



Big Data Analytics su dataset "Hotel Reviews" con Apache Spark

Giuseppe Pasquale Caligiure
Mat. 280867

Contesto di lavoro e obiettivo

L'archivio “Hotel Reviews” contiene 515.738 recensioni di hotel di lusso in Europa, raccolte da Booking.com

- Ogni recensione presenta 17 campi
- Formato File: csv
- Dimensioni File: circa 238 MB
- Reperibile su: [515K Hotel Reviews Data in Europe](#)

Obiettivo: realizzare un'app interattiva che consenta di effettuare interrogazioni sul dataset per **estrarre insight significativi dai dati**, sfruttando le potenzialità di elaborazione distribuita del framework Spark

Architettura Frontend/Backend

Backend (PySpark): la logica di lavoro delle interrogazioni è stata realizzata sfruttando il motore di calcolo distribuito in-memory di PySpark, cioè le API Python per Apache Spark, ed è contenuta nel file “queries.py”

- Ogni query accetta un DataFrame Spark in input e restituisce un DataFrame Spark contenente i risultati
- Altre tecnologie utilizzate: Spark MLlib, Scikit-learn (in UDF Pandas)

Frontend (Streamlit): l’interfaccia utente è gestita da una **web-app** interattiva realizzata con il framework Streamlit.

- La UI consente all’utente di selezionare un’analisi da effettuare, invoca la corrispondente funzione dal backend e infine visualizza i risultati ottenuti.
- Altre tecnologie utilizzate: Altair e PyDeck per mappe geospaziali, Pandas per grafici interattivi

Analisi dei dati

Le query implementate esplorano diversi metodi di analisi dei dati, fra cui:

- **Analisi Descrittiva:** sintetizzare e riassumere caratteristiche di dati storici (metriche chiave come media e deviazione standard).
- **Analisi Diagnostica:** confrontare i dati, scoprire correlazioni e capire l'impatto di un fattore sull'altro (regressione) in modo da identificare le cause reali dei fenomeni osservati.
- **Analisi Predittiva:** prevedere risultati (clustering/trend), basandosi sull'uso di algoritmi che imparano dagli schemi presenti nei dati passati per proiettarli nel futuro.

1. Trend Recensioni (Time Series)

Obiettivo: Analizzare il trend temporale dei punteggi degli hotel, utilizzando la **Regressione Lineare Score vs Tempo**, per identificare quali alberghi stanno migliorando o peggiorando nel tempo.

Casi d'uso: Identificare "stelle nascenti" o hotel decadenti nonostante un alto punteggio medio storico.

Opzioni Query

Scelgi la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Trend Recensioni (Time Series)

Analizza il trend temporale dei punteggi di ogni hotel, utilizzando la Regressione Lineare.

L'obiettivo è capire se un hotel sta migliorando o peggiorando nel tempo.

Venne identificato se la reputazione dell'hotel è:

- In crescita: quando la pendenza della retta di regressione lineare è positiva.
- In decrescita: quando la pendenza della retta di regressione lineare è negativa.
- Stabile: quando la pendenza della retta di regressione lineare è zero.

Legenda Campi:

- Trend_Slope : Pendenza della retta di regressione lineare.
- Review_Count : Numero totale di recensioni dell'hotel.
- Average_Score_Calculated : Punteggio medio calcolato sulle recensioni presenti nel dataset.
- Average_Score : Punteggio medio calcolato su tutte le recensioni ricevute dall'hotel nell'ultimo anno (anche quelle che non sono presenti nel dataset).
- First_Review_Date : Data della prima recensione dell'hotel.
- Last_Review_Date : Data della ultima recensione dell'hotel.

Nota: Il calcolo del trend viene effettuato tenendo conto solo delle recensioni presenti nel dataset che hanno una data valida. Inoltre, vengono esclusi gli hotel con meno di 30 recensioni valide.

Minimo numero di recensioni per albergo

30

Calcola Trend per tutti gli Hotel

1. Trend Recensioni (Time Series): Risultati

- Top 10 Hotel in crescita/calo ordinati per **trend slope**
- Grafico **Distribuzione Trend vs Punteggio medio** (consente di individuare visivamente gli alberghi migliori o peggiori e leggerne le caratteristiche)



Top 10 Hotel in Crescita

Hotel_Name	Trend_Slope	Review_Count	Average_Score_Calculated	Average_Score	First_Review_Date	Last_Review_Date
1076 The Curtain	0.0064	30	8.8833	9.1	2017-05-25	2017-08-03
931 Chasse Hotel	0.0059	139	9.0344	8.9	2017-03-27	2017-08-02
125 Hotel Park Lane Paris	0.0041	154	8.6358	8.7	2016-05-23	2017-08-03
1096 Villa Lutet Port Royal	0.004	47	6.3851	7	2015-05-15	2017-08-02
853 NIX Milan	0.0035	180	8.5956	8.8	2017-03-25	2017-08-03
352 Hotel Capitol Milano	0.0031	66	8.1242	8.3	2015-09-22	2017-07-05
1298 Lansbury Heritage Hotel	0.0031	40	9.5175	9.4	2017-06-25	2017-08-02
1114 Best Western Le 18 Paris	0.0031	86	7.4291	7.6	2016-07-22	2017-07-20
265 Bob Hotel by Elegancia	0.003	49	8.9398	9	2017-03-12	2017-08-02
90 Hilton London Euston	0.0025	470	7.1209	7.4	2015-08-04	2017-08-02

Top 10 Hotel in Calo

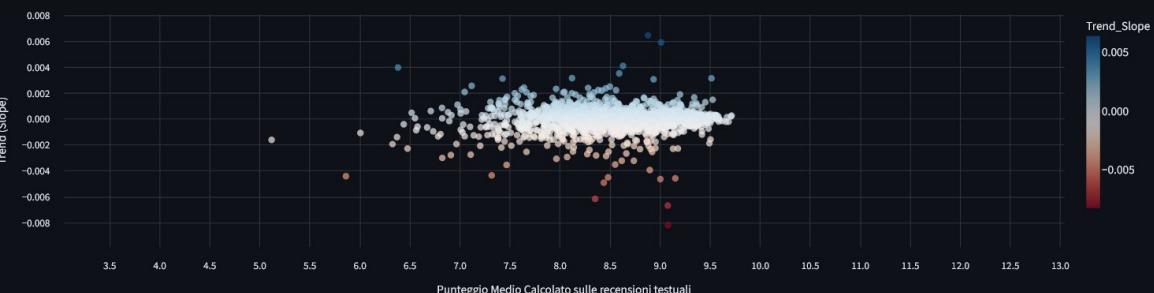
Hotel_Name	Trend_Slope	Review_Count	Average_Score_Calculated	Average_Score	First_Review_Date	Last_Review_Date
632 Okko Hotels Paris Porte De Versailles	-0.0082	43	9.0637	9.2	2017-06-09	2017-07-31
1019 Le Tuilie Hotel	-0.0067	57	9.0807	9.3	2017-03-25	2017-08-02
380 Marin Waterloo	-0.0062	84	8.3548	8.6	2017-05-29	2017-08-03
917 Arthotel ANA WestBahn	-0.0049	67	8.4418	8.3	2017-03-02	2017-08-03
1025 Maison Albar Hotel Paris C Line	-0.0047	62	9.0065	8.9	2018-12-25	2017-08-03
1297 La Villa Neumann	-0.0046	41	9.1561	9.2	2017-01-30	2017-08-02
1314 Majestic Hotel Spa	-0.0045	38	8.4042	8.7	2015-08-18	2017-07-29
904 Villa Eugenie	-0.0044	62	5.0845	6.8	2015-08-05	2017-08-02
830 London Suites	-0.0044	114	7.3211	7.4	2016-07-03	2017-08-02
183 Park Plaza London Park Royal	-0.004	197	8.991	8.9	2017-03-10	2017-08-03

Distribuzione Trend vs Punteggio Medio

In questo grafico è possibile osservare la distribuzione dei trend in relazione al punteggio medio degli hotel.

- ^ In alto si trovano gli hotel con trend positivo (in crescita)
- v In basso gli hotel con trend negativo (in calo).
- → A destra si trovano gli hotel con punteggio medio alto
- ← A sinistra gli hotel con punteggio medio basso.
- Gli hotel con punteggio medio alto e trend positivo sono i migliori hotel.

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative all'hotel.



2. Analisi Influenza Tag

Obiettivo: Determinare quali fattori (es. "Single Room", "No Window") impattano positivamente o negativamente sul punteggio che i recensori assegnano agli hotel.

