

Using Data Analysis to Predict Attendance for NHL Regular Season Games

Brian Macdonald¹
Michael Peterson²
James Cifu¹

¹Florida Panthers Hockey Club, Sunrise, FL
²Tampa Bay Lightning Hockey Club, Tampa, FL



NESSIS, 09/23/2017

Twitter: @bmacNHL, @FlaPanthers, @TBLightning,
#NESSIS

Attendance Model



Goal:

- ▶ Develop a model for predicting attendance for games using only information that is known before tickets go on sale.



Attendance Model

Goal:

- ▶ Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- ▶ Which games should be in which tiers for variable pricing?



Attendance Model

Goal:

- ▶ Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- ▶ Which games should be in which tiers for variable pricing?
- ▶ What kinds of things could we request when the league is developing the schedule?



Attendance Model

Goal:

- ▶ Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- ▶ Which games should be in which tiers for variable pricing?
- ▶ What kinds of things could we request when the league is developing the schedule?
 - ▶ **Specific question:** Do we prefer good team on a Saturday and bad team during the week, or a good team during the week and a bad team on Saturday?"



Attendance Model

Goal:

- ▶ Develop a model for predicting attendance for games using only information that is known before tickets go on sale.

This could help answer questions like:

- ▶ Which games should be in which tiers for variable pricing?
- ▶ What kinds of things could we request when the league is developing the schedule?
 - ▶ **Specific question:** Do we prefer good team on a Saturday and bad team during the week, or a good team during the week and a bad team on Saturday?"
 - ▶ What do we want Thanksgiving week?



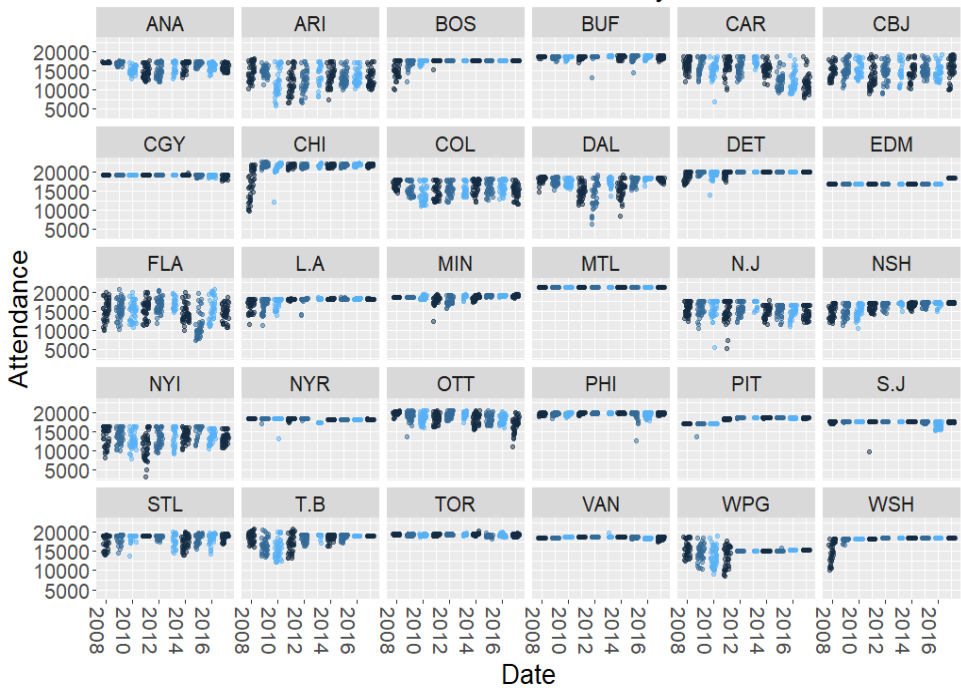
First, let's plot some raw data.
Attendance* by game, from 2007-08 to 2016-17, for all 30
teams.



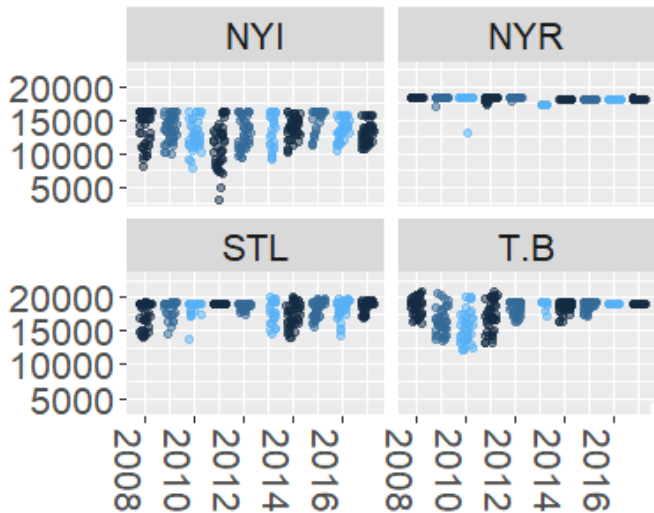
First, let's plot some raw data.
Attendance* by game, from 2007-08 to 2016-17, for all 30
teams.

