

## 1. Component 2: Visualització tipus storytelling

### 1.1 Contextualització, objectiu i usuari

#### 1.1.1 Missatge central

La història que es vol explicar és la següent:

**Les cancel·lacions no són aleatòries: responen a patrons sistemàtics i segmentables (origen, tipus de viatge, tipus d'hotel i preu), i això obre la porta a decisions concretes de revenue management i política comercial.**

El relat parteix d'un problema real del sector: *una reserva no és un ingrés fins que es materialitza*. La cancel·lació és un risc operatiu i financer perquè desestabilitza la planificació (ocupació, personal, inventari d'habitacions) i pot provocar pèrdua d'ingressos o sobrecostos (recol·locacions, overbooking defensiu, ajustos de tarifes a última hora).

A partir del conjunt de dades de demandes hoteleres (2015–2017) procedent de sistemes PMS reals, es proposa una narrativa que ajudi a respondre una pregunta directiva:

**Quins segments de clients i quins tipus de reserves són “més fràgils” i, per tant, requereixen polítiques més estrictes o estratègies específiques?**

#### 1.1.2 Usuari objectiu (persona) i necessitat

L'usuari ideal és un perfil professional del sector hoteler amb responsabilitat sobre decisions comercials:

- **Revenue manager / direcció comercial:** necessita segmentar el risc per ajustar dipòsits, condicions de cancel·lació i estratègies de preus.
- **Direcció d'operacions:** necessita anticipar fluctuacions d'ocupació per planificació de recursos.
- **Analista de negoci / BI:** vol una explicació interpretativa, replicable, no només gràfics bonics.

A diferència d'una anàlisi purament exploratòria, el *storytelling* exigeix:

1. un **conflicte clar** (incertesa per cancel·lacions),
2. una **tensió que creix** (descoberta de patrons i segments de risc),
3. una **resolució accionable** (què faria un hotel amb això).

### 1.2 Presentació del conjunt de dades i síntesi crítica de l'anàltica visual

#### 1.2.1 Origen i naturalesa de les dades

El dataset conté reserves de dos hotels portuguesos (un *City Hotel* a Lisboa i un *Resort Hotel* a l'Algarve) per arribades entre juliol de 2015 i agost de 2017. Les variables provenen de bases

de dades PMS (taula de reserves i *change logs*) i han estat dissenyades per evitar *data leakage* (valors capturats el dia anterior a l'arribada per a reserves creades abans de l'arribada). Això és rellevant perquè quan analitzem cancel·lacions, no totes les variables són “neutres”: part de la informació canvia amb el temps i amb l'operativa hotelera.

### 1.2.2 Neteja de dades (per què és important pel relat)

L'analítica visual prèvia (el fitxer R que ens venia donat) va detectar valors extrems o inconsistents (p. ex. valors impossibles d'hostes, ADR negatius o desmesurats) i va justificar la seva correcció/eliminació. Aquesta fase és essencial en un storytelling honest perquè:

- evita construir una història sobre soroll o errors;
- permet que els patrons narrats siguin atribuïbles al comportament real i no a anomalies.

També es van crear transformacions orientades a interpretabilitat narrativa (no només estadística), per exemple:

- agrupacions de durada d'estada i d'historial de cancel·lacions,
- variables sintètiques com *first-time visitor* i *booking\_changed*,
- tipologies de viatge (*tipo*) basades en caps de setmana, feina, descans i paquets.

Aquesta decisió connecta amb la teoria de data storytelling: no és suficient visualitzar; cal construir **significat** i **context** per a un públic concret.

### 1.2.3 Troballes clau que motiven la història (amb xifres)

Els resultats que actuen com a “motor narratiu” són:

**1) La cancel·lació és freqüent i desigual.** En el dataset net, la taxa global de cancel·lació és aproximadament del **37,5%**. Però aquesta mitjana amaga diferències molt fortes:

- **City Hotel:** **42,2%** de cancel·lacions
- **Resort Hotel:** **28,1%** de cancel·lacions

**2) L'origen és un factor estructural.** Segmentant per origen:

- **Portugal:** **58,1%** de cancel·lacions
- **Internacional:** **23,7%** de cancel·lacions

Aquesta asimetria és coherent amb la interpretació qualitativa: el turista domèstic té menys cost de replanificació i més flexibilitat.

**3) El risc depèn del tipus de viatge i del context (interaccions).** Per exemple, la taxa de cancel·lació a **City Hotel + Portugal** pot superar el **70%** en segments com *rest* i *work+rest*, mentre que a **Resort Hotel + Internacional** molts segments es mantenen per sota del **20%**. Això demostra que no n'hi ha prou amb mirar una variable: cal explicar **interaccions**.

**4) El preu (ADR) també forma part del relat.** L'ADR mitjà varia per tipologia i segment. Per exemple, a **Resort Hotel + Portugal**, els *package* tenen ADR mitjà molt alt (per damunt de **130€**), mentre que *weekend* és molt més baix (entorn de **80€**). En canvi, a **City Hotel + Internacional** l'ADR és més homogeni entre tipologies (*estandardització* de demanda urbana).

