

Per què les reserves locals cancel·len més?

Una història de dades sobre risc de cancel·lació en reserves hoteleres a Portugal (2015–2017)

Acte 1 — Context

Tot i que el dataset inclou dos tipus d'hotel, el volum de reserves no és equilibrat. El City Hotel concentra aproximadament dos terços de les reserves en tots els anys analitzats, fet que amplifica qualsevol risc associat a aquest segment.

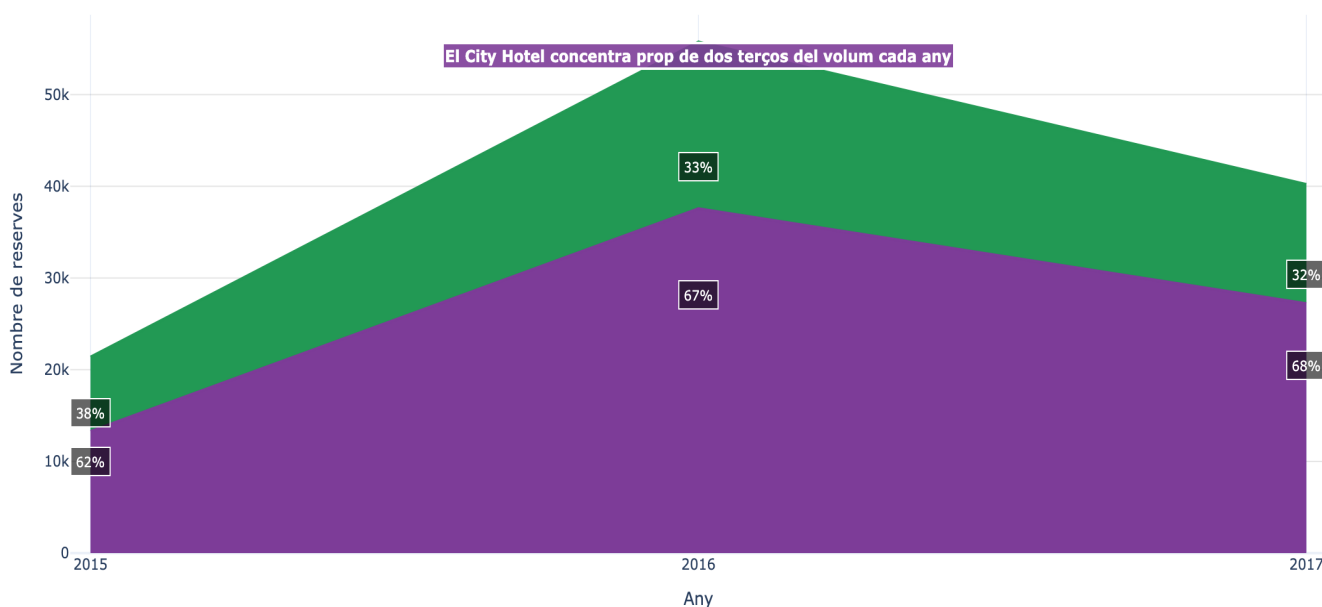
Aquesta distribució desigual implica que qualsevol diferència de comportament associada al City Hotel tindrà un impacte desproporcionat sobre el risc global del sistema.

Visualització temporal: Stacked Area Chart que permet visualitzar simultàniament el pes relatiu de cada hotel i la seva evolució temporal, mostrant que la dominància del City Hotel és consistent al llarg dels anys.

Distribució del volum de reserves per tipus d'hotel (2015–2017)

