



Formula 1

Time Series Forecasting





Introductions



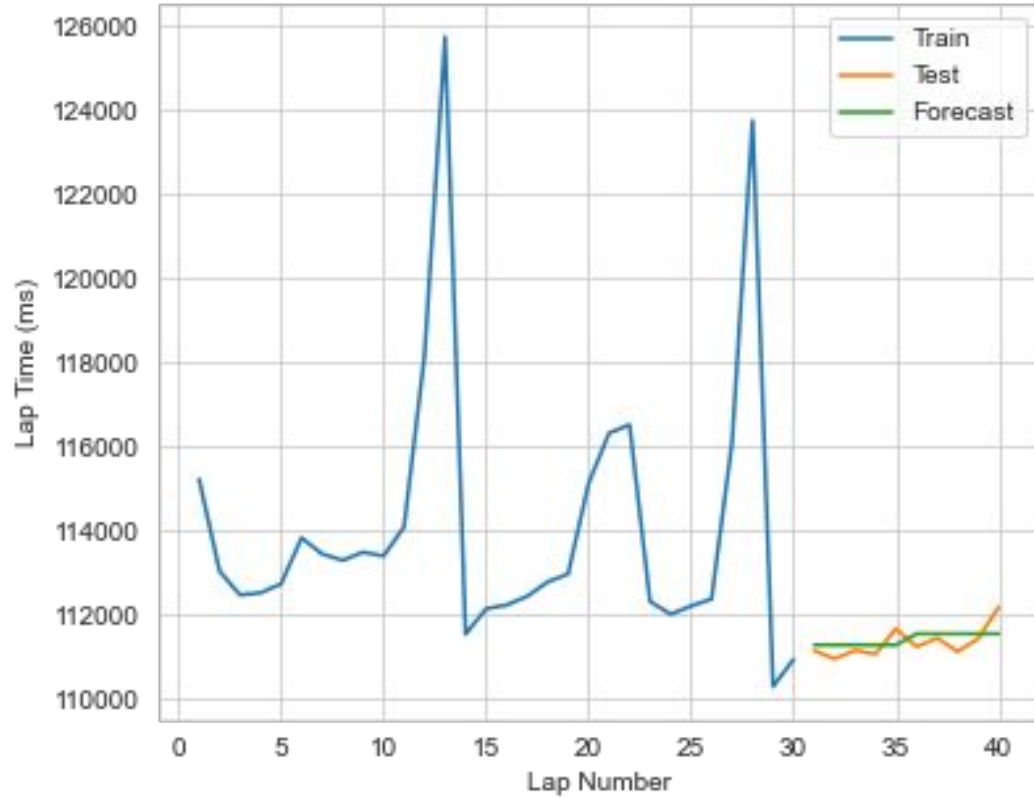
Agenda

- Objective
- Best Outcome
- How I Got There
 - Process Overview
 - Understanding F1
 - Data Understanding & Preparation
 - Time Series Analysis
 - Key Models & Walk-Forward Forecasts
- Future Improvements

Objective

Forecast lap times for a horizon of 5 laps

Best Model



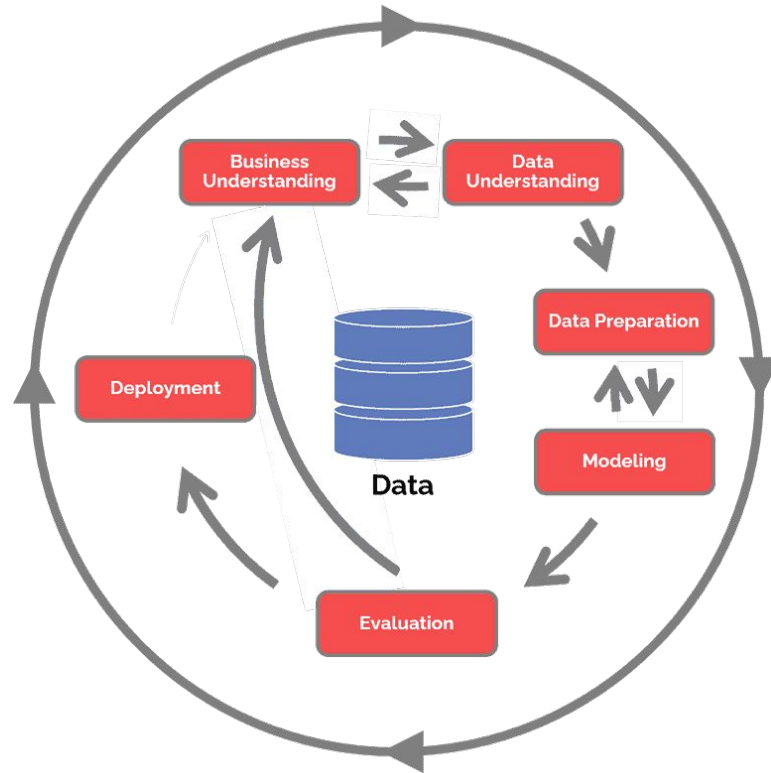


How I Got Here



The Process

CRISP-DM





Understanding Formula 1




MATT DAMON CHRISTIAN BALE
FORD v **FERRARI**




Formula One

- 10 teams x 2 drivers x 40-80 laps
- 3 Practices + 3 Qualifying Rounds or Sprints + 1 Grand Prix
- Qualifiers determine grid positions
- Pit stops - 2.5 seconds x 1-3 per race
- Safety car periods & penalties
- Every 100 milliseconds count
- Best strategy updated real-time wins



Understanding & Preparing Data



Datasets Review

Table	Table Type	Data Type
Lap Times	Fact	Time Series
Pit Stops	Fact	Cross-Sectional
Races	Dimensions	Cross-Sectional
Results	Fact	Cross-Sectional
Sprint Results	Fact	Cross-Sectional
Qualifying	Fact	Cross-Sectional
Status	Dimensions	Cross-Sectional

Datasets Review

Table	Table Type	Data Type
Lap Times	Fact	Time Series
Pit Stops	Fact	Cross-Sectional
Races	Dimensions	Cross-Sectional
Results	Fact	Cross-Sectional
Sprint Results	Fact	Cross-Sectional
Qualifying	Fact	Cross-Sectional
Status	Dimensions	Cross-Sectional

Data Preparation

Action	Examples
Drop Columns	Time, Duration
Type Conversion	Date
Derivatives	Pitted, Year
Join Tables	Lap Times + Pit Stop on RaceID, DriverID, Lap
Null Values	Pit stops
Driver-Race Selection	Max Verstappen @ 2023 Belgian Grand Prix
Train-Test Split	Varying window widths
Error Measurement	Dataframe -> CSV

Final DataFrame

	milliseconds	stop_count	pit_stop_milliseconds	pitted
lap				
40	112166	0.0000	0.0000	0
41	111796	0.0000	0.0000	0
42	115220	3.0000	23156.0000	1
43	129564	0.0000	0.0000	0
44	107305	0.0000	0.0000	0

