Hospitality analytics report

Hotel RevPAR Performance Readout

A monthly STR-style performance readout for an urban hotel, staged as Hotel Meridian Lisbon: occupancy, ADR, RevPAR, and market share indices against a competitive set, built from 119,390 real bookings through a bronze, silver, and gold Delta pipeline on Databricks.

Prepared by	Leah Hall, Leah Kat Insights
Date	July 2026
Contact	leah@leahkatinsights.com, leahkatinsights.com

Table of Contents

Introduction . . . 3

1 Executive summary . . . 3

Key findings . . . 3

2 Background and objective . . . 4

3 Data and method . . . 4

Pipeline . . . 4

Measures . . . 4

Competitive set simulation . . . 5

Testing . . . 5

4 Results . . . 5

Market position . . . 5

Where the occupancy gap sits . . . 6

5 Limitations . . . 7

6 Recommendations . . . 7

Measures to track . . . 8

A Technical stack and repository . . . 8

B Glossary . . . 9

About the author . . . 9

Introduction

This report presents a monthly performance readout for a single urban hotel against a four-property competitive set, following STR reporting conventions: occupancy, ADR, RevPAR, and the MPI, ARI, and RGI share indices, with segment mix and booking pace.

The subject property's bookings are real. The competitive set is simulated from the subject's own history with seeded, reproducible perturbations, and every exhibit built on it is labeled as such. Booking pace uses only real lead-time data.

The pipeline, tests, and exhibit code are public on GitHub.

1. Executive summary

The property earns more per available room than its competitive set. Over the trailing twelve months its RevPAR index is 103.5, and it has stayed above 100 in every month. The gain comes from rate, not volume: the ADR index is 106.1 while the occupancy index is 97.5.

> **Recommendation to the General Manager**
> Hold rate. The rate premium more than covers the occupancy shortfall, and the shortfall is specific: midweek, short-lead corporate business that a discounting competitor is winning. Close it with targeted corporate rate fences and a two-week booking window offer rather than a general rate cut.

Key findings

103.5	**RevPAR index above 100 in all twelve months.** The property is gaining share of revenue in its market, and the position is stable rather than a one-month spike.
106.1	**The premium is rate-led.** ADR index above 100 throughout the period means the market keeps paying the property's rate; this is not a rate ceiling problem.
97.5	**Occupancy lags the comp set.** The gap is concentrated midweek (MPI 94.6) rather than at weekends (MPI 96.6), which points to transient corporate demand.
79.8%	**Bookings are arriving earlier.** 79.8% of final room nights were on the books 30 days out for 2017 arrivals, against 76.1% a year earlier. Pickup inside 30 days fell from 23.9% to 20.2% of final room nights, so the late-window demand the midweek gap depends on is thinner than last year.

A general rate cut would give up revenue the property already earns to buy occupancy it can win at lower cost. Section 6 sets out the recommended actions and the measures to track.

2. Background and objective

Hotel revenue teams read their market position through STR reports: a monthly comparison of the subject property against a self-selected competitive set on occupancy, average daily rate, and revenue per available room, expressed as indices where 100 means parity with the comp set. This project reproduces that readout from raw bookings, end to end.

The objective was twofold. First, produce the report a General Manager actually receives each month, with the standard indices, segment mix, booking pace, and a recommendation. Second, build it the way a production lakehouse would be built: a medallion pipeline on Databricks with data-quality gates and automated tests, so the readout can be rerun every month without manual work.

The source is the public Hotel Booking Demand dataset, 119,390 bookings for two Portuguese hotels between 2015 and 2017. The city hotel is used as the subject and staged as Hotel Meridian Lisbon. No public competitive set exists for it, so one is simulated and disclosed throughout.

3. Data and method

Pipeline

Raw bookings move through four stages. Each stage writes a Delta table in Unity Catalog, and data-quality checks stop the run on any failure, so a bad month cannot reach the readout.

Table 1. Pipeline stages

Stage	What happens	Output
Bronze	Raw CSV loaded from a Unity Catalog volume and written unchanged with an ingestion timestamp.	Bronze Delta table
Silver	Arrival dates assembled from year, month, and day fields; cancellations typed; zero-guest and negative-rate rows removed; one €5,400 ADR outlier capped; market segments normalized to five values. Data-quality checks run here.	Clean bookings
Comp set	Four competitor hotels simulated from the subject's daily history with different pricing behavior, including one discounter. Aggregated room-weighted, following STR practice.	Daily comp set KPIs
Gold	Bookings expanded to stay nights for daily occupancy, ADR, and RevPAR. Monthly rollups add MPI, ARI, and RGI. Booking pace measured at 120, 60, and 30 days before arrival.	Monthly readout tables

Measures

Table 2. KPI definitions used in this report

Measure	Definition
Occupancy	Rooms sold divided by rooms available.

