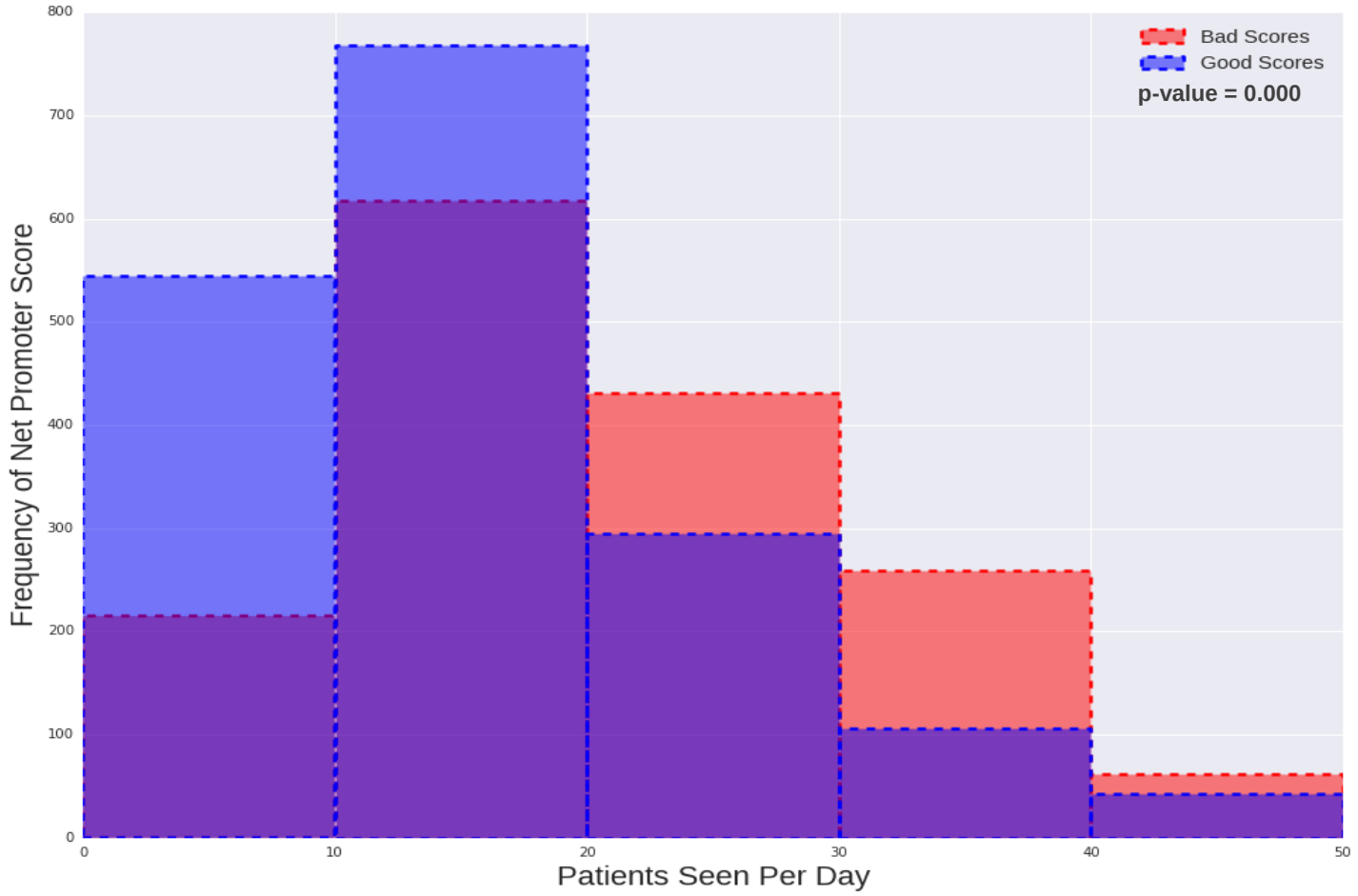
Polling Patients: Elevating the Healthcare Experience

ANJALI SHAH

INSIGHT PROJECT PRESENTATION

Finding the Optimal Number of Patients Per Provider

Objective: Finding the optimal number of patients per provider to get good scores



Approach

- Net Promoted Score (NPS) was binarized into good (>=90th percentile) and bad (<90th percentile)
- Non-parametric Mann-Whitney U test was performed to compare if the means of good and bad score distributions were significantly different

