Using Data Analysis to Predict Attendance for NHL Regular Season Games

Brian Macdonald\textsuperscript{1}
Michael Peterson\textsuperscript{2}
James Cifu\textsuperscript{1}

\textsuperscript{1}Florida Panthers Hockey Club, Sunrise, FL
\textsuperscript{2}Tampa 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.

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?
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?
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
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
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)
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.
Effect of Day Of Week on Attendance

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

Month

Effect of Month on Attendance

<table>
<thead>
<tr>
<th>Month</th>
<th>Announced Attendance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apr</td>
<td>1057</td>
</tr>
<tr>
<td>Mar</td>
<td>852</td>
</tr>
<tr>
<td>Feb</td>
<td>350</td>
</tr>
<tr>
<td>Jan</td>
<td>54</td>
</tr>
<tr>
<td>Dec</td>
<td>-553</td>
</tr>
<tr>
<td>Nov</td>
<td>-803</td>
</tr>
<tr>
<td>Oct</td>
<td>-957</td>
</tr>
</tbody>
</table>

Attendance increases over the course of the season
Attendance increases over the course of the season
Effect of Away on Attendance

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Away Team

Effect of Away on Attendance

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away

-1000  0  1000

Det NYR PIT BOS CHI MTL TOR PHI WSH N.J VAN S.J L.A BUF T.B FLA COL WPG NYI EDM ANA CGY ATL DAL MIN CBJ CAR PHX NSH OTT STL

Announced Attendance

Away
Holidays

Effect of Holiday on Attendance

- Columbus Day: -146
- Dec 22: 491
- Dec 23: 1049
- Dec 26: 1214
- Dec 27: 2092
- Dec 28: 2232
- Dec 29: 2918
- Dec 30: 2110
- Dec 31: 2111
- Thanksgiving Mon: 681
- Thanksgiving Tue: 150
- Thanksgiving Wed: 110
- Thanksgiving Thu: 1748
- Thanksgiving Fri: 2570
- Thanksgiving Sat: 580
- Thanksgiving Sun: 1237
- Thanksgiving Day: 1361
- Presidents Day: 1006
- None: 630
- Veterans Day: 177
- MLK Day: 564
- Valentine's Day: 846
- Election Day: -2000
- Good Friday: 0
- Columbus Day: 2000
- Season Opener: 1748

Announced 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
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.

Record:

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

game

2015-12-29 vs MTL
2016-01-02 vs NYR
2016-03-19 vs DET
2016-04-02 vs MTL
2016-02-08 vs PIT
2016-01-22 vs CHI
2016-03-12 vs PHI
2015-11-21 vs NYR
2016-03-07 vs BOS
2016-04-09 vs CAR
2016-01-23 vs TBL
2016-02-04 vs DET
2015-11-27 vs NYI
2016-02-15 vs PIT
2016-03-29 vs TOR
2016-01-26 vs TOR
2016-02-20 vs WPG
2015-12-27 vs CBJ
2015-12-20 vs VAN
2016-03-31 vs NJD
2015-10-30 vs BOS
2015-10-10 vs PHI
2015-02-13 vs NSH
2016-03-10 vs OTT
2016-01-03 vs MIN
2015-12-22 vs OTT
2015-01-18 vs EDM
2016-02-12 vs STL
2016-02-18 vs SJ S
2016-02-25 vs ARI
2015-12-10 vs WSH
2015-11-16 vs TBL
2015-12-08 vs OTT
2015-10-31 vs WSH
2015-11-23 vs LAK
2015-10-27 vs COL
2015-11-10 vs COY
2015-11-19 vs ANA
2015-10-17 vs DAL
2015-10-15 vs BUF
2015-11-12 vs BUF

Tickets purchased as of end of pre-sale

0
2500
5000
7500
10000
Total tickets vs tickets 60 days out
Total tickets vs tickets 60 days out

Total tickets sold vs Tickets sold with 60 days left

- 2015-10-10 vs PHI
- 2016-04-02 vs MTL

Season:
- 1415
- 1516
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
<table>
<thead>
<tr>
<th>Region</th>
<th>Lead scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Broward</td>
<td>88% (n=768)</td>
</tr>
<tr>
<td>Palm Beach</td>
<td>89% (n=226)</td>
</tr>
<tr>
<td>Miami−Dade</td>
<td>90% (n=136)</td>
</tr>
<tr>
<td>Other FL counties</td>
<td>82% (n=27)</td>
</tr>
<tr>
<td>Northeast States</td>
<td>88% (n=26)</td>
</tr>
<tr>
<td>Canada</td>
<td>82% (n=11)</td>
</tr>
</tbody>
</table>
Lead scores

<table>
<thead>
<tr>
<th>Region</th>
<th>Renewal Rate</th>
<th>Count (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canada</td>
<td>82%</td>
<td>(n=11)</td>
</tr>
<tr>
<td>Northeast States</td>
<td>88%</td>
<td>(n=26)</td>
</tr>
<tr>
<td>Other FL counties</td>
<td>82%</td>
<td>(n=27)</td>
</tr>
<tr>
<td>Miami–Dade</td>
<td>90%</td>
<td>(n=130)</td>
</tr>
<tr>
<td>Palm Beach</td>
<td>89%</td>
<td>(n=226)</td>
</tr>
<tr>
<td>Broward</td>
<td>88%</td>
<td>(n=768)</td>
</tr>
</tbody>
</table>
Win% in games attended

Renewal Rate

<table>
<thead>
<tr>
<th>Wins Att P</th>
<th>Wins %</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6</td>
<td>73%</td>
<td>11</td>
</tr>
<tr>
<td>0.5</td>
<td>93%</td>
<td>592</td>
</tr>
<tr>
<td>0.4</td>
<td>89%</td>
<td>341</td>
</tr>
<tr>
<td>0.3</td>
<td>84%</td>
<td>144</td>
</tr>
<tr>
<td>0.2</td>
<td>70%</td>
<td>66</td>
</tr>
<tr>
<td>0.1</td>
<td>70%</td>
<td>30</td>
</tr>
<tr>
<td>0.0</td>
<td>33%</td>
<td>24</td>
</tr>
</tbody>
</table>
Average total goals in games attended

Renewal Rate

<table>
<thead>
<tr>
<th>Tot Goals Att P</th>
<th>0.8</th>
<th>1.6</th>
<th>2</th>
<th>2.4</th>
<th>2.8</th>
<th>3.2</th>
<th>3.6</th>
<th>4</th>
<th>4.4</th>
<th>4.8</th>
<th>5.2</th>
<th>5.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>93% (n=117)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>91% (n=242)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94% (n=302)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94% (n=125)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84% (n=147)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>94% (n=53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>78% (n=72)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>84% (n=31)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>76% (n=45)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59% (n=17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>71% (n=14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>58% (n=12)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Proportion of 1-goal games in games attended

<table>
<thead>
<tr>
<th>Renewal Rate</th>
<th>Proportion of Games</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>21% (n=14)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>0.2</td>
<td>0.25</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Number of games attended: 372