Ralph Lauren Sales Prediction

Data Science Institute
COLUMBIA UNIVERSITY

Qi Feng, Yawen Han, Jiaying Jin, Ruizhi Zhang
Mentors: Rohit Cherian (RL), Ami Pascale (RL), Sining Chen (DSI)

Data Science Capstone
Project with
RALPH LAUREN

Background and Objective

Are external factors such as events able to affect long-term trends in sales? Answering this question is crucial for hindsighting and determining future outlook of the fashion store. To explore the impact of events on sales, major event features were engineered to predict sales at Orlando FOA store from April, Fiscal Year 2018 to September, Fiscal Year 2019. Key insights are provided for future empirical research and development of sustainable competitive marketing strategies.

Events Data Exploration: Web Application

Events data were retrieved and processed from *PredictHQ.* Major features include category, rank, venue (distance, capacity) and date (duration, periodicity, time). A Bokeh web application was developed for self-service exploration.
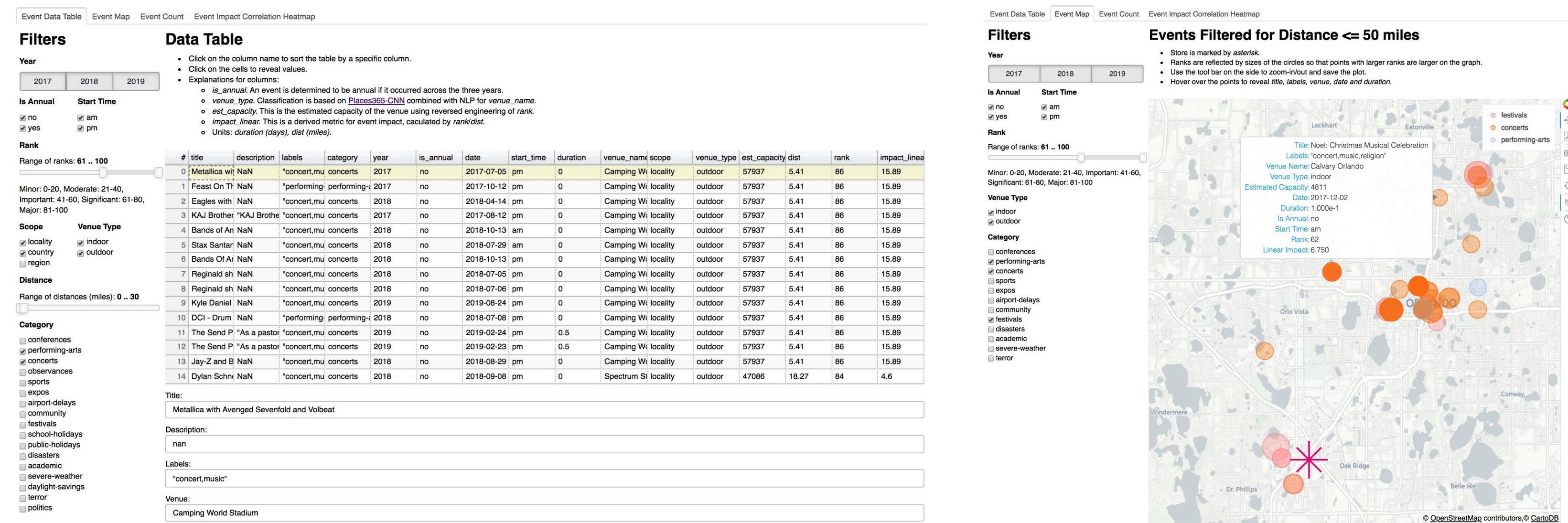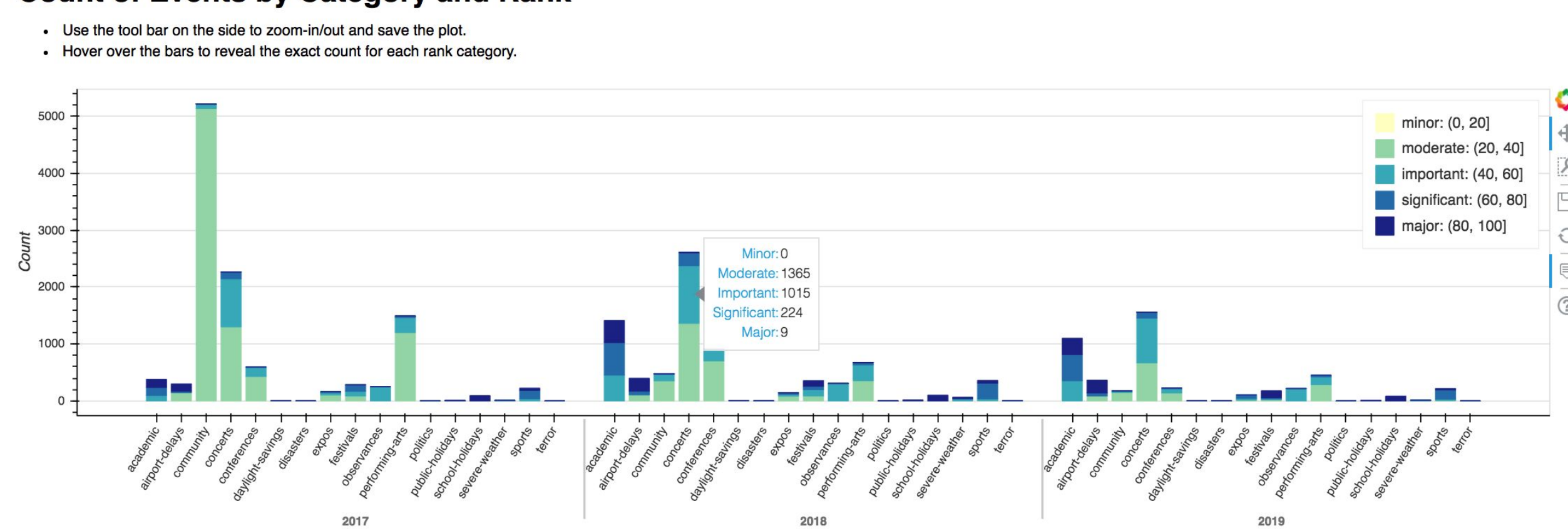