*Announced attendance, as published on nhl.com

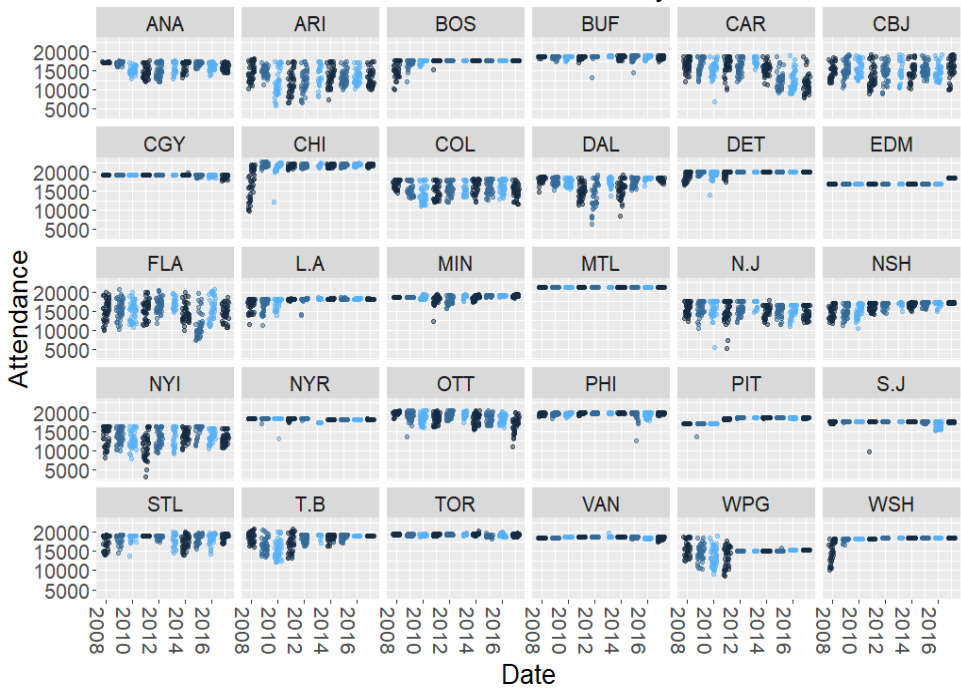
Announced Attendance from 2007-2017 by team



Snippet



Announced Attendance from 2007-2017 by team





Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
 - ▶ European games



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
 - ▶ European games
 - ▶ Blizzards



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
 - ▶ European games
 - ▶ Blizzards
4. Use several predictor variables (next slide)



Attendance Data and Model

Two observations

1. For many teams, attendance is relatively flat.
2. Winning matters. See BOS, CHI, LA, NSH, TB, WSH.

Model

1. Remove the teams that have flat attendance.
2. That leaves us with ANA, CAR, CBJ, COL, DAL, FLA, NJ, NSH, NYI, OTT, PHX, STL, and TB.
3. Remove a few games unusual characteristics.
 - ▶ European games
 - ▶ Blizzards
4. Use several predictor variables (next slide)
5. Announced attendance is outcome we're trying to predict



Predictors

- ▶ home team, away team



Predictors

- ▶ home team, away team
- ▶ day of week, month



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)
- ▶ points during previous year for home/away (lag)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)
- ▶ points during previous year for home/away (lag)
- ▶ year-to-date points relative to average for home/away.



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)
- ▶ points during previous year for home/away (lag)
- ▶ year-to-date points relative to average for home/away.
- ▶ day and month interaction (Sundays different in fall?)



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)
- ▶ points during previous year for home/away (lag)
- ▶ year-to-date points relative to average for home/away.
- ▶ day and month interaction (Sundays different in fall?)
- ▶ home team and day interaction



Predictors

- ▶ home team, away team
- ▶ day of week, month
- ▶ holiday (Columbus Day, Thanksgiving week, etc., or none.)
- ▶ season opener (Y or N)
- ▶ same division (Y or N)
- ▶ same conference (Y or N)
- ▶ points during previous year for home/away (lag)
- ▶ year-to-date points relative to average for home/away.
- ▶ day and month interaction (Sundays different in fall?)
- ▶ home team and day interaction
- ▶ home team and month interaction (snowbird months good for us?)

