Dataset

Dataset folder path

D:\Thesis prototype\Dataset

Configuration

Train ratio (time-based)

0.80

Train: 80% | Test: 19%

Season period

7

ACF/PACF lags

30

Model parameters

ARIMA/ARIMAX order (p,d,q)

(1, 1, 1)

SARIMA/SARIMAX order (P,D,Q,s)

(1, 1, 1)

Seasonal order (P,D,Q,s)

(1, 1, 1, 7)

Exogenous variables

- [x] Use oil price (dcoilwtico)
- [x] Use holiday_any
- [x] Use holiday_event_count
- [x] Use is_weekend

Prophet

- [x] Enable Prophet model

Prophet interval width

0.8

Outputs

Save folder name

thesis_outputs

- [x] Enable saving PNG/CSV

IT Prototype: Big Data Time-Series Forecasting Analysis and Visualization

Prototype interface for thesis demonstration (Seasonal Naive, ARIMA/SARIMA, optional Prophet).

1. Dataset Overview

Days (daily series)	Start date	End date	Mean transactions
1,682	2013-01-01	2017-08-15	84113.52

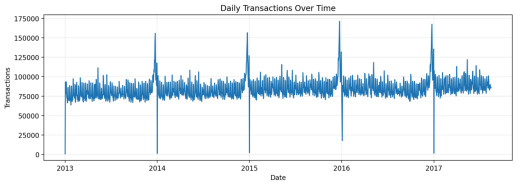
1.1 Data Preview

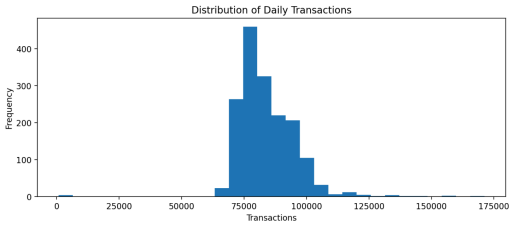
	date	transactions	dcoilwtico	holiday_any	holiday_event_count	day_of_week	is_weekend	log_tra
0	2013-01-01 00:00:00	770	93.14	1	1	1	0	
1	2013-01-02 00:00:00	93215	93.14	0	0	2	0	
2	2013-01-03 00:00:00	78504	92.97	0	0	3	0	
3	2013-01-04 00:00:00	78494	93.12	0	0	4	0	
4	2013-01-05 00:00:00	93573	93.12	1	46	5	1	
5	2013-01-06 00:00:00	90464	93.12	0	0	6	1	
6	2013-01-07 00:00:00	75597	93.2	0	0	0	0	
7	2013-01-08 00:00:00	72325	93.21	0	0	1	0	
8	2013-01-09 00:00:00	71971	93.08	0	0	2	0	
9	2013-01-10 00:00:00	66383	93.81	0	0	3	0	

Exogenous columns used:


```
[
  0 : "dcoilwtico"
  1 : "holiday_any"
  2 : "holiday_event_count"
  3 : "is_weekend"
]
```