Figure 1. Web Application: Event Data Table, Map and Count Bar Plots

Sales Data Cleaning: Noise Patterns Removal in Time Series

Noise patterns (weekly, yearly, holidays & promotions) were removed to eliminate undesirable impacts on sales other than events using *Automatic Forecasting Algorithm*.
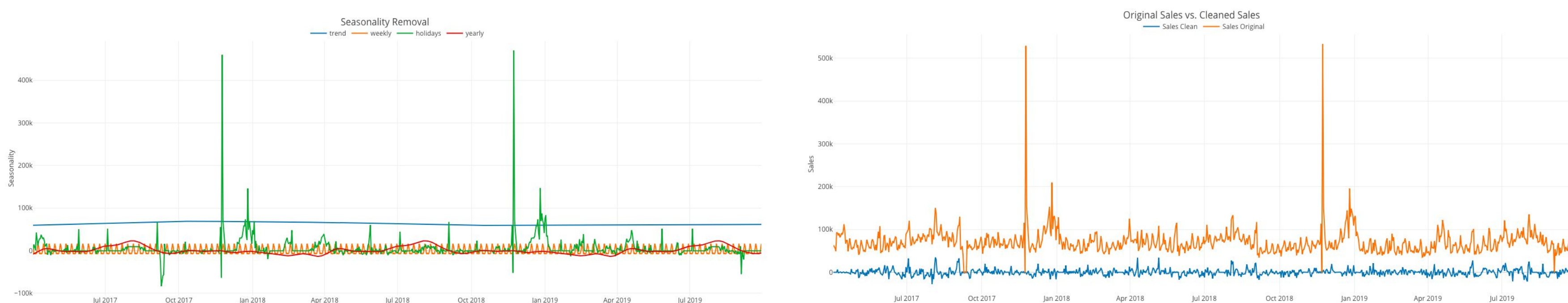

Figure 2.   (a) Noise Patterns   (b) Original Sales vs. Cleaned Sales

Modeling

A linear impact feature was derived using the formula $\sum_{j \in C} \sum_{i \in E} \frac{(rank)_{ij}}{(distance)_{ij}}$, where C is the category and E is the event. The impact feature was then feeded into the models.