Casi d'uso: Identificare le caratteristiche più apprezzate o criticate dai clienti degli hotel.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Analisi Influenza Tag

Questa query analizza l'impatto dei tag che appaiono nelle recensioni (es. "double bedroom", "no windows", ecc.), determinando quali **caratteristiche degli hotel** influenzano positivamente le recensioni (sono associate a voti più alti) e quali influenzano negativamente le recensioni (sono associate a voti più bassi).

Inoltre, per ogni tag viene calcolato un **indice di affidabilità** che tiene conto della frequenza e della deviazione standard dei voti, premiando i tag con elevata frequenza e deviazione standard ridotta (cioè i tag che appaiono spesso e in recensioni con voti simili e coerenti fra loro).

1. **Map:** Espplode la lista dei tag di ogni recensione.
2. **Reduce:** Aggrega per tag calcolando voto medio, frequenza e deviazione standard.
3. **Analisi di ogni tag:**
 - **Average_Score:** punteggio medio delle recensioni associate al tag.
 - **Count:** frequenza del tag.
 - **Impact:** differenza tra il punteggio medio del tag e il voto medio globale (calcolato su tutte le recensioni).
 - **StdDev_Score:** deviazione standard dei punteggi del tag:
 - < 1.0: **Molto Attendibile.** C'è forte consenso (quasi tutte le recensioni hanno lo stesso punteggio).
 - 1.0 - 2.0: **Normale.** C'è una naturale variabilità umana, ma il trend è chiaro
 - > 2.0: **Disperso/Controverso.** C'è grande variabilità umana (le recensioni hanno voti molto diversi).
 - **Reliability Index:** Indice euristico $\frac{1}{\text{StdDev}} * \log(\text{Count})$ che premia coerenza e frequenza.
 - **Weighted Impact:** $\text{Impact} * \text{Reliability Index}$.

Minimo numero di occorrenze per tag

50

Analizza Influenza Tag

2. Analisi Influenza Tag: Risultati

- Top 10 Tag Positivi/negativi ordinati per **Weighted Impact**, cioè un indice euristico che valuta l'impatto di un tag, pesandolo per la sua "affidabilità", cioè premiando tag i cui voti hanno una **deviazione standard bassa** e una **frequenza alta**.
- Grafico **Affidabilità vs Impatto**: consente di individuare i tag che influenzano maggiormente la valutazione degli hotel e confrontarlo a colpo d'occhio con il grado di attendibilità

Impact = Average Score - Global Average

Reliability Index = $(1 / (\text{StdDev} + 0.1)) * \log(\text{Count})$

Weighted Impact = Impact * Reliability Index



Media Globale Voti (calcolata su tutte le recensioni):

8.40

👉 Top 10 Tag Positivi (per Weighted Impact)

Tag più affidabili che alzano il voto.

Single_Tag	Average_Score	Count	Impact	Reliability_Index	Weighted_Impact
0 Camper Room	9.73	225	1.33	9.04	12.04
1 Kings Junior Suite	9.65	55	1.26	6.95	8.76
2 Luxury Double or Twin Room	9.68	68	1.28	6.78	8.67
3 Double or Twin Room Allergy Free	9.69	50	1.30	5.97	7.74
4 Double Room XL	9.67	58	1.28	5.49	7.00
5 Deluxe Double Room 1 2 Adults	9.53	305	1.13	5.73	6.48
6 Small Double Room Annex building	9.56	72	1.17	5.14	6.01
7 Double Room Annex	9.43	187	1.03	5.68	5.88
8 Delightful Queen Room	9.56	64	1.17	5.03	5.87
9 City King Room	9.30	739	0.91	6.47	5.86

👉 Top 10 Tag Negativi (per Weighted Impact)

Tag più affidabili che abbassano il voto.

Single_Tag	Average_Score	Count	Impact	Reliability_Index	Weighted_Impact
660 Standard Double Room without Window	7.03	2424	-1.36	3.63	-4.96
659 Cabin Single Room	6.63	195	-1.76	2.64	-4.65
658 Standard Double Room No Window	6.98	470	-1.41	3.14	-4.44
657 Eiffel Tower View King Room	6.65	109	-1.75	2.18	-3.82
657 Eiffel Tower View King Room	6.65	109	-1.75	2.18	-3.82
656 Single Guest Room	6.90	109	-1.50	2.53	-3.79
655 King Hilton Guest Room	7.20	646	-1.19	3.09	-3.68
654 Superior Suite with 1 Double Bed and 1 Single Bed	7.01	181	-1.39	2.64	-3.65
653 Small Single Room	7.12	312	-1.27	2.74	-3.50
652 Standard Double Room with View Terrace	7.13	142	-1.27	2.74	-3.48
651 Standard Room with 1 Double Bed	7.47	879	-0.92	3.61	-3.34

Grafico: Affidabilità vs Impatto

- Asse X (Impact): Quanto il tag sposta il voto (Desta=Positivo, Sinistra=Negativo).
- Asse Y (Reliability): Quanto è "solido" il dato (Alto=Molto affidabile, Basso=Incerto).
- Obiettivo: Cerca i tag negativi in alto a destra (vincitori sicuri) e in alto a sinistra (problematici).

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative al tag.



3. Analisi Bias Nazionalità

Obiettivo: Individuare, se esistono, le nazionalità che tendono a dare voti più alti o bassi rispetto alla media.

Casi d'uso: Profilazione per nazionalità e identificazione di bias culturali che peggiorano o migliorano l'esperienza dei clienti.

Opzioni Query

- Scgli la query da eseguire:
- Trend Recensioni (Time Series)
 - Analisi Influenza Tag
 - Analisi Bias Nazionalità
 - Analisi Competitività Locale
 - Segmentazione Hotel (K-Means)
 - Migliori Hotel per Nazione
 - Locals vs Tourists - Preferenze e distribuzione clienti
 - Analisi Stagionale & Target
 - Analisi Durata Soggiorno
 - Analisi Esperienza Recensore



Hotel Reviews Analytics con Spark

Query selezionata: Analisi Bias Nazionalità

Questa query cerca di identificare se esistono nazionalità tendenzialmente più generose o severe nei voti. Inoltre, calcola un **Sentiment Ratio** basato sul rapporto tra parole positive e totali usate nelle recensioni.

Legenda campi:

- **Reviewer Nationality:** Nazionalità del recensore.
- **Average Score:** Voto medio assegnato da recensori di stessa nazionalità.
- **Count:** Numero di recensioni per nazionalità.
- **Score Deviation:** Differenza tra il voto medio della nazionalità e la media globale dei voti di tutte le recensioni.
- **Sentiment Ratio:** percentuale di parole positive sul totale delle parole usate nelle recensioni di stessa nazionalità (indicatore di "positività" nel testo).
- **Total Words Avg:** Lunghezza media delle recensioni di stessa nazionalità (quanto sono "verbosi").

Minimo numero recensioni per nazionalità

Analizza Bias Nazionalità

8.40

🧐 I Più Critici (Voti Bassi)

Reviewer_Nationality	Average_Score	Count	Score_Deviation	Sentiment_Ratio	Total_Words_Avg
140 Mongolia	7.25	39	-1.14	0.42	20.72
139 Syria	7.52	33	-0.87	0.24	36.85
138 Libya	7.60	72	-0.80	0.43	23.06
137 Seychelles	7.60	21	-0.80	0.43	30.14
136 Bangladesh	7.61	151	-0.79	0.46	26.85
135 Iran	7.73	1086	-0.67	0.47	26.62
134 Algeria	7.78	100	-0.62	0.44	21.55
133 Tanzania	7.81	58	-0.59	0.42	26.14
132 Jordan	7.81	757	-0.59	0.43	27.30
131 Pakistan	7.81	916	-0.59	0.49	27.81

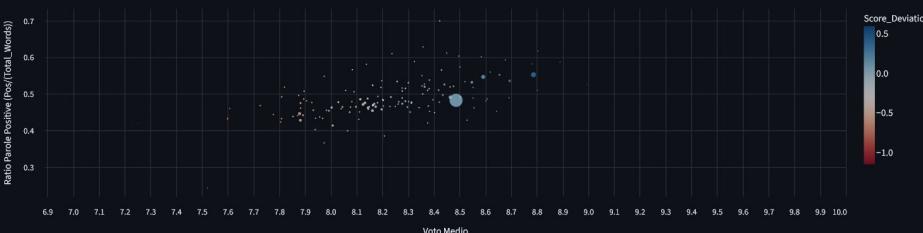
🧐 I Più Generosi (Voti Altì)

Reviewer_Nationality	Average_Score	Count	Score_Deviation	Sentiment_Ratio	Total_Words_Avg
0 Kyrgyzstan	9.00	21	0.60	0.53	37.05
1 Liechtenstein	8.89	20	0.49	0.59	31.60
2 Puerto Rico	8.80	180	0.41	0.62	29.08
3 Panama	8.80	122	0.41	0.51	25.35
4 United States of America	8.79	35437	0.39	0.55	42.02
5 Honduras	8.78	22	0.39	0.58	29.77
6 Israel	8.69	6610	0.30	0.54	33.10
7 El Salvador	8.69	24	0.30	0.49	23.08
8 Trinidad and Tobago	8.68	154	0.28	0.59	38.40
9 New Zealand	8.65	3237	0.26	0.55	35.05

Correlazione Voto vs Positività Testo (Corenza delle recensioni per nazionalità)

Chi dà voti alti scrive davvero cose positive? (In alto a destra = Sì, In basso a destra = No)

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative alla nazionalità.



