Analysis by: Valeri Aladashvili
LinkedIn Profile

Hotel Booking Data Visualization & Insights

Project Overview

Explored and visualized patterns in a hotel booking dataset (~119k rows, 2015–2017) to uncover drivers of cancellations. Focused on lead time distribution, seasonality effects, and channel differences using advanced formatting, trend lines, and segmented tables for actionable business insights.

Skills Demonstrated

- Advanced Conditional Formatting (color scales for heatmaps)
- Sparkline / Trend Lines (in-cell visualizations for monthly patterns)
- Pivot Tables & Segmented Analysis (by status, lead time category, month, channel)
- Data Binning & Distribution Insights (short/medium/long lead time categorization)
- Visual Storytelling (heatmaps + notes to highlight seasonality and risk factors)

Key Analytical Findings

- Long lead times strongly correlate with higher cancellation risk.
- Peak summer months show massive volume increase, especially for long-lead bookings.
- Agent channel: highest volume but longest lead times; Corporate/Direct: shorter and more reliable.

Business Recommendations

- Offer short-lead incentives and stricter policies for long-lead/agent bookings in summer.
- Prioritize direct/corporate channels for lower risk and higher retention.
- Implement predictive overbooking using seasonality and channel patterns.

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

Full data, calculations, and interactive workbooks are available on my GitHub

Monthly Booking Volume by Reservation Status and Lead Time

Heatmap table with conditional formatting and trend lines. Long lead time bookings (both Canceled and Check-Out) spike dramatically in peak summer months (June–August), indicating higher volume but also elevated cancellation risk during high-demand periods.

Reservation Status	Lead Time	January	February	March	April	May	June	July	August	September	October	November	December
Canceled	Long	12	45	111	182	355	395	259	325	568	287	97	59
Canceled	Medium	242	653	575	824	555	549	1086	1230	400	618	316	499
Canceled	Short	52	58	43	39	83	43	70	59	27	51	28	36
Check-out	Long	11	26	114	195	614	530	324	463	753	573	35	35
Check-out	Medium	658	1089	1431	1642	1305	1117	2231	2208	1014	1253	1111	1218
Check-out	Short	1199	1193	1028	713	616	391	582	586	335	751	830	764
No-show	Long		2	2		6	3		5	2	2		
No-show	Medium	9	19	13	9	14	11	17	13	9	10	15	13
No-show	Short	10	18	19	5	11	6	4	5		10	5	24

For reservations with a Long lead time, the volume of both Canceled and Check-out reservations significantly increases during the peak summer months (June, July, August)

Lead Time by Status

The distribution of cancellations is heavily skewed to the right. The highest volume of cancellations (over 1000) occurs in the very first bin, which is the [0,15] day range (reservations cancelled within 15 days of arrival). This indicates that the largest number of cancellations are last-minute or short-notice. Beyond this initial spike, the number of cancellations steadily decreases as the lead time (number of days) increases. Cancellations for bookings made more than 400 days in advance are extremely rare. The distribution for Kept reservations is also heavily skewed right. The largest peak occurs in the [0,15] day bin (over 4500 reservations). The volume of kept reservations drops sharply in the subsequent bins, falling close to 1000 in the [30,45] day range, and then decreases consistently further to the right. This shows that the majority of kept reservations are short-notice.