Time Series Analysis

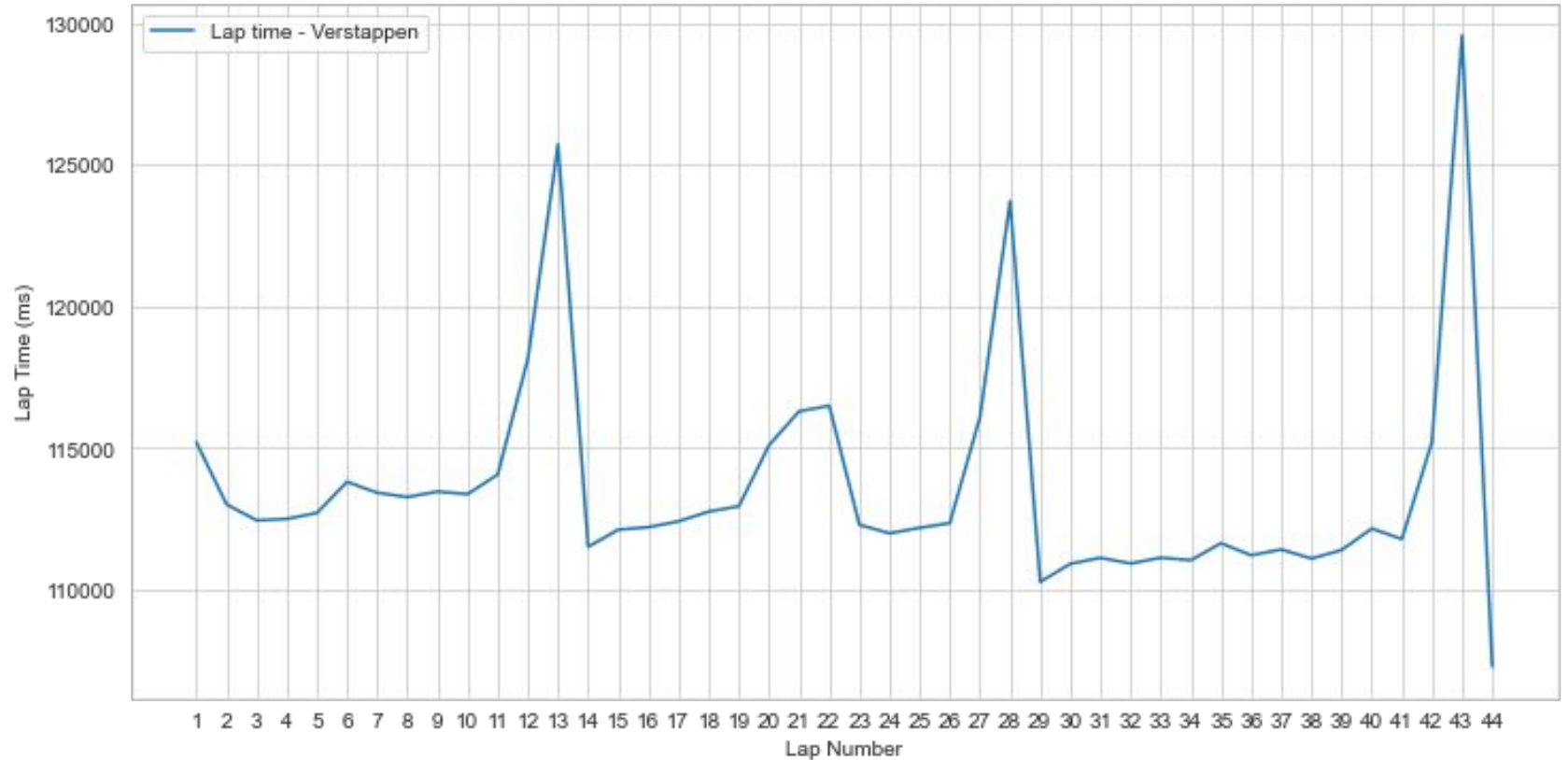
Time Series Characteristics

1. Level
2. Trend
3. Seasonality
4. Exogenous Effects
5. Stationarity
6. Autocorrelation

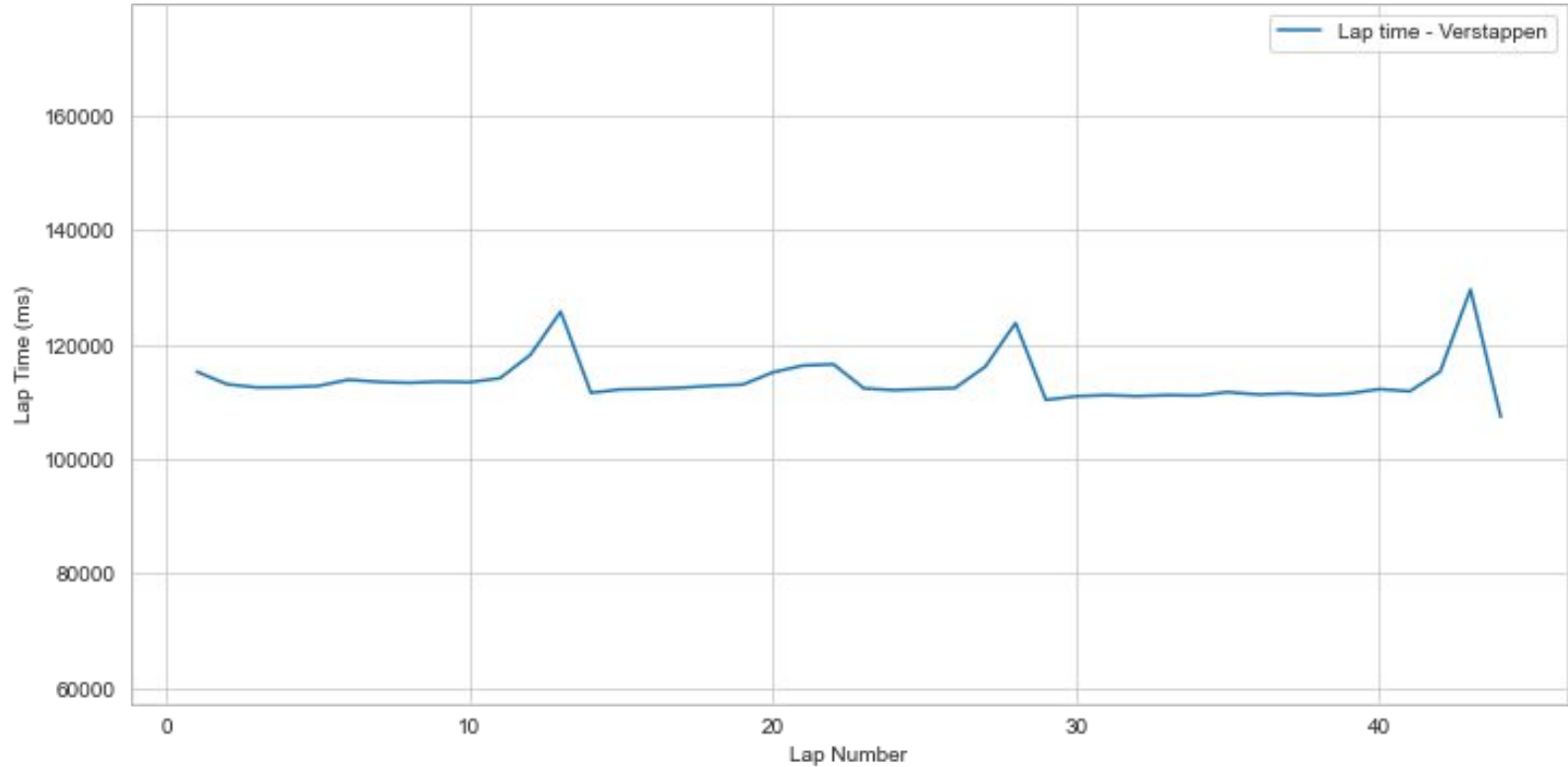
Analysis Methods

1. Visual Analysis
2. Seasonal Decomposition
3. Statistical Testing
4. Auto Correlation Analysis
5. Box Cox Transformation & Differencing