Actionable Insights

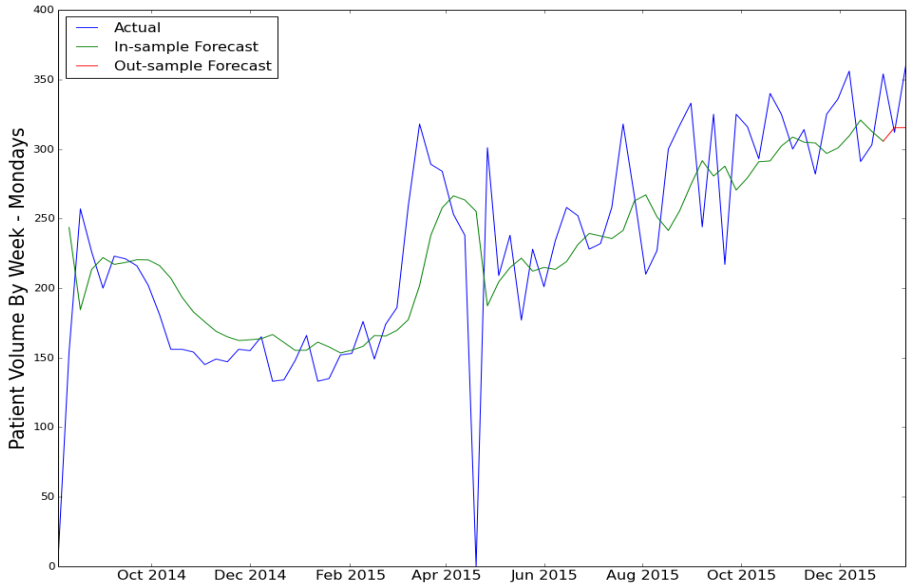
- Statistically significant difference between good and bad score distributions (p-value = 0.000)
- Patients per provider in a day (95% CL):
  - ✓ Good Scores: 14 – 15 patients
  - ✓ Bad Scores: 19 – 20 patients

Forecasting Patient Volume (Weekly Data)

Objective: Predicting patient volumes to determine optimal number of providers per site

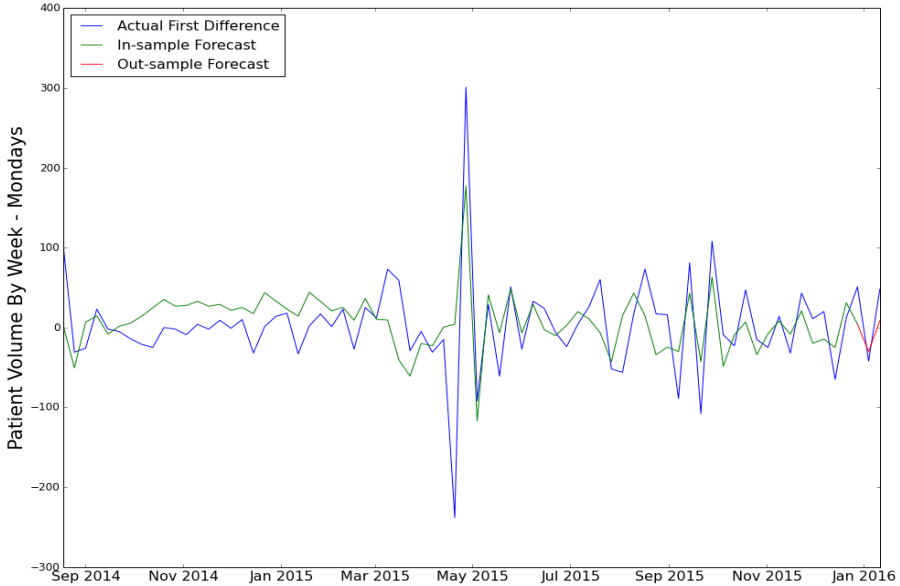
ARIMA (p=2,d=0,q=1) Model

(Non-Stationary Model)



ARIMA (p=2,d=1,q=1) Model

(Stationary Model)



