
"I have seen the future and it is very much like the present, only longer."

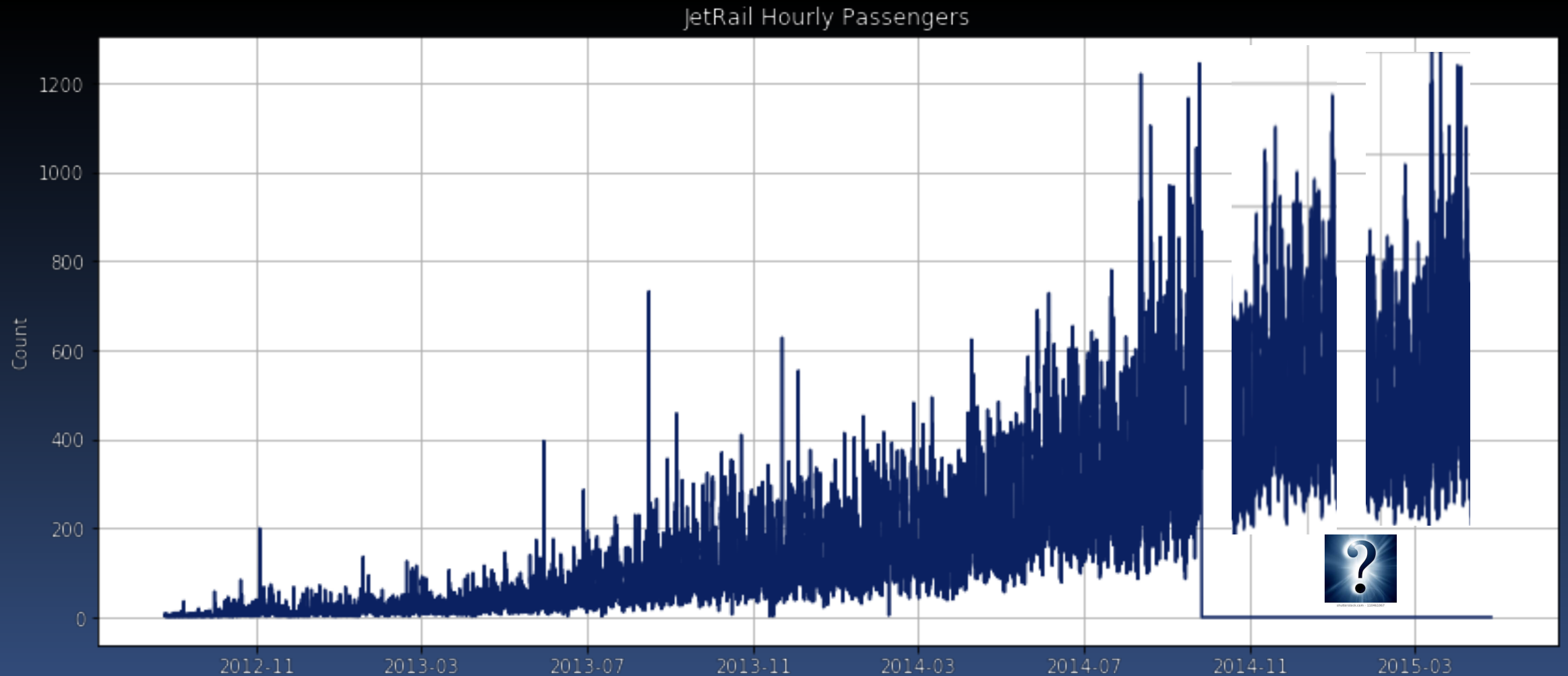
—Kehlog Albran, *The Profit*



JetRail Ridership Forecast

TIME SERIES USING AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA)

DATS6501 / Data Science Capstone / Student: Barry Blackburn / Professor: Amir Jafari, PhD / December 7, 2018