Interpretation of regression model results



- ▶ Impact that each of these variables have on attendance,

Interpretation of regression model results



- ▶ Impact that each of these variables have on attendance, **independent of all other variables.**

Interpretation of regression model results



- ▶ Impact that each of these variables have on attendance, **independent of all other variables.**

- ▶ For example, we find the effect of day, controlling for all of the other variables in our model

Interpretation of regression model results



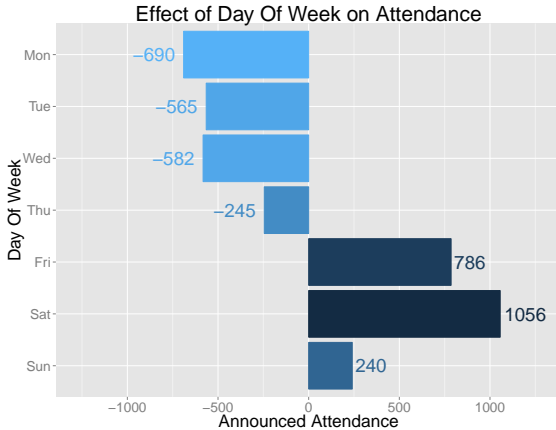
- ▶ Impact that each of these variables have on attendance, **independent of all other variables.**

- ▶ For example, we find the effect of day, controlling for all of the other variables in our model

- ▶ That's an important point. Example: If teams schedule big opponents on the weekend, then the effect of a weekend game could be overstated if we just look at day and ignore opponent.

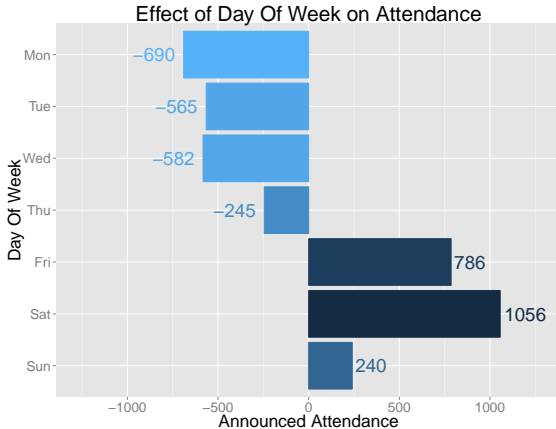


Example: day of week





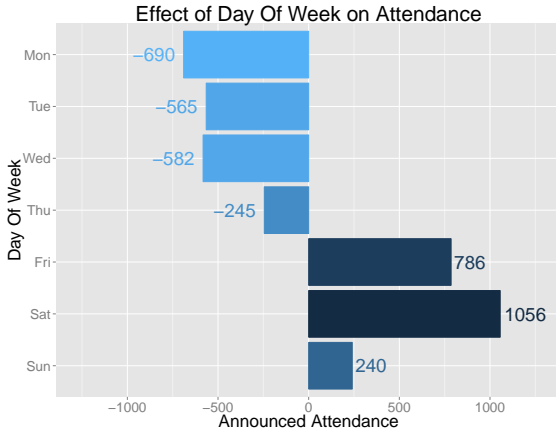
Example: day of week



1. Attendance on Saturday is expected to be 1,056 higher than average, "holding all other variables constant."



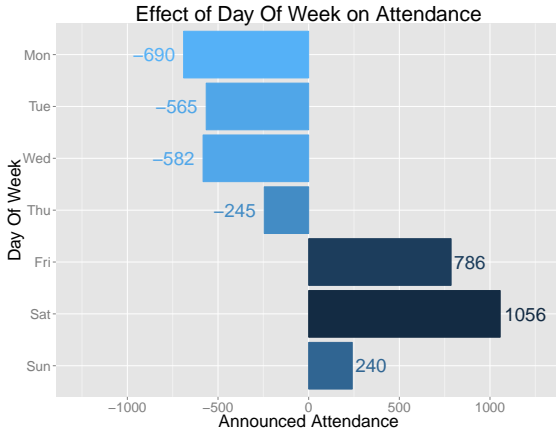
Example: day of week



1. Attendance on Saturday is expected to be 1,056 higher than average, "holding all other variables constant."
2. The difference between Saturday and Monday is expected to be 1,746 ($1,056 + 690$).