Approach

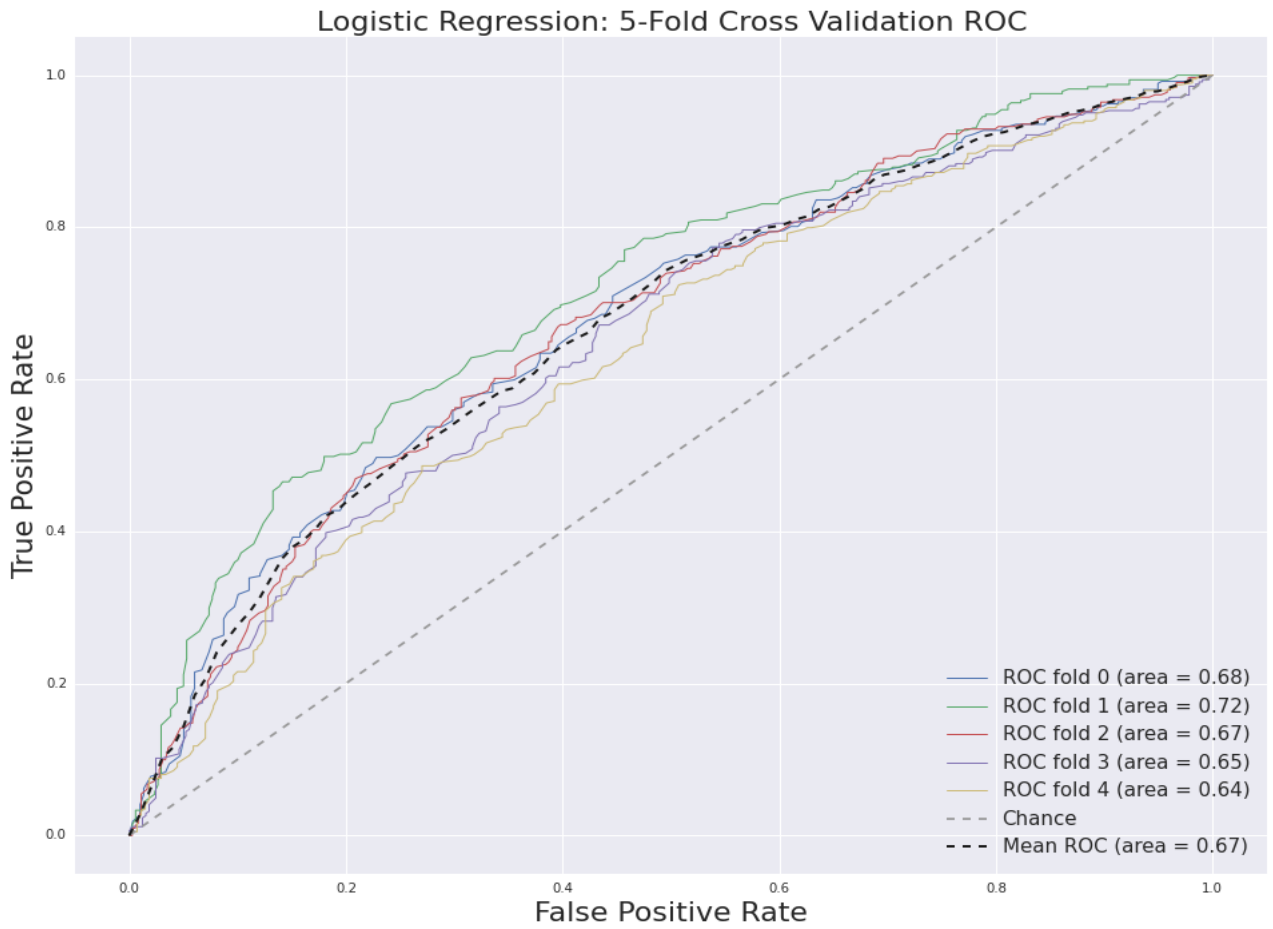
- Used ACF and PACF plots to determine order of ARIMA model
- Applied an iterative approach to reduce prediction errors

Actionable Insights

- Mean absolute percentage error = ~17%
- Predicted volume outside sample (2016-01-18) is ~330 patients

Classifying Scores by Weekday & Patients Per Day

Objective: Finding probability of good scores based on patient-per-provider and day-of-week features



Approach
• Four different classification models were fitted to the training dataset • Total running time and performance (ROC AUC) were measured on test dataset to determine the best model

Model	Total Running Time	ROC AUC using 10-fold CV
Logistic Regression	0.004	0.68
Naïve Bayes	0.002	0.64
Support Vector Classification (SVC)	0.16	0.67
Random Forest	0.03	0.62

Selected model based on performance and total running time

Putting It All Together

Final Product

Statistical Mann-Whitney U Test
Time Series Analysis
Logistic Regression Model

http://anjalibshah.github.io/Elevating-Healthcare-Experience/

How will it help the startup and elevate patients' healthcare experience?

Optimum Management

Better Experience

Happy Customers

About Me

PhD

- Biomedical Informatics (Rutgers University)

Masters and Bachelors

- Computer Science and Engineering

Professional Experience

- 10+ years of professional experience across education, healthcare, financial services, and telecom sectors

