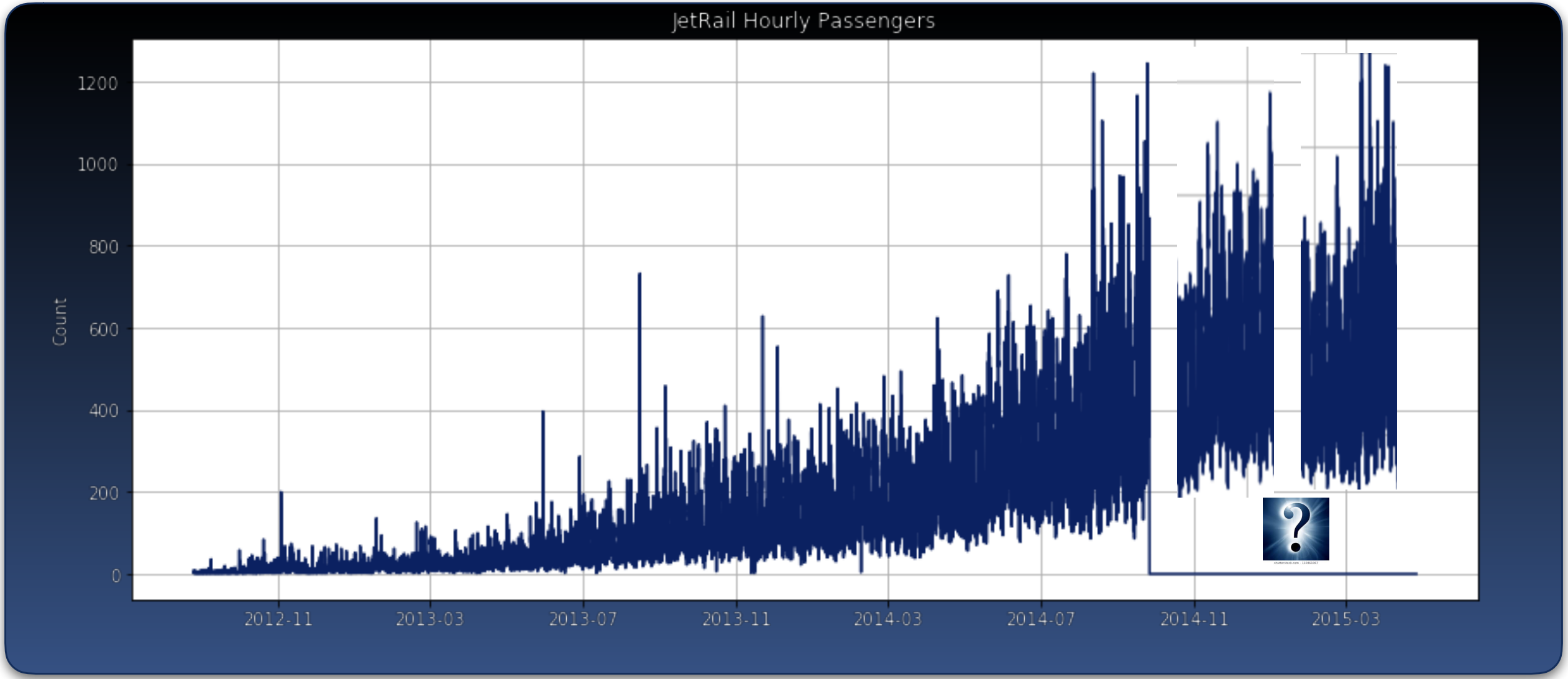
"I have seen the future and it is very much like the present, only longer."