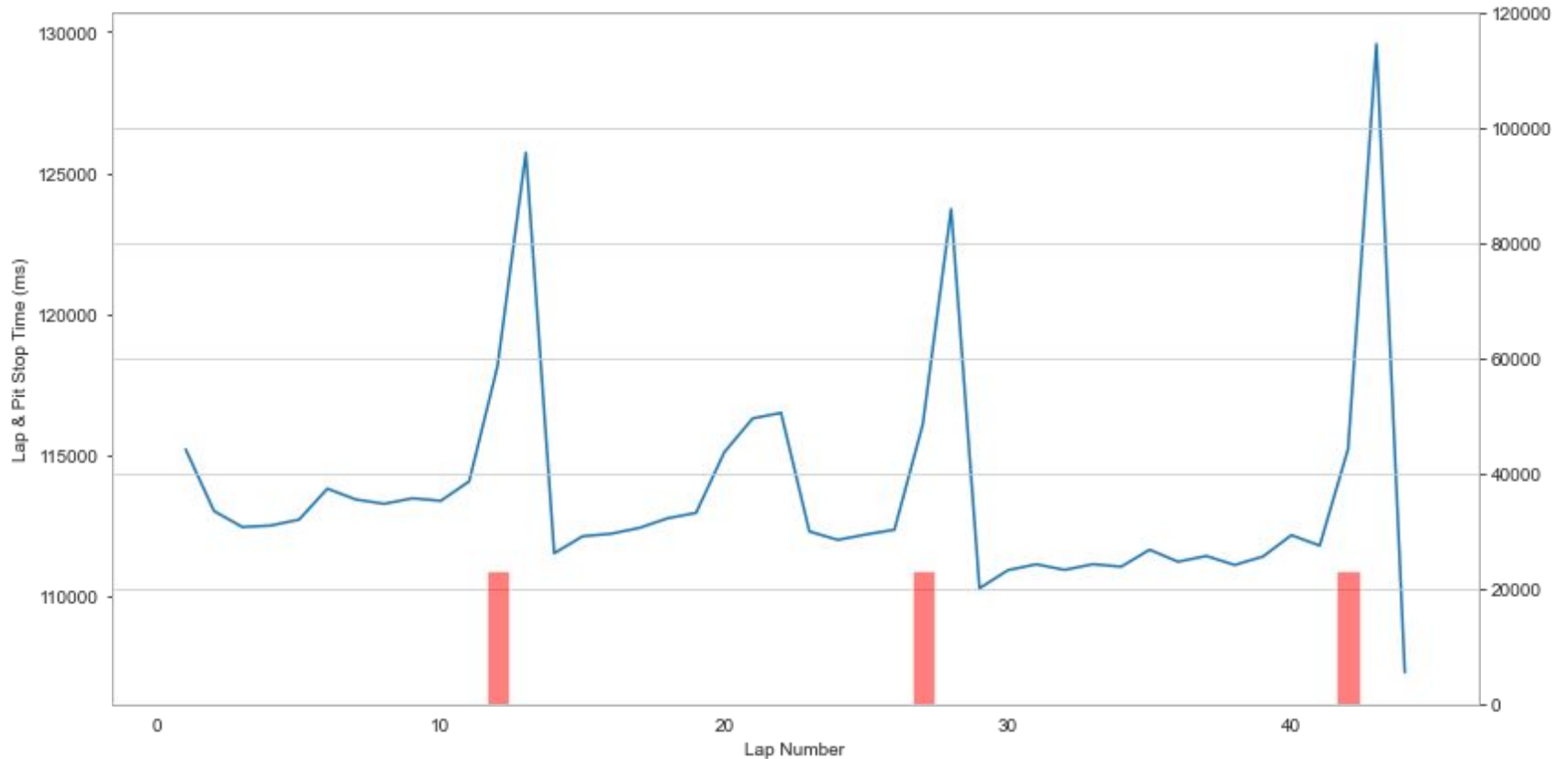
Verstappen @ Belgian 2023 Grand Prix



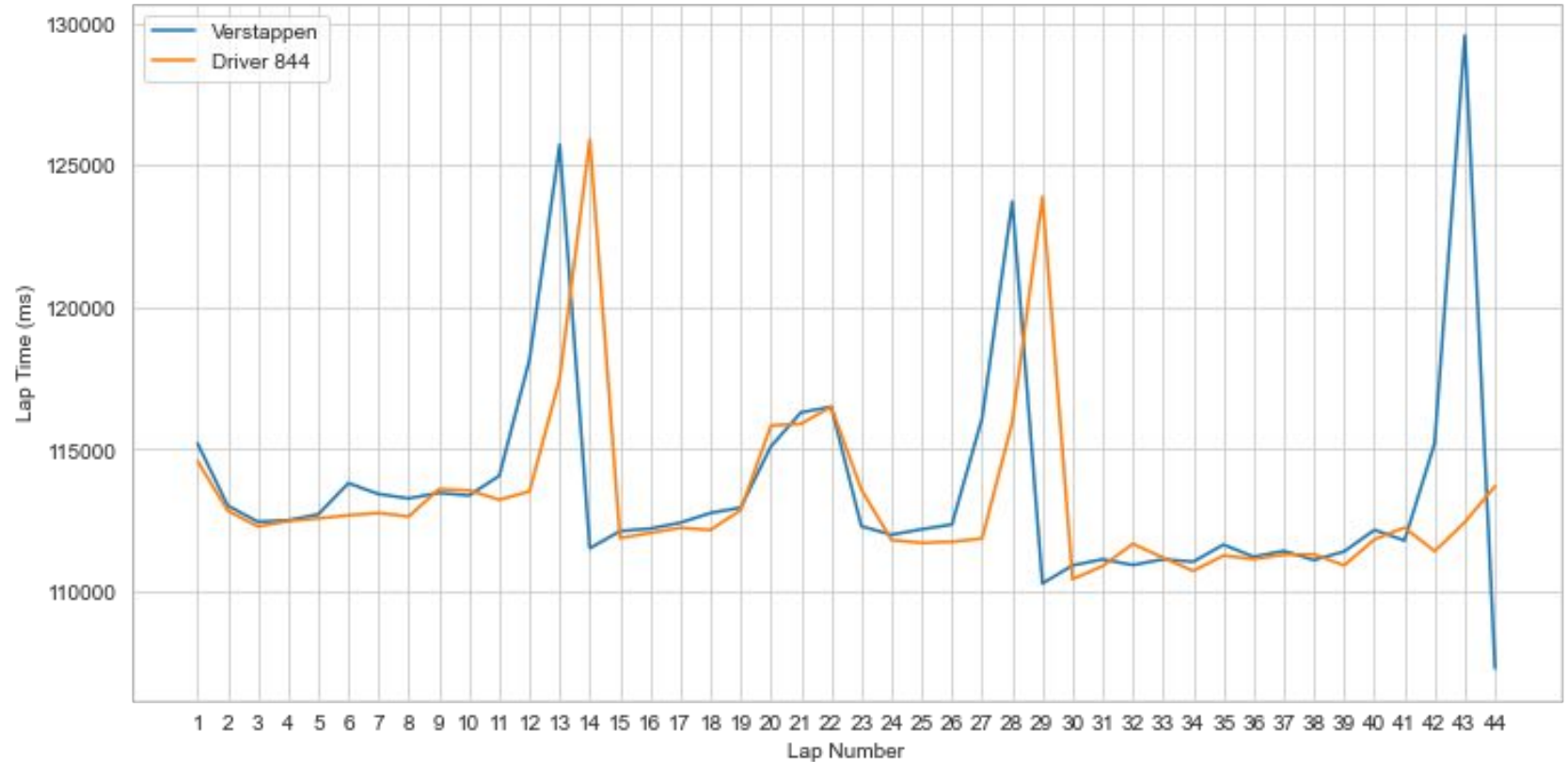
Verstappen @ Belgian 2023 Grand Prix



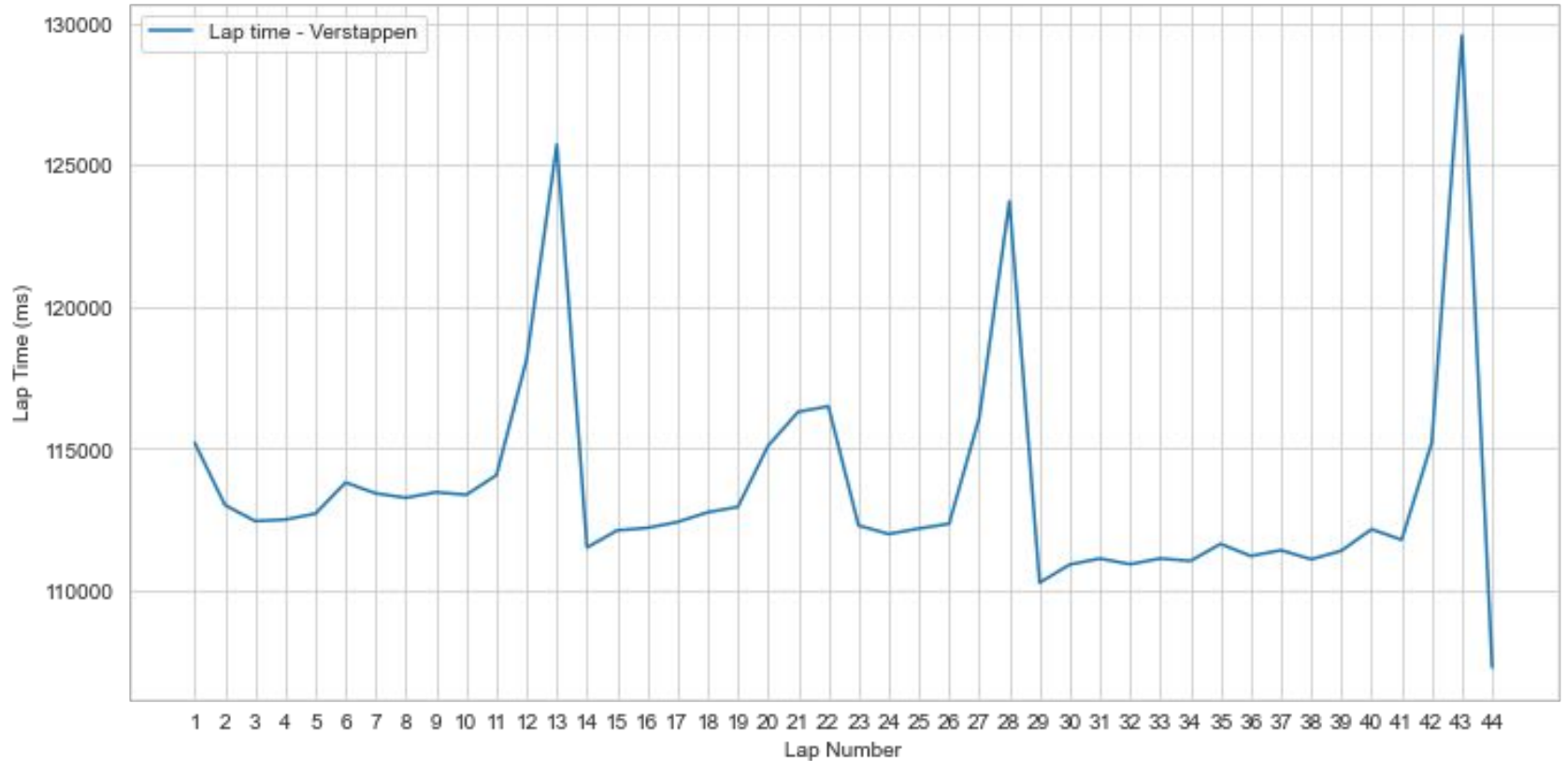
Verstappen @ Belgian 2023 Grand Prix



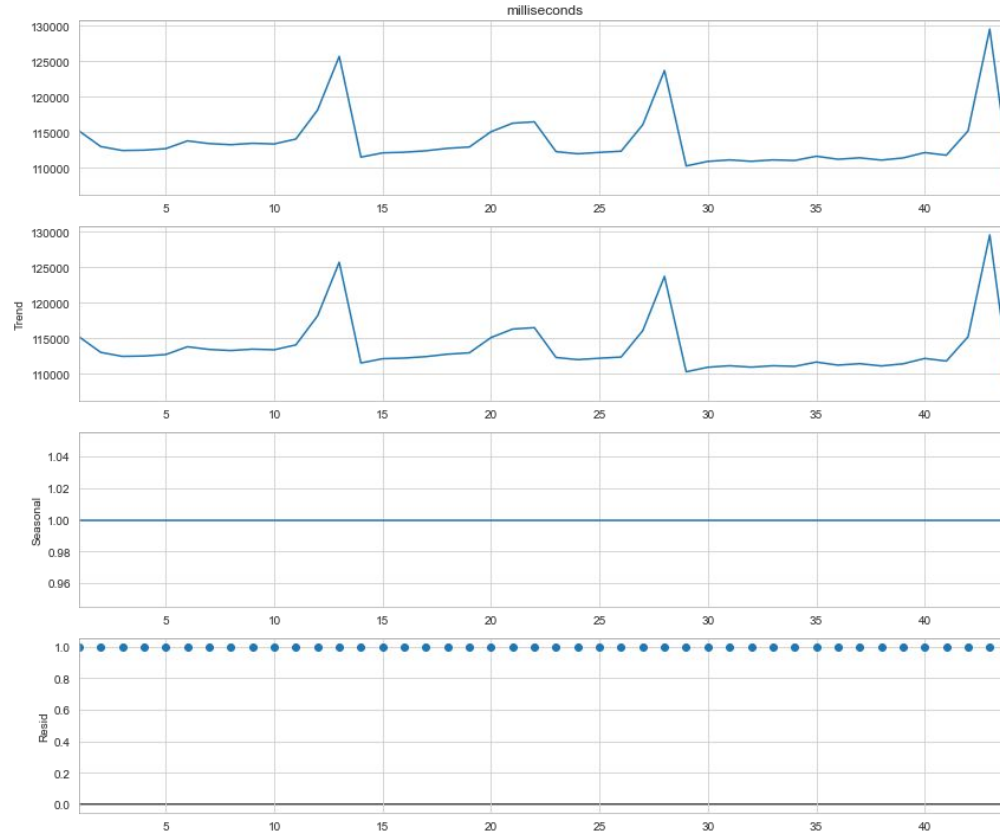
Verstappen vs Driver 844