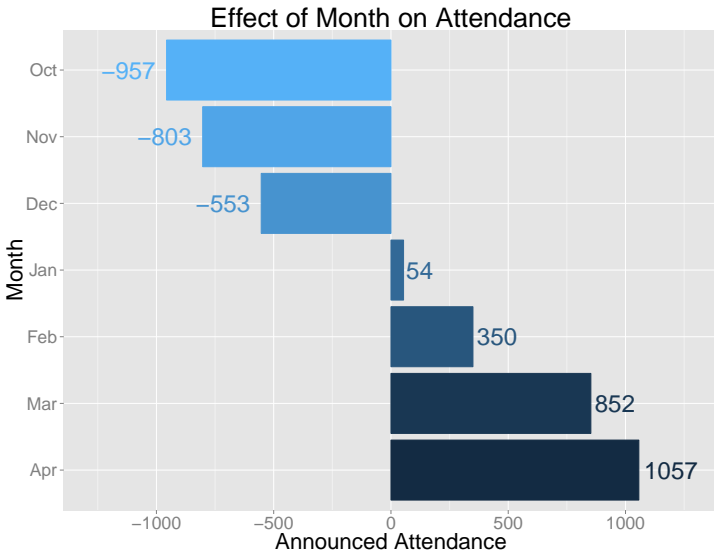


Example: day of week

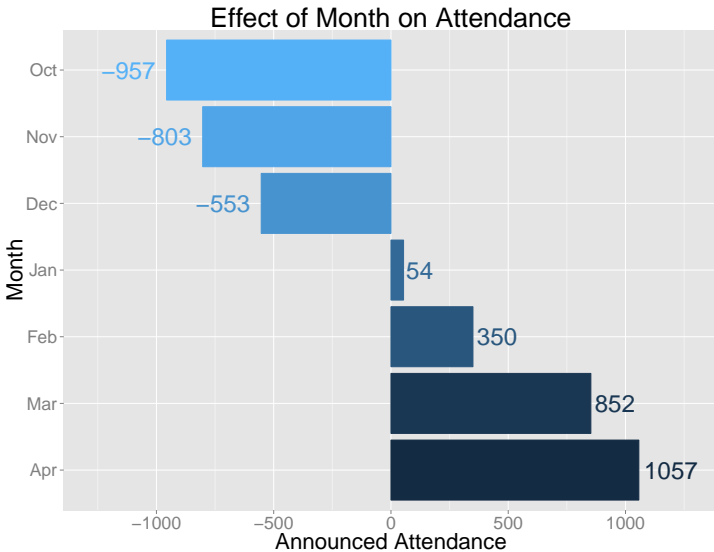


1. Attendance on Saturday is expected to be 1,056 higher than average, "holding all other variables constant."
2. The difference between Saturday and Monday is expected to be 1,746 ($1,056 + 690$).
3. Not surprising. Stuff we knew. But now we've quantified.

Month

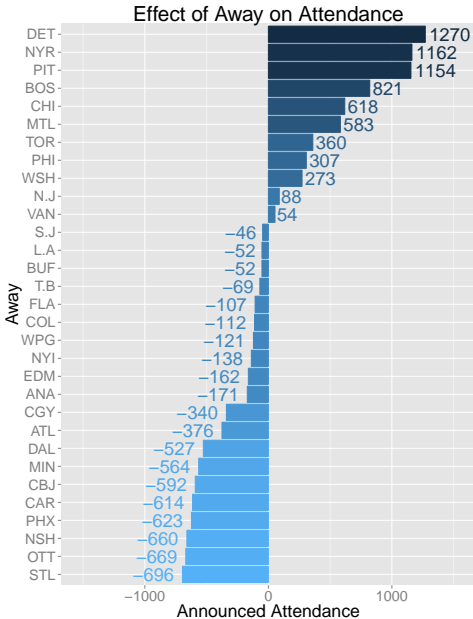


Month



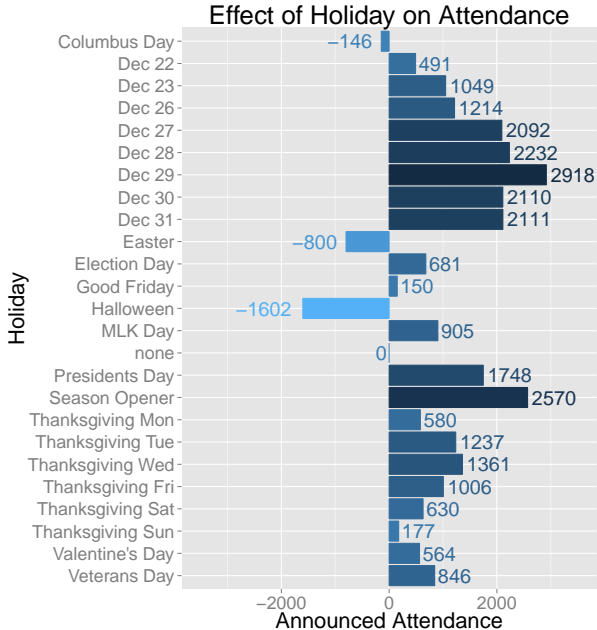
- ▶ Attendance increases over the course of the season