1. Feature Pre-screening

- **Multicollinearity check:** No significant inter-associations were observed.
- **Univariate F-test:** *Expos* and *festivals* appeared to be the most influential categories.
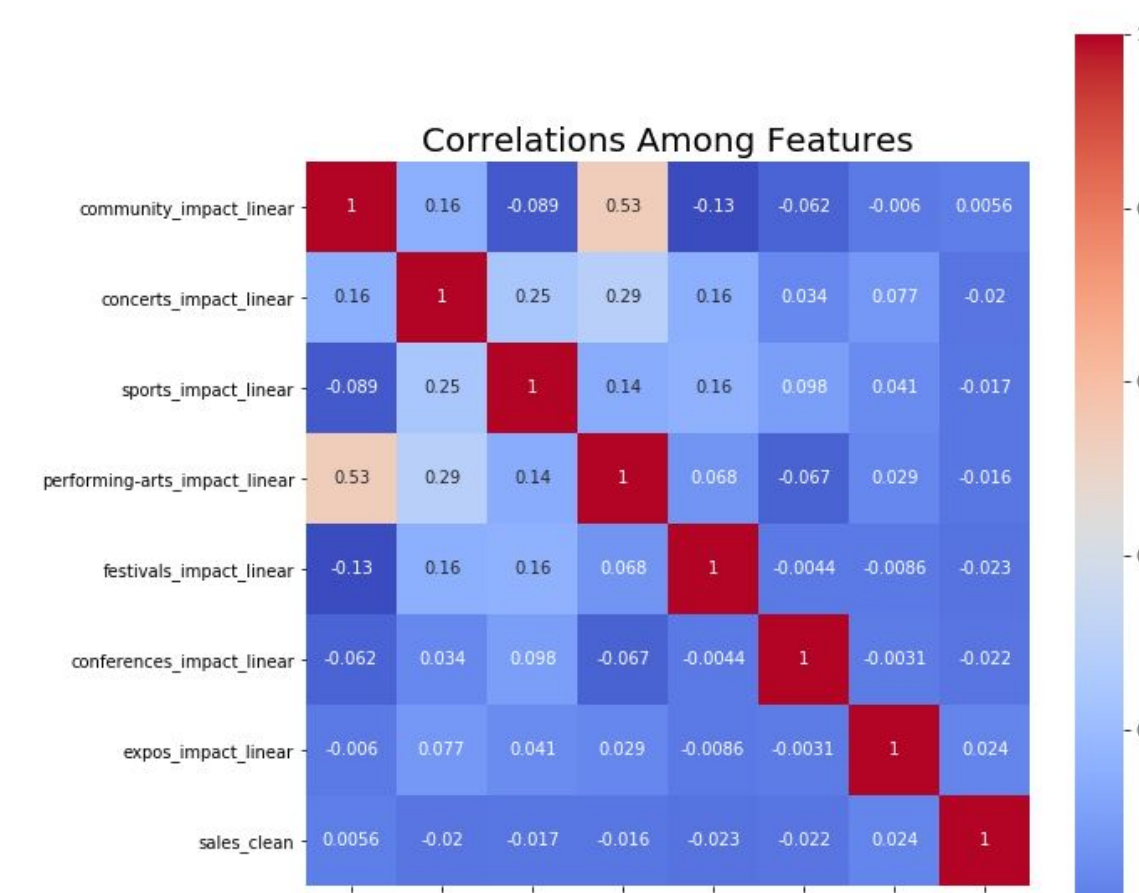

	feature	F-score	p-value
6	expos_impact_linear	0.528210	0.467548
4	festivals_impact_linear	0.467840	0.494157
5	conferences_impact_linear	0.427650	0.513310
1	concerts_impact_linear	0.360024	0.548643
2	sports_impact_linear	0.259874	0.610330
3	performing-arts_impact_linear	0.228434	0.632803
0	community_impact_linear	0.028356	0.866312

Table 1. F-test Result (sorted in descending F-score)