—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/

TOOLS

METHODOLOGY

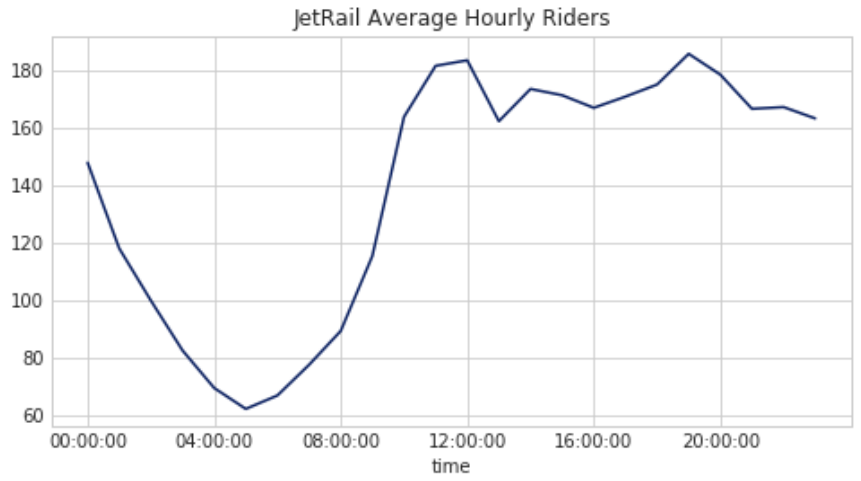
1. EDA - Analyze time series frequencies & averages
2. Conduct tests for Stationarity
3. Evaluate model parameters
4. Evaluate SARIMAX models
5. Execute model forecast
6. Analyze Results
7. Build contest submission file
8. Submit forecast for grade

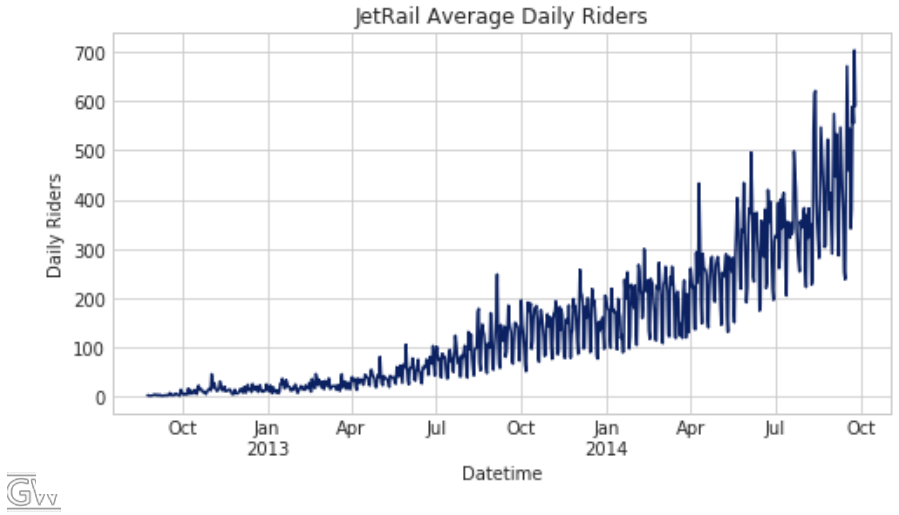
TERMINOLOGY

- Time Series
  - Sequential set of data points measured over successive times $\{x(t),\ t = 0,1,2,...\}$
  - Continuous vs Discrete, Uni vs Multivariate
- Components
  - Frequency
  - Trend, Seasonality, Cyclicality, Irregularities
  - Additive vs Multiplicative
- Stationarity
- Stochastic - unconditional joint probability does not change when shifted in time

