Data Analytics for Resort Revenue Management

Aliaksandr Nekrashevich, Yuri Levin, Guang Li, Mikhail Nediak

Business Context

- Golf Resort: **hotel, golf, pub, bar, shop**





- Regular Revenues (**certain**) + Ancillary Revenues (**uncertain**)
- Open from **May** to **October**
- Price categories are **fixed** in advance:

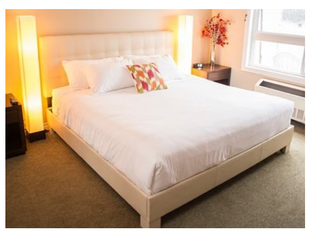
$$early = late \leq spring = fall \leq middle \leq peak$$

May → October

early	spring	middle	peak	middle	fall	late

Motivation

- A guest books a **single hotel night**
- Maximum price is **$3K**, minimum is **$250**
- Then spends **$5K** for a bottle of wine









- **Who are such clients?** The resort likes them!
- Opportunity: **10 guests**, **5 days** ⟹ **$250K** within a **week**!

Case Study

1. Inputs: prices, capacities, individual ancillary spending

2. Output: RM system

3. Decision-Making Mechanism: smart personalized bid-pricing

Datasets:

- golf resort (confidential), 3 years
- resort & hotel (public), 3 years

Restrictions:

- free upgrades, no downgrades
- prices are fixed prior to the beginning of the season





Reference:

Antonio N., de Almeida A., Nunes L. (2019) Hotel booking demand datasets. Data in Brief 22:41 – 49, ISSN 2352-3409

Current Practice

- FCFS (first come, first served) – benchmark
- **Pros**: easy to implement, intuitive
- **Cons**: no personalization, doesn't account for ancillary revenues, cancellations, no-shows
- **Opportunities**: ancillary profits, customer-centric view, personalized decisions, free upgrades



Randomized LP:

- **computationally** tractable
- induced bid prices are easy to **interpret**
- better than deterministic LP as it captures the **entire distributions of RVs**
- **easy to show** on sampled and historical trajectories



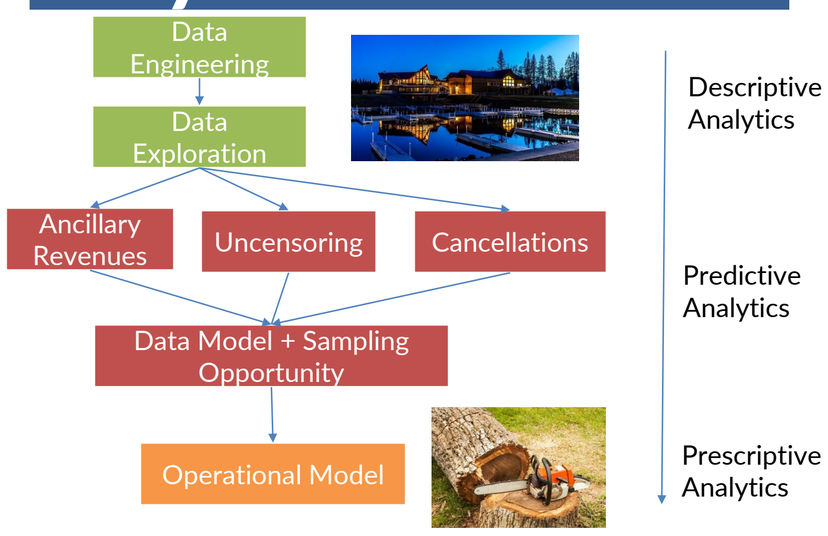
Reference:

Talluri K., Van Ryzin G. (1999) A randomized linear programming method for computing network bid prices. Transportation Science 33(2):207–216

Punchline

1. Real data, 3-year booking period, ancillary values, free upgrades
2. Complete Revenue Management System prototype
3. Profit lift from 1.3% to 14.1 %
4. Personalized booking decisions

