**THE CASE**

**Our business**: We are a building contractor company which constructs 5-star hotels for our client Hilton Hotels.

**Objective**: Propose to build a new Hilton Hotel in Netherlands or in the region close to this country.

**Conditions and requisites**:

* There should already exist 5-star hotels in the province and city where we will propose to construct. The hotels have to be aimed at high-end markets.
* The chosen city needs to have 3 working 5-star hotels.
* Hilton hotels aims for higher end-pricing than average 5-star hotels.
* Gather information regarding the potential competitors.
* Create a diagram which can help the client define a price per night for the new hotel.

**STEP 1: UPLOADING THE DATA FILES ONTO HDFS**

HDFS

Uploading the two files to HDFS (Main info file and stars file):

Graphical user interface, application

Description automatically generated

Graphical user interface, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated

**STEP 2: CREATING TABLES ON HIVE**

HIVE

Creating an external table in HIVE called “hotels”:

Graphical user interface, text, application

Description automatically generated

Testing the table was created:

Graphical user interface, application, table

Description automatically generated

Now we’ll create an internal table stored in ORC. The table is called “myhotels”:

Graphical user interface, text, email

Description automatically generated

Checking that “myhotels” table was created:

Graphical user interface, text, application

Description automatically generated

Now we will pass all the data from the external table “hotels” to the internal table “myhotels”:

Graphical user interface, text, application

Description automatically generated

Checking that the internal table “myhotels” is now populated with data:

Table

Description automatically generated

**STEP 3: CREATING DATAFRAME AND VIEWS ON SPARK**

ZEPPELIN

Now we’ll create a new data frame on Zeppelin (called “hotel\_Stars”) for a secondary table which contains the star listing for each hotel:

Graphical user interface

Description automatically generated with low confidence

Checking the schema of the new dataframe:



Checking that the data was loaded onto Zeppelin notebook:

Background pattern

Description automatically generated with low confidence

Creating a tempview for “hotel\_Stars” dataframe. The tempview will be named “starsView”



Checking that “starsView” is now retrievable:

Table

Description automatically generated

Checking we can retrieve the internal table created in Hive:

A picture containing graphical user interface

Description automatically generated

**STEP 4: MACRO LEVEL ANALYSIS USING %SPARK2.SQL INTERPRETER**

Provincial and City analysis

SQL QUERIES ON ZEPPELIN

MACRO LEVEL ANALYSIS - PROVINCIAL

Finding out which provinces already have 5-star hotels. We want to make sure there is an existing market for this target segment (we don’t want to build a hotel where there is no demand for this type of hotels):

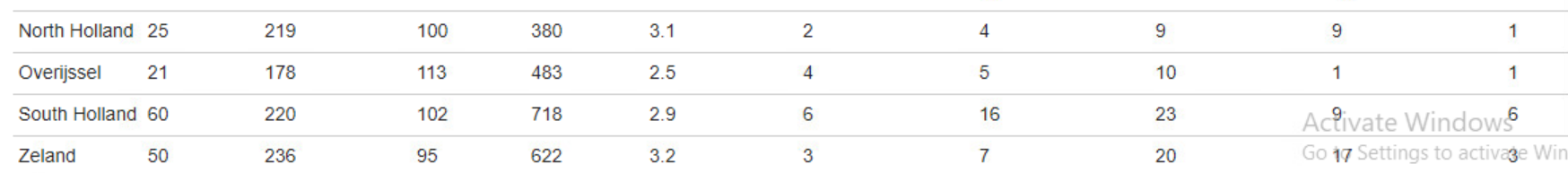
**Provincial metrics:**

Table

Description automatically generated with low confidence

Graphical user interface

Description automatically generated with low confidence



Quantity and distribution of hotel-star-rating per province:

Chart, bar chart

Description automatically generated

We notice there are 4 provinces with significant number of 5-star hotels:

Quantity of 5-star hotels at these preliminary provinces:

Amsterdam province – 8 hotels

Maastricht province – 13 hotels

Gueldres province – 6 hotels

South Holland province – 6 hotels

Average prices at these 4 preliminary provinces:

We can see that average prices are highest at Maastricht (503 CAD per night) and Amsterdam (328 CAD per night). Both of these provinces have the highest average rate.

On the other hand Gueldres has an average rate of 244 CAD per night and South Holland an average rate of 220 CAD, which is not very different from the average country price of 230 CAD per night, as shown in the query below:

National metrics:

Graphical user interface, text, application

Description automatically generated

We can also see there is an interesting secondary match for these provinces with acceptable amounts of 4-star hotels, suggesting there is in fact enough demand for high end hotels at these provinces:

Quantity of 4-star hotels at these preliminary provinces:

Amsterdam – 9 hotels

Maastricht – 5 hotels

Gueldres – 13 hotels

South Holland – 9 hotels

MACRO LEVEL ANALYSIS – CITY

Now we know that our potential cities to build in have to be within the provinces of Amsterdam, Maastricht, Gueldres and South Holland. Initially, it seems that Amsterdam and Maastricht seem to be a little more interesting with average night spending high above the national average rate, which is not the case for Gueldres and South Holland.