3. Analisi Bias Nazionalità: Risultati

- Top 10 Nazionalità più critiche/generose
- **Score Deviation** = deviazione dal voto globale medio
- **Sentiment Ratio** = rapporto fra numero di parole positive e negative utilizzate nelle recensioni.
Indica la **coerenza** dei recensori: chi è mediamente più generoso dovrebbe utilizzare più parole positive che negative, e viceversa.
- Grafico **Correlazione Voto vs Positività Testo**: consente di valutare visivamente la coerenza dei recensori per ogni nazionalità

4. Analisi Competitività Locale

Obiettivo: Confrontare le performance di un hotel rispetto ai suoi concorrenti geograficamente più vicini.

Casi d'uso: Analisi della competitività fra attività geograficamente vicine.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Analisi Competitività Locale

Confronta ogni hotel con i suoi vicini nel raggio di N km, individuando:

- **Gemme Locali (Outperformers):** Hotel con punteggio superiore alla media della zona.
- **Hotel sotto la media (Underperformers):** Hotel con punteggio inferiore alla media della zona.

Logica di funzionamento:

1. Calcola la distanza fra ogni hotel e i suoi competitor usando la **Formula di Haversine** (con coordinate geografiche in latitudine e longitudine)
2. Esclude i competitor che si trovano fuori dal raggio di ricerca specificato
3. Analizza le seguenti statistiche per ogni hotel:
 - **Average_Score** = Media dei voti dell'hotel (valore basato sulle recensioni raccolte nell'ultimo anno)
 - **Neighborhood_Avg_Score** = Media dei voti dei competitor nella zona
 - **Score_Delta** = Differenza di punteggio tra l'hotel e la **media dei competitor** (se il delta è positivo, l'hotel è meglio dei competitor)
 - **Competitor_Count** = Numero di competitor nella zona

Raggio di ricerca (km)

Minimo numero di competitor nella zona

Analizza Competitività

4. Analisi Competitività Locale: Risultati

- Top 10 Outperformer/Underperformer
- **Score Delta** = differenza fra il punteggio di un hotel e il punteggio medio del suo vicinato
- Distanza fra hotel calcolata con Formula di Haversine, a partire dalle coordinate geografiche.
- Grafico **Performance Relativa**: consente di confrontare visivamente le performance di tutti gli hotel



Top 10 Gemme Locali (Meglio dei competitor vicini)

	Hotel_Name	Average_Score	Neighborhood_Avg_Score	Score_Delta	Competitor_Count
0	Milestone Hotel Kensington	9.50	8.23	1.27	100
1	Ritz Paris	9.80	8.59	1.21	230
2	Nu Hotel	8.90	7.82	1.08	5
3	Hotel The Peninsula Paris	9.50	8.44	1.06	129
4	Waldorf Astoria Amsterdam	9.50	8.48	1.02	65
5	Haymarket Hotel	9.60	8.59	1.01	134
6	Pillows Anna van den Vondel Amsterdam	9.40	8.39	1.01	47
7	Hotel Eiffel Blomet	9.40	8.39	1.01	46
8	Charlotte Street Hotel	9.50	8.49	1.01	136
9	H tel de La Tamise Esprit de France	9.60	8.61	0.99	228

Top 10 Sotto la Media (Peggio dei competitor vicini)

	Hotel_Name	Average_Score	Neighborhood_Avg_Score	Score_Delta	Competitor_Count
1376	Hotel Liberty	5.20	8.42	-3.22	41
1375	Hotel Cavendish	6.40	8.53	-2.13	120
1374	The Tophams Hotel	6.60	8.71	-2.11	88
1373	Savoy Hotel Amsterdam	6.40	8.35	-1.95	41
1372	Britannia International Hotel Canary Wharf	7.10	8.96	-1.86	8
1371	Bloomsbury Palace Hotel	6.80	8.53	-1.73	119
1370	Villa Eugenie	6.80	8.47	-1.67	105
1369	Gran Hotel Barcino	7.00	8.67	-1.67	119
1368	Park Lane Mews Hotel	7.00	8.61	-1.61	137
1367	Best Western Maitrise Hotel Edgware Road	6.60	8.18	-1.58	78

Grafico: Performance Relativa

Confronto tra il voto dell'hotel (Asse X) e il voto medio della zona (Asse Y).

- **Sotto la diagonale:** L'hotel è meglio della zona (Gemma Locale)
- **Sopra la diagonale:** L'hotel è peggio della zona (Underperformer)

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative all'hotel.



5. Segmentazione Hotel (K-Means Clustering)

Obiettivo: Raggruppare gli hotel in cluster omogenei (utilizzando K-Means Clustering), sulla base delle seguenti feature: Punteggio Medio, Popolarità (numero recensioni), Verbosità delle recensioni, Posizione geografica, Bias di nazionalità.

Casi d'uso: Segmentazione di marketing, raccomandazioni di hotel simili.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Segmentazione Hotel (K-Means)

Identifica gruppi di hotel simili in base alle loro caratteristiche. Questo metodo *unsupervised* permette di scoprire pattern nascosti nel dataset.

Scegli le caratteristiche da utilizzare per il raggruppamento:

Numero di Cluster (gruppi)

Feature Selezionate:

Punteggio (Avg Score) Popolarità (Num. Recensioni) Verbosità (Lunghezza Recensioni) Posizione (Lat/Lng) Profilo Nazionalità

Esegui Segmentazione

5. Segmentazione Hotel (K-Means Clustering): Risultati

Analisi delle caratteristiche di ogni cluster

- Es. Clustering con tutte le feature:

 Analisi dei Gruppi Identificati

	Cluster Name	Count	Avg_Score	Total_Reviews	Avg_Pos_Words	Avg_Neg_Words
0	Cluster 0	207	8.50	1425	19.5	18.3
1	Cluster 1	844	8.50	800	18.8	17.2
2	Cluster 2	297	8.43	1418	18.4	17.5
3	Cluster 3	127	8.28	4124	17.2	19.6

- Es. Clustering solo per punteggio:

Punteggio (Avg Score) Popolarità (Num. Recensioni) Verbosità (Lunghezza Recensioni) Posizione (Lat/Lng) Profilo Nazionalità 