FREQUENCIES

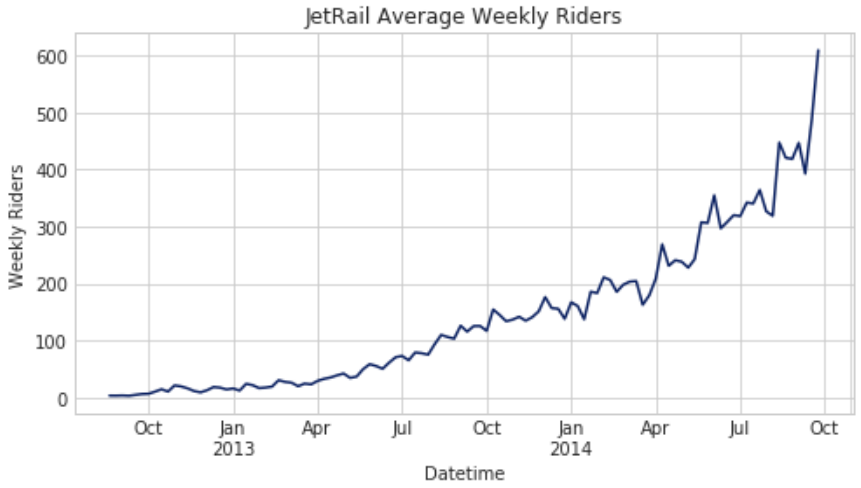
Hourly

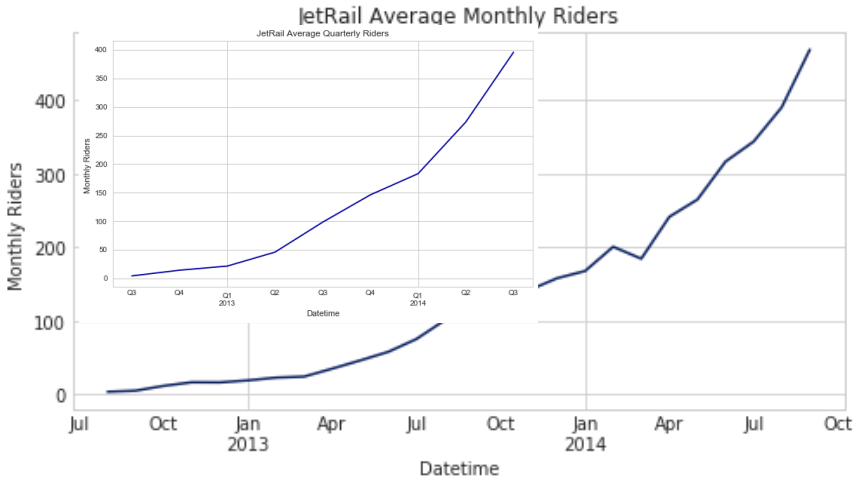




Daily