PROBLEM

- Whether to invest in a new tech transportation company - JetRail
- JetRail uses jet propulsion technology to run rails and move people at a high speed
- Investment decision requires accurate projection of 1 million passengers
- Given: 18,288 hourly rider counts over 25 months (24 x7)
- Predict: 5,112 subsequent hourly rider counts over 7 months
- Evaluate: Root Mean Square Error (RMSE)

DATA

TRAIN

8/25/12 - 9/25/14

TEST

9/26/14 - 4/26/15

TRAIN			TEST	
ID	Datetime	Count	ID	Datetime
0	25-08-2012 00:00	8	18288	26-09-2014 00:00
1	25-08-2012 01:00	2	18289	26-09-2014 01:00
2	25-08-2012 02:00	6	18290	26-09-2014 02:00
3	25-08-2012 03:00	2	18291	26-09-2014 03:00
4	25-08-2012 04:00	2	18292	26-09-2014 04:00
18283	25-09-2014 19:00	868	23395	26-04-2015 19:00
18284	25-09-2014 20:00	732	23396	26-04-2015 20:00
18285	25-09-2014 21:00	702	23397	26-04-2015 21:00
18286	25-09-2014 22:00	580	23398	26-04-2015 22:00
18287	25-09-2014 23:00	534	23399	26-04-2015 23:00

2,541,266

?,???,???

[HTTPS://DATAHACK.ANALYTICSVIDHYA.COM/CONTEST/PRACTICE-PROBLEM-TIME-SERIES-2/](https://datahack.analyticsvidhya.com/contest/practice-problem-time-series-2/)

TOOLS



METHODOLOGY

- 1.EDA - Analyze time series frequencies & averages
- 2.Conduct tests for Stationarity
- 3.Evaluate model parameters
- 4.Evaluate SARIMAX models
- 5.Execute model forecast
- 6.Analyze Results
- 7.Build contest submission file
- 8.Submit forecast for grade

TERMINOLOGY

- Time Series

- Sequential set of data points measured over successive times $\{x(t), t = 0, 1, 2, \dots\}$
- Continuous vs Discrete, Uni vs Multivariate

- Components

- Frequency
- Trend, Seasonality, Cyclicity, Irregularities
- Additive vs Multiplicative