Verstappen @ Belgian 2023 Grand Prix



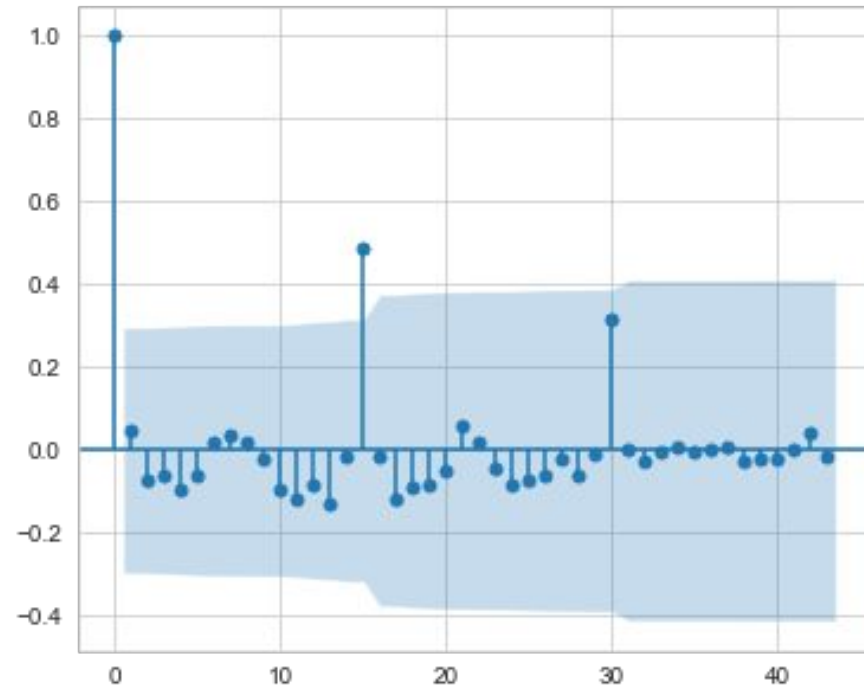
Seasonal Decomposition



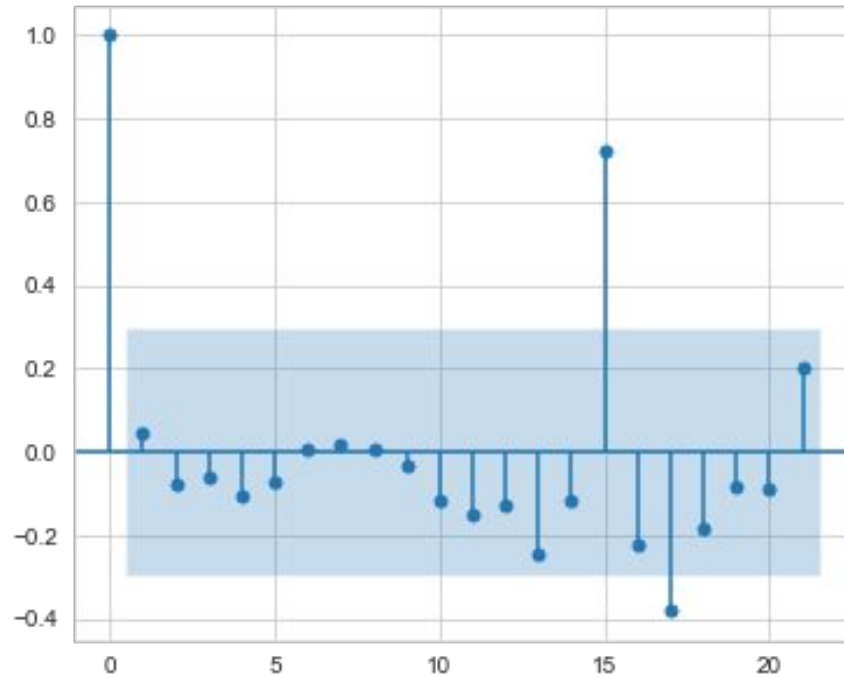
Stationarity Tests

Test	Null Hypothesis	P-Value	Conclusion
AD Fuller	Non-Stationary	2.57 e-07	Stationary
KPSS	Stationary	0.1	Stationary

