Queries on booking dataset

Sample json looks like this:

{

"pageurl": "https://www.booking.com/hotel/in/treebo-trip-daisey-dee.en-gb.html?label=gen173nr-1DCAQoggJCDHNlYXJjaF9jb29yZ0gJWARolQKIAQGYAQm4ARjIAQ\_YAQPoAQH4AQOIAgGoAgS4ArzlsvMFwAIB&sid=b369ec9535ceb9bef18e93d6d1377489&all\_sr\_blocks=513689101\_179845172\_2\_1\_0&checkin=2020-03-16&checkout=2020-03-17&dest\_id=7467&dest\_type=region&group\_adults=2&group\_children=0&hapos=27&highlighted\_blocks=513689101\_179845172\_2\_1\_0&hpos=2&no\_rooms=1&sr\_order=popularity&sr\_pri\_blocks=513689101\_179845172\_2\_1\_0\_\_119461&srepoch=1584181959&srpvid=ce584a239277011d&ucfs=1&from=searchresults;highlight\_room=&;selected\_currency=INR;changed\_currency=1;top\_currency=1",

"record": {

"uniq\_id": "0ef15af30235dc2a267ec37ff9c5c0ee",

"hotel\_id": "5136891",

"hotel\_name": "Treebo Trip Daisey Dee",

"review\_count": "27",

"rating\_count": "8.8",

"default\_rank": "30",

"price\_rank": "37",

"ota": "booking.com",

"room\_type": [

{

"room\_type\_name": "Standard Double Room",

"room\_type\_price": 1338,

"room\_type\_occupancy": 2,

"room\_type\_breakfast": "breakfast",

"room\_type\_cancellation": "free\_cancellation",

"availability": [

{

"from": "2025-01-20",

"to": "2025-01-25"

},

{

"from": "2025-01-12",

"to": "2025-01-14"

},

{

"from": "2025-01-01",

"to": "2025-01-07"

},

{

"from": "2025-01-28",

"to": "2025-01-31"

},

{

"from": "2025-01-11",

"to": "2025-01-18"

}

]

},

{

"room\_type\_name": "Standard Double Room",

"room\_type\_price": 1136,

"room\_type\_occupancy": 1,

"room\_type\_breakfast": "breakfast",

"room\_type\_cancellation": "free\_cancellation",

"availability": [

{

"from": "2025-02-06",

"to": "2025-02-11"

},

{

"from": "2025-02-04",

"to": "2025-02-09"

},

{

"from": "2025-02-02",

"to": "2025-02-05"

}

]

},

{

"room\_type\_name": "Deluxe Double Room",

"room\_type\_price": 1640,

"room\_type\_occupancy": 2,

"room\_type\_breakfast": "breakfast",

"room\_type\_cancellation": "free\_cancellation",

"availability": [

{

"from": "2025-01-18",

"to": "2025-01-23"

},

{

"from": "2025-01-16",

"to": "2025-01-22"

}

]

},

{

"room\_type\_name": "Deluxe Double Room",

"room\_type\_price": 1338,

"room\_type\_occupancy": 1,

"room\_type\_breakfast": "breakfast",

"room\_type\_cancellation": "free\_cancellation",

"availability": [

{

"from": "2025-01-25",

"to": "2025-02-01"

},

{

"from": "2025-01-04",

"to": "2025-01-11"

},

{

"from": "2025-01-21",

"to": "2025-01-27"

},

{

"from": "2025-01-28",

"to": "2025-02-01"

},

{

"from": "2025-01-28",

"to": "2025-01-29"

},

{

"from": "2025-02-18",

"to": "2025-02-20"

},

{

"from": "2025-02-10",

"to": "2025-02-14"

},

{

"from": "2025-02-23",

"to": "2025-03-01"

}

]

}

],

"checkin\_date": "2020-03-16",

"crawled\_date": "2020-03-14 10:59:45 +0000",

"city": "Paris"

}

}

**1. Search**

**Business Question 1**: Retrieve all accommodations with a nightly price between $100 and $200.

**CouchDB Query**:

{

"selector": {

"record.room\_type": {

"$elemMatch": {

"room\_type\_price": {

"$gte": 100,

"$lte": 200

}}}}}

From terminal:

time curl -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {

"record.room\_type": {

"$elemMatch": {

"room\_type\_price": {

"$gte": 100,

"$lte": 200

}

}

}

},

"limit": 1000

}

'