We want to make sure there are more than 3 five-star hotels present in the city we want to propose the investment, so we filter accordingly in the query below:

Graphical user interface, text, application

Description automatically generated

With this query we were able to filter down 5 city 5 options with more than 3 5-star hotels:

1. Amsterdam City Center (Amsterdam province)
2. Maastricht City Center (Maastricht province)
3. Wijck (Maastricht province)
4. Voorthuizen (Gueldres province)
5. Delft (South Holland province)

These are the average, min and max prices for all hotels in these cities:

Chart, bar chart

Description automatically generated

As our client, Hilton Hotels, is a high-end hotel, even for 5-star hotel standards, prices can range between 700-1000 CAD per night. Therefore, we want to select the two cities with highest average price in their 5 star-hotels. We must also discard the option of building in Maastricht City Centre because our client already has a hotel in that area, so we discard it in the ‘where’ clause as shown below:

Graphical user interface, text, application

Description automatically generated

Proportion of cities with most 5-star hotels:

Graphical user interface, application

Description automatically generated

5-star-hotels average price per city:

Chart, bar chart, histogram

Description automatically generated

As Amsterdam City Center and Wijck have the highest average and maximum prices of 5-star hotel rooms of the 4 cities, we will select these two cities to analyze in detail the hotels and prices available at these 2 selected cities.

At the same time, Hilton Hotels has a historical record of having 9.2 ratings from reviews (out of 10). Both in Amsterdam City Center and Wijck the average rating of reviews is lower than 8.7, thus possibly creating a favorable situation for Hilton Hotels.

**STEP 5: MICRO LEVEL ANALYSIS USING HBASE AND ZEPPELIN**

Hotel by hotel analysis – Competition analysis

HBASE

MICRO LEVEL ANALYSIS – HOTEL BY HOTEL IN THE 2 SELETED CITIES

Creating a “hotel” table in Hbase

A picture containing text

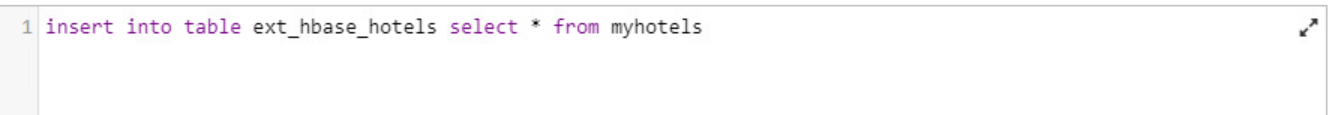
Description automatically generated

Create “ext\_hbase\_hotels” table in Hive that maps to HBase table “hotel”:

Graphical user interface, text, application, email

Description automatically generated

Inserting all data from the internal table myhotels o “ext\_hbase\_hotels”



The we scan ‘hotel’ to be sure data has been exported to hbase:

Text

Description automatically generated

Now we will fetch the details of the 7 five-star hotels in Amsterdam City Center and the 6 five-star hotels at Wijck.

Amsterdam City Center:

Text

Description automatically generated with medium confidence

Table

Description automatically generated

Table

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated with low confidence

Text

Description automatically generated with low confidence

Text

Description automatically generated

WIJCK CITY:

Text

Description automatically generated

Text, table

Description automatically generated with medium confidence

Graphical user interface, text, table

Description automatically generated

Text

Description automatically generated with medium confidence

We could also look at this information from Zeppelin:

Graphical user interface, text, application

Description automatically generated

Table

Description automatically generated with medium confidence

Price of each 5-star hotel in 2 groups: Amsterdam City Center and Wijck

Chart

Description automatically generated

Positive linear correlation between ratings and price. If we know that our average pricing is 9.2, this could help us fix the price per night for the new Hilton Hotel:

Application

Description automatically generated with medium confidence

**CONCLUSION**

The project was aimed to find the best locations for the construction of five-star hotels by Hilton group in Netherlands. The hospitality industry is dynamic in Netherlands as cities like Amsterdam, Maastricht retains the top spots as the most attractive European cities for hotel investments as it attracts foreign tourists travelling in Europe. Based on our analysis using hive and zeppelin, we found that Amsterdam City Center (Amsterdam province), Maastricht City Center (Maastricht province), Wijck (Maastricht province), Voorthuizen (Gueldres province), Delft (South Holland province) are the most favourable locations for constructing new five-star hotel as there already exists a market for high end hotels. But we ruled out the possibility of constructing a new hotel in Maastricht city center as Hilton group is already entered the market. From our further analysis based on average price and ratings using zeppelin, we found that the Amsterdam city centre and Wijck in Amsterdam province have higher average price and the hotels in these cities have average ratings lower than 8.7, thus, there is a high probability of attracting more people by providing them with exotic amenities and customer service. By analysis on h base we are able to find the major competitors of Hilton in these regions. The major competitors can be Andaz Amsterdam by Hyatt and Hotel V Nesplein in Amsterdam City Centre and Bourgogne suite in Wijck. Also we use zeppelin in our analysis of fixing the price of rooms per night in Hilton.