Former Employers

Goldman Sachs

TEMPLE UNIVERSITY

BNP PARIBAS

RUTGERS

My Passion

Travel and…

Scrapbooking

Appendix

Algorithm and Data Analysis Approach

Algorithm Stages and Pipeline

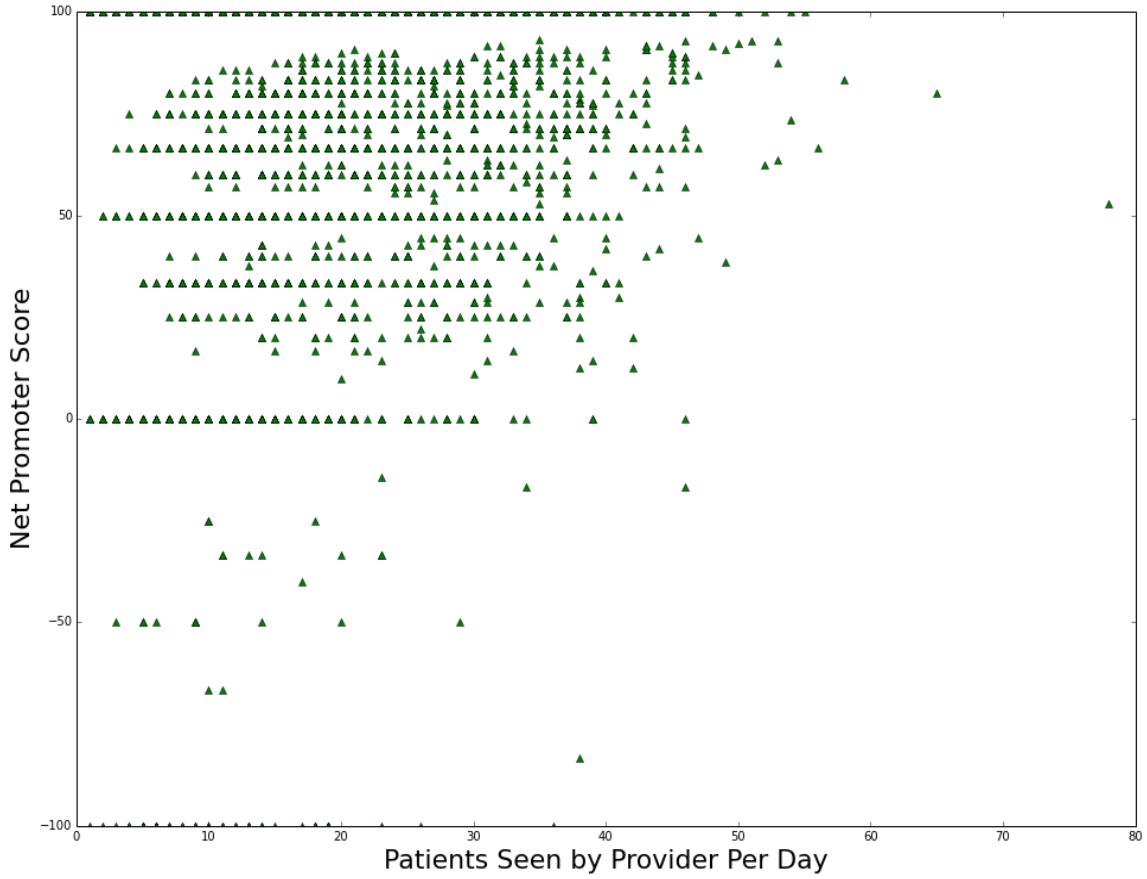
Features

- Customer ID
- Site ID
- Provider ID
- Patient ID
- Visit ID
- Date Seen
- Score

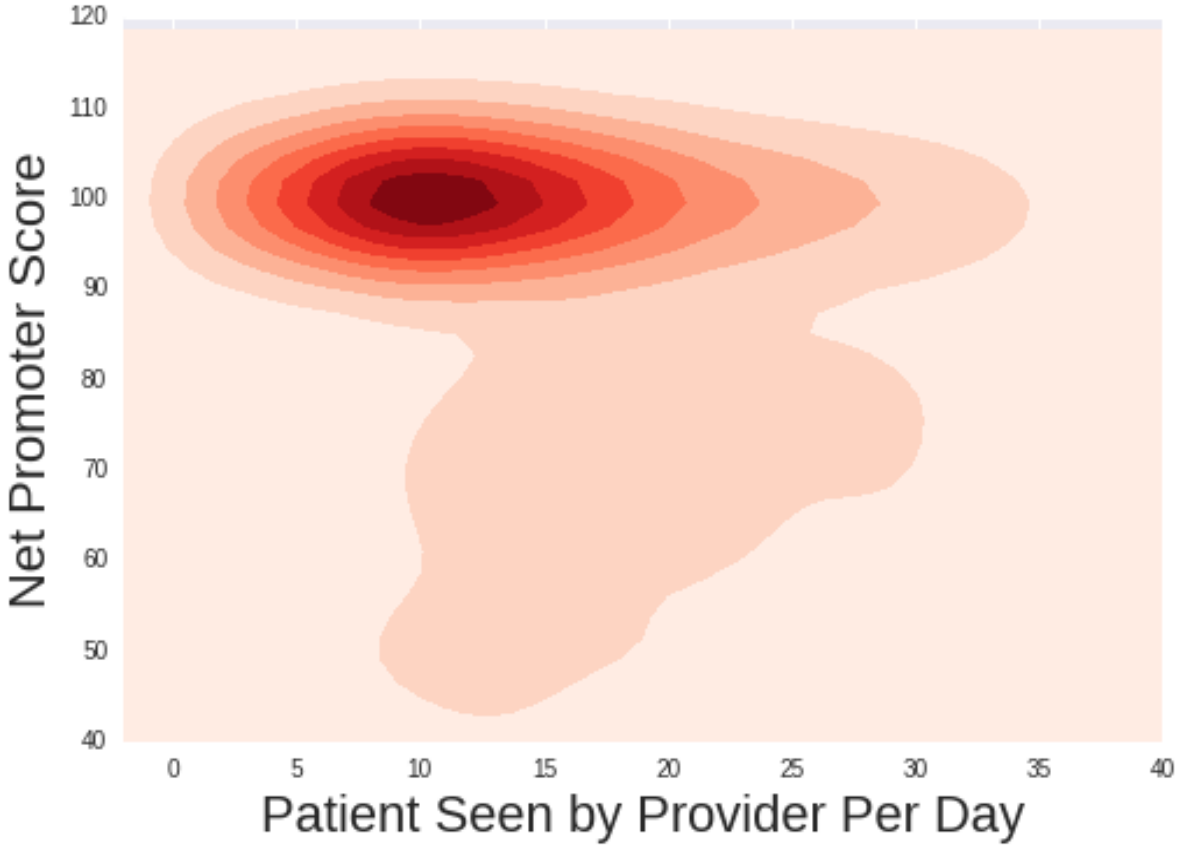
1. Patients seen by provider per day
2. Day of the week
3. Net Promoter Score (NPS)