**Couchbase Query**:

SELECT \*

FROM accommodations

WHERE ANY room IN record.room\_type SATISFIES room.room\_type\_price BETWEEN 100 AND 200 END;

To run from terminal:

cbq -u Administrator -p password -s "SELECT \*

FROM booking.\_default.\_default

WHERE ANY room IN record.room\_type SATISFIES room.room\_type\_price BETWEEN 100 AND 200 END;"

**Business Question 2**: Find accommodations with an average rating of 4.5 or higher.

**CouchDB Query**:  
 {

  "selector": {

    "record.rating\_count": { "$gte": 4.5 }

  }

}

From terminal:

time curl -s -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {

"record.rating\_count": {

"$gte": "4.5"

}

},

"limit": 20000

}' | jq '.docs | length'

**Couchbase Query**:

SELECT \*

FROM \_default

WHERE TO\_NUMBER(record.rating\_count) >= 4.5;

From terminal:

cbq -u Administrator -p password -s "SELECT \*

FROM booking.\_default.\_default

WHERE TO\_NUMBER(record.rating\_count) >= 4.5;"

**Business Question 3**: Search for accommodations that are available between "2024-01-01" and "2024-01-15".

**CouchDB Query**:

{ "selector": {

"record.room\_type": {

"$elemMatch": {

"availability": {

"$elemMatch": {

"from": {

"$lte": "2025-01-15"

},

"to": {

"$gte": "2025-01-01"

}}}}}}}

From terminal:

time curl -s -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {

"record.room\_type": {

"$elemMatch": {

"availability": {

"$elemMatch": {

"from": {

"$lte": "2025-01-15"

},

"to": {

"$gte": "2025-01-01"

} }}}}

}, "limit":40000

}'

**Couchbase Query**:

SELECT \*

FROM \_default

WHERE ANY room IN record.room\_type SATISFIES

ANY avail IN room.availability

SATISFIES avail.`from` <= "2025-01-15" AND avail.`to` >= "2025-01-01"

END

END;

From terminal:

cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT \*

FROM booking.\_default.\_default

WHERE ANY room IN record.room\_type SATISFIES

ANY avail IN room.availability

SATISFIES avail.\`from\` <= '2025-01-15' AND avail.\`to\` >= '2025-01-01'

END

END;

"

**2. Page**

**Business Question 1**: Show the first 10 accommodations sorted by price.

**CouchDB Query**:

Query doesn’t work unless we create an index :

{

"index": {

"fields": ["record.room\_type.0.room\_type\_price"]

},

"name": "room\_type\_price\_index",

"type": "json"

}

Then:

{

"selector": {},

"sort": [

{ "record.room\_type.0.room\_type\_price": "asc" }

],

"limit": 10

}

From terminal:

time curl -s -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {},

"sort": [

{ "record.room\_type.0.room\_type\_price": "asc" }

],

"limit": 10

}'

**Couchbase Query**:

SELECT \*

FROM \_default

ORDER BY record.room\_type[0].room\_type\_price ASC

LIMIT 10;

From terminal:

cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT \*

FROM booking.\_default.\_default

ORDER BY record.room\_type[0].room\_type\_price ASC

LIMIT 10;

"

**Business Question 2**: Display the second page of accommodations with 5 results per page.

**CouchDB Query**:

Also needs an index (using the same previous one)

{

"selector": {},

"sort": [

{

"record.room\_type.0.room\_type\_price": "asc"

}

],

"limit": 5,

"skip": 5

}

time curl -s -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {},

"sort": [

{

"record.room\_type.0.room\_type\_price": "asc"

}

],

"limit": 5,

"skip": 5

}'

**Couchbase Query**:

SELECT \*

FROM \_default

ORDER BY record.room\_type[0].room\_type\_price ASC

LIMIT 5 OFFSET 5;

From Terminal:

cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT \*

FROM booking.\_default.\_default

ORDER BY record.room\_type[0].room\_type\_price ASC

LIMIT 5 OFFSET 5;

"

**4. Aggregate**

**Business Question 1**: Calculate the minimum nightly price of accommodations by city.

**CouchDB Query**: (Requires creating a view):

First, create map function and view  
  
curl -X PUT "http://Admin:password@localhost:5984/booking/\_design/aggregations" \