Away Team



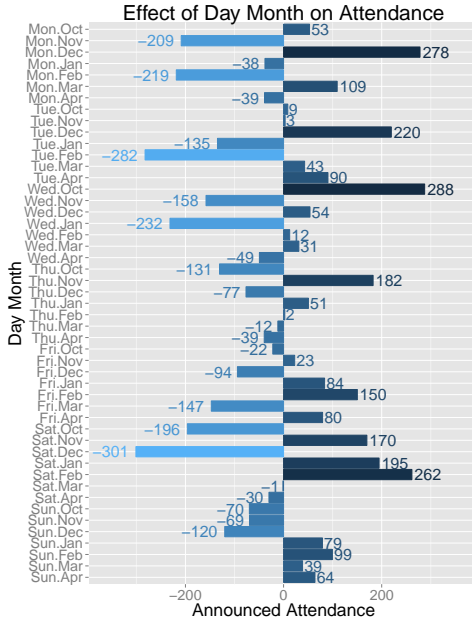


Holidays





Day-month combinations



Opponent-day combinations, Other Notes



Opponent-day:

- ▶ Good team on Sat and bad team on Tue, or good team on Tue and bad team on Sat

Opponent-day combinations, Other Notes



Opponent-day:

- ▶ Good team on Sat and bad team on Tue, or good team on Tue and bad team on Sat

- ▶ No evidence that there is a difference between these two, in terms of attendance.

Opponent-day combinations, Other Notes



Opponent-day:

- ▶ Good team on Sat and bad team on Tue, or good team on Tue and bad team on Sat
- ▶ No evidence that there is a difference between these two, in terms of attendance.
- ▶ Revenue, however, is another story.

Opponent-day combinations, Other Notes



Opponent-day:

- ▶ Good team on Sat and bad team on Tue, or good team on Tue and bad team on Sat
- ▶ No evidence that there is a difference between these two, in terms of attendance.
- ▶ Revenue, however, is another story.

Record:

- ▶ Record matters for both home team and away team.

Opponent-day combinations, Other Notes



Opponent-day:

- ▶ Good team on Sat and bad team on Tue, or good team on Tue and bad team on Sat
- ▶ No evidence that there is a difference between these two, in terms of attendance.
- ▶ Revenue, however, is another story.

Record:

- ▶ Record matters for both home team and away team.
- ▶ Last year's record matters too.