Esegui Segmentazione

 Analisi dei Gruppi Identificati

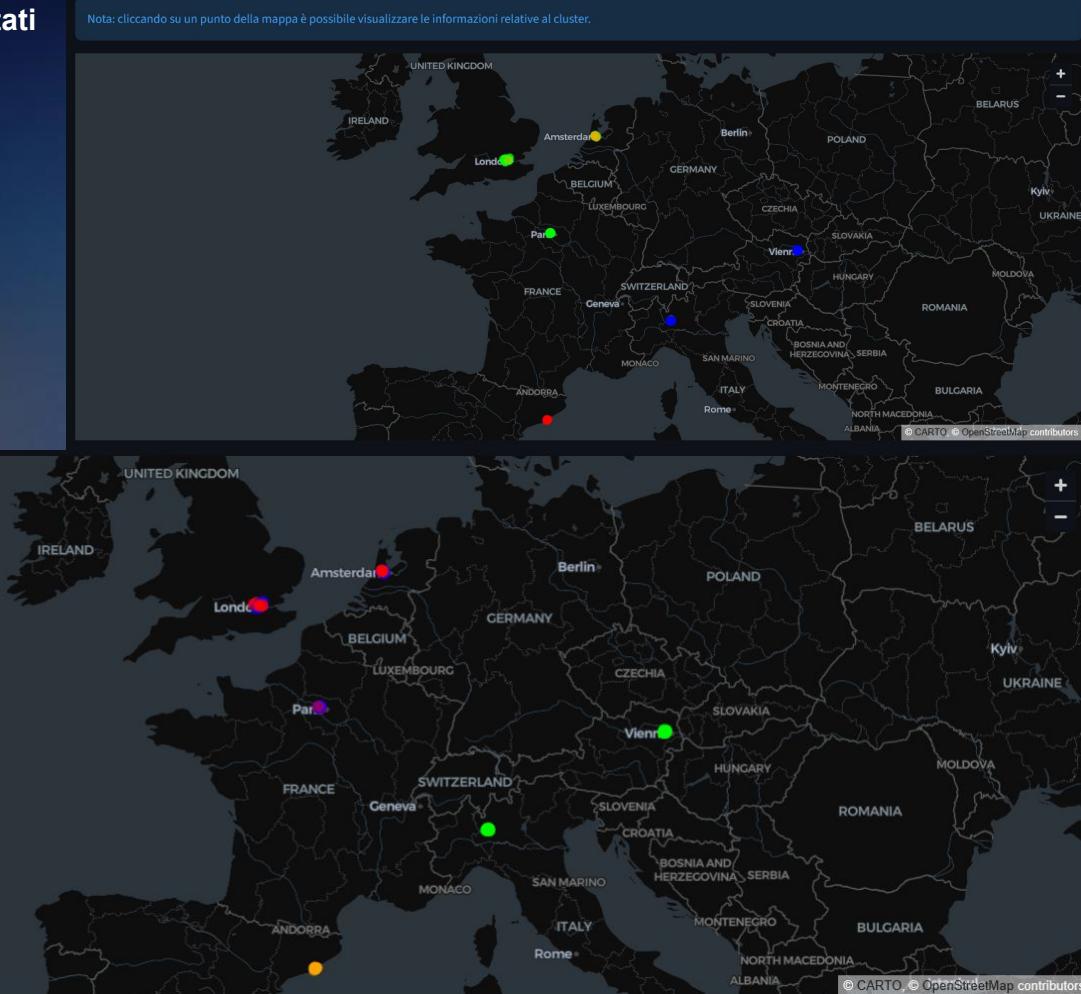
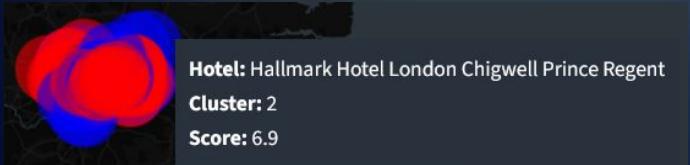
	Cluster Name	Count	Avg_Score	Total_Reviews	Avg_Pos_Words	Avg_Neg_Words
0	Cluster 0	618	8.35	1472	17.8	18.4
1	Cluster 1	579	9.00	1027	21.3	14.4
2	Cluster 2	39	6.97	1676	13.3	27.7
3	Cluster 3	256	7.77	1430	15.5	21.5

Visualizza come i cluster sono distribuiti geograficamente.

5. Segmentazione Hotel (K-Means Clustering): Risultati

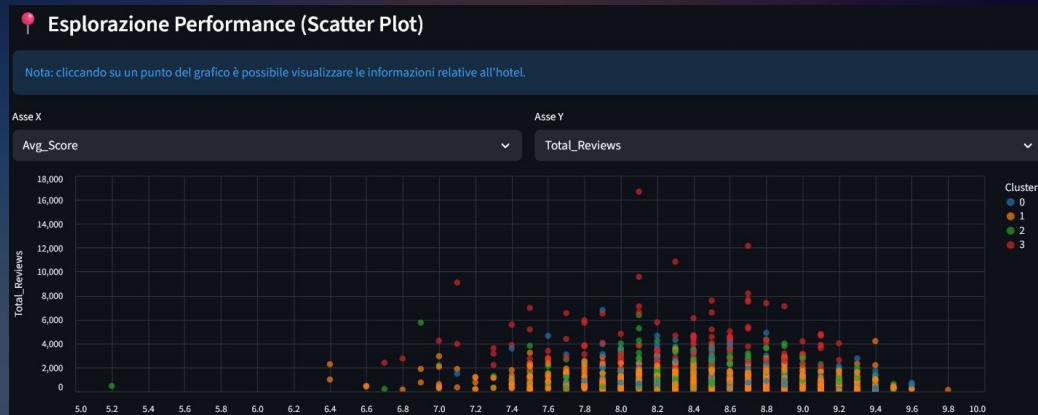
Mappa interattiva della distribuzione geografica degli hotel, colorati per cluster di appartenenza.

- Mappa 1: clustering con tutte le feature
- Mappa 2: clustering solo per punteggio e posizione



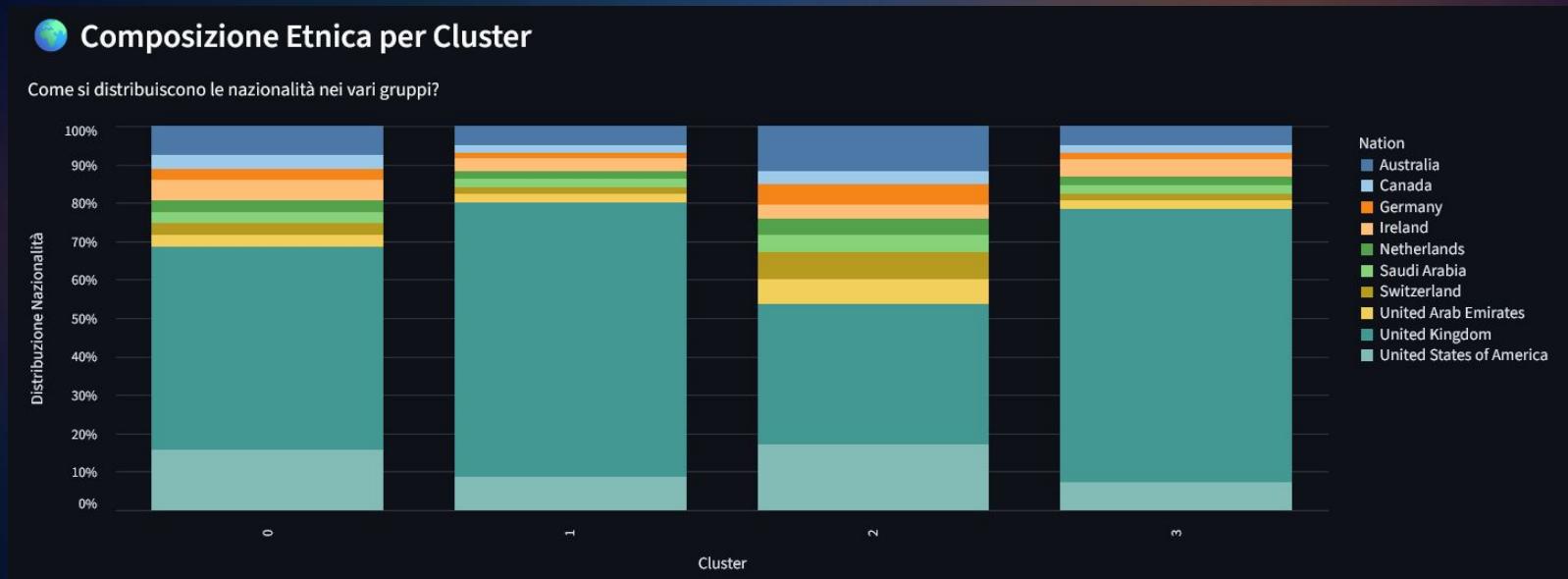
5. Segmentazione Hotel (K-Means Clustering): Risultati

- Grafico interattivo per l'esplorazione delle performance dei cluster, con possibilità di confronto.



5. Segmentazione Hotel (K-Means Clustering): Risultati

- Grafico della distribuzione delle nazionalità nei cluster (es. clustering con tutte le feature)



6. Migliori Hotel per Nazione

Obiettivo: Identificare le eccellenze alberghiere suddivise per paese.

Casi d'uso: Utenti che cercano i migliori hotel in una specifica destinazione turistica.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore



Hotel Reviews Analytics con Spark

Query selezionata: Migliori Hotel per Nazione

Questa query consente di individuare i **migliori hotel per ogni nazione**. Il criterio di ranking utilizzato è il **punteggio medio** ottenuto nelle recensioni dei clienti, inoltre in caso di parità si predilige l'hotel con il **numero totale di recensioni** più elevato.