**5) Temporalitat i estacionalitat segmentada.** Els gràfics temporals indiquen comportaments diferents: a City Hotel domina la demanda internacional gairebé tot l'any, mentre que a Resort Hotel hi ha una estacionalitat més marcada i un pes domèstic més rellevant fora de temporada alta. Això reforça que “**quan**” es reserva i “**qui**” reserva s’han d’explicar plegats.

### 1.3 Storyboarding complet (guió gràfic) per a Flourish Studio

#### 1.3.1 Estructura narrativa (Freytag aplicada a dades)

S’aplica l’esquema de la [Piràmide de Freytag](#) (exposició → tensió → clímax → resolució):

- **Exposició:** quant gran és el problema? (cancel·lacions globals)
- **Tensió:** descobrim que no és homogeni (hotel i origen)
- **Clímax:** apareixen segments “crítics” (interacció origen × hotel × tipo)
- **Resolució:** com transformar-ho en decisions (polítiques, preus, dipòsits, segmentació)

#### 1.3.2 Storyboard en escenes (scrollytelling 5–7 minuts)

N.	Objectiu narratiu	Què diu el text	Visual (Flourish)
1	Hook / problema	“Una reserva no és ingrés fins que arriba. Quantes es perdren pel camí?”	KPI + barra 100% (cancel·lades vs no)
2	Context dataset	“Dues realitats: Lisboa vs Algarve, 2015–2017.”	Dues targetes + mini timeline
3	Tensió 1: hotel	“El risc no és igual: ciutat cancela més que resort.”	Barres apilades 100% per hotel
4	Tensió 2: origen	“Portugal és molt més volàtil que l'internacional.”	Barres apilades 100% per origen
5	Escalada: interacció	“No és una variable: és la combinació la que importa.”	Heatmap: hotel × origen × tipo (cancel rate)
6	Clímax: segments crítics	“Hi ha segments que superen el 70% i d'altres sota el 10–20%.”	Small multiples / heatmap anotat
7	Explicació econòmica	“El compromís canvia: ADR i tipus de viatge.”	Box/violin d'ADR per tipo + facetes

8	Temporalitat	“Quan reserves? Canvia per hotel i origen.”	Línies setmanals + fletes hotel
9	Resolució accionable	“Què faria un hotel?” (polítiques, dipòsits, comunicació)	Decisions per segment

---

## 2. Guió del vídeo de visualització

### Diapositiva 1 — Títol i plantejament del problema

Quan una reserva no esdevé estada: com segmentar el risc de cancel·lació en dades hoteleres

Les cancel·lacions de reserves hoteleres són un problema recurrent dins del sector. Una reserva que es cancel·la no és només una habitació buida, sinó incertesa, pèrdua potencial d'ingressos i dificultats de planificació. L'objectiu d'aquesta visualització és entendre si aquest comportament respon a patrons identificables o si és simplement aleatori.

### Diapositiva 2 — Magnitud del fenomen

Quan observem el conjunt global de reserves, es fa evident que una part important acaba sent cancel·lada. Aquesta visió general ens indica que no estem davant d'un fenomen marginal, sinó d'un comportament prou freqüent com per justificar una anàlisi més profunda.

### Diapositiva 3 — Primer tall: tipus d'hotel

Si comencem a desglossar les dades segons el tipus d'hotel, apareix el primer patró clar. Els hotels urbans presenten taxes de cancel·lació més elevades que els hotels resort. Això suggereix que el context del viatge influeix en el nivell de compromís del client.

### Diapositiva 4 — Incorporant l'origen del client

El següent pas és preguntar-nos qui són aquests clients que cancel·len més sovint. Quan distingim entre reserves nacionals i internacionals, observem que el comportament no és homogeni. Les reserves nacionals tendeixen a mostrar una major taxa de cancel·lació, especialment en entorns urbans.

### Diapositiva 5 — Interacció entre hotel, origen i tipus de viatge

A mesura que combinem més dimensions, el comportament es torna més complex però també més informatiu. Els viatges associats a descans o paquets presenten una major inestabilitat, sobretot quan es combinen amb hotels urbans i clients nacionals. Aquí ja no parlem d'un únic factor, sinó de la interacció entre diversos elements.

### Diapositiva 8 — Diferències entre països

L'origen geogràfic del client aporta una nova capa d'interpretació. Quan analitzem els principals països d'origen, apareixen diferències significatives tant en volum com en comportament de

cancel·lació. Això reforça la idea que el context cultural i logístic del viatge té un impacte rellevant, i el fet que els clients nacionals o de païssos més propers, són els que tenen més tendència a cancel·lar.