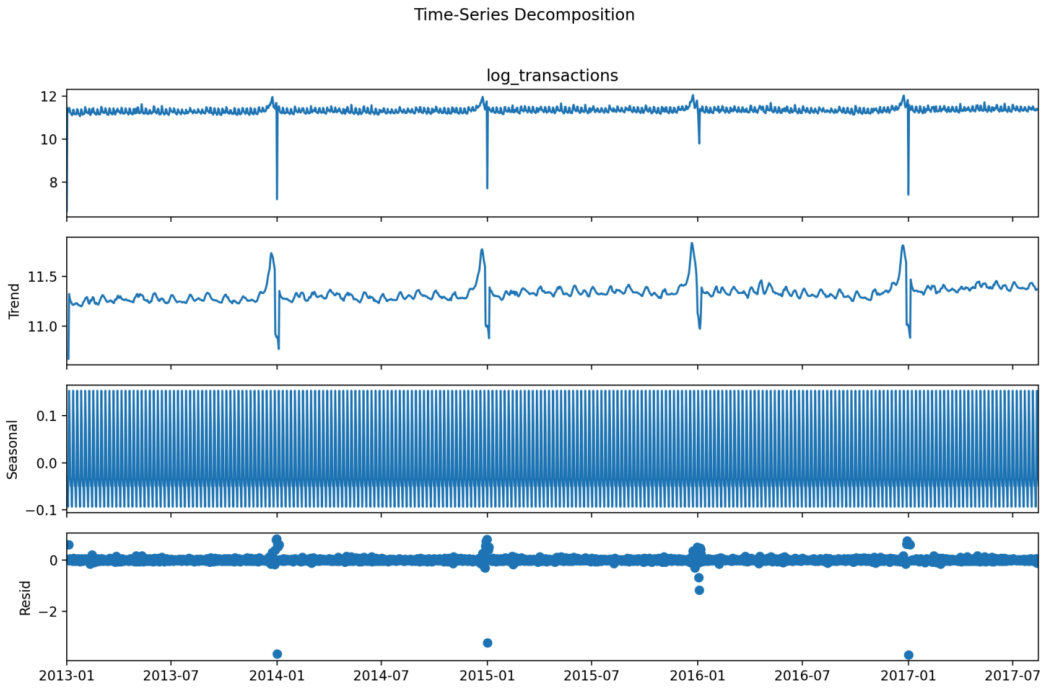

2. Exploratory Data Analysis



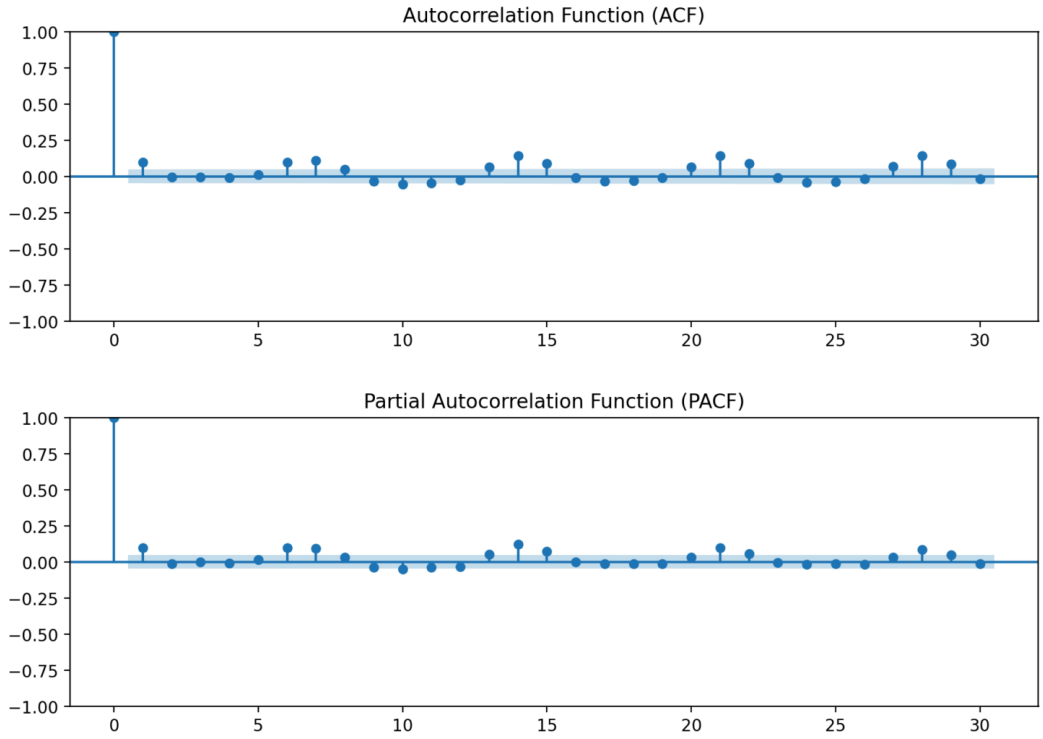


3. Seasonality and Correlation Analysis

3.1 Time-Series Decomposition



3.2 ACF and PACF Diagnostics



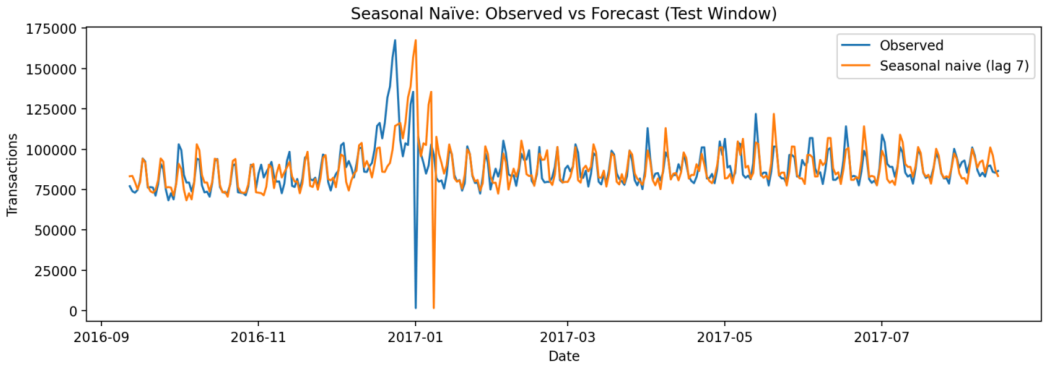
4. Train/Test Split (Time-Based)