Numero di migliori hotel da visualizzare per ogni nazione:

3

- +

[Analizza per Nazione](#)

6. Migliori Hotel per Nazione: Risultati

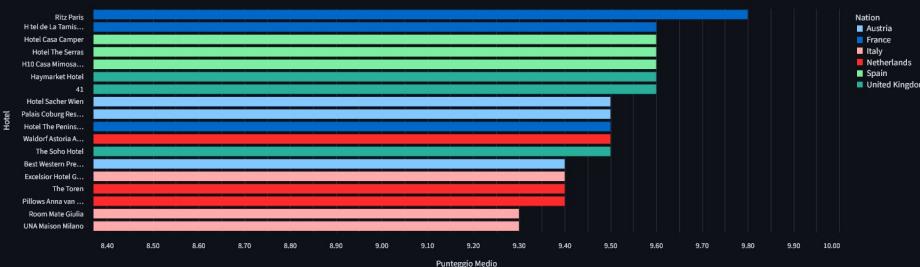
- Top N hotel per ogni nazione presente nel dataset (ordinamento alfabetico per nazione, modificabile dall'utente)
- Grafico della distribuzione dei punteggi
- Mappa geografica interattiva che mostra la posizione degli hotel



Top 3 Hotel per Nazione

Nation	Hotel_Name	Average_Score	Total_Number_of_Reviews	Hotel_Address	lat	lng
0 Austria	Hotel Sacher Wien	9.5	632	Philharmoniker Stra e 4 O1 Innere Stadt 1010 Vienna Austria	48.2039	16.3698
1 Austria	Palais Coburg Residenz	9.5	98	Coburgbastei 4 O1 Innere Stadt 1010 Vienna Austria	48.2059	16.3677
2 Austria	Best Western Premier Kaiserhof Wien	9.4	1353	Frankenberggasse 10 O4 Winden 1040 Vienna Austria	48.1975	16.3696
3 France	Ritz Paris	9.8	122	15 Place Vend me 1 arr 75001 Paris France	48.8679	2.329
4 France	H tel de La Tamise Esprit de France	9.6	166	4 rue d Alger 1st arr 75001 Paris France	48.8649	2.3296
5 France	Hotel The Peninsula Paris	9.5	275	19 avenue Kleber 16th arr 75116 Paris France	48.8711	2.2934
6 Italy	Excelsior Hotel Gallia Luxury Collection Hotel	9.4	1345	Piazza Duca D'Aosta 9 Central Station 20124 Milan Italy	45.4857	9.202
7 Italy	Room Mate Giulia	9.3	2011	Silvio Pellico 4 Milan City Center 20121 Milan Italy	45.4651	9.1885
8 Italy	UNA Maison Milano	9.3	1108	Via Mazzini 4 Milan City Center 20123 Milan Italy	45.4633	9.1884
9 Netherlands	Waldorf Astoria Amsterdam	9.5	443	Herengracht 542 556 Amsterdam City Center 1017 CG Amsterdam Netherlands	52.3648	4.8969

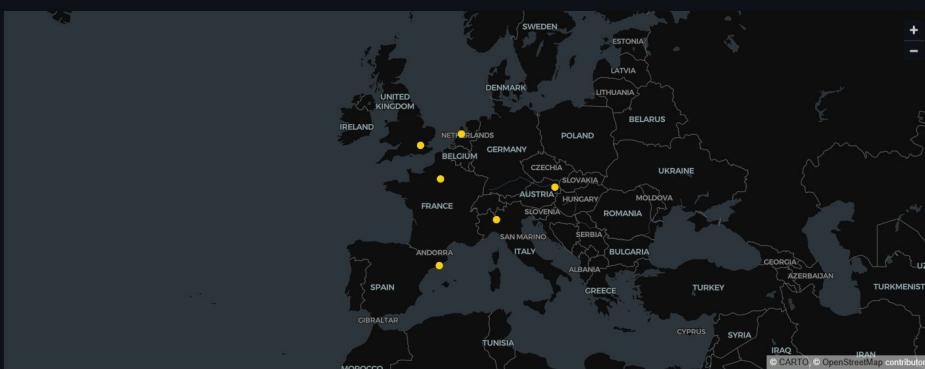
Distribuzione dei Punteggi degli Hotel



Mappa dei Migliori Hotel

Nota: cliccando su un punto della mappa è possibile visualizzare le informazioni relative all'hotel.

Visualizzazione geografica di 18 hotel (quelli con coordinate valide).



7. Locals vs Tourists - Preferenze e distribuzione clienti

Obiettivo: Analizzare la differenza di percezione tra chi visita il proprio paese (Local) e chi viene dall'estero (Tourist).

Casi d'uso: Identificazione di "Trappole per turisti" o "Preferiti dai locali" (hotel che potrebbero offrire un'esperienza più autentica).

Opzioni Query

Scgli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Locals vs Tourists - Preferenze e distribuzione clienti

Locals' Favorites vs Tourist Traps

Analizza come cambia la percezione degli hotel tra chi è del posto (Locals) e i turisti internazionali (Tourists).

- Locals (Recensori Locali): Recensori che hanno la stessa nazionalità della nazione in cui si trova l'hotel.
- Tourists (Recensori Turisti): Recensori di nazionalità diversa.

Indicatori:

- Local_Avg_Score: Voto medio dato dai locali.
- Tourist_Avg_Score: Voto medio dato dai turisti.
- Preference_Delta (Local - Tourist):
 - Valori POSITIVI (> 0): L'hotel piace più ai Locali (possibile "Gemma Nasosta" o esperienza autentica).
 - Valori NEGATIVI (< 0): L'hotel piace più ai Turisti (possibile "Tourist Trap" o standard internazionale generico).

Distribuzione Nazionalità

Visualizza graficamente la composizione delle nazionalità dei visitatori per ciascun hotel.

Minimo recensioni per gruppo (Local & Tourist)

20

Analizza Preferenze

Hotel preferiti dai Locali (Rispetto ai Turisti)

Hotel dove il voto dei locali supera di più quello dei turisti.

Hotel_Name	Hotel_Nation	Local_Avg_Score	Tourist_Avg_Score	Preference_Delta	Top_Nationalities
0 Simply Rooms Suites	United Kingdom	7.91	6.67	1.24	Spain (2%), United Kingdom (78%)
1 London Suites	United Kingdom	7.62	6.74	0.89	Italy (2%), France (2%), Austria (3%), United Kingdom (65%), Australia (4%)
2 Crowne Plaza London Docklands	United Kingdom	8.88	8.04	0.85	United Kingdom (8%), Ireland (2%)
3 Sheraton Grand London Park Lane	United Kingdom	8.44	7.61	0.83	Turkey (2%), United Arab Emirates (6%), Kuwait (5%), Saudi Arabia (7%), Australia (5%), United Kingdom (80%)
4 DoubleTree by Hilton London West End	United Kingdom	8.43	7.63	0.80	United Kingdom (80%)
5 The Westbury A Luxury Collection Hotel Mayfair London	United Kingdom	8.49	7.72	0.77	United Arab Emirates (2%), United States of America (2%), South Africa (2%), Kuwait (4%), Saudi Arabia (4%), United Kingdom (11%), United States of America (65%), Israel (2%), Australia (2%), United States of America (5%)
6 DoubleTree by Hilton Hotel London Marble Arch	United Kingdom	8.43	7.67	0.76	Saudi Arabia (4%), Kuwait (5%), Israel (4%), Ireland (2%), Australia (2%), United States of America (5%)
7 The Dorchester Dorchester Collection	United Kingdom	9.52	8.77	0.75	Saudi Arabia (4%), Kuwait (4%), United Arab Emirates (5%), United States of America (6%), United Kingdom (11%), United States of America (5%)
8 Holiday Inn London Oxford Circus	United Kingdom	7.70	6.96	0.75	Kuwait (3%), United Arab Emirates (2%), United Kingdom (57%), India (2%), Ireland (2%), Israel (2%)
9 Novotel London Paddington	United Kingdom	8.80	8.06	0.73	United States of America (5%), Ireland (2%), United Kingdom (51%), Saudi Arabia (3%), New Zealand (2%)

Hotel preferiti dai Turisti (Rispetto ai Locali)

Hotel dove il voto dei turisti supera di più quello dei locali.

