

# Modeling Assessment

## Demand Analytics

### Motivation

The goal of this assessment is to build a ticket listing and pricing forecast model. The position we are hiring for will involve similar data and analysis in a production and team-based environment.

### Data

Our provided data will be used for both training your model and testing it on holdout data. Data consists of publicly available SeatGeek ticket sales listings and mean ticket listing prices by day. Los Angeles Dodgers, Oklahoma City Dodgers (AAA), and Los Angeles Philharmonic events at Dodger Stadium, Chickasaw Bricktown Ballpark, and Walt Disney Concert Hall are included. The data were collected for all upcoming events every day from 5/20/17 through 8/31/17. The `tickets_listed` and `mean_listing_price` fields are populated with “NA” values for dates after 8/1/17; these are the holdout dates and response variables for you to predict. Fields in the dataset:

- `event_id`: Unique ID for each event
- `listing_date`: Date when the listings were collected
- `event_listing_date_id`: Unique ID for each event and listing date; the data is unique by this field
- `taxonomy`: Event classification
- `event_title`: Event title
- `event_datetime`: Datetime of event in local timezone
- `tickets_listed`: The number of tickets listed
- `mean_listing_price`: The average listing price of tickets listed
- `performer_1`: The first performer at the event
- `performer_2`: The second performer at the event (if applicable)
- `performer_3`: The third performer at the event (if applicable)
- `performer_4`: The fourth performer at the event (if applicable)
- `venue_name`: The name of the venue of the event

### Directions

Please submit the following three items:

1. A .tsv (tab separated value) file of the assessment data with missing “NA” `tickets_listed` and `mean_listing_price` records populated by your model’s predicted values. No other changes to the other original assessment data should be included. The name of the file should have a suffix with an underscore and your last name, e.g. `assessment_data_gershenfeld.tsv`.
2. The code you used to build your model and generate predicted values. R code is preferred. This code should be fully reproducible and any additional data sources you used should be included.
3. Any graphics or supporting documentation describing your model and relevant patterns within the data. This is optional if not already included and commented in the main analysis code.

You are welcome to join our provided data with other publicly available demand data. As there may be many potentially relevant external data sources and the primary focus of this assessment is on your modeling, we

do not recommend spending a significant amount of time on external data collection. If impactful, providing an example external data source and comments on others you may like to test in the future would be appropriate. Please do not share the provided data or this document with others.

## Evaluation

Your submission will be evaluated on three criteria:

1. **Accuracy** — the mean absolute percent error for both response variables in the hold-out period. A particular focus will be the change in tickets listed in the days immediately prior to Los Angeles Dodgers events and appropriately accounting for variation between events.
2. **Scalability** — the ability for your model framework to robustly handle the three different performers included here and potential capability to accurately capture demand effects across many more diverse ticketed events, performers, and venues.
3. **Interpretability** — the ability to understand the model functionality and underlying data patterns for troubleshooting and improving the model, collaborative model development, and building confidence and buy-in with your desired user and audience.