

REPORT

SUMMARY

Analysis of historical tour logs and the Lead Singer's notes identified the primary drivers of Crowd Energy. The data indicates that venues behave differently, meaning a standardized approach is ineffective. A Random Forest Regressor was trained to predict future energy levels, providing a data-driven basis for optimizing setlists, pricing, and scheduling.

VENUE - SPECIFIC FINDINGS

V_Alpha : The Holy Grounds

- Best Day - Sunday
- Best Moon Phase - Waxing Gibbous
- Best Weather - Cloudy
- Best Outfit - Leather
- Openers dont affect energy by much
- Best Time - 12:00 Noon

V_Beta: The Vampire's Den

- Best Day - Sunday
- Best Moon Phase - Waning Gibbous
- Best Weather - Rainy
- Best Outfit - Denim
- Openers dont affect energy by much
- Best Time - 23:00
- At night it has the highest energy

V_Gamma: The Snob Pit

- Best Day - Saturday
- Best Moon Phase - Waxing Gibbous
- Best Weather - All are pretty same but Rainy
- Best Outfit - Spandex/Leather
- Energy highly affected by openers
- Best Time - 15:00
- People here see expensive tickets as exclusive
- Energy not much affected by volume level

V_Delta: The Mosh Pit

- Best Day - Sunday
- Best Moon Phase - Last Quarter
- Best Weather - Cloudy
- Best Outfit - Denim
- Openers dont affect energy by much
- Best Time - 16:00
- Avg crowd energy is not much affected by ticket prices
- As volume increases , Energy increases

TESTING SINGER'S THEORY

- Tuesdays are a curse - WRONG
- Goths only come alive at night - CORRECT
- Full Moon = Magic - PARTIALLY CORRECT
- Rain **definitely** sucks - WRONG
- .The Spandex always feel right - True for Gamma and Delta -PARTIALLY CORRECT
- .Mosh(Delta) kids dont care about opener -CORRECT
- Snobs(Gamma) might care about openers - CORRECT
- Time of day Matters - CORRECT

MODEL CHOICE JUSTIFICATION

Linear Regression, Gradient Boosting, and Random Forest were evaluated to determine the most effective predictive model. Linear Regression was established as a baseline but proved unsuitable due to the dataset's inherent non-linear relationships. Consequently, ensemble methods were prioritized. Comparative analysis confirmed that the Random Forest Regressor delivered superior performance, achieving the lowest Root Mean Squared Error (RMSE) on the test set.

To further enhance predictive accuracy, the model underwent hyperparameter tuning using GridSearchCV. A 5-fold cross-validation strategy was employed to systematically test various combinations of parameters—specifically `n_estimators`, `max_depth`, and `min_samples_split`.