

TOURISM Hotel Booking

Rebecca Bou Zerdan





TABLE OF CONTENTS

01

Topic Overview

04

Analysis

02

Assumptions

05

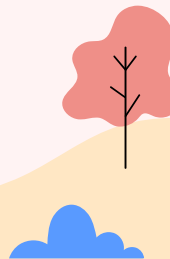
Prediction

03

Web Scrapping

06

Conclusion





01

TOPIC OVERVIEW

TOURISM SECTOR

It has been long known that tourism has been one of the major contributors in the Lebanese economy along the banking sector. Thus, with disintegration of the banking sector, how can we support our tourism sector to support the whole toll?



PROCESS

Web Scrapping

Using python, panda, and
Beautiful Soup.

1

Analysis

Using pyspark.

2

Prediction.

Using pyspark, including
VectorAssembler,
LinearRegression, and
Pipeline.

3





02

ASSUMPTIONS

ASSUMPTIONS

- Chosen City: Beirut
- Chosen Date: 5th August to 10th August, 2024
- Chosen Room: 1 room for 2 adults
- Chosen Pricing: all pricing included





03

WEB SCRAPING

Overview



Input

- URLs of all pages which include hotels found in the city of Beirut



Output

- Scraped data inserted in a csv file which is used for further analysis. This csv file included useful data related to every hotel in beirut.



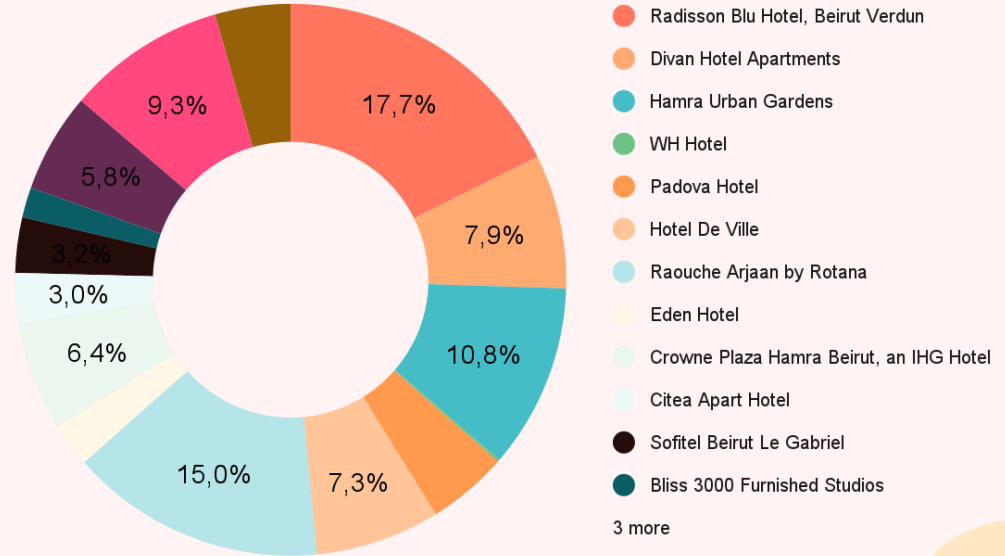
04

ANALYSIS

ANALYSIS

Since according to research around 40% of people who book an hotel room leave a review, I performed the needed calculation to estimate the number of bookings so far for every hotel. This would give me an idea about which hotels are most visited and conclude from its respective features, the reason for that.