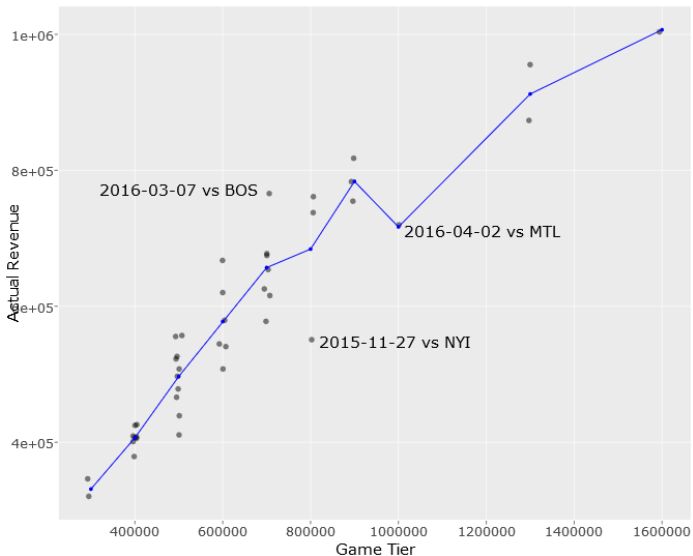
Using prediction model to tier games



Using prediction model to tier games



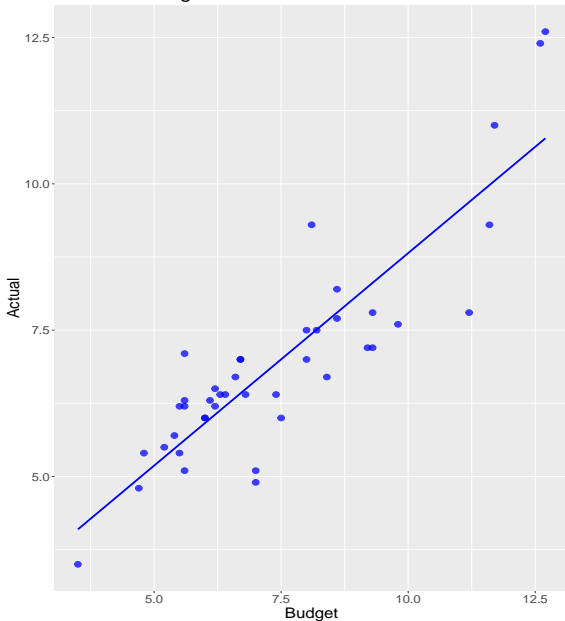
Actual Revenue vs Game Tier



Actual vs Budget for 16-17



Actual vs Budget for 16-17



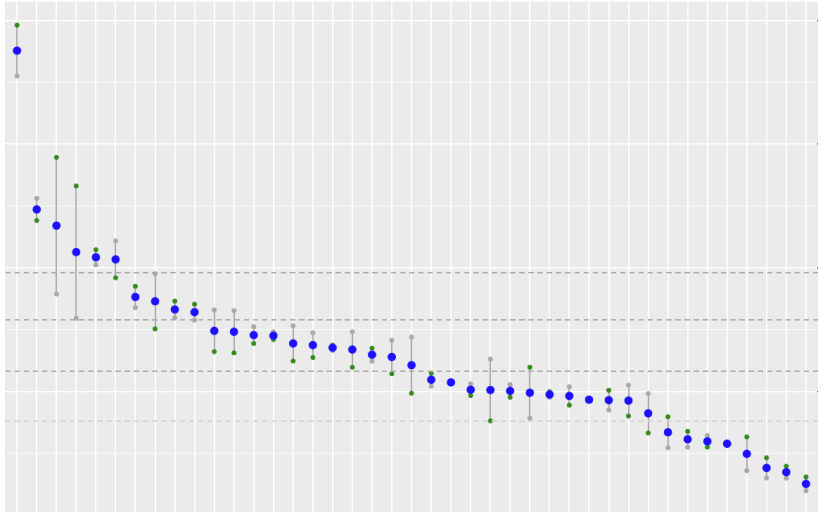


Internal data

1. Model using public data (2007-08 to 2016-17)
2. Model using internal data (only 2014-15 to 2016-17, but can use ticket prices and revenue)
3. Average