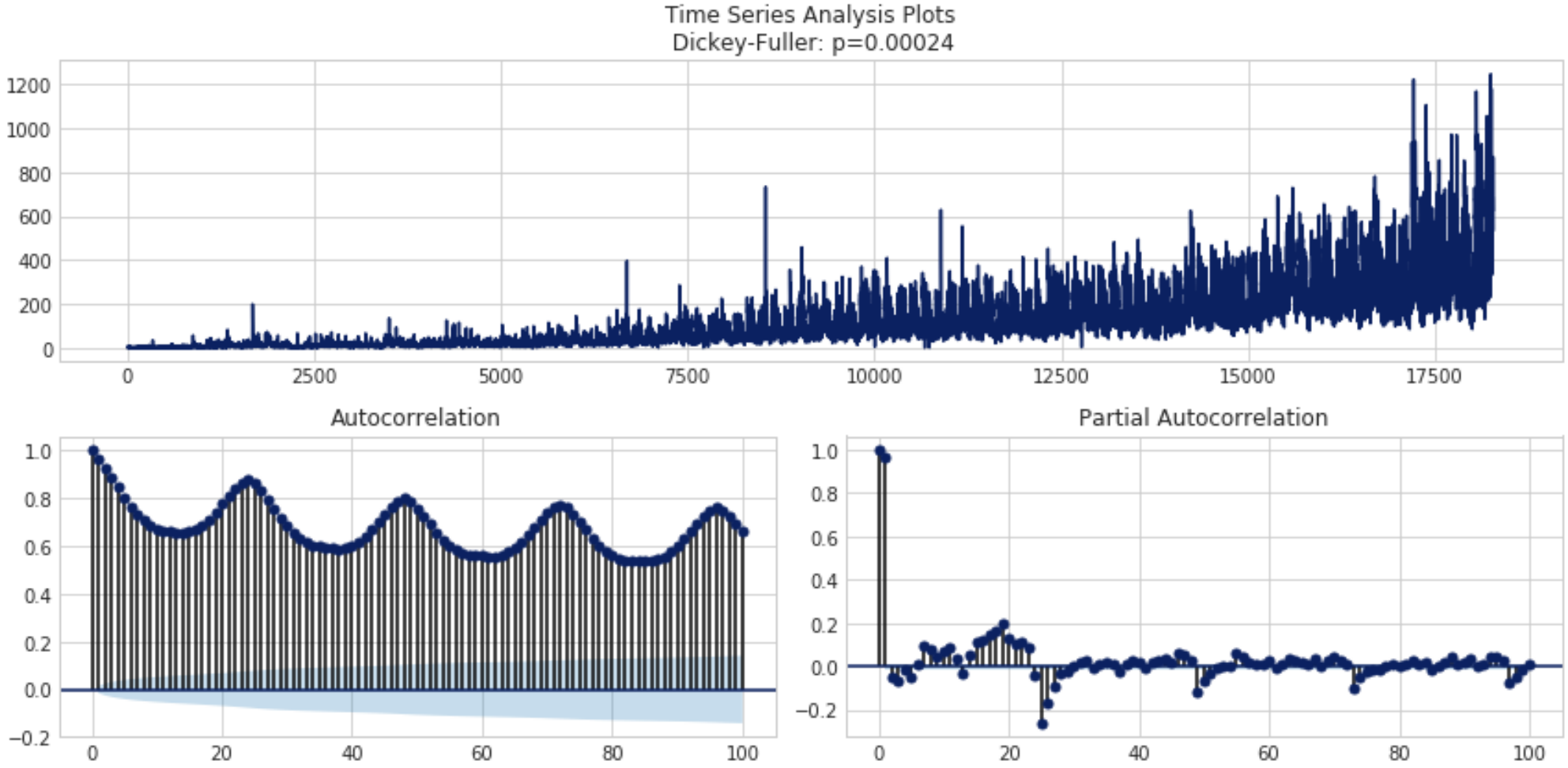
Weekly



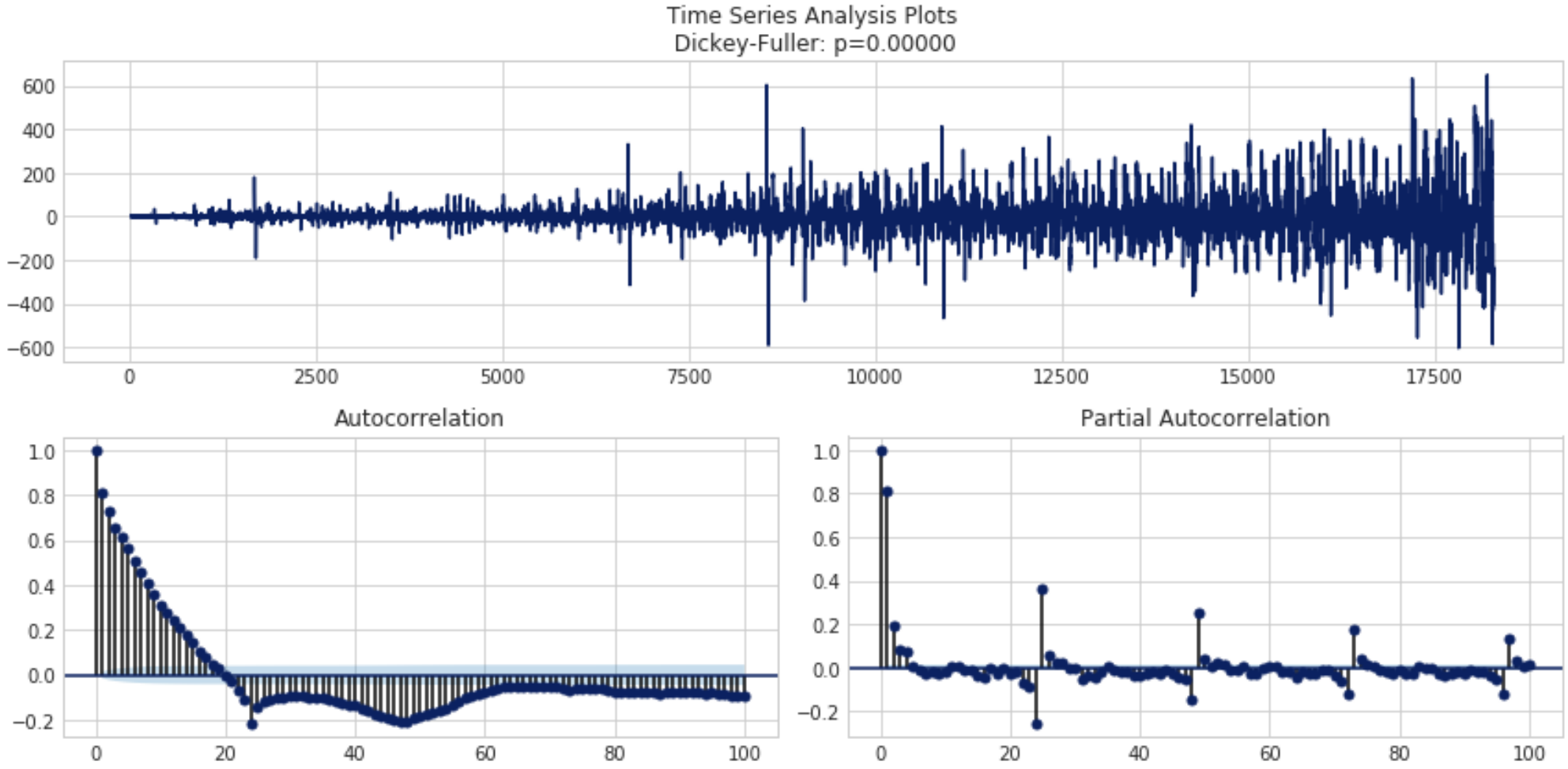


Quarterly / Monthly

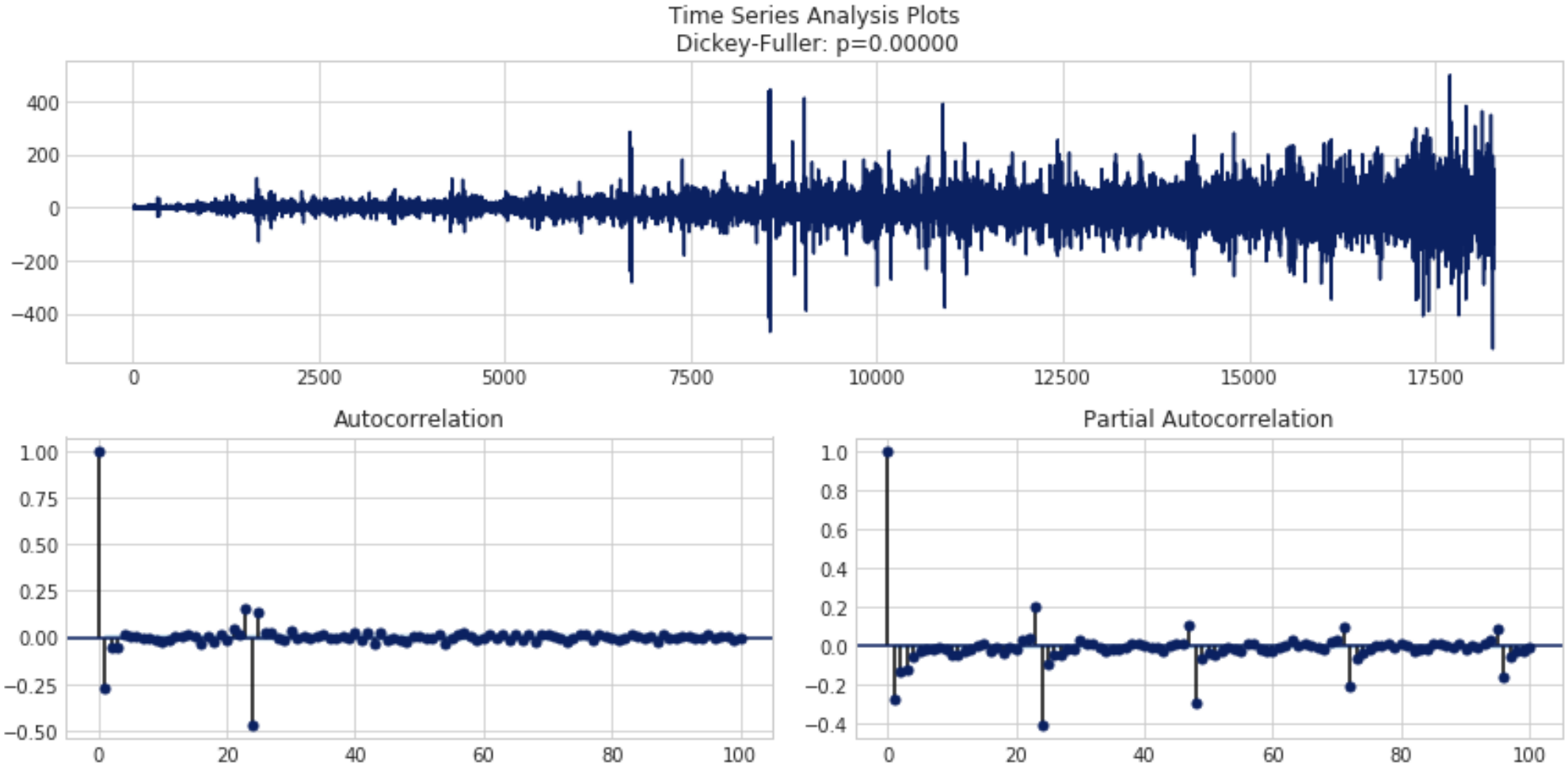