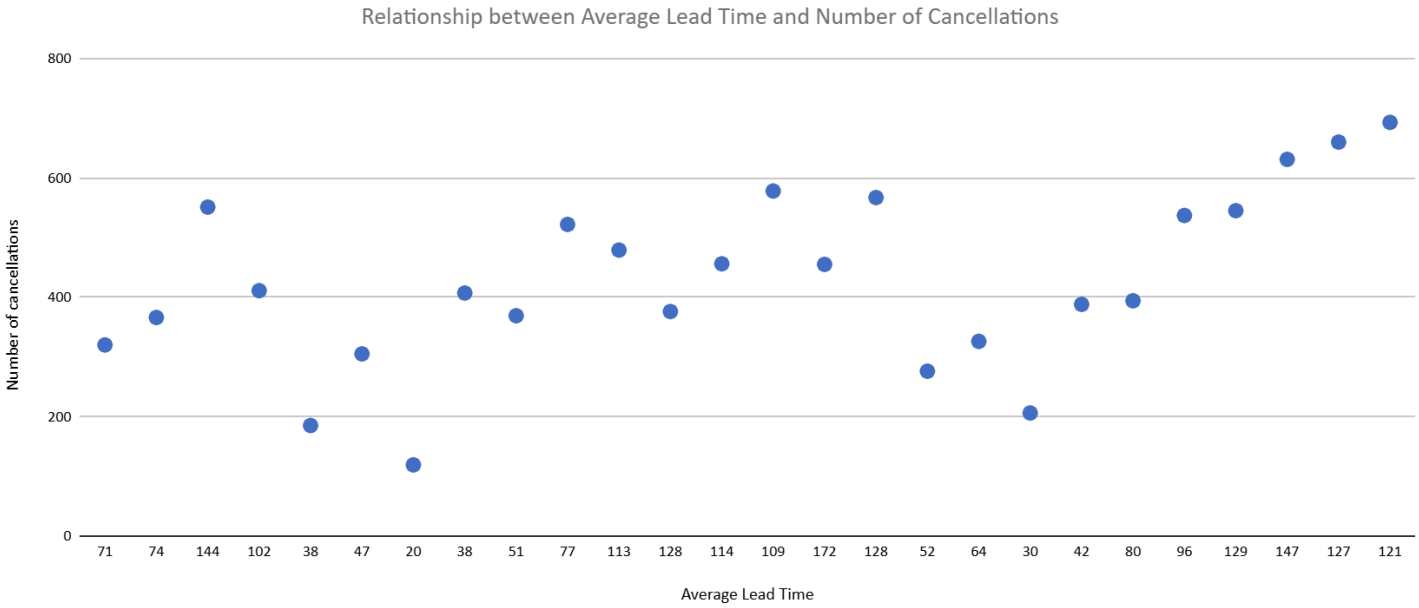
Cancelled	Kept
85	342
75	737
23	7
60	13
96	14
45	14
40	0
43	9
45	35
47	68
3	18
71	37
63	68
62	37
101	12
51	0
48	7
368	37
81	72
79	72

More ..

Lead Time to Cancellations

The scatterplot shows a positive correlation between the average lead time and the number of cancellations. This means that as the average time between booking and arrival increases, the total number of cancellations also tends to increase.

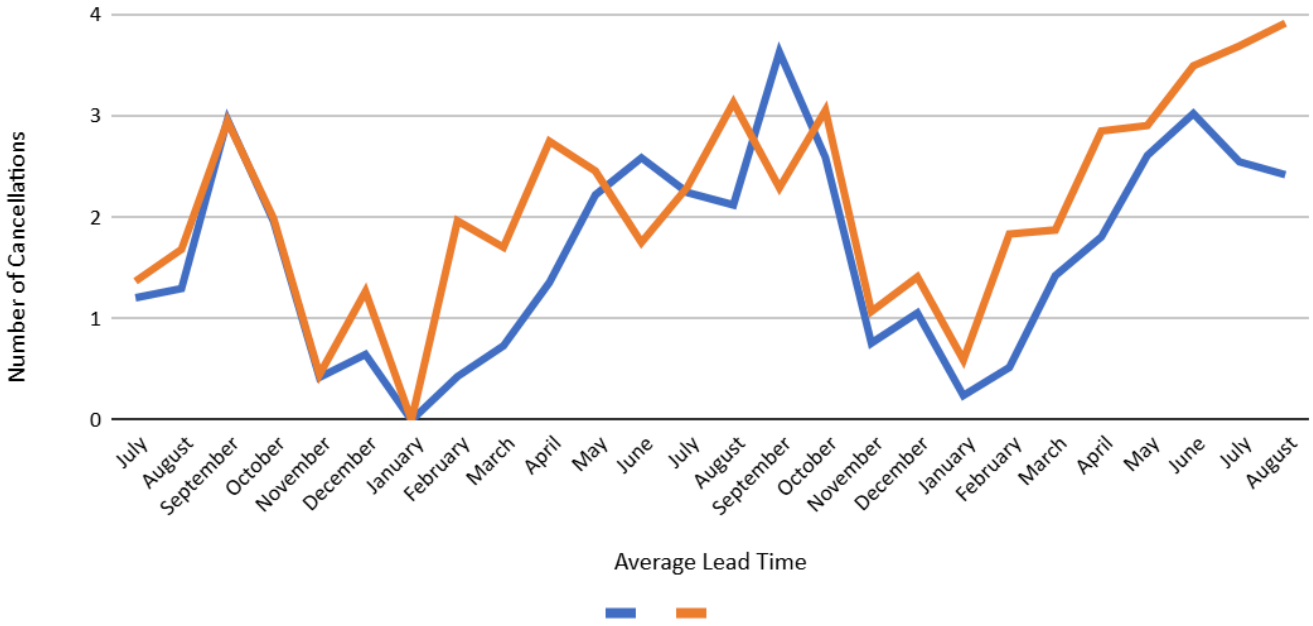
The correlation is not perfectly strong, but the general trend suggests that months with longer average lead times generally experience a higher volume of cancellations.