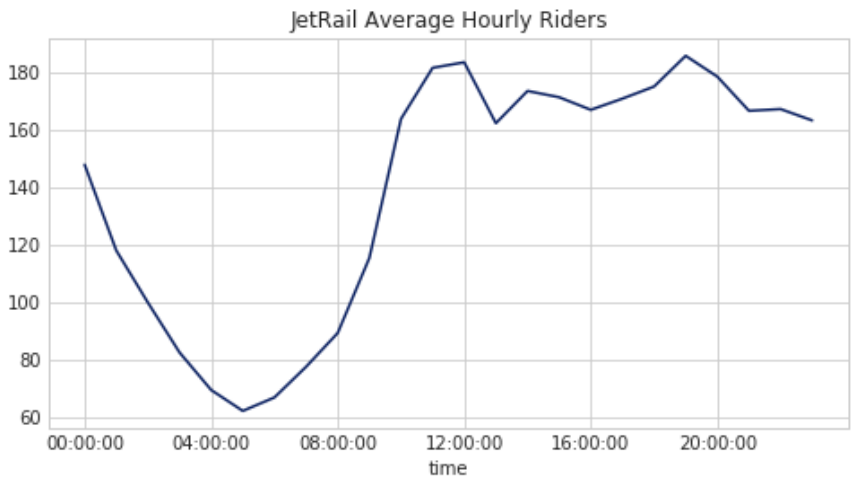
GW

- Stationarity

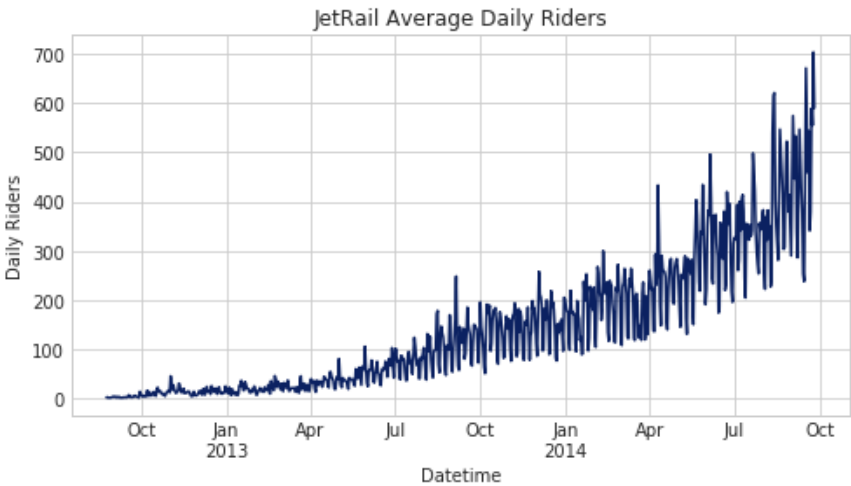
- Stochastic - unconditional going probability does not change when shifted in time