STATIONARITY TEST #1



STATIONARITY TEST #2



STATIONARITY TEST #3



MODEL SUMMARY

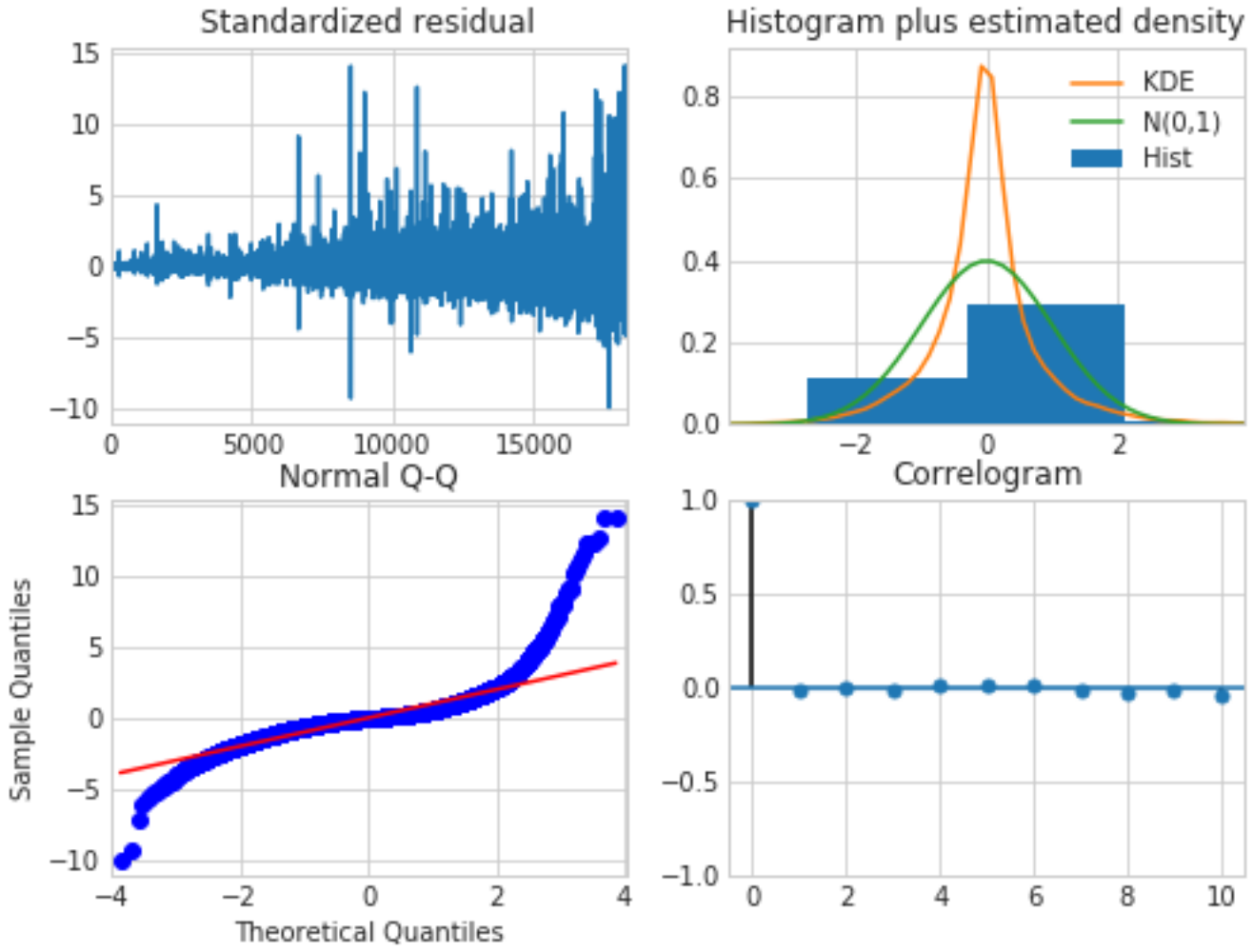
Statespace Model Results

Dep. Variable:	Count	No. Observations:	18288
Model:	SARIMAX(3, 1, 3)x(5, 1, 2, 24)	Log Likelihood	-89110.817
Date:	Sat, 08 Dec 2018	AIC	178249.633
Time:	16:46:23	BIC	178359.010
Sample:	0	HQIC	178285.576
	- 18288		
Covariance Type:	opg		

