

Analysis by: Valeri Aladashvili  
[LinkedIn Profile](#)

## Hotel Booking Cancellation Analysis (Excel Project)

**Executive Summary:** The objective of this analysis was to explore hotel booking data, identify cancellation patterns, and evaluate factors like lead time, seasonality, and special requests. Completed as part of Codecademy's Excel for Data Analysis course.

### Key Analytical Findings:

**Superior Cancellation Risk with Long Lead Times:** Lead time is heavily right-skewed; longer lead times correlate with higher cancellation rates.

**Seasonal & Volume Spikes:** Cancellations and reservations peak in summer months (June-August), especially for long lead time bookings.

**City vs Resort Hotels:** City hotels show higher cancellation rates.

**Distribution Channels:** Agent channels have longest lead times; Corporate/Direct are shorter.

**Special Requests Correlation:** More special requests link to kept bookings and higher revenue.

**Data Cleaning Insights:** Handled missing/suspect data; excluded 2020 outliers for accuracy.

**Strategic Recommendation:** Reduce cancellations with short-lead incentives, flexible policies for high-risk bookings, and predictive overbooking based on seasonality and channel.

**Data Source:** Hotel booking dataset (2015-2017).

[Full data, calculations, and interactive workbooks are available on my GitHub](#)

## Data Cleaning / Change Log

Missing and Suspect Data Summary

	status	arrival	number of special requests
missing data	Yes	No	No
suspect data	No	Yes	Yes

Monthly Cancellation Rates by Hotel Type

Hotel Type	Status	Jan	Feb	March	April	May	June	July	August	September	October	November	December
city hotel	Blank	0%	0%	0%	0%	0%	0%	0%	0%	37%	0%	0%	0%
city hotel	Canceled	24%	21%	23%	31%	30%	32%	26%	25%	14%	29%	24%	25%
city hotel	Check-Out	38%	38%	42%	36%	39%	40%	38%	39%	19%	39%	40%	35%
city hotel	No-Show	1%	2%	1%	1%	1%	1%	1%	0%	0%	0%	1%	1%
resort hotel	Blank	0%	0%	0%	0%	0%	0%	0%	0%	14%	0%	0%	0%
resort hotel	Canceled	5%	9%	7%	9%	8%	9%	11%	12%	5%	9%	6%	9%
resort hotel	Check-Out	32%	29%	26%	23%	21%	19%	25%	23%	10%	23%	29%	30%
resort hotel	No-Show	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%
Total													
Canceled		29%	31%	31%	40%	39%	41%	37%	37%	20%	37%	30%	34%

Data Quality Inspection and Cancellation Patterns

Combined view of missing/suspect data check and monthly cancellation rates by hotel type. City hotels show significantly higher cancellation rates (peaking at 41% in summer) vs. resort hotels, highlighting the need for targeted retention strategies.

## Special Requests and Revenue by Year with Outlier Handling

hotel type	status	date	day	number of special requests
r	Canceled	8/12/2015	Wed	5
c	Check-Out	10/22/2015	Thu	5
c	Check-Out	10/22/2015	Thu	5
r	Check-Out	10/22/2015	Thu	5
c	Canceled	6/30/2016	Thu	5
c	Check-Out	8/22/2016	Mon	5
r		9/21/2016	Wed	5
c		9/26/2016	Mon	5
c	Canceled	10/6/2016	Thu	5
c	Canceled	10/6/2016	Thu	5
c	Canceled	10/7/2016	Fri	5
r	Check-Out	10/20/2016	Thu	5
r	Canceled	10/23/2016	Sun	5
c	Canceled	11/7/2016	Mon	5
c	Canceled	11/7/2016	Mon	5
c	Canceled	11/29/2016	Tue	5
r	Check-Out	12/5/2016	Mon	5
c	Check-Out	12/17/2016	Sat	5
c	Check-Out	2/25/2017	Sat	5
c	Canceled	4/8/2017	Sat	5
r	Check-Out	5/3/2017	Wed	5
c	Check-Out	5/14/2017	Sun	5
c	Canceled	5/14/2017	Sun	5
c	Canceled	5/15/2017	Mon	5
r	Canceled	5/20/2017	Sat	5
r	Check-Out	5/20/2017	Sat	5
c	Canceled	5/27/2017	Sat	5
c	Check-Out	6/2/2017	Fri	5
c	Check-Out	6/10/2017	Sat	5
r	Canceled	6/10/2017	Sat	5
r	Check-Out	7/2/2017	Sun	5
c	Check-Out	7/11/2017	Tue	5

c	Check-Out	7/12/2017	Wed	5
r	Check-Out	7/31/2017	Mon	5
c	Canceled	7/31/2017	Mon	5
c	Canceled	8/12/2017	Sat	5
c	Check-Out	8/13/2017	Sun	5
c	Check-Out	8/21/2017	Mon	5
r	Canceled	8/25/2017	Fri	5
r	Check-Out	8/28/2017	Mon	5

### **Practical Example – High Special Request Bookings**

Filtered view of bookings with 5 special requests (post-cleaning). Demonstrates text trimming (c/r codes) and filtering; high-request bookings often result in Check-Out, indicating upselling/revenue potential.

Year	Total special requests	Revenue		Average cost per special request -->	\$20
2015	8551	\$171,020			
2016	30878	\$617,560			
2017	28782	\$575,640			
2020	105	\$2,100			
	Average without 2020:	\$454,740			
NOTE: Excluded 2020 data from calculations due to suspect outliers (unrealistic number of special requests)					

Outlier Handling & Revenue Metrics

Yearly special requests and revenue summary. Excluded 2020 outlier to calculate reliable \$20 average revenue per special request. Sheet protected for data integrity.