El City Hotel concentra de manera consistent la major part del volum

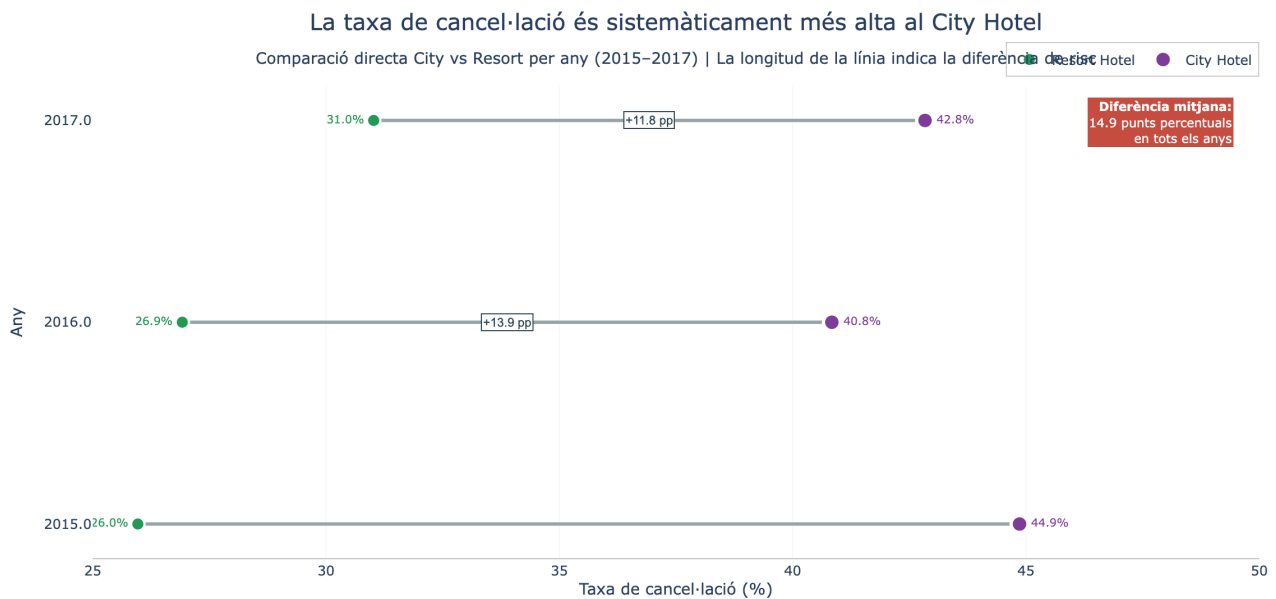
Resort Hotel City Hotel



Acte 2 — Tensió

En tots els anys analitzats, la taxa de cancel·lació del City Hotel supera la del Resort Hotel. La diferència es manté estable al llarg del temps, cosa que indica que no es tracta d'un fenomen puntual, sinó d'un patró estructural.

Visualització de diferència: Dumbbell Plot que permet visualitzar directament la diferència de risc entre els dos tipus d'hotel.



La diferència no només és constant, sinó també rellevant en magnitud.

La persistència d'aquesta diferència al llarg dels anys analitzats suggereix que el risc de cancel·lació no respon a fluctuacions puntuals, sinó a un patró estructural associat al tipus d'hotel.