-H "Content-Type: application/json" \

-d '{

"\_id": "\_design/aggregations",

"views": {

"min\_price\_by\_city": {

"map": "function (doc) { if (doc.record && doc.record.room\_type && doc.record.city) { doc.record.room\_type.forEach(function(room) { if (room.room\_type\_price && !isNaN(parseFloat(room.room\_type\_price))) { emit(doc.record.city, parseFloat(room.room\_type\_price)); } }); } }",

"reduce": "\_stats"

}

}

}'

Then, get city aggregations

curl -X GET "http://Admin:password@localhost:5984/booking/\_design/aggregations/\_view/min\_price\_by\_city?reduce=true&group=true"

But, we get result for all stats (avg,max,min,..) not only what we need

**Couchbase Query**:

SELECT doc.record.city, min(room.room\_type\_price) AS min\_price

FROM \_default AS doc

UNNEST doc.record.room\_type AS room

GROUP BY doc.record.city;

In terminal:

time cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT doc.record.city, min(room.room\_type\_price) AS min\_price

FROM booking.\_default.\_default AS doc

UNNEST doc.record.room\_type AS room

GROUP BY doc.record.city;

"

**Business Question 2**: Count the number of accommodations available in each city.

**CouchDB Query**: (Requires creating a view):  
  
 Use a reduce function to sum the counts.

time curl -X PUT "http://Admin:password@localhost:5984/booking/\_design/city\_counts" \

-H "Content-Type: application/json" \

-d '{

"\_id": "\_design/city\_counts",

"views": {

"count\_by\_city": {

"map": "function(doc) { if (doc.record && doc.record.city) { emit(doc.record.city, 1); } }",

"reduce": "\_count"

}

}

}'

Then,

time curl -X GET "http://Admin:password@localhost:5984/booking/\_design/city\_counts/\_view/count\_by\_city?reduce=true&group=true"

**Couchbase Query**:  
  
SELECT record.city, COUNT(\*) AS accommodation\_count

FROM \_default

GROUP BY record.city;

In terminal:

cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT record.city, COUNT(\*) AS accommodation\_count

FROM booking.\_default.\_default

GROUP BY record.city;"

**Business Question 3**: Group accommodations by rating and count the total in each group.

**CouchDB Query**: (Requires creating a view):  
  
 Use a reduce function to sum the counts.

time curl -X PUT "http://Admin:password@localhost:5984/booking/\_design/rating\_counts" \

-H "Content-Type: application/json" \

-d '{

"\_id": "\_design/rating\_counts",

"views": {

"count\_by\_rating": {

"map": "function(doc) { var rating = doc.record && doc.record.rating\_count ? parseFloat(doc.record.rating\_count) : null; emit(rating, 1); }",

"reduce": "\_count"

}

}

}'

Then,

time curl -X GET "http://Admin:password@localhost:5984/booking/\_design/rating\_counts/\_view/count\_by\_rating?reduce=true&group=true"

**Couchbase Query**:

SELECT TO\_NUMBER(record.rating\_count) AS rating, COUNT(\*) AS rating\_count

FROM \_default

GROUP BY TO\_NUMBER(record.rating\_count);

In terminal:

time cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT TO\_NUMBER(record.rating\_count) AS rating, COUNT(\*) AS rating\_count

FROM booking.\_default.\_default

GROUP BY TO\_NUMBER(record.rating\_count);"

**5. Report**

**Business Question 1**: Generate a report of bookings made in December 2024.

**CouchDB Query**:  
{

"selector": {

"record.checkin\_date": {

"$gte": "2024-12-01",

"$lte": "2024-12-31"

} } }

In Terminal:

time curl -s -X POST http://Admin:password@localhost:5984/booking/\_find \

-H "Content-Type: application/json" \

-d '{

"selector": {

"record.checkin\_date": {

"$gte": "2024-12-01",

"$lte": "2024-12-31"

}

}, "limit":30000

}

'

**Couchbase Query**:  
 SELECT \*

FROM \_default

WHERE record.checkin\_date BETWEEN "2024-12-01" AND "2024-12-31";

In terminal:

time cbq -u Administrator -p password -e http://localhost:8093 -s "

SELECT \*

FROM booking.\_default.\_default

WHERE record.checkin\_date BETWEEN '2024-12-01' AND '2024-12-31';"