Predictions for 17-18



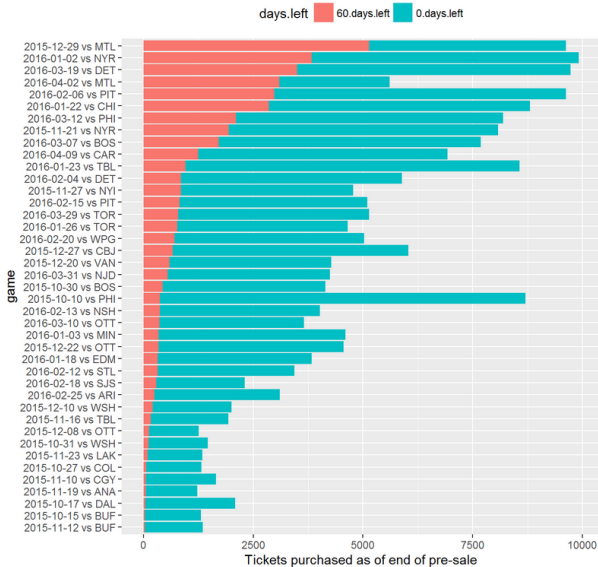
Total tickets and tickets 60 days out





Total tickets and tickets 60 days out

Tickets sold with 60 days left and total tickets, 1516



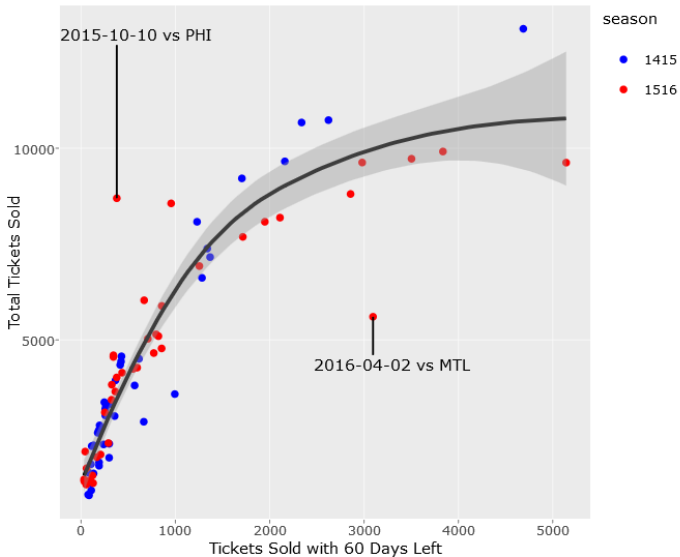
Total tickets vs tickets 60 days out





Total tickets vs tickets 60 days out

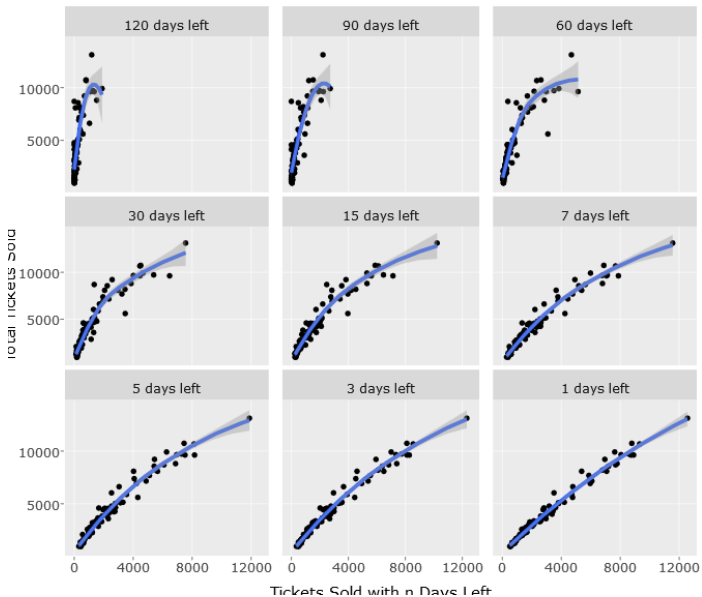
Total tickets sold vs Tickets sold with 60 days left



Similar relationship for n days out



Similar relationship for n days out



Joining two models

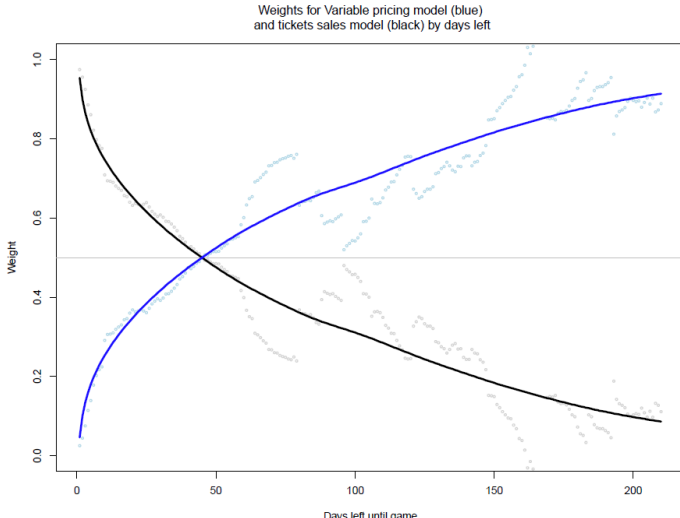


Use both models. Ticket sales model gets better as game approaches



Joining two models

Use both models. Ticket sales model gets better as game approaches





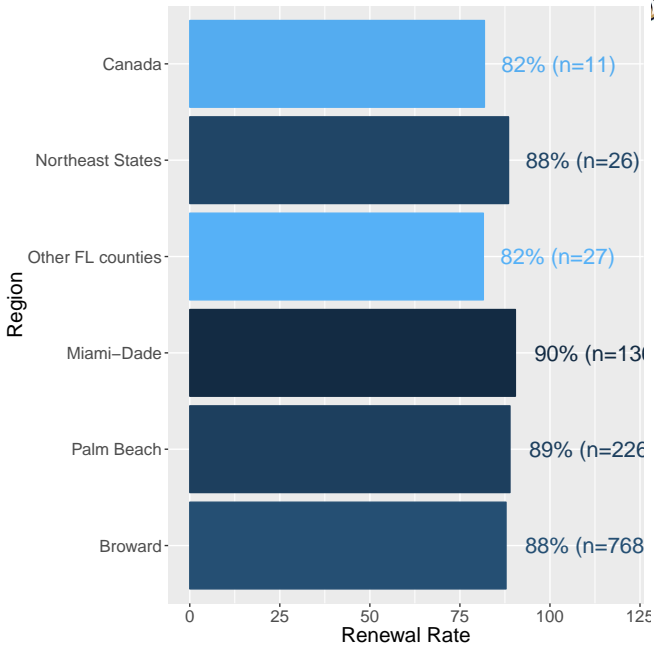
Fin.