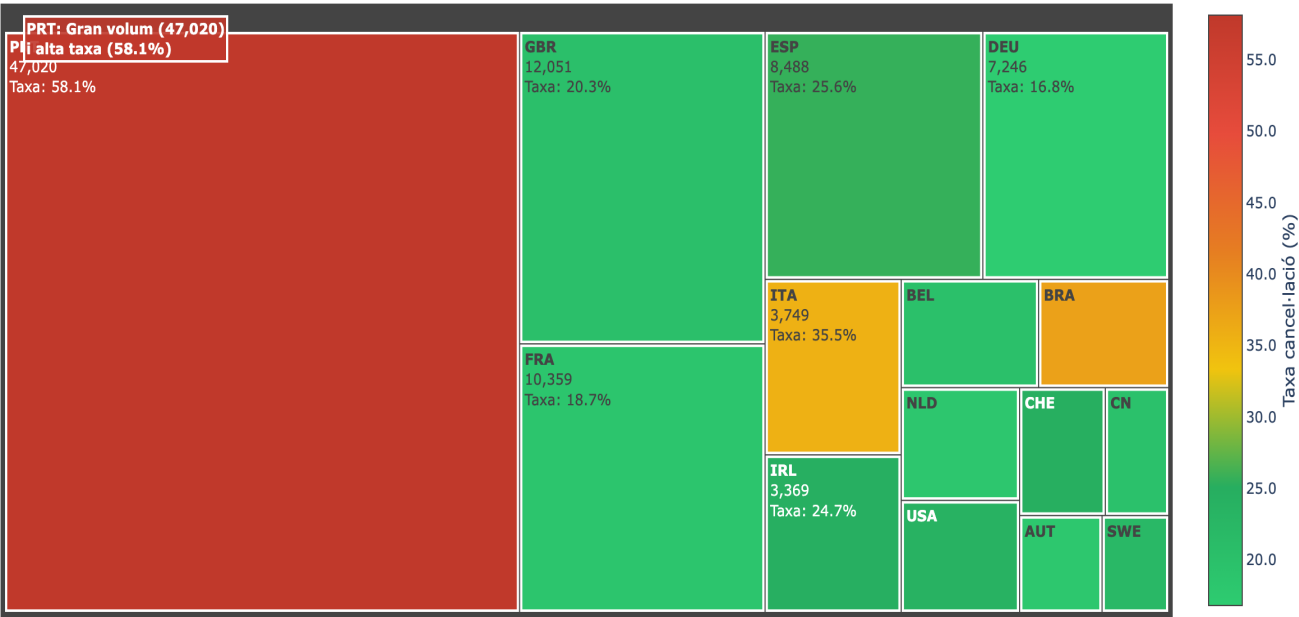
Acte 3 — Clímax

Aquí apareix un patró especialment rellevant des del punt de vista operatiu: el país amb més reserves (Portugal) també presenta una taxa de cancel·lació més alta que la majoria de mercats internacionals. El Treemap mostra simultàniament el volum de reserves (àrea) i la taxa de cancel·lació (color), revelant l'impacte agregat del risc.