### **Diapositiva 7 — Diferències entre païssos i tipus de viatge**

No obstant, quan observem aquesta distribució segons el tipus de viatge i hotel, observem com els patrons són similars a quan observavem totes les cancel·lacions, el que ens indica que aquestes cancel·lacions tenen més dependència del tipus de viatge i hotel que del païs, encara que les cancel·lacions nacionals continuen sent majoritaries.

### **Diapositiva 8 — El paper del preu en el context de la reserva**

Intuïtivament, podríem pensar que un preu més alt implica un major compromís i menys cancel·lacions. Tanmateix, quan el preu es posa en relació amb el tipus d'hotel i l'origen del client, aquesta idea es matisa. El preu, per si sol, no explica el comportament; Observem com en els tipus d'hotel i origen del viatger, les tendències de preus són similars, però no sembla haver una correlació amb les cancel·lacions per tipus de viatge que em vist anteriorment. Per tant, és el context en què es produeix la reserva el que determina el risc real.

### **Diapositiva 9 — La dimensió temporal**

El temps introduceix una nova perspectiva en la història. Al llarg de l'any, les cancel·lacions no es distribueixen de manera uniforme. Apareixen patrons estacionals que reflecteixen canvis en la demanda i en el tipus de client segons el moment del calendari. Podem observar com en general les cancel·lacions al hotel de ciutat són més estables, encara que cauen a Internacional al 3er trimestre, mentre que hi ha un clar patró estacional als messos de més calor a les cancel·lacions del restor.

### **Diapositiva 10 — L'efecte del lead time**

Un dels factors més determinants emergeix quan considerem el temps d'antelació amb què es fan les reserves. Com més llunyana és la data d'arribada en el moment de reservar, més alta és la probabilitat de cancel·lació. Aquest resultat apunta directament al nivell de compromís del client.

### **Diapositiva 11 — El comportament passat del client**

El comportament previ del client completa el relat. Els clients que han cancel·lat reserves anteriorment presenten una probabilitat molt més elevada de tornar a cancel·lar. Això suggerix que la cancel·lació no sempre és una decisió puntual, sinó sovint un patró recurrent.

### **Diapositiva 12 — Síntesi dels principals patrons**

Arribats a aquest punt, els patrons clau es fan evidents. Les cancel·lacions són més freqüents en hotels urbans, en reserves nacionals, en viatges associats a paquets i descans, en reserves fets amb molta antelació i clients que ja han cancel·lat previament.

Aquests comportaments només emergeixen clarament quan s'analitzen les dades de manera conjunta.

### Diapositiva 13 — Dades, eines i metodologia

Pel que fa a l'anàlisi visual, s'han utilitzat diferents tipus de gràfics segons l'objectiu de cada vista. Els gràfics de barres i heatmaps permeten comparar ràpidament percentatges de cancel·lació entre segments, mentre que els Violin Graphs ajuden a analitzar la variabilitat del preu.

La paleta de colors s'ha mantingut coherent al llarg de tota la visualització. Els tons càlids s'utilitzen per representar cancel·lacions, mentre que els tons més foscos o neutres indiquen reserves no cancel·lades. Aquesta coherència cromàtica facilita la interpretació i redueix la càrrega cognitiva. El contrari passa als gràfics de heatmap, per donar contrast a la resta de gràfics, utilitzem una escala de blaus per donar intensitat a les cancel·lacions.

La tipografia és simple i llegible, prioritant la claredat per sobre de l'estètica decorativa. Els textos són breus i funcionen com a guia, no com a explicació exhaustiva.

Finalment, la interactivitat és un element clau del disseny. Filtres, pestanyes i animacions permeten que l'usuari explori els patrons de manera progressiva com tipus d'hotel o origen dels viatgers, transformant una anàlisi complexa en una experiència comprensible i narrativa.

### Diapositiva 14 — Conclusions finals

Per concloure, aquesta visualització mostra que les cancel·lacions no són esdeveniments aïllats, sinó el resultat de patrons de comportament i context clarament identificables. L'ús de storytelling interactiu permet explorar la complexitat, comparar segments i entendre el risc de manera intuïtiva. Més enllà de l'estadística descriptiva, aquest enfocament transforma dades complexes en coneixement útil per a la presa de decisions.

En un context real, aquests insights permetrien definir polítiques de cancel·lació diferenciades, ajustar estratègies de preus i millorar la planificació operativa de l'hotel, que era l'objectiu de la visualització.

## 3. Bibliografia

### Referències

- [1] Antonio, N., de Almeida, A., & Nunes, L. (2019). *Hotel booking demand datasets*. Data in Brief, 22, 41–49.
- [2] Ramos, C. M. Q., Ashqar, R. I., & Contreiras, A. (2025). *Design of Dashboards for CRM Associated with Health and Wellbeing Tourism*. HCII 2024, LNCS 15376.
- [3] Jiménez, E. (2025). *Visual analytics of hotel bookings data* (informe d'analítica visual i resultats). Document de pràctica.