Measure	Definition
ADR	Room revenue divided by rooms sold (average daily rate).
RevPAR	Room revenue divided by rooms available; equal to occupancy times ADR.
MPI	Subject occupancy divided by comp set occupancy, times 100 (market penetration index).
ARI	Subject ADR divided by comp set ADR, times 100 (average rate index).
RGI	Subject RevPAR divided by comp set RevPAR, times 100 (RevPAR generation index).

Competitive set simulation

Because the dataset covers a single property, the four-hotel comp set is generated from the subject's own daily occupancy and rate history using seeded perturbations (seed 42) so that every run reproduces the same market. One competitor is configured to discount midweek. The simulation is labeled on every exhibit that uses it. Booking pace and segment data for the subject are real and unsimulated.

Testing

Nineteen pytest tests cover cleaning rules, the comp set simulation, KPI arithmetic, and the readout tables. The same transforms run locally without Databricks through a single script, which is how the tests execute.

4. Results

Market position

Table 3 and Figure 1 summarize the last six months. The subject sits above its comp set on RevPAR in every month, and the composition of that result is consistent: rate above market, occupancy below it.

Table 3. Trailing twelve-month share indices

Index	Value	Position
RGI, RevPAR index	103.5	Above comp set
ARI, rate index	106.1	Above comp set
MPI, occupancy index	97.5	Below comp set
MPI, midweek	94.6	Below comp set
MPI, weekend	96.6	Below comp set

September 2016 to August 2017. Comp set simulated from source data.

STAR-style performance summary, subject vs competitive set

Hotel Meridian Lisbon (subject property), last six months

	Mar 17	Apr 17	May 17	Jun 17	Jul 17	Aug 17
Occupancy	89.8%	94.9%	93.7%	93.1%	96.2%	96.6%
ADR	€92	€124	€133	€130	€125	€135
RevPAR	€83	€117	€124	€121	€121	€130
Comp Occupancy	91.6%	95.7%	94.8%	94.1%	96.6%	96.4%
Comp ADR	€87	€117	€126	€123	€119	€127
Comp RevPAR	€80	€112	€119	€116	€115	€123
MPI	98.0	99.1	98.9	98.9	99.6	100.2
ARI	105.7	106.0	105.5	105.5	105.3	105.7
RGI	103.6	105.1	104.3	104.3	104.9	105.9

Source: Kaggle Hotel Booking Demand (real bookings). Comp set simulated from source data.

Figure 1. Monthly readout, subject versus comp set, last six months

Source: Kaggle Hotel Booking Demand. Comp set simulated from source data.

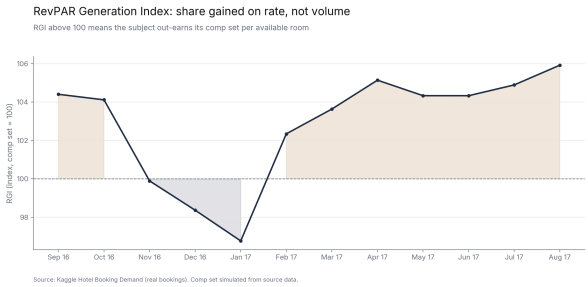


Figure 2. RGI, trailing twelve months

Source: Comp set simulated from source data.

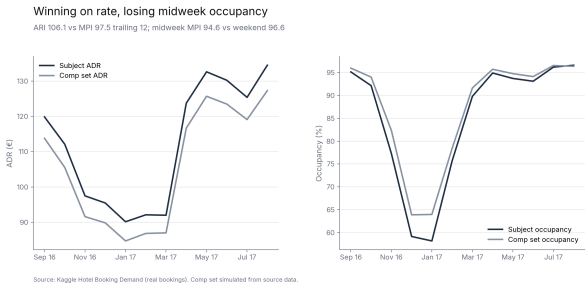


Figure 3. ADR and occupancy versus comp set

Source: Comp set simulated from source data.

Figure 2 shows share gained on rate across the year with no month below the 100 line. Figure 3 places the two components side by side: the ADR premium holds through both shoulder seasons, while occupancy tracks below the comp set for most of the period.

Where the occupancy gap sits

The occupancy shortfall is not spread evenly. Midweek MPI runs roughly two points below weekend MPI, and the segment mix shows the subject under-indexed on transient corporate business relative to the comp norm. Booking pace confirms the demand in question books inside two weeks of arrival.

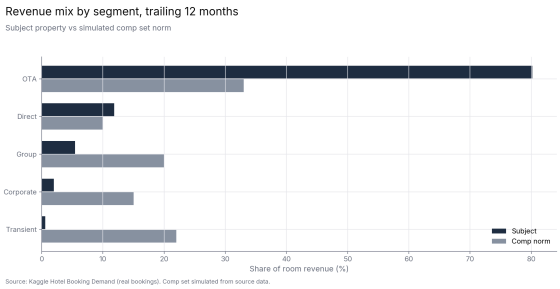


Figure 4. Revenue mix by segment versus comp norm

Source: Subject mix is real. Comp norm simulated from source data.