Visualització d'impacte: Treemap que permet visualitzar simultàniament el volum (àrea) i la taxa de cancel·lació (color), fent evident que Portugal no només té una taxa alta, sinó també un impacte operatiu significatiu per la seva gran quantitat de reserves.

Impacte de risc per mercat (Treemap)

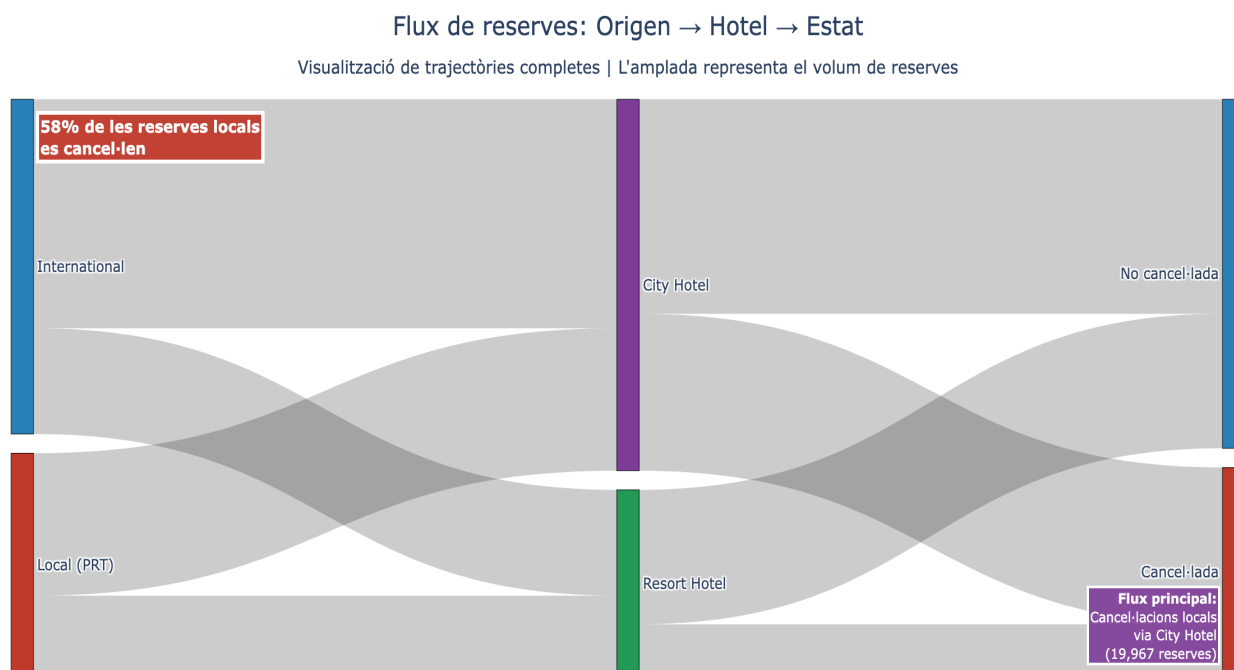
Àrea = volum de reserves | Color = taxa de cancel·lació | PRT destaca per volum i risc