System Overview



Ancillary Value

Algorithm	Target	RMSE	MAE	Samples Over	Avg Prediction
Ord. Lst. Sqrs.	id	566.153	388.682	0.594	448.715
Ridge Regr.	id	566.126	388.667	0.594	448.69
Ridge Regr.	log	703.862	414.981	0.382	72.197
Lasso Regr.	id	565.589	388.258	0.605	448.652
Lasso Regr.	log	726.728	426.625	0.382	53.803
Orth. Match. Purs.	id	570.682	394.731	0.623	448.891
Orth. Match. Purs.	log	706.987	417.286	0.38	68.003
Supp. Vect. Mach.	id	585.164	371.252	0.494	309.157
Supp. Vect. Mach.	log	615.838	385.463	0.498	249.145
XGBoost Regr.	id	552.295	368.838	0.613	450.662
XGBoost Regr.	log	638.291	356.352	0.476	228.765
Neural Network	id	560.371	380.824	0.588	441.375
Neural Network	log	704.383	413.607	0.383	70.982
Ada Boosting	id	607.316	451.843	0.713	600.041
Ada Boosting	log	719.472	422.714	0.378	56.492
Gradient Boosting	id	555.643	376.097	0.614	447.872
Gradient Boosting	log	681.726	402.568	0.389	98.017
Random Forest	id	556.526	372.668	0.623	447.471
Random Forest	log	643.986	355.012	0.455	197.978
					Avg Test: 443.651

Cancellations

Whether Cancellation Happens

Algorithm	Accuracy
LogisticRegression-l2	0.782
LogisticRegression-l1	0.782
LogisticRegression-elnet	0.782
kNN	0.807
GaussianNaiveBayes	0.753
QDA	0.761
LDA	0.781
RandomForest-50	0.831
AdaBoosting	0.784
Gradient Boosting	0.787

Duration Before Cancellation

Algorithm	MAE	MSE
kNN	27.836	3757.968
OLS	48.371	4191.695
Lasso	48.378	4192.015
Ridge	48.381	4191.076
Random Forest[mae]	28.748	2499.991
Random Forest[mse]	27.419	2645.419
AdaBoosting	47.948	4055.869
Gradient Boosting	44.593	3671.713

Uncensoring

Notation

λ_{dw}- probability of arrival of demand on period d to stay at week w, Bernoully RV

p_{scl}- probability that stay starts at weekday s with room type c for duration l days

a^t-probability that arrival happened at uncertain situation observed in period t

Incomplete Information Log-Likelihood

$$LLI = \sum_{(d,w)\in Arr,(s,c,l)=Dest(d,w)} (\log \lambda_{dw} + \log p_{scl}) + \sum_{(d,w)\in NArr} \log(1-\lambda_{dw}) + \sum_{(d,w)\in QArr} \log(1-\lambda_{dw}+\lambda_{dw} \sum_{full(d,w)} p_{scl})$$

Complete Information Log-Likelihood

$$LLC = \sum_{(d,w)\in Arr,(s,c,l)=Dest(d,w)} (\log \lambda_{dw} + \log p_{scl}) + \sum_{(d,w)\in NArr} \log(1-\lambda_{dw}) + \sum_{(d,w)\in QArr} (1-a^t)\log(1-\lambda_{dw}) + a^t \log(\lambda_{dw} \sum_{full(d,w)} p_{scl})$$

EM Algorithm (iterate)

1. Given λ, p, recompute a^t.
2. Given a^t, find optimal λ, p by convex optimization of complete information log-likelihood

Extensions

1. Parametric intensity by supposing $\lambda = Sigmoid(w_1 f_1 + \cdots + w_n f_n)$, where w – weights, f – features
2. Sequential choice, locally and globally.
   - Allows multiple parameterizations
   - Various ways to state the complete information log-likelihood
   - For scalable situations, requires permutation decompositions
3. Natural opportunity to compare with alternating optimization

Reference:

- Garrett van Ryzin, Gustavo Vulcano (2017) An expectation-maximization method to estimate a rank-based choice model of demand. Operations Research 65(2):396–407

Results – Private, GLM

Ancillary Value Model	Ratio 2016	Ratio 2017	Ratio 2018
Gradient Boosting – Log	1.024	1.019	1.033
Random Forest	1.032	1.027	1.04
Ridge Regression	1.03	1.025	1.035
Extreme Gradient Boosting	1.035	1.03	1.041
Adaptive Boosting – Log	1.023	1.019	1.032
LASSO Regression – Log	1.023	1.016	1.031
Neural Network – Log	1.023	1.019	1.033
Extreme Gradient Boosting – Log	1.03	1.021	1.034
Ridge Regression – Log	1.024	1.018	1.032
Neural Network	1.03	1.027	1.037
Gradient Boosting	1.032	1.028	1.039
Random Forest	1.027	1.019	1.034
LASSO Regression	1.029	1.025	1.036
Adaptive Boosting	1.033	1.03	1.039

Uncensoring Baseline – Generalized Linear Models

- Log at suffix tells that log-transformed targets were learned and predicted
- The ratio describes the performance of our approach relative to the FCFS (first come, first served) policy
- Optimistic performance is highlighted in yellow, and pessimistic is highlighted in red