Split index: **1345** | Train size: **1345** | Test size: **337**

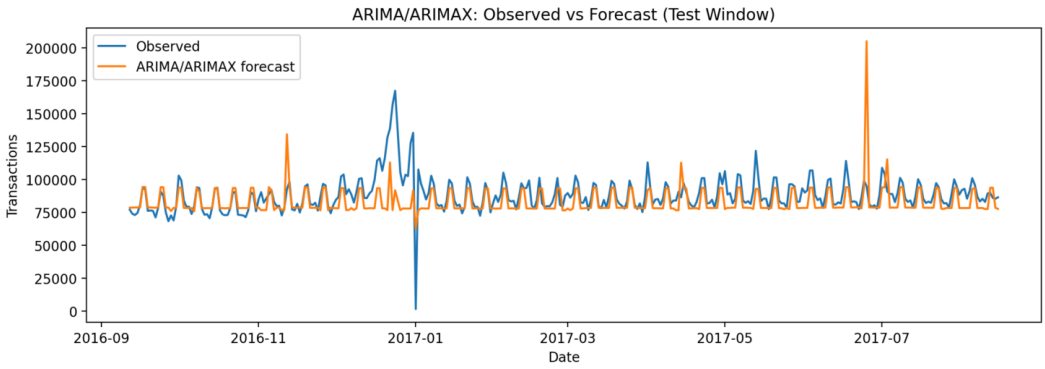
	date	transactions	dcoilwtico	holiday_any	holiday_event_count	day_of_week	is_weekend	log_tra
1345	2016-09-12 00:00:00	77125	46.28	0	0	0	0	
1346	2016-09-13 00:00:00	74032	44.91	0	0	1	0	
1347	2016-09-14 00:00:00	73096	43.62	0	0	2	0	
1348	2016-09-15 00:00:00	74741	43.85	0	0	3	0	
1349	2016-09-16 00:00:00	80789	43.04	0	0	4	0	
1350	2016-09-17 00:00:00	94298	43.04	0	0	5	1	
1351	2016-09-18 00:00:00	92375	43.04	0	0	6	1	
1352	2016-09-19 00:00:00	76255	43.34	0	0	0	0	
1353	2016-09-20 00:00:00	76572	43.85	0	0	1	0	
1354	2016-09-21 00:00:00	76260	45.33	0	0	2	0	