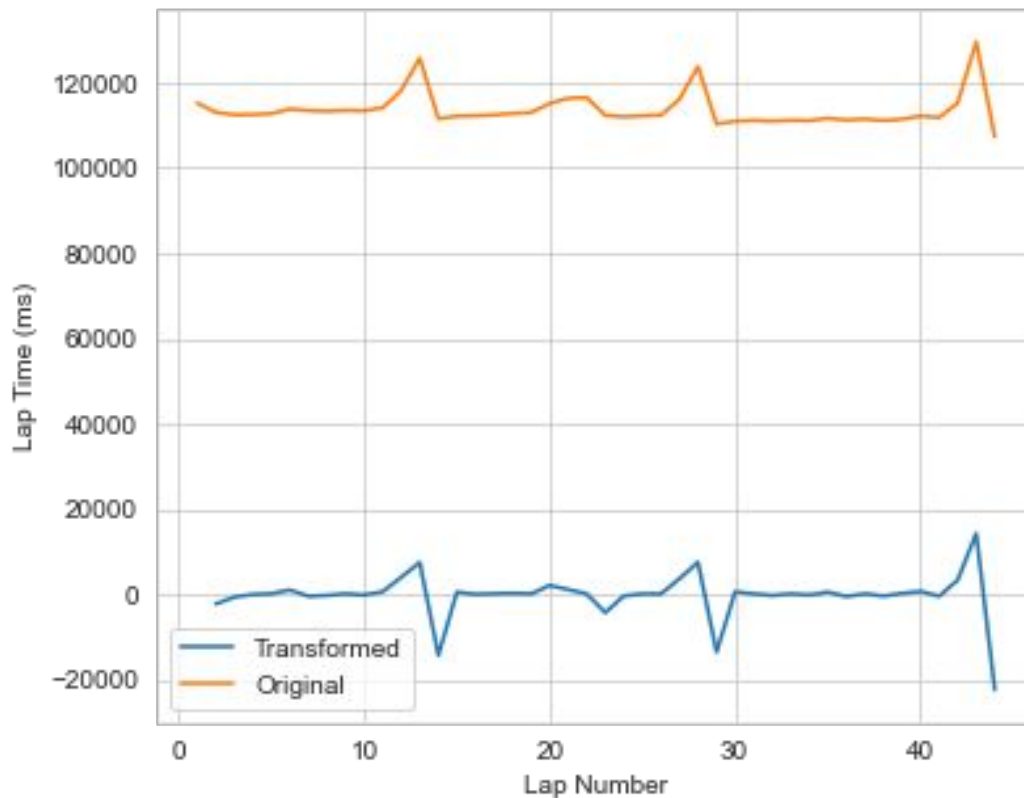
ACF



PACF



Box-Cox & Differencing Transformation



Analysis Conclusion

- ✓ Level
- × Trend
- × Seasonality
- ✓ Exogenous Effects
- ✓ Stationarity
- ✓ Autocorrelation

Key Models

Approach

1. Simple -> Complex
2. Baseline -> Grid Search
3. Two Family of Models:
 - Exponential Smoothing
 - Autoregressive
4. Evaluation Metrics:
 - Root Mean Squared Error (RMSE)
 - Mean Absolute Percentage Error (MAPE)
5. Walk Forward Forecasts

Models

Model	Type	Captures
Simple Exponential Smoothing	Decompositional	Level

Models

Model	Type	Captures
Simple Exponential Smoothing	Decompositional	Level
Holt's Exponential Smoothing	Decompositional	Level + Trend

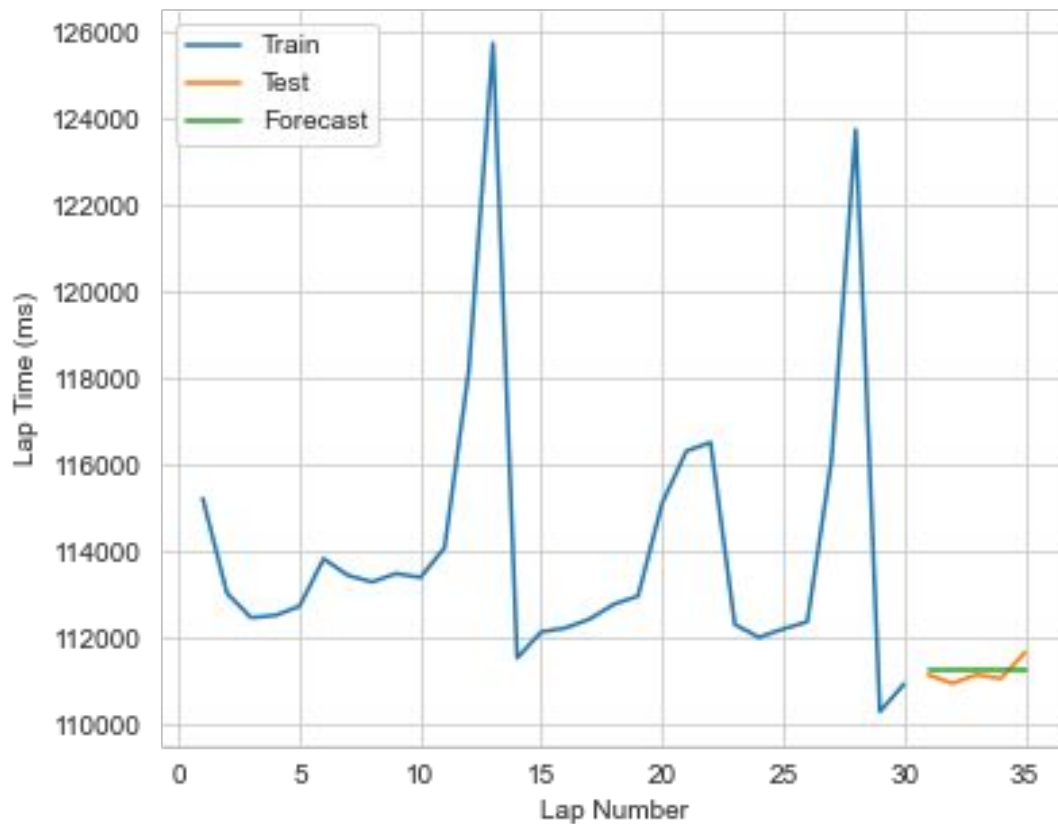
Models

Model	Type	Captures
Simple Exponential Smoothing	Decompositional	Level
Holt's Exponential Smoothing	Decompositional	Level + Trend
Holt Winter's Exponential Smoothing	Decompositional	Level + Trend + Seasonality