Results – Private, EM

Intensity Form	Intensity Stratification	Split Stratification	Ratio 2016	Ratio 2017	Ratio 2018
Constant	Weekday-Common	Common-Season	1.029	1.041	1.024
Constant	Common-Season	Common-Common	1.037	1.034	1.028
Constant	Common-Common	Common-Season	1.031	1.033	1.036
Linear Sigmoid	Common-Common	Common-Common	1.034	1.023	1.031
Linear Sigmoid	Common-Season	Common-Season	1.035	1.04	1.028
Constant	Common-Season	Common-Season	1.03	1.026	1.034
Linear Sigmoid	Weekday-Season	Common-Common	1.022	1.033	1.023
Linear Sigmoid	Weekday-Common	Common-Common	1.024	1.031	1.021
Linear Sigmoid	Common-Season	Common-Common	1.033	1.024	1.023
Constant	Weekday-Season	Common-Common	1.032	1.029	1.027
Linear Sigmoid	Weekday-Season	Common-Season	1.031	1.032	1.027
Constant	Weekday-Season	Common-Season	1.035	1.039	1.029
Constant	Common-Common	Common-Common	1.03	1.023	1.02
Constant	Weekday-Common	Common-Common	1.021	1.025	1.022
Linear Sigmoid	Common-Common	Common-Season	1.022	1.035	1.029
Linear Sigmoid	Weekday-Common	Common-Season	1.037	1.035	1.029

1. Intensity describes the probability that some client arrives at each booking period
2. Split probabilities correspond to the reserved product (arrival, departure, service level)
3. Intensity is modeled as a fixed constant parameter (Constant Intensity) or linear function of period features with logistic activation (Linear Sigmoid)
4. Stratification describes the number of different intensities or splits correspondingly. For Weekday-Season stratification, parameters are different for each reservation weekday and between arrival seasons. Common stratification = No stratification.

Results – Semi-Real

Year	Ancillary	Target	Ratio Resort	Ratio City
2015	Ridge Regression	Log	1.084	1.045
2016	Ridge Regression	Log	1.067	1.047
2017	Ridge Regression	Log	1.081	1.059

GLM uncensoring

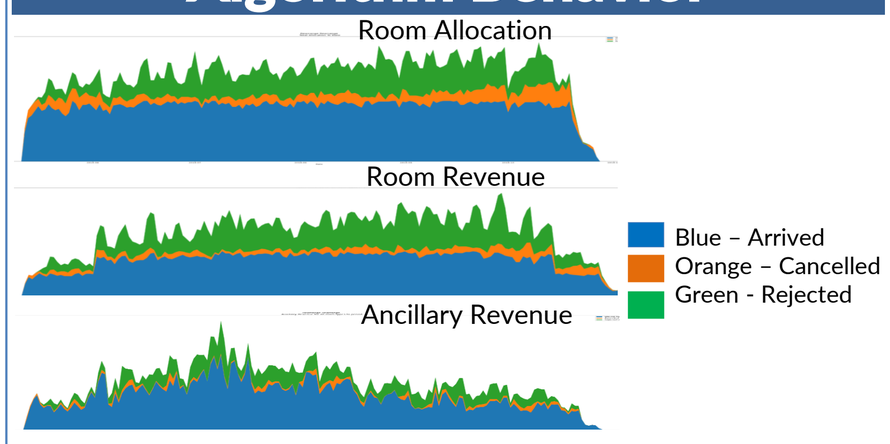
Intensity Form	Year	Ratio Resort	Ratio City
Linear Sigmoid	2015	1.067	1.013
Constant	2015	1.11	1.072
Constant	2016	1.141	1.085
Linear Sigmoid	2016	1.09	1.05
Constant	2017	1.121	1.057
Linear Sigmoid	2017	1.081	1.039

EM uncensoring

Semi-Real Data:

- Demand information is real
- Ancillary model is taken from private data (ridge regression, distorted by multiplying coefficients on random numbers)
- The ratio describes the performance of our approach relative to the FCFS (first come, first served) policy
- Top-3 ratios for EM uncensoring and the top-1 ratio for GLM uncensoring are highlighted

Algorithm Behavior



- The upper shape in the graph represents the complete demand or revenue on the table correspondingly
- The flat, static blue shape represents reliable and stable capacity utilization for room allocation
- Majority of ancillary revenue is captured. Ancillary spending is highest during the peak seasons

CCECE-2024
Kingston, ON, Canada

IEEE

Smith School of Business | Queen's University

Contact Information:
aliaksandr.nekrashevich@queensu.ca