Hotel_Name	Hotel_Nation	Local_Avg_Score	Tourist_Avg_Score	Preference_Delta	Top_Nationalities
458 Austria Trend Hotel Bosel Wien	Austria	6.76	8.18	-1.41	Poland (4%), Croatia (2%), United Kingdom (6%), Bulgaria (2%), Israel (2%), Belgium (2%), Czech Republic (1%), United Kingdom (11%), Italy (2%), United States of America (6%), United Kingdom (4%)
457 Rosa Grand Milano Starhotels Collezione	Italy	7.39	8.46	-1.07	United Arab Emirates (8%), Israel (2%), United Kingdom (11%), Italy (2%), Kuwait (7%), Qatar (3%)
456 Hampshire Hotel The Manor Amsterdam	Netherlands	7.49	8.54	-1.05	Ireland (5%), Australia (3%), Netherlands (7%), United States of America (6%), United Kingdom (11%), United States of America (5%)
455 Arthotel ANA Boutique Six	Austria	7.11	8.10	-0.99	Russia (2%), Croatia (2%), Hungary (3%), Israel (2%), Italy (3%), Poland (2%), United Kingdom (11%)
454 Savoy Hotel Amsterdam	Netherlands	5.33	6.09	-0.76	France (2%), United States of America (4%), Germany (4%), Netherlands (10%), Israel (2%), United Kingdom (11%)
453 The Pillar Hotel	United Kingdom	8.15	8.88	-0.73	United Kingdom (66%), Israel (18%), United States of America (3%)
452 Eurostars Grand Marina Hotel GL	Spain	7.74	8.47	-0.73	Saudi Arabia (9%), United Kingdom (33%), Kuwait (4%), Australia (2%), Ireland (2%), Spain (2%)
451 Catalonia Atenas	Spain	7.43	8.09	-0.66	Ireland (6%), United Kingdom (48%), Spain (2%), United States of America (5%), Netherlands (2%)
450 Courtyard by Marriott Amsterdam Arena Atlas	Netherlands	7.95	8.57	-0.62	United Arab Emirates (2%), United States of America (5%), Turkey (2%), Germany (6%), Netherlands (2%)
449 Ramada Apollo Amsterdam Centre	Netherlands	7.71	8.31	-0.60	Netherlands (2%), United States of America (5%), Australia (2%), United Kingdom (49%), Ireland (1%)

7. Locals vs Tourists: Risultati

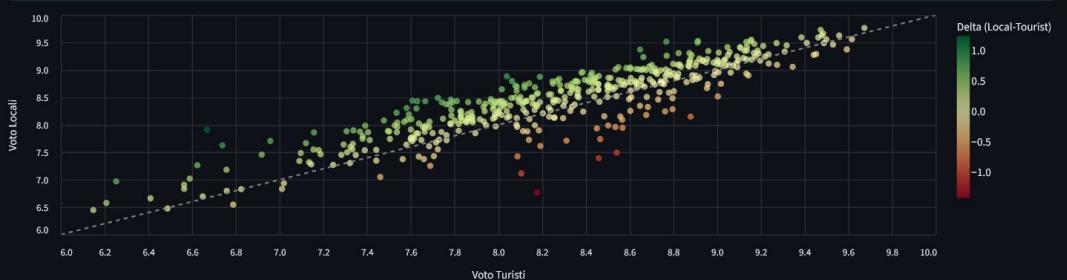
- Top hotel preferiti dai locali (Preference Delta positivo) e dai turisti (Preference Delta negativo), ordinati in modo da visualizzare per primi gli hotel che hanno le maggiori discrepanze di voto
- Grafico Discrepanza Voti (Local vs Tourist)

Grafico: Discrepanza Voti

Confronto diretto tra Voto Local (Y) e Voto Tourist (X).

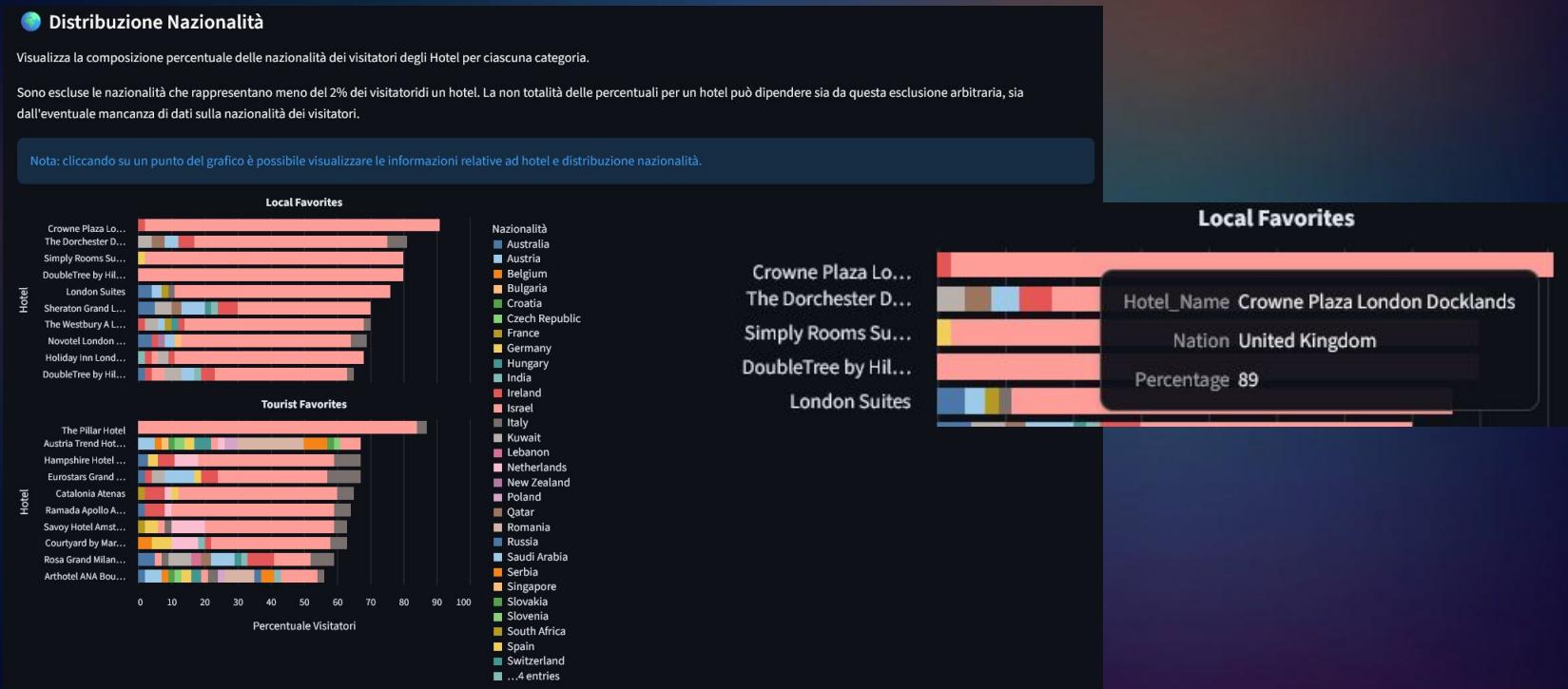
- Punti Sopra la diagonale: Meglio per i Locali.
- Punti Sotto la diagonale: Meglio per i Turisti.

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative all'hotel.



7. Locals vs Tourists: Risultati

- Grafico Distribuzione Nazionalità per visualizzare la composizione percentuale delle nazionalità dei visitatori degli Hotel



8. Analisi Stagionale & Target

Obiettivo: Analizzare come varia il gradimento in base alla stagione e al tipo di viaggio (Leisure/Business/Solo/Family/Group/Couple).

Casi d'uso: Analisi delle Performance degli hotel in specifiche stagioni (es. hotel ottimi per l'inverno ma carenti in estate), analisi dell'apprezzamento degli hotel da parte di specifiche tipologie di viaggiatori, pianificazione di viaggi in base alle esigenze dell'utente.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Analisi Stagionale & Target

  **Analisi Stagionale e Profilazione Viaggiatore**

Scopri quando andare e con chi. Questa query analizza come cambiano le performance degli hotel in base alla **Stagione** e al **Tipo di Viaggiatore**.

- **Season:** Inverno, Primavera, Estate, Autunno (derivata dalla data delle recensioni).
- **Traveler Type:** Families, Couples, Business, Solo, Group (derivato dai Tag).

Selezione Stagioni:

- Winter
- Spring
- Summer
- Autumn

Selezione Tipo Viaggiatore:

- Family
- Couple
- Business
- Solo
- Group
- Leisure

Minimo recensioni per includere un hotel nella ricerca:

10

← →

Analiza Stagionalità

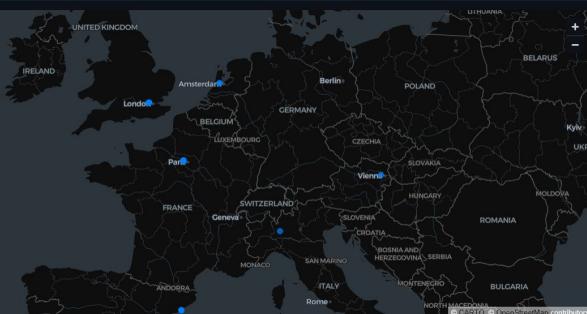
8. Analisi Stagionale & Target: Risultati

- Lista degli hotel migliori in base alle stagioni e alle tipologie di viaggio selezionate
- Mappa interattiva che mostra la posizione geografica dei migliori hotel corrispondenti ai criteri della ricerca



Mappa dei migliori risultati (Winter, Spring, Summer, Autumn - Family, Couple, Business, Solo, Group, Leisure)

Nota: cliccando su un punto della mappa è possibile visualizzare le informazioni relative all'hotel.