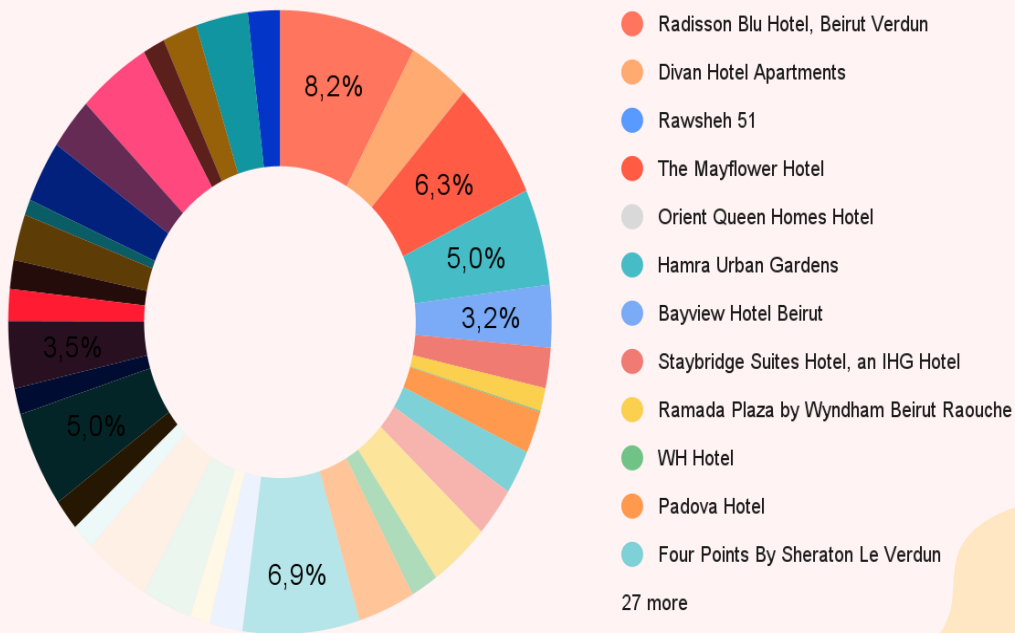
Number of Bookings per Hotel



ANALYSIS

The values scraped from Booking.com are restricted to the bookings done online. But to get an overview of the bookings in general in the city of Beirut which would give us insights of the tourism sector in beirut, I calculated the total number of bookings based. Based on research only 57% of hotel bookings are done online. (I disregarded the fact that there other booking websites since our main focus is on Booking.com)

Total Number of Bookings per Hotel



ANALYSIS

Nowadays, people have become more concerned with their impact on the society and world. Hence, when searching for hotels, a new factor was taken into consideration which is sustainability. The sustainability level is not sufficient to choose the best hotel, rating matters also since this represents people's point of views through on ground experience. A condition was created to categorize the hotels as Highly Recommended, Moderately Recommended, Recommended, Slightly Recommended, and Not Recommended.



```
graph TD; A[Eden Hotel] --> B[Slightly Recommended]
```

Eden Hotel

**Slightly
Recommended**

ANALYSIS

It is worth noting that people would like to have an idea about the location of the hotel they are booking and whether the respective hotel satisfies the purpose of the trip. Hence, metric cards should be used to display the name of the hotel along with its location with respect to the center of the city of Beirut.



```
graph TD; A[Eden Hotel] --> B[5.8 km from centre]
```

Eden Hotel

**5.8 km from
centre**

ANALYSIS

It is worth noting that people would like to have an idea about the location of the hotel they are booking and whether the respective hotel satisfies the purpose of the trip. Hence, metric cards should be used to display the name of the hotel along with its location with respect to the center of the city of Beirut. When the value is 0, this implies that no info is provided about it and hence assumed to be nonexistent and replaced with a value of zero.

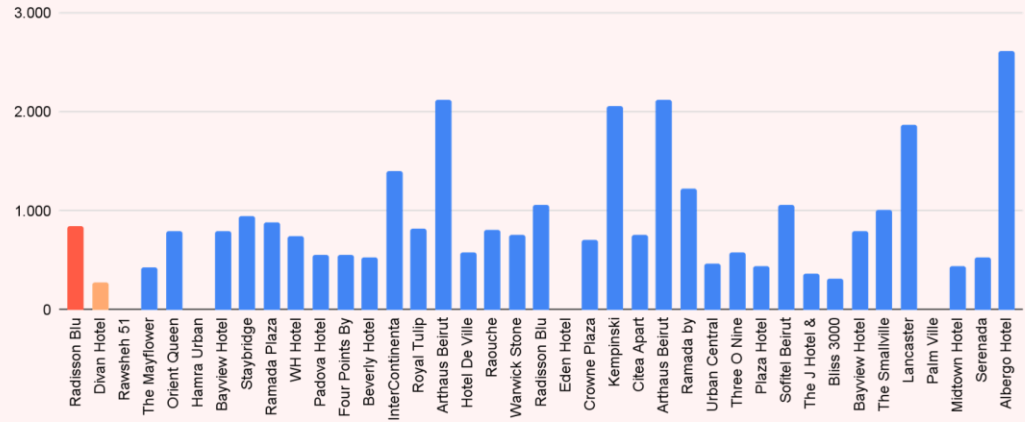


Eden Hotel

0

ANALYSIS

One of the major factors that tourists look at is costs and that's why they care about checking which is the best option for them.





05

PREDICTION

Flow of Prediction



Choose Columns Useful for Prediction

"number_reviews",
"total_bookings", "rating",
"price", "number_bookings",
"charges_converted",
"average_price"



Assemble them Together

Placed as part of features.



Apply Linear Regression

The outcome needed is the
expected_number_of_bookings

Flow of Prediction



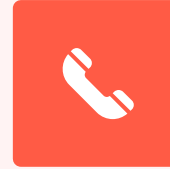
Create Pipeline

This includes the formation of the pipeline including the chosen columns for analysis and expected column.



Training

Input Data for training the linear regression model.



Testing

Test the model on the current data.



06

CONCLUSION



Make Tourism Flourish Again

Targets

- Where to invest?
- How to invest?
- What are tourists looking for?





Thank you !!