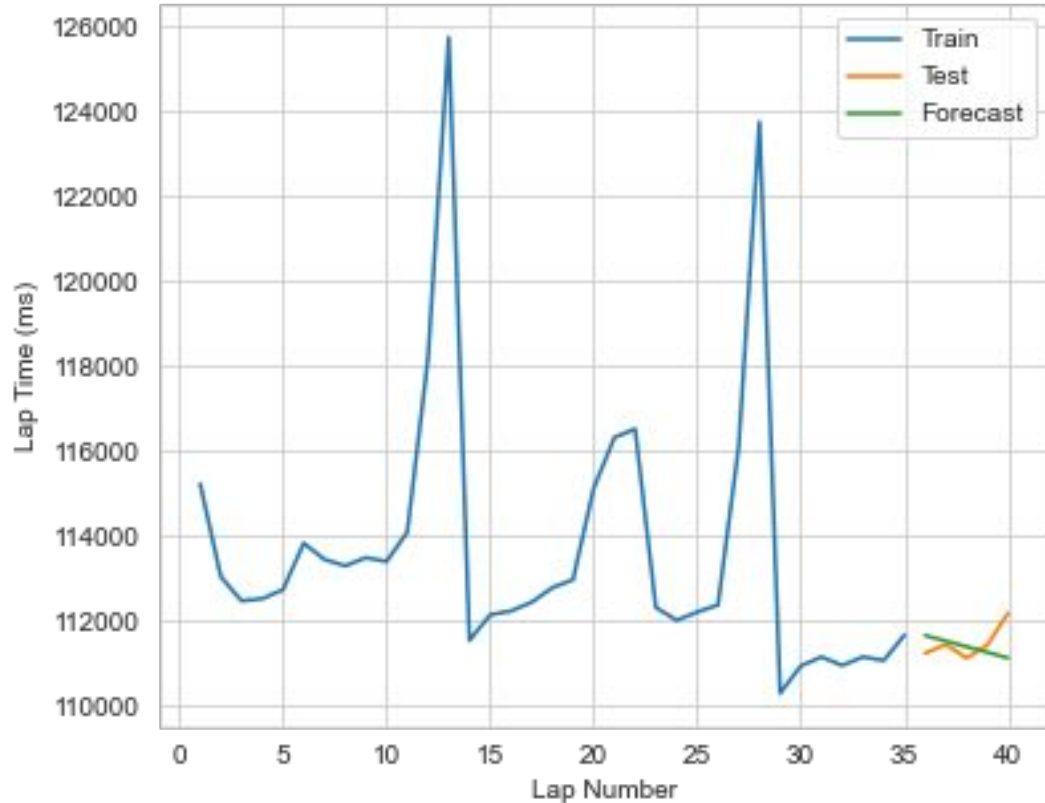
Models

Model	Type	Captures
Simple Exponential Smoothing	Decompositional	Level
Holt's Exponential Smoothing	Decompositional	Level + Trend
Holt Winter's Exponential Smoothing	Decompositional	Level + Trend + Seasonality
Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX)	Regression	Correlations + Trend + Exogenous Variable

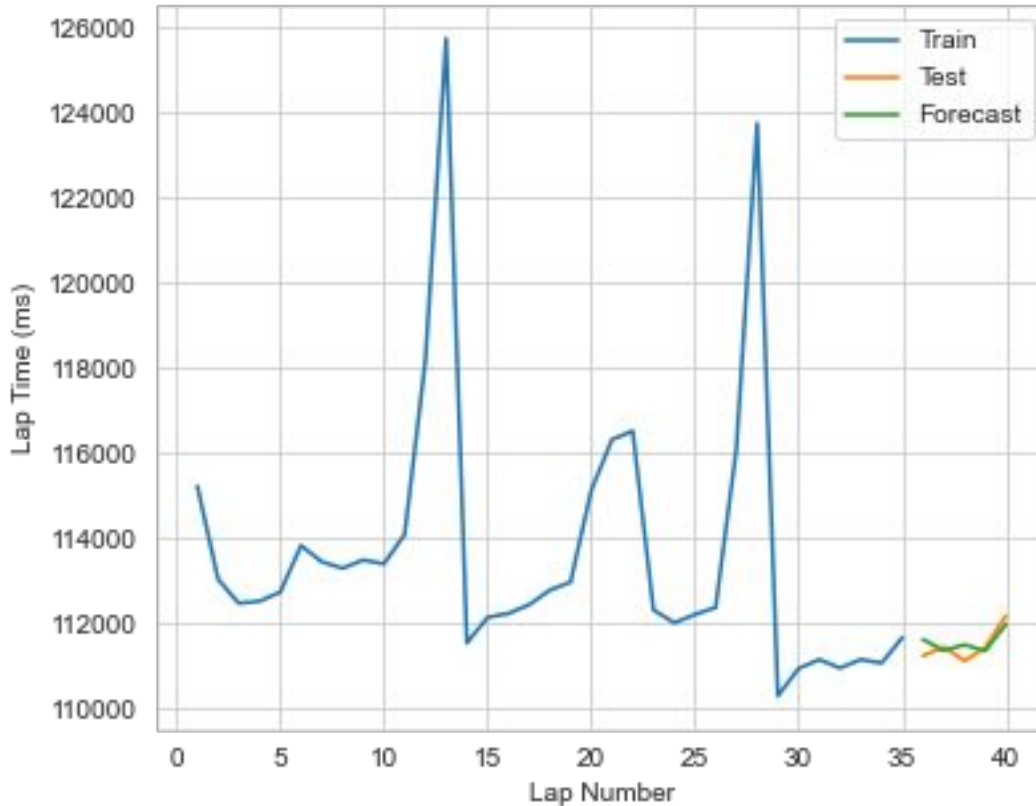
SES ($\alpha=0.8$)



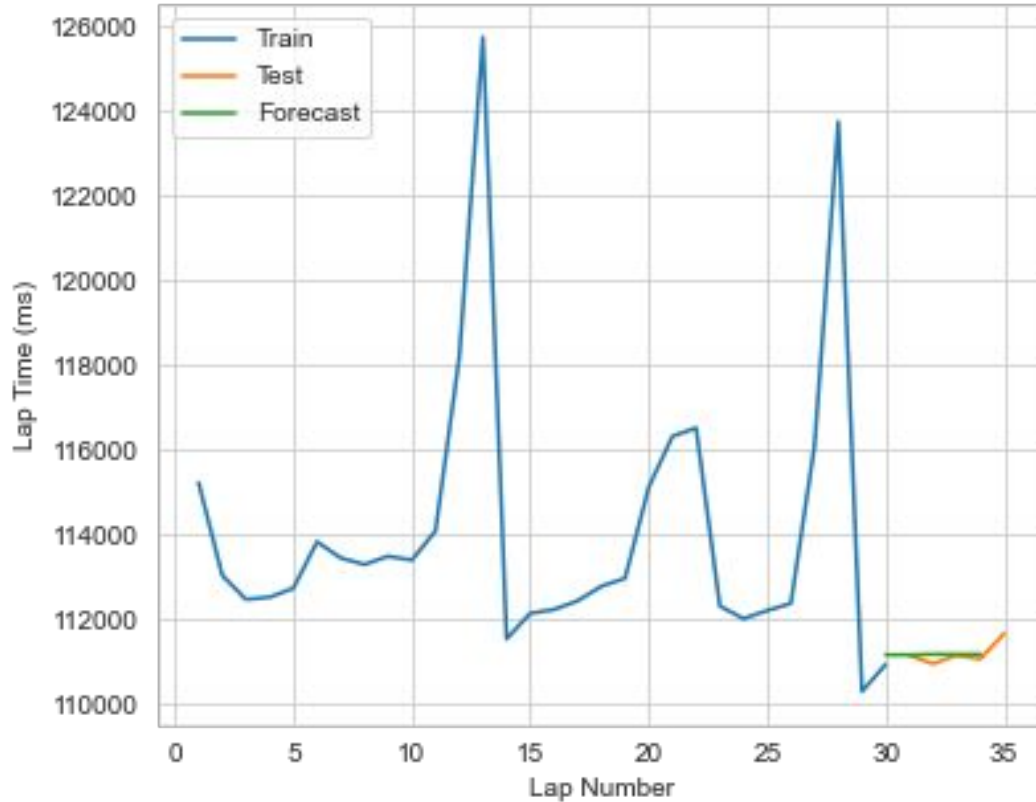
Holt's ($\alpha=0.2$ | $\beta=0.1$)