🏆 Top Hotel: Summer & Couple, Leisure

I migliori hotel per la combinazione selezionata.

Hotel_Name	Season	Traveler_Type	Avg_Score	Review_Count	Nation
6 Hotel Le Six	Summer	Couple	9.85	11	France
11 Pillows Anna van den Vondel Amsterdam	Summer	Leisure	9.84	10	Netherlands
41 Hotel Casa Camper	Summer	Couple	9.79	56	Spain
47 Hotel Sacher Wien	Summer	Couple	9.78	22	Austria
87 Covent Garden Hotel	Summer	Leisure	9.74	14	United Kingdom
92 Hotel Casa Camper	Summer	Leisure	9.74	88	Spain
105 Hotel Eiffel Blomet	Summer	Leisure	9.72	12	France
114 Lansbury Heritage Hotel	Summer	Couple	9.72	10	United Kingdom
115 The Beaumont Hotel	Summer	Leisure	9.72	20	United Kingdom
116 Waldorf Astoria Amsterdam	Summer	Leisure	9.72	36	Netherlands

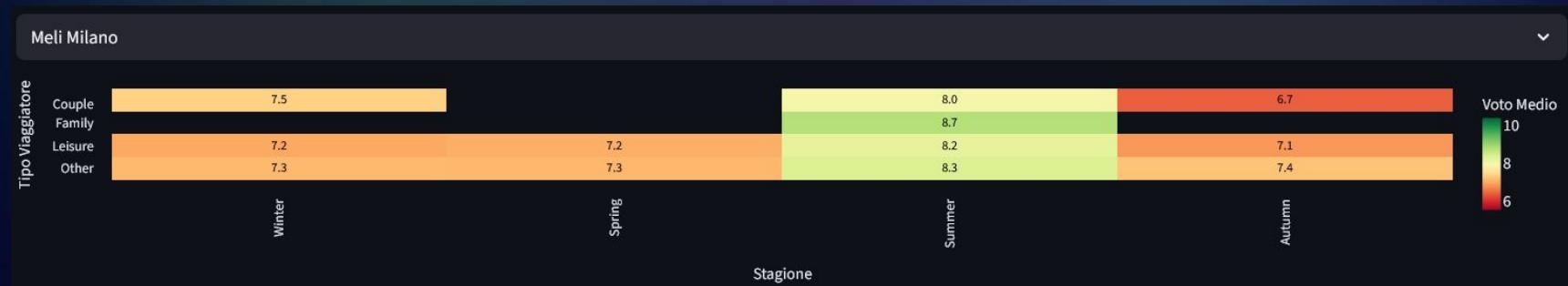
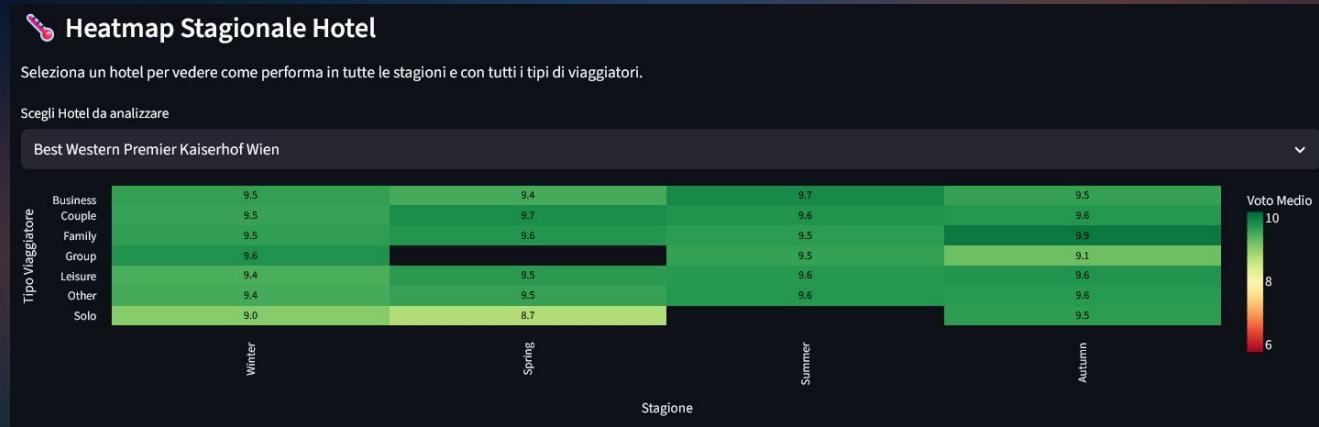
🏆 Top Hotel: Winter, Spring, Summer, Autumn & Family, Couple, Business, Solo, Group, Leisure

I migliori hotel per la combinazione selezionata.

Hotel_Name	Season	Traveler_Type	Avg_Score	Review_Count	Nation
2 Hotel The Peninsula Paris	Autumn	Couple	9.88	13	France
3 Banks Mansion All Inclusive Hotel	Winter	Group	9.87	12	Netherlands
4 Le Narcisse Blanc Spa	Winter	Couple	9.87	18	France
6 Hotel Le Six	Summer	Couple	9.85	11	France
8 Best Western Premier Kaiserhof Wien	Autumn	Family	9.85	11	Austria
9 Hotel de La Tamise Esprit de France	Winter	Couple	9.85	13	France
10 Hotel Spadari Al Duomo	Winter	Family	9.84	10	Italy
11 Pillows Anna van den Vondel Amsterdam	Summer	Leisure	9.84	10	Netherlands
12 H10 Casa Mimosa 4 Sup	Winter	Solo	9.84	10	Spain
14 The Wellesley Knightsbridge a Luxury Collection Hotel Londra	Winter	Leisure	9.83	12	United Kingdom

8. Analisi Stagionale & Target: Risultati

- Heatmap Stagionale Hotel per visualizzare dettagliatamente le performance di un hotel in ogni possibile combinazione di stagione e tipologia di viaggio



9. Analisi Durata Soggiorno

Obiettivo: Analizzare la correlazione tra la durata del soggiorno (numero di notti) e il livello di soddisfazione dei clienti.

Casi d'uso: Ottimizzazione delle offerte per soggiorni lunghi e corti, grazie a statistiche che mostrano se i soggiorni lunghi tendono ad avere recensioni peggiori o migliori.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Analisi Durata Soggiorno

Analisi dell'impatto della Durata del Soggiorno

Analizza come varia la soddisfazione dei clienti in base alla durata del loro soggiorno. Le durate sono categorizzate in:

- Short Stay: 1-3 notti
- Medium Stay: 4-7 notti
- Long Stay: 8+ notti

Questa analisi aiuta a capire quali hotel sono più adatti a toccate e fuga o a lunghe vacanze.

Minimo recensioni per categoria/hotel

20

Analizza Soggiorni

9. Analisi Durata Soggiorno: Risultati

- **Score Delta:** indicatore globale delle performance (indica quanto il punteggio medio di ogni categoria di recensioni si discosta dalla media globale)

Es. Se le recensioni di categoria "Long Stay" hanno un delta negativo, significa che, mediamente, i clienti che soggiornano a lungo negli hotel sono più insoddisfatti rispetto al gradimento medio globale.

- Classifica dei **Top 5 hotel per ogni categoria di soggiorno**, ordinati per punteggio medio (lo stesso hotel può apparire in più classifiche)



9. Analisi Durata Soggiorno: Risultati

- Grafico della distribuzione degli hotel in relazione a: **Durata media dei soggiorni dei clienti vs Punteggio medio dell'hotel.**

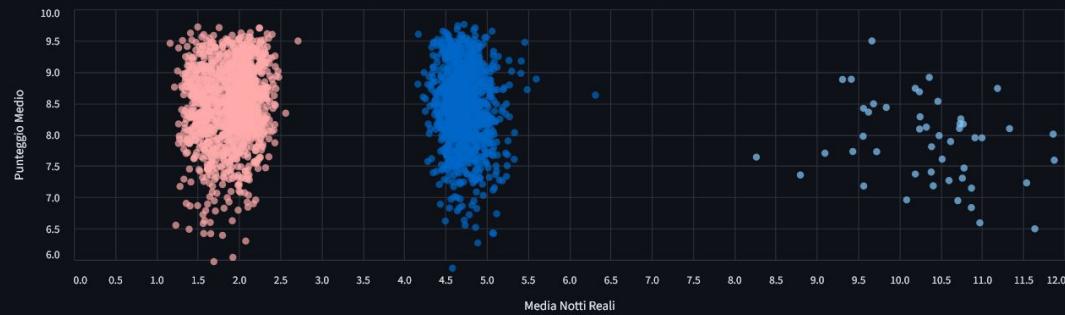
Da questa rappresentazione è possibile valutare quali sono le durate dei soggiorni più frequenti e come sono correlate con la soddisfazione dei clienti.