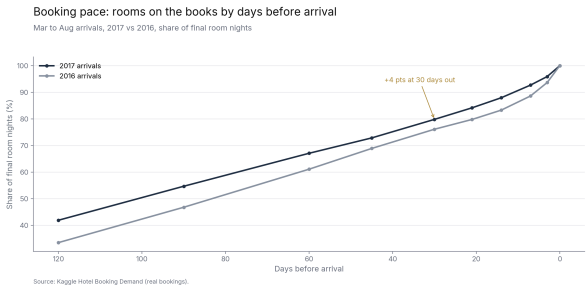


Figure 5. Booking pace, 2017 versus 2016 arrivals

Source: Real lead-time data from source bookings; no simulation.

Pace is building earlier than last year: 79.8% of final room nights were on the books at 30 days for 2017 arrivals, against 76.1% for 2016. The share of room nights picked up inside the final 30 days fell from 23.9% to 20.2%. The midweek gap therefore sits in a window that is getting thinner, which argues for a targeted offer in that window rather than a change to public rate.

5. Limitations

- **The competitive set is simulated.** Index values describe the subject against a constructed market, not a real STR comp set. The method is the point of the exercise; the specific index values are illustrative.
- **Single property, historical period.** The data covers 2015 to 2017 for one hotel. Seasonality and segment behavior may differ for other markets and years.
- **No cost or profit data.** The readout measures top-line share. The recommendation assumes the rate premium is worth more than the lost occupancy on a contribution basis, which would be confirmed with departmental profit figures.
- **Segment normalization.** Source market segments were collapsed to five values for comparability, which hides some detail within the corporate and group categories.

6. Recommendations

The analysis supports holding public rate and addressing the midweek gap with targeted, fenced offers that are invisible to leisure demand.

1. **Fence, do not discount.** Introduce a corporate midweek rate with a Sunday-through-Thursday stay pattern requirement, available through corporate channels only.
2. **Open a 14-day advance offer on midweek dates.** Size it to the thinner 30-day pickup window so it captures short-lead demand without diluting rate on dates that already fill.
3. **Hold public rate.** Cutting it to chase MPI trades index points the property owns for index points it rents.

Measures to track

Table 4. Monitoring plan

Measure	Why it matters	Review
Midweek MPI	Direct test of whether the fenced offers close the gap.	Monthly
ARI	Confirms the rate premium holds while the offers run.	Monthly
Pickup inside 30 days	Shows whether the late-window demand responds to the advance offer.	Monthly, by arrival month
RGI	The summary measure; should remain above 100 as occupancy recovers.	Monthly

Appendix A. Technical stack and repository

Table 5. Stack

Layer	Tool
Storage	Delta Lake tables in Unity Catalog (bronze, silver, gold)
Transform	PySpark notebooks on Databricks; the same modules run locally
Simulation	Seeded numpy perturbations, labeled on every exhibit
Tests	pytest, 19 tests across cleaning, comp set, KPIs, and readout
Exhibits	matplotlib, pandas

Repository layout

```
data/          source data and download instructions (raw CSV not committed)
notebooks/     01 bronze, 02 silver, 03 comp set simulation, 04 gold
scripts/       run_local.py, the same transforms without Databricks
src/revpar/    cleaning, comp set simulation, KPI, and readout logic
tests/         19 pytest tests
exhibits/      make_exhibits.py and the five PNGs
```

Appendix B. Glossary

ADR	Average daily rate: room revenue divided by rooms sold.
ARI	Average rate index: subject ADR as a percentage of comp set ADR.
Booking pace	Share of final room nights already on the books at a given number of days before arrival.
Comp set	Competitive set: the group of hotels a property benchmarks itself against.
Medallion	A lakehouse layout with bronze (raw), silver (clean), and gold (reporting) tables.
MPI	Market penetration index: subject occupancy as a percentage of comp set occupancy.
Pickup	Room nights booked within a window, here the final 30 days before arrival.
RevPAR	Revenue per available room: room revenue divided by rooms available.
RGI	RevPAR generation index: subject RevPAR as a percentage of comp set RevPAR.
STR	The hospitality benchmarking firm whose monthly report format this readout follows.

About the author

Leah Hall is a data and analytics professional based in Seal Beach, California. She builds data pipelines, forecasting models, and financial analyses for ecommerce, finance, and hospitality, holds the Certification in Hotel Industry Analytics (CHIA), and publishes the work at leahkatinsights.com.

Project page	leahkatinsights.com/projects.html#project-05
Source code	github.com/leahkatinsights/hotel-revpar-readout
Email	leah@leahkatinsights.com
Website	leahkatinsights.com
LinkedIn	linkedin.com/in/leah-hall
GitHub	github.com/leahkatinsights

Portfolio project on public data. The subject property's bookings are from the Kaggle Hotel Booking Demand dataset. The competitive set is simulated from that data and does not represent a real market; Hotel Meridian Lisbon is a staging name. Code and tests are on GitHub.