Visualization of the Distribution

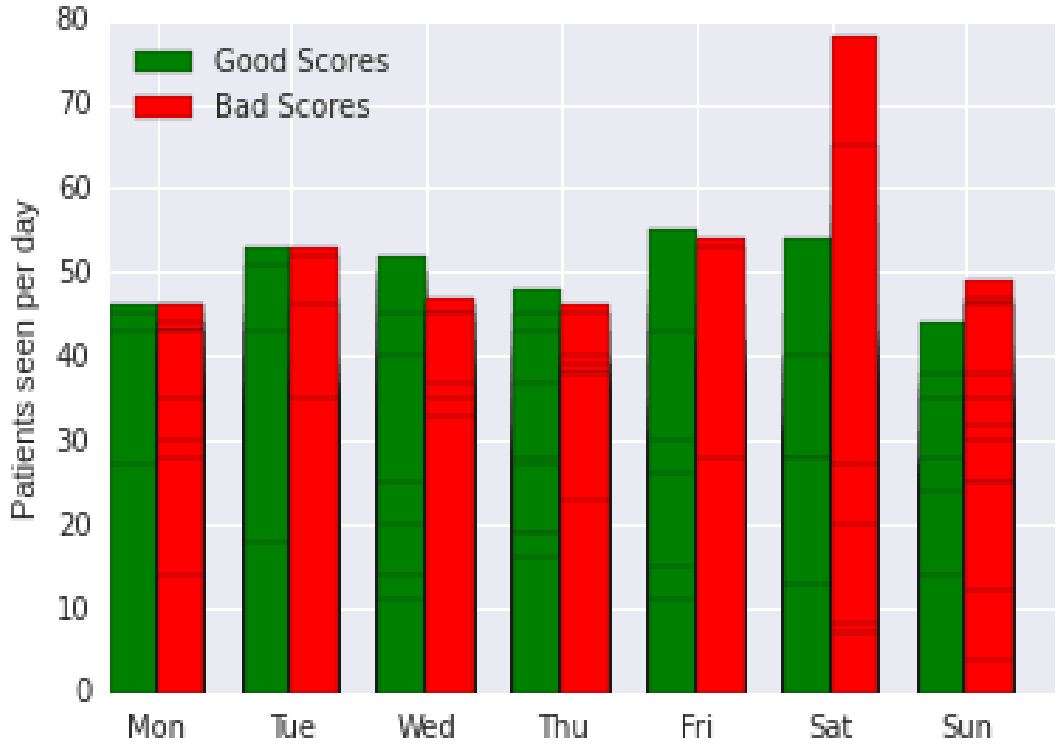
Scatterplot Distribution



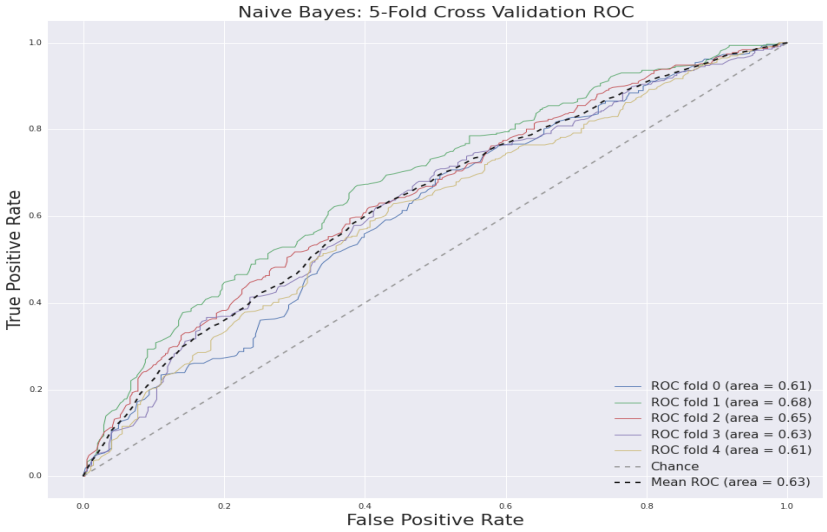
Kernel Density Estimation

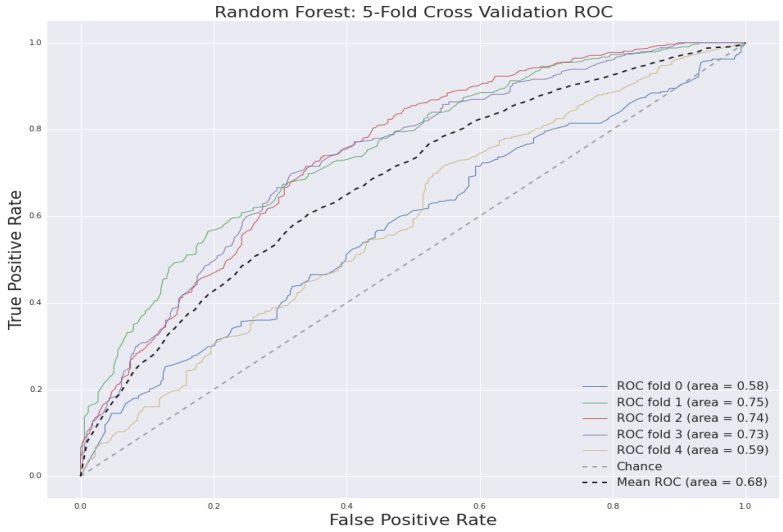


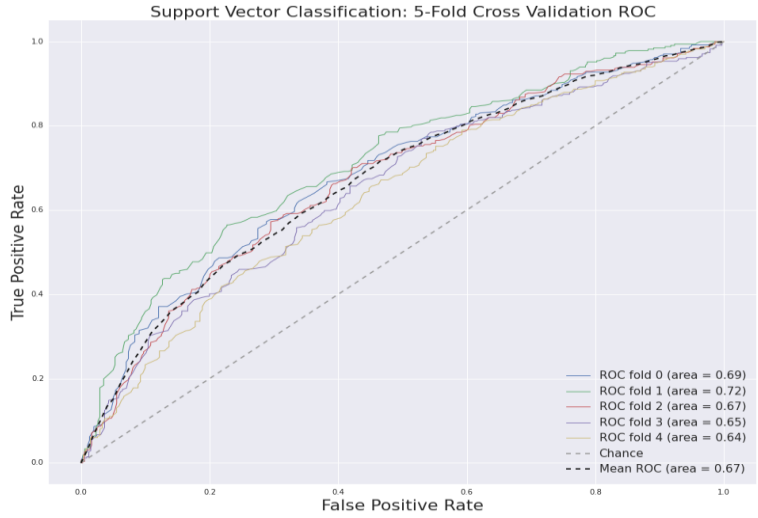
Classifying Scores by Weekday & Patients Per Day