FREQUENCIES

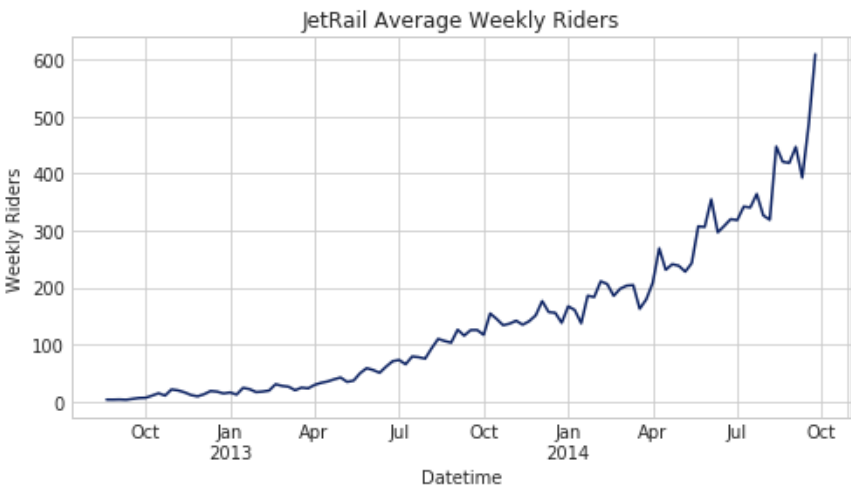
Hourly



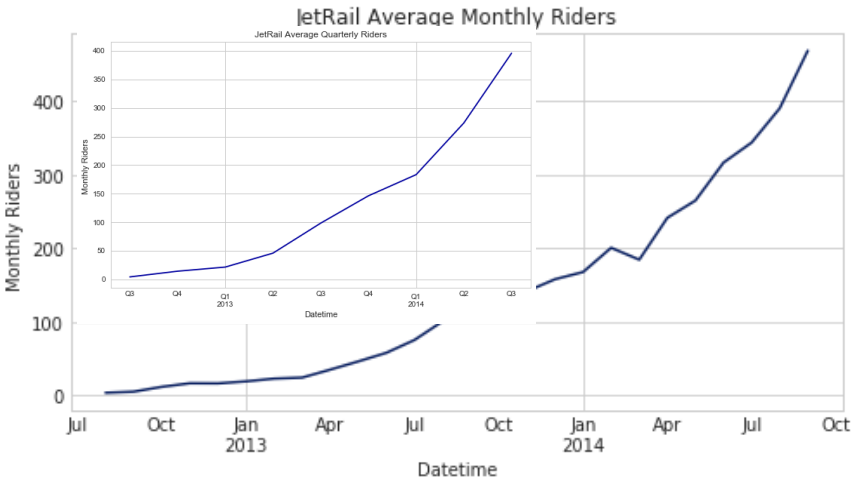
Daily