5. Forecasting Models

5.1 Seasonal Naïve Baseline

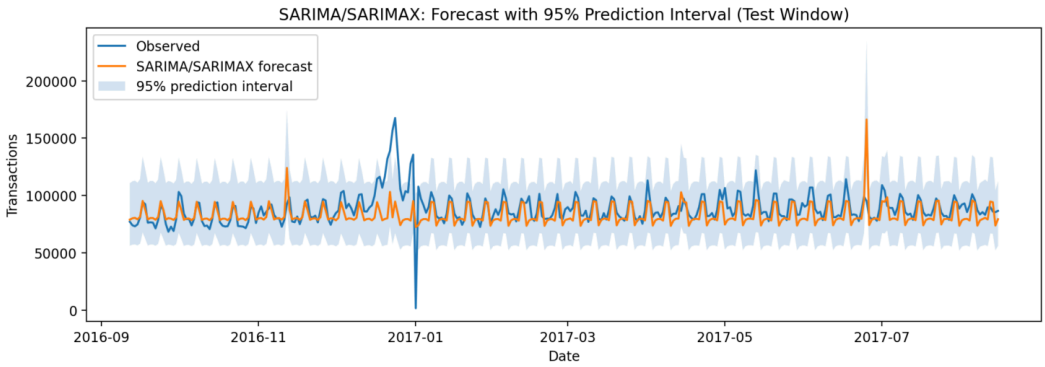


5.2 ARIMA / ARIMAX



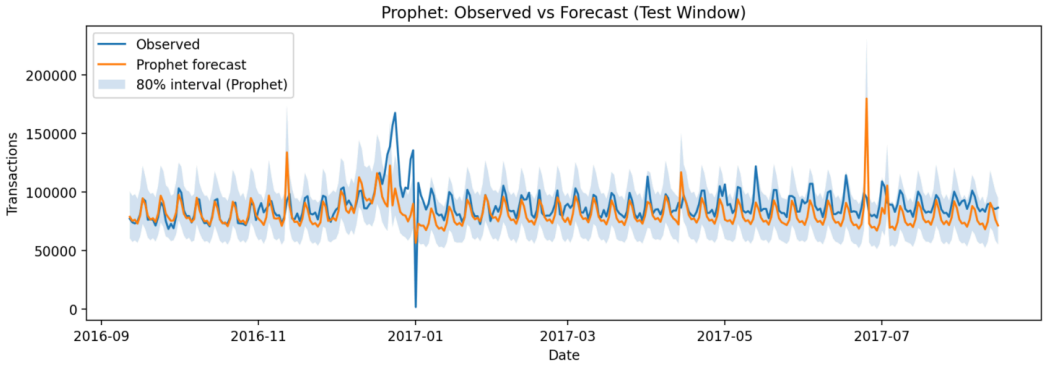
> ARIMA/ARIMAX Model Summary

5.3 SARIMA / SARIMAX (with prediction interval)



SARIMA/SARIMAX Model Summary

5.4 Prophet (with optional regressors)

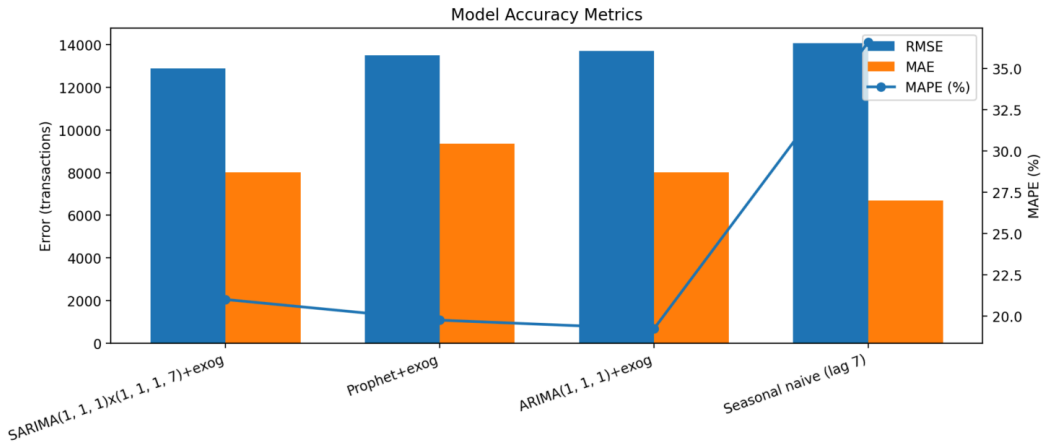


Prophet components (trend / weekly / yearly)

6. Model Evaluation

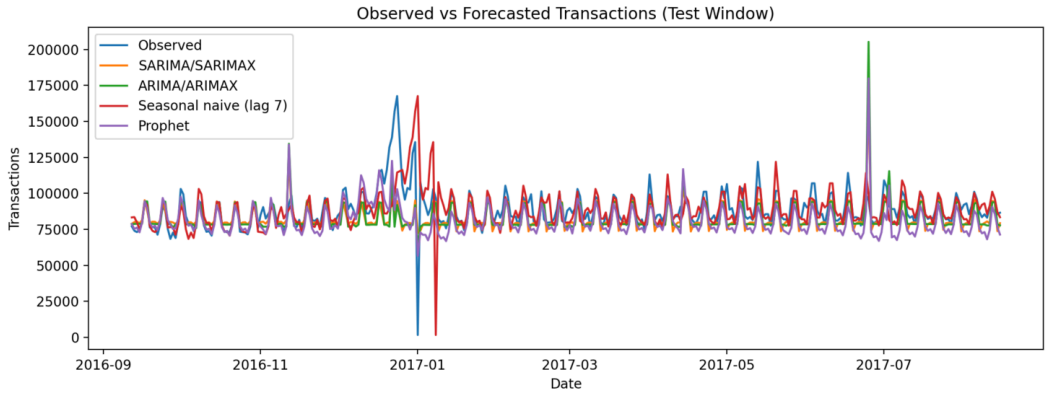
	Model	RMSE	MAE	MAPE (%)	AIC
2	SARIMA(1, 1, 1)x(1, 1, 1, 7)+exog	12895.2255	8006.9519	21.0042	-849
3	Prophet+exog	13501.2979	9363.9525	19.7528	
1	ARIMA(1, 1, 1)+exog	13708.3352	8003.226	19.2254	-380
0	Seasonal naive (lag 7)	14079.2291	6695.7537	36.5916	

