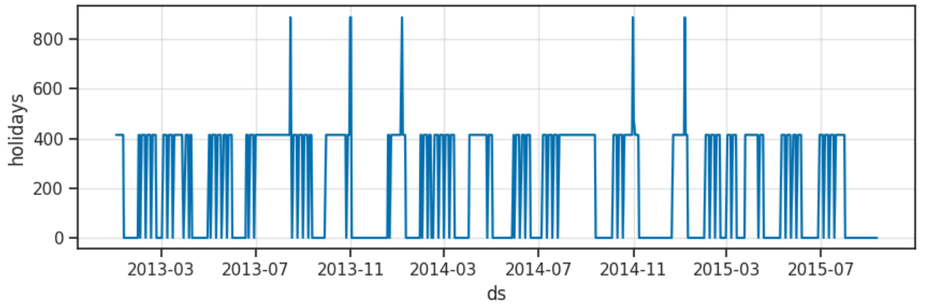
Prophet Output Analysis

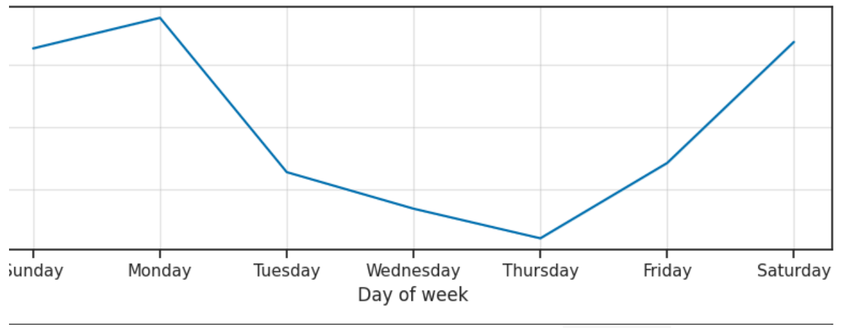
PROPHET FORECAST OUTPUT HIGHLIGHTS

- Store #1 shows linear decrease in sales
- Weekly peak: Monday sales spike
- Holiday influence: Christmas most active period
- Visualizations help validate model fit & patterns









Accuracy & Evaluation

FORECAST ACCURACY EVALUATION

- Used metrics:
- MAE, MSE, RMSE, MAPE
- Forecasts stored in 'yhat' column
- Comparison: 'sales' vs 'forecastt'
- Best ARIMA model MSE ≈ 9.4M (very high)

```
MAE: 674.1702896702294
RMSE: 832.6118394962039
MAPE: 14.747619716275343 %
Accuracy: 85.25238028372466 %
```

```
Best ARIMA(2, 0, 2) MSE=9414839.925
```

Why ARIMA Alone Fails

ARIMA MODEL LIMITATIONS

- Rossmann Sales = Complex time series:
- Trend + Seasonality
- Promotions, Holidays, Competition
- ARIMA without external features underperforms
- Better suited: SARIMA, Prophet, XGBoost, LSTM

SARIMA Forecasting (Optional)

TIME SERIES WITH SARIMA (BONUS)

- Captures seasonal trends
- 7 hyperparameters → Manual tuning needed
- Requires 4–5 full seasons of data
- Limited ability to handle external variables

Time Series Forecasting – Conclusion

KEY TAKEAWAYS

Advantages:

- Accounts for time trends, seasonalities, holidays.
- Prophet is intuitive and handles holidays well.

- Drawbacks:
- Limited handling of external interactions (Promo, Competition).
- SARIMA requires extensive data & manual tuning.
- ARIMA underperforms in real-world complexity.

OUR TEAM

Our team is a diverse blend of creative minds, strategic thinkers, and industry experts committed to propelling your business to new heights. With a passion for innovation and a collective dedication to excellence, we bring a wealth of experience and fresh perspectives to the project.

Nour Sameh

Amr Gaber

Ahmed Gamal

Ahmed khaled

Moustafa

Zeyad Tarek

THANK YOU

FOR YOUR ATTENTION

May 2025