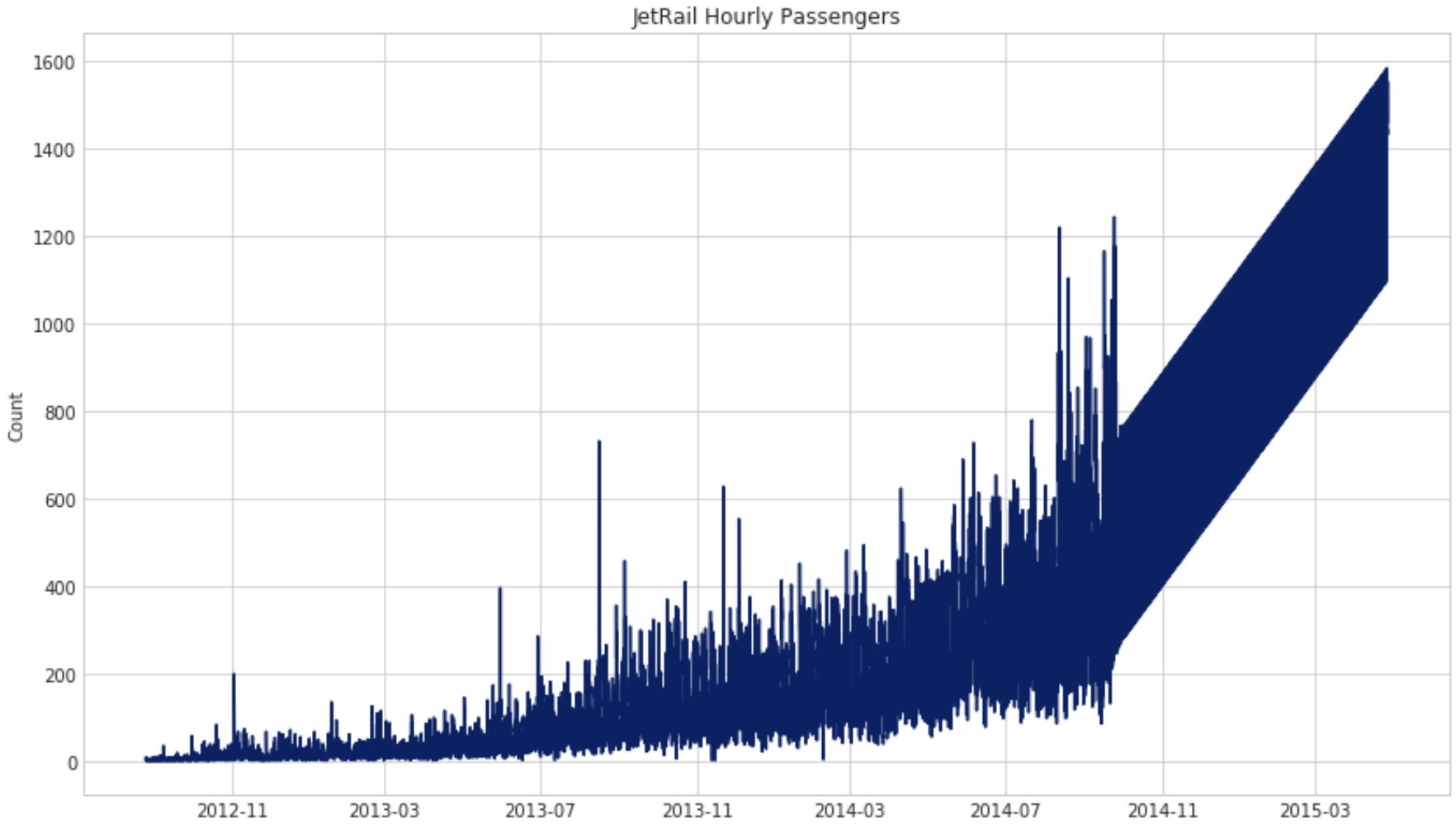
	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.2101	0.028	7.435	0.000	0.155	0.265
ar.L2	0.9474	0.025	37.174	0.000	0.897	0.997
ar.L3	-0.2426	0.015	-16.624	0.000	-0.271	-0.214
ma.L1	-0.5336	0.028	-19.056	0.000	-0.588	-0.479
ma.L2	-0.9845	0.033	-29.518	0.000	-1.050	-0.919
ma.L3	0.5221	0.018	29.003	0.000	0.487	0.557
ar.S.L24	-0.7844	0.356	-2.205	0.027	-1.482	-0.087
ar.S.L48	-0.0353	0.012	-3.065	0.002	-0.058	-0.013
ar.S.L72	-0.1049	0.022	-4.690	0.000	-0.149	-0.061
ar.S.L96	-0.0861	0.021	-4.115	0.000	-0.127	-0.045
ar.S.L120	-0.0302	0.017	-1.790	0.074	-0.063	0.003
ma.S.L24	-0.1236	0.356	-0.348	0.728	-0.820	0.573
ma.S.L48	-0.7636	0.333	-2.293	0.022	-1.416	-0.111
sigma2	1010.2187	3.304	305.789	0.000	1003.744	1016.694

Ljung-Box (Q):	192.00	Jarque-Bera (JB):	474958.53
Prob(Q):	0.00	Prob(JB):	0.00
Heteroskedasticity (H):	25.83	Skew:	2.18
Prob(H) (two-sided):	0.00	Kurtosis:	27.60

MODEL DIAGNOSTICS



PREDICTION - FORECAST



CONTEST

Analytics Vidhya
Learn everything about analytics

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?