Weekly

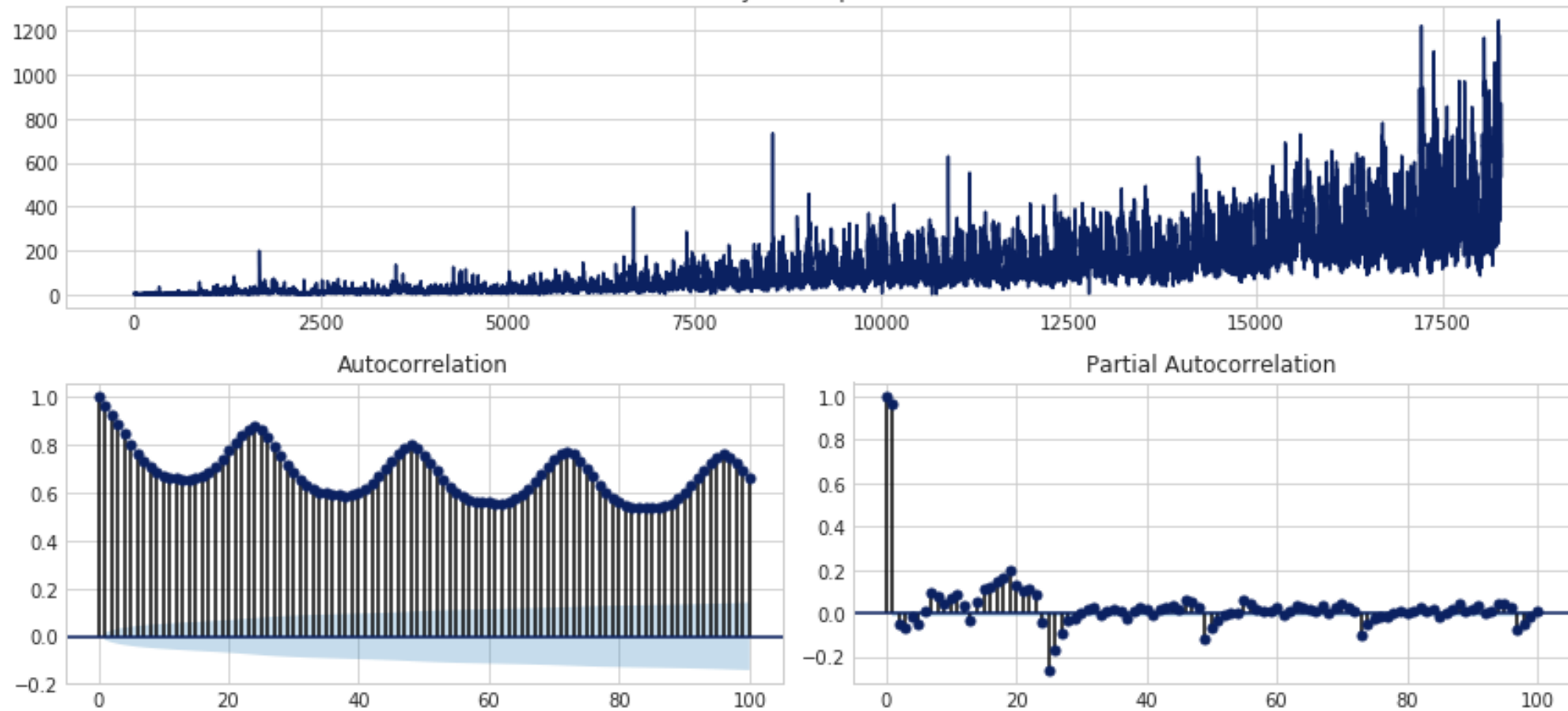


Quarterly / Monthly

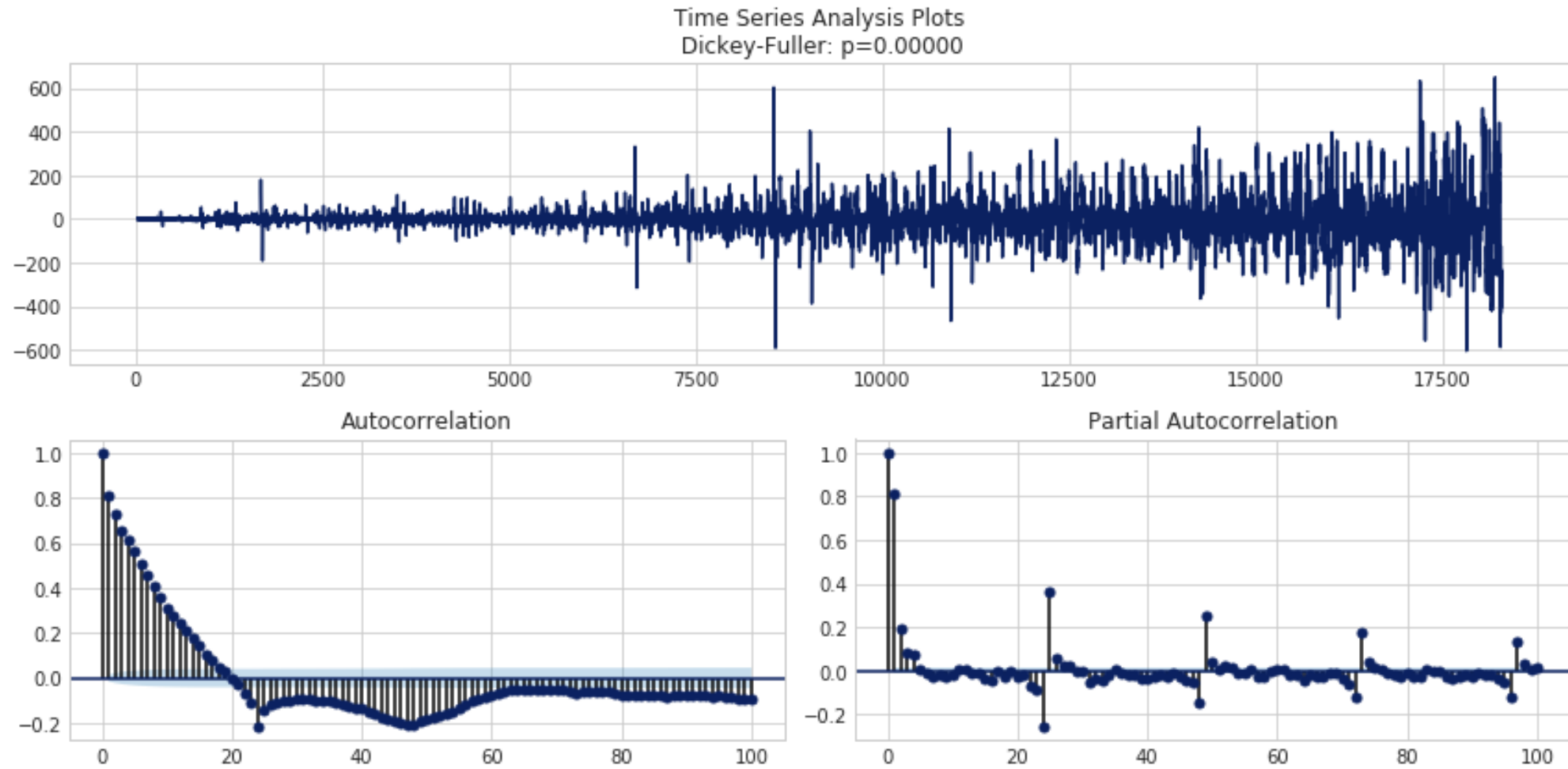