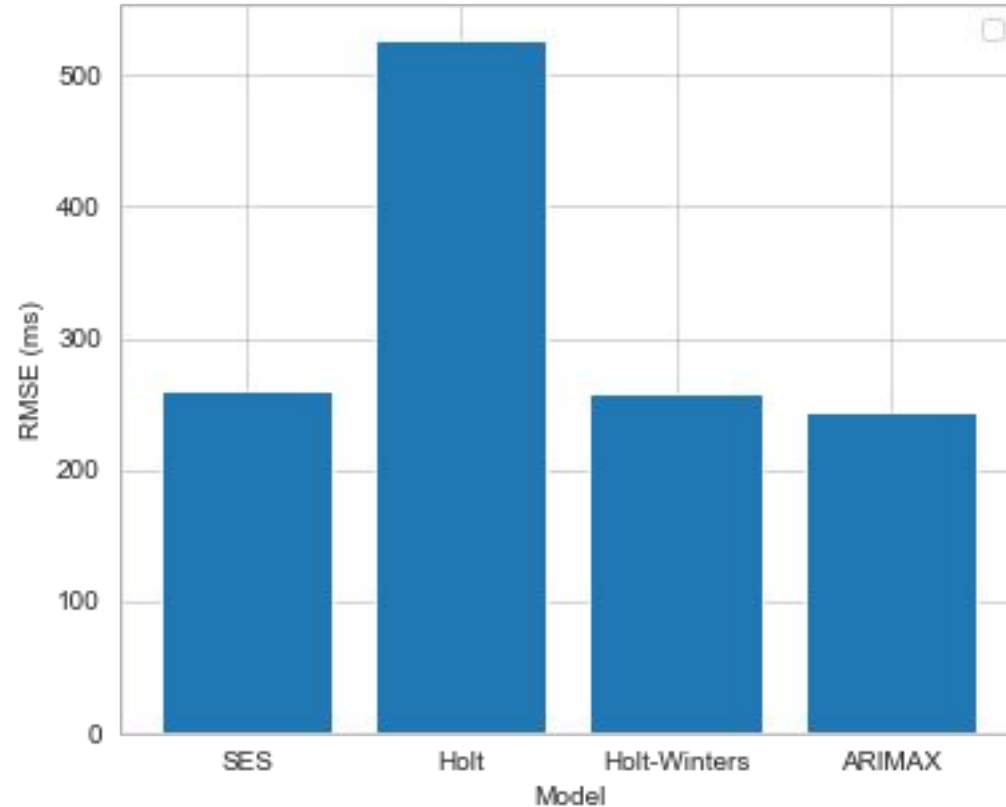
Holt Winter's ($\alpha=0.2$ | $\beta=0.04$ | $\gamma=0.01$)



ARIMAX (p=1 | d=1 | q=4)