Classifying Scores by Weekday & Patients Per Day

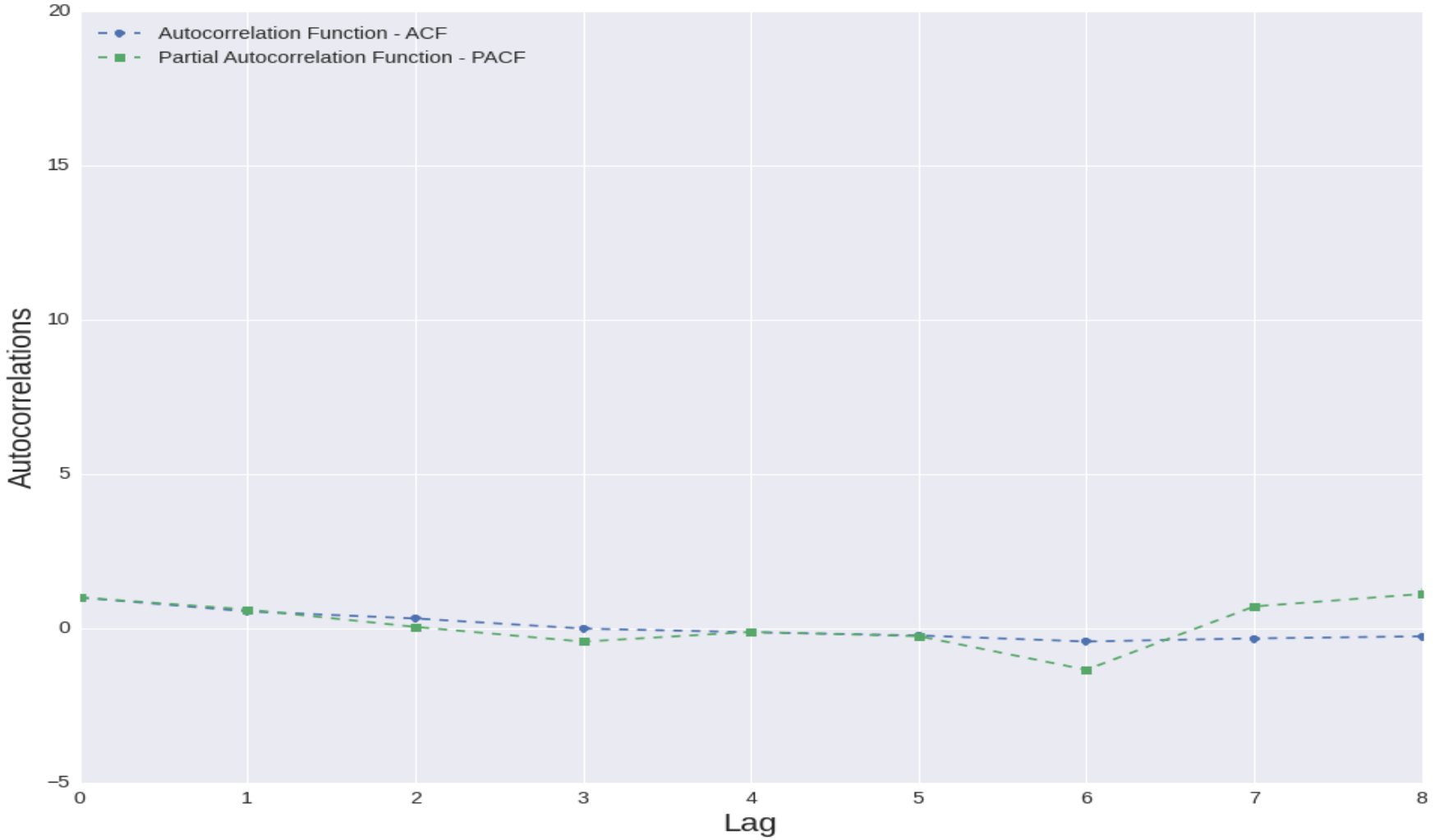




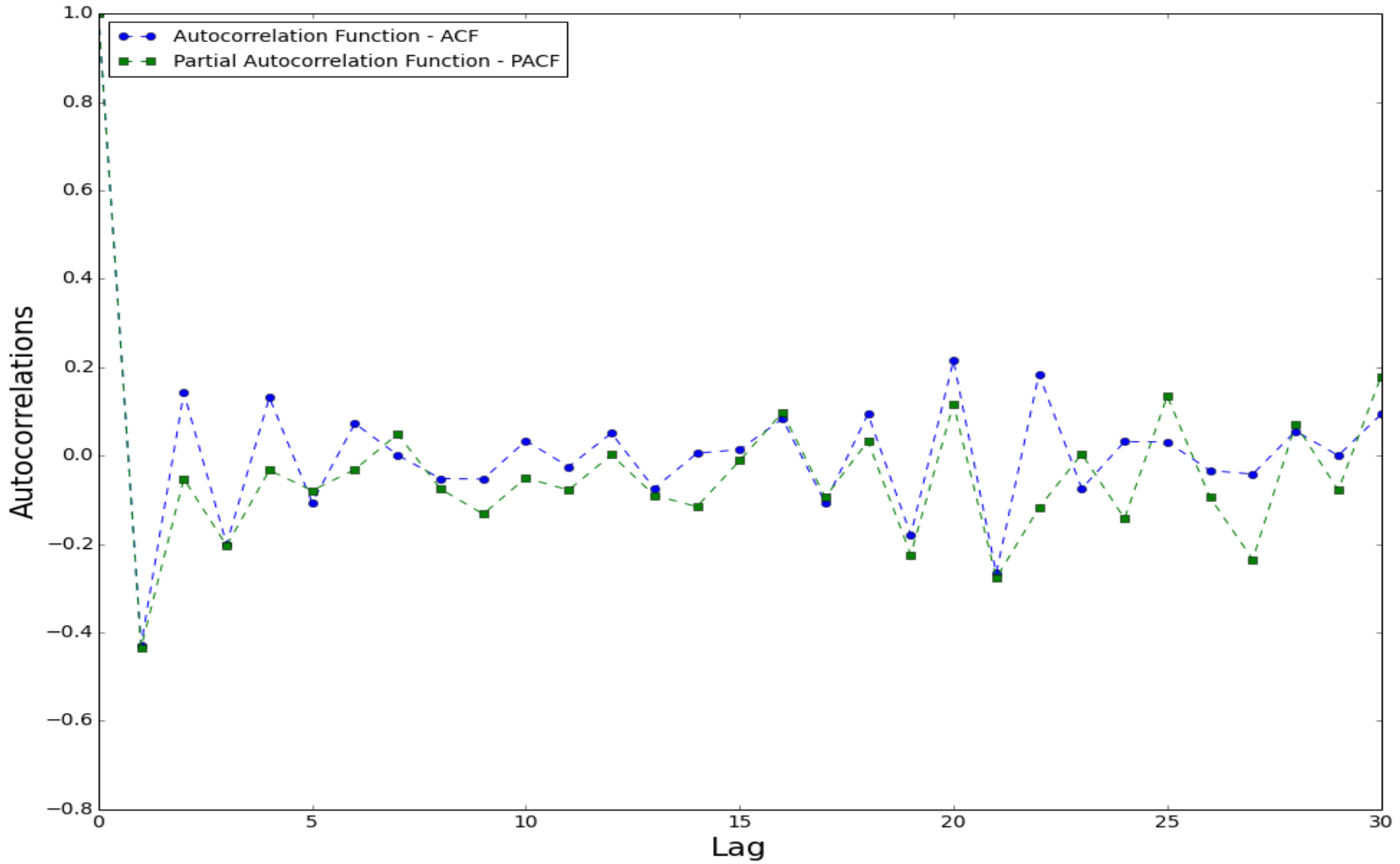


Logistic Regression Classification Summary	Precision	Recall	F1-score
Bad Score	0.60	0.45	0.52
Good Score	0.61	0.74	0.67
ROC AUC with 10-fold Cross-validation	0.68		