Es. Gli hotel con clienti che fanno soggiorni lunghi sono meno frequenti e le loro valutazioni e il loro punteggio è spesso inferiore alla media.

Dettagli Performance Hotel

Ogni punto è un hotel in una specifica categoria di durata.

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative all'hotel.



Stay_Category
● Long Stay (8+)
● Medium Stay (4-7)
● Short Stay (1-3)



10. Analisi Esperienza Recensore

Obiettivo: Valutare se il punteggio delle recensioni degli hotel cambia in base all'esperienza dei recensori (ad esempio se i recensori esperti sono più critici dei novizi).

Casi d'uso: Analisi della severità del voto in base dell'esperienza, in funzione di una ponderazione del peso delle recensioni in un sistema di ranking avanzato.

Opzioni Query

Scegli la query da eseguire:

- Trend Recensioni (Time Series)
- Analisi Influenza Tag
- Analisi Bias Nazionalità
- Analisi Competitività Locale
- Segmentazione Hotel (K-Means)
- Migliori Hotel per Nazione
- Locals vs Tourists - Preferenze e distribuzione clienti
- Analisi Stagionale & Target
- Analisi Durata Soggiorno
- Analisi Esperienza Recensore

Hotel Reviews Analytics con Spark

Query selezionata: Analisi Esperienza Recensore

L'Esperienza conta? Novizi vs Esperti

Analizza se il punteggio delle recensioni degli hotel cambia in base all'esperienza dei recensori.

I recensori sono considerati più o meno esperti in base al numero di recensioni scritte, venendo classificati nelle seguenti categorie:

- Novice (< 5 recensioni): novizio, viaggiatore occasionale.
- Intermediate (5-25 recensioni): viaggiatore regolare.
- Expert (> 25 recensioni): critico o frequent flyer.

Per ogni hotel, viene calcolato l'**Experience Gap (Expert - Novice)**, ovvero la differenza tra il punteggio medio assegnato dagli esperti e quello assegnato dai novizi:

- Valore Positivo: gli esperti danno voti più alti dei novizi (è un buon segnale, perché le recensioni dei viaggiatori esperti sono considerate più affidabili).
- Valore Negativo: gli esperti sono più critici dei novizi (potrebbe essere indice di un hotel che delude le aspettative più elevate).

Minimo recensioni per livello di esperienza

10

Analizza Esperienza Recensore

10. Analisi Esperienza Recensore: Risultati

- Gap Medio: differenza fra il voto medio assegnato dagli esperti e il voto medio assegnato dai novizi. Costituisce la metrica di riferimento per valutare quanto in media i recensori esperti sono più critici dei novizi, o viceversa.
- Classifiche degli hotel con la maggior discrepanza (Experience Gap) fra la valutazione degli esperti e quella dei novizi.

Con questi confronti è possibile valutare quanto in media l'esperienza dei recensori impatta sulla valutazione degli hotel e quanto sono concordi o discordi le opinioni di recensori con diversi livelli di esperienza.

Gap Medio (Esperti - Novizi): se negativo, gli esperti sono mediamente più severi dei novizi

-0.05

👉 I più criticati dagli Esperti (Gap Negativo)

Hotel_Name	Novice_Avg_Score	Intermediate_Avg_Score	Expert_Avg_Score	Experience_Gap	Total_Analyzed_Reviews	Novice_Count	Intermediate_Count	Expert_Count
0 Hotel Les Bulles De Paris	8.44	8.24	6.67	-1.78	159	67	80	12
1 Best Western Blue Tower Hotel	7.40	7.30	6.19	-1.21	845	582	236	27
2 Mercure Paris Montmartre Sac Coeur	8.27	8.46	7.09	-1.18	269	174	85	10
3 Hilton Diagonal Mar Barcelona	7.80	7.70	6.71	-1.09	140	80	50	10
4 Hilton London Euston	7.18	7.05	6.11	-1.07	470	337	122	11
5 Holiday Inn London West	8.04	7.89	6.98	-1.06	701	526	161	14
6 Hotel Sans Souci Wien	9.61	9.53	8.59	-1.03	235	120	98	17
7 Le Dokhan s a Tribute Portfolio Hotel	8.12	8.44	7.10	-1.02	68	37	20	11
8 Ilunion Almirante	7.08	7.06	6.09	-0.98	264	127	121	16
9 Hotel Serhs Rivoli Rambla	8.01	8.09	7.04	-0.97	357	207	133	17

👉 I più apprezzati dagli Esperti (Gap Positivo)

Hotel_Name	Novice_Avg_Score	Intermediate_Avg_Score	Expert_Avg_Score	Experience_Gap	Total_Analyzed_Reviews	Novice_Count	Intermediate_Count	Expert_Count
838 Hotel Cavendish	6.39	6.47	7.49	1.10	920	627	272	21
837 Banke Hotel	8.48	8.20	9.53	1.05	159	86	61	12
836 Hotel Alimara	7.84	7.70	8.88	1.04	195	114	71	10
835 Radisson Blu Champs Elyses Paris	7.24	7.47	8.18	0.94	152	87	55	10
834 Hilton London Green Park	7.21	7.38	8.14	0.93	484	346	126	12
833 Hotel Tiziano Park Vita Parcour Gruppo MiniHotels	7.77	8.17	8.67	0.90	61	27	24	10
832 Sloane Square Hotel	8.19	8.25	9.05	0.86	462	269	170	23
831 Grosvenor House Suites by Jumeirah Living	8.41	8.51	9.22	0.81	100	30	60	10
830 BEST WESTERN Maitrise Hotel Maida Vale	6.78	6.99	7.58	0.80	816	501	284	31
829 FourSide Hotel Vienna City Center	7.51	7.99	8.29	0.79	173	52	97	24

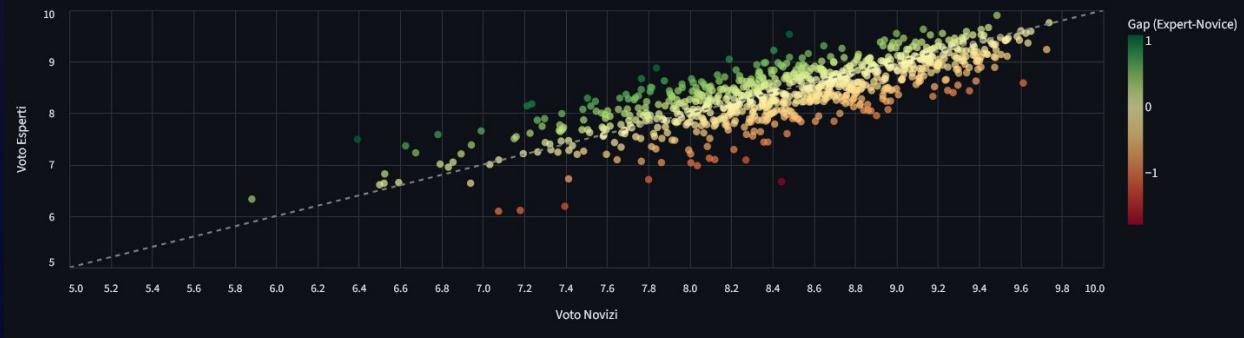
10. Analisi Esperienza Recensore: Risultati

- Grafico della distribuzione degli hotel in relazione al voto ottenuto dai novizi e dagli esperti,
Consente di valutare visivamente quanto sono frequenti gli hotel con alta o bassa discrepanza di voti.

Grafico: Novizi vs Esperti

Confronto voti: Novizi (X) vs Esperti (Y).

Nota: cliccando su un punto del grafico è possibile visualizzare le informazioni relative all'hotel.



4	Hotel_Name	Hotel Les Bulles De Paris
	Novice_Avg_Score	8.4447761194
	Expert_Avg_Score	6.66666666667
	Experience_Gap	-1.77810945274
	Total_Analyzed_Reviews	159

4	Hotel_Name	Hotel Cavendish
	Novice_Avg_Score	6.39314194577
	Expert_Avg_Score	7.49047619048
	Experience_Gap	1.0973342447
	Total_Analyzed_Reviews	920

Possibili Sviluppi Futuri

- **Analisi del Testo Avanzata:** implementare modelli NLP per Sentiment Analysis sulle recensioni testuali.
- **Streaming:** integrare Spark Streaming per elaborare recensioni in tempo reale.
- **Recommendation System:** sviluppare un sistema di raccomandazione basato sulla similarità utente-utente

Grazie per l'attenzione

HOTEL