Year	Arrival Month	Average lead time	Number of cancellations
2015	July	71	320
2015	August	74	366
2015	September	144	551
2015	October	102	411
2015	November	38	185
2015	December	47	305
2016	January	20	119
2016	February	38	407
2016	March	51	369
2016	April	77	522
2016	May	113	479
2016	June	128	376
2016	July	114	456
2016	August	109	578
2016	September	172	455
2016	October	128	567
2016	November	52	276
2016	December	64	326
2017	January	30	206
2017	February	42	388
2017	March	80	394
2017	April	96	537
2017	May	129	545
2017	June	147	631
2017	July	127	660
2017	August	121	693

Over Time

The line chart demonstrates a very strong positive correlation between Average Lead Time and the Number of Cancellations. Both lines show almost identical peaks (e.g., September/October and the following August) and valleys (e.g., November and January). This suggests that the relationship between lead time and cancellations is heavily influenced by seasonal factors, as both metrics rise and fall together throughout the year.



Normalized Average Lead Time vs Cancellations			
Year	Arrival Month	Average Lead Time	Cancellations
2015	July	1.2	1.4
2015	August	1.3	1.7
2015	September	3.0	2.9
2015	October	2.0	2.0
2015	November	0.4	0.4
2015	December	0.6	1.3
2016	January	0.0	0.0
2016	February	0.4	2.0
2016	March	0.7	1.7

2016	April	1.4	2.7
2016	May	2.2	2.5
2016	June	2.6	1.8
2016	July	2.2	2.3
2016	August	2.1	3.1
2016	September	3.6	2.3
2016	October	2.6	3.1
2016	November	0.8	1.1
2016	December	1.1	1.4
2017	January	0.2	0.6
2017	February	0.5	1.8
2017	March	1.4	1.9
2017	April	1.8	2.8
2017	May	2.6	2.9
2017	June	3.0	3.5
2017	July	2.5	3.7
2017	August	2.4	3.9

Hotel Bookings by Distribution Channel: Lead Time and Volume Insights

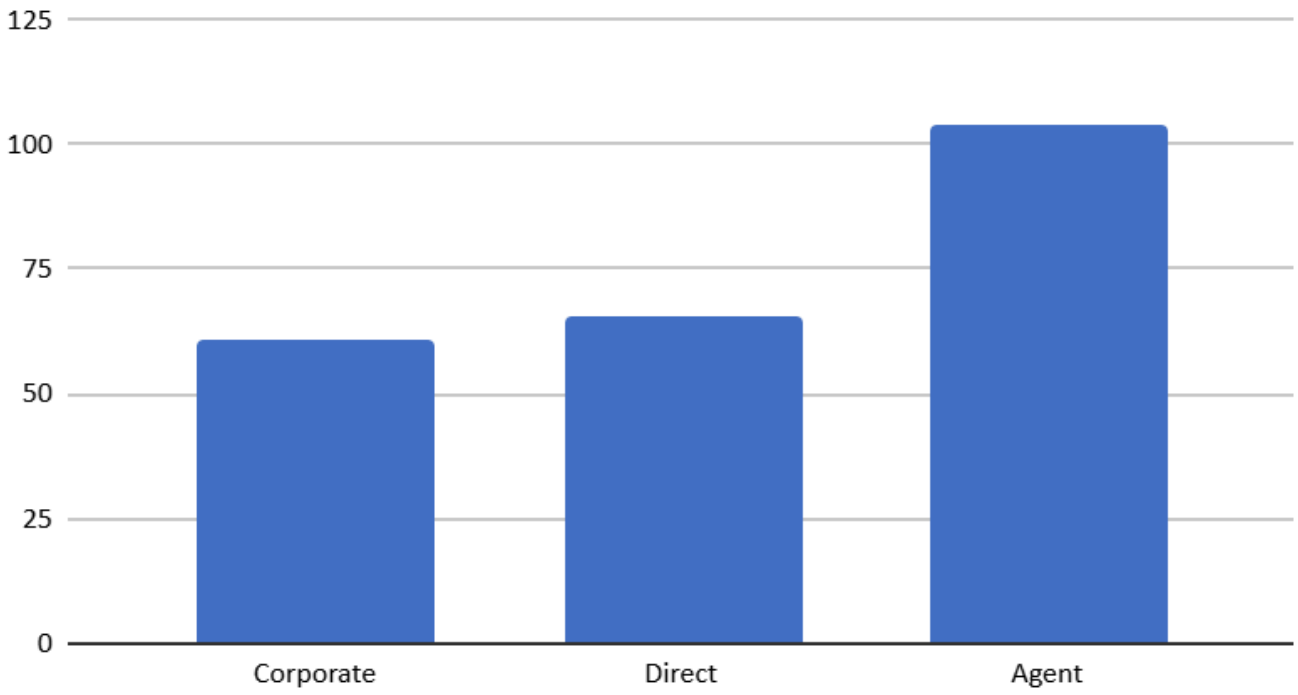
The analysis by distribution channel reveals three key insights:

Dominance of Agents (Total Volume): The Agent channel is the primary source of business, accounting for the highest volume of both kept (19,816) and cancelled (9,109) reservations. The Direct channel is the second largest, and Corporate is the smallest.

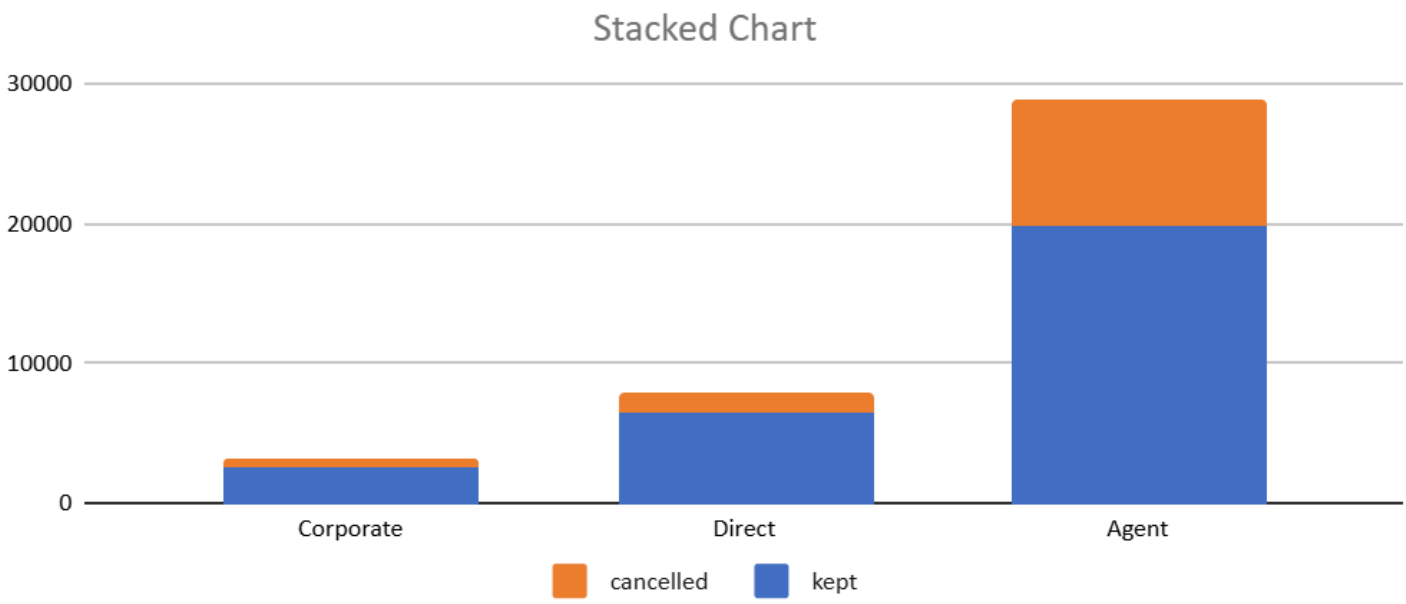
Longest Lead Time (Average Lead Time Chart): The Agent channel has a significantly longer Average Lead Time (104 days) compared to the Direct (65 days) and Corporate (61 days) channels. This explains why the overall cancellation rate is high for long lead times.

Lead Time Distribution (Percentage Chart): The Agent channel is the most reliant on Long lead times (46% of its business falls into the long category), while the Corporate channel is more evenly split between Long (35%) and Medium (33%) lead times. The Short lead time category is consistently the smallest across all channels (7-9%).

Distribution Channel	Average Lead Time
Corporate	61
Direct	65
Agent	104



	Corporate	Direct	Agent
kept	2581	6540	19816
cancelled	688	1325	9109



Lead Time	Corporate	Direct	Agent
Long	35%	27%	46%
Medium	33%	23%	34%
Short	9%	7%	7%