@bmacNHL

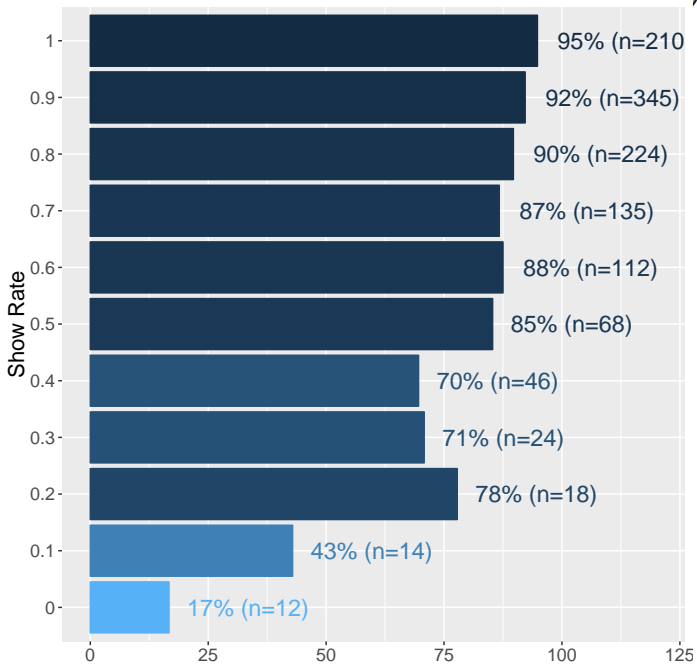
Lead scores



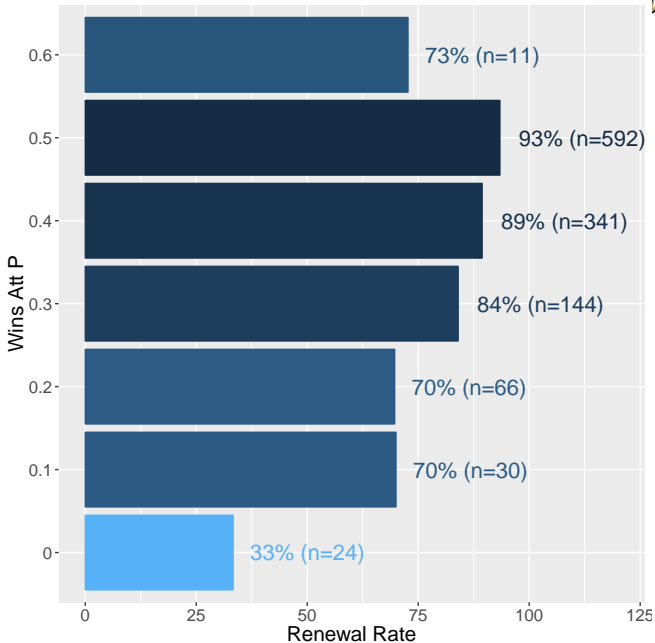
Lead scores



Show rate

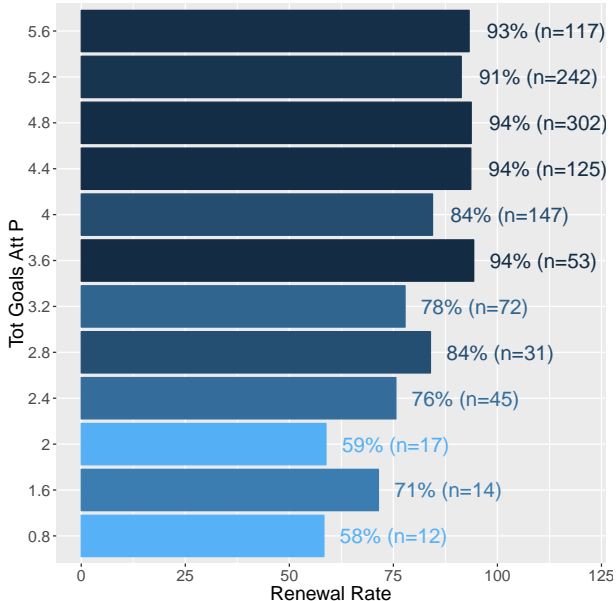


Win% in games attended





Average total goals in games attended



Proportion of 1-goal games in games attended