STATIONARITY TEST #1

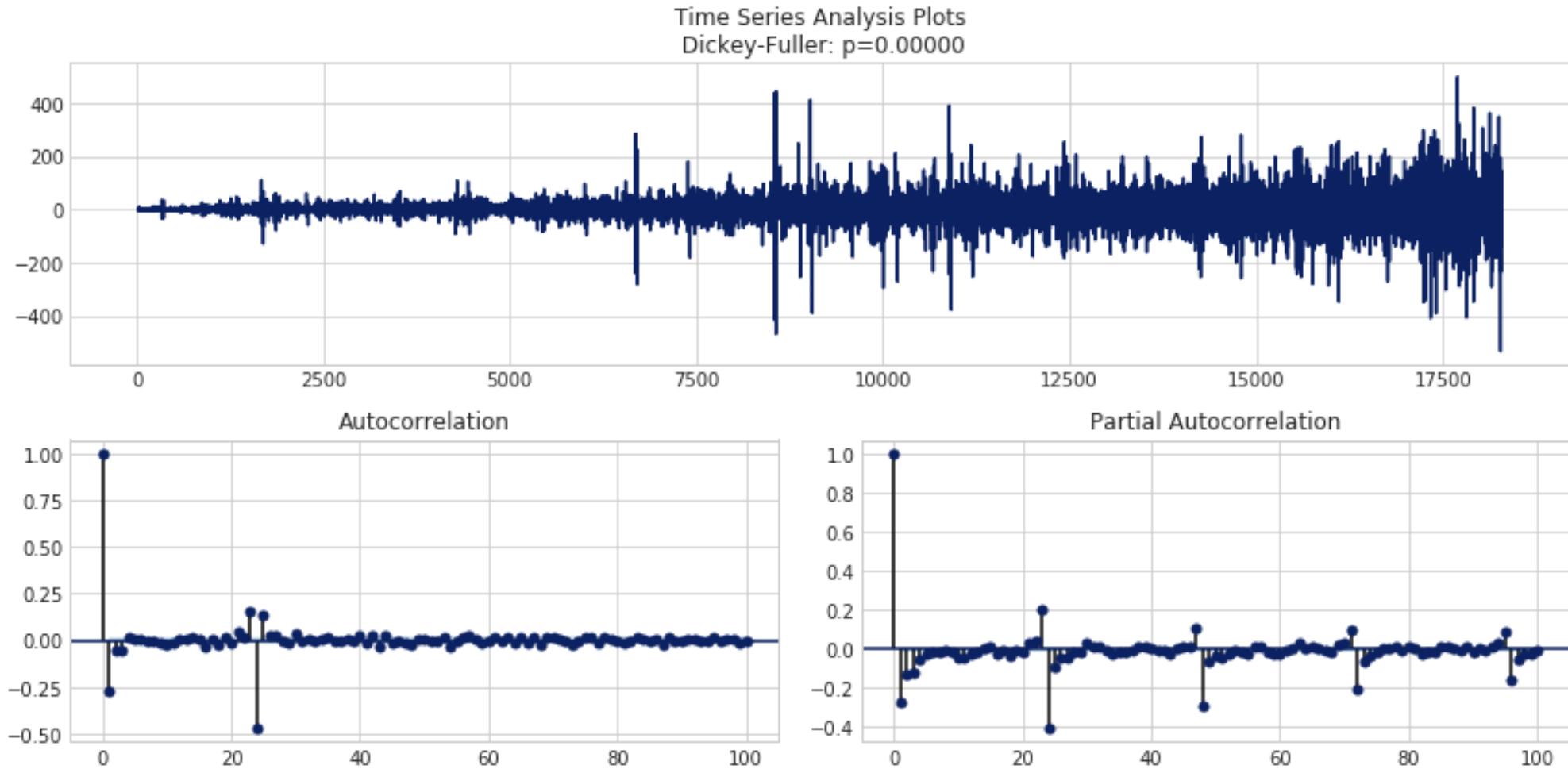
Time Series Analysis Plots
Dickey-Fuller: $p=0.00024$



