

README (Project Overview)

Justin Watkins | justin.watkins@braves.com

Jan 30th 2024

Contents

1	Project Overview	2
2	Deliverables Summary	2
2.1	Deliverables Format	2
3	Materials	3
3.1	data_sales	3
3.2	data_schedule	3
3.3	prediction_data	4
3.4	Joining the data sets	4
4	Contact Information	4

1 Project Overview

The goal of this project is to build an accurate model to estimate ticket sales for a future season. Ticket sales data is wide and typically not very long. This structure may lend itself to using specific techniques.

Estimating sales figures is a fundamental component of consumer pricing. Taking an algorithmic approach to pricing products can be a rote exercise once the mechanics of the industry are well understood. Producing a top-down approach to forecasting sales allows us to create a baseline we can use to construct the more granular bottom-up pricing models. This is where this project will focus.

There are no rules in terms of what tools or materials you can use. You can use any algorithm or ensemble of algorithms to construct your model. However, you will need to document your process.

All required data sets are included and outlined in the following sections. These data sets can be enriched further or used in their present form. Be sure to make sure you understand the data before applying a model.

This project has three deliverables. You or your team will be evaluated relative to the other students or teams in terms of how closely your model matches the ticket sales outcomes contained in the *prediction_data* folder. *Absolute error* will be used to rank each team in terms of performance.

2 Deliverables Summary

This project consists of three deliverables:

1. Predictions for ticket sales for each event contained in the *prediction_data* folder.
2. A Word document explaining your work.
3. The file or spreadsheet you used to construct your model.

The project will be evaluated based on how closely your model approximates the actual outcomes for event contained in the *prediction_data* folder.

2.1 Deliverables Format

Deliverable 1: A .csv file with the event name, the ticket estimate, and your name or group. See below.

Output Format: Format.

event_name (CHAR)	tickets (NUMERIC)	name (CHAR)
EXBBxxxx	xxxxx	xxxxx

Deliverable 2: A Word document with a brief explanation of your approach. Keep this document concise and brief, you are not being evaluated on its contents. This document should include:

- What steps did you take to evaluate and prepare the data before you began to model it?
- What tools did you use to build your model? Why? Include any external packages if using a language such as Python or R.
- What model did you use to construct your forecast? Why?
- What is the $RMSE^1$ of your model on test data? Please include summary output of the model that documents basic model features such as pvalues etc.
- Please give any other context that you find important.

¹https://en.wikipedia.org/wiki/Root-mean-square_deviation

- Were there any surprises in the data?
- Did anything stand out to you after constructing your model?
- Did you enrich the data further?
- Did you have to make manual adjustments to your outcome? Why?

Deliverable 3: The file or spreadsheet that you used to build and evaluate your model.

- Any manual adjustments to the output should be documented with an explanation in deliverable 2.

3 Materials

The following sections detail the included materials you will use to construct your forecasts. The data is located in the following folders.

3.1 data_sales

This folder contains ticket sales data for several seasons. All of this data is publicly available from a variety of sources. We disclose tickets sold after each game. There are three columns in each data set:

- **event_name (CHAR):** A unique identifier for each event. (These can repeat every other year)
- **Date (DATE):** The date of the game
- **Tickets (NUMERIC):** The number of tickets sold for each game

This data will need to be joined to each table in the *data_schedule* folder.

3.2 data_schedule

This folder contains event-level metadata for each game over several seasons. All of this data is publicly available from a variety of sources. This data has been compiled for convenience and could be reconstructed independently by anyone.

- **Season (CHAR):** The year the game took place.
- **Game_Number (NUMERIC):** The game sequence number.
- **Date (DATE):** The date the game occurred.
- **Time_of_year (CHAR):** Specific meta data pertaining to the event.
- **Day (CHAR):** The day of the week (Mon-Sun).
- **effective_day (CHAR):** A modified day of the week (Mon-Sun).
- **O_Team (CHAR):** The opposing team.
- **Time_Type2 (NUMERIC):** Did the game happen in the morning, afternoon, or evening?
- **Month (CHAR):** What month did this game occur?
- **Weekend (LOGICAL):** Did this game occur on a Fri, Sat, or Sun.
- **School (LOGICAL):** Are children in school?
- **start_of_homestand (LOGICAL):** Was this the first game in a homestand?
- **end_of_homestand (LOGICAL):** Was this the last game in a homestand?
- **Opening_Day (LOGICAL):** Was this the first game of the season?
- **Holiday (LOGICAL):** Did this game occur on a holiday such as Memorial Day?
- **Fourth_of_july (LOGICAL):** Did this game occur on July 4th?
- **Final_Game (LOGICAL):** Did this game occur on the final game of the season?
- **BobbleHead (LOGICAL):** Was there a bobblehead promotion?
- **replica (LOGICAL):** Was there a replica ring promotion?
- **Day_In_Homestand (NUMERIC):** The sequence of games in a homestand.
- **star_wars (LOGICAL):** Was there a Star Wars promotion?

- **concert (LOGICAL):** Was there a concert after the game?
- **christian_concert (LOGICAL):** Was there a Christian concert after the game?
- **playoff_odds (NUMERIC):** What is the likelihood that the team will appear in the playoffs? 0-100.
- **delete (LOGICAL):** Remove this game from your data set.
- **afternoon_weekday (LOGICAL):** Was this an afternoon game on a weekday?
- **afternoon_weekend (LOGICAL):** Was this an afternoon game on the weekend?
- **Week (NUMERIC):** The week number of the year.
- **weighted_odds (NUMERIC):** Weighted playoff odds.
- **first_homestand (LOGICAL):** Did this game occur during the first homestand of the season?
- **Time_Type (FACTOR):** Expanded game time classification.

3.3 prediction_data

This folder contains the data sets you will be forecasting. They mimic the data sets contained in *data_schedule* and *data_schedule*. However, they have been slightly modified. Not every feature is included since they would be impossible to know *a priori*

3.4 Joining the data sets

The data sets can be joined on *Date*. The best method will be to use a *left join*. It will look something like this:

```
left_join(data_sales,data_schedule,by = "Date")
```

4 Contact Information

Please feel free to contact me with any questions:

Justin Watkins | justin.watkins@braves.com

Include **EMORY_MSBA_PROJECT** in the subject line of your email.