6.1 Metric Comparison Visualization

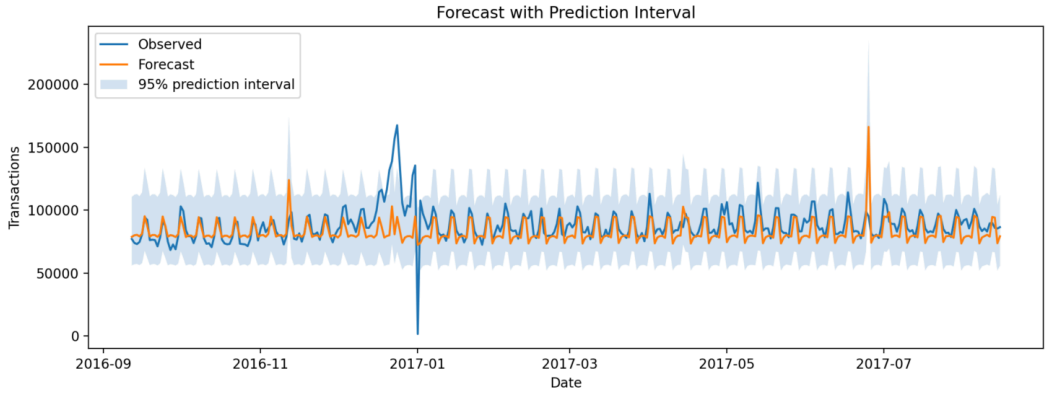


7. Forecast Visualization

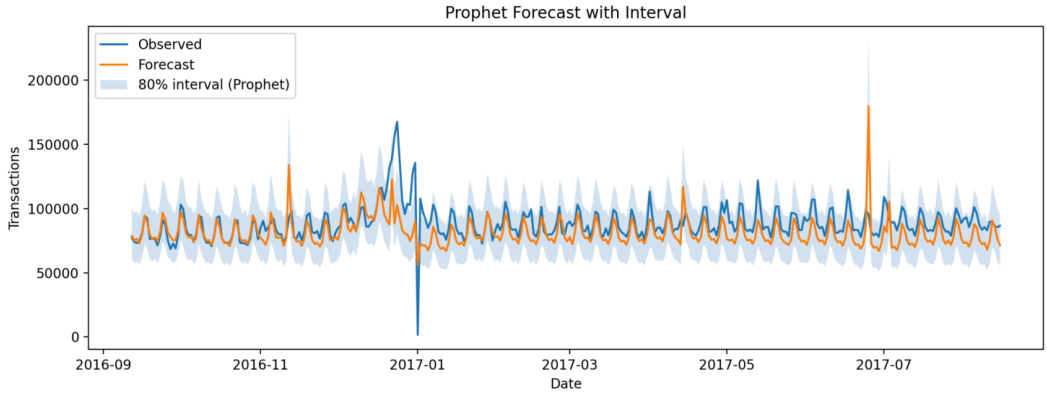
7.1 Model Comparison (Test Window)



7.2 Forecast with 95% Prediction Interval (SARIMA/SARIMAX)

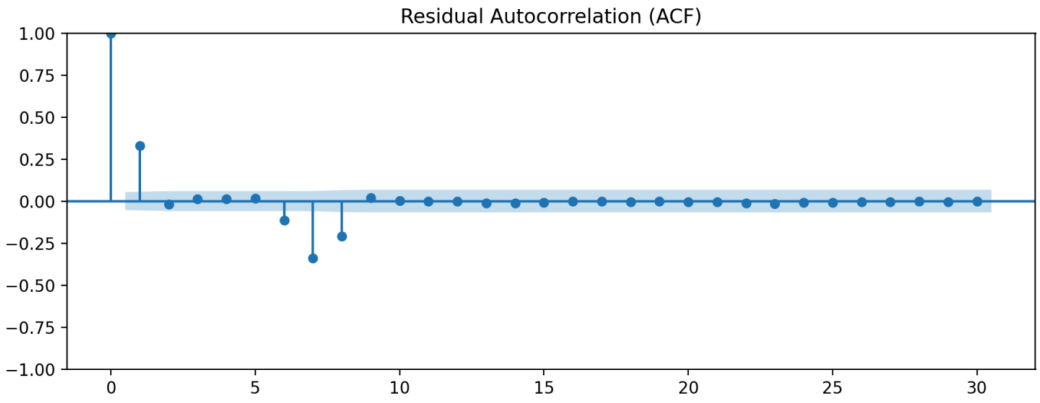


7.3 Forecast with Interval (Prophet)



8. Residual Diagnostics

8.1 Residual ACF (SARIMA/SARIMAX)



9. Export (Optional)

Save key figures (PNG)

Save evaluation table (CSV)

Output folder: D:\Thesis prototype\Dataset\thesis_outputs