Plots of ACF and PACF (Monthly)



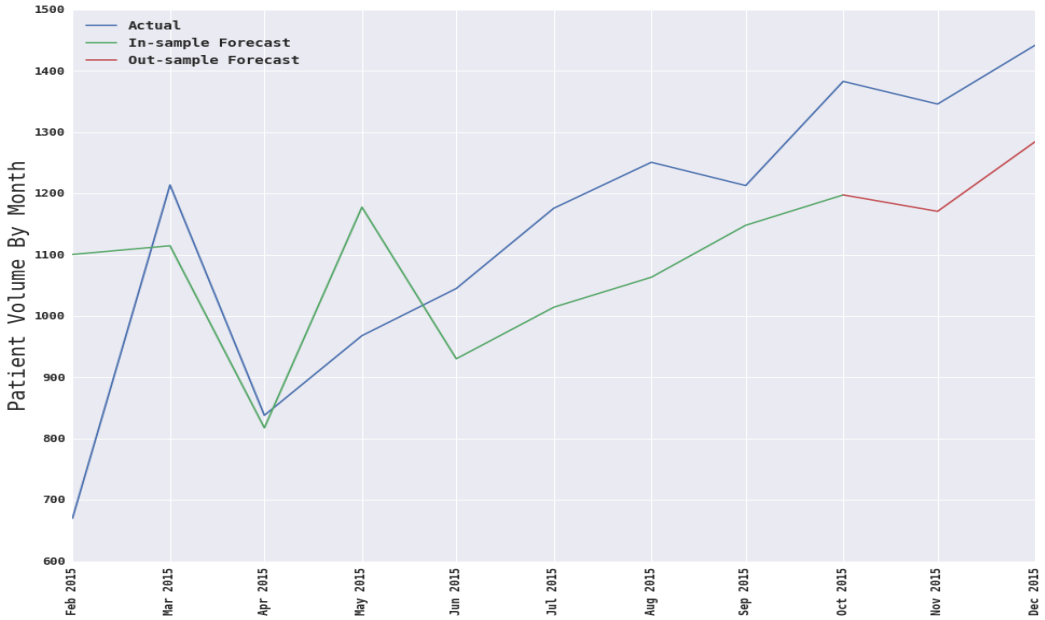
Plots of ACF and PACF (Weekly)



Forecasting Patient Volume (Monthly Data)

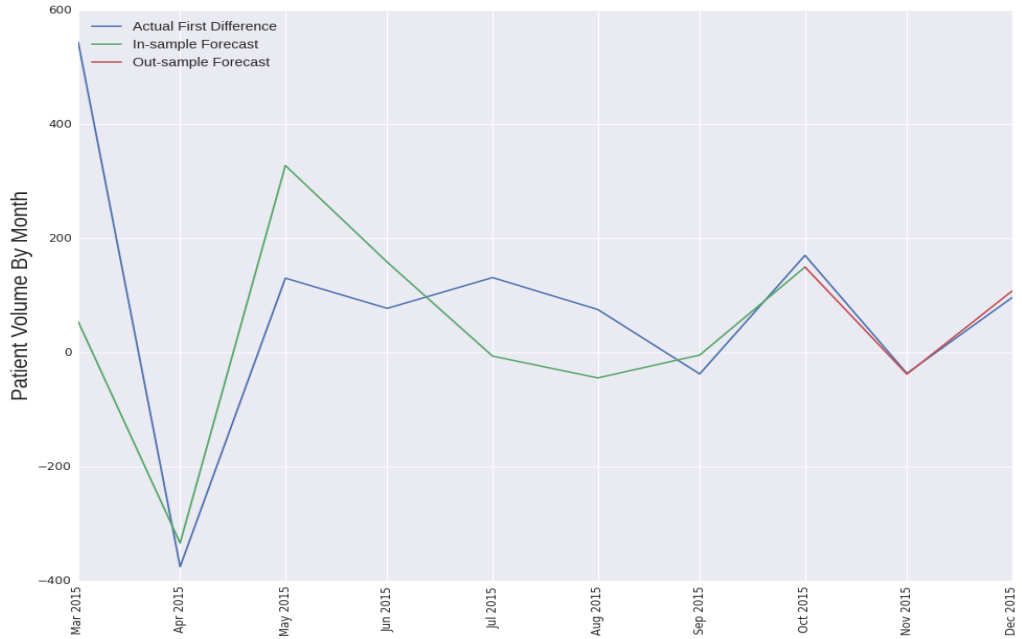
ARIMA (p=2,d=0,q=0) Model

(Non-Stationary Model)



ARIMA (p=2,d=1,q=0) Model

(Stationary Model)



Approach

- Used ACF and PACF plots to determine order of ARIMA model
- Applied an iterative approach to reduce prediction errors

Actionable Insights

- Mean absolute percentage error = ~15.5%
- Predicted volume outside sample (2016-01-31) is ~1487 patients

Forecasting Patient Volume (Suboptimal Fit)

ARIMA (1,0,0) Model for Monthly Data

ARIMA (0,0,1) Model for Weekly Data