Model Comparison



Walk Forward Forecasts

Methodology

Initial Forecast

1. Train model on first 35 laps
2. Forecast laps 36-40

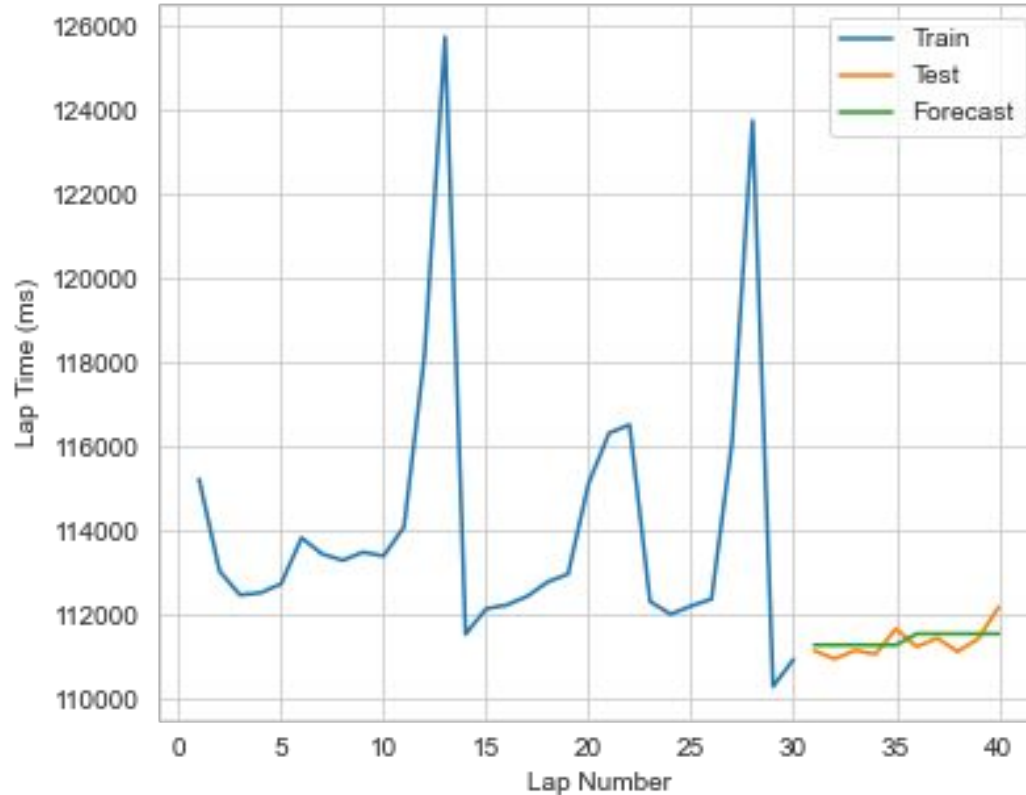
Expanding Forecast

1. Re-train model on first 40 laps
2. Forecast laps 40-45

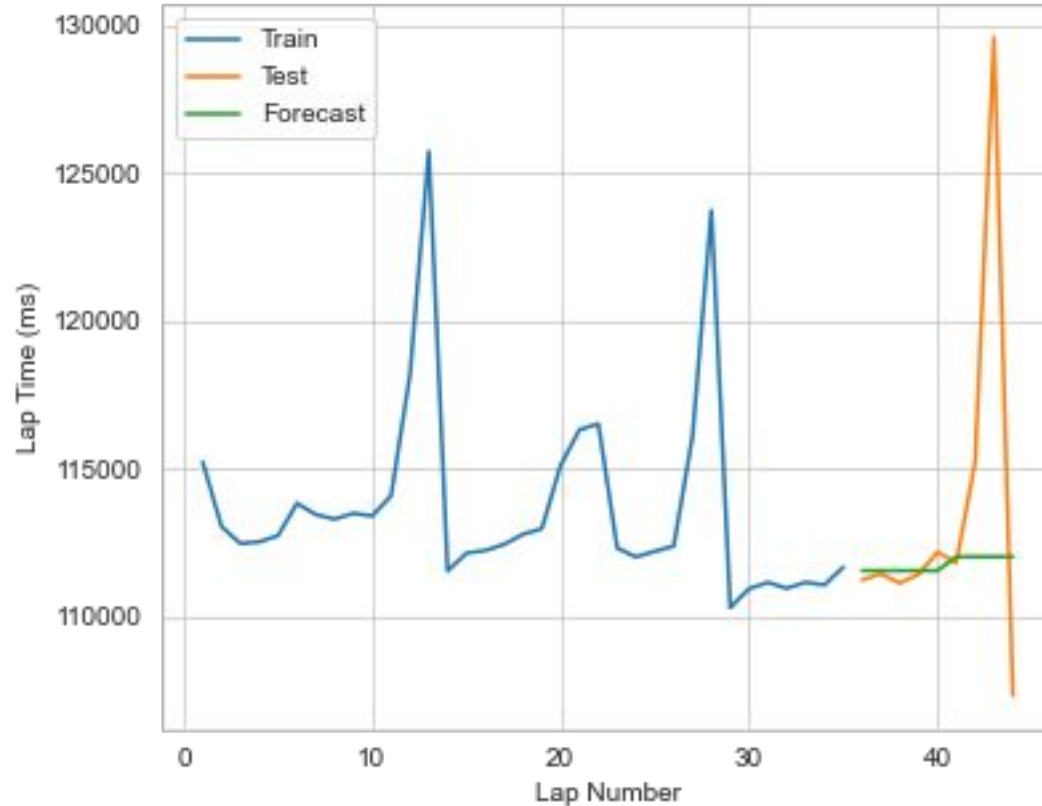
Models - Walk Forwards

1. Simple Exponential Smoothing (SES)
2. Holt Winter's
3. ARIMAX

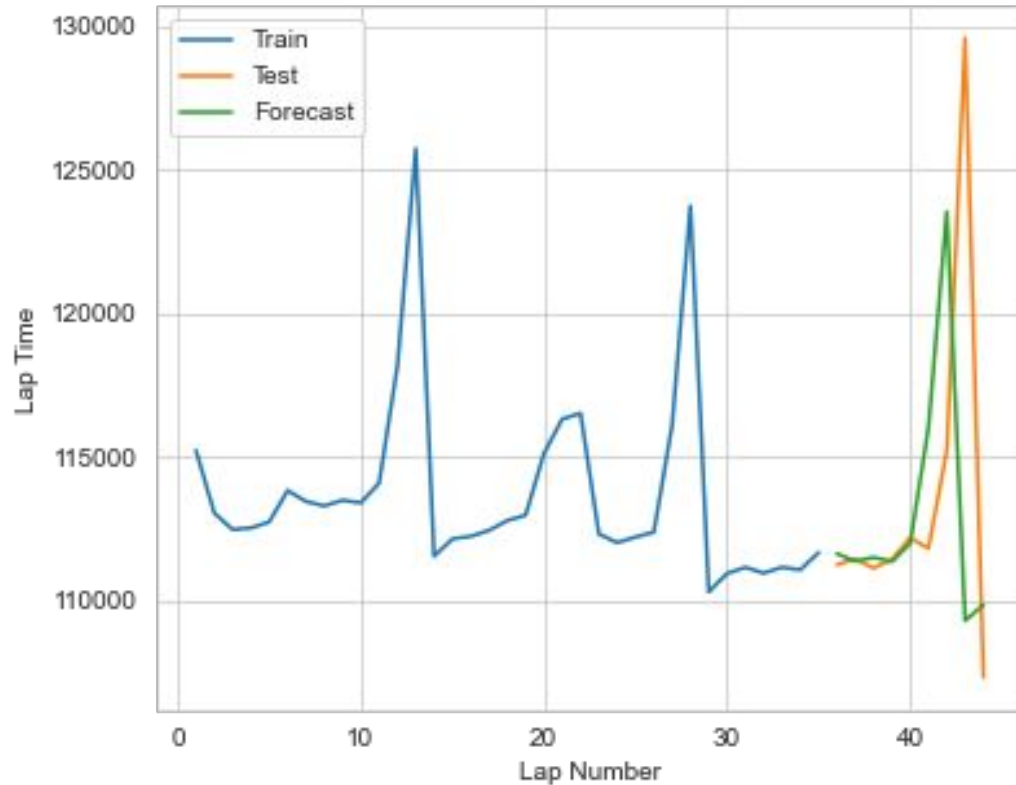
SES - Walk Forward Forecast



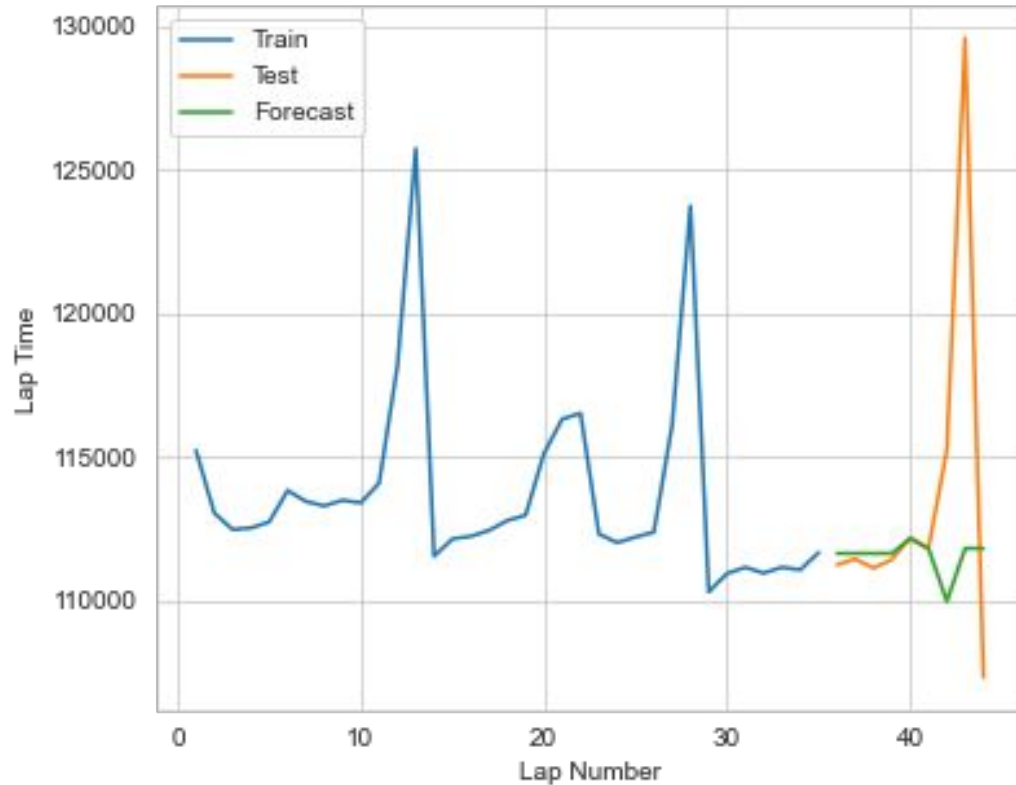
SES - Walk Forward Forecast



Holt Winter's - Walk Forward Forecast



ARIMAX - Walk Forward Forecast





Future Improvements



Improvements

1. Shift pit stops across laps
2. Remove pit stops prior to forecasting
3. Grid Search with Look Forward Forecast
4. Cross-Validation
5. Look under the hood
6. Explore other algorithms - LSTM, VAR, Random Forest, ensembles
7. Base forecasts on real-time explanatory variables (longitudinal data)

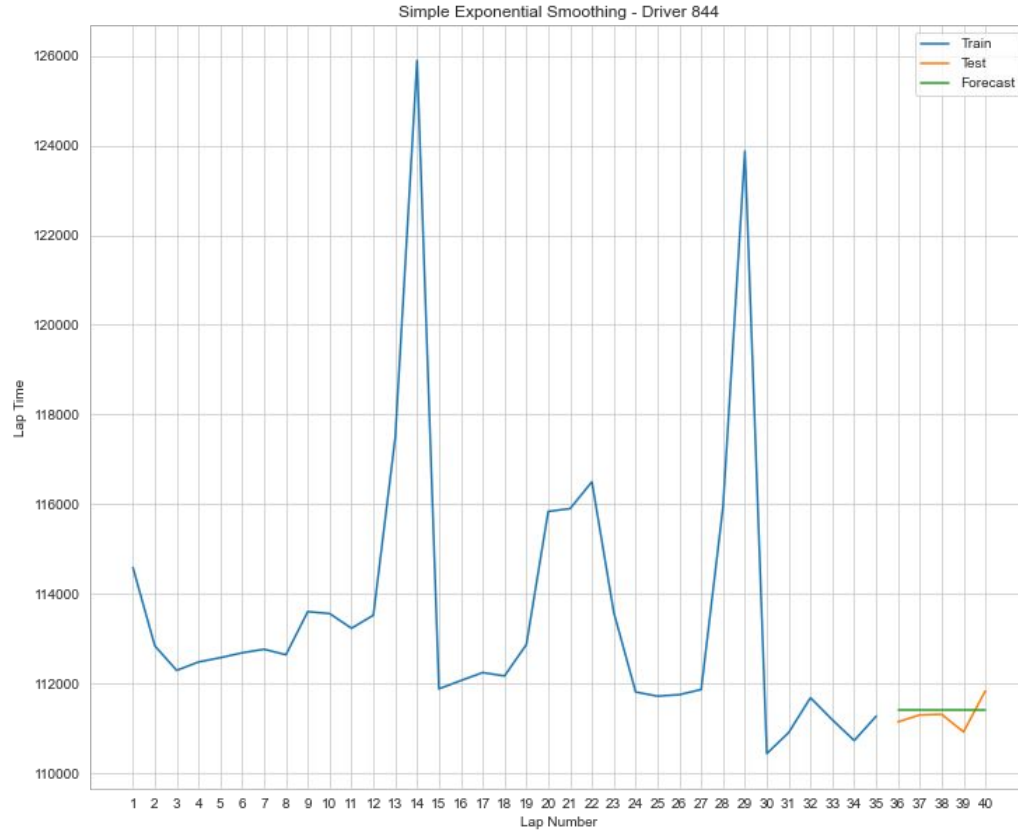


Questions?



Appendix

SES ($\alpha=0.8$) - Driver 844



ARIMAX (p=4 | d=1 | q=15) - Driver 844