Acte 4 — On passa exactament

El diagrama de Sankey mostra el flux complet de reserves des de l'origen fins a l'estat final. Això permet veure trajectòries completes i on es concentra el flux de cancel·lacions, no només comparacions estàtiques.

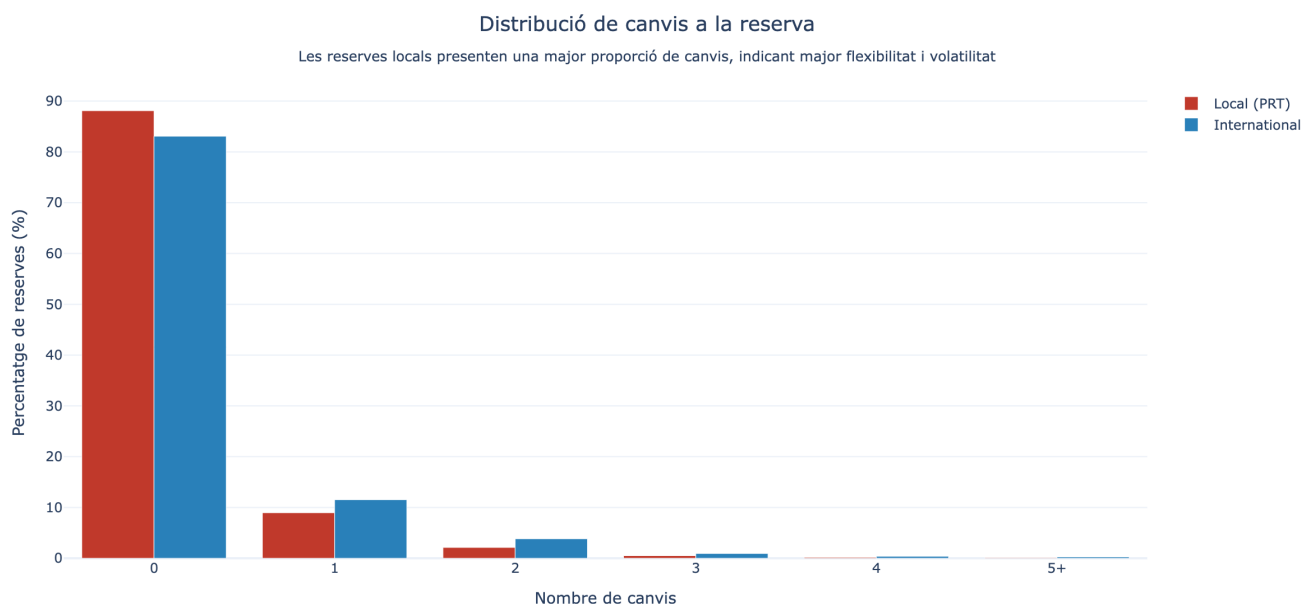
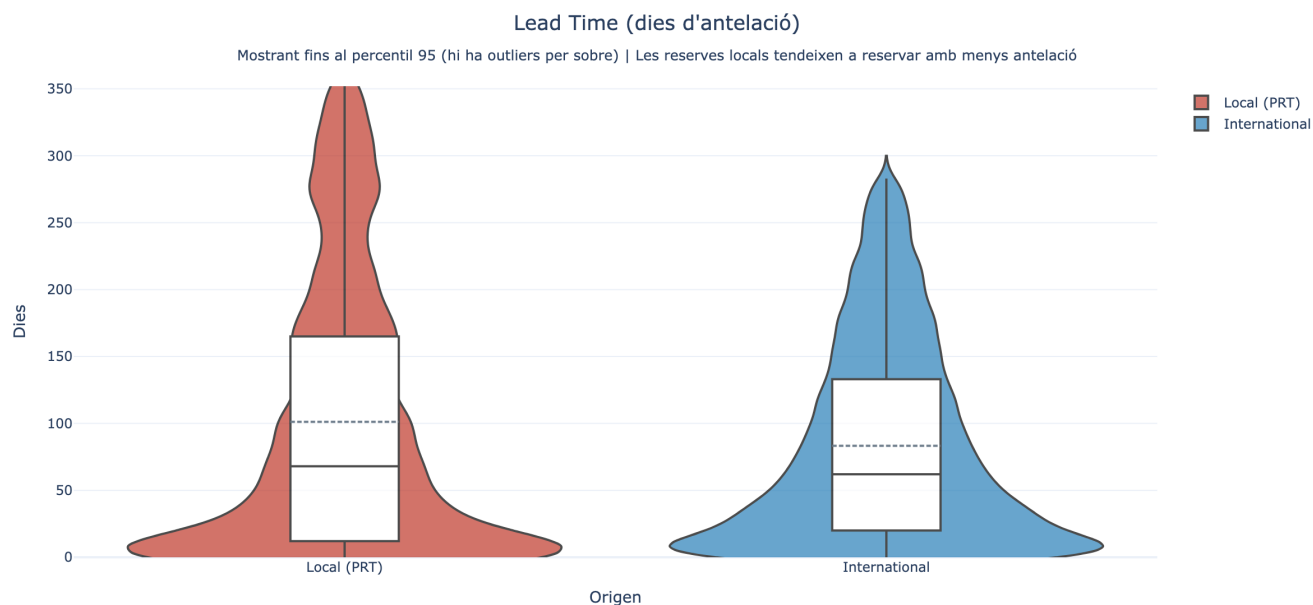
Visualització de flux: He utilitzat un Sankey perquè vull mostrar fluxos i composició, no només comparacions estàtiques. L'amplada de cada flux representa el volum de reserves.