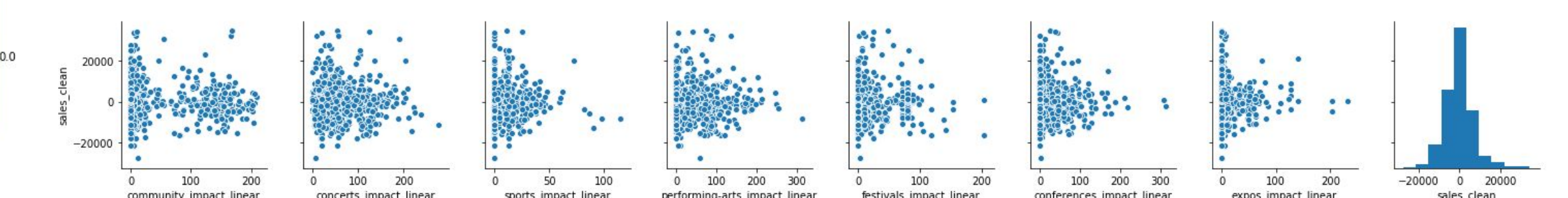**Figure 3. Correlation Heatmap   Figure 4. Pairplot for Sales vs. Event Category Impacts**

2. Regression

- **Random Forest** outperformed other regressors such as Gradient Boosting.
- **R-squared = 0.43, with a test set accuracy of 92.18%.**
- **Most important event categories:** *sports, community and performing-arts.*
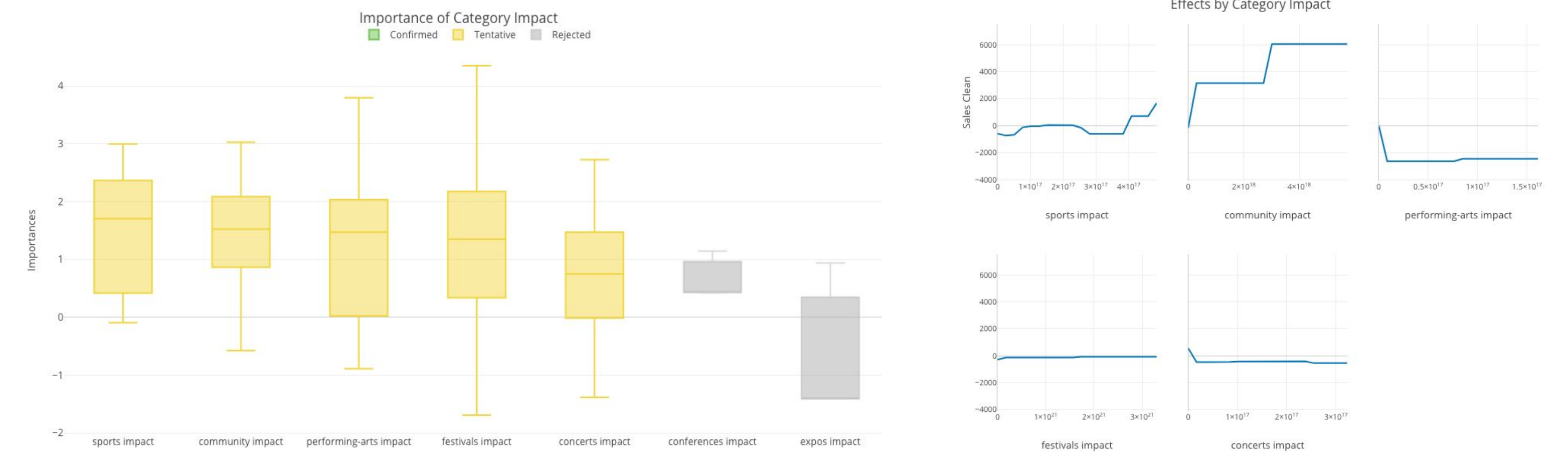

Figure 5. (a) Importance of Event Category Impacts   (b) Sales vs. Event Category Impact

Conclusion

The impact of events on sales is limited at Orlando FOA. We recommend to scale up the analysis to other stores to re-examine the effect of events, or switch the focus to other external factors like news to direct marketing strategies.

References

Sean J. Taylor, Benjamin Letham (2018) Forecasting at scale. The American Statistician 72(1):37-45. https://peerj.com/preprints/3190.pdf.