STATIONARITY TEST #2



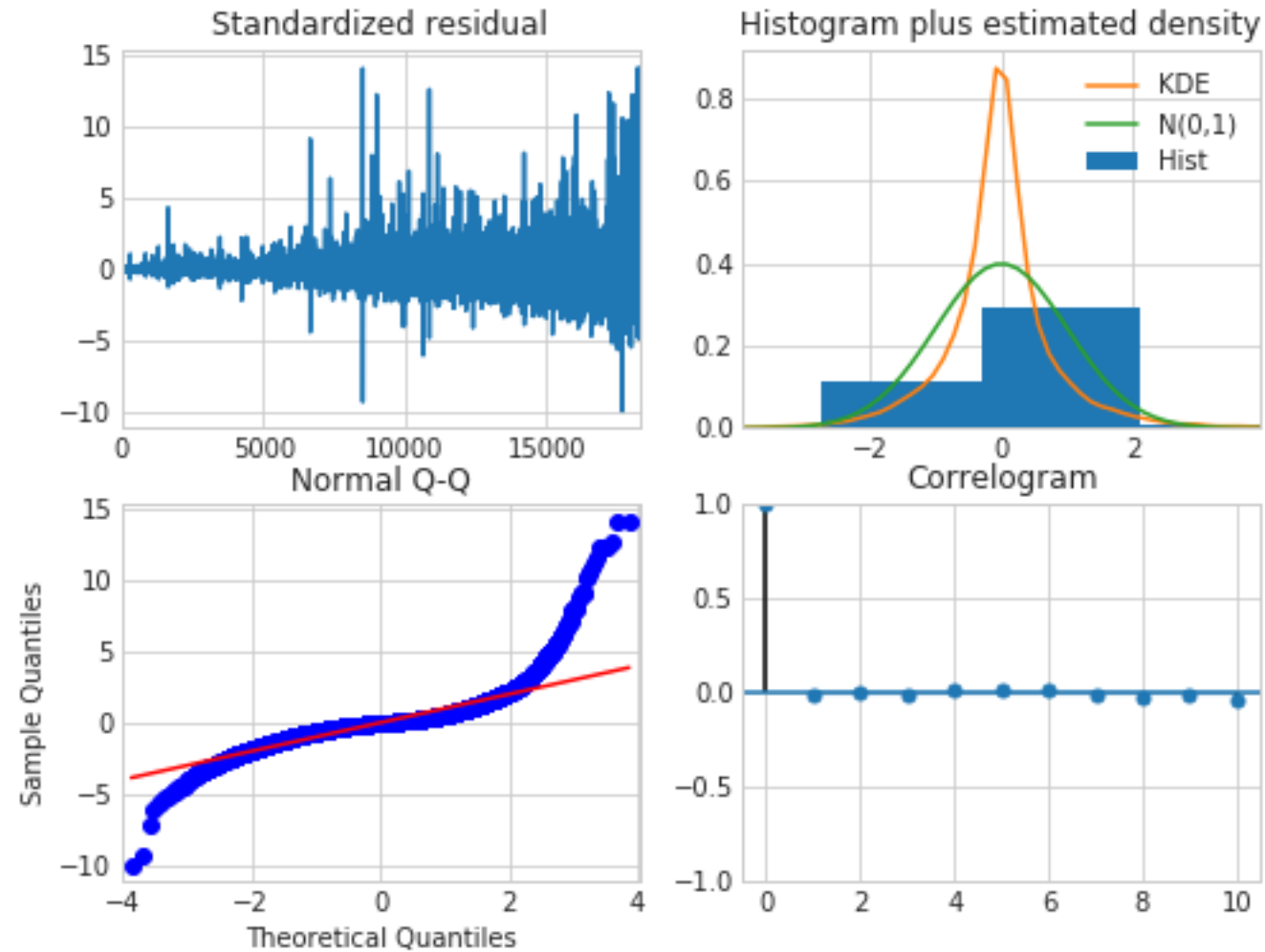
STATIONARITY TEST #3



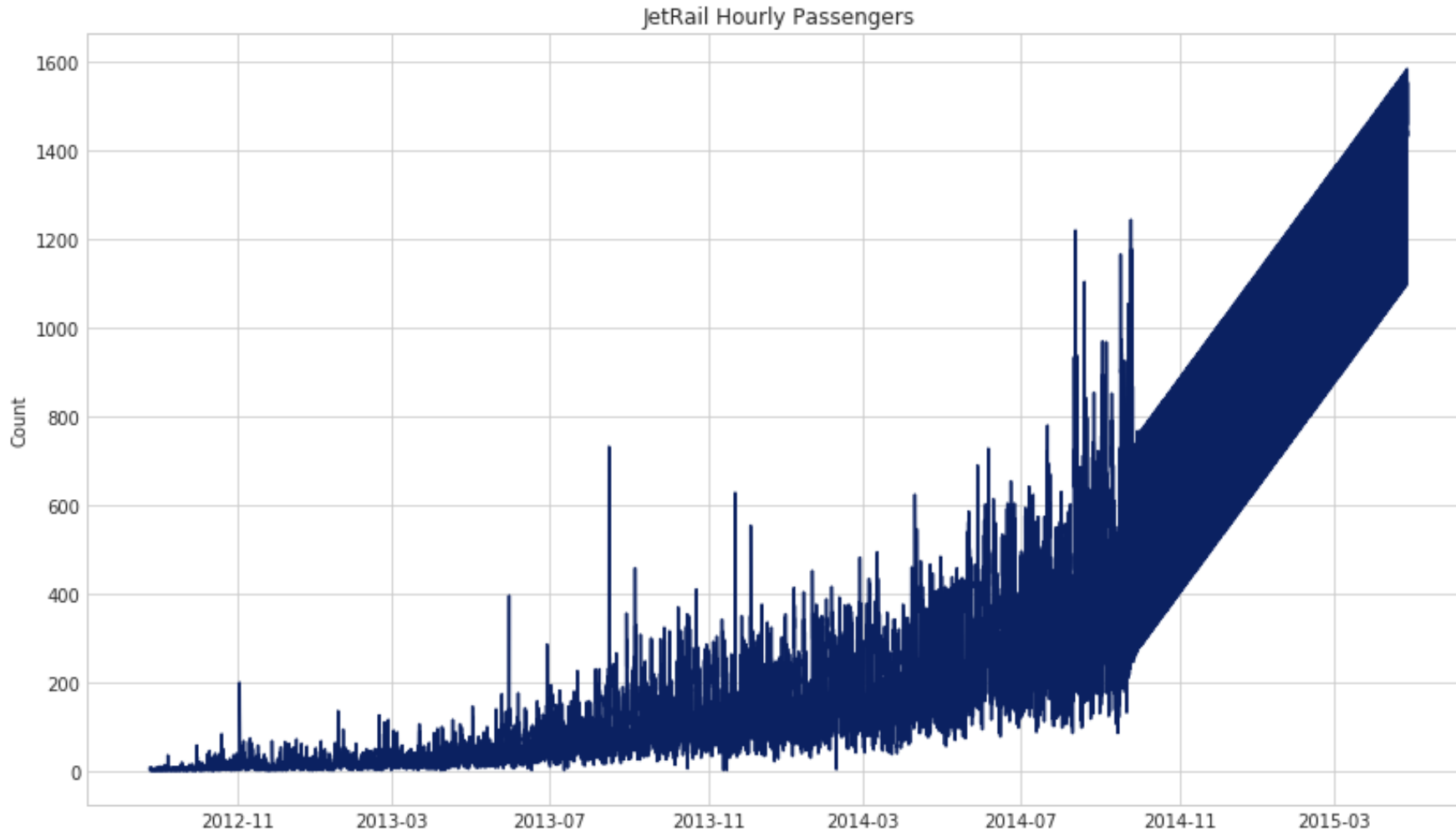
MODEL SUMMARY



MODEL DIAGNOSTICS



PREDICTION - FORECAST











CONTEST

[Blog](#)[Learn](#)[Engage](#)[Compete](#)[Get Hired](#)[View All Hackathons](#)

Home

Public Leaderboard - Practice Problem: Time Series

#	Name	Score	Submission Trend
1	 smibimj1995	132.2925845437	
108	 barry09	186.4531584689	
302	 k1917	269.1363009422	
453	 VijayashreeT	657.5227499233	

<https://datahack.analyticsvidhya.com/contest/practice-problem-time-series-2/lb?page=2>

QUESTIONS?