El diagrama de flux mostra que una part significativa de les cancel·lacions de reserves locals es canalitza a través del City Hotel, reforçant la interacció entre l'origen de la reserva i el tipus d'hotel.

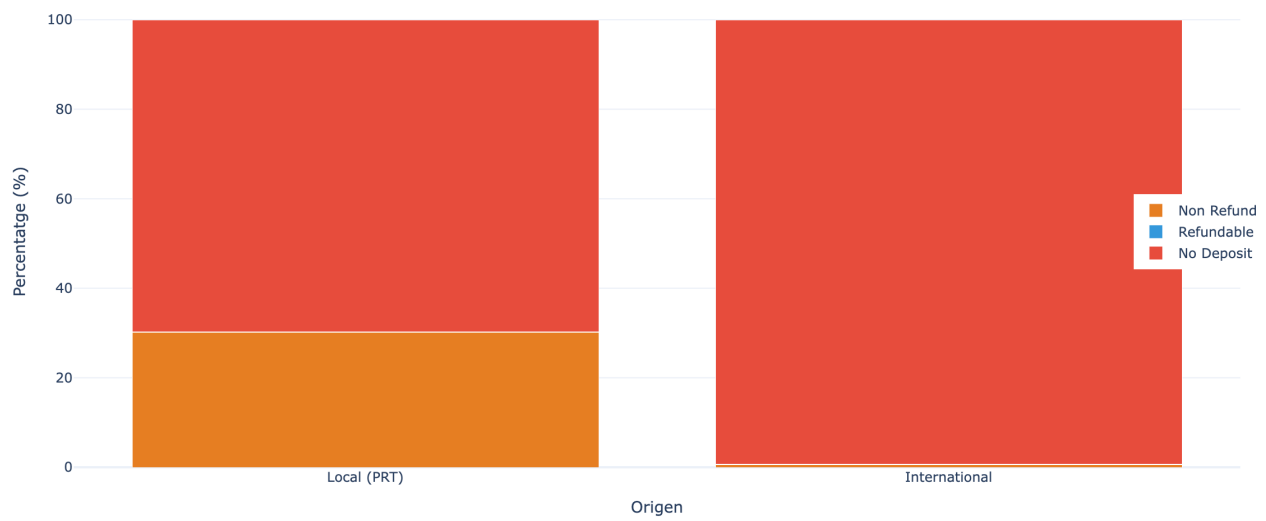
Acte 5 — Possible explicació

Una possible peça del trencaclosques és el comportament de reserva. Les reserves locals mostren patrons diferents: reserven amb menys antelació, fan més canvis a les reserves i utilitzen menys dipòsits Non Refund (que impliquen més compromís). Aquests factors poden estar associats a una major volatilitat i, per tant, més cancel·lacions.



Tipus de dipòsit per origen

Non Refund implica més compromís amb la reserva i està associat a menys cancel·lacions



Cal remarcar que aquestes relacions no impliquen una causalitat directa, sinó associacions observades en el comportament de reserva.

Sobre aquesta visualització

Metodologia

El conjunt de dades recull més de 117.000 reserves realitzades entre 2015 i 2017 en dos hotels de Portugal: un hotel urbà a Lisboa i un resort a l'Algarve.

Abans de construir la visualització final, s'ha realitzat una fase d'analítica visual exploratòria utilitzant el notebook proporcionat (Component 1 de la PAC), introduint diverses correccions per garantir la qualitat de les dades. Les dades netes generades en aquesta fase s'utilitzen directament per a la visualització narrativa, assegurant consistència entre l'anàlisi exploratòria i la visualització final.

Les correccions aplicades durant l'EDA inclouen:

- Eliminació d'outliers en adults, infants i bebès
- Eliminació de valors negatius o extremadament elevats a l'ADR
- Tractament coherent de valors nuls
- Construcció robusta de la variable temporal

Aquest procés assegura que les visualitzacions reflecteixin patrons reals de comportament i no errors de registre.

Eines i Tecnologies

La visualització s'ha creat utilitzant **Plotly**, una llibreria Python per a visualitzacions interactives. Aquesta elecció permet crear visualitzacions avançades com diagrames de Sankey, treemaps i violin plots, mantenint alhora la interactivitat necessària per a una experiència narrativa efectiva.

Les funcionalitats interactives emprades inclouen:

- **Tooltips informatius:** Al passar el ratolí sobre els elements es mostra informació detallada
- **Zoom i pan:** Permet explorar els gràfics en detall
- **Disseny responsive:** S'adapta a diferents mides de pantalla
- **Navegació interactiva:** Menú fixe i indicador de progrés per facilitar la navegació

Decisions de Disseny

Cada tècnica de visualització s'ha escollit en funció de la pregunta analítica que es vol respondre:

- **Stacked Area Chart:** Permet visualitzar simultàniament el pes relatiu i l'evolució temporal
- **Dumbbell Plot:** Visualitza directament la diferència de risc entre grups
- **Treemap:** Mostra simultàniament volum (àrea) i taxa de cancel·lació (color)
- **Sankey Diagram:** Visualitza fluxos i trajectòries completes
- **Violin Plot, Histograma i Barres apilades:** Permeten comparar distribucions, freqüències i composicions

La paleta de colors és consistent al llarg del dashboard per diferenciar hotels i origen del client, utilitzant una tipografia clara (Segoe UI) i s'han eliminat elements visuals que no aporten informació rellevant.

Objectiu i Usuari

L'objectiu de la visualització és explicar el risc de cancel·lació des d'un punt de vista operatiu, pensada per a perfils de gestió hotelera o anàlisi de negoci.

La història s'ha estructurat com una narrativa en cinc actes, seguint una progressió de context, tensió, clímax i possible explicació, facilitant una comprensió progressiva d'un fenomen complex.